



Deep Learning for Discussion-Based Cross-Domain Performance Prediction of MOOC Learners Grouped by Language on FutureLearn

Ismail Duru¹ · Ayse Saliha Sunar² · Su White³ · Banu Diri⁴

Received: 26 June 2020 / Accepted: 5 November 2020 / Published online: 6 January 2021
© King Fahd University of Petroleum & Minerals 2021

Abstract

Analysing learners' behaviours in MOOCs has been used to identify predictive features associated with positive outcomes in engagement and learning success. Early methods predominantly analysed numerical features of behaviours such as the page views, video views, and assessment grades. Analysing extracted numeric features using baseline machine learning algorithms performed well to predict the learners' future performance in MOOCs. We propose categorising learners by likely English language proficiency and extending the range of data to include the content of comment texts. We compare results to a model trained with a combined set of extracted features. Not all platforms provide this rich variety of data. We analysed a series of a FutureLearn language focused MOOCs. Our data were from discussions embedded into each lesson's content. Analysing whether we gained any additional insights, over 420,000 comments were used to train the algorithm. We created a method for identifying one's possible first language from their country. We found that using comments alone is a weaker predictive approach than using a combination including extracted features from learners' activities. Our study contributes to research on generalisability of learning algorithms. We replicated the method across different MOOCs—the performance varies on the model though it always remained over 50%. One of the deep learning architecture, Bidirectional LSTM, trained with discussions on the language learning 73% successfully predicted learners' performance on a different MOOC.

Keywords MOOCs · Deep learning · English as a second language · FutureLearn · Predictive models · Natural language processing

1 Introduction

Large-scale datasets generated from MOOC participants have stimulated research to analyse and categorise learners' behaviours. Statistical methods have been used to investigate aggregated data associated with learners' behaviours (learning analytics). The objective has been to better understand the learner and thus gain insights which may potentially help optimise learning. Data that tracks learning activities, used as a basis for learning analytics, is one of the biggest and the most reliable information sources associated with online participants [1,2].

Learning analytics when applied to MOOCs mainly uses three distinctive sources: (a) click-stream data gathered as users click on screen as they progress on the platform; (b) outcomes of course assignments and quizzes—if this exists; and (c) social data from activities such as writing comments, peer interactions, and contributions on other social media tools where applicable.

✉ Ismail Duru
ismail.duru@izu.edu.tr

Ayşe Saliha Sunar
assunar@beu.edu.tr

Su White
saw@ecs.soton.ac.uk

Banu Diri
diri@yildiz.edu.tr

¹ Department of Software Engineering, Istanbul Sabahattin Zaim University, 34303 Istanbul, Turkey

² Department of Computer Engineering, Bitlis Eren University, 13000 Bitlis, Turkey

³ Department of Electronic and Computer Science, University of Southampton, Southampton SO17 1BJ, UK

⁴ Department of Computer Engineering, Yildiz Technical University, 34220 Istanbul, Turkey



Depending on what they want to analyse, and what is available, studies have selected different combinations of source data. Studies including [3–5] use sequences of click-stream data to predict participants' future performance; other studies use social interactions on the courses to model their participants' behaviours to have insight into their course experience [6,7].

Methods have been refined, in an attempt to increase the value of information, and researchers have begun to investigate their datasets using artificial intelligence approaches of machine learning. Machine learning has established its value for predicting likely outcomes in other domains. Cobos and Olmos [8] first extended the method to MOOCs, predicting likely future performance and course dropouts [9,10].

In a previous, proof of concept study, we exploited commonly used machine learning algorithms to predict performance of participants categorised into the different language groups according to their first language in a single language MOOC on FutureLearn [11].

Our approach was motivated by the assertion that understanding the behaviours of learners for whom English is a second language (ESL) may have value when improving MOOC design. ESL speakers typically comprise the largest proportion of MOOC participants in a course. Identifying the behaviours and needs specific to ESL speakers can potentially help the MOOC authors and the platform designers create more effective environments for a greater proportion of participants to get the best possible learning experience. But first, we needed a technique to easily identify whether a learners' first language is likely to be English, and their likely level of English language proficiency.

Our proof of concept study was limited to only one MOOC course. It predefined set of search terms (*regular expressions*) to search and identify participants' language from their posts. We identified numbers of limitations: (i) We could improve our method to identify a participants' first language to be more precise; (ii) our dataset was relatively small—we experimented with only one iteration of a MOOC (understanding Language MOOC analysing the data from over 25,000 participants who posted over 50,000 comments); (iii) the machine learning techniques we used required extensive feature engineering.

The first study identifies learners' behaviour on posting comments to discussions differing according to the language group of the learner. Perhaps unsurprisingly, the findings show that learners whose first language is not English tend to write shorter comments in English than those whose first language is English. Therefore, we need a more sophisticated method to investigate the comment data. An artificial intelligence-based approach taken from machine learning was the most promising candidate. We proposed that deep learning methods could be a possible solution to overcome the identified limitations of our original study applied the

method to data drawn from the series of MOOCs on a range of different subjects.

Deep learning is a machine learning technique which is a sub-field of artificial intelligence using neural networks, so called because they use mathematical models structured in a way inspired by the neural system of the brain. To be effective, deep learning requires massive data sets.

Deep learning is not a new method, but early use was limited because the computational process requires powerful computers. Advances in computer architecture in the early 2000s [12] has provided more widespread availability of computers capable of processing the large data required.

The widespread use of deep learning has also been a consequence of the availability of large volumes of data. Deep learning requires larger data sets than other machine learning techniques. Since MOOCs produce large data generated from thousand of participants, we took our inspiration from examples in the literature of deep learning applied to MOOC research for a variety of purposes (i.e. [13,14], see Sect. 2.1 for more example).

When applying machine learning to MOOC data, the initial data (extracted features, comment texts, and others), forms an *input layer* of the neural network to which we can apply the learning algorithm. *Hidden layers* are an interim where the algorithm learns from the data, and finally an *output layer* is created generating the final prediction (classification). When the number of hidden layers is many, the neural network is called as *deep learning* [15].

In this new study, we refined our method not only using the extracted numeric features from learning activities plus the comments' content. The FutureLearn MOOC platform is designed with a social-constructivist approach. It promotes engaging with the course through conversations [16].

FutureLearn MOOCs typically generate large amounts of conversational data that we have used as a primary source to feed the deep learning model. These rich conversation data have been already used by different researchers to gain insight into the learners' social behaviour on FutureLearn [6,17–19].

Our method of comparing the deep learning models developed with different architectures and multiple types of input sets allow us to more thoroughly evaluate the prediction performance of the model across the language-based groups.

We extended our approach to incorporate larger datasets aggregated from a number of MOOCs on different subjects in order to evaluate its accuracy and show the potential for generalisability amongst different MOOC domains.

We aimed to exploit deep learning models for performance prediction of second language English speakers by using their posts to discussions gathered from a range of MOOCs without performing feature extraction. Therefore, we developed three deep learning models which are slightly re-structured based on the input data: (i) only comments, (ii) only extracted

features, and (iii) comments and extracted features together. The highlights of the models we have applied are as following:

- We developed deep learning models by using a selection of different deep learning architectures: convolutional neural network (CNN), long short-term memory (LSTM), bidirectional LSTM, gated recurrent unit (GRU), convolutional neural network and long short-term memory (CNN_LSTM). We trained the models with data aggregated from the language focused MOOC and tested them on the last run of the language focused MOOC and on a set of other MOOCs from different educational domains (see Fig. 7).
- Before training the deep learning model with only comments, we pre-processed the data (removing the punctuation and making all words lowercase) and then used the GloVe library [20] to let the model recognise the patterns in the comments posted to discussions.
- We compared the performance of each deep learning model with comments only, extracted features only, and comments and extracted features together.
- We compared the performance of the deep learning models for the learners from different language groups: English as an Official and Primary Language (EPL), English as an Official but not a Primary Language (EOL), and English as a Second Language (ESL).

Studies in the existing literature (see Sect. 2) generally train and test their algorithms within the same set of MOOCs. Our research differs in that it uses not only iterations of the same MOOC but also a selection of MOOCs from different subject domains. This approach was adopted so that we might also investigate the generalisability of the model as well as its accuracy. While the literature provides examples for identifying ESL learners in MOOCs and analysis to understand their engagement with the course [21,22], there is, to the best of our knowledge, no study which incorporates this approach with using deep learning for predicting their future behaviours. We anticipate that our work will stimulate further widespread research with this focus.

We identified the contributions of this paper as follows:

- Identifying language status: We have created a reference source in table format for the languages and the probability of the status of English in the countries where a non-English language is spoken (see Tables 1, 2 and 3).
- Predicting course engagement: A deep learning model which is trained only with comments is not as strongly predictive for course engagement as a model trained with both extracted features and comments.
- Consistent model accuracy irrespective of commenting behaviour: The deep learning model trained with only

extracted features was also trained and tested for each language group. The accuracy of the model is almost same (84%) for each language-based group of learners, whereas their behaviour in commenting was quite different [11].

- Generalisability of our approach: The model with only comments was trained with the seven instances of a language MOOC delivered on FutureLearn. The trained model with only comments was then tested on MOOCs from different scientific domains for predicting learners' performance. The accuracy of the model always remained above 70% for each course.
- Identifying best performing algorithm: Among the deep learning models applied, the model implemented with the gated recurrent unit (GRU) architecture always performed slightly better than others in each experiment.
- Identifying strong predictors: Quantity of the comments is more strongly predictive than the content of the comments posted to the discussions.
- Establishing reliability of each subset of data: The algorithms performed very similarly for each language group, namely EPL, EOL, and ESL. However, transfer learning which indicates the success of the trained algorithm on a different set of data, is slightly worse on the data produced by the ESL learners.

2 Related Studies

2.1 Deep Learning in MOOCs

A study conducted in 2018 by Dalipi et al. [23] examined how machine learning techniques are used in the literature for MOOC dropout prediction. The authors find that most studies use engagement patterns and clickstream data as predictive factors in machine learning models. The authors also identify that basic algorithms such as logistic regression, support vector machine, and decision tree are used much more frequently compared to probabilistic models, neural networks, and natural language processing techniques.

MOOCs are attracted by the growing number of people over the years. With the COVID-19 lockdowns since mid-March 2020, there are 18 million newly registered MOOC users in August 2020, according to the report of Class Central [24]. As the growing number of MOOC users generates ever larger amounts of data from learners' activities, we can use deep learning techniques to identify patterns of learner behaviour and thus potentially more accurately predict future achievements of the learners within MOOCs.

Hernández-Blanco et al. [25] present a systematic review of how deep learning has been used in educational data mining in general. Their study and other related studies show that deep learning is already being used by MOOCs to respond to

the needs of learners during their involvement in the course and in evaluating the social and design functions of the courses [14].

Existing studies using machine learning for prediction differ according to the following criteria: (i) objective, (ii) data resources used, (iii) feature engineering, (iv) deep learning architecture, and (v) performance of the model and achievements.

Wang et al. [26] show that the use of a deep learning approach combining convolutional and recurrent neural networks can be effective in predicting course dropouts since it can free the analysis of the costly, and often difficult, overhead of applying prior knowledge to the domain being investigated. The authors do not perform manual feature engineering, but instead they automatically extract the features from the raw data such as time, browser source, event type, and so on.

Sun et al. [27] take a different approach to predict dropouts in MOOCs by using the URLs in a recurrent neural network (RNN). RNNs are compatible with time series, and the sequence of URLs is threatened in their study as time series. The sequence of the URLs is embedded in the embedding layer and predicts the dropout rate. The deep learning algorithm is more powerful than the baseline algorithms.

Similarly, Körösi and Farkas [4] aim to predict learner performance by applying deep learning without feature engineering. The authors apply an RNN model to the raw clickstream data identifying some actions such as video stop, page open, video forward, and so on. According to the results, their model performed better than the baseline algorithms.

Deep learning can also be used with extracted features. Xiong et al. [28] extract a number of features from learners' course activities to train a deep learning model constructed with RNNs (the long-short term memory architecture) to predict dropouts at a given week. The performance of the algorithm varied depending on the definition of dropout.

Xing and Du [13] adopt deep learning methods to assess individual dropout probabilities that help personalise and prioritise interventions for at-risk students. The authors perform a feature extraction from the clickstream and discussion data. They conclude that the use of deep learning is the most successful approach to predict the probability of dropout for each individual in any week and develop a temporal prediction model by using numerical features such as number of assignments and number of active days.

Deep learning is not always used to predict course completion. Tang and Pardos [29] use this approach to predict which course page a learner will navigate next. After a deep learning implementation with 13 edX MOOCs, their proposed model usually performs better than the commonly used machine learning models.

Yang et al. [46] use a combination of click-stream video-watching data and assessment grades for grade prediction. They show that no single behaviour is particularly correlated with performance, but the combination of factors is effective for predictive analysis.

These examples use data such as the time spent on a resource, the number of times a particular page is viewed, or whether learners ever use specific tools. Our research uses an additional data source. We are particularly interested in the potential use of comments posted by learners to classify behaviour and predict the likely success of learners in completing the course.

2.2 A Clue for the Future: Posts in Discussions

Discussion forums provide us with richer data than simple numeric tracking like counting page views and logging, time spent reading an article or watching a video. Forums contain more nuanced information such as participants' reflections, opinions, questions and emotions about the course and the content. Therefore, researchers are interested in using content from discussion posts to understand the attitude of participants, and to predict dropouts in order to diagnose those who need help [31].

Although deep learning approaches are not the miracle solution that beats all other methods for predicting at-risk students in MOOCs, they are well-suited to the analysis of loose data such as students' posts in discussions. In addition to traditional and modern machine learning techniques, various applications of deep learning were applied to text data from discussions.

Chaplot et al. [32], for example, perform sentiment analysis of forum posts to determine the week in which the student would dropout. The authors use the sentiment analysis alongside numerical data. Research to determine the emotions of MOOC participants through their contributions remains a research interest [33]. Chen et al. [34] use a method of deep learning for sentiment classification of forum posts, and their results show that the semi-supervised model of deep learning is more successful than the traditional and deep supervised models.

Wei et al. [35] use deep learning to classify forum posts that express either confusion or urgency using cross-domain MOOC data and obtain promising results for further improvements. Sun et al. [36] use another method of deep learning, the recurrent neural network, to identify urgent posts that require immediate attention. In a similar study, Hararak et al. [37] design an automatic annotation tool to mark each question in forum posts so that it can be analysed in relation to the course success of the participants.

Chanaa and El Faddouli [38] apply BERT, which is a deep learning algorithm for multiple tasks of natural language processing, to the posts in the discussion forums to identify the confused learners. The preliminary results are promising, so it could be used in a larger model to intervene in the learning process.

The previous studies we identify draw on a range of different technical approaches to investigate large sets of educational data—from virtual learning environments and from MOOCs. In some cases, these methods have been used to predict outcomes for learners. When the predictive methods are applied to discussions in particular, they usually aim to perform a mood analysis or to identify urgency and confusion.

Doleck et al. [39] compare the different frameworks and libraries of deep learning applying to two different educational dataset, which are a MOOC dataset and an academic performance dataset. The authors conclude that deep learning methods are not always superior to basic models, which is also reflected in our literature analysis. The authors also emphasise that a large and well-prepared dataset is important for the application of deep learning methods. Furthermore, researchers should focus on interpretability and explanation of accuracy as criteria for the selection of computational techniques for educational research for practical reasons.

Considering the example from the literature, we observed that for a number of targets, the techniques of deep learning were applied to the different data sources with or without the use of feature engineering. We find that there is not enough research on the application of deep learning to conversational data without feature engineering, focusing on the primary languages of learners. This would be particularly beneficial for MOOC platforms that include social learning and therefore produce a large amount of conversation data that would provide information about learners' performance.

Our study was designed to compare the performance of the proposed different deep learning models already tested in different circumstances with a single large set of FutureLearn MOOC data. In this paper we try to contribute to research on the prediction of MOOC performance:

- using the body of forum discussions instead of numerical data
- to compare different deep learning models trained with different sets of input data and different deep learning architectures
- assessing cross-domain transfer learning i.e. using the trained algorithm within one MOOC to test it on another MOOC

3 Implementation

In order to implement the research, there are two main tasks in this study:

1. Intelligently and accurately allocate the learners into a language group based on their primary languages. Participants are placed in one of three groups: Those for whom English is an Official and Primary Language (EPL), those for whom English is an Official but not a Primary Language (EOL), and those for whom English is a Second Language (ESL).
2. Building deep learning models and pre-processing the data for experiments to evaluate the proposed models across the MOOCs in different domains and between our defined different language groups.

The rest of the section explains in detail how we implement these two tasks to enable us to conduct meaningful experiments.

To identify which language group a participant belongs to, we use pre-course survey information and discussion forum content. If a participant posts a comment which includes information about their primary language, nationality, or country they live in, it is possible for us to identify this information from comment text by using the regular expressions techniques, as explained in the next section in greater detail.

3.1 Categorisation of Learners into Language-Based Groups

As it is emphasised in Sect. 1, identifying the participants who speaks English as a second language is important to provide them with any necessary assistance they may need. In this study, there are three categories of learner by language:

1. **English as an official and primary languages (EPL)** Where participants' first language is English or they live in a country where the frequently spoken language and the official language is English i.e. British people or those from the US.
2. **English as an official but not a primary language (EOL)** Where participants' first language is not English and/or they live in a country where English is official but not primary language. There are some countries where English is an official language but not a dominantly spoken local language. For example, in India, which attracts large numbers of MOOC participants, there are a large number of official languages including English. While people from some states in India speak fluent Indian English, participants from other states are far less likely to be fluent in English especially if the local government does not favour English language education [40].



3. **English as a second languages (ESL)** Where participants neither have English as a first language nor do they live in an English-speaking country they are categorised as ESL, e.g. German, Japanese, or Turkish.

Identifying participants' first languages through their comments is a challenging task for the computer. We currently have two means of identifying a participant's primary language: (i) the location information they provided during a pre-course survey and (ii) their statements in the discussion forums.

However, the participants do not necessarily identify their first language, instead they talk about their nationality, their home country, or where they live. We therefore first need a set of rules to teach the machine which defined category includes which language, nationality, and place information.

The language information in the comment itself gives us a clue about use of English and thus the countries in which the detected language might be spoken. Unfortunately, there is no single authoritative source we could use for this purpose, thus, we combined existing categorisations identified via Wikipedia to create our set of reference tables (Tables 1, 2, and 3). These tables show the languages and the probability of the status of English in the countries where a non-English language is spoken. To create these tables, the following steps have been carried out:

1. Identify a list of countries where English is an official language¹
2. Identify a list of countries and their official and minority languages²
3. Identify a list of the largest cities of countries³. We need this because participants may not write the name of their country but instead write which city they are from. In these cases, we have used the list to identify the country participants are from in order to identify their language group.
4. Identify a list of the nationalities, their countries and the languages spoken⁴. Some people refer to their nationality rather than their country. We have used this list to identify a participant's country and thereby identify their language.

The first table (Table 1) lists languages frequently spoken in countries where English is an official and/or primary lan-

Table 1 Languages frequently spoken in countries where English is an official and/or primary language (criteria at least 75%)

Spoken language	EPL—frequency (%)
Irish	100
Yoruba	100
Igbo	100
Manx	100

Table 2 Percentages of the most dominant EOL language group (at least 75%) for the languages different than English

Spoken language	EOL—frequency (%)
Hindi	100
Urdu	100
Tamil	100
Bangla	100
Sindhi	100
Afrikaans	100
Tagalog	100
Shona	100
Telugu	100
Marathi	100
Zulu	100
Burmese	100

guage. For example, if a participant stated their first language is Yoruba, we will categorise them as EPL participants.

Participants who use languages like Hindu are categorised in EOL since the language is predominant in locations where English is an official language although not the primary language spoken in the region. Table 2 shows the list of languages categorised in EOL.

Identifying participants who belong to the English as a second language group is rather complicated. Some languages are spoken across a number of different countries, but the status of the English language varies across those countries. For example, Spanish is an official language in more than 20 countries. We used Wikipedia's list of official languages by country as our reference to help build this list⁵.

Our approach here is that despite English not having an official status in those Spanish speaking countries. Nonetheless, English is effectively a secondary language in the 87% of the countries where Spanish is spoken. Therefore, if a participant says that Spanish is their first language, they are categorised in ESL because there is an 87% chance that they have ESL ability.

¹ https://en.wikipedia.org/wiki/List_of_territorial_entities_where_English_is_an_official_language.

² https://en.wikipedia.org/wiki/List_of_official_languages_by_country_and_territory.

³ https://en.wikipedia.org/wiki/List_of_largest_cities.

⁴ <http://www.vocabulary.cl/Basic/Nationalities.htm>.

⁵ https://en.wikipedia.org/wiki/List_of_countries_where_Spanish_is_an_official_language.

Table 3 Percentages of the most dominant ESL language group (threshold at least 75%) for the languages different than English

Spoken language	ESL—frequency
Portuguese	100
Spanish	87
Chinese	80
Indonesian	100
Valencian	100
Italian	86
Dutch	100
German	100
Greek	75
Korean	100
Persian	100
Galician	100
Kurdish	100
Bahasa	100
Tatar	100
Quechua	100

Another example is the Persian language. English is a secondary language in every country in which Persian is spoken. So, if a participant's first language is Persian, we assume 100% they are ESL participants. This is because English is a secondary language for at least 75% of a country's population; therefore, we have assumed that the participant belongs to the listed language group in the third table (Table 3).

Having established a fundamental approach to identifying language-based groups and created the reference source, we were ready to begin analysing the comment data. The rest of the section explains what constitutes our data and how we applied the regex method to identify language-based groups.

Table 4 shows the total number of participants, comments, participants' posted comments and survey replies in each iteration of understanding language MOOC delivered by the University of Southampton via FutureLearn. Only 5.87% (12,545) of participants provided their country information in the pre-survey (Table 4).

However, 34,166 (15.99%) participants provided comments which could potentially be analysed to identify their linguistic groups. The advantage of using this data lies in the potential to improve the accuracy of allocating participants to the different language groups and increase the volume of candidate comments.

Searching via regular expressions (regex) is a technique which has been used to tag the words in a sentence in a way that enables the researchers to match patterns in other sentences [41]. Regular expressions offer a route to powerful, flexible, and efficient text processing. We use expressions like

a general pattern notation allow us to describe and derive meaning from texts [42].

Regular expressions have been used to identify hashtags, keywords or certain words from texts of comments and transcripts. Various researchers have already analysed MOOC content using regular expressions. For example, Acosta and Otero [43] exploit regular expressions to automatically assess learners' answers to open questions in a MOOC. Shukla and Kakkar [44] use regular expressions to identify the noun chunks in a video transcript of a MOOC to extract keywords which may be helpful for assessing the content of video. An et al. [45] used the regular expressions with hashtags to link internal and external resources in MOOCs.

We use regular expressions to identify sets of words linked to individuals' primary languages. In our study, if a participant's language group identified in the MOOC's pre-survey is the same as the language group identified from the comments in discussion forums, then no changes occur. Otherwise, the allocated language group of the participant is updated according to the information identified from the comments.

To improve the accuracy of identification, we created an algorithm combining searching for regular expression patterns with a similarity metric to extract the information about participants' country, city, nationality and language from their comments in order to allocate their language groups whenever possible.

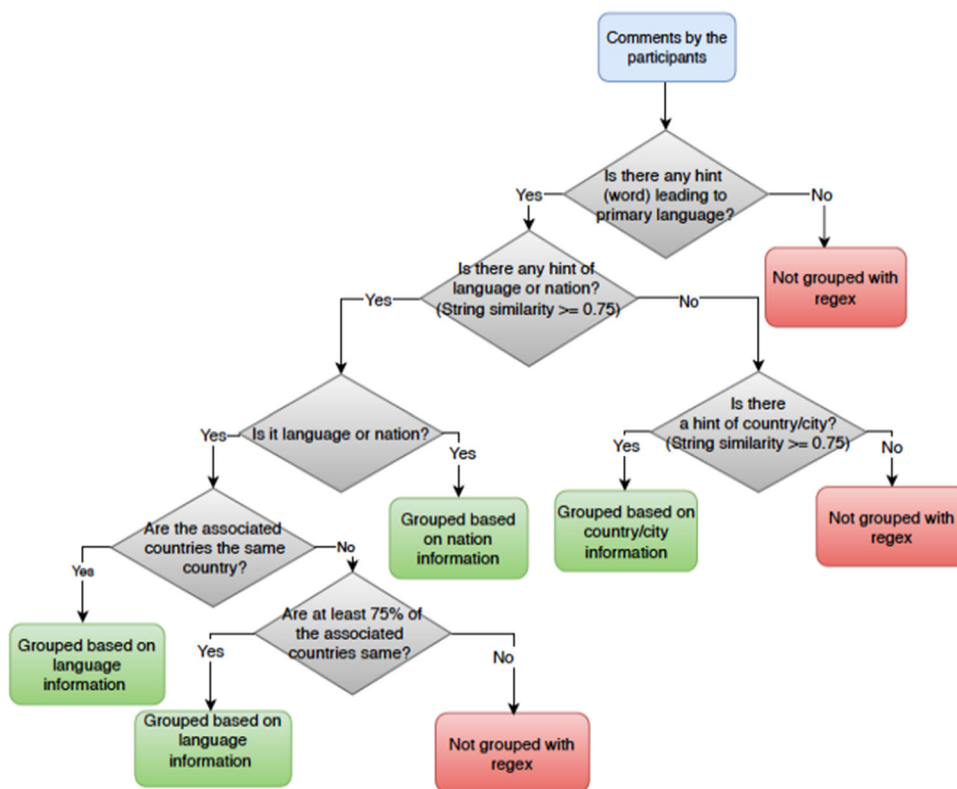
Figure 1 shows the flowchart of our method to identify language groups as using regular expressions and similarity metric. The algorithm checks all the written comments to extract information regarding their language, country/city, or nation. To eliminate the information loss during this process, the algorithm also checks the string similarity in case there are misspelled words. For example, if someone wrote *Trukish* instead of *Turkish*, the algorithm is able to identify these misspelled words as long as they are 75% similar to the original word. Also, when the extracted nation or language is one of the languages spoken in different countries based on the Tables 1, 2 and 3, the algorithm checks whether these countries are mostly (75%) the same country. Then, the algorithm checks the information with participants already grouped by pre-survey. If the participant has not already been classified during the pre-survey or if it is classified different from the result of the regular expressions method, the algorithm updates the participant's language group.

Across the set of MOOC runs examined (Understanding Language: Learning and Teaching from the first run to the seventh), 31,478 participants wrote 421,059 comments. With our method with regular expression, we detected 344,260 comments (81.8% of all comments) we could analyse and identified the language group of 21,621 participants (68.7% of all participants).

Figure 2 shows one of the regular expression patterns (C3) used to identify learners' country. For example, a com-

Table 4 Number of participants and comments in understanding language (UL) MOOCs

	Participants	Comments	Commented Participants	Participants replied to pre-surveys
UL 1	58,781	145,425	11,315 (19.25%)	1793 (3.05%)
UL 2	41,912	86,139	7585 (18.10%)	1516 (3.6%)
UL 3	44,293	58,285	5199 (11.74%)	1207 (2.7%)
UL 4	25,597	50,332	4164 (16.27%)	3305 (12.9%)
UL 5	19,866	37,637	2431 (12.24%)	2292 (11.5%)
UL 6	10,275	18,616	1319 (12.84%)	1019 (9.9%)
UL 7	12,893	24,941	2153 (16.70%)	1413 (10.96%)
Overall	213,617	421,375	34,166 (15.99%)	12,545 (5.87%)

Fig. 1 Flowchart of regular expressions (regex) model for language group identification

ment including a sentence ‘I came from northern Turkey.’ is matched with the pattern shown in Fig. 2.

Figure 3 shows one of the regular expression patterns (L3) used to identify learners’ language. For example, a comment including a sentence ‘English language is my first tongue.’ is matched with the pattern shown in Fig. 3.

Figure 4 shows one of the regular expression patterns (N1) used to identify learners’ nation. For example, a comment including a sentence ‘I am a typical Japanese.’ is matched with the pattern shown in Fig. 4.

Table 5 displays the percentages of language groups identified from pre-survey and regular expression patterns.

3.2 Building Deep Learning Models and Pre-processing Data

A deep learning model consists of structured hierarchical layers [15] computers learn complicated concepts by building them starting with the simpler layer.

We based our investigation of the MOOC data on methods which were successful in previous studies and compared the performance of each approach: convolutional neural network (CNN), long short-term memory (LSTM), bidirectional LSTM, gated recurrent unit (GRU), convolutional neural network and long short-term memory (CNN_LSTM). The rest of this section explains the logic and application of the selected models.

<p>Fig. 2 Sample regular expression pattern (C3) to identify country</p>	<pre>\\b(?:i we i s* s*a?m? i s*am im (?:(my s+name s+is s+ w+)) s+[^\\.]*?(?:c[oa]me s+)?from(?:s+here)?(?:s+the)?s+(?:east(?:ern)? north(?:ern)? west(?:ern)? south(?:ern)? s+)?(?:of s+)?(\\w{2,}(?!\\bmy))(?:s+republic)?</pre>
<p>Fig. 3 Sample regular expression pattern (L3) to identify language</p>	<pre>\\b(\\w{4,})\\s+(?:\\.(.*\\)\\s+)?(?:l[ae]nguage s+)?(?:is s+ as s+)?(?:also s+)?(?:ma?y s+)?(?:first 1st mother native of+icial main national primary best)[\\s\\]*(?:t[ou]{1,2}n[gq](?:[ue]{1,2})? l[ae]nguan?ge speaker)</pre>
<p>Fig. 4 Sample regular expression pattern (N1) to identify nation</p>	<pre>\\b(?:i s* s*m i s*am im we am)\\s+(?:a s+ an s+ are s+)?(?:typical s+)?(\\w{4,})\\b[\\.]*?(?!\\s+\\w+[eoir])</pre>

3.2.1 Pre-processing Data and Word Embedding

Some pre-processing of data was needed. This included removing URLs (since they are not actual words), converting all words into lower case to avoid duplication of the same word, and deleting posts shorter than 100 characters as they are not long enough to reveal usable information.

A text can be analysed letter by letter, word by word, or sentence by sentence. Processing data at the letter level increases the complexity, processing at sentence level can result in overlooking important data. In our study, we chose to process text data at the word level.

In addition to the pre-processed text data, some of the extracted features from our previous work [11] have been used in order to compare the performance of the algorithms. The extracted features used in the developed deep learning model are categorised in two areas, Step Activity and Comments, presented in Table 6.

In order to process the text after the pre-processing phase, the GloVe word-vector library has been used. The GloVe library represents over 400,000 words in English language with vectors in 300 dimensions. We have replaced the words in posts with these vectors defined in GloVe. As a consequence, we have an n-length vector for each post including n number of words.

The model is going to embed the words on the *embedding layer* to be trained on the hidden layers later. In this layer, the most frequently used 20,000 words in posts are stored as vectors and then converted to vectors. The reasons why we need this layer instead of giving the word-vectors directly to the neural network layer as input are:

- it is faster.
- less burden on GPU. As explained in Sect. 1, deep learning requires a powerful machine in terms of memory and speed. Therefore, it is crucial to use the memory efficiently for the machine performance.

When the posts are ready to be processed, the neural network layers process the data to produce the prediction result for the optimal class of a participant.

Figure 5 shows the general flowchart of the system including the data pre-processing, embedding layers and the specific deep learning architectures. The FutureLearn MOOC platform offers three types of data source: (i) text of comment data collected through discussion forums, (ii) follow data which indicates the Twitter-like friendships between the learners, and (iii) course activities such as opening a page, completing a task and so on.

After cleaning and pre-processing the data, our model does two things: (i) extracts features by learning analytics and (ii) transforms the comment into the vectors to be used in the embedding and deep learning structure as an input.

The model then applies a final training, concatenating the extracted features and the output of the deep learning training. Finally, the model produces the predictions as an output for learners’ performance in the course.

The rest of the section explains how different models are constructed and how they handle the data.

The deep learning models that we developed differ from each other in terms of

1. input data: comments only, extracted features only, and comments and extracted features together
2. used deep learning architecture in hidden layers: CNN, RNN or hybrid

We now explain the different architectures that are used in our study.

3.2.2 Convolutional Neural Networks (CNNs)

CNNs have hidden layers which can learn the internal structure of data and detect the patterns by applying filters. Although this approach to pattern recognition is particularly successful in image processing, it is also applied to natural language processing tasks to filter text with sequences of words/sentences [35]. CNNs have also been used for other textual purposes including sentiment analysis [46].

Figure 6 displays our developed deep learning model which is implemented with the convolutional layers and ‘comments and comments+extracted features’ as inputs. The

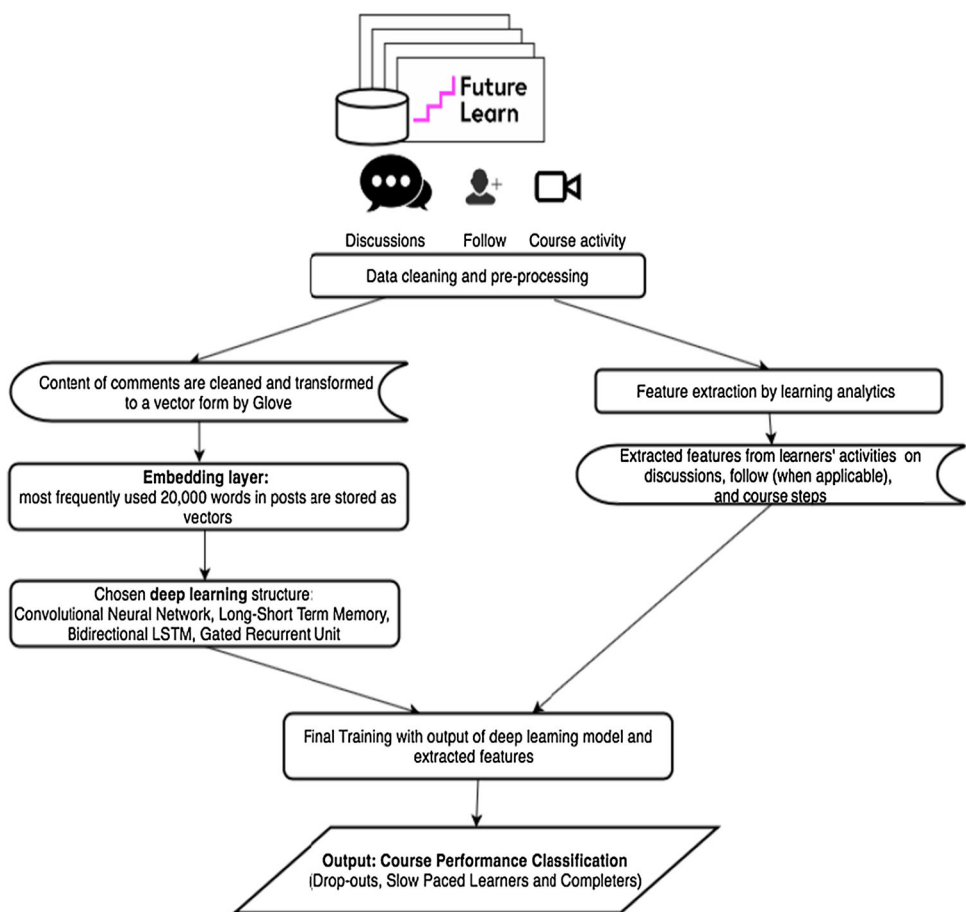
Table 5 Number of people whose language group identified by different methods

	Total people	EPL	EOL	ESL
People identified ONLY with pre-survey	9853	1760 (17.69%)	1281 (12.86%)	6908 (69.43%)
People identified ONLY with regex	18,928	3036 (16.04%)	1935 (10.22%)	13,957 (73.74%)
People identified and updated with combined method (pre-survey and regex)	2692	684 (25.00%)	254 (9.28%)	1798 (65.72%)

Table 6 Extracted features from the learners’ course activities

Step activity	Comments
1. Number of steps opened	1. Number of comments posted
2. Number of steps opened but not completed	2. Number of comments initiated
3. Total completed Week 1 steps	3. Number of replies
4. Total completed Week 2 steps	4. Average length of all comments
5. Total completed Week 3 steps	5. Average length of initiated comments
6. Total completed Week 4 steps	6. Average length of replies
	7. Average numbers of likes that a comment received

Fig. 5 Flowchart of the deep learning model with comments and extracted features for predicting learners’ performance



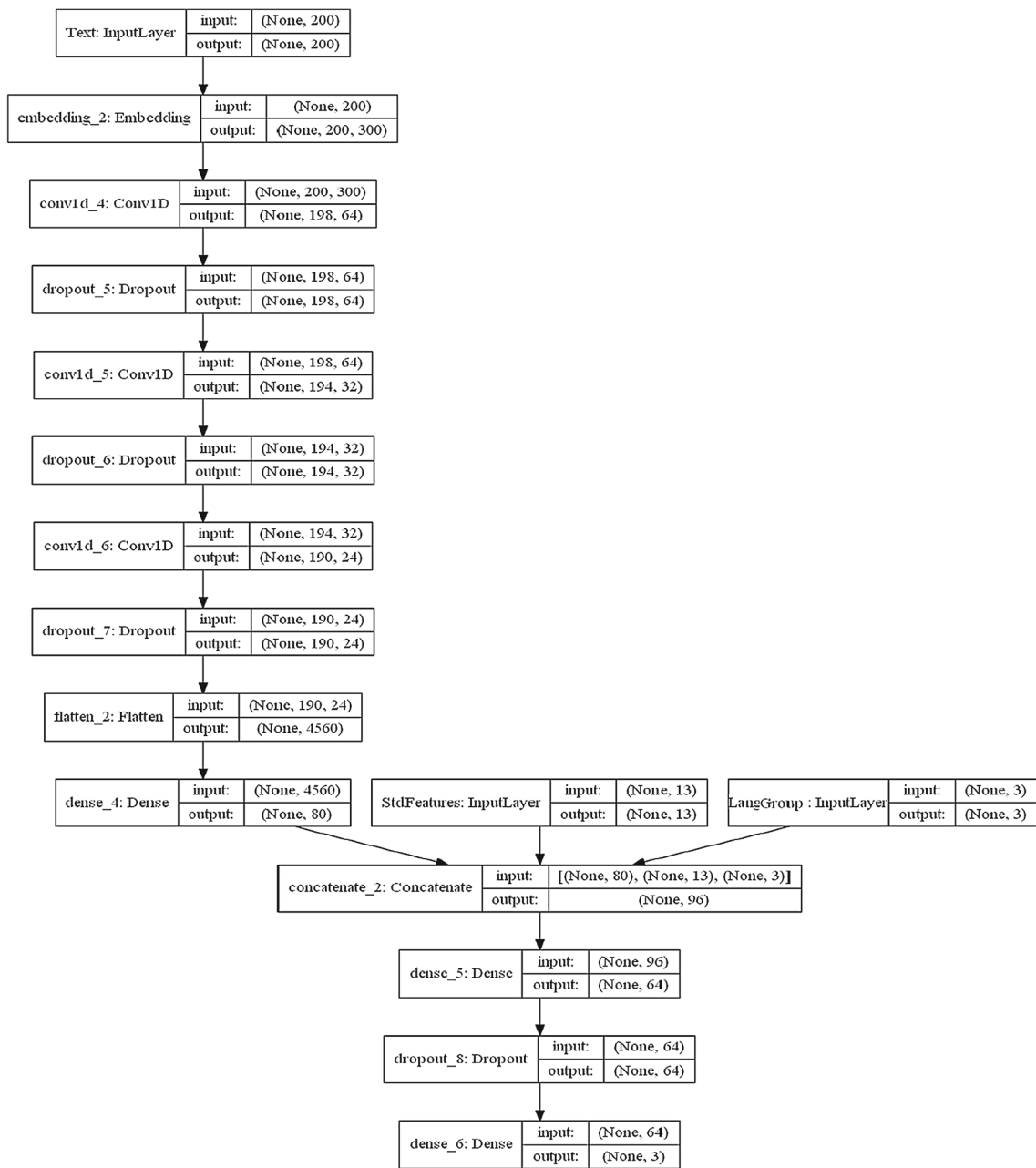


Fig. 6 The proposed deep learning model implemented with ‘comments and comments + extracted features’ as inputs and with convolutional layers as a deep learning architecture

models are various depending on the inputs and the implemented deep learning architecture.

In our model, the first layer consecutively runs to search three neighbour words, the second is to search five neighbour words, the third is to search seven neighbour words, respectively. After each layer, a *dropout* layer is run for better learning where the model drops some results that are unnecessary. The model then produces an output of training to be used as an input again for the classification in the output layer of the model.

3.2.3 Recurrent Neural Networks (RNNs)

In RNNs, the algorithm keeps the previous input in mind and consistently gives feedback after each run. Since this architecture enables the model to deal with sequences and lists, RNNs are convenient for modelling text [25], this is used in our research with posts (text body of comments) in discussion threads. There are different kinds of application of RNNs, we have used three of them explained below.

Long Short-Term Memory (LSTM) Using LSTM introduces memory cells which model long-range dependencies, especially useful for sequence learning problems involving textual analysis. In each cell, there are three gates: input, output and forget gates. According to the weights appointed to dependencies, the cell decides when and which information stored is going to be forgotten. Therefore, it carries the information from the past to the future for a longer time. It is different in this way from the convolutional layer as convolutional layer decides which information is going to be dropped certainly at the end of each run.

Bidirectional LSTM In addition to the traditional LSTMs, bidirectional LSTM carries the information between the past and the future forward and backward.

Gated Recurrent Unit (GRU) This is the simpler version of LSTM as it has only two gates instead of three gates.

3.2.4 Hybrid Models

Combining more than one deep learning architectures as hybrid models is also one of the methods for possibly more effective training and accurate results. The hybrid model used in this study is the combination of the convolutional neural network and the long short term memory models (CNN_LSTM).

4 Experiment Results

4.1 Experiment Design

Experiments have been carried out in order to compare the deep learning models in the cross-domain MOOCs.

We trained the models with the data generated from the first seven runs of the FutureLearn's language MOOC entitled Understanding Language (dated back from 2014 to 2018). Then, we tested the deep learning models on 20% of the first seven MOOCs, the eighth run of the same MOOC and further runs, on the same platform, of the Exploring Our Oceans MOOC (four weeks length), and the Web Science MOOC (six weeks length).

Figure 7 shows four different experiment scenarios. These four scenarios show the features and courses used in the training and testing phases. Models used in the scenarios vary according to the data used in training. In the first three experiments, comments and extracted features were used as input data. In the fourth experiment, only extracted features were used as input data. In all scenarios, data from the first seven runs of the Understanding Language: Learning and Teaching MOOC was used for training. Also, in the first three experiments, two deep learning model with comments only, and comments and extracted features together were implemented with different architectures (Convolution, LSTM,

Bidirectional LSTM, GRU and CNN+LSTM). However, in the fourth scenario, a deep learning model with only extracted features used for training were trained and tested.

Data was pre-processed before entering the models. For example, we have removed data of participants whose language group could not be identified and participants whose end-of-course performance information was unknown.

In the first experiment in Fig. 7, we used 80% training and 20% test data with a random selection from the commenting participants in the first seven versions of the Understanding Language: Learning and Teaching MOOC.

In the second scenario in Fig. 7, we used the data of the participants who commented in the 'Understanding Language 8: Learning and Teaching'. This course is the eighth iteration of the MOOC series used.

In the third scenario in Fig. 7, we used two MOOCs from different domains as test data from the MOOC series used in training. One of the courses used for the test is the 'Web Science-1', the MOOC for Science, and the second is the 'Exploring Our Oceans-4', the MOOC for natural sciences.

In the fourth scenario in Fig. 7, different from the first three scenarios, we split the data based on the language groups and only the extracted features were used in both training and testing. Three different test datasets were used. The first test dataset was obtained by allocating 20–80% of the same MOOC series. The second test dataset was created from different iterations of the same MOOC series. The third test dataset was created from different types of courses.

We observed the highest accuracy in performance prediction of learners when only extracted features are used. We first implemented models with data from all three English language groups in all scenarios.

According to the results we obtained, we found that the estimation made with the extracted features in the fourth scenario showed the best result. Therefore, unlike other scenarios, in the fourth scenario, in order to observe the change in the performance estimation success of the language groups, we created the data belonging to each language group EPL (English as Primary Language), EOL (English as an Official Language) and ESL (English as a second language) and performed performance estimation using extracted features of these courses separately.

4.2 Experiment I: Training and Testing Within the Same MOOC

In the first experiment in Table 7, we calculated the results of the performance predictions of learners on the first seven versions of 'Understanding Language: Learning and Teaching MOOC' series (80% train 20% test) with two separate deep learning models (comments only, comments and extracted features) and five different architectures (Convolution, LSTM, Bidirectional LSTM, GRU and 'CNN + LSTM'


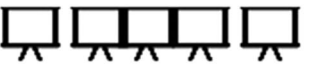




















	Train dataset	Test dataset	Features
Experiment I	 LMOOC 1st run LMOOC 7th run	 LMOOC 1st run LMOOC 7th run	 Comments  Comments +  Extracted features
Experiment II	 LMOOC 1st run LMOOC 7th run	 LMOOC 8th run	 Comments  Comments +  Extracted features
Experiment III	 LMOOC 1st run LMOOC 7th run	 Ocean MOOC Web Science MOOC	 Comments  Comments +  Extracted features
Experiment IV	 LMOOC 1st run LMOOC 7th run divided into the language groups as 	 LMOOC 1st run LMOOC 7th run  LMOOC 8th run  Ocean MOOC Web Science MOOC divided into the language groups as 	 Extracted features

Fig. 7 Design of different experimental settings

architectures using Convolution and LSTM). According to the results in Table 7, when 20% of the same course series was used as test data, the highest accuracy in learners’ performance estimation was obtained when only comments were used in the Convolution layer architecture.

In the learners’ performance estimation with ‘comments + extracted features’, we found that the highest learners’ performance estimation accuracy was achieved with GRU

architecture. In all tested architectures, learners’ performance prediction accuracy with ‘comment+extracted features’ was higher than learners’ performance estimation with ‘comments only’. From this result, we can say that analysing comments alone is insufficient to predict learners’ performance in the course.

Table 7 Performance prediction results with different deep learning models (Experiment I)

	Comments (%)	Extracted features and comments (%)
Convolution	68.05	82.95
LSTM	61.62	85.92
Bidirectional LSTM	60.76	84.62
GRU	60.49	87.05
CNN and LSTM	66.99	81.70

Bold values indicate the highest accuracy

Table 8 Results of performance prediction categorised by deep learning model (Experiment II)

	Comments (%)	Extracted features and comments (%)
Convolution	55.88	69.91
LSTM	51.36	76.92
Bidirectional LSTM	48.19	75.79
GRU	51.58	84.62
CNN and LSTM	52.71	72.62

Bold values indicate the highest accuracy

4.3 Experiment II: Transfer Learning 1.0: Training with a MOOC and Testing with Another Iteration of the Same MOOC

In the second experiment, Table 8 shows how we calculated the results of the learners' performance predictions in a MOOC with two separate deep learning models (comments only, comments and standard features) and five different tested architectures (Convolution, LSTM, Bidirectional LSTM, GRU and 'CNN + LSTM' architectures using Convolution and LSTM). The first seven versions of the 'Understanding Language: Learning and Teaching' MOOC series were used as training dataset. The eighth version of the course was used as test dataset. The results presented in Table 8, show that as in the first experiment, when the eighth version of the same course series was used as test data, the highest accuracy in learners' performance estimation was achieved using the convolution layer on the 'comments only' learners' performance prediction. In the learners' performance prediction with 'comment + standard features', we found that the highest learners' performance estimation accuracy was achieved with the GRU architecture. In all tested architectures, learners' performance estimation accuracy with 'comment + extracted features' was higher than learners' performance estimation with 'comments only'.

Table 9 Results of performance prediction categorised by Deep Learning model (Experiment III - Exploring Our Oceans 4)

	Comments (%)	Extracted features and comments (%)
Convolution	57.69	74.04
LSTM	66.35	74.04
Bidirectional LSTM	73.08	75.96
GRU	68.27	77.89
CNN and LSTM	63.46	71.15

Bold values indicate the highest accuracy

Table 10 Results of performance prediction categorised by Deep Learning model (Experiment III—Web Science 1)

	Comments (%)	Extracted features and comments (%)
Convolution	57.79	70.93
LSTM	58.82	74.39
Bidirectional LSTM	57.79	73.70
GRU	59.52	79.24
CNN and LSTM	55.02	71.28

Bold values indicate the highest accuracy

4.4 Experiment III: Transfer Learning 2.0: Training with a Series of MOOC and Testing with Another MOOC from a Different Domain

In the third experiment, learners' performance predictions results with two separate deep learning models (comments only, comments and extracted features) and five different tested architectures (convolution, LSTM, bidirectional LSTM, GRU and 'CNN + LSTM' architectures using convolution and LSTM) are given. The first seven versions of the 'Understanding Language: Learning and Teaching' MOOC series were used as a training dataset. 'Web Science-1' and 'Exploring Our Oceans-4' were used as test dataset. Results are shown in Tables 9 and 10. 'Exploring Our Oceans-4' MOOC, one of the MOOCs used for testing, is a four-week course just like the 'Understanding Language: Learning and Teaching' series MOOCs, while the 'Web Science-1' MOOC is a six-week MOOC. Due to this difference, we used the first four weeks of the 'Web Science-1' course.

The results presented in Table 9, show that when the 'Exploring Our Oceans-4' course was used as test data, the highest accuracy in learners' performance estimation was achieved with the bidirectional LSTM architecture using 'comments only'. In the learners' performance prediction with 'comment + extracted features', we found that the highest learners' performance estimation accuracy was achieved with the GRU architecture. In all tested architectures, learn-

ers' performance estimation accuracy with 'comment + extracted features' was higher than learners' performance estimation with 'comments only'.

In the results presented in Table 10, we see that when the 'Web Science-1' course was used as test data, the highest accuracy in learners performance estimation was obtained with GRU architecture for both 'comments only' and 'comment + extracted features' learners performance predictions. In all tested architectures, learners performance estimation accuracy with 'comment + extracted features' was higher than learners performance estimation with 'comments only'.

4.5 Experiment IV: Deep Learning Model with Extracted Features Applied to Different Language Groups by Primary Languages

In this case, the first seven editions of the Understanding Language: Learning and Teaching MOOC were used as training data. We trained the deep learning models with participants' extracted features. Table 11 shows the test results using the data of three different MOOC courses: i) the eighth version of the same Understanding Language MOOC, a science course Web Science-1, and an environmental course Exploring Our Oceans-4.

The results presented in Table 11 show that we can observe that the success of the learners' performance prediction does not show a significant decrease when the course data used for testing changes. We can also see that the accuracy of learners' performance estimation of different English language groups is very close to each other. It is observed that the performance prediction accuracy of the extracted features does not vary much according to the type of course and always gives good results.

5 Discussions

This study addressed two research areas: identifying a participant's language from their comments and predicting their future performance in the course based on learners' comments by using deep learning.

For identifying a participant's language from their comments, our proposed method uses regular expression pattern searching. We can clearly identify language-based groups as long as they have mentioned their primary language, city/country they are from, or their nationality during the conversations they had on the platform. However, considering the small proportion of participants joining the discussions, the chance that a participant gives at least one of these items of information is very low. Therefore, there is a need for another method to identify the participants' language-based groups no matter what they wrote about to the discussion board. For example, Mahmoud et al. [47] used natural language

Table 11 Results of performance prediction categorised by different language groups (Experiment IV)

	Overall (%)	EPL (%)	EOL (%)	ESL (%)
Training and Testing within the same MOOC	83.47	84.31	84.48	84.44
Training with a MOOC and Test with another iteration of the same MOOC	83.78	84.72	84.84	84.55
Training with a MOOC and Test with another MOOC on a different domain-1 (exploring our oceans-4)	83.78	84.02	84.84	84.55
Training with a MOOC and Test with another MOOC on a different domain-2 (web science-1)	83.51	84.72	84.04	83.60

processing techniques to identify how similar two written documents, which can be modified to group the comments.

Different deep learning models have been developed to compare performance across a number of MOOCs from different domains, either by solely using comment texts or taking comment texts together with the extracted features. Even though working with raw data, i.e. comment texts, lessens the workload of feature engineering, the difference between the performance of models sometimes reached over 20%.

We expected that the content of comments with or without combining with other features would be valuable to better understand and predict the users' future performance on MOOCs. The results show that using solely the content of the comment is not strongly predictive compared to the results from quantitative features extracted from learning activities. However, the results show that combining both is more promising for prediction, which can be investigated and improved by further studies.

In our previous studies, we were only able to train and test the algorithms on one case of the Understanding Language MOOC delivered on FutureLearn by University of Southampton. Since transferring learning is a crucial task in machine learning and potentially important in to enable generalised application in distance learning, we tested the developed models on MOOCs from different domains. It was

observed that even though the accuracy of the model fell by around 10%, it is still above 70% for each model, which is a very promising result.

However, it needs to be noted that the FutureLearn MOOC platform is designed according to the social-constructivist learning theory [16]. Therefore the design of MOOCs on FutureLearn facilitates posting comments and interacting with other fellow learners. Another MOOC platform which only provides Q&A forums may not generate sufficient information about users from their comments so in that context the accuracy of the model might be much lower.

6 Conclusions

In this study, we aimed to exploit deep learning models for performance prediction of different of second language English speakers categorised by their likely linguistic skills. We based our predictions, without performing feature extraction, on the content their posts to discussions gathered from a range of MOOCs.

Our research categorises learners into three language groups: those who have English as an Official and Primary Language (EPL), those for whom English is an Official but not a Primary Language (EOL), and those for whom English is a Second Language (ESL).

Using regular expressions is a handy approach to recognise the words and phrases associated with language, nationality, and country, which helping us to identify primary languages. Our research showed that deciding the primary language is a challenging task even though we have information regarding country or nationality. To overcome this difficulty we created a reference list for the languages and the probability of the status of English in the countries where a non-English language is spoken. In addition, we defined regular expression rules for identifying certain sentences giving a hint to the languages.

Having identified the language-based groups, the next task in the research was building deep learning models and testing them. We built a deep learning model by using solely comment texts. For the comparison, we also built a deep learning model by using a combination of comment texts and extracted features from user activities during the course.

We implemented five different deep learning architectures in the developed deep learning models: convolutional neural network (CNN), long short-term memory (LSTM), bidirectional LSTM, gated recurrent unit (GRU), and a hybrid of convolutional neural network and long short-term memory (CNN_LSTM).

According to the results, the deep learning models we implemented performed better with the combination of comment texts and extracted features than if they were constructed based solely on comment texts. When we tested

the models using MOOCs from different domains, the deep learning with extracted features-only performed better.

However, some of the deep learning models with the combination of comment texts and extracted features performed better than the extracted features-only models when training and testing within the same series of a MOOC or testing on a different iteration. In addition, the trained algorithms performed reasonably well on MOOCs from different domains which shows the generalisibility of our model.

Acknowledgements The authors would like to thank M. Emir Ozcevik, who is a former undergraduate student in Yildiz Technical University (YTU), for his technical help in the process analysis. We also would like to thank Assoc. Prof. Gulustan Dogan and Gozde Merve Demirci for their help. This work has been done under the project issued 01/04/2016 DOP05 in YTU, the TUBITAK 2214-A project issued 1059B141601346 and TUBITAK 2228-B scholarship issued 1059B281304197. The dataset used in this paper is provided by the University of Southampton for the ethically approved collaborative study (ID: 23593).

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Khalil, M., Taraghi, B., Ebner, M.: Engaging learning analytics in MOOCs: the good, the bad, and the ugly. In: International Conference on Education and New Developments 2016 (2016)
2. Muslim, A.; Chatti, M.A.; Guesmi, M.: Open learning analytics: a systematic literature review and future perspectives. In: Artificial Intelligence Supported Educational Technologies, pp. 3–29. Springer (2020)
3. Yu, C.H.; Wu, J.; Liu, A.C.: Predicting learning outcomes with MOOC clickstreams. *Educ. Sci.* **9**(2), 104 (2019)
4. Kőrösi, G.; Farkas, R.: MOOC performance prediction by deep learning from raw clickstream data. In: International Conference on Advances in Computing and Data Sciences, pp. 474–485. Springer (2020)
5. Lemay, D.J.; Doleck, T.: Grade prediction of weekly assignments in MOOCs: mining video-viewing behavior. *Educ. Inf. Technol.* **25**(2), 1333–1342 (2020)
6. Sunar, A.S.; Abbasi, R.A.; Davis, H.C.; White, S.; Aljohani, N.R.: Modelling MOOC learners' social behaviours. *Comput. Hum. Behav.* (2018)
7. Wang, W.; Guo, L.; He, L.; Wu, Y.J.: Effects of social-interactive engagement on the dropout ratio in online learning: insights from MOOC. *Behav. Inf. Technol.* **38**(6), 621–636 (2019)
8. Cobos, R.; Olmos, L.: A learning analytics tool for predictive modeling of dropout and certificate acquisition on MOOCs for professional learning. In: 2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), pp. 1533–1537. IEEE (2018)
9. Moreno-Marcos, P.M.; Muñoz-Merino, P.J.; Maldonado-Mahauad, J.; Pérez-Sanagustín, M.; Alario-Hoyos, C.; Kloos, C.D.: Temporal analysis for dropout prediction using self-regulated learning strategies in self-paced MOOCs. *Comput. Educ.* **145**, 103728 (2020)



10. Chen, C.; Sonnert, G.; Sadler, P.M.; Sasselov, D.D.; Fredericks, C.; Malan, D.J.: Going over the cliff: MOOC dropout behavior at chapter transition. *Distance Educ.* **41**(1), 6–25 (2020)
11. Duru, I.; Sunar, A.S.; White, S.; Diri, B.; Dogan, G.: A case study on English as a second language speakers for sustainable MOOC study. *Sustainability* **11**(10), 2808 (2019)
12. Schmidhuber, J.: Deep learning in neural networks: an overview. *Neural Netw.* **61**, 85–117 (2015)
13. Xing, W.; Du, D.: Dropout prediction in MOOCs: using deep learning for personalized intervention. *J. Educ. Comput. Res.* **57**(3), 547–570 (2019)
14. Aljohani, T.; Cristea, A.I.: Predicting learners' demographics characteristics: deep learning ensemble architecture for learners' characteristics prediction in MOOCs. In: *Proceedings of the 2019 4th International Conference on Information and Education Innovations*, pp. 23–27. ACM (2019)
15. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y.: *Deep Learning*, vol. 1. MIT Press, Cambridge (2016)
16. Ferguson, R.; Clow, D.; Beale, R.; Cooper, A.J.; Morris, N.; Bayne, S.; Woodgate, A.: Moving through MOOCs: pedagogy, learning design and patterns of engagement. In: *Design for Teaching and Learning in a Networked World*, pp. 70–84. Springer (2015)
17. Tubman, P.; Oztok, M.; Benachour, P.: Being social or social learning: a sociocultural analysis of the Futurelearn MOOC platform. In: *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*, pp. 1–2. IEEE (2016)
18. Shi, L.; Cristea, A.I.; Toda, A.M.; Oliveira, W.: Social engagement versus learning engagement an exploratory study of Futurelearn learners. In: *2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, pp. 476–483. IEEE (2019)
19. O'Riordan, T.; Millard, D.E.; Schulz, J.: Is critical thinking happening? Testing content analysis schemes applied to MOOC discussion forums. *Comput. Appl. Eng. Educ.* (2020)
20. Pennington, J.; Socher, R.; Manning, C.D.: Glove: global vectors for word representation. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543 (2014)
21. Uchidiuno, J.; Ogan, A.; Yarzabinski, E.; Hammer, J.: Understanding ESL students' motivations to increase MOOC accessibility. In: *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*, pp. 169–172 (2016)
22. Reilly, E.D.; Williams, K.M.; Stafford, R.E.; Corliss, S.B.; Walkow, J.C.; Kidwell, D.K.: Global times call for global measures: investigating automated essay scoring in linguistically-diverse MOOCs. *Online Learn.* **20**(2), 217–229 (2016)
23. Dalipi, F.; Imran, A.S.; Kastrati, Z.: MOOC dropout prediction using machine learning techniques: Review and research challenges. In: *2018 IEEE Global Engineering Education Conference (EDUCON)*, pp. 1007–1014. IEEE (2018)
24. Shah, D.: By the numbers: MOOCs during the pandemic. *Class Central* (2020). <https://www.classcentral.com/report/mooc-stats-pandemic/>. Accessed 18 Sept 2020
25. Hernández-Blanco, A.; Herrera-Flores, B.; Tomás, D.; Navarro-Colorado, B.: A systematic review of deep learning approaches to educational data mining. *Complexity* **2019** (2019)
26. Wang, W.; Yu, H.; Miao, C.: Deep model for dropout prediction in MOOCs. In: *Proceedings of the 2nd International Conference on Crowd Science and Engineering*, pp. 26–32. ACM (2017)
27. Sun, D.; Mao, Y.; Du, J.; Xu, P.; Zheng, Q.; Sun, H.: Deep learning for dropout prediction in MOOCs. In: *2019 Eighth International Conference on Educational Innovation through Technology (EITT)*, pp. 87–90. IEEE (2019)
28. Xiong, F.; Zou, K.; Liu, Z.; Wang, H.: Predicting learning status in MOOCs using LSTM. In: *Proceedings of the ACM Turing Celebration Conference-China*, p. 74. ACM (2019)
29. Tang, S.; Pardos, Z.A.: Personalized behavior recommendation: a case study of applicability to 13 courses on edX. In: *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, pp. 165–170. ACM (2017)
30. Yang, T.Y.; Brinton, C.G.; Joe-Wong, C.; Chiang, M.: Behavior-based grade prediction for MOOCs via time series neural networks. *IEEE J. Sel. Top. Signal Process.* **11**(5), 716–728 (2017)
31. Almatrafi, O.; Johri, A.; Rangwala, H.: Needle in a haystack: identifying learner posts that require urgent response in MOOC discussion forums. *Comput. Educ.* **118**, 1–9 (2018)
32. Chaplot, D.S.; Rhim, E.; Kim, J.: Predicting student attrition in MOOCs using sentiment analysis and neural networks. *AIED Workshops* **53**, 54–57 (2015)
33. Xing, W.; Tang, H.; Pei, B.: Beyond positive and negative emotions: looking into the role of achievement emotions in discussion forums of MOOCs. *Internet High. Educ.* **43**, 100690 (2019)
34. Chen, J.; Feng, J.; Sun, X.; Liu, Y.: Co-training semi-supervised deep learning for sentiment classification of MOOC forum posts. *Symmetry* **12**(1), 8 (2020)
35. Wei, X.; Lin, H.; Yang, L.; Yu, Y.: A convolution-LSTM-based deep neural network for cross-domain MOOC forum post classification. *Information* **8**(3), 92 (2017)
36. Sun, X.; Guo, S.; Gao, Y.; Zhang, J.; Xiao, X.; Feng, J.: Identification of urgent posts in MOOC discussion forums using an improved RCNN. In: *2019 IEEE World Conference on Engineering Education (EDUNINE)*, pp. 1–5. IEEE (2019)
37. Harrak, F.; Luengo, V.; Bouchet, F.; Bachelet, R.: Towards improving students' forum posts categorization in MOOCs and impact on performance prediction. In: *Proceedings of the Sixth (2019) ACM Conference on Learning@ Scale*, pp. 1–4 (2019)
38. Chanaa, A.; El Faddouli, N.E.: BERT and prerequisite based ontology for predicting learner's confusion in MOOCs discussion forums. In: *International Conference on Artificial Intelligence in Education*, pp. 54–58. Springer (2020)
39. Doleck, T.; Lemay, D.J.; Basnet, R.B.; Bazalais, P.: Predictive analytics in education: a comparison of deep learning frameworks. *Educ. Inf. Technol.* **25**(3), 1951–1963 (2020)
40. Thiyagarajan, K.: Higher education and practice of English in India. *Lang. India* **8**, 8 (2008)
41. Bird, S.; Klein, E.: *Regular Expressions for Natural Language Processing*. University of Pennsylvania, Philadelphia (2006)
42. Friedl, J.E.: *Mastering Regular Expressions*. O'Reilly Media, Inc., Newton (2002)
43. Acosta, E.S.; Otero, J.J.E.: Automated assessment of free text questions for MOOC using regular expressions. *Inf. Resour. Manag. J. IRMJ* **27**(2), 1–13 (2014)
44. Shukla, H.; Kakkar, M.: Keyword extraction from educational video transcripts using NLP techniques. In: *2016 6th International Conference-Cloud System and Big Data Engineering (Confluence)*, pp. 105–108. IEEE (2016)
45. An, Y.H.; Chandrasekaran, M.K.; Kan, M.Y.; Fu, Y.: The MUIR Framework: Cross-linking MOOC resources to enhance discussion forums. In: *International Conference on Theory and Practice of Digital Libraries*, pp. 208–219. Springer (2018)
46. Rani, S.; Kumar, P.: Deep learning based sentiment analysis using convolution neural network. *Arab. J. Sci. Eng.* **44**(4), 3305–3314 (2019)
47. Mahmoud, A.; Zrigui, M.: Sentence embedding and convolutional neural network for semantic textual similarity detection in Arabic language. *Arab. J. Sci. Eng.* **44**(11), 9263–9274 (2019)

