

Received May 23, 2020, accepted June 7, 2020, date of publication June 16, 2020, date of current version January 11, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3002791

Academic Performance Prediction Based on Multisource, Multifeature Behavioral Data

LIANG ZHAO¹, KUN CHEN¹, JIE SONG¹, XIAOLIANG ZHU¹, JIANWEN SUN¹, BRIAN CAULFIELD², (Member, IEEE), AND BRIAN MAC NAMEE²

¹National Engineering Laboratory for Educational Big Data, National Engineering Research Center for E-learning, Central China Normal University, Wuhan 430079, China

²The Insight Center for Data Analytics, University College Dublin, Dublin 4, Ireland

Corresponding author: Liang Zhao (liang.zhao@mail.cnu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB1401300 and Grant 2017YFB1401303, in part by the National Natural Science Foundation of China under Grant 61977030, and in part by the Fundamental Research Funds for the Central University under Grant 20205170443.

ABSTRACT Digital data trails from disparate sources covering different aspects of student life are stored daily in most modern university campuses. However, it remains challenging to (i) combine these data to obtain a holistic view of a student, (ii) use these data to accurately predict academic performance, and (iii) use such predictions to promote positive student engagement with the university. To initially alleviate this problem, in this article, a model named Augmented Education (AugmentED) is proposed. In our study, (1) first, an experiment is conducted based on a real-world campus dataset of college students ($N = 156$) that aggregates multisource behavioral data covering not only online and offline learning but also behaviors inside and outside of the classroom. Specifically, to gain in-depth insight into the features leading to excellent or poor performance, metrics measuring the linear and nonlinear behavioral changes (e.g., regularity and stability) of campus lifestyles are estimated; furthermore, features representing dynamic changes in temporal lifestyle patterns are extracted by the means of long short-term memory (LSTM). (2) Second, machine learning-based classification algorithms are developed to predict academic performance. (3) Finally, visualized feedback enabling students (especially at-risk students) to potentially optimize their interactions with the university and achieve a study-life balance is designed. The experiments show that the AugmentED model can predict students' academic performance with high accuracy.

INDEX TERMS Academic performance prediction, behavioral pattern, digital campus, machine learning (ML), long short-term memory (LSTM).

I. INTRODUCTION

As an important step to achieving personalized education, academic performance prediction is a key issue in the education data mining field. It has been extensively demonstrated that academic performance can be profoundly affected by the following factors:

- **Students' Personality** (e.g., neuroticism, extraversion, and agreeableness) [1]–[4];
- **Personal Status** (e.g., gender, age, height, weight, physical fitness, cardiorespiratory fitness, aerobic fitness, stress, mood, mental health, intelligence, and executive functions) [1]–[12];

The associate editor coordinating the review of this manuscript and approving it for publication was Shadi Alawneh¹.

- **Lifestyle Behaviors** (e.g., eating, physical activity, sleep patterns, social tie, and time management) [7]–[28]; and
- **Learning Behaviors** (e.g., class attendance, study duration, library entry, and online learning) ([7], [8], [23]–[26],[28]–[38]).

For example, [2] investigated the incremental validity of the Big Five personality traits in predicting college GPA. [21] demonstrated that physical fitness in boys and obesity status in girls could be important factors related to academic achievement. Meanwhile, [22] showed that a regular lifestyle could lead to good performance among college students. [24] showed that the degree of effort exerted while working could be strongly correlated with academic performance. Additionally, [32] showed that compared with high- and medium-achieving students, low-achieving students were

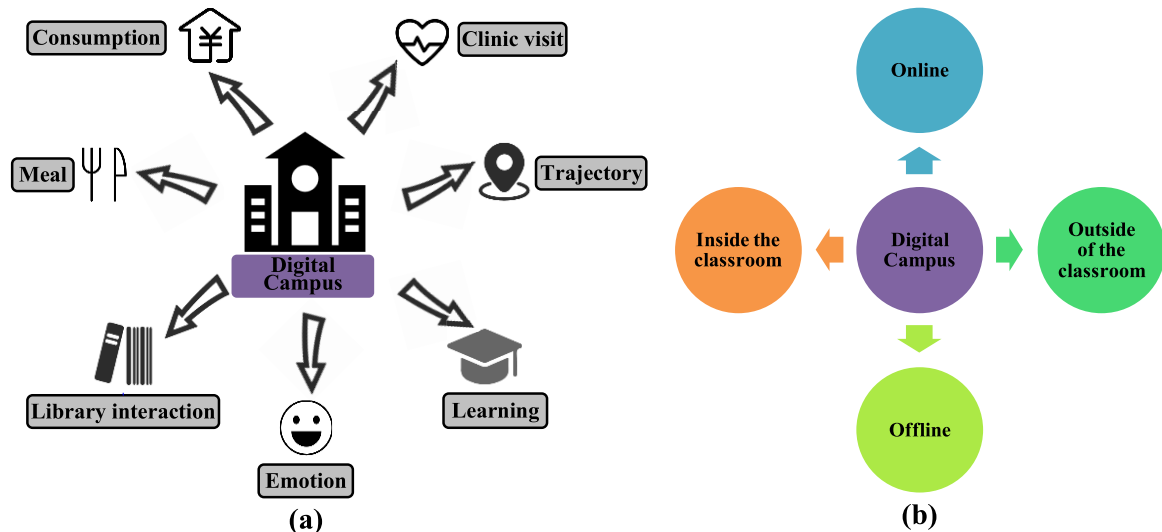


FIGURE 1. Digital data remaining on a modern campus: (a) Multisource; (b) Multispace, covering not only online and offline learning but also students' behaviors inside and outside of the classrooms.

less emotionally engaged throughout the semester and tended to express more confusions during the final stage of the semester.

By analyzing the effect of the factors influencing academic performance, many systems using data to predict academic performance have been developed in the literature [1]–[4], [7], [8], [13]–[19], [22]–[27], [29]–[31], [33], [34], [37]–[41]. For instance, in [8], academic performance was predicted based on passive sensing data and self-reports from students' smart phones. In [23], a multitask predictive framework that captures intersemester and intermajor correlations and integrates student similarity was built to predict students' academic performance. In [34], based on homework submission data, the academic performance of students enrolled in a blended learning course was predicted.

According to their predicted academic performance, early feedbacks and interventions could be individually applied to at-risk students. For example, in [33], to help students with a low GPA, basic interventions are defined based on GPA predictions. However, the research on the feedback/intervention is still in the early stage, its achievements are relatively few.

In recent years, compared with primary and secondary education (i.e. K12) [6], [10], [12], [17], more and more attentions have been paid to the academic performance prediction for higher education [7]–[9], [14], [15], [22]–[25], [27], [28], [30]–[32], [36]–[38]. The reasons contributing to this phenomenon warrant further investigation and might include the following. First, for college students on a modern campus, life involves a combination of studying, eating, exercising, socializing, etc. (see Fig. 1) [7], [8], [22]–[25], [27], [42], [43]. All activities that students engage in (e.g., borrowing a book from the library) leave a digital trail in some database. Therefore, it is relatively easy to track college students' behaviors, e.g. online learning behaviors captured from massive open online

courses (MOOC) and small private online courses (SPOC) platforms [30]–[32], [36]–[38]. Second, given the diverse range of activities listed above, it could be difficult for college students to maintain balanced, self-discipline, well-being university experiences, including excellent academic performance.

Although many academic performance prediction systems have been developed for college students, the following challenges persist: (i) capturing a sufficiently rich profile of a student and integrating these data to obtain a holistic view; (ii) exploring the factors affecting students' academic performance and using this information to develop a robust prediction model with high accuracy; and (iii) taking advantage of the prediction model to deliver personalized services that potentially enable students to drive behavioral change and optimize their study-life balance.

To address these challenges, four representative prediction systems (including one online system and three offline systems) are summarized in Table 1. We first discuss the online prediction system, System A [32] (proposed by Z. Liu). This system is relatively simple because its data is only captured from either SPOC or MOOC. Regarding the latter three offline prediction systems, i.e., Systems B ~ D [8], [22], [24] (proposed by R. Wang, Y. Cao, and Z. Wang respectively), the number of data sources is reduced, while the corresponding scale size rapidly increases; Unfortunately, the number of different types of behaviors that could be considered is decreased. Ideally, multisource data at a medium/large scale could help lead to a better prediction system design. However, in practice, due to limitations, such as computing capability, either data diversity or the sample size is sacrificed during the system design process.

To initially alleviate the challenges mentioned above, a model named Augmented Education (AugmentED) is

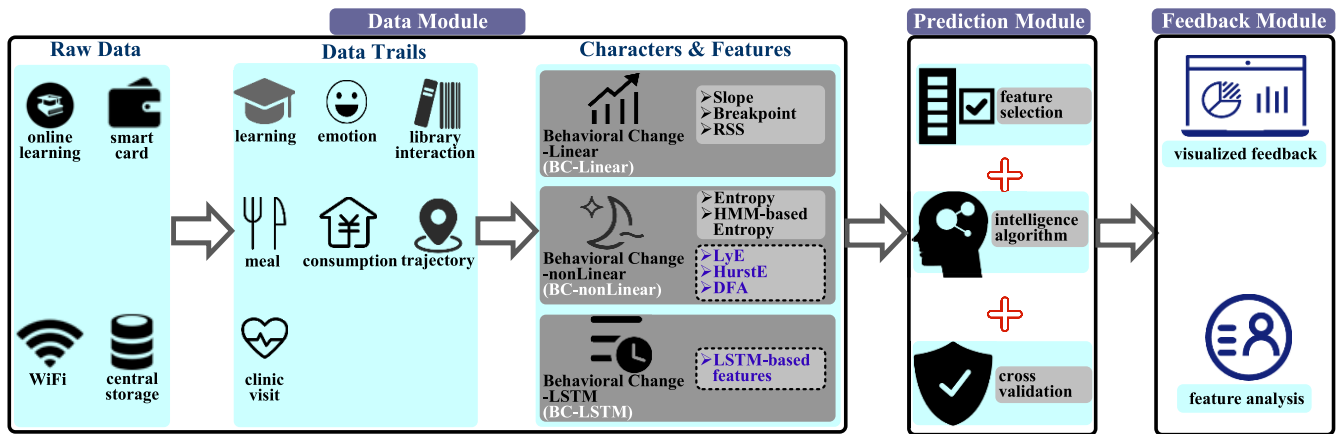


FIGURE 2. Overview of AugmentedED. In the data module, the features blocked in dashed boxes (including LyE, HurstE, DFA, and LSTM-based features) are proposed in our study, to the best of our knowledge, which is used for the first time in student’s behavioral analysis.

TABLE 1. Four typical prediction systems (proposed by previous researchers).

	Systems	Scale Size (N)	Data Source	Behaviors
On-line	System A (single-source + medium scale) [32]	243	SPOC/MOOC	<ul style="list-style-type: none"> ● online study ● discussions on the forum
Off-line	System B (multisource + small scale) [8]	30	<ul style="list-style-type: none"> ● wearable sensors ✧ smart phone ✧ accelerometer ✧ light sensor ✧ microphone ✧ GPS/Bluetooth ● self-reports ✧ SurveyMonkey ✧ mobile EMA 	<ul style="list-style-type: none"> ● activity ● conversation ● sleeping ● location ● socializing ● exercising ● mental health ● stress ● mood
	System C (almost multisource + medium scale) [24]	528	<ul style="list-style-type: none"> ● WiFi ● campus network ● smart card ● class schedule 	<ul style="list-style-type: none"> ● usage of smart card ✧ showering ✧ eating ✧ consumption ● trajectory ✧ wake-up time... ● network ✧ network cost...
	System D (single-source + large scale) [22]	18960	smart card	<ul style="list-style-type: none"> ● usage of smart card ✧ showering ✧ eating ✧ library entry-exit ✧ fetching water

proposed in this article. As shown in Fig. 2, this model mainly consists of the following three modules: (1) a **Data Module** in which multisource data on campus covering a large variety of data trails are aggregated and fused, and the characteristics/features that can represent students’ behavioral change from three different perspectives are evaluated; (2) a **Prediction Module** in which academic performance prediction is considered a classification problem that is solved by machine learning (ML)-based algorithms; and (3) a **Feedback Module** in which visualized feedback is delivered individually based on the predictions made and feature analysis. Finally, AugmentedED is examined using a real-world dataset of 156 college students.

The remainder of this article is organized as follows. In Section II, a literature review is given. In Section III, the methodology of AugmentedED is described in detail. In Section IV, the experimental results are discussed and analyzed. Finally, a brief conclusion is given in Section V.

II. RELATED WORK

A. FEATURE EXTRACTION

Feature evaluation plays an important role in designing prediction systems. Features that measure the various behavioral patterns can enhance our understanding of how a student’s behavior changes as the semester progresses. In this part, on the one hand, previous features that quantify students’ behavioral patterns are summarized; On the other hand, new features worthy of inclusion are also introduced.

In general, behavioral change can be quantified by the following three groups of metrics.

1) BEHAVIORAL CHANGE-LINEAR (BC-LINEAR)

Traditionally, behavioral change is mainly quantified by two linear metrics: behavioral slope and behavioral breakpoint.

First, the behavioral slope can be captured by computing the slope of the behavioral time series of each student using a linear regression [8]. The value of the slope indicates the direction and strength of the behavioral changes, e.g., a positive slope with a greater absolute value indicates a faster increase in behavioral change [8]. Given a mid-term day during the semester [8], both the pre-slope and post-slope can be calculated to represent the students’ behavioral change during the first and second halves of the semester, respectively.

Second, the behavioral breakpoint can be captured by computing the rate of behavioral changes occurring across the semester. The value of the breakpoint identifies the day during the semester before and after which a student’s behavioral patterns differed. Two linear regressions can be used to fit a behavioral time series and then use the Bayesian information criterion (BIC) to select the best breakpoint [8]. If a single

regression algorithm is selected, the breakpoint can be set to the last day.

2) BEHAVIORAL CHANGE-NONLINEAR (BC-NONLINEAR)

In recent years, nonlinear metrics have been increasingly applied to time series analysis [22], [44]–[59].

Regarding the students' behavioral time series, nonlinear metrics have been used to discover nonlinear behavioral patterns. We consider entropy an example. In [22], entropy is proposed to quantify the regularity/orderliness of students' behaviors, and it was demonstrated that a small entropy value generally leads to high regularity and high academic performance. Another example is entropy calculated based on a Hidden Markov Model (HMM) analysis [44], which is called *HMM-based entropy* for simplicity in our study. HMM-based entropy is proposed to quantify the uncertainty/diversity of students' behaviors, e.g., the uncertainty between the transition of different behaviors and the various activities that a behavior exhibits. In [44], HMM-based entropy is evaluated by the following two steps: (i) extracting the hidden states of a behavioral time series by HMM [45], [46]; and (ii) subsequently calculating the HMM-based entropy of the extracted hidden states.

To further recognize students' activities and discover their nonlinear behavioral patterns, the following three new metrics, which have not been applied in students' behavioral time series analysis previously, are also worth to be studied.

- **Lyapunov Exponent (LyE)** [47]–[51] is a measure of the stability of a time series. For example, in [47], LyE is used to quantify the stability of a gait time series, and the results demonstrate that a time series with a large LyE value is less stable than a series with a small LyE value, i.e., generally, a large LyE value indicates high instability. Therefore, in gait analyses, LyE is considered a stability risk indicator for falls [47] that can distinguish healthy subjects from those at a high risk of falling.
- **Hurst Exponent (HurstE)** [52]–[54] is a measure of predictability (in some studies, it is also called long-term memory) of a time series. For example, in [53], HurstE is applied to quantify the predictability of a financial time series, and the results demonstrate that a time series with a large HurstE value can be predicted more accurately than a series with a HurstE value close to 0.5.
- **Detrended Fluctuation Analysis (DFA)** [54]–[57] is a measure of the long-range correlation (also called statistical self-affinity or long-range dependence) of a time series [56]. For example, in [56], DFA is used to quantify the long-range correlation of a heart rate time series, and it is demonstrated that a time series with a small DFA value indicates less long-range correlation behavior than a series with a large DFA value. Therefore, in heart rate analyses, DFA is considered a long-range correlation indicator that can distinguish healthy subjects from those with severe heart disease [56].

In summary, the above three nonlinear metrics can measure the *stability*, *predictability*, and *long-range correlation* of

a time series. Although these metrics have already been extensively applied in time series analyses, e.g., gait time series [47], in this study, for the first time, they are used in a behavioral time series analysis. These metrics can enhance our understanding of not only whether a student's behavior is stable, predictable, and long-range correlated, but also how good a student's behavior is (e.g., self-discipline).

3) BEHAVIORAL CHANGE-LSTM (BC-LTSM)

Features represent temporal change over time is also worthy of study. Such features can be extracted by long short-term memory (LSTM) [58], which in this article is called LSTM-based features for short. LSTM-based features have been applied in many fields, including for example emotion recognition [59], [60], traffic forecast [61] and video action classification [62]. However, these features have not been applied in lifestyle behavioral analysis previously.

B. PREDICTION ALGORITHMS

In general, academic performance prediction can be considered either a regression or a classification problem. A wide variety of algorithms have been used/proposed in literatures to predict academic performance.

For example, in [8], *Lasso* (least absolute shrinkage and selection operator) regularized linear regression model, proposed by Tibshirani [63] in 1996, is used to predict academic performance. In [24], four supervised learning algorithms (consisting of support vector machine (SVM), logistic regression (LR), decision tree and naïve Bayes) are used to classify students' performance. In [22], RankNET, a neural network method proposed by Burges *et al.* [64] in 2015, is used to predict the ranks of students' semester grades. Similarly, in [27], a layer-supervised MLP-based method is proposed for academic performance prediction. In [32], a temporal emotion-aspect model (TEAM), modeling time jointly with emotions and aspects extracted from SPOC platform, is proposed to explore the effect of most concerned emotion-aspects as well as their evolutionary trends on academic achievement. In [65], four classification methods (consisting of Naïve-Bayes, SMO, J48, and JRip) are used to predict students' performance by considering student heterogeneity.

In general, due to the lack of open-access, large-scale, and multisource data sets in the education field, on the one hand, to some extent, it is impossible to compare the performances of the existing academic performance prediction algorithms; On the other hand, the algorithms proposed in this field are relatively simple, which are mainly based on basic statistics models (e.g. ANOVA and Post hoc tests) or ML algorithms (e.g. SVM and LR).

C. MULTISOURCE AND MULTIFEATURE

It has been verified in many literatures that the predictive power could be improved by multisource data and multifeatured fusion. For example, it is demonstrated that the performances of predicting both at-risk students [65] and stock market [66] could be improved by combining multi-source

TABLE 2. Characteristics and features evaluated in our study.

Raw Data				BC-Linear (Behavioral Change-Linear)			BC-nonLinear (Behavioral Change-nonLinear)					BC-LSTM (Behavioral Change-LSTM)	Basic info.	
	Data Source	Data Label	Data Content	Slope	Breakpoint	RSS	Entropy	HMM-based Entropy	LyE	HurstE	DFA	LSTM-based Features	Freq	Duration
On-line	SPOC	D ₁	online study	√	√	√	√		√	√	√	√	√	√
			emotion	√	√	√								
Off-line	Smart Card	D ₂	borrowing a book	√	√	√	√	√	√	√	√	√	√	
			library entry	√	√	√	√	√	√	√	√		√	√
			eating	√	√	√	√	√	√	√	√		√	√
			breakfast	√	√	√	√	√	√	√	√		√	√
			consumption	√	√	√	√	√	√	√	√		√	√
			clinical visit	√	√	√	√	√	√	√	√		√	√
	WiFi	D ₃	study area	√	√	√	√		√	√	√	√	√	√
			relaxation area	√	√	√	√		√	√	√		√	√
Central Storage				<ul style="list-style-type: none"> ● academic records: academic history and class schedule ● personal information: gender, age, and grade 										

Note: (i) “√” denotes features/characteristics extracted; (ii) “emotion” represents the students’ emotion patterns on SPOC forums in terms of engagement behaviors and the following three different emotions [36-38]: positivity, negativity, and confusion; (iii) “Freq” represents the total times of a behavior throughout the semester; and (iv) “grade” represents the year when a student starts his/her college study.

data. Similarly, in [22], [23], the performances of academic performance prediction are improved by combing traditional diligence features with orderliness (and sleep patterns) features. In [67], the accuracy of scholars’ scientific impact prediction is improved by using multi-field feature extraction and fusion. In [68], a contrast experiments of eleven different feature combinations were conducted, demonstrating that the performances of sentiment classification can be improved by multifeatured fusion.

However, we note that multisource and/or multifeature data cannot always guarantee a higher predictive power. For instance, [69] shows that the results of predictive modeling, notwithstanding the fact that they are collected within a single institution, strongly vary across courses. Actually, compared with single course, the portability of the prediction models across courses (multisource data) is lower [69]. Therefore, the effect of multisource and multifeature data needs to be varied in experiments.

III. METHODOLOGY

In our study, academic performance prediction is considered as a classification problem. According to the high-low discrimination index proposed by Kelley [41], academic performance is divided into *low*-, *medium*-, and *high*- groups. Given a digital campus dataset, according to Fig. 2, the main task is to first extract features from the raw multisource data; then select the features that are strongly correlated with academic performance and use these features to train the classification algorithm; and finally provide visualized feedback based on the prediction results.

In this section, the three modules designed in AugmentED (see Fig. 2) are described in detail.

A. DATA MODULE

A flowchart of this module is shown in Fig. 3, which includes the following three parts.

1) RAW DATA

Permission to access the raw data was granted by the Academic Affairs Office of our university. The raw dataset used in our study was captured from students engaging in the course of “*Freshman Seminar*” during the fall semester of 2018-2019. The “*Freshman Seminar*” was chosen for the following reasons: (1) more students were enrolled in this course ($N = 156$) than other comparable courses, and (2) these 156 students were more active on our self-developed SPOC platform, thus providing abundant valuable behavioral data. Our dataset consists of the following four data sources (see Table 2):

- **SPOC Data.** Two different types of data were collected on the SPOC platform. The first type is log files, which are recorded when a student logs in or out of the system, and the second type is posts on the SPOC discussion forum, which records discussions related to students’ learning experience.
- **Smart Card Data.** Similar to most modern universities, in our university, all students have a campus smart card registered under their real name. The usage of this smart card, such as for borrowing books from the library, entering the library, consuming meals in campus cafeterias, shopping on campus, or making an appointment with the school clinic, is captured daily.
- **WiFi Data.** There are approximately 3000 wireless access points at our university, covering most areas of campus. Once a student passes by one of these points, the MAC address of his/her device (e.g., tablet, laptop, or smart phone) can be recorded [40]. In our study, to distinguish among diverse behaviors, the entire campus is divided into several different areas, including a study area and a relaxation/dormitory area.
- **Central Storage Data.** As shown in Table 2, other features used in our study, including the students’ personal

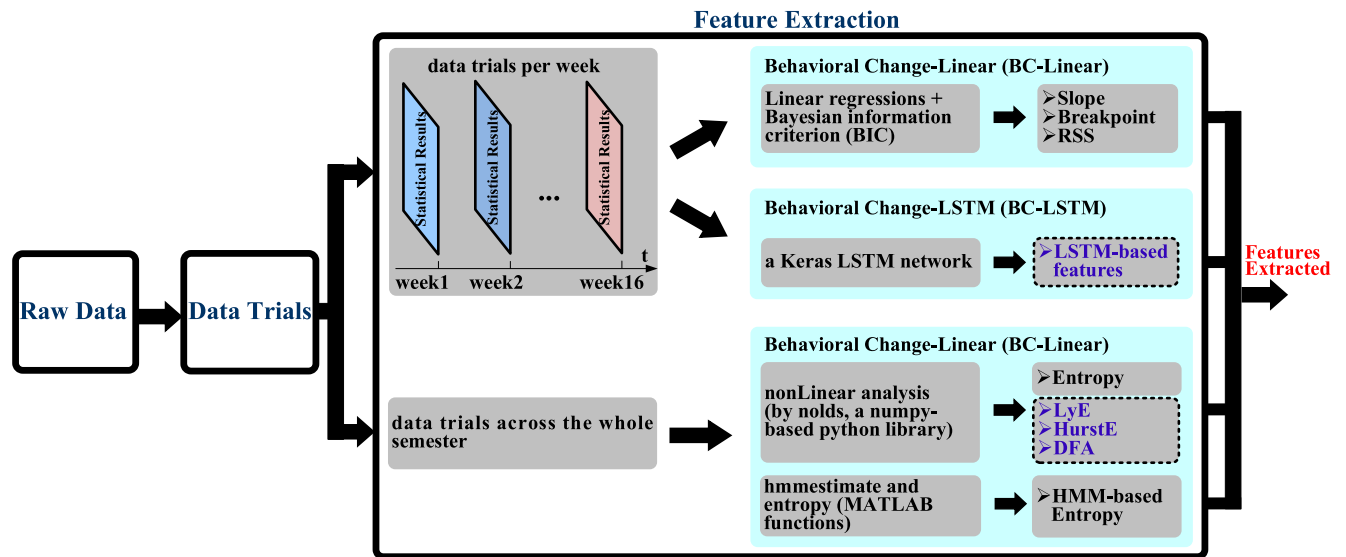


FIGURE 3. Flowchart of the data module.

information and academic records, are recorded by the central storage system of our university. For simplicity, the former three data sources are designated D_1 , D_2 , and D_3 , see Table 2. To evaluate the effect of multisource data on the academic performance prediction, which is similar to the studies introduced in Section II.C, contrast experiments of different data source combinations were conducted in our study (see Section IV). To be specific, based on D_1 , D_2 , and D_3 , in total, the following seven data combinations could be obtained: D_1 , D_2 , D_3 , D_1+D_2 , D_1+D_3 , D_2+D_3 , and $D_1+D_2+D_3$. The latter data source, i.e., Central Storage (which is relatively static and simple), is considered fundamental information shared by all seven combinations.

In our study, privacy protection is seriously considered, and all students' identifying information is anonymized. The infringement of students' privacy is avoided during both the data collection period and data analysis period. First, the student IDs are already pseudonymous in our raw data. Moreover, the resolution of the students' spatial-temporal trajectory is reduced. All information regarding the exact date/area showing when/where a behavior occurred is removed. Therefore, it would be reasonably difficult to reidentify individuals through our dataset.

2) DATA TRIALS

In our study, to initially understand how a student's behavior changes as the semester progresses, on the one hand, data trails across the whole semester is processed and organized in chronological order, including when, where and how a behavior occurs; On the other hand, data trails per week is summarized according to preliminary statistics, including the flowing information in each week, e.g., how often a behavior occurs (i.e. total frequency), how long does a behavior last (i.e. duration), and how much money does a student need.

Regarding the **SPOC data** (D_1), online learning is quantified by (i) *learning frequency and duration*, which are extracted from the raw log files; and (ii) *online learning emotion*, which is extracted from the discussion forum. Regarding the **Smart Card data** (D_2), multiple behaviors are involved, e.g. library interaction (including borrowing a book and library entry), see Table 2. Regarding the **WiFi data** (D_3), first, student's trajectory is calculated, mainly including when a student comes to a place; how often does he/she visit this place (i.e. frequency); how long does he/she stay there (i.e. duration). Second, attendance is calculated by combining WiFi data with class schedules. Specifically, to distinguish among behavioral patterns during different periods, three types of durations (namely, durations *on working days*, *on weekends*, and *throughout the semester*) and two types of attendances (namely, attendance *during the final study week* and attendance *throughout the semester*) are evaluated in our study.

3) FEATURE EXTRACTION

To gain a deeper insight into students' behavioral patterns, as summarized in Section II.A, in our study behavioral change is evaluated by linear, nonlinear, and deep learning (LSTM) methods, see Fig. 3.

- **BC-Linear.** Similar to the traditional approach, linear behavioral change is quantified by behavioral slope and behavioral breakpoint. Students behavioral series are fitted by two linear regressions, subsequently the optimized breakpoint is selected by BIC and behavioral slopes are calculated. Additionally, to further measure the amount of variance in the dataset that is not explained by the traditional regression model, the residual sum of squares (RSS) is also evaluated (see Table 2). In our

study, those linear metrics are mainly calculated by the python model *sklearn.linear_model*.

- **BC-nonLinear.** Similar to the traditional approach, first, entropy and HMM-based entropy are evaluated in our study, measuring the regularity and diversity of campus lifestyles respectively. Notably, the hidden states are numerically extracted by the MATLAB function *hmmestimate*, then the HMM-based entropy of the extracted hidden states is evaluated by the MATLAB function *entropy*. Second, to further discover nonlinear behavioral patterns, the following three nonlinear metrics are proposed and extracted for the first time: LyE, HurstE, and DFA, measuring the *stability*, *predictability*, and *long-range correlation* of campus lifestyles respectively. In our study, four nonlinear metrics (entropy, LyE, HurstE, and DFA) are evaluated by a numpy-based python library, i.e. *nolds*, based on the 0&1 sequence (see Appendix A).
- **BC-LSTM.** LSTM-based features representing dynamic changes in temporal behavioral patterns are calculated as follows. First, as input information, data trails from multiple behaviors are organized together week by week, see Fig. 3. In each week, the basic information of all multiple behaviors involved in our study is summarized, including for example how many times having breakfast and borrowing books from library *etc.* occurred respectively. Subsequently, this weekly information is fitted into a *Keras* LSTM network, then features representing the weekly behavioral patterns that might change throughout semester are extracted.

B. PREDICTION MODULE

The main task of this module is to select features and use these features to train the prediction algorithm.

1) FEATURE SELECTION

In our study, 708 different types of features are extracted, including 510 linear features, 119 nonlinear features, 50 LSTM-based features, and 29 basic features (including e.g. frequency and duration, gender, age, and grade). For instance, because multiple behaviors are involved in our study, there are 20 DFA related features in total to quantify *long-range correlation* for each behavior individually (e.g. library entry).

The distributions of the evaluated features and GPA are spread in different value scopes. Therefore, to eliminate a potential effect on the correlation analysis, both the features and GPA are normalized by min-max normalization. Additionally, to improve the performance of the prediction algorithms, the top 130 features with the most significant effect on academic performance are selected by the *SelectKBest* function in a python library named *scikit-learn*.

2) PREDICTION ALGORITHM

Subsequently, the selected features are used to train the ML-based classification algorithm for the academic performance

prediction. Specifically, in our study, five ML algorithms are applied, including RF (random forest), GBRT (gradient boost regression tree), KNN (k-nearest neighbor), SVM, and XGBoost (extreme gradient boosting). The hyperparameters of the ML and LSTM algorithms are optimized by *Grid-SearchCV* in *scikit-learn*.

3) CROSS VALIDATION

Our dataset is divided into a training set and a test set at the ratio of 7:3. The classification algorithm is first trained and then applied to the test set to predict academic performance. Finally, the robustness of the algorithm is tested by 10-fold cross validation.

C. VISUALIZATION MODULE

The main task of this module is to provide personalized feedback, including GPA prediction and a visualized summary of the students' behavioral patterns.

IV. EXPERIMENTAL RESULTS

In this section, first, the experimental results of AugmentED is presented and analyzed. Second, to evaluate the effectiveness of multisource and multifeature, contrast experiments are conducted, and the corresponding results are discussed. Finally, visualized feedback offered to students are designed.

A. PREDICTION RESULTS

The experimental results of AugmentED are shown in the last five rows of Table 3 (i.e., RF*, GBRT*, KNN*, SVM* and XGBoost*), which are highlighted in bold. Five indexes (accuracy, precision, recall, f1, and AUC) are used to evaluate the performance.

Notely, AugmentED is proposed based on

- multisource data**, i.e. $D_1+D_2+D_3$ (including SPOC, Smart Card, and WiFi data);
- multiple features**, i.e. C-III (including BC-Linear, BC-nonLinear, and BC-LSTM features).

* in Tab III denotes that C-III feature combination is used in the corresponding ML algorithms for academic performance prediction.

From Table 3, it can be seen that, first, the academic performance can be predicted by AugmentED with quite high accuracy. Second, the performance of the five different ML algorithms (RF*, GBRT*, KNN*, SVM* and XGBoost*) are similar, which can all lead to a good prediction result. To clarify, we consider the case of precision values, see the 5th column of Table 3. The precision values of five ML algorithms are 0.873, 0.877, 0.863, 0.889, and 0.871 respectively, indicating that (i) its minimum value is 0.863, i.e. the precision of AugmentED is no less than 86.3%; (ii) the difference between the minimum and maximum values is 0.026, which is quite small, i.e. AugmentED is independent of ML algorithms.

TABLE 3. Prediction Results (the average classification results of 10-Folder cross validation).

Data	Feature	Algorithm	accuracy	precision	recall	f1	AUC
D_1 (SPOC)	C-I	RF	0.604	0.642	0.604	0.602	0.698
		GBRT	0.584	0.643	0.584	0.581	0.684
		KNN	0.500	0.497	0.500	0.473	0.627
		SVM	0.539	0.590	0.539	0.534	0.641
		XGBoost	0.521	0.586	0.521	0.520	0.605
	C-II	LSTM	0.488	0.517	0.488	0.491	0.583
	C-III	RF*	0.776	0.806	0.776	0.774	0.807
		GBRT*	0.764	0.784	0.764	0.761	0.800
		KNN*	0.801	0.825	0.801	0.802	0.840
		SVM*	0.795	0.818	0.795	0.792	0.836
XGBoost*		0.789	0.823	0.789	0.787	0.836	
D_2 (Smart Card)	C-I	RF	0.507	0.538	0.507	0.496	0.600
		GBRT	0.487	0.510	0.487	0.481	0.604
		KNN	0.462	0.523	0.462	0.454	0.617
		SVM	0.528	0.571	0.528	0.515	0.637
		XGBoost	0.508	0.522	0.508	0.495	0.617
	C-II	LSTM	0.436	0.467	0.436	0.425	0.552
	C-III	RF*	0.742	0.795	0.742	0.746	0.808
		GBRT*	0.737	0.778	0.737	0.733	0.791
		KNN*	0.733	0.780	0.730	0.729	0.771
		SVM*	0.719	0.744	0.719	0.712	0.773
XGBoost*		0.733	0.760	0.733	0.733	0.767	
D_3 (WiFi)	C-I	RF	0.424	0.408	0.424	0.395	0.511
		GBRT	0.437	0.473	0.437	0.428	0.532
		KNN	0.399	0.424	0.399	0.396	0.531
		SVM	0.391	0.432	0.391	0.387	0.480
		XGBoost	0.425	0.383	0.425	0.395	0.503
	C-II	LSTM	0.413	0.398	0.413	0.380	0.512
	C-III	RF*	0.545	0.620	0.545	0.539	0.639
		GBRT*	0.558	0.600	0.558	0.548	0.650
		KNN*	0.501	0.591	0.501	0.487	0.607
		SVM*	0.487	0.404	0.487	0.416	0.590
XGBoost*		0.546	0.624	0.546	0.529	0.646	
D_1+D_2 (SPOC + Smart Card)	C-I	RF	0.616	0.660	0.616	0.608	0.688
		GBRT	0.603	0.676	0.603	0.605	0.692
		KNN	0.534	0.608	0.534	0.526	0.626
		SVM	0.608	0.661	0.608	0.617	0.683
		XGBoost	0.565	0.603	0.565	0.556	0.641
	C-II	LSTM	0.493	0.520	0.493	0.488	0.579
	C-III	RF*	0.814	0.839	0.814	0.815	0.855
		GBRT*	0.809	0.843	0.809	0.809	0.853
		KNN*	0.815	0.839	0.815	0.815	0.857
		SVM*	0.821	0.848	0.821	0.821	0.862
XGBoost*		0.833	0.860	0.833	0.826	0.813	
D_1+D_3 (SPOC + WiFi)	C-I	RF	0.616	0.650	0.616	0.614	0.697
		GBRT	0.616	0.664	0.616	0.612	0.702
		KNN	0.538	0.594	0.538	0.529	0.626
		SVM	0.602	0.640	0.602	0.598	0.690
		XGBoost	0.573	0.606	0.573	0.568	0.626
	C-II	LSTM	0.488	0.500	0.488	0.469	0.559
	C-III	RF*	0.800	0.867	0.800	0.799	0.849
		GBRT*	0.793	0.851	0.793	0.795	0.851
		KNN*	0.814	0.859	0.814	0.811	0.857
		SVM*	0.807	0.843	0.807	0.799	0.843
XGBoost*		0.807	0.843	0.807	0.803	0.867	
D_2+D_3 (Smart Card + WiFi)	C-I	RF	0.584	0.612	0.584	0.567	0.662
		GBRT	0.571	0.584	0.571	0.560	0.626
		KNN	0.551	0.593	0.551	0.558	0.641
		SVM	0.545	0.550	0.545	0.530	0.579
		XGBoost	0.559	0.563	0.559	0.544	0.579
	C-II	LSTM	0.449	0.469	0.449	0.434	0.551
	C-III	RF*	0.781	0.819	0.781	0.782	0.837
		GBRT*	0.781	0.801	0.781	0.779	0.816
		KNN*	0.755	0.793	0.755	0.754	0.793
		SVM*	0.762	0.794	0.762	0.763	0.815
XGBoost*		0.762	0.801	0.762	0.767	0.813	

$D_1+D_2+D_3$ (SPOC + Smart Card + WiFi)	C-I	RF	0.630	0.679	0.630	0.630	0.699
		GBRT	0.637	0.671	0.637	0.620	0.703
		KNN	0.604	0.647	0.604	0.596	0.669
		SVM	0.635	0.696	0.635	0.644	0.716
	C-II	XGBoost	0.608	0.670	0.608	0.598	0.691
		LSTM	0.501	0.544	0.501	0.473	0.578
	C-III	RF*	0.852	0.873	0.852	0.844	0.857
		GBRT*	0.852	0.877	0.852	0.852	0.876
		KNN*	0.847	0.863	0.847	0.841	0.851
		SVM*	0.866	0.889	0.866	0.865	0.872
XGBoost*		0.859	0.871	0.859	0.850	0.874	

Note: (i) $D_1+D_2+D_3$ is the multiple data source used in AugmentED to predict academic performance, including SPOC, Smart Card and WiFi data; (ii) The rows highlighted in light blue, light pink, light green, denote the three following feature or feature combinations are used for prediction, C-I (BC-linear and BC-nonlinear features), C-II (only BC-LSTM, i.e., LSTM-based features), C-III (BC-Linear, BC-nonLinear, and BC-LSTM features).

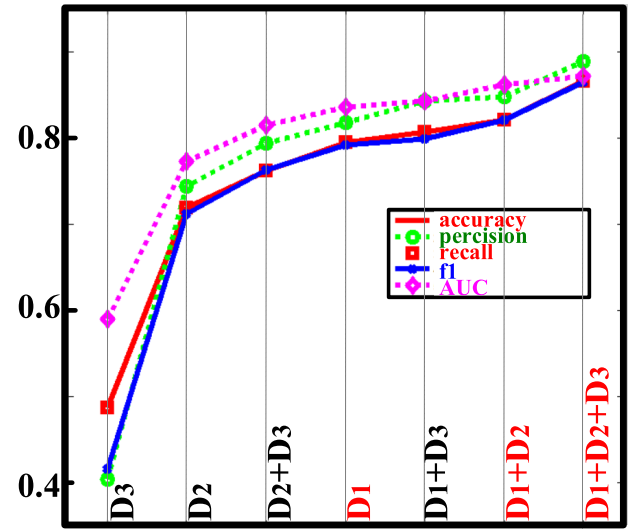


FIGURE 4. Comparisons of the SVM* performance of different data source combinations.

B. COMPARATIVE EXPERIMENTS

In this part, contrast experiments are conducted to evaluate the prediction effect of multisource and multifeature combinations.

1) MULTISOURCE

Comparisons of the performance of different data source combinations are conducted, see the 1st column of Table 3. As shown in Table 3, a large number of multiple data sources can lead to a more accurate prediction result.

To clarify, we consider the case of SVM*, from D_1 to D_1+D_2 and $D_1+D_2+D_3$ (see Tables 3 and Fig. 4), all five evaluation indexes significantly improves with the types of data sources increases. Specifically, (i) the accuracy values of D_1 , D_1+D_2 , and $D_1+D_2+D_3$ are 0.795, 0.821, 0.866, respectively; (ii) the precision values are 0.818, 0.848 and 0.889; (iii) the recall values are 0.795, 0.821 and 0.866; (iv) the f1 values are 0.792, 0.821 and 0.865; and (v) the AUCs value are 0.836, 0.862 and 0.872. It is verified that

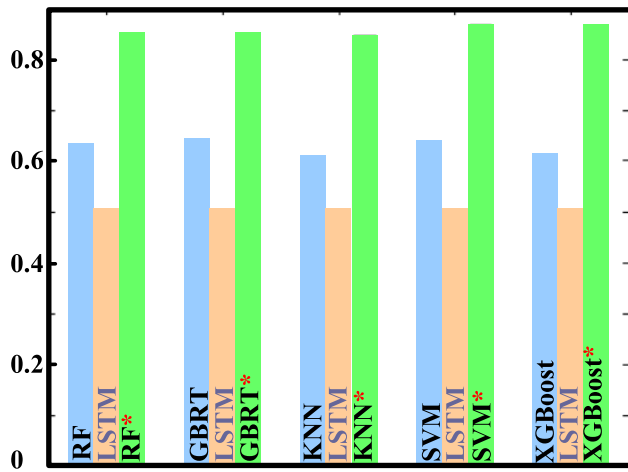


FIGURE 5. Comparisons of the accuracy values of three feature combinations in the (D₁+D₂+D₃) dataset.

multisource data can enhance the in-depth insight gained into students’ behavioral patterns.

2) MULTIFEATURE

Comparisons of the performance of three different feature combinations (C-I, C-II, C-III) are also conducted, see the 2nd column of Table 3 and Fig. 5.

- C-I (including BC-linear and BC-nonlinear features), see the rows of Table 3 highlighted in *light blue*. Its corresponding MLs are denoted as RF, GBRT, KNN, SVM, and XGboost;
- C-II (only including BC-LSTM features), see the rows of Table 3 highlighted in *light pink*;
- C-III (including BC-Linear, BC-nonLinear, and BC-LSTM features), see the rows of Table 3 highlighted in *light green*. Its corresponding MLs are denoted as RF*, GBRT*, KNN*, SVM*, and XGboost*.

As shown in Fig. 5 and Table 3, all five evaluation indexes (accuracy, precision, recall, f1, and AUC) of C-III are significantly higher than those of C-I and C-II. To clarify, we consider the case of SVM* in the (D₁+D₂+D₃) dataset, the accuracy value of SVM* is 0.866, which is much higher than that of SVM and LSTM (i.e., 0.635 and 0.501, respectively), see the 4th column of Table 3. This result indicates that the multifeature combination proposed in our study (i.e. C-III) can significantly improve the predictive power.

C. IDENTIFICATION OF AT-RISK STUDENTS BASED ON THE PREDICTOR

The prediction result obtained by AugmentED can be used to identify at-risk students, i.e., determine whether a student belongs in a low performance group. It could be quite helpful for early warning and feedback to be provided to at-risk students before the final exam week.

To illuminate how AugmentED could potentially help students optimize their college lifestyles and consequently

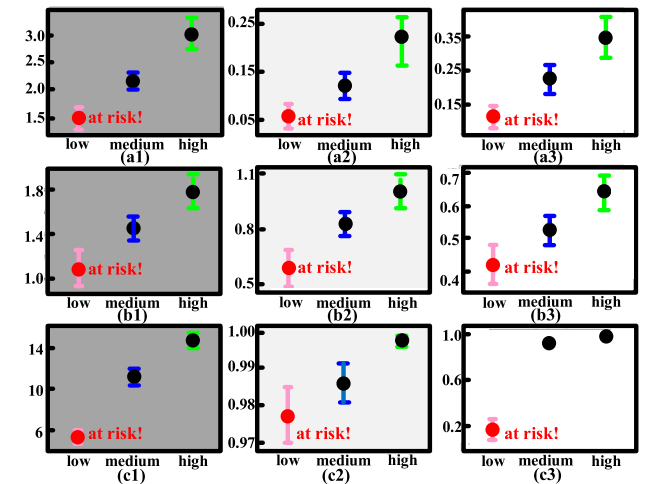


FIGURE 6. Example of a feedback given to one at risk student, including the average values and 95% confidence intervals of the following nine assistant indicators from the low-, medium-, and high- performance groups: First, (a1) D-Linear, (a2) D-postRSS, (a3) D-preSlope are the indicators representing (weighted) *linear*, *RSS in post-semester*, *slope in pre-semester* patterns of all behaviors (rather than one single behavior) respectively; Second, (b1) D-nonLinear; (b2) D-Entropy; (b3) D-DFA are the indicators representing (weighted) *nonlinear*, *entropy*, and *DFA* patterns of all behaviors. Finally, (c1) D-LSTM is the indicators representing the temporal pattern of all behaviors, while (c2) LSTM-49 and (c3) LSTM-1 are two of the 50 features extracted by our LSTM network.

improve their academic performance, a feedback example delivered to one at risk student is shown in Fig. 6.

We note that except for the prediction result itself, the extracted features that are strongly correlated with academic performance can also be taken as assistant indicators, to identify at-risk students. Traditionally, those features can be selected by either statistical analysis (e.g. by ANOVA) or ML algorithms (e.g. feature importance returned by RF). We recall that in our study, multiple behaviors are involved, and each behavior is quantified by a plenty of -linear, -nonlinear, and -LSTM features. Therefore, a particular feature (e.g. entropy) of one single behavior (e.g. either having breakfast or learning online) might not make sense to gain a comprehensive evaluation of student’ behavioral patterns. From this perspective, in Fig. 6, nine assistant indicators are calculated and plotted.

We begin by discussing the indicators of *-linear*, *-nonlinear*, and *-LSTM* features (see Appendix B), which are denoted as D-linear, D-nonLinear and D-LSTM respectively, representing the (weighted) linear, nonlinear and temporal pattern of all multiple behaviors involved in our study (rather than one single behavior). Regarding these three indicators,

- The average values and 95% confidence intervals (from the low-, medium-, and high- academic performance groups) are plotted in the left column of Fig. 6.
- The Pearson correlation between the indicators and academic performance is calculated, see the 2nd, 5th, and 8th rows of Table 4 which are highlighted in light gray.

TABLE 4. Correlation Coefficient and P-value.

	Assistant Indexes	Correlation coefficient	P-value
Fig.6(a1)	D-Linear	0.534	$7.18e^{-13}$
Fig.6(a2)	D-postRSS	0.366	$2.65e^{-06}$
Fig.6(a3)	D-preSlope	0.425	$3.12e^{-08}$
Fig.6(b1)	D-nonLinear	0.392	$4.28e^{-07}$
Fig.6(b2)	D-entropy	0.402	$2.02e^{-07}$
Fig.6(b3)	D-DFA	0.345	$1.05e^{-05}$
Fig.6(c1)	D-LSTM	0.703	$1.32e^{-24}$
Fig.6(c2)	LSTM-49	0.254	0.001
Fig.6(c3)	LSTM-1	0.734	$1.23e^{-27}$

Furthermore, six more indicators are calculated and provided as supplementary, see the 2nd and 3rd columns of Fig. 6. The Pearson correlation between these indicators and academic performance is also calculated and listed in Table 4.

From Table 4 it can be seen that all the nine indicators are strongly correlated with academic performance. Additionally, in Fig. 6, the apparent distinction among three academic performance groups demonstrates that all the nine indicators can offer strong support in at-risk student identification.

To clarify, we consider the case of D-linear. On the one hand, its average values and 95% confidence intervals from low-, medium-, and high- academic performance groups are $(1.457 \pm 0.199, 2.160 \pm 0.193, 3.035 \pm 0.341)$, see Fig. 6(a1), indicating clear separation. On the other hand, its correlation coefficient is 0.534, see the 3rd row of Table 4, i.e., this indicator is significantly correlated with academic performance. Therefore, D-linear can be taken as an indicator to explore which student is at risk because of the low performance he/she will achieve.

V. CONCLUSION AND FUTURE WORK

As an important issue in the education data mining field, academic performance prediction has been studied by many researchers. However, due to lack of richness and diversity in both data sources and features, there still exist a lot of challenges in prediction accuracy and interpretability. To initially alleviate this problem, our study aims at developing a robust academic performance prediction model, to gain an in-depth insight into student behavioral patterns and potentially help students to optimize their interactions with the university.

In our study, a model named AugmentED is proposed to predict the academic performance of college students. Our contributions in this study are related to three sources. First, regarding data fusion, to the best of our knowledge, this work is the first to capture, analyze and use multisource data covering not only online and offline learning but also campus-life behaviors inside and outside of the classroom for academic performance prediction. Based on these multisource data, a rich profile of a student is obtained. Second, regarding the feature evaluation, behavioral change is evaluated by linear, nonlinear, and deep learning (LSTM) methods respectively, which provides a systematical view of students' behavioral patterns. Specifically, it is the first time that three novel nonlinear metrics (LyE, HurstE, and DFA) and LSTM are applied in students' behavioral time series analysis. Third, our experimental results demonstrate that

AugmentED can predict academic performance with quite high accuracy, which help to formulate personalized feedback for at-risk (or unself-disciplined) students.

However, there are also some limitations in our study. To gain a multisource dataset, we sacrificed the scale the dataset by only using student-generated data within a single course. This limitation might have a certain negative influence on the generalization of AugmentED. Furthermore, in this study, we mainly focus on behavioral change. Other characteristics/features (e.g., peer effect, sleep) that are worthy of consideration were not evaluated in this study.

In conclusion, our study is based on a complete passive daily data capture system that exists in most modern universities. This system can potentially lead to continual investigations on a larger scale. The knowledge obtained in this study can also potentially contribute to related research among K-12 students.

APPENDIX A

To evaluate the four nonlinear metrics (entropy, LyE, HurstE, and DFA) of the time series, we concentrate on the precise time of day during which the behaviors occurred. Therefore, in our study, the involved time is first converted to a discrete time sequence. Then, according to the represented discrete time sequence, the raw behavioral time series data are converted to the 0&1 sequence as follows:

STEP 1. TIME DATA REPRESENTATION

The time data are converted to a discrete sequence with a normalized time interval by the following three steps:

- **Step 1.1.** The entire semester was from 01/09/2018 (September 1st) to 20/01/2019 (January 20st) and includes a total of 140 days. Thus, each day can be numbered from 1 to 140, resulting in a discrete sequence $\{p_1, p_2, \dots, p_i\} = \{1, 2, \dots, 140\}$, where i denotes the i^{th} day in the semester;
- **Step 1.2.** We divide each day into 48 time bins such that each bin spans 30 minutes. Subsequently, every bin is encoded from 1 to 48, i.e., $\{q_1, q_2, \dots, q_j\} = \{1, 2, \dots, 48\}$, where j denotes the j^{th} time bin. For example, “0:01—0:30” is the 1st time bin, “0:31—1:00” is the 2nd bin, etc.
- **Step 1.3.** By combining the sequences of days and time bins, the time during the spring semester is mapped to a discrete time sequence with length N_t , i.e., $\{T_1, T_2, \dots, T_{ij}\} = \{1, 2, \dots, N_t\}$, where

$$T_{ij} = (p_i - 1) \times 48 + q_j, \quad (\text{A-1})$$

And $N_t = 6720$. Specifically, if the time is “03/09/2018, 10:24”, i.e., $p_i = 3$ and $q_j = 21$ (the 21st time bin of the 3rd day), according to Eq. A-1, $T_{ij} = 2 \times 48 + 21 = 117$, i.e., “03/09/2018, 10:24” is encoded by 117.

STEP 2. BEHAVIORAL DATA REPRESENTATION

Following the time data representation, the raw behavioral data are converted to a 0&1 sequence by the following two steps:

- **Step 2.1.** First, a zero sequence X_{ij} with length N_t is generated, and
- **Step 2.2.** If a behavior occurs at time T_{ij} , the T_{ij}^{th} element of the corresponding discrete behavioral sequence X_{ij} is set to 1, i.e., $X_{ij} = 1$. For instance, if a student has a meal at “03/09/2018, 10:24” (where $T_{ij} = 117$), the 117th element of the discrete meal sequence is set to 1, i.e., $X_{ij} = 117$.

This process can be described as follows:

$$X_{ij} = \begin{cases} 1, & \text{if a behavior happens at time } T_{ij} \\ 0, & \text{otherwise,} \end{cases} \quad (\text{A-2})$$

where $X_{ij} \in [0,1]$. According to Eq. A-2, all behavioral data listed in Table 2 (including SPOC online study, borrowing a book, library entry, meal consumption, breakfast consumption, consumption, clinical visits, and WiFi data in the study and relaxation areas) are converted to discrete behavioral sequences.

APPENDIX B

Regarding the nine assistant indicators described in Section IV.C, the former seven are calculated according to [24]; while the latter two (LSTM-49, LSTM-1) are selected from the extracted 50 LSTM-features without any further processing.

The fundamental mathematical approach to calculate the former seven indicators is the same. The major similarity between these indicators is that they all represent certain property of all multiple behaviors involved in our study. The major difference is the input features used for calculation. To clarify, in this section the mathematical approach to the calculation of D-linear is given.

- **Step 1.** The score of each linear feature (e.g. slope) for each student is calculated as follows,

$$Score_k^n = \begin{cases} (N - Rank(x_n))/N, & Corr(X_k) > 0 \\ Rank(x_n)/N, & Corr(X_k) < 0 \end{cases} \quad (\text{B-1})$$

We assume that there are N students and K extracted features in total. $Corr(X_k)$ is the Pearson correlation coefficient between the k_{th} feature X_k and students' academic performance, where $k \leq K$. $Rank(x_n)$ means the ranking of the n_{th} student's (denoted as u_n , where $n \leq N$) feature among all students. For example, there are three students (u_1, u_2, u_3), and their k_{th} feature (e.g. slope value of having breakfast) are (0.8, 0.4, 0.6), then we have $Score_k^1 = 0$, $Score_k^2 = 0.667$, and $Score_k^3 = 0.333$ because $Corr(X_k) > 0$.

- **Step 2.** The indicator of linear feature group, D-Linear, is calculated by utilizing the feature scores as follows,

$$\sum_{k=1}^K (|Corr(X_k)| * Score_k^n) \quad (\text{B-2})$$

We note that essentially D-Linear is the weighted mean of all linear feature scores, and its weights are the correlation coefficients.

REFERENCES

- [1] A. Furnham and J. Mosen, “Personality traits and intelligence predict academic school grades,” *Learn. Individual Differences*, vol. 19, no. 1, pp. 28–33, 2009.
- [2] M. A. Conard, “Aptitude is not enough: How personality and behavior predict academic performance,” *J. Res. Personality*, vol. 40, no. 3, pp. 339–346, Jun. 2006.
- [3] T. Chamorro-Premuzic and A. Furnham, “Personality predicts academic performance: Evidence from two longitudinal university samples,” *J. Res. Personality*, vol. 37, no. 4, pp. 319–338, Aug. 2003.
- [4] R. Langford, C. P. Bonell, H. E. Jones, T. Poulou, S. M. Murphy, and E. Waters, “The WHO health promoting school framework for improving the health and well-being of students and their academic achievement,” *Cochrane Database Systematic Rev.*, vol. 4, no. 4, 2014, Art. no. CD008958.
- [5] A. Jones and K. Issroff, “Learning technologies: Affective and social issues in computer-supported collaborative learning,” *Comput. Edu.*, vol. 44, no. 4, pp. 395–408, May 2005.
- [6] A. G. van der Niet, E. Hartman, J. Smith, and C. Visscher, “Modeling relationships between physical fitness, executive functioning, and academic achievement in primary school children,” *Psychol. Sport Exerc.*, vol. 15, no. 4, pp. 319–325, Jul. 2014.
- [7] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell, “StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones,” in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. (UbiComp Adjunct)*, Seattle, WA, USA, 2014, pp. 3–14.
- [8] R. Wang, G. Harari, P. Hao, X. Zhou, and A. T. Campbell, “SmartGPA: How smartphones can assess and predict academic performance of college students,” in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput. (UbiComp)*, Osaka, Japan, 2015, pp. 295–306.
- [9] M. T. Trockel, M. D. Barnes, and D. L. Egget, “Health-related variables and academic performance among first-year college students: Implications for sleep and other behaviors,” *J. Amer. College Health*, vol. 49, no. 3, pp. 125–131, Nov. 2000.
- [10] D. M. Hansen, S. D. Herrmann, K. Lambourne, J. Lee, and J. E. Donnelly, “Linear/nonlinear relations of activity and fitness with children's academic achievement,” *Med. Sci. Sports Exerc.*, vol. 46, no. 12, pp. 2279–2285, 2014.
- [11] A. K. Porter, K. J. Matthews, D. Salvo, and H. W. Kohl, “Associations of physical activity, sedentary time, and screen time with cardiovascular fitness in United States adolescents: Results from the NHANES national youth fitness survey (NNYFS),” *J. Phys. Activity Health*, vol. 14, no. 7, pp. 506–512, 2017.
- [12] K. N. Aadland, Y. Ommundsen, E. Aadland, K. S. Brønneck, A. Lervåg, G. K. Resaland, and V. F. Moe, “Executive functions do not mediate prospective relations between indices of physical activity and academic performance: The active smarter kids (ASK) study,” *Frontiers Psychol.*, vol. 8, p. 1088, Jun. 2017.
- [13] M. Credé, S. G. Roch, and U. M. Kieszczynka, “Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics,” *Rev. Educ. Res.*, vol. 80, no. 2, pp. 272–295, Jun. 2010.
- [14] S. P. Gilbert and C. C. Weaver, “Sleep quality and academic performance in university students: A wake-up call for college psychologists,” *J. College Student Psychotherapy*, vol. 24, no. 4, pp. 295–306, Sep. 2010.
- [15] A. Wald, P. A. Muennig, K. A. O'Connell, and C. E. Garber, “Associations between healthy lifestyle behaviors and academic performance in U.S. undergraduates: A secondary analysis of the American college health association's national college health assessment II,” *Amer. J. Health Promotion*, vol. 28, no. 5, pp. 298–305, 2014.
- [16] P. Scanlon, “Identifying the impact of friends on their peers academic performance,” in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, San Francisco, CA, USA, Aug. 2016, pp. 1368–1369.
- [17] E. L. Faught, J. P. Ekwari, D. Gleddie, K. E. Storey, M. Asbridge, and P. J. Veugelers, “The combined impact of diet, physical activity, sleep and screen time on academic achievement: A prospective study of elementary school students in Nova Scotia, Canada,” *Int. J. Behav. Nutrition Phys. Activity*, vol. 14, no. 1, p. 29, Dec. 2017.

- [18] E. L. Faught, D. Gleddie, K. E. Storey, C. M. Davison, and P. J. Veugelaers, "Healthy lifestyle behaviours are positively and independently associated with academic achievement: An analysis of self-reported data from a nationally representative sample of Canadian early adolescents," *PLoS ONE*, vol. 12, no. 7, Jul. 2017, Art. no. e0181938.
- [19] V. Kassarnig, E. Mones, A. Bjerre-Nielsen, P. Sapiezynski, D. D. Lassen, and S. Lehmann, "Academic performance and behavioral patterns," *EPJ Data Sci.*, vol. 7, no. 1, p. 10, Dec. 2018.
- [20] J. E. Donnelly, C. H. Hillman, J. L. Greene, D. M. Hansen, C. A. Gibson, D. K. Sullivan, J. Poggio, M. S. Mayo, K. Lambourne, A. N. Szabo-Reed, S. D. Herrmann, J. J. Honas, M. R. Scudder, J. L. Betts, K. Henley, S. L. Hunt, and R. A. Washburn, "Physical activity and academic achievement across the curriculum: Results from a 3-year cluster-randomized trial," *Preventive Med.*, vol. 99, pp. 140–145, Jun. 2017.
- [21] N. Morita, T. Nakajima, K. Okita, T. Ishihara, M. Sagawa, and K. Yamatsu, "Relationships among fitness, obesity, screen time and academic achievement in Japanese adolescents," *Physiol. Behav.*, vol. 163, pp. 161–166, Sep. 2016.
- [22] Y. Cao, J. Gao, D. Lian, Z. Rong, J. Shi, and Q. Wang, "Orderliness predicts academic performance: Behavioural analysis on campus lifestyle," *J. Roy. Soc. Interface*, vol. 15, no. 146, 2018, Art. no. 20180210.
- [23] H. Yao, D. Lian, Y. Cao, Y. Wu, and T. Zhou, "Predicting academic performance for college students: A campus behavior perspective," *ACM Trans. Intell. Syst. Technol.*, vol. 1, no. 1, pp. 1–20, 2019.
- [24] Z. Wang, X. Zhu, J. Huang, X. Li, and Y. Ji, "Prediction of academic achievement based on digital campus," in *Proc. Int. Conf. Educ. Data Mining*, Buffalo, NY, USA, 2018, pp. 266–272.
- [25] I. Pytlarz, S. Pu, and M. Patel, "What can we learn from college students' network transactions? Constructing useful features for students prediction," in *Proc. 11th Int. Conf. Educ. Data Mining*, Buffalo, NY, USA, 2018, pp. 444–448.
- [26] S. Ahmad, K. Li, A. Amin, M. S. Anwar, and W. Khan, "A multilayer prediction approach for the student cognitive skills measurement," *IEEE Access*, vol. 6, pp. 57470–57484, 2018.
- [27] S. Qu, K. Li, S. Zhang, and Y. Wang, "Predicting achievement of students in smart campus," *IEEE Access*, vol. 6, pp. 60264–60273, 2018.
- [28] N. Alalwan, W. M. Al-Rahmi, O. Alfarraj, A. Alzahrani, N. Yahaya, and A. M. Al-Rahmi, "Integrated three theories to develop a model of factors affecting students' academic performance in higher education," *IEEE Access*, vol. 7, pp. 98725–98742, 2019.
- [29] T. Phan, S. G. McNeil, and B. R. Robin, "Students' patterns of engagement and course performance in a massive open online course," *Comput. Edu.*, vol. 95, pp. 36–44, Apr. 2016.
- [30] S. Liu, X. Peng, H. N. H. Cheng, Z. Liu, J. Sun, and C. Yang, "Unfolding sentimental and behavioral tendencies of learners' concerned topics from course reviews in a MOOC," *J. Educ. Comput. Res.*, vol. 57, no. 3, pp. 670–696, Jun. 2019.
- [31] S. Helal, J. Li, L. Liu, E. Ebrahimi, S. Dawson, D. J. Murray, and Q. Long, "Predicting academic performance by considering student heterogeneity," *Knowl.-Based Syst.*, vol. 161, pp. 134–146, Dec. 2018.
- [32] Z. Liu, C. Yang, S. Rüdian, S. Liu, L. Zhao, and T. Wang, "Temporal emotion-aspect modeling for discovering what students are concerned about in online course forums," *Interact. Learn. Environments*, vol. 27, nos. 5–6, pp. 598–627, Aug. 2019.
- [33] A. Zollanvari, R. C. Kizilirmak, Y. H. Kho, and D. Hernandez-Torrano, "Predicting students' GPA and developing intervention strategies based on self-regulatory learning behaviors," *IEEE Access*, vol. 5, pp. 23792–23802, 2017.
- [34] A. Akram, C. Fu, Y. Li, M. Y. Javed, R. Lin, Y. Jiang, and Y. Tang, "Predicting students' academic procrastination in blended learning course using homework submission data," *IEEE Access*, vol. 7, pp. 102487–102498, 2019.
- [35] L. Gao, Z. Zhao, L. Qi, Y. Liang, and J. Du, "Modeling the effort and learning ability of students in MOOCs," *IEEE Access*, vol. 7, pp. 128035–128042, 2019.
- [36] Z. Liu, H. N. H. Cheng, S. Liu, and J. Sun, "Discovering the two-step lag behavioral patterns of learners in the college SPOC platform," *Int. J. Inf. Commun. Technol. Edu.*, vol. 13, no. 1, pp. 1–13, Jan. 2017.
- [37] Z. Liu, W. Zhang, H. N. H. Cheng, J. Sun, and S. Liu, "Investigating relationship between discourse behavioural patterns and academic achievements of students in SPOC discussion forum," *Int. J. Distance Edu. Technol.*, vol. 16, no. 2, pp. 37–50, Apr. 2018.
- [38] Z. Liu, N. Pinkwart, H. Liu, S. Liu, and G. Zhang, "Exploring students' engagement patterns in SPOC forums and their association with course performance," *EURASIA J. Math., Sci. Technol. Edu.*, vol. 14, no. 7, pp. 3143–3158, May 2018.
- [39] B. Kim, B. Vizitei, and V. Ganapathi, "GritNet: Student performance prediction with deep learning," in *Proc. 11th Int. Conf. Educ. Data Mining*, Buffalo, NY, USA, 2018, pp. 625–629.
- [40] S. Sahebi and P. Brusilovshky, "Student performance prediction by discovering inter-activity relations," in *Proc. 11th Int. Conf. Educ. Data Mining*, Buffalo, NY, USA, 2018, pp. 87–96.
- [41] T. L. Kelley, "The selection of upper and lower groups for the validation of test items," *J. Educ. Psychol.*, vol. 30, no. 1, pp. 17–24, 1939.
- [42] J. Heo, S. Yoon, W. S. Oh, J. W. Ma, S. Ju, and S. B. Yun, "Spatial computing goes to education and beyond: Can semantic trajectory characterize students?" in *Proc. 5th ACM SIGSPATIAL Int. Workshop Anal. Big Geospatial Data (BigSpatial)*, Burlingame, CA, USA, 2016, pp. 14–17.
- [43] J. Heo, H. Lim, S. B. Yun, S. Ju, S. Park, and R. Lee, "Descriptive and predictive modeling of student achievement, satisfaction, and mental health for data-driven smart connected campus life service," in *Proc. 9th Int. Conf. Learn. Analytics Knowl.*, Tempe, AZ, USA, Mar. 2019, pp. 531–538.
- [44] X. Zhang, G. Sun, Y. Pan, H. Sun, Y. He, and J. Tan, "Students performance modeling based on behavior pattern," *J. Ambient Intell. Humanized Comput.*, vol. 9, no. 5, pp. 1659–1670, Oct. 2018.
- [45] S. R. Eddy, "Hidden Markov models," *Current Opinion Struct. Biol.*, vol. 6, no. 3, pp. 361–365, 1996.
- [46] *Hmmestimate*. [Online]. Available: <https://ww2.mathworks.cn/help/stats/hidden-markov-models-hmm.html>
- [47] J. Howcroft, J. Kofman, and E. D. Lemaire, "Review of fall risk assessment in geriatric populations using inertial sensors," *J. Neuroeng. Rehabil.*, vol. 10, no. 1, pp. 1–12, 2013.
- [48] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series," *Phys. D, Nonlinear Phenomena*, vol. 16, no. 3, pp. 285–317, 1985.
- [49] M. T. Rosenstein, J. J. Collins, and C. J. D. Luca, "A practical method for calculating largest Lyapunov exponents from small data set," *Expert Syst. Appl.*, vol. 29, no. 3, pp. 506–514, 1993.
- [50] S. M. Bruijn, D. J. J. Bregman, O. G. Meijer, P. J. Beek, and J. H. van Dieën, "Maximum Lyapunov exponents as predictors of global gait stability: A modelling approach," *Med. Eng. Phys.*, vol. 34, no. 4, pp. 428–436, May 2012.
- [51] N. Guler, E. Ubeyli, and I. Guler, "Recurrent neural networks employing Lyapunov exponents for EEG signals classification," *Expert Syst. Appl.*, vol. 29, no. 3, pp. 506–514, Oct. 2005.
- [52] H. E. Hurst, "A suggested statistical model of some time series which occur in nature," *Nature*, vol. 180, no. 4584, p. 494, 1957.
- [53] B. Qian and K. Rasheed, "Hurst exponent and financial market predictability," in *Proc. 2nd IASTED Int. Conf. Financial Eng. Appl.*, Cambridge, MA, USA, 2004, pp. 356–362.
- [54] R. Weron, "Estimating long-range dependence: Finite sample properties and confidence intervals," *Phys. A, Stat. Mech. Appl.*, vol. 312, nos. 1–2, pp. 285–299, Sep. 2002.
- [55] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, "Mosaic organization of DNA nucleotides," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 49, no. 2, pp. 1685–1689, Feb. 1994.
- [56] C. K. Peng, S. Havlin, E. H. Stanley, and A. L. Goldberger, "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 5, no. 1, pp. 82–87, 1995.
- [57] R. Hardstone, S. S. Poil, G. Schiavone, R. Jansen, V. V. Nikulin, and H. D. Mansvelder, "Detrended fluctuation analysis: A scale-free view on neuronal oscillations," *Frontiers Physiol.*, vol. 3, pp. 1–12, Nov. 2012.
- [58] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [59] S. Alhagry, A. Fahmy, and R. El-Khoribi, "Emotion recognition based on EEG using LSTM recurrent neural network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 10, pp. 355–358, 2017.
- [60] M. Wöllner, A. Metallinou, F. B. Eyben Schuller, and S. Narayanan, "Context-sensitive multimodal emotion recognition from speech and facial expression using bidirectional LSTM modeling," in *Proc. Interspeech, Conf. Int. Speech Commun. Assoc. (ISCA)*, Makuhari, Japan, 2010, pp. 2362–2365.

- [61] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: A deep learning approach for short-term traffic forecast," *IET Intell. Transp. Syst.*, vol. 11, no. 2, pp. 68–75, Mar. 2017.
- [62] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, "Action classification in soccer videos with long short-term memory recurrent neural networks," in *Proc. Int. Conf. Artif. Neural Netw.*, Thessaloniki, Greece, 2010, pp. 154–159.
- [63] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc., Ser. B (Methodol.)*, vol. 58, no. 1, pp. 267–288, Jan. 1996.
- [64] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender, "Learning to rank using gradient descent," in *Proc. 22nd Int. Conf. Mach. Learn. (ICML)*, Bonn, Germany, 2005, pp. 89–96.
- [65] S. Helal, J. Li, L. Liu, E. Ebrahimie, S. Dawson, D. J. Murray, and Q. Long, "Predicting academic performance by considering student heterogeneity," *Knowl.-Based Syst.*, vol. 161, pp. 134–146, Dec. 2018.
- [66] X. Zhang, S. Qu, J. Huang, B. Fang, and P. Yu, "Stock market prediction via multi-source multiple instance learning," *IEEE Access*, vol. 6, pp. 50720–50728, 2018.
- [67] Z. Wu, W. Lin, P. Liu, J. Chen, and L. Mao, "Predicting long-term scientific impact based on multi-field feature extraction," *IEEE Access*, vol. 7, pp. 51759–51770, 2019.
- [68] L. Qiu, Q. Lei, and Z. Zhang, "Advanced sentiment classification of tibetan microblogs on smart campuses based on multi-feature fusion," *IEEE Access*, vol. 6, pp. 17896–17904, 2018.
- [69] R. Conijn, C. Snijders, A. Kleingeld, and U. Matzat, "Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS," *IEEE Trans. Learn. Technol.*, vol. 10, no. 1, pp. 17–29, Jan. 2017.



XIAOLIANG ZHU received the B.S. degree from the Hefei University of Technology, Hefei, China, in 1996, and the M.S. and Ph.D. degrees from the Huazhong University of Science and Technology, Wuhan, China, in 2003 and 2006, respectively.

He is currently an Associate Professor with the National Engineering Laboratory for Educational Big Data, Central China Normal University, China. His research interests include affective computing and education data mining.



JIANWEN SUN received the Ph.D. degree in educational technology from Central China Normal University, Wuhan, China, in 2011.

He has been an Associate Professor with the National Engineering Research Center for E-Learning since 2014. He joined the National Engineering Laboratory for Educational Big Data in 2018. His current research interests include smart classrooms, the brain-computer interface, and knowledge tracing.



LIANG ZHAO received the Ph.D. degree in electronic engineering from the University College Dublin (UCD), Dublin, Ireland, in 2015.

From 2015 to 2017, she held a postdoctoral position with The Insight Centre for Data Analytics, UCD, Dublin, Ireland. She has been an Associate Professor with Central China Normal University since 2017. She has authored 14 journal articles (including ten SCI articles) and 18 conference papers. Her research interests include data analysis and prediction, time and frequency analysis, and digital holography.



KUN CHEN received the B.E. degree in automation from Huazhong Agricultural University, Wuhan, China, in 2018. He is currently pursuing the M.E. in computer science and technology with Central China Normal University, China.

His main research interest is educational data mining.



JIE SONG received the B.E. degree in communications engineering from Guangxi University, Nanning, China, in 2017. She is currently pursuing the M.E. degree in communication and information system with Central China Normal University, Wuhan, China.

Her main research interest is educational data mining.



BRIAN CAULFIELD (Member, IEEE) received the Ph.D. degree in medical science from the University College Dublin (UCD), Ireland, in 1992.

He has been a Professor with UCD since 2014. He is currently the Director of The Insight Center for Data Analytics, UCD, and a Lead Investigator with the Applied Research for Connected Health Centre. His research program is focused on exploiting technological advances to enhance human performance in the fields of connected

health and sport. He has coauthored over 180 research publications and six patents.



BRIAN MAC NAMEE received the Ph.D. degree in computer science from the Trinity College Dublin in 2004.

In 2015, he joined the School of Computer Science, University College Dublin, as a Lecturer. He is currently the Centre Director for the SFI Centre for Research Training in Machine Learning and a Funded Investigator with The Insight Centre for Data Analytics and VistaMilk SFI Research Centre. He has authored or coauthored extensively

in machine learning, predictive analytics, and information visualization—a recent highlight is the textbook *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples and Case Studies* (MIT Press, 2015). His research focuses on machine learning, predictive analytics, data visualization, and augmented reality.

...