Early Prediction of Students at Risk of Failing a Face-to-Face Course in Power Electronic Systems

Raúl Alcaraz, Arturo Martínez-Rodrigo, Roberto Zangróniz, *Member, IEEE*, and José J. Rieta

***Abstract*—Early warning systems (EWSs) have proven to be useful in identifying students at risk of failing both online and conventional courses. Although some general systems have reported acceptable ability to work in modules with different characteristics, those designed from a course-specific perspective have recently provided better outcomes. Hence, the main goal of this work is to design a tailored EWS for a conventional course in power electronic circuits. For that purpose, effectiveness of some common classifiers in predicting at-risk students has been analyzed. Although slight differences in their performance have only been noticed, an ensemble classifier combining outputs from several of them has provided to be the best performer. As a major contribution, a novel weighted voting combination strategy has been proposed to exploit global information about how basic prediction algorithms perform on several time points during the semester and diverse subsets of student-related features. Predictions at five critical points have been analyzed, revealing that the end of the fourth week is the optimal time to identify students at risk of failing the course. At that moment, accuracies about 85–90% have been reached. Moreover, several scenarios with different subsets of student-related attributes have been considered in every time point. Besides common parameters from student's background and continuous assessment, novel features estimating student's performance progression on weekly assignments have been introduced. The proposal of this set of new input variables is another key contribution, because they have allowed to improve more than 5% predictions of at-risk students at every time point.**

***Index Terms*—Early warning system, at-risk students, performance prediction, educational data mining, power electronic systems.**

I. Introduction

NOWADAYS, academic processes generate a huge amount of data, mainly pertaining to the interaction between students and teachers, as well as between students and their learning environment [1]. This information is being thoroughly explored to gain novel knowledge about how students learn in a variety of scenarios, and then in the improvement of the quality of educative systems by providing timely support to learners, instructors and administrators [2]. In fact, the application of well-known data mining techniques to educational data is an emerging research field, which is referred to as Educational Data Mining (EDM) [1]. More precisely, the International Educational Data Mining Society defines EDM as "an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in." [3].

Although many application areas can be found within this research field, one of the most important is to predict student's academic performance [4]. This kind of prediction has been developed at different levels of granularity [5]. Whereas student's score on a specific activity or test is predicted at task level, student's success/failure or end-of-semester grade are forecasted at course level [4]. Additionally, prediction of student's dropout in a degree program has also been widely addressed [4]. In all cases, an early forecast is desired to enable proactive teaching actions aimed at providing students with sufficient support to improve their performance and avoid their attrition [6], [7]. To this respect, some recent experiments have proven that tools such as intelligent tutoring systems, early warning systems (EWSs), and recommender systems can be very useful in higher education [8].

Given the large dropout rates exhibited by online courses, EWSs have been mostly established in this kind of teaching [9]. In fact, detailed parametrization of the student's access pattern to e-learning platforms has allowed to successfully identify learners at risk of quitting [10] and, then implement instructional interventions for their retention [11]. However, the systems based on this approach cannot be directly extrapolated to face-to-face teaching, because in this case most activities are developed outside an e-learning platform [7]. Hence, in the last years great efforts have been made to design effective EWSs for conventional courses, and particularly for disciplines traditionally presenting high failure rates, such as programming [12], maths [13], [14], physics [15], and engineering [7]. In this context, learning of low-performance students can still be improved through proper instructional interventions, including one-on-one tutoring and review of key concepts after class, assignment of extra homework, provision of remedial lessons, or supervised revision of previously learned concepts in prerequisite courses [15].

Within the area of engineering, courses dealing with the bases of power electronic systems are daunting for many students, since they often struggle with understanding the complex physical phenomena involved in the operation of

Manuscript received MMM DDD, YYY; revised MMM DDD, YYY; accepted MMM DDD, YYY. Date of publication MMM DDD, YYY; date of current version MMM DDD, YYY. This work was supported by the Research Group in Electronic, Biomedical and Telecommunication Engineering through the University of Castilla–La Mancha and the European Regional Development Fund under Grant 2018/11744. *(Corresponding author: Raúl Alcaraz.)*

R. Alcaraz and R. Zangróniz are with the Department of Electronic Engineering, University of Castilla-La Mancha, 16071 Cuenca, Spain (e-mail: raul.alcaraz@uclm.es; roberto.zangroniz@uclm.es).

A. Martínez-Rodrigo is with the Department of Computer Systems, University of Castilla-La Mancha, 16071 Cuenca, Spain (e-mail: arturo.martinez@uclm.es).

J.J. Rieta is with BioMIT.org, Electronic Engineering Department, Universitat Politecnica de Valencia, 46022 Valencia, Spain (e-mail: jjrieta@upv.es).

Digital Object Identifier XXXX.

many circuits [16]. Indeed, many learners are not sufficiently familiar with the high level of abstractness required to handle non-visible phenomena, like electrostatic, magnetic, electromagnetic, or thermal fields, as well as non-visible quantities, such as magnetic flux [16], [17]. Hence, they usually need to dedicate a huge amount of time and effort to understand the key concepts of the course, and this extra work leads many students to quit at an early stage or not be able to acquire sufficient knowledge to pass. As an example, the course under study in the present work, entitled *Power Electronic Systems*, has presented failure rates between 45–60% in the last offerings, such as Table I shows. Therefore, this and other similar courses could strongly benefit from the use of an EWS to prevent most low-performance students from failing.

So far, a few EWSs have proven a high level of generalization to reliably predict at-risk students in courses of different disciplines and with diverse learning settings [18]–[20]. However, in the last years many works have reported that EWSs designed from a course-specific perspective achieve significantly higher accuracy in student's performance predictions [7], [21]–[23]. To this respect, the differences among courses in terms of structure, intended learning objectives, planned activities, and grading schemes have been suggested to have a high impact on the selection of the most predictive input variables for each particular case [23], [24]. In view of this context and considering that no tailored EWS has been still presented in the field of power electronics, the present work explores how effective some common machine learning algorithms are in predicting student's performance in a face-to-face course dealing with basic notions of power systems. Moreover, an ensemble classifier combining outputs from several of them through a novel weighted voting strategy is also proposed. This method has reported the best performance, overcoming basic classification techniques, as well as homogeneous and heterogeneous ensemble classifiers.

A key point in the design of every EWS is its ability to operate at an early stage, when students have still not quitted the course [18]. Although some works have reported accurate predictions at the end of the semester, they are of very limited use when it comes to establish an EWS [13]. At the opposite extreme, some authors have tried to predict at-risk students before the course starts [15], [25], [26]. In this case, only personal information and student's academic performance in past modules can be used to make predictions. However, the fact that every information about how students learn during the course is discarded could explain the limited predictive ability about 60–70% reported by previous works [7]. Consequently, a trade-off between accuracy and early prediction has been recommended for a successful implementation of every EWS [13]. Hence, to attain this compromise, several predictions have been conducted in five critical times along the semester.

Adequate selection of student-related features also plays a main role in reaching accurate predictions [27]. To this respect, student's performance data available during the course have been recently identified as the factors able to provide the earliest and most accurate predictions of learners at risk of failing conventional modules [7]. In fact, cumulative and average grades on quizzes, homework assignments, hands-on experiments, and mid-term exams have provided high predictive abilities in the first weeks of the semester [12], [13], [15], [28], [29]. However, it has still not been explored whether progression of the student's grades on successive assessment tasks can improve predictions. The temporal trend exhibited by these scores could be associated with student's engagement and progressive learning, both aspects being particularly relevant in courses where cumulative knowledge is essential. This is the case of *Power Electronic Systems*, since the basic concepts covered at the beginning of the course are definitely required to later understand how most power circuits work [16]. Hence, another major contribution of the present study is that a set of novel features are proposed to quantify student's performance progression on weekly assignments. The role of these new input variables in predicting at-risk learners is also assessed in all the time points tested during the semester.

As a summary, the present study focuses on the following three research questions:

- How effective are common machine learning techniques to identify students at risk of failing a conventional course in power electronic systems?
- What is the optimal time point to predict at-risk students along the course?
- Can student's performance progression on weekly assignments provide useful predictive information?

II. Materials and Methods

A. Main Features of the Course

Power Electronic Systems is a compulsory second-year course in the Telecommunications Engineering Degree Program at Technical School of Cuenca, University of Castilla-La Mancha (UCLM), Spain. It exposes students for the first time to fundamental concepts of three-phase voltages and powers, transformers, single-phase and three-phase rectifiers, isolated and non-isolated dc-dc converters, single-phase and three-phase inverters, and photovoltaic installations.

The course consists of 4 theory European Credit Transfer System (ECTS) credits and 2 practice ECTS credits, thus requiring 40 lecture hours and 20 laboratory (lab) hours, structured over 15 weeks. Students are therefore required to attend three 90 min sessions per week, two for theory lectures and one for hands-on experiments. Regarding assessment, 50% of the end-of-semester mark is distributed throughout the course to foster student's engagement [13], [30], [31]. Thus, the grading scheme awards 5% of the end-of-semester mark for in-class activities (ICAs), 5% for homework assignments (HAs), 20% for lab activities (LAs), 20% for a mid-term exam (MTE), and 50% for the final exam. The MTE is held in week 9, thus covering the topics addressed from the beginning of the course to week 8. Similarly, the final exam is always scheduled for two weeks after the course finishes.

The student's continuous assessment is planned on a weekly basis. Thus, each week starts with two lectures, where the instructor teaches theory concepts using PowerPoint presentations and solves some illustrative problems on a blackboard.

TABLE I
NUMBER OF STUDENTS ENROLLED IN THE EXPERIMENT, AND PASS AND FAILURE RATES FOR EACH OFFERING OF THE COURSE

Offering	# Students	Failure Rate	Pass Rate
2010/11	44	61.36%	38.64%
2011/12	50	58.00%	42.00%
2012/13	48	45.84%	54.16%
2013/14	44	56.82%	43.18%
2014/15	46	47.82%	52.18%
2015/16	40	47.50%	52.50%
2016/17	46	43.48%	56.52%
2017/18	44	59.10%	40.90%

In the last 45 minutes of the second class period, students are asked to solve a problem on their own. Although they do not work in group, they can share and discuss possible solutions with peers. This ICA gives students the opportunity to apply the gained knowledge, as well as to resolve their particular doubts. In fact, the instructor will walk around the classroom during the activity to monitor learners' progress and provide individual feedback. On the other hand, to encourage learners to continue studying the covered materials, they are also required to solve a HA (consisting in another problem) outside class and within a period of two days. It should be noted that these ICAs and HAs were designed by the same instructor for all weeks with an homogenous level of difficulty.

For each week, a LA is also scheduled, where students are required to simulate via Matlab/Simulink and build on a breadboard some circuits for their analysis and testing. The instructor will provide learners with sufficient guidance, in form of direct instructions, to complete the intended tasks. Moreover, students will work in permanent groups of two, and will have to submit their simulations and measurements, along with their conclusions about the operation of the tested circuits, before the end of the session. All these weekly tasks (i.e., ICAs, HAs and LAs), as well as the mid-term and final exams, will be corrected by the instructor within the same week of submission, and scores will be assigned on a scale of 0 to 10. These grades will then be used to make predictions of at-risk students at the end of some weeks (see Section II-E).

B. Participants and Their Pass/Fail Classification

A total of 362 students were enrolled in the experiment from eight consecutive offerings of the course, such as Table I shows. No significant differences were noticed in the number of students registered for each offering, as well as in pass and failure rates. Moreover, it is worth noting that the course was always taught the same three days a week, in the same time-slots, and by the same instructor. Also, the covered content remained unaltered, and the topics were introduced in the same order. Likewise, the same e-learning platform (i.e. Moodle) was used to create similar learning environments, where students could download teaching material and upload their results for HAs and LAs.

To pass the course, students had to satisfy two requirements: (i) an end-of-semseter score equal or greater than five points, and (ii) a grade on the final exam equal or greater than four points. No minimum marks were required for ICAs, HAs and LAs, because UCLM regulation constrains compulsory attendance in undergraduate courses. No requisite was also set to pass the MTE, because it was considered as a good learning opportunity, where students received valuable feedback about their progress. Accordingly, 172 students passed the course and the remaining 190 failed it. This binary classification (i.e., pass/fail) was used as output variable to train and test the predictive models obtained from the classifiers described in the next two subsections.

C. Common Prediction Methods

In the last years a broad variety of classification algorithms have been used to predict student's performance in higher education [4], [32]. However, no method able to successfully work on every scenario has still been found, thus compelling to look for the optimal technique in each particular case [5]. Accordingly, several basic and advanced classifiers widely used in EDM contexts have been tested in the present work. For that purpose, statistics and machine learning functionalities offered by Matlab R2019a (The MathWorks Inc., Natick, Massachusetts, United States) were used.

Briefly, *Logistic Regression* (LR) is a statistical technique estimating an output variable as a linear combination of several input attributes [33]. Its goal is to predict the probability of occurrence of an event by fitting data into a logistic function. Thus, in the training stage a coefficient regression, representing its degree of contribution to the output, is computed for each input variable. Given its relative simplicity and reliability, this method has been broadly used to predict student's performance [7], [15], [29], [34]. As in these previous works, the algorithm was trained without any kind of regularization.

Another simple algorithm commonly employed in EDM applications is *Naive Bayes Classifier* (NBC) [35]. This method is based on well-known Bayes' Theorem, thus classifying data according to the highest probability of belonging to a particular category [33]. It assumes independence among input variables, but this condition is often infringed, including the scenario of student's performance prediction considered here. Nonetheless, even in that case, the algorithm is usually able to provide predictions with similar accuracy to more complex classifiers [36]. Continuous input variables were modeled as gaussian distributions, because a Kolmogorov-Smirnov test proved that most of them were normally distributed. For categorical predictors, a multinomial distribution was considered.

The combination of a set of rules in a tree form has also proven to be an effective and easily interpreted classifier in many situations, including EDM [34]. The basic idea behind this *Decision Tree* (DT) is to split the data based on one variable until every node only presents objects from a category, or all input variables have been used [33]. To pick the feature that best separates the data at each level, a score function must be selected to estimate the impurity of all possible nodes. Although there exist several approaches for that purpose, the common Gini's index was used, such as in other previous works [7], [34]. Moreover, to control the tree growth, the maximum number of splits and the minimum number of instances per leaf were set to 10 and 2, respectively [34].

A more complex algorithm is *Support Vector Machine* (SVM), which classifies data from two categories by finding the linear decision boundary (i.e., the hyperplane) that best discerns between them [33]. This method hence turns classification into an optimization problem, where the best hyperplane is the one providing the largest margin between two categories. If data are not linearly separable, a kernel function can then be used to provide dissociation between categories in a higher dimension. However, because simple kernels can increase generalizability of the predictive models [37], a linear kernel with a box constraint parameter of 1 was used in this study [7].

An *Artificial Neural Network* is a classifier inspired by the human brain. Indeed, it consists of a set of interconnected nodes or neurons, which relate the inputs to the desired outputs [33]. Each neuron performs tasks of information processing by converting received inputs into processed outputs. Depending on these tasks and how neurons learn, several variants can be found. A *Multi-Layer Perceptron* (MLP) with two hidden layers and a number of nodes just the half of input variables, a sigmoid activation function, and a learning rate of 0.1 was used, according to previous studies also dealing with student's performance prediction [7].

Unlike the previous techniques, a *K–Nearest Neighbor* (KNN) algorithm is a non-parametric classifier. In this case, a discriminant model is not trained with parameters, but each object is classified by majority voting from its K neighbors [33]. This approach assumes that objects near each other are similar, thus playing a key role in how the distance among them is computed. Although a variety of distance metrics exist, the well-known Euclidean distance was employed here. Moreover, information from the five nearest neighbors, i.e., from the five students with the most similar scores, was considered to predict at-risk students, such as in [7].

In addition to these basic classifiers, advanced homogeneous ensemble algorithms have also been used in EDM applications, such as Random Forest (RF) and AdaBoost (AB) [32], [38], [39]. These methods attempt to create a strong classifier by majority voting from a set of weak learners, which are differently derived for each case. Precisely, RF is based on iteratively constructing a set of decision trees by randomly selecting a portion of the input features, as well as by resampling the training dataset [37]. In this way, an ensemble of randomized, independent, and equally weighted trees are obtained. A similar approach is also used in AB, but in this case each new classification model is influenced and weighted by the performance of those built previously. Thus, a tree is firstly designed from the training data, and then a second model is created to correct the errors from the first one. Models are iteratively added until the training set is perfectly predicted, or a maximum number of models are reached [37]. As for DT, both RF and AB were trained with a Gini's index to split data, a maximum number of splits of 10, and a minimum number of instances per leaf of 2. Also, according to previous works [34], [40], the number of trained trees was 10 for both methods.

D. Heterogeneous Ensemble Classifiers Based on Voting

Compared to AB and RF, heterogeneous ensemble algorithms have provided better classification results in a variety of fields [41], [42]. These classifiers are based on combining outputs from different base algorithms, thus exploiting diversity in their results and obtaining a strong generalization [43], [44]. In this case, the combination strategy plays a crucial role, and several alternatives have been introduced. A popular option is majority voting, because it is simple and reaches good results in many contexts [45], including student's performance prediction [7]. Thus, by combining the six basic classifiers previously described, i.e., LR, NBC, DT, SVM, MLP, and KNN, a majority voting ensemble (MVE) approach was constructed.

A disadvantage of this combination strategy is that all base classifiers are treated equally, sometimes leading to suboptimal predictions [44], [45]. To palliate this issue, assignation of specific weights to each base algorithm has been proposed [44]. Unfortunately, no approach has still been designed to estimate these weights when predicting student's performance, and therefore a well-known weighted voting ensemble (WVE) algorithm has been studied in the present work. The weights were computed from the accuracy reported by each basic algorithm on only one classification task and a defined subset of input variables, such as described in [44]. This approach could overtrain the ensemble for the classification context, and hence a novel combination strategy has also been proposed. This new scheme is based on the results of a Friedman test, which compares the performance of the base classifiers on several time points during the semester and different subsets of student-related features. More details about this statistical test, the analyzed time points, and the subsets of student-related attributes (used as input variables) will be described in the next three subsections (i.e., in Sections II-E, II-F, and II-G).

Briefly, a Friedman test separately sorts the performance of each algorithm on all the considered classification tasks and reports an average rank for each one. Scaling these values between 0 and 1 through a min-max normalization (which will be referred to as $\overline{\mathcal{R}_i}$), the specific weight (w_i) for each base classifier was obtained through an inverse exponential rule, i.e., $w_i = A \cdot e^{-\overline{\mathcal{R}}_i}$, where A was computed to meet the condition that $\sum_{i=1}^{6} w_i = 1$. Note that i ranges from 1 to 6, because the six basic algorithms previously described, i.e., LR, NBC, DT, SVM, MLP, and KNN, were combined. Finally, the weights w_i were rounded to the nearest multiple of five, and a final decision was then adopted by majority voting. This approach will henceforth be referred to as the proposed ensemble classifier (PEC).

E. Control Points

To determine how early at-risk students could be identified, several predictive models were trained and tested at different critical times during the course, which will hereafter be referred to as control points (CPs). Thus, because predictions as early as possible are desired to enable teaching actions aimed at preventing low-performance students from premature quitting [34], the first CP (CP1) was established just before the beginning of the semester. The second CP (CP2) was established at week 4, since the basic concepts about three-phase circuits and transformers, which are essential to understand more advanced power systems (e.g., rectifiers and inverters) [16], are covered within this time period.

During the next four weeks, the main notions about non-controlled and controlled rectifiers, as well as their applications, are mainly described. Bearing the complex operation of these circuits in mind [17], the third CP (CP3) was then set at week 8. The MTE is just held one week later, thus locating the forth CP (CP4) at this moment. Finally, the last CP (CP5) was established at week 13, once the main concepts about dc-dc converters and inverters are covered. The remainder of the course only presents the application of these circuits in the design of isolated and grid-connected photovoltaic installations and, therefore, the more challenging topics have been taught at that point. Although low-performance students are very close to fail the course at week 13, extremely accurate predictions may be expected.

F. Input Variables

The student-related attributes used to feed the prediction algorithms were selected from two categories. Firstly, some features were chosen from the student's background. Thus, final scores achieved in all the first-year courses, along with their average, were collected on a continuous scale of 0 to 10. These courses are *Electronic Components and Circuits* (ECC), *Electronic Devices and Subsystems* (EDS), *Fundamentals of Physics I* (FPI), *Fundamentals of Physics II* (FPII), *Fundamentals of Mathematics I* (FMI), *Fundamentals of Mathematics II* (FMII), *Fundamentals of Mathematics III* (FMIII), *Signals and Systems* (SS), *Computing* (COM), and *Business Management* (BM). All are compulsory and consist of 6 ECTS credits (4 theory and 2 practice). Additionally, the number of these courses that students passed, failed and quitted were also considered as categorical variables, which took integer values between 0 and 10.

In CP1, these input variables from the student's background could only be explored. Although UCLM regulation restricts the possibility of establishing prerequisite courses, those introducing essential notions to understand the operation of most power electronic circuits received special attention. Indeed, three different scenarios were analyzed, such as Table II summarizes. Thus, whereas the student's scores on the two and three first-year courses more closely related to *Power Electronic Systems* were considered in the two first cases, respectively (i.e., grades on ECC and EDS in scenario #1, and on ECC, EDS, and FPII in scenario #2), all features were analyzed in the last experiment (i.e., scenario #3).

On the other hand, the most predictive variables from the student's background were complemented in CP2–CP5 with three different subsets of features extracted from the student's continuous assessment along the semester (see Table II). More precisely, mean student's scores on the ICAs, HAs, and LAs completed until each CP were firstly considered (scenario #1). According to previous works [7], [14], [15], these attributes were defined as

$$\overline{ICA_{\mathcal{W}}} = \frac{1}{\mathcal{W}} \sum_{i=1}^{\mathcal{W}} ICA_i, \tag{1}$$

$$\overline{HA_{\mathcal{W}}} = \frac{1}{\mathcal{W}} \sum_{i=1}^{\mathcal{W}} HA_i, \text{ and} \tag{2}$$

$$\overline{LA_{\mathcal{W}}} = \frac{1}{\mathcal{W}} \sum_{i=1}^{\mathcal{W}} LA_i, \tag{3}$$

where $\mathcal{W}$ refers to the week when each CP was established, i.e., 4, 8, 9, and 13 for CP2, CP3, CP4, and CP5, respectively, and ICA_i, HA_i, and LA_i to the student's scores on the corresponding tasks for i-th week. Note that these parameters were continuous and ranged from 0 to 10.

Instead of these average grades, student's performance progression on the same weekly tasks for each time period was analyzed in scenario #2 (see Table II). Thus, in this case the following indices were considered as input variables:

$$\Delta ICA_{\mathcal{W}} = \frac{1}{\mathcal{W}-1} \sum_{i=1}^{\mathcal{W}-1} ICA_{i+1} - ICA_i, \tag{4}$$

$$\Delta HA_{\mathcal{W}} = \frac{1}{\mathcal{W}-1} \sum_{i=1}^{\mathcal{W}-1} HA_{i+1} - HA_i, \text{ and} \tag{5}$$

$$\Delta LA_{\mathcal{W}} = \frac{1}{\mathcal{W}-1} \sum_{i=1}^{\mathcal{W}-1} LA_{i+1} - LA_i. \tag{6}$$

It should be noted that these indices were continuous and ranged from -10 to 10.

To compute these six features in each CP, those activities that were not submitted on time were discarded. Nonetheless, the number of these unreported ICAs, HAs, and LAs were also quantified as categorical variables, which took integer values between 0 and $\mathcal{W}$. As can be seen in Table II, these input variables were analyzed in scenario #3 of CP2–CP5, along with the six previously defined attributes from the student's continuous assessment and the most predictive ones from the student's background. Finally, remark that the student's score on the MTE was also considered as an input in the three scenarios analyzed in CP4 and CP5. As all student's marks, it was quantified on a continuous scale of 0 to 10.

G. Assessment of Prediction Outcomes

The performance of all predictive models has been summarized in terms of two metrics. Thus, accuracy (Acc) was firstly estimated as the fraction of students correctly classified, i.e.,

$$Acc = \frac{TP + TN}{TP + FN + TN + FP}, \tag{7}$$

where TP (*true positives*) and FN (*false negatives*) are the number of failing students correctly and incorrectly identified, and TN (*true negatives*) and FP (*false positives*) are the number of passing students properly and improperly classified, respectively. Although this index provides a global overview about how a predictive model works, it does not give specific information about the percentage of students rightly discerned for each group. However, this information is relevant in the present study, because the main goal is to identify at-risk learners. Hence, the rate of failing students correctly identified, commonly known as sensitivity (Se), was also computed as

$$Se = \frac{TP}{TP + FN}. \tag{8}$$

TABLE II
INPUT FEATURES CONSIDERED IN EACH CP AND SCENARIO TO PREDICT STUDENTS AT RISK OF FAILING THE COURSE

CP	Scenario #1	Scenario #2	Scenario #3
1	Grade on ECC	Grade on ECC	Grade on ECC
	Grade on EDS	Grade on EDS	Grade on EDS
		Grade on FPII	Grade on FPII
			Grade on FPI
			Grade on FMI
			Grade on FMII
			Grade on FMIII
			Grade on COM
			Grade on SS
			Grade on BM
			Mean score
			# courses passed
			# courses failed
			# courses quitted
2	Grade on ECC	Grade on ECC	Grade on ECC
	Grade on EDS	Grade on EDS	Grade on EDS
	Grade on FPII	Grade on FPII	Grade on FPII
	$\overline{ICA_4}$	ΔICA_4	$\overline{ICA_4}$
	$\overline{HA_4}$	ΔHA_4	$\overline{HA_4}$
	$\overline{LA_4}$	ΔLA_4	$\overline{LA_4}$
			ΔICA_4
			ΔHA_4
			ΔLA_4
			# unreported ICAs
			# unreported HAs
			# unreported LAs
3	Grade on ECC	Grade on ECC	Grade on ECC
	Grade on EDS	Grade on EDS	Grade on EDS
	Grade on FPII	Grade on FPII	Grade on FPII
	$\overline{ICA_8}$	ΔICA_8	$\overline{ICA_8}$
	$\overline{HA_8}$	ΔHA_8	$\overline{HA_8}$
	$\overline{LA_8}$	ΔLA_8	$\overline{LA_8}$
			ΔICA_8
			ΔHA_8
			ΔLA_8
			# unreported ICAs
			# unreported HAs
			# unreported LAs
4	Grade on ECC	Grade on ECC	Grade on ECC
	Grade on EDS	Grade on EDS	Grade on EDS
	Grade on FPII	Grade on FPII	Grade on FPII
	$\overline{ICA_9}$	ΔICA_9	$\overline{ICA_9}$
	$\overline{HA_9}$	ΔHA_9	$\overline{HA_9}$
	$\overline{LA_9}$	ΔLA_9	$\overline{LA_9}$
	Grade on MTE	Grade on MTE	ΔICA_9
			ΔHA_9
			ΔLA_9
			# unreported ICAs
			# unreported HAs
			# unreported LAs
			Grade on MTE
5	Grade on ECC	Grade on ECC	Grade on ECC
	Grade on EDS	Grade on EDS	Grade on EDS
	Grade on FPII	Grade on FPII	Grade on FPII
	$\overline{ICA_{13}}$	ΔICA_{13}	$\overline{ICA_{13}}$
	$\overline{HA_{13}}$	ΔHA_{13}	$\overline{HA_{13}}$
	$\overline{LA_{13}}$	ΔLA_{13}	$\overline{LA_{13}}$
	Grade on MTE	Grade on MTE	ΔICA_{13}
			ΔHA_{13}
			ΔLA_{13}
			# unreported ICAs
			# unreported HAs
			# unreported LAs
			Grade on MTE

On the other hand, although resubstitution validation has been widely used in EDM contexts, training and testing with the same data often leads the predictive model to be overfitted [46]. A well-known procedure to handle this issue and generalize the performance of every classifier is K-fold cross-validation [46]. Thus, a 10-fold cross-validation approach was used, where data were randomly rearranged to ensure that every subset was sufficiently representative of the whole [46]. In each one of the 10 iterations, the basic classifiers and homogeneous ensembles were trained and tested with separate, non-overlapped subsets of samples. Moreover, for each iteration, the weights in the MVE and PEC algorithms were computed from the values of Acc obtained by the six basic classifiers (i.e., LR, NBC, DT, SVM, MLP, and KNN) on the training samples, according to the approaches described in Section II-D. Thus, as a result of this validation, 10 independently trained and validated prediction models were obtained for every basic, homogeneous ensemble, and heterogeneous ensemble classifier.

Finally, some statistical tests were also conducted on the values of Acc obtained by the classifiers. More precisely, to compare the performance of every classifier in two different conditions (e.g., two different scenarios within a CP), a paired Wilcoxon signed rank test was used [47]. Moreover, a Friedman test was employed for a global comparison of the performance of the classifiers on all the analyzed CPs and scenarios [47]. For both tests, a statistical significance p lower than 0.05 was considered as significant. Also, average ranks computed by the Friedman test were used to sort out classifiers from the best performer to the lowest one [48], as well as to design the PEC (such as previously described in Section II-D).

III. RESULTS

Cross-validated classification outcomes obtained by the prediction models built in CP1 are shown in Table III. A significant increase in values of Acc and Se can be noticed from scenario #1 to #2 for all the prediction algorithms. Indeed, improvements about 4–5% were reported when student's score on FPII (scenario #2) was added to those on ECC and EDS (scenario #1). Also, a paired Wilcoxon test always reported values of statistical significance $p < 0.05$, when the performance of every classifier was assessed in both scenarios. On the contrary, compared to scenario #2, no significantly larger, or even lower, values of Acc and Se were noticed for all the classifiers in scenario #3, i.e., when all input variables from the student's background were considered to predict at-risk students. In this case, a paired Wilcoxon test provided values of $p > 0.05$ for every classifier.

On the other hand, Tables IV–VII display prediction results achieved by the models built in CP2–CP5, respectively. In general terms, regardless of the tested scenario, a stepwise increase in both Acc and Se was observed for all the classifiers as the semester progressed. Thus, improvements about 10–12% were noticed from CP1 to CP2, about 3–4% from CP2 to CP3, about 1–2% from CP3 to CP4, and about 0.5–1% from CP4 to CP5. As a graphical summary, Fig. 1 shows how Acc and Se evolve over the CPs in scenario #2 for the common LR approach. Nonetheless, it should be noted that, when the performance of every classifier was compared in two

TABLE III
PREDICTION RESULTS OBTAINED BY THE MODELS BUILT IN CP1

Algorithm	Scenario #1		Scenario #2		Scenario #3	
	Acc	Se	Acc	Se	Acc	Se
LR	0.699	0.711	0.776	0.763	0.765	0.758
NBC	0.704	0.689	0.779	0.753	0.776	0.758
DT	0.718	0.689	0.782	0.747	0.757	0.747
SVM	0.729	0.721	0.773	0.768	0.773	0.763
MLP	0.721	0.705	0.757	0.758	0.768	0.768
KNN	0.715	0.711	0.787	0.768	0.790	0.774
RF	0.737	0.744	0.798	0.806	0.792	0.811
AB	0.742	0.722	0.794	0.789	0.781	0.764
MVE	0.735	0.729	0.780	0.777	0.788	0.799
WVE	0.727	0.736	0.784	0.782	0.786	0.797
PEC	**0.761**	**0.748**	**0.811**	**0.827**	**0.806**	**0.816**

TABLE IV
PREDICTION RESULTS OBTAINED BY THE MODELS BUILT IN CP2

Algorithm	Scenario #1		Scenario #2		Scenario #3	
	Acc	Se	Acc	Se	Acc	Se
LR	0.820	0.837	0.873	0.900	0.878	0.905
NBC	0.831	0.821	0.867	0.884	0.865	0.879
DT	0.829	0.842	0.873	0.868	0.873	0.874
SVM	0.843	0.858	0.887	0.895	0.892	0.889
MLP	0.826	0.853	0.873	0.884	0.881	0.879
KNN	0.829	0.847	0.890	0.895	0.896	0.889
RF	0.844	0.842	0.894	0.895	0.902	**0.925**
AB	0.849	0.848	0.891	0.887	0.896	0.908
MVE	0.845	0.832	0.883	0.875	0.896	0.918
WVE	0.843	0.840	0.891	0.873	0.898	0.920
PEC	**0.862**	**0.859**	**0.904**	**0.904**	**0.911**	0.918

TABLE V
PREDICTION RESULTS OBTAINED BY THE MODELS BUILT IN CP3

Algorithm	Scenario #1		Scenario #2		Scenario #3	
	Acc	Se	Acc	Se	Acc	Se
LR	0.878	0.853	0.923	0.932	0.920	0.921
NBC	0.867	0.858	0.925	0.947	0.917	**0.942**
DT	0.873	0.863	0.920	0.916	0.909	0.916
SVM	0.881	0.874	0.939	0.947	0.931	0.937
MLP	0.876	0.874	0.925	0.932	0.920	0.932
KNN	0.883	0.868	0.936	0.942	0.925	0.937
RF	0.888	0.897	0.930	0.944	0.922	0.932
AB	0.885	0.882	0.933	0.918	0.933	0.925
MVE	0.888	0.890	0.936	0.937	0.929	0.930
WVE	0.888	0.895	0.933	0.939	0.931	0.928
PEC	**0.907**	**0.910**	**0.950**	**0.955**	**0.938**	**0.942**

TABLE VI
PREDICTION RESULTS OBTAINED BY THE MODELS BUILT IN CP4

Algorithm	Scenario #1		Scenario #2		Scenario #3	
	Acc	Se	Acc	Se	Acc	Se
LR	0.906	0.895	0.953	0.937	0.945	0.921
NBC	0.898	0.884	0.948	0.958	0.942	0.953
DT	0.892	0.863	0.956	0.937	0.950	0.942
SVM	0.903	0.879	0.959	0.954	0.956	0.937
MLP	0.912	0.900	0.936	0.963	0.939	**0.958**
KNN	0.903	0.895	0.950	0.958	0.945	0.937
RF	0.901	0.900	0.962	**0.986**	0.952	0.941
AB	0.907	0.893	0.960	0.952	0.950	0.944
MVE	0.911	0.893	0.960	0.973	0.952	0.941
WVE	0.915	0.888	0.957	0.970	0.955	0.946
PEC	**0.925**	**0.923**	**0.975**	0.962	**0.968**	0.950

successive CPs in the same scenario, statistically significant differences ($p < 0.05$) were only noticed from CP1 to CP2 for all the algorithms.

Apart from this increasing trend over time, for CP2–CP5

TABLE VII
PREDICTION RESULTS OBTAINED BY THE MODELS BUILT IN CP5

Algorithm	Scenario #1		Scenario #2		Scenario #3	
	Acc	Se	Acc	Se	Acc	Se
LR	0.912	0.905	0.956	0.963	0.953	0.947
NBC	0.901	0.895	0.961	0.958	0.959	0.974
DT	0.909	0.889	0.959	0.942	0.953	0.937
SVM	0.914	0.889	0.964	0.974	0.953	0.974
MLP	0.923	0.895	0.961	0.937	0.963	0.953
KNN	0.914	0.921	0.959	0.958	0.948	0.968
RF	0.912	0.909	0.961	0.945	0.959	0.956
AB	0.910	0.909	0.961	0.945	0.956	**0.975**
MVE	0.922	0.895	0.957	0.949	0.961	0.962
WVE	0.922	0.892	0.959	0.960	0.957	0.964
PEC	**0.935**	**0.958**	**0.976**	**0.988**	**0.970**	0.957

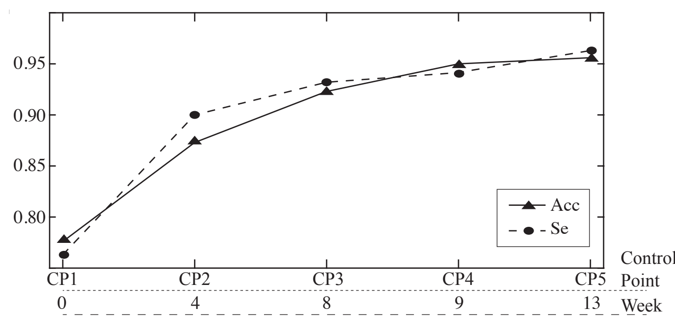

Fig. 1. Evolution over the CPs of the values of Acc (solid line) and Se (dotted line) obtained by the common LR method for the scenario #2.

similar variations in values of Acc and Se were also observed among the three analyzed scenarios. More precisely, the best prediction outcomes were always reported in scenario #2, i.e., when the proposed variables $\Delta ICA_{\mathcal{W}}$, $\Delta HA_{\mathcal{W}}$, and $\Delta LA_{\mathcal{W}}$ were jointly considered with the most predictive features from the student's background. In fact, compared to this case, for every CP a decrease by about 5% was observed both in Acc and Se when the prediction algorithms were fed with the attributes $\overline{ICA_{\mathcal{W}}}$, $\overline{HA_{\mathcal{W}}}$, and $\overline{LA_{\mathcal{W}}}$ in scenario #1. Moreover, a paired Wilcoxon test always provided values of statistical significance $p < 0.05$, when the performance of every classifier was assessed in both scenarios. Finally, only negligible improvements (lower than 1%) were sometimes seen in Acc and Se from scenario #2 to #3, i.e., when all the attributes from the student's continuous assessment were tested. Indeed, the performance of any classifier was not statistically different for both scenarios, being always statistical significance $p > 0.05$.

For every CP and scenario, no great differences were observed among the classifiers in terms of Acc and Se. However, a Friedman test rejected the null hypothesis, thus pointing out that some algorithms performed better than others. Then, a post-hoc analysis with Bonferroni correction was conduced, and adjusted p-values are shown in Table VIII. As can be seen, among the basic classifiers, SVM provided a significantly better performance than LR, NBC, and DT. Moreover, although no significant differences were noticed between SVM, KNN, and MLP, average ranks computed by the Friedman test suggest that SVM (rank of 5.67) performed better than KNN (rank of 6.63), and MLP (rank of 7.57). Regarding RF and AB, both reported a significantly better performance

TABLE VIII
ADJUSTED p-VALUES OBTAINED BY A POST HOC ANALYSIS CONDUCTED ON RESULTS OF A FRIEDMAN TEST

Algorithm	LR	NBC	DT	SVM	MLP	KNN	RF	AB	MVE	WVE	PEC
LR	—	9.123	9.780	0.054	2.259	0.475	0.001	0.003	0.003	0.002	<0.001
NBC	9.123	—	8.905	0.076	2.709	0.613	0.001	0.004	0.005	0.004	<0.001
DT	9.780	8.905	—	0.050	2.155	0.445	0.001	0.003	0.003	0.002	<0.001
SVM	0.054	0.076	0.050	—	1.167	4.248	2.478	3.935	4.090	3.637	0.001
MLP	2.259	2.709	2.155	1.167	—	4.409	0.064	0.154	0.166	0.132	<0.001
KNN	0.475	0.613	0.445	4.248	4.409	—	0.507	0.987	1.044	0.879	<0.001
RF	0.001	0.001	0.001	2.478	0.064	0.507	—	7.621	7.412	8.044	0.070
AB	0.003	0.004	0.003	3.935	0.154	0.987	7.621	—	9.780	9.561	0.027
MVE	0.003	0.005	0.003	4.090	0.166	1.044	7.412	9.780	—	9.342	0.025
WVE	0.002	0.004	0.002	3.637	0.132	0.879	8.044	9.561	9.342	—	0.032
PEC	<0.001	<0.001	<0.001	0.001	<0.001	<0.001	0.070	0.027	0.025	0.032	—

than LR, NBC, and DT. Also, both homogeneous ensembles proved to be better performers than SVM, KNN, and MLP, but no statistically significant differences were observed. Indeed, average ranks for RF and AB were 4.27 and 4.67, respectively. Very similar results to these two classifiers were also reported by the heterogeneous ensambles MVE and WVE, which exhibited average ranks of 4.63 and 4.57, respectively. Moreover, no statistically significant differences between heterogeneous and homogeneous ensembles were noticed. Nonetheless, the PEC significantly improved the performance of all previous classifiers, reporting a rank of 1 and statistically significant differences from most of them.

IV. DISCUSSION

The main findings obtained in the study are next discussed in response to the raised research questions. Additionally, other relevant aspects are also covered in the last two subsections.

A. *How Effective Are Common Machine Learning Techniques to Identify Students at Risk of Failing a Conventional Course of Power Electronic Systems?*

All the analyzed classifiers have proven to be rather effective in identifying students at risk of failing the course. In fact, values of Acc and Se between 70 and 81% were observed even when only attributes from the student's background were considered (see Table III). According to some previous works [13], [15], [25], this outcome reinforces the idea that past achievements can be useful in predicting future student's performance. Nonetheless, the predictive abilities observed in CP1 are considerably higher than those reported in other previous studies [13], [15], [25], [26]. This contrast could be explained by the large differences existing among the works. To this respect, Badr *et al.* [25] successfully identified academic performance for 65% of students enrolled in a programming course, but only grades on previous subjects of math and english were combined by a naive approach based on association rules. Similarly, a simple decision tree, built with several features from the student's background, reported values of Acc lower than 68% when predicting student's scores on final theory and lab examinations for a course in basic programming [26]. Also, using cumulative grade point average and individual scores on four pre-requisites courses for a subject in engineering dynamics as input variables, Huang and Fang [15] yielded values of Acc about 60% with several classifiers. Finally, Howard *et al.* [13] noticed large values of mean absolute error when the end-of-semester student's score on a blended course in statistic was estimated through regression techniques.

As expected, student's grades on the first-year courses more closely related to *Power Electronic Systems* reported more predictive information than those on the remaining past modules. Thus, only student's marks on ECC and EDS achieved values of Acc and Se between 70 and 76% (see Table III, scenario #1). These two courses introduce learners to the electrical circuit theory and operation of some basic components and subsystems, such as resistors, capacitors, inductors, diodes, transistors, and operational amplifiers. Additionally, when student's score on FPII was considered as input variable, both Acc and Se significantly increased by about 5% in scenario #2. This outcome is not surprising, since this module covers the electromagnetism theory, which plays a key role in most power electronic circuits [16]. The remaining features extracted from the student's background did not make predictive models more accurate and were hence discarded.

When this set of three attributes (i.e., grades on ECC, EDS, and FPII) was expanded with variables from the student's continuous assessment, values of Acc and Se larger than 85% were noticed in CP2 (see Table IV, scenario #2). Although comparison should be carefully established, because course characteristics, prediction methods, and input variables change from study to study, this outcome is similar to that reported by other previous works [7], [12]–[14], [28], [29], as will be discussed at the end of Section IV-B. Of note is also that, for every CP, no relevant differences were seen among the performance reported by all the classifiers. However, a Friedman test has suggested that some algorithms globally performed better than others, the ranking from the best performer to the lowest one being PEC, RF, WVE, MVE AB, SVM, KNN, MLP, NBC, LR, and DT.

These outcomes can be assessed in the light of the well-known dilemma of bias-variance in machine learning [37]. Whereas bias can be considered as the error provoked by inaccurate assumptions in the classifier, variance is the algorithm sensitivity to small changes in the training dataset. Both components determine the classification error for every algorithm, and are moreover interlinked, such that reducing one will result in increasing the other [37]. Hence, the limited differences in values of Acc and Se observed for all the classifiers suggest that, in general terms, they reached a good trade-off between bias and variance. Nonetheless, those with low bias and high variance (i.e., SVM, KNN, and MLP) performed slightly better than those with high bias and low variance (i.e., LR, and NBC). This finding points out that LR and NBC are too simple to capture all underlying data patterns, because algorithms presenting a simple structure often exhibit high bias and low variance [7], [37]. On the other hand, although DT is also characterized by low bias and high variance, its performance was slightly worse than the one reported by SVM, KNN, and MLP. This outcome could be due to the high variance presented by the classifier, thus requiring a larger dataset for its optimal training [7], [37]. Indeed, RF and AB have provided a significantly better performance than DT. These homogeneous ensembles are characterized by a reduced variance, since they are based on combining several trees.

Taking advantage of the variance reduction achieved by combining all the basic classifiers, MVE and WVE also provided a very similar performance to RF and AB. Nonetheless, these heterogeneous ensembles also exploit the diversity in the classification results obtained by the basic algorithms [43], [44]. To this respect, no relevant differences were observed among the methods in terms of Se, but slightly higher values were seen for SVM, MLP, and KNN than for LR, NBC, and DT. Moreover, LR, NBC, and DT trended to better work in predicting students passing the course. Hence, this complementarity between both groups of techniques could also explain the improvement exhibited by MVE and WVE to predict at-risk students regarding the basic classifiers.

Despite the highly accurate predictions provided by the four ensembles, the PEC was the best performer. As Table VIII shows, its performance was statistically better than most of the remaining algorithms. Moreover, to the best of our knowledge, ranks obtained by a Friedman test have not been previously used in any field to weigh each base model composing a voting ensemble classifier. This novel combination strategy is able to exploit information about the performance of each base classifier on different CPs and diverse subsets of student-related variables (i.e., scenarios), thus preventing overtraining for only one classification task [49]. In fact, for every cross-validation iteration, the basic algorithms were firstly trained on the learning samples in every CP and scenario, the weights w_i were then computed as described in Section II-D, and finally the PEC was separately validated on the testing samples in each one of the 15 cases under analysis (i.e., five CPs and three scenarios). In so doing, overtraining on a single scenario was avoided and unbiased classification in every conducted analysis was achieved. This broad-based combination strategy could explain the better results obtained by the PEC in every CP and scenario with regard to the remaining classifiers (see Tables III–VII). In addition, because no heterogeneous ensemble classifiers based on weighted voting have been previously proposed for student's performance prediction, the PEC could be considered as a relevant contribution of the present work.

Finally, remark that only variations about 3–4% between values of Acc and Se can be seen in Tables III–VII for all the conducted analyses and tested classifiers. The fact that the dataset was highly balanced, presenting a similar number of students passing and failing the course, could explain such minor differences. Although a predictive model maximizing right identification of at-risk students (i.e., Se) is pursued, Acc still has to be maintained as high as possible to avoid too many false alarms [7], [50]. To this respect, misclassification of passing students is not critical, because they would benefit from receiving additional teaching support to improve their performance. However, if this prediction error is too high, many resources could be wasted on students who do not need that special attention [34].

B. What Is the Optimal Time Point to Predict At-risk Students Along the Course?

A key goal of the present study was to determine the optimal time to implement an EWS able to provide students with a real chance to improve their performance. As previously described, values of Acc and Se between 70 and 81% were reported only making use of attributes extracted from the student's background. However, these predictions still misclassified roughly one quarter of students at risk of failing the course (see Table III). Interestingly, the inclusion of features from the student's continuous assessment in the predictive models increased values of Acc and Se by about 20–25%, thus reaching rates larger than 93.5% in CP5 (see Table VII). This good outcome is in line with other previous works [13], [28], and could be explained by the fact that the end-of-semester grade is computed as a linear combination of the student's scores on the course activities and exams [7]. Nonetheless, predictions at week 13 could be too late to start effective instructional measures for low-performance students [13].

Comparing the prediction results obtained in all CPs, it might be noticed that there is little benefit in waiting beyond the forth week to predict at-risk students. In fact, whereas a statistically significant increase in values of Acc and Se by about 10–12% was observed from CP1 to CP2, only improvements between 0.5 and 4% were seen later (see Tables III–VII). This small, non-significant rise in values of Acc and Se for CP3–CP5 could not overcome the additional effort required to collect data after week 4, as well as the delay in predicting student's performance. On the other hand, although results have not been presented in Section III, a non-statistically significant increase in values of Acc and Se by about 1–2% in week 2 and a statistically significant rise by about 3–5% in week 3 were noticed regarding CP1 (i.e., before the course starts). However, compared with week 3, statistically significant improvements about 6–9% were still observed in CP2 (i.e., in week 4). As a consequence, the end of the fourth week seems to be the optimal time point to implement an EWS in the examined course, *Power Electronic Systems*. This finding is consistent with the fact that the fundamental notions used throughout the semester to describe most power electronic circuits and systems are covered until that moment. Hence, students presenting difficulties in this stage will be potential candidates to fail the course.

An outcome in the same line has also been reported by other previous works. After building several regression models on a week-by-week cumulative basis, Howard *et al.* [13] have concluded that the optimal time to predict at-risk students in a blended course in statistics was week 5. Also, making use of a similar approach, Lu *et al.* [14] have reported week 6 as the critical point to establish an EWS in a blended course in calculus. For a conventional module in engineering basis, Marbouti *et al.* [7], [28] have also found that about 90% of low-performance learners can be correctly identified at week 4. Similarly, Costa *et al.* [12] have obtained predictive abilities between 80 and 90% in the fourth week of a face-to-face course in programming, when data were preprocessed and classifiers were fine-tuned. Finally, feeding regression techniques with student's scores on activities completed in the first quarter of a course in mechanics of materials, Sadati and Libre [29] have also shown an accurate estimation of the student's end-of-semester grade.

C. Can Student's Performance Progression on Weekly Assignments Provide Useful Predictive Information?

According to some recent works [7], [13], [15], [28], the previously described improvement in the prediction results from CP1 to CP2–CP5 also highlights the relevant role that continuous assessment plays in predicting at-risk students in face-to-face courses. This kind of assessment has proven to be effective in engaging learners over the entire course, thus limiting the number of those only studying during the weeks before mid-term or final exams [12], [31]. Hence, even if weekly assignments are graded with a lower weight than in the present study (i.e., 50%), they could still contribute to improve early identification of at-risk students. Nonetheless, whereas cumulative and average grades on successive tasks have been traditionally used as input variables [7], [13], [15], [28], the novel features proposed to quantify the student's performance progression on weekly assignments have revealed more predictive information. In fact, compared to mean marks on ICAs, HAs, and LAs, the proposed indices $\Delta ICA_{\mathcal{W}}$, $\Delta HA_{\mathcal{W}}$, and $\Delta LA_{\mathcal{W}}$ have revealed statistically significant improvements about 5% in values of Acc and Se for every CP and prediction algorithm (see Tables IV–VII).

This outcome might be indicative of a better capability of these novel attributes to quantify the student's learning progression along the semester. To this respect, Figure 2(a) shows how weekly grades on ICAs, HAs, and LAs for passing students are maintained or slightly decreased, thus suggesting strong engagement during the entire course and adequate understanding of the main concepts. Contrarily, a markedly decreasing trend is noticed in Figure 2(b) for failing learners, thus pointing to a loss of interest in the course or a poor progressive learning. In line with these observations, Sadati and Libre [29] have also reported that students more involved in weekly activities become more successful on the final exam. Moreover, the notable difference in the behavior exhibited by both groups of students agrees with the nature of the course under study, where knowledge accumulation plays a key role. Indeed, if students do not understand the main principles about three-phase systems and how some basic circuits work (e.g., transformers), they will probably fail at learning more advanced systems, such as electrical energy converters [16].

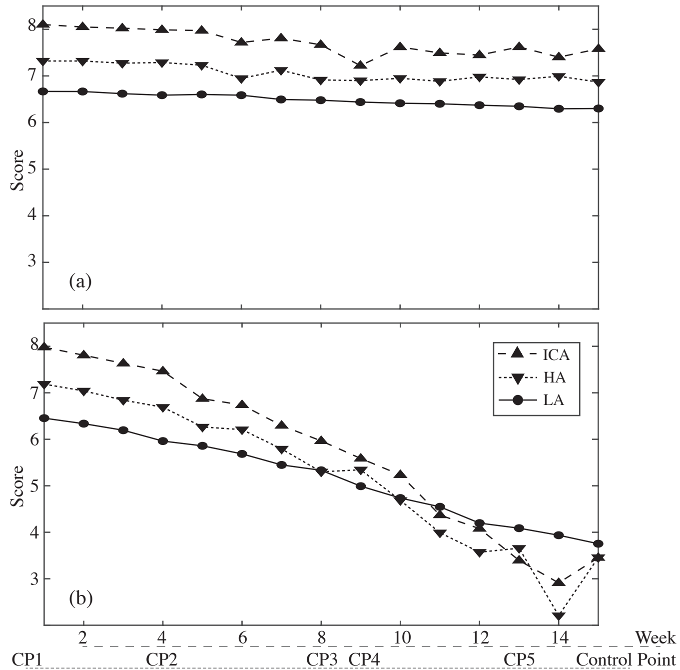

Fig. 2. Weekly mean scores on ICAs, HAs, LAs for students (a) passing and (b) failing the course.

It should also be noted that most predictive student-related attributes in each CP were manually derived, because previous works have suggested that student's performance predictions can be maximized when domain knowledge is used as support to select the best performing set of input data [27]. Moreover, a broad variety of previous works have also made use of this approach [8], [15], [29], [51]. Nonetheless, automatic selection of features in each CP was also explored. To this respect, a sequential forward selection approach was used to choose the subset of input variables minimizing the prediction error for each classifier. As expected, the selected features were always the same as those derived manually. More precisely, the inputs automatically chosen in CP1 were student's grades on ECC, EDS, and FPII. In addition to these variables, the proposed attributes $\Delta ICA_{\mathcal{W}}$, $\Delta HA_{\mathcal{W}}$, and $\Delta LA_{\mathcal{W}}$, along with student's mark on the MTE (if possible), were selected for the remaining CPs.

D. Other Findings About Input and Output Variables

The features extracted from the student's continuous assessment were part of the end-of-semester grade, which could have impacted on the classification results described in previous sections. Hence, all the conducted analyses were also repeated using student's grade on the final exam as output variable. The obtained outcomes in this case are presented in the Appendix.

Briefly, regarding the use of the end-of-semester grade as output variable, a decrease by about 5% was noticed in the values of Acc and Se for every scenario, CP, and classifier. In fact, a paired Wilcoxon signed rank test provided statistically significant differences in all the analyses, when the performance of every classifier was compared for both output variables. Nonetheless, the main findings seen in each CP still remained. Indeed, student's grades on ECC, EDS, and FPII showed to be more predictive than the remaining variables extracted from the student's background (see Table IX). Similarly, the proposed features $\Delta ICA_{\mathcal{W}}$, $\Delta HA_{\mathcal{W}}$, and $\Delta LA_{\mathcal{W}}$ also reported a better performance than the remaining ones used to quantify how student's learning progressed in CP2–CP5 (see Tables X–XIII). Moreover, a Friedman test also provided that the ranking from the best performer to the lowest one continued to be the PEC, RF, WVE, MVE, AB, SVM, KNN, MLP, NBC, LR, and DT. Finally, a post-hoc analysis with Bonferrini correction also yielded a similar distribution of adjusted p-values for every pairwise comparison of the classification algorithms (see Table XIV).

These outcomes are not completely surprising, because the final exam contributes 50% to the end-of-semester grade. Moreover, it is reasonable to think that most students obtaining a good performance in weekly tasks (i.e., ICAs, HWs, and LAs) will have no difficulty in passing the final exam. Nonetheless, considering data from the student's continuous assessment both in input and output variables seems to introduce a bias of about 5% in classification rates, regardless of the used prediction model.

E. Limitations and Future Research

The course-specific perspective used to design the pursued EWS for *Power Electronic Systems* has allowed to reach accurate predictions of at-risk students, but it hampers the resulting system extrapolation to other education scenarios. Nonetheless, the most relevant contributions presented in this work may still be of interest for many researchers. Thus, because this is the first study proposing an EWS for a course dealing with power electronic circuits and systems, other instructors teaching similar subjects would have a reference about how accurate and early predictions of at-risk students can be achieved. Additionally, the set of new features proposed to quantify the weekly progression of the student's learning, i.e., $\Delta ICA_{\mathcal{W}}$, $\Delta HA_{\mathcal{W}}$, and $\Delta LA_{\mathcal{W}}$, could also be useful to early predict at-risk students in other STEM and non-STEM courses, especially in those where knowledge accumulation is key to reach the intended learning objectives. Although students in many courses are not required to complete weekly assignments, variation between consecutive grades on regular activities could still be informative of how student's learning progresses along the semester.

Likewise, the PEC could also be useful in other STEM and non-STEM courses, since it exploits global information about how several basic algorithms perform on a variety of scenarios, thus reducing overfitting and reaching strong generalizations. In this case, the analysis of several CPs along the semester and diverse subsets of student-related variables must firstly be analyzed in one or several offerings of the specific course to train the algorithm. However, because this initial analysis is required to design whatever EWS from scratch, it does not involve any limitation for the successful performance of the PEC in other educational contexts.

An important point to emphasize is that all students registered in the course *Power Electronic Systems* from its first offering back in 2010 have been included in the present study. Although a sample size larger than 362 students could be desired to obtain more robust conclusions from a statistical point of view, the analyzed database could still be considered as representative. Indeed, students were collected from several consecutive offerings, but the learning setting always remained approximately constant, as described in Section II-B. Moreover, the dataset seems to be sufficiently large to study the multivariate effect of the small set of features studied here. To this respect, a too small dataset, which could prevent an acceptable performance from a statistical technique, has been defined by a ratio of sample size to number of variables lower than 20 [52], [53]. However, Table II shows that in the present study the greatest number of input variables analyzed in scenarios #1, #2, and #3 for every CP were 7, 7, and 14, respectively, thus that ratio being always notably larger than 20. Nonetheless, students registered in future offerings of the course will be used to validate the PEC in an upcoming work. For this study, the algorithm will be prospectively used without additional training. Although 10 prediction models were obtained during cross-validation, in all iterations the weights w_i obtained for the base classifiers were very similar, and those with the highest repetition frequency will be used to make prospective predictions in future offerings. These weights were 35, 20, 15, 10, 10, and 10% for SVM, KNN, MLP, DT, NBC, and LR, respectively.

On the other hand, all classifiers were trained with hyperparameters manually selected from previous works. However, because their fine-tuning could slightly improve student's performance predictions [12], automatic optimization approaches will be used for that purpose in the future. Additionally, regression techniques have also been used for identifying at-risk students in a variety of courses [13]–[15], [51]. In this case, a continuous variable has to be considered as output, thus final exam or end-of-semester marks being commonly predicted. However, this approach presents an inherent difficulty to discern between two groups of students, because estimation error in the final grade could have a notable impact on classification [54]. Nonetheless, given the promising results obtained by previous works [13], [15], [51], some techniques, including principal component regression, support vector regression, gaussian regression, and regression trees, will be eventually analyzed.

Regarding the input variables used to feed the classifiers, only academic features have been considered in every CP. They are the most common variables in previous works, but in the last years some student's demographic and socio-economic factors are receiving particular attention [4]. Similarly, because learning is a complex process involving a variety of psychological factors, such as self-efficacy, emotional state, motivation, and interest (among others), these have been recently suggested to improve identification of at-risk students in some courses [4], [55]. Hence, although most of these non-academic factors are not collected in daily teaching practice and time-consuming surveys are required [12], [56], some will be analyzed in future offerings of the course. Moreover, data from student's interaction with the used e-learning platform will also be included in these further studies.

Finally, note that identification of at-risk students is only the first step to implement an EWS, and further research is required about what kind of support should be provided to low-performance students. To this respect, some experiments will be promptly conducted with instructional interventions, such as review of key concepts after class via interactive readings, and one-on-one tutoring attendance for 15 minutes per week.

V. Conclusions

With the goal of establishing a tailored EWS, the ability of some common classifiers to predict students at risk of failing a conventional course in power electronic systems has been analyzed. Although no remarkable differences in the performance of all algorithms have been observed, homogeneous and heterogeneous ensemble classifiers have reported statistically better predictions than other more basic techniques. Nonetheless, a heterogeneous ensemble classifier based on a novel combination strategy of several basic classifiers by weighted majority voting has shown to be the best performer, reaching classification results comparable to previous works dealing with courses from other fields. Additionally, the end of the forth week has been identified as the optimal time in the semester to make predictions, thus leaving sufficient time to support low-performance students with instructional measures. At that moment, values of accuracy about 87–90% and 81–85% have been obtained when the end-of-semester and final exam marks have been used as output variables, respectively. To reach these outcomes, a set of novel features quantifying student's performance progression on weekly assignments have played a key role. In fact, they have revealed more predictive information than student's average grades typically used in continuous assessment. The time-dependent information estimated by these novel input variables could also be useful in other STEM and non-STEM courses, because they seem to be able to closely estimate student's learning progression along the course.

Appendix
Results for Final Exam Score as Output Variable

As described in Section II-B of the paper, students passed the course if they reached an end-of-semester grade equal or higher than 5 points, and a score on the final exam equal or greater than 4 points. Hence, this last value was considered to generate labels when student's grade on the final exam was used as output variable. Accordingly, 188 students passed the final exam and the remaining 174 failed it.

The prediction outcomes obtained by the models built in CP1–CP5 are shown in Tables IX–XIII, respectively. Moreover, because a Friedman test rejected the null hypothesis, a post-hoc analysis with Bonferroni correction was conducted and adjusted values of statistical significance are presented in Table XIV.

TABLE IX
Prediction Results Obtained by the Models Built in CP1 When the Student's Grade on the Final Exam Was Used As Output Variable

Algorithm	Scenario #1		Scenario #2		Scenario #3	
	Acc	Se	Acc	Se	Acc	Se
LR	0.654	0.659	0.729	0.728	0.725	0.742
NBC	0.664	0.670	0.721	0.735	0.725	0.701
DT	0.668	0.672	0.731	0.740	0.737	0.737
SVM	0.673	0.685	0.739	0.746	0.732	0.715
MLP	0.671	0.675	0.726	0.738	0.728	0.743
KNN	0.666	0.672	0.732	0.742	0.732	0.748
RF	0.693	0.675	0.747	0.750	0.749	0.750
AB	0.690	0.699	0.739	0.748	0.745	0.740
MVE	0.673	0.689	0.731	0.735	0.734	0.750
WVE	0.681	0.685	0.729	0.731	0.737	0.754
PEC	**0.708**	**0.713**	**0.768**	**0.763**	**0.768**	**0.756**

TABLE X
Prediction Results Obtained by the Models Built in CP2 When Student's Grade on the Final Exam Was Used As Output Variable

Algorithm	Scenario #1		Scenario #2		Scenario #3	
	Acc	Se	Acc	Se	Acc	Se
LR	0.769	0.778	0.819	0.826	0.822	0.814
NBC	0.785	0.751	0.815	0.817	0.818	0.800
DT	0.762	0.747	0.828	0.820	0.820	0.824
SVM	0.781	0.784	0.828	0.838	0.835	0.847
MLP	0.774	0.760	0.821	**0.845**	0.816	0.833
KNN	0.766	0.772	0.816	0.824	0.829	0.814
RF	0.789	0.797	0.836	**0.845**	0.839	0.828
AB	0.779	0.794	0.832	0.843	0.832	0.833
MVE	0.791	0.792	0.836	0.826	0.839	0.871
WVE	0.793	0.788	0.839	0.831	0.841	**0.874**
PEC	**0.808**	**0.803**	**0.853**	**0.845**	**0.853**	0.863

TABLE XI
Prediction Results Obtained by the Models Built in CP3 When Student's Grade on the Final Exam Was Used As Output Variable

Algorithm	Scenario #1		Scenario #2		Scenario #3	
	Acc	Se	Acc	Se	Acc	Se
LR	0.828	0.832	0.873	0.852	0.879	0.864
NBC	0.822	0.828	0.873	0.862	0.872	0.883
DT	0.816	0.802	0.870	0.859	0.867	0.860
SVM	0.822	0.822	0.881	0.906	0.886	**0.912**
MLP	0.826	0.834	0.885	0.902	0.881	0.873
KNN	0.826	0.844	0.881	0.864	0.879	0.885
RF	0.839	0.830	0.888	0.899	0.890	0.903
AB	0.832	0.824	0.881	0.893	0.886	0.901
MVE	0.836	0.848	0.881	0.890	0.879	0.881
WVE	0.838	0.846	0.885	0.893	0.876	0.887
PEC	**0.858**	**0.858**	**0.901**	**0.912**	**0.901**	0.908

References

[1] A. Dutt, M. A. Ismail, and T. Herawan, "A systematic review on educational data mining," *IEEE Access*, vol. 5, pp. 15 991–16 005, Jan. 2017.

[2] B. Bakhshinategh, O. R. Zaiane, S. ElAtia, and D. Ipperciel, "Educational data mining applications and tasks: A survey of the last 10 years," *Educ. Inf. Technol.*, vol. 23, no. 1, pp. 537–553, Jan. 2018.

[3] (Apr. 6, 2020) Homepage of the International Educational Data Mining Society. [Online]. Available: http://www.educationaldatamining.org

[4] A. Hellas, P. Ihantola, A. Petersen, V. V. Ajanovski, M. Gutica, T. Hynninen, A. Knutas, J. Leinonen, C. Messom, and S. N. Liao, "Predicting academic performance: A systematic literature review," in *Proc. 23rd Ann. ACM Conf. Innovation and Technology Computer Science Education*, 2018, pp. 175–199.

[5] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining," *Comput. Educ.*, vol. 113, pp. 177–194, Oct. 2017.

[6] P. A. Murtaugh, L. D. Burns, and J. Schuster, "Predicting the retention of university students," *Res. Higher Educ.*, vol. 40, no. 3, pp. 355–371, Jun. 1999.

TABLE XII
PREDICTION RESULTS OBTAINED BY THE MODELS BUILT IN CP4 WHEN STUDENT'S GRADE ON THE FINAL EXAM WAS USED AS OUTPUT VARIABLE

Algorithm	Scenario #1		Scenario #2		Scenario #3	
	Acc	*Se*	*Acc*	*Se*	*Acc*	*Se*
LR	0.845	0.850	0.891	0.884	0.893	0.892
NBC	0.852	0.840	0.893	0.891	0.883	0.871
DT	0.857	0.830	0.891	0.901	0.889	0.894
SVM	0.857	0.887	0.901	0.889	0.907	0.906
MLP	0.852	0.843	0.891	0.903	0.895	0.894
KNN	0.860	0.862	0.896	0.907	0.891	0.897
RF	0.852	0.882	0.901	0.910	0.909	0.908
AB	0.858	**0.897**	0.907	0.887	0.906	0.908
MVE	0.863	0.840	0.905	0.922	0.902	0.899
WVE	0.857	0.850	0.907	**0.929**	0.902	0.897
PEC	**0.874**	0.865	**0.913**	0.918	**0.915**	**0.925**

TABLE XIII
PREDICTION RESULTS OBTAINED BY THE MODELS BUILT IN CP5 WHEN STUDENT'S GRADE ON THE FINAL EXAM WAS USED AS OUTPUT VARIABLE

Algorithm	Scenario #1		Scenario #2		Scenario #3	
	Acc	*Se*	*Acc*	*Se*	*Acc*	*Se*
LR	0.862	0.835	0.898	0.901	0.896	0.871
NBC	0.862	0.878	0.898	0.885	0.898	0.902
DT	0.859	0.869	0.889	0.894	0.891	0.889
SVM	0.869	0.885	0.901	0.906	0.904	0.906
MLP	0.862	0.858	0.901	0.899	0.901	0.902
KNN	0.864	0.871	0.904	0.894	0.894	0.906
RF	0.871	0.890	0.916	0.904	0.917	0.909
AB	0.867	0.888	0.911	0.909	0.907	0.904
MVE	0.871	0.844	0.908	0.911	0.904	0.915
WVE	0.869	0.851	0.908	0.906	0.909	0.918
PEC	**0.882**	**0.901**	**0.931**	**0.933**	**0.932**	**0.930**

[7] F. Marbouti, H. A. Diefes-Dux, and K. Madhavan, "Models for early prediction of at-risk students in a course using standards-based grading," *Comput. Educ.*, vol. 103, pp. 1–15, Dec. 2016.

[8] A. Sandoval, C. Gonzalez, R. Alarcon, K. Pichara, and M. Montenegro, "Centralized student performance prediction in large courses based on low-cost variables in an institutional context," *Internet Higher Educ.*, vol. 37, pp. 76–89, Apr. 2018.

[9] I. Lykourentzou, I. Giannoukos, V. Nikolopoulos, G. Mpardis, and V. Loumos, "Dropout prediction in e-learning courses through the combination of machine learning techniques," *Comput. Educ.*, vol. 53, no. 3, pp. 950–965, Nov. 2009.

[10] J. Kuzilek, M. Hlosta, D. Herrmannova, Z. Zdrahal, and A. Wolff, "OU Analyse: analysing at-risk students at The Open University," *Learn. Analytics Rev.*, pp. 1–16, 2015.

[11] C. Burgos, M. L. Campanario, D. de la Pena, J. A. Lara, D. Lizcano, and M. A. Martínez, "Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout," *Comput. Elect. Eng.*, vol. 66, pp. 541–556, Feb. 2018.

[12] E. B. Costa, B. Fonseca, M. A. Santana, F. F. de Araújo, and J. Rego, "Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses," *Comput. Human Behavior*, vol. 73, pp. 247–256, Aug. 2017.

[13] E. Howard, M. Meehan, and A. Parnell, "Contrasting prediction methods for early warning systems at undergraduate level," *Internet Higher Educ.*, vol. 37, pp. 66–75, Apr. 2018.

[14] O. H. Lu, A. Y. Huang, J. C. Huang, A. J. Lin, H. Ogata, and S. J. Yang, "Applying learning analytics for the early prediction of students' academic performance in blended learning," *J. Educational Technol. Soc.*, vol. 21, no. 2, pp. 220–232, Apr. 2018.

[15] S. Huang and N. Fang, "Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models," *Comput. Educ.*, vol. 61, pp. 133–145, Feb. 2013.

[16] P. Bauer and V. Fedak, "Philosophy of interactive e-learning for power electronics and electrical drives: A way from ideas to realization," *J. Power Electron.*, vol. 10, no. 6, pp. 587–594, Nov. 2010.

[17] R. Chibante, C. Vaz De Carvalho, and P. Ferreira, "A simulation tool to promote active learning of controlled rectifiers," *Comput. Appl. Eng. Educ.*, vol. 26, no. 3, pp. 688–699, Feb. 2018.

[18] S. M. Jayaprakash, E. W. Moody, E. J. Lauría, J. R. Regan, and J. D. Baron, "Early alert of academically at-risk students: An open source analytics initiative," *J. Learn. Analytics*, vol. 1, no. 1, pp. 6–47, May 2014.

[19] A. Essa and H. Ayad, "Student success system: risk analytics and data visualization using ensembles of predictive models," in *Proc. 2nd Int. Conf. Learning Analytics and Knowledge*, 2012, pp. 158–161.

[20] K. E. Arnold and M. D. Pistilli, "Course signals at purdue: Using learning analytics to increase student success," in *Proc. 2nd Int. Conf. Learning Analytics and Knowledge*, 2012, pp. 267–270.

[21] Y. Cui, F. Chen, and A. Shiri, "Scale up predictive models for early detection of at-risk students: a feasibility study," *Inf. Learn. Sci.*, vol. 121, no. 3/4, pp. 97–116, 2020.

[22] L. P. Macfadyen and S. Dawson, "Mining LMS data to develop an early warning system for educators: A proof of concept," *Comput. Educ.*, vol. 54, no. 2, pp. 588–599, Feb. 2010.

[23] D. Gašević, S. Dawson, T. Rogers, and D. Gasevic, "Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success," *Internet Higher Educ.*, vol. 28, pp. 68–84, Jan. 2016.

[24] A. Wolff, Z. Zdrahal, A. Nikolov, and M. Pantucek, "Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment," in *Proc. 3rd Int. Conf. Learning Analytics and Knowledge*, 2013, pp. 145–149.

[25] G. Badr, A. Algobail, H. Almutairi, and M. Almutery, "Predicting students' performance in university courses: a case study and tool in KSU mathematics department," *Procedia Comput. Sci.*, vol. 82, pp. 80–89, May 2016.

[26] M. Ayub and O. Karnalim, "Predicting outcomes in introductory programming using J48 classification," *World Trans. Eng. Technol. Educ.*, vol. 15, no. 2, pp. 132–136, Jul. 2017.

[27] N. Tomasevic, N. Gvozdenovic, and S. Vranes, "An overview and comparison of supervised data mining techniques for student exam performance prediction," *Comput. Educ.*, vol. 143, p. 103676, Jan. 2020.

[28] M. F. Marbouti and H. A. Diefes-Dux, "Building course-specific regression-based models to identify at-risk students," in *Proc. 122nd ASEE Anun. Conf. and Expo.*, 2015, p. 11229.

[29] S. Sadati and N. A. Libre, "Development of an early alert system to predict students at risk of failing based on their early course activities," in *Proc. 124th ASEE Ann. Conf. and Expo.*, 2017, p. 18736.

[30] J. S. Cole and S. W. Spence, "Using continuous assessment to promote student engagement in a large class," *Eur. J. Eng. Educ.*, vol. 37, no. 5, pp. 508–525, Sep. 2012.

[31] N. Holmes, "Student perceptions of their learning and engagement in response to the use of a continuous e-assessment in an undergraduate module," *Assessment Eval. Higher Educ.*, vol. 40, no. 1, pp. 1–14, Jan. 2015.

[32] F. Ofori, E. Maina, and R. Gitonga, "Using machine learning algorithms to predict students' performance and improve learning outcome: A literature based review," *J. Inf. Technol.*, vol. 4, no. 1, pp. 33–55, Mar. 2020.

[33] G. Bonaccorso, *Machine learning algorithms*. Birmingham, UK: Packt Publishing Ltd, 2017.

[34] Y.-H. Hu, C.-L. Lo, and S.-P. Shih, "Developing early warning systems to predict students' online learning performance," *Comput. Human Behavior*, vol. 36, pp. 469–478, Jul. 2014.

[35] M. Mimis, M. El Hajji, Y. Es-Saady, A. O. Guejdi, H. Douzi, and D. Mammass, "A framework for smart academic guidance using educational data mining," *Educ. Inf. Technol.*, vol. 24, no. 2, pp. 1379–1393, Mar. 2019.

[36] J. D. Rennie, L. Shih, J. Teevan, and D. R. Karger, "Tackling the poor assumptions of naive bayes text classifiers," in *Proc. 20th Int. Conf. Machine Learning*, 2003, pp. 616–623.

[37] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers, 2005.

[38] C. Romero and S. Ventura, "Educational data mining: a review of the state of the art," *IEEE Trans. Systems, Man, and Cybernetics, Part C: Appl. and Rev.*, vol. 40, no. 6, pp. 601–618, Nov. 2010.

[39] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," *Artif. Intell. Rev.*, vol. 26, no. 3, pp. 159–190, Nov. 2006.

TABLE XIV
ADJUSTED p-VALUES OBTAINED BY A POST HOC ANALYSIS CONDUCTED ON RESULTS OF A FRIEDMAN TEST. STUDENT'S GRADE ON THE FINAL EXAM WAS USED AS OUTPUT VARIABLE

Algorithm	LR	NBC	DT	SVM	MLP	KNN	RF	AB	MVE	WVE	PEC
LR	—	7.412	8.905	0.064	3.637	1.958	<0.001	0.002	0.003	0.001	<0.001
NBC	7.412	—	8.472	0.023	2.155	1.044	<0.001	<0.001	0.001	<0.001	<0.001
DT	8.905	8.472	—	0.042	2.956	1.524	<0.001	0.001	0.001	0.001	<0.001
SVM	0.064	0.023	0.042	—	0.693	1.524	0.297	2.831	3.784	2.477	0.001
MLP	3.637	2.155	2.956	0.693	—	6.999	0.001	0.039	0.070	0.030	<0.001
KNN	1.958	1.044	1.524	1.524	6.999	—	0.003	0.123	0.208	0.098	<0.001
RF	<0.001	<0.001	<0.001	0.297	0.001	0.003	—	2.709	1.958	3.085	0.879
AB	0.002	<0.001	0.001	2.831	0.039	0.123	2.709	—	8.472	9.342	0.050
MVE	0.003	0.001	0.001	3.784	0.070	0.208	1.958	8.472	—	7.831	0.027
WVE	0.001	<0.001	0.001	2.477	0.030	0.098	3.085	9.342	7.831	—	0.064
PEC	<0.001	<0.001	<0.001	0.001	<0.001	<0.001	0.879	0.050	0.027	0.064	—

[40] G. P. Kumari and M. U. Rani, "A study of AdaBoost and bagging approaches on student dataset," *Int. Res. J. Adv. Eng. Sci.*, vol. 2, no. 2, pp. 375–380, 2017.

[41] M. M. Idrees, L. L. Minku, F. Stahl, and A. Badii, "A heterogeneous online learning ensemble for non-stationary environments," *Knowl.-Based Syst.*, vol. 188, p. 104983, 2020.

[42] J. L. Fernández-Alemán, J. M. Carrillo-de Gea, M. Hosni, A. Idri, and G. García-Mateos, "Homogeneous and heterogeneous ensemble classification methods in diabetes disease: a review," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.*, 2019, pp. 3956–3959.

[43] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Mach. Learn.*, vol. 51, no. 2, pp. 181–207, 2003.

[44] L. I. Kuncheva and J. J. Rodríguez, "A weighted voting framework for classifiers ensembles," *Knowl. Inf. Syst.*, vol. 38, no. 2, pp. 259–275, 2014.

[45] A. Husin, "Designing multiple classifiers combination. A survey," *J. Theor. Appl. Inf. Technol.*, vol. 97, no. 20, pp. 2386–2405, 2019.

[46] Y. Jung and J. Hu, "A K-fold averaging cross-validation procedure," *J. Nonparametric Statist.*, vol. 27, no. 2, pp. 167–179, Feb. 2015.

[47] J. Derrac, S. García, D. Molina, and F. Herrera, "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms," *Swarm Evol. Computation*, vol. 1, no. 1, pp. 3–18, Mar. 2011.

[48] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian J. Statist.*, vol. 6, no. 2, pp. 65–70, 1979.

[49] R. P. Duin, "The combining classifier: to train or not to train?" in *Object recognition supported by user interaction for service robots*, vol. 2, Quebec City, Quebec, Canada, 2002, pp. 765–770.

[50] W. Chen, C. G. Brinton, D. Cao, A. Mason-singh, C. Lu, and M. Chiang, "Early detection prediction of learning outcomes in online short-courses via learning behaviors," *IEEE Trans. Learn. Technol.*, vol. 12, no. 1, pp. 44–58, Jan. 2019.

[51] G. Kostopoulos, S. Kotsiantis, C. Pierrakeas, G. Koutsonikos, and G. A. Gravvanis, "Forecasting students' success in an open university," *Int. J. Learn. Technol.*, vol. 13, no. 1, pp. 26–43, May 2018.

[52] A. Zollanvari, R. C. Kizilirmak, Y. H. Kho, and D. Hernández-Torrano, "Predicting students' GPA and developing intervention strategies based on self-regulatory learning behaviors," *IEEE Access*, vol. 5, pp. 23 792–23 802, Aug. 2017.

[53] V. N. Vapnik, *Statistical Learning Theory*. New York, USA: Wiley & Sons-Interscience, 1998.

[54] Z.-H. Zhou and M. Li, "Semisupervised regression with cotraining-style algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 11, pp. 1479–1493, Nov. 2007.

[55] L. Yu, C. Lee, H. Pan, C. Chou, P. Chao, Z. Chen, S. Tseng, C. Chan, and K. Lai, "Improving early prediction of academic failure using sentiment analysis on self-evaluated comments," *J. Comput. Assisted Learn.*, vol. 34, no. 4, pp. 358–365, Mar. 2018.

[56] R. Aldunate and M. Nussbaum, "Teacher adoption of technology," *Comput. Human Behavior*, vol. 29, no. 3, pp. 519–524, May 2013.

Raúl Alcaraz received the M.Sc. degree in Electronic Eng. from the University of Alcalá de Henares, Madrid, Spain, in 2005, and the Ph.D. degree in Electronic Eng. from the Technical University of Valencia, Spain, in 2008. He is an Associate Professor at the Department of Electrical, Electronic, Automatic Control and Communications, University of Castilla-La Mancha, Spain. His research interests include biomedical signal processing and educational data mining.

Arturo Martínez-Rodrigo received the M.Sc. degree in Telecommunications Eng. from the University of Castilla-La Mancha, Spain, in 2010, and the Ph.D. degree in Medical Care Research from the University of Castilla-La Mancha, Spain, in 2013. Since 2015 he is Assistant Professor at the department of Computer Science, University of Castilla-La Mancha, Spain. He has taught several subjects related to computation, statistics, and communications. His research interests include signal processing applied to biomedical signal and education.

Roberto Zangróniz (M'03) received the M.Sc. degree in Electronic Eng. from the University of Valladolid, Spain, in 1998, and the Ph.D. degree in Advanced Computing Eng. from the Universidad de Castilla-La Mancha, Spain, in 2019. He is currently an Associate Professor at the Department of Electrical, Electronic, Automatic Control and Communications, University of Castilla-La Mancha, Spain. His research interests include biomedical signal processing and educational data mining.

José Joaquín Rieta received the M. Eng. degrees in Image and Sound Eng. from the Universidad Politécnica de Madrid, Spain, in 1991, the M.Sc. degree in Telecommunications and the Ph.D. degree in Biomedical Signal Processing from the Universidad Politécnica de Valencia, Spain, in 1996 and 2003, respectively. Since 1994 he has been a lecturer at the Electronic Engineering Department of the Universidad Politécnica de Valencia, where he is now Full Professor in Biomedical Engineering. He has taught several subjects related to Electronic and Biomedical Instrumentation, and has been the author of several academic publications in these areas.