Computers & Education 61 (2013) 133–145

Contents lists available at SciVerse ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models

Shaobo Huang, Ning Fang*

Department of Engineering Education, Utah State University, Logan, UT 84322, USA

ARTICLE INFO

Article history:
Received 11 February 2012
Received in revised form
29 August 2012
Accepted 30 August 2012

Keywords:
Applications in subject areas
Postsecondary education
Teaching/learning strategies

ABSTRACT

Predicting student academic performance has long been an important research topic in many academic disciplines. The present study is the first study that develops and compares four types of mathematical models to predict student academic performance in engineering dynamics – a high-enrollment, high-impact, and core course that many engineering undergraduates are required to take. The four types of mathematical models include the multiple linear regression model, the multilayer perception network model, the radial basis function network model, and the support vector machine model. The inputs (i.e., predictor variables) of the models include student's cumulative GPA, grades earned in four pre-requisite courses (statics, calculus I, calculus II, and physics), and scores on three dynamics mid-term exams (i.e., the exams given to students during the semester and before the final exam). The output of the models is students' scores on the dynamics final comprehensive exam. A total of 2907 data points were collected from 323 undergraduates in four semesters. Based on the four types of mathematical models and six different combinations of predictor variables, a total of 24 predictive mathematical models were developed from the present study. The analysis reveals that the type of mathematical model has only a slight effect on the average prediction accuracy (APA, which indicates on average how well a model predicts the final exam scores of all students in the dynamics course) and on the percentage of accurate predictions (PAP, which is calculated as the number of accurate predictions divided by the total number of predictions). The combination of predictor variables has only a slight effect on the APA, but a profound effect on the PAP. In general, the support vector machine models have the highest PAP as compared to the other three types of mathematical models. The research findings from the present study imply that if the goal of the instructor is to predict the average academic performance of his/her dynamics class as a whole, the instructor should choose the simplest mathematical model, which is the multiple linear regression model, with student's cumulative GPA as the only predictor variable. Adding more predictor variables does not help improve the average prediction accuracy of any mathematical model. However, if the goal of the instructor is to predict the academic performance of individual students, the instructor should use the support vector machine model with the first six predictor variables as the inputs of the model, because this particular predictor combination increases the percentage of accurate predictions, and most importantly, allows sufficient time for the instructor to implement subsequent educational interventions to improve student learning.

1. Introduction

1.1. Importance and difficulty of engineering dynamics

Engineering dynamics is a high-enrollment, high-impact, and core course that many engineering undergraduates are required to take. This sophomore-level course covers numerous foundational engineering concepts (e.g., rectilinear and curvilinear motion, displacement and velocity, force and acceleration, work and energy, impulse and momentum, and vibrations), and encompasses fundamental building

* Corresponding author. Tel.: +1 435 7972948; fax: +1 435 7979093.
E-mail address: ning.fang@usu.edu (N. Fang).

0360-1315/$ – see front matter
http://dx.doi.org/10.1016/j.compedu.2012.08.015

blocks that are essential for advanced studies in subsequent courses such as machine design, advanced structural design, and advanced dynamics (Beer et al., 2009; Hibbeler, 2009).

However, dynamics is also widely regarded as "one of the most difficult courses that engineering students encounter during their undergraduate study" (Magill, 1997). To succeed in this difficult course, students must not only have a deep conceptual understanding of abstract concepts that underlie various dynamics problems, but also must have solid mathematical modeling skills to generate correct solutions to those problems (Njock-Libii, 2010; Self, Wood, & Hansen, 2004). Barrett et al. (2010) reported that, in the 2009 standard Fundamentals of Engineering examination given in the USA, the national average score on the dynamics exam was only 53%. In a recent survey conducted at Utah State University by the authors of this paper, students were asked to share their perspectives about dynamics. More than 60% of the students surveyed used phrases such as "much harder than statics," "extremely difficult," "very challenging," and "I am afraid of it." Students often drop out of engineering because they fail dynamics, the last pre-professional gateway course before entering a professional engineering program.

1.2. Usefulness of predicting student academic performance

Predicting student academic performance has long been an important research topic in many academic disciplines (e.g., Cohen, Manion, & Morrison, 2007; Grudnitski, 1997; Pokay & Blumenfeld, 1990; Ransdell, 2001; Ting, 2001). Based on the results of a predictive model, the instructor can take proactive measures (Veenstra, Dey, & Herrin, 2008; Ware & Galassi, 2006) to improve student learning, especially for those low-performance students. For example, if a model predicts that a student in a class would earn a score below 50 (out of 100) in the final comprehensive exam, the student would be identified as "potentially" low performance. The instructor can then take proactive measures and implement effective instructional interventions, such as one-on-one tutoring and review of important concepts after class, assigning extra technical problems, providing remedial lessons, and asking the student to review previously learned concepts in pre-requisite courses.

The results of a predictive model can also be used to encourage those "potentially" low-performance students to develop a better learning strategy. The prediction results might help students develop a good understanding of how well, or how poorly, they would perform in a course; and therefore "force" students to rethink the way in which they have been learning (McKeachie, Pintrich, Lin, & Smith, 1986).

1.3. Mathematical techniques used in predictive modeling

A variety of mathematical techniques have been employed in predictive modeling, including both traditional linear regression (e.g., Ayan & Garcia, 2008; Cios, Pedrycz, & Swiniarski, 1998) and modern data-mining techniques (e.g., Imbrie, Lin, Reid, & Malyscheff, 2008; Lykourentzou, Giannoukos, & Mpardis, 2009; Vandamme, Meskens, & Superby, 2007). The Appendix A provides a detailed description of four types of representative mathematical techniques employed in this study. A mathematical model generally consists of a set of mathematical formulas that describe the quantitative (i.e., numerical) relationships between dependent variables (i.e., outputs) and independent variables (i.e., inputs, or predictor variables). The model is validated if it makes accurate predictions, that is, the error between the predicted and actual values is within a certain, pre-defined small range.

For example, Lykourentzou et al. (2009) employed neural network and multiple linear regression techniques to predict student achievement in e-learning courses. In their study, students took four multiple-choice tests: mc1, mc2, mc3, and mc4. The dataset of 27 students (or 85% of the class) in a 2006 semester were used to train the models, and another dataset of 25 students in a 2007 semester were used to validate the models. They found that, in terms of the mean absolute error, predictions from the neural network models were more accurate than those of the multiple linear regression models.

In another example, Vandamme et al. (2007) made early prediction of students' academic success in the first academic year. A total of 533 students from three universities were classified into three achievement categories: low-risk, medium-risk, and high-risk students. The mathematical techniques used in the Vandamme et al. (2007) study included decision trees, neural networks, and linear discriminant analysis. Their results showed that linear discriminant analysis had the highest rate of correct classifications based on the collected samples. However, none of the three models had a high rate of correct classification. They found that a larger sample size was needed to increase the rate of correct classification for each model.

1.4. Innovation of the present study

The overall goal of the present study is to predict student academic performance in an engineering dynamics course by developing a set of validated mathematical models and then identifying the most appropriate model(s) for use in prediction. Four types of mathematical modeling techniques [multiple linear regression (MLR), multilayer perception (MLP) network, radial basis function (RBF) network, and support vector machine (SVM)] and six combinations of predictor variables were used to develop a total of 24 predictive mathematical models based on the dataset collected from 323 undergraduates in four semesters. The outputs of the models are the students' scores on the dynamics final comprehensive exam. The inputs of the models (i.e., predictor variables) are the student's cumulative GPA (X_1); grades earned in four pre-requisite courses: statics (X_2), calculus I (X_3), calculus II (X_4), and physics (X_5); and scores on three dynamics mid-term examinations (X_6, X_7, X_8). The six combinations of predictor variables were:

- X_1 alone
- X_1, X_2, X_3, X_4, and X_5
- X_6 alone
- X_1, X_2, X_3, X_4, X_5, and X_6
- X_1, X_2, X_3, X_4, X_5, X_6, and X_7
- X_1, X_2, X_3, X_4, X_5, X_6, X_7, and X_8

These combinations were selected from a practical consideration of what predictors can be reasonably employed to predict the students' scores on the dynamics final comprehensive exam. For example, one can use the student's cumulative GPA (X_1) alone as the predictor variable, or one can use X_1 as well as the students' grades earned in four pre-requisite courses (X_2, X_3, X_4, and X_5) as the predictor variables. However, it would be unreasonable if one selects X_1 and X_2 as the combination of predictor variables because X_2 corresponds to only one of the four pre-requisite courses.

The four types of mathematical modeling techniques (MLR, MLP, RBF, and SVM) were chosen in the present study because they are most widely employed in engineering research and engineering education research. Data drawn from databases, including the Education Resources Information Center, Science Citation Index, Social Science Citation Index, Engineering Citation Index, Academic Search Premier, the ASEE annual conference proceedings (1995–2011), and the ASEE/IEEE Frontier in Education conference proceedings (1995–2011), were examined. The results show that the present study is the first study that develops and compares four types of mathematical models to predict student academic performance in engineering dynamics. No other engineering educators, or other education researchers, have developed any predictive model for any engineering dynamics course. In the authors' previous work (Fang & Lu, 2010; Huang & Fang, 2010), only two types of mathematical models were dealt with, and the sample size of data collected was limited in up to two semesters only.

1.5. Research questions of the present study

The present study has the following two research questions:

Research question No. 1: Among the four types of predictive mathematical models investigated in the present study, what particular type is the most appropriate for use in the prediction of student academic performance in engineering dynamics?

Research question No. 2: Among the six combinations of predictor variables investigated in the present study, what particular combination is the most appropriate for use in the prediction of student academic performance in engineering dynamics?

In the following sections of the paper, data collection and each predictor variable are described first. Then, the method of data pre-processing is introduced, followed by a descriptive analysis of the collected data. The four types of mathematical models, the six combinations of predictor variables, and the two criteria for assessing the prediction accuracy of the models are described in detail. The limitations of the present study are also discussed. Finally, the answers to the two research questions are summarized at the end of the paper.

2. Data collection and description of predictor variables

Data were collected from a total of 323 undergraduate students who took dynamics in four semesters: 128 students in Semester #1, 58 students in Semester #2, 53 students in Semester #3, and 84 students in Semester #4. The summary of student demographics is shown in Fig. 1, where "MAE" stands for Mechanical and Aerospace Engineering, "CEE" represents Civil and Environmental Engineering, and "Others" include Biological Engineering, General Engineering, Pre-engineering, undeclared, or non-engineering majors. As seen from Fig. 1, the majority of the students were either Mechanical and Aerospace Engineering majors or Civil and Environmental Engineering majors.

For each of the 323 students, nine data points (Y, X_1, X_2, X_3, …, X_8) were collected, where Y is the score on the dynamics final comprehensive exam; X_1 is the cumulative GPA; X_2, X_3, X_4, and X_5 are the grades in four pre-requisite courses, statics, calculus I, calculus II, and physics, respectively; X_6, X_7, and X_8 are the scores on three dynamics mid-term exams. Therefore, a total of $323 \times 9 = 2907$ data points were collected in the present study. These particular variables (Y, X_1, X_2, X_3, …, X_8) are described as follows:

- X_1 (cumulative GPA) is a comprehensive measurement of a student's problem-solving skills.
- X_2 (statics grade) was included because numerous concepts of statics (such as free-body diagram, force equilibrium, and moment equilibrium) are employed in dynamics.
- X_3 and X_4 (calculus I and II grades) measure a student's mathematical skills needed to solve calculus-based dynamics problems.
- X_5 (physics grade) measures a student's basic understanding of physical concepts and principles behind various dynamics phenomena.
- X_6 (score on dynamics mid-term exam #1) measures a student's problem-solving skills concerning "kinematics of a particle" and "kinetics of a particle: force and acceleration."

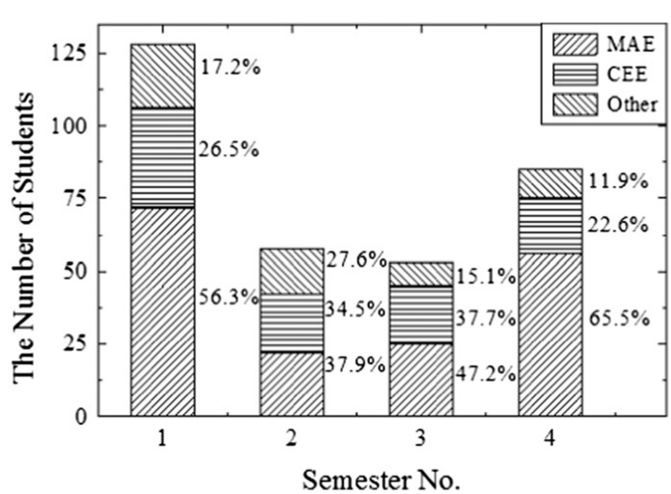

Fig. 1. Student demographics.

Table 1
Normalization of the raw data.

Variables	Initial value of data	Normalized value of data
X_1 cumulative GPA	0.00–4.00 (numerical value)	Initial value/4
X_2 statics grade	Letter grade A, A−, B+, B, etc.	Initial value/4
X_3 calculus I grade	Letter grade A, A−, B+, B, etc.	Initial value/4
X_4 calculus II grade	Letter grade A, A−, B+, B, etc.	Initial value/4
X_5 physics grade	Letter grade A, A−, B+, B, etc.	Initial value/4
X_6 dynamics mid-exam #1 score	0.00–15.00 (numerical value)	Initial value/15
X_7 dynamics mid-exam #2 score	0.00–16.00 (numerical value)	Initial value/16
X_8 dynamics mid-exam #3 score	0.00–15.00 (numerical value)	Initial value/15
Y dynamics final exam score	0.00–100.00 (numerical value)	Initial value/100

- X_7 (score on dynamics mid-term exam #2) measures a student's problem-solving skills concerning "kinetics of a particle: work and energy" and "kinetics of a particle: impulse and momentum."
- X_8 (score on dynamics mid-term exam #3) measures a student's problem-solving skills on "planar kinetics of a rigid body" and "planar kinetics of a rigid body: force and acceleration."

The dynamics final comprehensive exam (Y) covers all the above-listed dynamics topics as well as three additional topics that students learned after mid-term exam #3. The three additional topics included "planar kinetics of a rigid body: work and energy," "planar kinetics of a rigid body: impulse and momentum," and "vibration."

3. Data pre-processing and descriptive analysis

The collected raw data ($Y, X_1, X_2, X_3, \ldots, X_8$) were initially in different scales of measurement: X_1 varies from 0.0 to 4.0; X_2, X_3, X_4, and X_5 are letter grades varying from A to F; X_6 and X_8 vary from 0 to 15; X_7 varies from 0 to 16; and Y varies from 0 to 100. Before being used for developing a predictive mathematical model, the collected raw data must be pre-processed to avoid the cases in which one variable receives a higher or lower weight for its coefficient due to its initial low or high scale of measurements.

Data pre-processing was conducted in the following way: First, all letter grades of X_2, X_3, X_4, and X_5 were converted into the corresponding numerical values using the following scales: A = 4.00; A− = 3.67; B+ = 3.33; B = 3.00; B− = 2.67; C+ = 2.33; C = 2.00; C− = 1.67; $D+ = 1.33$; $D = 1.00$; $F = 0.00$. Second, the numerical values of all data were normalized, so each datum varied within the same scale from 0 to 1, as shown in Table 1. The normalized value of data was calculated through dividing the initial value of the data by its range. For instance, the range of GPA that a student could receive was 4.00. If one student earned a GPA of 3.55, then the student's normalized GPA would be $3.55 \div 4.00 = 0.89$.

Table 2 shows the results of descriptive analysis of the normalized data in four semesters. As seen in Table 2, most variables of X_1–X_8 and Y in Semesters #2 and #3 had lower means and higher standard deviations, and some variables in Semester #4 had higher means and lower standard deviations. For example, compared to students in Semester #1 as a whole, students in Semesters #2 and #3 had lower cumulative GPAs, lower statics scores, lower dynamics mid-exam #3 scores, and higher standard deviations in GPA, statics, and dynamics mid-exam #3 scores; while students in Semester #4 had higher cumulative GPA, higher statics scores, higher physics scores, and lower standard deviations in GPA, statics, and physics scores.

Figs. 2–5 further show the histograms (frequency distributions) of students' dynamics final exam scores in Semesters #1–4, respectively, where the difference of students' dynamics final exam scores in four semesters can be seen clearly. In short, students in Semesters #2–4 were diverse in their academic performance. Therefore, Semesters #2–4 provided excellent cases to validate the generalizability of the predictive models developed from the dataset collected in Semester #1.

4. Predictive modeling of student academic performance in dynamics

4.1. Combinations of predictor variables

The full dataset collected in Semester #1 ($n = 128$) were used to develop a set of predictive mathematical models. The datasets collected in Semesters #2–#4 were then employed to validate the predictive models. Four types of mathematical modeling techniques, including

Table 2
Descriptive statistics of the normalized data.

Variable	Semester #1		Semester #2		Semester #3		Semester #4	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Cumulative GPA	0.86	0.096	0.81	0.112	0.84	0.106	0.86	0.088
Statics grade	0.81	0.189	0.67	0.206	0.77	0.243	0.84	0.173
Calculus I grade	0.76	0.186	0.76	0.193	0.72	0.194	0.73	0.183
Calculus II grade	0.78	0.183	0.73	0.200	0.71	0.209	0.77	0.191
Physics grade	0.79	0.160	0.74	0.187	0.75	0.167	0.86	0.119
Dynamics mid-exam #1 score	0.79	0.158	0.71	0.185	0.73	0.152	0.78	0.128
Dynamics mid-exam #2 score	0.78	0.137	0.78	0.144	0.73	0.152	0.72	0.143
Dynamics mid-exam #3 score	0.85	0.124	0.81	0.150	0.77	0.152	0.83	0.134
Dynamics final exam score	0.72	0.167	0.69	0.158	0.66	0.177	0.71	0.165

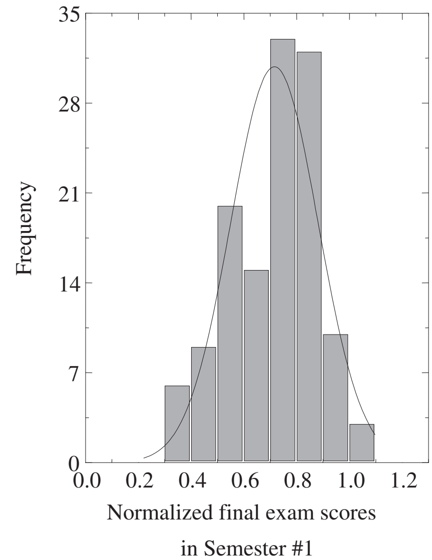

Fig. 2. Histogram of students' normalized scores on the dynamics final exam in Semester #1 ($n = 128$).

multiple linear regression (MLR), multilayer perception (MLP) network, radial basis function (RBF) network, and support vector machine (SVM), were used to develop the predictive models by using six combinations of predictor variables. This results in a total of $4 \times 6 = 24$ predictive mathematical models. The six combinations of predictor variables are listed below:

I X_1 used as the predictor
II X_1, X_2, X_3, X_4, and X_5 used as predictors
III X_6 used as the predictor
IV X_1, X_2, X_3, X_4, X_5, and X_6 used as predictors
IV X_1, X_2, X_3, X_4, X_5, X_6, and X_7 used as predictors
V X_1, X_2, X_3, X_4, X_5, X_6, X_7, and X_8 used as predictors

The first (X_1) and the second (X_1, X_2, X_3, X_4, and X_5) combinations of predictor variables do not include students' dynamics mid-term exam scores. Therefore, the predictive models developed with the first or the second combination of predictor variables can be employed even before the dynamics course starts. The predictive models with predictor combinations Nos. III–VI can only be developed as the dynamics course proceeds because the models require students' dynamics mid-term exam scores as inputs. For example, X_6 would not become incorporated until the end of the first quarter of the semester, and X_7 not until the middle of the semester, while X_8 would not come into play

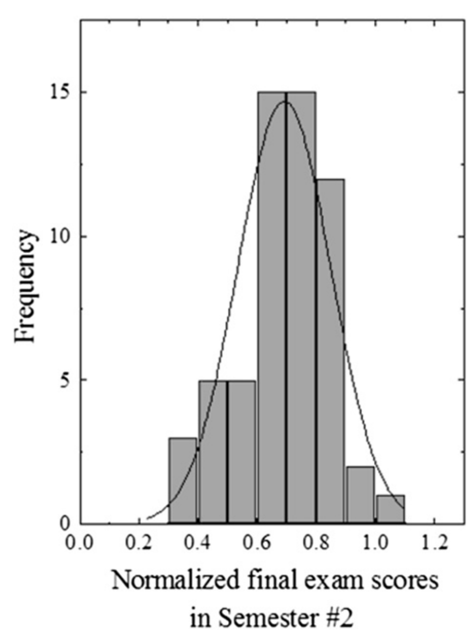

Fig. 3. Histogram of students' normalized scores on the dynamics final exam in Semester #2 ($n = 58$).

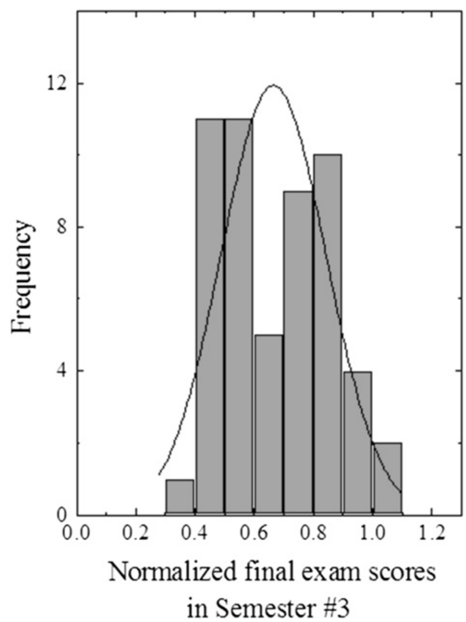

Fig. 4. Histogram of students' normalized scores on the dynamics final exam in Semester #3 ($n = 53$).

until the last quarter of the semester. The instructor may choose different predictor combinations during different periods in a semester based on the needs of each course.

4.2. Criteria for examining the prediction accuracy of each predictive model

The prediction accuracy of each of the 24 predictive models was examined by using the following two criteria:

(1) Average prediction accuracy (APA). It indicates on average how well the model predicts the final exam scores of all students in the dynamics course. The APA for the final exam scores was calculated as:

$$APA = 1 - \frac{1}{n} \cdot \sum_{i=1}^{n} \left| \frac{P_i - A_i}{A_i} \right| \times 100\% \tag{1}$$

where n is the total number of cases (students); P_i is the predicted final exam score of the ith student in the class ($i = [1, n]$); and A_i is the actual final exam score of the ith student. The higher the APA, the better the model.

(2) Percentage of accurate predictions (PAP). The PAP among all predictions was calculated as the number of accurate predictions divided by the total number of predictions. In the present study, an accurate prediction was defined as the prediction in which the predicted value is within 90–110% of the actual value (namely, the prediction error is ±10%). The higher the PAP, the better the model.

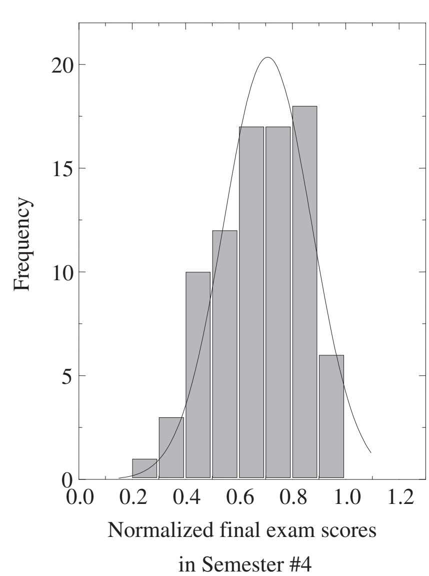

Fig. 5. Histogram of students' normalized scores on the dynamics final exam in Semester #4 ($n = 84$).

4.3. Mathematical modeling and algorithms

Multiple regression is a "logical extension" of simple linear regression based on the least square principle (Field, 2005). Multiple regression takes into account the effect of multiple independent variables (i.e., predictors variables) on a dependent variable and determines a quantitative relationship between them. In the present study, the multiple linear regression (MLR) models were developed using a commercial statistical software package SPSS Base 18.0. The statistical significance threshold was set to be 0.05. The MLR models contain all eight predictor variables (X_1–X_8) in order to compare the prediction accuracy of the MLR models with that of the other three types of models.

Neural networks refer to a set of interconnected units/neurons that function in parallel to complete a global task. In the present study, two most commonly used types of neural networks are employed: multilayer perception (MLP) and radial basis function (RBF) networks. The MLP model is a multilayer, feed-forward network model using error back propagation as the learning method (Haykin, 1999). Its performance is influenced by the number of hidden layers, units in hidden layers, activation function, weight, and the learning rate. The RBF model is a three-layer, feed-forward network model using RBF as activation function in the hidden layer and linear function as activation function in the output layer (Huang, Saratchandran, & Sundararajan, 2005). Its performance is mainly influenced by the number of units in the hidden layer. In the present study, the MLP and RBF models were developed using a commercial statistical software package SPSS Neural Network. The default values of relevant parameters of SPSS Neural Network, such as the minimum relative change in training error, the minimum relative change in training error ratio, and the maximum training epochs, were adopted and automatically optimized with specific criteria. The Appendix A of this paper provides a detailed elaboration of the MLP and RBF algorithms.

The support vector machine (SVM) model is a new machine learning approach based on statistics learning theory and the principle of structural risk minimization (Vapnick, 1995). SVM has advantages in global optimization, generalization ability, and learning with small size samples. The performance of SVM is influenced by penalty factor and kernel parameter. In the present study, the SVM models were developed using MATLAB codes. Genetic algorithms were also employed to optimize two important parameters C and σ^2 in the SVM models (Pai & Hong, 2005). The Appendix A of this paper also provides a detailed elaboration of the SVM algorithm employed in this study.

5. Results and analysis

5.1. Mathematical equations of the predictive models

Because the algorithms of MLP, RFB, and SVM models are complex, no simple mathematical equations can be provided in this paper to show what these three types of models look like. The multiple linear regression (MLR) models are simple and have the following explicit mathematical equations:

MLR model No. 1 (using X_1 as the predictor):

$$Y = 0.047 + 0.781X_1 \tag{2}$$

MLR model No. 2 (using X_1 to X_5 as predictors):

$$Y = 0.022 + 0.715X_1 + 0.034X_2 - 0.063X_3 - 0.077X_4 + 0.204X_5 \tag{3}$$

MLR model No. 3 (using X_6 as the predictor):

$$Y = 0.334 + 0.487X_6 \tag{4}$$

MLR model No. 4 (using X_1 to X_6 as predictors):

$$Y = -0.053 + 0.567X_1 - 0.025X_2 - 0.041X_3 - 0.101X_4 + 0.191X_5 + 0.334X_6 \tag{5}$$

MLR model No. 5 (using X_1 to X_7 as predictors):

$$Y = -0.079 + 0.502X_1 - 0.036X_2 - 0.036X_3 - 0.090X_4 + 0.186X_5 + 0.303X_6 + 0.138X_7 \tag{6}$$

MLR model No. 6 (using X_1 to X_8 as predictors):

$$Y = -0.369 + 0.515X_1 - 0.097X_2 + 0.024X_3 - 0.085X_4 + 0.149X_5 + 0.233X_6 - 0.001X_7 + 0.556X_8 \tag{7}$$

Table 3 shows the R-square and standardized coefficients β of each MLR model. As seen from R-square values in Table 3, the MLR models explain 20.1%–44.7% of the change in students' dynamics final exam scores. For MLR model No. 6 that contains all eight predictor variables,

Table 3
Standardized coefficients of the MLR models.

MLR model no.	Predictor	R^2	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
1	X_1	0.201	0.448[a]	–	–	–	–	–	–	–
2	X_1–X_5	0.238	0.410[a]	0.039	−0.07	−0.085	0.196[b]	–	–	–
3	X_6	0.212	–	–	–	–	–	0.461[a]	–	–
4	X_1–X_6	0.311	0.325[a]	−0.028	−0.045	−0.111	0.183[b]	0.315[a]	–	–
5	X_1–X_7	0.320	0.288[a]	−0.041	−0.040	−0.099	0.178[b]	0.286[a]	0.113[b]	–
6	X_1–X_8	0.447	0.295[a]	−0.110	0.027	−0.093	0.142[b]	0.220[a]	−0.001	0.413[a]

[a] Correlation is significant at the 0.01 level (2-tailed).
[b] Correlation is significant at the 0.05 level (2-tailed).

Table 4
Prediction accuracy of the four types of mathematical models with X_1as the predictor.

Model type	Average prediction accuracy (%)					Percentage of accurate predictions (%)				
	Sem. #1	Sem. #2	Sem. #3	Sem. #4	Four-sem. average	Sem. #1	Sem. #2	Sem. #3	Sem. #4	Four-sem. average
MLR	88.3	90.2	87.7	88.1	88.6	50.0	56.9	54.7	53.6	53.8
MLP	88.1	90.1	87.7	88.1	88.5	48.4	58.6	52.8	53.6	53.4
RBF	87.8	89.8	87.8	87.7	88.3	50.0	62.1	50.9	52.4	53.9
SVM	80.5	90.4	87.3	87.9	86.5	51.6	62.1	50.9	50.0	53.7

the most important predictor variables that affect prediction accuracy are: dynamics mid-term exam #3 ($\beta_8 = 0.413$), cumulative GPA ($\beta_1 = 0.295$), dynamics mid-term exam #1 ($\beta_6 = 0.220$), and physics ($\beta_5 = 0.142$). The most important predictor variables also vary with different types of mathematical models employed (MLR, MLP, RFB, or SVM).

5.2. Effect of the type of mathematical model on prediction accuracy

Tables 4–9 show the average prediction accuracy (APA) and the percentage of accurate predictions (PAP) of the four types of mathematical models with six combinations of predictor variables. Because four semesters were involved, the four-semester averages data are also included in Tables 4–9.

As seen from Tables 4–9, the four types of mathematical models have the APA of 81%–91%, and the PAP of 40%–72%. For different types of models, the four-semester average of APA varies slightly within 1.8%, 0.6%, 0.5%, 2.1%, 1.4%, and 0.8% in Tables 4–9, respectively. And the four-semester average of PAP varies also slightly within 0.5%, 2.0%, 3.6%, 3.8%, 5.8%, and 4.5% in Tables 4–9, respectively. This means that the type of mathematical model has only a slight effect on prediction accuracy. If the combination of predictor variables is determined, the instructor can choose any type of mathematical model to make predictions without causing a significant difference in prediction accuracy.

5.3. Effect of the combination of predictor variables on prediction accuracy

To more clearly show the effect of the combination of predictor variables on prediction accuracy, the four-semester averages data in Tables 4–9 were employed to make Table 10. As seen from Table 10, the combination of predictor variables has only a slight effect on the average prediction accuracy for all four types of models employed. For the MLR models, the four-semester average of APA varies within 1.4% (from 88.3% to 89.7%) for the six combinations of predictor variables; for the MLP models, 1.4% (from 88.2% to 89.6%); for the RBF models, 1.8% (from 88.3% to 89.9%); and for the SVM models, 3.6% (from 87.5% to 90.1%).

However, the combination of predictor variables has a profound effect on the percentage of accurate predictions. As seen from Table 10, for the MLR models, the four-semester average of PAP varies from 49% (with X_6 as the predictor) to 61.3% (with X_1 to X_8 as predictors) – a change of 12.3%; for the MLP models, from 48.9% to 59.5% – a change of 10.6%; for the RBF models, from 51.5% to 61.9% – a change of 11.4%; and for the SVM models, from 52.5% to 64% – a change of 11.5%.

From a close examination of the PAP data listed in Table 10, the following observations are also made:

(1) In terms of four-semester averages, the PAP varies very slightly (less than 1.3%) among the four types of mathematical models when using X_1 as the predictor or using X_1–X_5 as the predictors. This means that adding X_2, X_3, X_4, and X_5 (which are student grades in four pre-requisite courses) to a mathematical model does not make the model more accurate than the model with only X_1 (cumulative GPA) as the predictor. If the instructor needs to predict an individual student's performance in dynamics from the student's previous track record only, the instructor just needs to collect the student's GPA data and then use the GPA data in a mathematical model to make predictions.
(2) Using X_6 (the score on dynamics mid-term exam #1) alone as the predictor generally generates the lowest PAP for all four types of mathematical models. This means that the instructor should not only rely on X_6 (the score on dynamics mid-term exam #1) to make predictions. Other factors should also be taken into account when making predictions.
(3) The PAP increases when more predictor variables (X_1–X_6, X_1–X_7, and X_1–X_8) are included in the models, as seen from the last three columns in Table 10. When all predictor variables (X_1–X_8) are included in the models, the highest APAs (59.5%–64% of four-semester averages) are generated. In most cases, the support vector machine (SVM) models produce the highest PAPs as compared to the other three types of mathematical models. For example, as shown in Table 10, the four-semester averages of APAs of the SVM models are 59.1%, 61.3%, and 64.0%, respectively, for the predictor combinations of X_1–X_6, X_1–X_7, and X_1–X_8.

Table 5
Prediction accuracy of the four types of mathematical models with X_1–X_5as predictors.

Model type	Average prediction accuracy (%)					Percentage of accurate predictions (%)				
	Sem. #1	Sem. #2	Sem. #3	Sem. #4	Four-sem. average	Sem. #1	Sem. #2	Sem. #3	Sem. #4	Four-sem. average
MLR	88.6	90.1	87.8	88.1	88.7	54.7	62.1	49.1	53.6	54.9
MLP	88.6	90.1	87.6	88.0	88.6	54.7	60.3	45.3	51.2	52.9
RBF	88.1	90.0	87.4	86.7	88.1	50.8	62.1	49.1	51.2	53.3
SVM	80.4	90.6	87.3	87.9	86.6	56.3	65.5	47.2	47.6	54.2

Table 6
Prediction accuracy of the four types of mathematical models with X_6as the predictor.

Model type	Average prediction accuracy (%)					Percentage of accurate predictions (%)				
	Sem. #1	Sem. #2	Sem. #3	Sem. #4	Four-sem. average	Sem. #1	Sem. #2	Sem. #3	Sem. #4	Four-sem. average
MLR	88.2	89.9	86.7	88.5	88.3	54.7	53.4	40.4	47.6	49.0
MLP	88.2	90.0	86.8	87.8	88.2	51.6	56.9	39.6	47.6	48.9
RBF	88.0	90.0	87.0	88.3	88.3	51.6	58.6	43.4	52.4	51.5
SVM	85.7	88.8	87.2	88.2	87.5	59.4	56.9	42.5	51.2	52.5

Table 7
Prediction accuracy of the four types of mathematical models with X_1–X_6 as predictors.

Model type	Average prediction accuracy (%)					Percentage of accurate predictions (%)				
	Sem. #1	Sem. #2	Sem. #3	Sem. #4	Four-sem. average	Sem. #1	Sem. #2	Sem. #3	Sem. #4	Four-sem. average
MLR	89.5	89.9	88.4	89.0	89.2	58.6	62.1	50.9	53.6	56.3
MLP	89.4	90.1	88.4	89.2	89.3	57.0	60.3	49.1	54.8	55.3
RBF	88.4	90.6	88.3	89.0	89.1	50.0	63.8	52.8	57.1	55.9
SVM	82.0	90.3	88.5	89.1	87.5	64.8	63.8	52.8	54.8	59.1

5.4. Implications

From the research findings described above, the following two most important conclusions can be made: (1) The type of mathematical model has only a slight effect on the average prediction accuracy and the percentage of accurate predictions. (2) The combination of predictor variables has only a slight effect on the average prediction accuracy but a profound effect on the percentage of accurate predictions. In general, the support vector machine (SVM) models produce the highest PAP as compared to the other three types of mathematical models. Note that the average prediction accuracy is important for predicting the average academic performance of a dynamics class as a whole, and that accurate prediction is important for individual students in the class. The conclusions made above imply that:

(1) If the goal of the instructor is to predict the average academic performance of his/her dynamics class as a whole, the instructor can choose the simplest mathematical model (i.e., MLR model) and the simplest predictor combination (i.e., X_1 cumulative GPA alone), which is Equation (2) in the present study. There is no need to adopt complex mathematical models and include any other predictor variables in the model.
(2) If the goal of the instructor is to predict the academic performance of individual students, the instructor should use the SVM model with X_1–X_6 (student's cumulative GPA; grades in statics, calculus I, calculus II, and physics; and the first dynamics mid-term exam score) as predictor variables because this particular predictor combination increases the percentage of accurate predictions and also allows sufficient time for the instructor to implement subsequent educational interventions. Although mathematically, the combination of all predictor variables (X_1–X_8) yields the highest PAP, the instructor would have to wait until the dynamic mid-term exams #2 and #3 (i.e., X_7 and X_8) are complete – leaving insufficient time for the instructor to implement educational interventions to improve student performance.

6. Limitations of the present study

Compared to the high average prediction accuracy (APA) of 81%–91%, the four types of mathematical models have a relatively low percentage of accurate predictions (PAP) of 40%–72% depending on a particular type of mathematical model and a particular combination of predictor variables employed. The relatively low PAP is associated with the limitations of the present study as described in the following paragraphs.

First, the predictive models developed in the present study only take into account eight cognitive factors (X_1–X_8). A significant amount of research (e.g., Graaff, Saunders-Smits, & Nieweg, 2005; Lin, Imbrie, & Reid, 2009; Pintrich & DeGroot, 1999; Ransdell, 2001; Riding & Rayner, 1998; Tracey & Sedlacek, 1984) has suggested that learning is an extremely complex process involving many psychological factors such as learning styles, self-efficacy, achievement goals, motivation, interest, and teaching and learning environment. These psychological factors will be considered in our future modeling work to develop a more accurate predictive model.

Table 8
Prediction accuracy of the four types of mathematical models with X_1–X_7 as predictors.

Model type	Average prediction accuracy (%)					Percentage of accurate predictions (%)				
	Sem. #1	Sem. #2	Sem. #3	Sem. #4	Four-sem. average	Sem. #1	Sem. #2	Sem. #3	Sem. #4	Four-sem. average
MLR	89.5	90.7	88.4	89.3	89.5	60.2	63.8	50.9	52.4	56.8
MLP	89.5	90.5	88.4	88.7	89.3	57.0	60.3	50.9	53.6	55.5
RBF	88.7	90.9	88.4	88.7	89.2	50.8	69.0	52.8	51.2	56.0
SVM	85.1	91.1	88.9	89.5	88.7	67.2	65.5	52.8	59.5	61.3

Table 9
Prediction accuracy of the four types of mathematical models with X_1–X_8 as predictors.

Model type	Average prediction accuracy (%)					Percentage of accurate predictions (%)				
	Sem. #1	Sem. #2	Sem. #3	Sem. #4	Four-sem. average	Sem. #1	Sem. #2	Sem. #3	Sem. #4	Four-sem. average
MLR	90.3	90.5	88.2	89.8	89.7	65.6	56.9	58.5	64.3	61.3
MLP	90.3	90.6	88.0	89.6	89.6	66.4	56.9	52.8	61.9	59.5
RBF	89.1	91.0	89.0	90.4	89.9	57.0	72.4	56.6	65.5	62.9
SVM	90.2	90.9	89.0	90.3	90.1	64.1	69.0	62.3	60.7	64.0

Table 10
Effect of the combination of predictor variables on prediction accuracy.

Model type	Four-semester average of the average prediction accuracy (%)						Four-semester average of the percentage of accurate predictions (%)					
	X_1	X_1–X_5	X_6	X_1–X_6	X_1–X_7	X_1–X_8	X_1	X_1–X_5	X_6	X_1–X_6	X_1–X_7	X_1–X_8
MLR	88.6	88.7	88.3	89.2	89.5	89.7	53.8	54.9	49.0	56.3	56.8	61.3
MLP	88.5	88.6	88.2	89.3	89.3	89.6	53.4	52.9	48.9	55.3	55.5	59.5
RBF	88.3	88.1	88.3	89.1	89.2	89.9	53.9	53.3	51.5	55.9	56.0	62.9
SVM	86.5	86.6	87.5	87.5	88.7	90.1	53.7	54.2	52.5	59.1	61.3	64.0

Second, the grades that a student earned in pre-requisite courses (i.e., X_2–X_5) might not truly reflect the student's knowledge of those topics. A student may have taken pre-requisite courses years ago. By the time the student takes dynamics, their knowledge of pre-requisite courses may have improved. For example, some students took calculus courses more than two semesters before they took dynamics, and got only a C− in the calculus final exam. However, they may have received more practice with calculus problems through some other courses, such as a calculus-based physics course, and it is possible that they would now understand calculus at a level higher than their below-average grade would suggest.

Third, the present study made no differentiation between norm-referenced and criterion-referenced scores in the data collected. Different predictor variables have been determined using different criteria. A student who earns 60 (out of 100) in a criterion-referenced system may receive an A in a norm-referenced system (Gronlund & Waugh, 2009). Thus, a student who received an A in a pre-requisite course might not truly understand the given topics as well as his/her grade indicates, and may receive a low grade in the dynamics course.

7. Conclusions

Predicting student academic performance has long been an important research topic in many academic disciplines (e.g., Cohen et al., 2007; Grudnitski, 1997; Pokay & Blumenfeld, 1990; Ransdell, 2001; Ting, 2001). Extensive literature review shows that the present study is the first study that develops and compares four types of mathematical models to predict student academic performance in engineering dynamics – a high-enrollment, high-impact, and core course that many engineering undergraduates are required to take. The four types of predictive mathematical models are the multiple linear regression model, the multilayer perception network model, the radial basis function network model, and the support vector machine model.

Based on 2907 data points collected from 323 undergraduates in four semesters, 24 predictive mathematical models are developed in the present study. The analysis reveals that the type of mathematical model has only a slight effect on the average prediction accuracy and the percentage of accurate predictions. The combination of predictor variables has only a slight effect on the APA but a profound effect on the PAP. The answers to the two research questions of the present study are summarized in the following paragraphs.

Research question No. 1: Among the four types of predictive mathematical models investigated in the present study, what particular type is the most appropriate for use in the prediction of student academic performance in engineering dynamics?

Answer: If the goal of the instructor is to predict the average academic performance of his/her dynamics class as a whole, the instructor can choose the simplest multiple linear regression model, because the four types of mathematical models do not have a significant difference in terms of the average prediction accuracy. However, if the goal of the instructor is to predict the academic performance of individual students, the instructor should choose the support vector machine model because the SVM model generally yields the highest percentage of accurate prediction among the four types of mathematical models.

Research question No. 2: Among the six combinations of predictor variables investigated in the present study, what particular combination is the most appropriate for use in the prediction of student academic performance in engineering dynamics?

Answer: If the goal of the instructor is to predict the average academic performance of his/her dynamics class as a whole, the instructor can choose X_1 (cumulative GPA) as the only predictor variable along with the multiple linear regression model. Adding more predictor variables does not help improve the average prediction accuracy of any of the four mathematical models that were tested in this study. However, if the goal of the instructor is to predict the academic performance of individual students, the instructor should choose X_1–X_6 (i.e., student's cumulative GPA; grades in statics, calculus I, calculus II, and physics; and the first dynamics mid-term exam score) as predictor variables along with the support vector machine model. Mathematically, the combination of all predictor variables (X_1–X_8) yields the highest PAP. However, the instructor would have to wait until the dynamic mid-term exams #2 and #3 (i.e., X_7 and X_8) are complete, leaving insufficient time for the instructor to implement educational interventions.

Finally, the predictive models developed in the present study only take into account eight cognitive factors. To increase the percentage of accurate prediction, psychological factors such as learning styles, self-efficacy, achievement goals, motivation, interest, and teaching and learning environment will be considered in our future modeling work.

Appendix A

A1. Mathematical modeling with the multilayer perception (MLP) network

MLP network, also known as multilayer feed-forward neural network, has been most widely used in prediction due to its good ability to deal with "functional mapping problems" in which one needs to identify how input variables affect output variables (Maimon, 2008; Zhang, Patuwo, & Hu, 1997). One of its key learning methods is error back propagation.

Fig. A1 shows the schematic diagram of the MLP neural network that contains an input layer, one or more hidden layers, and an output layer. Each layer consists of a set of interconnected neurons. The neurons, which include nonlinear activation functions, learn from experience without an explicit mathematical model about the relationship between inputs and outputs (Haykin, 1999). Sample data enter the network via the input layer, and exit from the output layer after being processed by each hidden layer. Each layer can only influence the one next to it. If the output layer does not yield the expected results, the errors go backward and distribute to the neurons. Then, the network adjusts weights to minimize errors.

Several factors affect the prediction accuracy of the MLP network, such as the number of layers, units in the hidden layers, activation function, weight, and learning rate. Increasing the number of layers and units may improve the prediction accuracy of the MLP network; however, it also increases complications and training time. Initial weight determines whether the network can reach a global minimum. The learning rate determines how much the weight is changed each time.

A2. Mathematical modeling with the radial basis function (RBF) network

RBF network is a three-layer feed-forward network that has a good capability to approximate complex nonlinear mapping directly from the input–output data (Huang et al., 2005). Different from the MLP network, the RBF network takes the RBF function as the activation function in the hidden layer, and a linear function as the activation function in the output layer (Maimon, 2008). The RBF network approach can estimate any continuous function (including nonlinear functions) and has a good capability for generalization.

The prediction accuracy of the RBF network is mainly affected by the number of units in the hidden layer. If the number is too small, the network is too simple to reflect the objective. However, if the number is too large, over-fit may occur and the generalization capability of the RBF network would decline.

It must be pointed out that although neural networks (MLP and RFB) are good for learning and modeling, one shortcoming of neural networks is over fitting. When over fitting occurs, the predictive capability of a neural network model decreases (Fulcher, 2008). This means that the neural network model is highly accurate only when the training dataset is used, but prediction falters if other dataset is included. To avoid over fitting, it is necessary to prune the model, that is, separate the data that are used for building the predictive model into the training and testing datasets, and use the testing dataset to modify the model to prevent over fitting. In this way, the prediction accuracy of the neural network model can be improved when dealing with different datasets (Linoff & Berry, 2011).

A3. Mathematical modeling with support vector machine (SVM)

SVM is a learning system developed by Vapnick (1995) based on the structural risk minimization (SRM) principle. Compared to the traditional empirical risk minimization (ERM) principle, which minimizes the errors in training data, SRM minimizes an upper bound on the expected risk. This feature enables SVM to be more accurate in generalization.

The SVM method was first used to handle classification problems (pattern recognition) by mapping nonlinear functions into linear functions "in a high dimensional feature space" (Cristianini & Taylor, 2000). However, by introducing a loss function, an SVM model can also be applied to regression problems as well. For regression purposes, ε – insensitive loss function is often used (Deng & Tian, 2004; Stitson, Weston, Gammerman, & Vapnik, 1996). ε is a small number that makes the predictive error (difference between the predicted value $f(x)$ and the actual value y) ignorable. In general, ε is set as a small positive number or zero, for example, 0.001. Equation (A1) and Fig. A2 illustrate the ε – insensitive loss function.

$$L_\varepsilon = |y - f(x,\omega)|_\varepsilon = \begin{cases} 0 & \text{for} |y - f(x,\omega)| \le \varepsilon \\ |y - f(x,\omega)| - \varepsilon & \text{otherwise} \end{cases} \tag{A1}$$

where ω is the parameter to identify, and ε is a user-defined precision parameter.

Given a set of data $\{x_i, y_i\}, i = 1, \cdots, n, x_i \in R^d, y_i \in R$, where$R^d$ is a Euclidean space, the linear regression function commonly used is given by Smola and Scholkopf (2004):

$$f(x) = (w \cdot x) + b \tag{A2}$$

where b is a model parameter. Considering the fitting error, two slack variables $\xi_i \ge 0$ and $\xi_i^* \ge 0$ are introduced. To minimize the ε – insensitive loss function $\|\omega\|^2/2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$, the equivalent primal optimization problem is

$$\min_{\omega \in R^n, b \in R} \quad \|\omega\|^2 \big/ 2 + C \sum_{i=1}^{l} \left(\xi_i + \xi_i^*\right) \tag{A3}$$

$$\text{subject to} \quad \begin{cases} y_i - \omega \cdot x_i - b \le \varepsilon + \xi_i \\ \omega \cdot x_i + b - y_i \le \varepsilon + \xi_i^* \end{cases} \quad i = 1, 2, \cdots, l$$

where constant $C > 0$. Constant C measures "the trade-off between complexity and losses" (Cristianini & Taylor, 2000) and stands for the penalty on the sample data which has a larger error than ε.

The kernel function $k(\cdot)$ is introduced to map a nonlinear regression problem in a low-dimensional space into a linear regression problem in a high-dimensional space (Collobert & Bengio, 2001). The Gaussian kernel $K(x, y)$ is one of the kernels that are most commonly used in SVM regression and is expressed as (Chapelle, Vapnik, Bousquet, & Mukherjee, 2002; Hong & Hwang, 2003):

$$K(x,y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}} \tag{A4}$$

where σ is a model parameter. The regression function at a given point is calculated by:

$$f(x) = \sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right)K(x_i \cdot x) + b \tag{A5}$$

where α is a Lagrange multiplier.

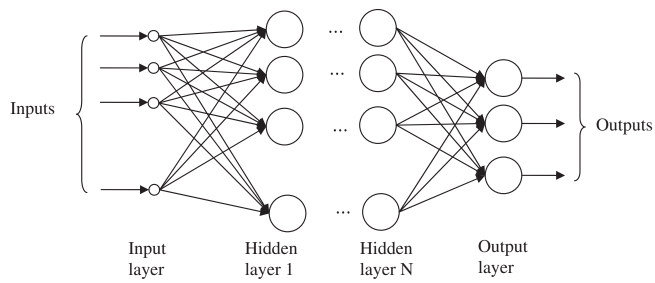

Fig. A1. Schematic diagram of an MLP neural network.

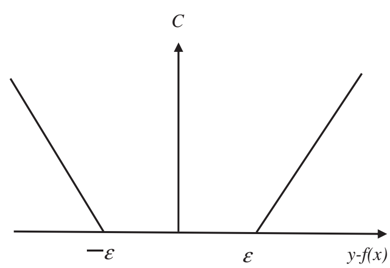

Fig. A2. The ε – insensitive loss function.

References

Ayan, M. N. R., & Garcia, M. T. C. (2008). Prediction of university students' academic achievement by linear and logistic models. *The Spanish Journal of Psychology, 11*, 275–288.

Barrett, S. F., LeFevre, E. W., Steadman, J. W., Tietjen, J. S., White, K. R., & Whitman, D. L. (2010). *Using the fundamentals of engineering (FE) examination as an outcomes assessment tool*. Clemson, SC: National Council of Examiners for Engineering and Surveying.

Beer, F. P., Johnston, E. R., Eisenberg, E. R., Clausen, W. E., Mazurek, D., & Cornwell, P. J. (2009). *Vector mechanics for engineers: Statics and dynamics* (9th ed.). New York, NY: McGraw-Hill.

Chapelle, O., Vapnik, V., Bousquet, O., & Mukherjee, S. (2002). Choosing multiple parameters for support vector machines. *Machine Learning, 46*(1–3), 131–159.

Cios, K. J., Pedrycz, W., & Swiniarski, R. M. (1998). *Data mining methods for knowledge discovery*. Boston, MA: Kluwer.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). Oxon, UK: Routledge.

Collobert, R., & Bengio, S. (2001). SVM torch: support vector machines for large-scale regression problems. *Journal of Machine Learning Research, 1*, 143–150.

Cristianini, N., & Taylor, J. S. (2000). *An introduction to support vector machines and other kernel-based learning methods*. New York, NY: Cambridge University.

Deng, N., & Tian, Y. (2004). *A new method of data mining: Support vector machine*. Beijing, China: Science Press.

Fang, N., & Lu, J. (2010). A decision tree approach to predictive modeling of student performance in engineering dynamics. *International Journal of Engineering Education, 26*(1), 87–95.

Field, A. (2005). *Discovering statistics using SPSS*. London, UK: Sage.

Fulcher, J. (2008). Artificial neural networks. In J. Flucher, & L. C. Jain (Eds.), *Computational intelligence: A compendium* (pp. 26–30). Berlin, Germany: Springer.

Graaff, E., Saunders-Smits, G. N., & Nieweg, M. R. (2005). *Research and practice of active learning in engineering education*. Amsterdam, The Netherlands: Palls.

Gronlund, N. E., & Waugh, C. K. (2009). *Assessment of student achievement*. Boston, MA: Allyn & Bacon.

Grudnitski, G. (1997). A forecast of achievement from student profile data. *Journal of Accounting Education, 15*(4), 549–558.

Haykin, S. (1999). *Neural network: A comprehensive foundation* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.

Hibbeler, R. C. (2009). *Engineering mechanics dynamics* (12th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Hong, D. H., & Hwang, C. (2003). Support vector fuzzy regression machines. *Fuzzy Sets and Systems, 138*(2), 271–281.

Huang, S., & Fang, N. (2010). Prediction of student academic performance in an engineering dynamics course: development and validation of multivariate regression models. *International Journal of Engineering Education, 26*(4), 1008–1017.

Huang, G., Saratchandran, P., & Sundararajan, N. (2005). A generalized growing and pruning RBF neural network for function approximation. *IEEE Transactions on Neural Networks, 16*(1), 57–67.

Imbrie, P. K., Lin, J. J., Reid, K., & Malyscheff, A. (2008). Using hybrid data to model student success in engineering with artificial neural networks. In *Proceedings of the Research in Engineering Education Symposium, July 7–10, Davos, Switzerland*.

Lin, J. J., Imbrie, P. K., & Reid, K. (2009). Student retention modeling: an evaluation of different methods and their impact on prediction results. In *Proceedings of the Research in Engineering Education Symposium, July, Palm Cove, QLD*.

Linoff, G. S., & Berry, M. J. A. (2011). Artificial neural networks. In R. Elliott, A. O. Tulton, D. Scribner, & P. Lowell (Eds.), *Data mining techniques for marketing, sales, and customer relationship management* (pp. 211–255). Indianapolis, IN: Wiley.

Lykourentzou, L., Giannoukos, L., & Mpardis, G. (2009). Early and dynamic student achievement prediction in e-learning courses using neural networks. *Journal of the American Society for Information Science and Technology, 60*(2), 372–380.

McKeachie, W. J., Pintrich, P. R., Lin, Y. G., & Smith, D. (1986). *Teaching and learning in the college classroom: A review of the research literature*. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning.

Magill, M. A. (1997). Classroom models for illustrating dynamics principles, part I-particle kinematics and kinetics. In *Proceeding of American Society of Engineering Education Annual Conference and Exposition, Milwaukee, WI* (pp. 15–18). Washington, DC: ASEE.

Maimon, O. Z. (2008). Neural network methods. In O. Maimon, & L. Rokach (Eds.), *Soft computing for knowledge discovery and data mining*. New York, NY: Springer.

Njock-Libii, J. (2010). Using microsoft windows to compare the energy dissipated by old and new tennis balls. In *Proceedings of the 2010 National Conference of the American Society for Engineering Education, June 20–23, Louisville, KY.*

Pai, P. F., & Hong, W. C. (2005). Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms. *Electric Power Systems Research, 74*(3), 417–425.

Pintrich, P. R., & DeGroot, E. V. (1999). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*, 33–40.

Pokay, P., & Blumenfeld, P. C. (1990). Predicting achievement early and late in the semester: the role of motivation and use of learning strategies. *Journal of Educational Psychology, 82*(1), 41–50.

Ransdell, S. (2001). Predicting college success: the importance of ability and non-cognitive variables. *International Journal of Educational Research, 35*(4), 357–364.

Riding, R., & Rayner, S. (1998). *Cognitive styles and learning strategies: Understanding style differences in learning and behavior.* London, UK: David Fulton.

Self, B. P., Wood, J. J., & Hansen, D. (2004). Teaching undergraduate kinetics using LEGO mindstorms race car competition. In *Proceeding of the 2004 National Conference of the American Society for Engineering Education, July, Salt Lake City, UT.*

Smola, A. J., & Scholkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing, 14*(3), 199–222.

Stitson, M. O., Weston, J. A. E., Gammerman, A. V., & Vapnik, V. (1996). *Theory of support vector machines* (Report No. CSD-TR-96–17). London, UK: Royal Holloway University.

Ting, S. R. (2001). Predicting academic success of first-year engineering students from standardized test scores and psychosocial variables. *International Journal of Engineering Education, 17*, 75–80.

Tracey, T. J., & Sedlacek, W. E. (1984). Noncognitive variables in predicting academic success by race. *Measurement and Evaluation in Guidance, 16*(4), 171–178.

Vandamme, J. P., Meskens, N., & Superby, J. F. (2007). Predicting academic performance by data mining methods. *Education Economics, 15*(4), 405–419.

Vapnick, V. N. (1995). *The nature of statistical learning theory*. New York, NY: Springer-Verlag.

Veenstra, C. P., Dey, E. L., & Herrin, G. D. (2008). Is modeling of freshman engineering success different from modeling of non-engineering success? *Journal of Engineering Education, 97*(3), 467–479.

Ware, W. B., & Galassi, J. P. (2006). Using correlational and prediction data to enhance student achievement in K-12 schools: a practical application for school counselors. *Professional School Counseling, 9*(5), 344–356.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1997). Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting, 14*, 35–62.