IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 14, 2020, accepted July 13, 2020, date of publication July 21, 2020, date of current version August 5, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3010938

Modern Artificial Intelligence Model Development for Undergraduate Student Performance Prediction: An Investigation on Engineering Mathematics Courses

RAVINESH C. DEO [1], (Senior Member, IEEE), ZAHER MUNDHER YASEEN [2], NADHIR AL-ANSARI [3], THONG NGUYEN-HUY [4], TREVOR ASHLEY MCPHERSON LANGLANDS [1], AND LINDA GALLIGAN [1]

[1] Faculty of Health, Engineering and Sciences, School of Sciences, University of Southern Queensland, Toowoomba, QLD 4350, Australia
[2] Sustainable Developments in Civil Engineering Research Group, Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam
[3] Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, 971 87 Lulea, Sweden
[4] Centre for Applied Climate Sciences, University of Southern Queensland, Toowoomba, QLD 4350, Australia

Corresponding authors: Ravinesh C. Deo (ravinesh.deo@usq.edu.au) and Zaher Mundher Yaseen (yaseen@tdtu.edu.vn)

This work was supported in part by the University of Southern Queensland, through the School of Sciences Quartile 1 Challenge Internal Grant, for this research study under Grant 2018, and in part by the Office for the Advancement of Learning and Teaching: Technology Demonstrator Project (Artificial intelligence as a predictive analytics framework for learning outcomes assessment and student success was granted by Human Research Ethics to securely and ethically utilize official examiner return data) under Grant H18REA236.

ABSTRACT A computationally efficient artificial intelligence (AI) model called Extreme Learning Machines (ELM) is adopted to analyze patterns embedded in continuous assessment to model the weighted score (WS) and the examination (EX) score in engineering mathematics courses at an Australian regional university. The student performance data taken over a six-year period in multiple courses ranging from the mid- to the advanced level and a diverse course offering mode (i.e., on-campus, ONC, and online, ONL) are modelled by ELM and further benchmarked against competing models: random forest (RF) and Volterra. With the assessments and examination marks as key predictors of WS (leading to a grade in the mid-level course), ELM (with respect to RF and Volterra) outperformed its counterpart models both for the ONC and the ONL offer. This generated relative prediction error in the testing phase, of only 0.74%, compared to about 3.12% and 1.06%, respectively, while for the ONL offer, the prediction errors were only 0.51% compared to about 3.05% and 0.70%. In modelling the student performance in advanced engineering mathematics course, ELM registered slightly larger errors: 0.77% (vs. 22.23% and 1.87%) for ONC and 0.54% (vs. 4.08% and 1.31%) for the ONL offer. This study advocates a pioneer implementation of a robust AI methodology to uncover relationships among student learning variables, developing teaching and learning intervention and course health checks to address issues related to graduate outcomes, and student learning attributes in the higher education sector.

INDEX TERMS Education decision-making, extreme learning machine, student performance modelling, AI in higher education, engineering mathematics.

I. INTRODUCTION

Over last three decades enormous growth in modelling and computational technologies has occurred, improving data analytics and computing resources to produce significant innovations [1]. Many statistical and mathematical modelling tools, including the autoregressive integrated moving average, linear regression and the partial and ordinary differential equations have long been the standard to understand causal inference and the relationships among variables. However,

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

data-driven models, focusing on artificial intelligence (AI) have recently been developed and are adopted in a wide range of fields, *e.g.*, education, medical sciences, healthcare, business intelligence, engineering, climate and environmental studies [2]–[5]. Investigators in many of these fields are constantly attempting to employ contemporary AI modelling approaches to develop, evaluate and implement modern-day decision systems.

Generally, AI falls in sub-category of smart algorithms implemented either in standalone models or within an integrated (hybridized) computer systems model [6]. They can demonstrate quite efficiently a human-like systems-thinking approach for practical decision-making processes. AI algorithms typically have a self-learning ability that enables them to capture and disseminate data patterns, synthesize information, self-correct and impute missing data and automatically perform an analysis of concealed patterns in complex variables [7]. These skills include feature extraction and the modelling the non-linear (and largely complex) relationships within relatively large and inter-related datasets. AI techniques are suited to the analysis of complex patterns existent between multivariate predictors and an objective variable. While recent studies showing their relevance in teaching and learning [8]–[11] and recent acceleration in their applications, the adoption of such techniques in education sector has been rather slow. This is despite recent studies showing their relevance in teaching and learning space [12], [13].

Similar to several sub-sectors in education, most universities today are under significant pressure to adopt evidence-based approaches in developing strategic measures to reduce failure and drop rates and improve progression rates, with their consequential benefits to improve graduate attributes and the quality of teaching and learning [14], [15]. On 1 January 2018, the Australian Government implemented key performance metrics to make the funding to universities contingent on their teaching performances [16]. The Government also requires greater transparency and reporting benchmarks on student experiences and demands more accountability of universities for the students they enroll. Accordingly, the effectiveness of operational decisions in the education sector could then be enhanced through a well-augmented, evidence-based practice. This would perhaps include investigating how the historical student learning datasets can be modelled with some sort of automated intelligent system to further support the key institutional decision-making processes.

Until recently, universities have predominantly employed manual reporting methods (*e.g.*, surveys) to collect evidence of student satisfaction and their performance outputs. A handful of universities in Australia have also developed strategies for the effective use of data, such as the University of New England's *Early Alert Program*, to track student learning, with the purpose of improving their retention and success rates [17]. Many other institutions are only beginning to consider the use of more sophisticated analytical methodologies to support and improve their learning and teaching outcomes [18]. A relatively new field in evidence-based practice is the 'learning analytics', which involves the measurement, collection, analysis and reporting of datasets about learners and their learning contexts to enhance one's understanding of these processes, and to also optimise the environments in which they occur [19].

Where many universities are still at the stage of considering costly integrations linking multiple datasets, incorporation of AI technologies as an ancillary modelling tool represents an advanced learning analytic methodology capable of bypassing the procedural and policy barriers of traditional learning analytics tools [20]. AI enables the implementation of data-driven decisions for operational purpose to address the challenges related to teaching and learning. When applied by educational practitioners to the student performance data, AI offers an objective methodology to collate and model such information and develop intelligent management systems that rapidly and efficiently assess student success. Such systems aim to identify possible indicators of student failure and attrition [21]. This system, where human biases are potentially eliminated from the decisions we make by considering the weighted risks and likelihoods, can be used to notify learning processes to the Faculty, course examiners and students, as a proactive forewarning tool of possible failures [22]–[24].

Measurement of student performance is expressed numerically based on assessments (*e.g.*, quiz marks) and the examination scores (*EX*), which are assigned a final weighted score (*WS*) to generate a student's grade. Continuous assessments are typically adopted as ongoing indicators of learning with the feedback in teaching activities (*e.g.*, quizzes) providing an early indication of the effectiveness of teaching and the improvements that are necessary to design a responsive teaching platform. The examination, in an undergraduate course, is often allotted a large proportion of a *WS*, to generate a final grade. It is therefore somewhat logical to construe that a predictive model, *via* a data mining approach, incorporating the assessments relative to the examination results or a grade, may aid in improving the course outcomes and such data can be utilized to design actionable insights to improve student learning [25]. The credibility of conventional methods to classify and grade academic performance is questionable given the nonlinear dependence of student performance on the assessments and the *EX*. That is, predictive models that employ mathematical and statistical techniques on student performance where the assumptions (*e.g.*, linearity, data distribution and model inputs) are forced may not be an optimal way to evaluate encapsulate human knowledge, skills acquired through a learning process or the most desirable learning outcomes (*e.g.*, the grade).

Studies are adopting AI-based learning analytics to model student performances. Gokmen *et al.*, (2010) used a fuzzy logic model to evaluate a laboratory course, reporting the model's advantages to be automation, flexibility, and a larger number of performance evaluation options relative to a classical approach adhering to the static mathematical calculations. Yadav and Singh proposed a fuzzy system for academic evaluation, reporting its flexibility and reliability, including the suitability not only for laboratory applications, but also in theoretical lessons, and in online and distance education [25]. In another study, authors used assessments comprised of six components (*i.e*., interface module, domain knowledge, inference engine, student module, mentor module, and pedagogical module) where an inference engine modelled the students' group classification from on-line pre-test examinations before starting a practical worksheet [26]. Their model was used as a flexible automated tool to classify learning groups based on the objectives of subjects and real-time performances compared to their t-scores. To study student performances in engineering, Bhatt and Bhatt developed a fuzzy model where practical components and results were compared to the outputs of classical methods [27]. Hwang and Yang used a fuzzy logic to assess student attentiveness, showing the model's ability to prevent erroneous judgments [28].

Another popular algorithm used to model student performance is the artificial neural network (ANN) [29]. An ANN model has the ability to handle complex data, analyze interrelated variables, learn nonlinear relationships between inputs and controlled or uncontrolled targets and check for patterns using nonlinear regressions [30]. Naik & Ragothaman, (2004) modelled student success for a Master of Business Administration program using neural networks. Gorr, Nagin, & Szczypula, (1994) compared ANNs to statistical models in their ability to predict grade point averages. The study of Naser *et al.*, applied ANNs to model the performances in courses for a Faculty of Engineering and Information Technology [29]. Huang and Fang predicted academic performances in engineering dynamics where four types of mathematical models served as a basis for comparison [33]. Stevens *et al.*, studying the performance of ANN-based modelling included a combinational incremental ensemble classifier to predict student performances in distance education [34]. A suite of ANN models was adopted to predict student performances with the data from Moodle logs [35].

Despite the widespread of ANN model in education sector, one of its drawbacks is the need for iterative tuning of hyper-parameters, its slow response based on the gradient learning algorithm and a relatively lower accuracy compared to some of the other modern AI algorithms [36], [37]. Hence, the enthusiasm to explore more advanced and reliable AI models for student's performance prediction is an ongoing research endeavor. Recently, in an effort to improve the ANN model, the newer version demoted as Extreme Learning Machines (ELM) that has a Single Hidden Layer of Feedforward Neural Network (SLFN) was proposed [38] and later, the purely randomized neuronal hidden layers were also formulated in this method [39]. Importantly, the ELM model has significant merits, producing greater accuracy and a faster learning speed than the ANN or the fuzzy logic model [39]. The ELM model also proved to be easier to implement given its randomized single hidden layers, and has thus attracted the attention of many researchers [40], [41]. Based on reported literature, the potential of using ELM model is yet to be explored, especially in modelling engineering mathematics student performances or its general application in the higher education sector.

In this paper, the skill of AI-based models, compared to the conventional modelling approaches, was investigated by developing a novel ELM method drawing upon the indicators of teaching and learning success and also exploring for the first time, its efficacy in predicting student performances in engineering mathematics courses. The objectives of the present study are as follows.

i. To construct an ELM model trained with multivariate predictors (*e.g*., continuous assessments) from mid- to advanced-mathematics courses in an Australian regional university, and assess its efficacy in the prediction of a *WS* versus teaching and learning indicators;
ii. To test the sensitivity of using continuous assessment marks on the relative contributions to the *EX* and the *WS*;
iii. To evaluate the model's accuracy versus the competing models, *e.g*., random forest and Volterra models.

To satisfy these objectives, student performance records spanning over six years (2013-2018) for a first-year engineering mathematics course and over five-years (2014-2018) for a second-year engineering mathematics course were drawn. To ensure credibility of all modelling data adopted to develop the proposed ELM model, the results were acquired from the Official Examiner Return repositories, stored in Faculty of Health, Engineering and Sciences under School of Sciences examination results, at University of Southern Queensland (USQ), Australia.

II. MATERIALS AND MODELS

In this section, the context of student performance data, the study, including the approach used in model design and the performance evaluation criteria are presented.

A. DATA AND STUDY CONTEXT

To design the proposed artificial intelligence models (*i.e*., ELM, RF and Volterra) for student performance prediction, we take the specific case of engineering mathematics student performance in Australian regional university. The present study employs independently analyzed and modelled data from both a *first* and *second* year engineering mathematics course, to provide a comparative platform for ELM modelling methodology. Data comprised of continuous internal assessments (*i.e*., quizzes & assignments), the final examination score (*EX*) and the weighted score (*WS*) (*i.e*., overall mark out of 100% necessary to attain a passing grade). These were for ENM1600 *Engineering Mathematics* and ENM2600 *Advanced Engineering Mathematics* the courses taught and administered by School of Sciences under Faculty of Health, Engineering and Science at University of Southern Queensland. Both courses are important service components of Bachelor of Engineering as well as a number of other programs, such as the Master of Science, Graduate Diploma and Graduate Certificate coursework programs majoring in mathematics, data science, statistics, and computer science. ENM1600 (but not ENM2600) is also a core course in surveying programs at undergraduate level, providing a diverse range of data where the prescribed ELM model was developed and tested for its ability to model examination marks and weighted scores.

Considering that USQ is a global leader in distance and online education and operates autonomously as an on-campus and a face to face teaching and research institution, the engineering mathematics student performance data from two different modes of course offering, ONL (online) and ONC (on-campus), were considered. The relevant data for the period 2013–2018 for ENM1600 and 2014-2018 for ENM2600, representing the total available data for these courses given their respective initial offers in 2013 and 2014, was acquired. Both ENM1600 and ENM2600 were developed as part of a major update and revision of previous mathematics syllabus to meet the program accreditation requirements under Engineers Australia. Therefore, the inclusion of ENM1600 and ENM2600 data in developing and evaluating ELM model is expected to make a significant contribution to future decision-making in these, and other courses and programs.

Prior to gaining access to, or processing student performance data, the relevant Human Ethics approval (#H18REA236) was promulgated, in strict accordance with requirements of Australian Code for Responsible Conduct of Research, 2018, and National Statement on Ethical Conduct in Human Research, 2007. Ethical application required disclosure of all relevant details about the proposed project to Ethics Committee and the perceived benefits and risk. As the project was purely quantitative and data-driven models did not draw upon any of student's personal record (*i.e.*, name, gender, socio-economic status) an expedited ethical approval marked this project as a 'low risk' teaching and learning investigation. Following this ethical approval, all student performance data were accessed from Examiner Returns, which are the official results provided to the Faculty after moderation process to facilitate grade release to students. In accordance with ethical standards, any form of student attributes such as names, gender and other personal attribute or identifiers were removed prior to data processing.

Since AI models are purely data-driven, they can face challenges with respect to the use of fragmented (or missing) data when used as an input for any predictive model. Accordingly, a preliminary quality checking procedure was undertaken. Any incomplete record, where for a given row of data a particular continuous assessment item or a *WS* was missing, was deleted entirely. Similarly, if a student's mark for at least one assessment piece (*e.g.*, Assignment 1) was missing the data for that student was considered incomplete, and thus discarded. Despite loss of some data from the original records, that may affect the ability of any model to predict a failing grade, this procedure ensured that the biases were reduced by using the records where every internal assessment data point (per student) used to construct a model had a corresponding *WS* value.

In Table 1 we report basic statistics of first and second year engineering mathematics data used to construct AI models. It is important to note that the two levels of courses have a different number of continuous assessments (*i.e.*, ENM1600 with three Assignments and two Quizzes; ENM2600 with only two Assignments and two Quizzes).

B. DESCRIPTION OF THE MODELING FRAMEWORK

In this sub-section, theoretical details of AI models developed for prediction of student performance are described.

1) THE OBJECTIVE MODEL: *EXTREME LEARNING MACHINES*

The Extreme Learning Machines (ELM) model consists of a single hidden layer feed-forward neural network. It has three distinct phases: *input*, where predictors and target variable are incorporated into the algorithm, *learning*, where data features are extracted and modelled nonlinearly to generate weights and biases, and *output*, where modelled data are transformed to their corresponding real values [41], [42]. In the learning phase, ELM adopts a least square approach, relying on weights and biases, to obtain a closed form of solution to the problem of interest.

One remarkable feature of ELM compared with the other AI models such as fuzzy logic, Support Vector Machines and ANN, is its ability to randomly generate weights and biases for a pre-defined dataset and a sufficiently large neural network [40]. Through its generally simplified modelling framework, ELM can lead to a highly accurate solution within a relatively short model execution time, albeit, also producing greater accuracy [41]. In hidden layer, a matrix pseudoinverse (*i.e.*, the Moore-Penrose inverse) is employed to generate the objective solution, avoiding the iterative training process as with the case of ANNs. This forces the solution to collapse to a local, rather than a global minimum [38]. Consistent with the universal approximation theory, this enables the ELM model to converge quickly, and exhibit a superior generalization. The modelling system also resolves issues of local minima and has a negligible over-fitting problem as separate training, validation, and testing datasets are employed [39].

The present study capitalizes on the relatively successful implementation of ELM model in science, arts and economics [43] and argues this model as a potentially new contribution for implementation in higher education sector. The ELM model is designed to train a multi-dimensional dataset with N pairs of training data $(\boldsymbol{X_i}, \boldsymbol{Y_i})$ with $\boldsymbol{X_i}$ (*i.e.*, predictors = continuous internal assessment marks, including Assignments marks and Quiz scores) of dimension D (*i.e.*, the number of predictors) and the 1-dimensional $\boldsymbol{Y}_i$ (i.e., the target variable denoted as the *EX* or *WS*).

Figure 1 shows a basic topological structure of ELM. Mathematically, ELM is written as follows.

TABLE 1. Statistics of first and second year engineering mathematics course data (2013–2018). The predictors (model's input) variables are: A_1 Assignment 1, A_2 Assignment 2, A_3 Assignment 3, Q_1 Quiz 1, Q_2 Quiz 2, while the target (weighted score, *WS*) represents the overall percentage score used to determine course grade attained at the University of Southern Queensland.

Statistical Property	Q_1 / 50	A_1 / 100	A_2 / 100	Q_2 / 50	A_3 / 100	EX / 600	Weighted Score (*WS*) /100%
ENM1600 Engineering Mathematics On-Campus (ONC)							
Mean	41.68	89.16	80.48	43.20	77.13	331.54	66.33
St. Dev.	7.95	13.08	17.27	8.27	22.05	137.19	16.76
Minimum	0.00	0.00	0.00	2.50	1.00	0.01	7.00
Maximum	50.00	100.00	100.00	50.00	100.00	600.00	99.30
Skewness	-1.36	-2.23	-1.39	-1.81	-1.13	0.08	-0.09
Flatness	2.51	6.91	2.11	3.54	0.53	-0.62	-0.33
ENM1600 Online (ONL)							
Mean	40.97	90.13	79.54	42.04	76.17	382.74	70.55
St. Dev.	8.00	11.55	15.96	8.43	21.07	133.32	17.38
Minimum	0.00	22.00	0.00	0.00	4.00	10.00	12.55
Maximum	50.00	100.00	100.00	50.00	100.00	600.00	100.00
Skewness	-1.26	-2.09	-1.13	-1.64	-1.11	-0.20	-0.40
Flatness	1.76	5.49	1.26	3.26	0.69	-0.71	-0.29

Statistical Property	Q_1 / 50	A_1 / 150	Q_2 / 50	A_2 / 150		EX / 600	Weighted Score (*WS*) /100%
ENM2600 Advanced Engineering Mathematics (ONC)							
Mean	46.63	121.08	44.75	129.55		295.11	63.74
St. Dev.	6.21	27.16	8.32	24.77		147.44	17.10
Minimum	10.00	0.01	1.50	0.01		2.00	13.00
Maximum	50.00	150.00	50.00	150.00		600.00	100.00
Skewness	-2.81	-1.65	-2.02	-2.31		0.24	0.06
Flatness	9.89	3.08	3.99	6.48		-0.88	-0.54
ENM2600 ONL							
Mean	45.12	115.50	38.90	117.27		302.02	61.91
St. Dev.	6.85	28.74	10.01	29.29		136.54	17.02
Minimum	10.00	1.00	0.00	1.00		0.01	11.00
Maximum	50.00	150.00	50.00	150.00		600.00	100.00
Skewness	-1.84	-1.18	-0.98	-1.16		0.09	-0.11
Flatness	3.80	1.09	0.59	1.02		-0.76	-0.41

For $i = 1, 2 \ldots N$, where N represents the student performance data the single feedforward hidden layer designated with L hidden neurons, $\Omega_L(\boldsymbol{x})$, can be expressed as [39]:

$$\Omega_L(x) = \sum_{i=1}^{L} h_i(x).\omega_i \quad (1)$$

In Eq. (1), $\omega = [\omega_1, \omega_2 \ldots \omega_L]^T$ is a weighted vector (or matrix) connecting the hidden layer with the output layer, $h_i(\boldsymbol{x})$ is the hidden neurons representing randomized hidden layer features based on how well target (*i.e.*, Grade) is related (linearly or nonlinearly) to each predictor (*e.g.*, Assignment), and $h(\boldsymbol{x}_i)$ is the i^{th} hidden neuron.

The feature space, that encloses the hidden neurons $h_i(x)$, can be defined as:

$$h_i(x) = \Phi(a_i, b_i, X) \quad \text{and} \quad a_i \in R^d,\ b_i \in R \quad (2)$$

The nonlinear piecewise-continuous activation function $h_i(\boldsymbol{x})$ is defined using neuron parameters $(\mathbf{a}, b)$ that satisfy universal approximation theorem: $\Phi(\mathbf{a}_i, b_i, \boldsymbol{X})$. In the present paper, an optimal ELM model was designed by evaluating a set of popular hidden layer activation functions (as feature estimation strategy) to fit the predictors to target variable [39]. These functions are as follows:

Tangent Sigmoid:

$$\Phi(a, b, X) = \frac{2}{1 + exp(-2(-aX + b))} \quad (3)$$

Logarithmic Sigmoid:

$$\Phi(a, b, X) = \frac{2}{1 + exp(-aX + b)} \quad (4)$$

Hard Limit:

$$\Phi(a, b, X) = 1 \quad \text{if } aX + b > 0 \quad (5)$$

Triangular Basis:

$$\Phi(a, b, X) = 1 - |-aX + b| \quad \text{if } -1 \leq -aX + b \leq 1, \text{ or 0 otherwise} \quad (6)$$

Radial Basis:

$$\Phi(a, b, X) = exp\left(-(-aX + b)^2\right) \quad (7)$$

In feature space (*i.e.*, the hidden layer), the ELM model approximation error is minimized when solving for the input weights that connect the hidden layer, using a least square method [39]:

$$\underset{\omega \in R^{L\times M}}{\text{minimize}} \|H\omega - T\|^2 \quad (8)$$

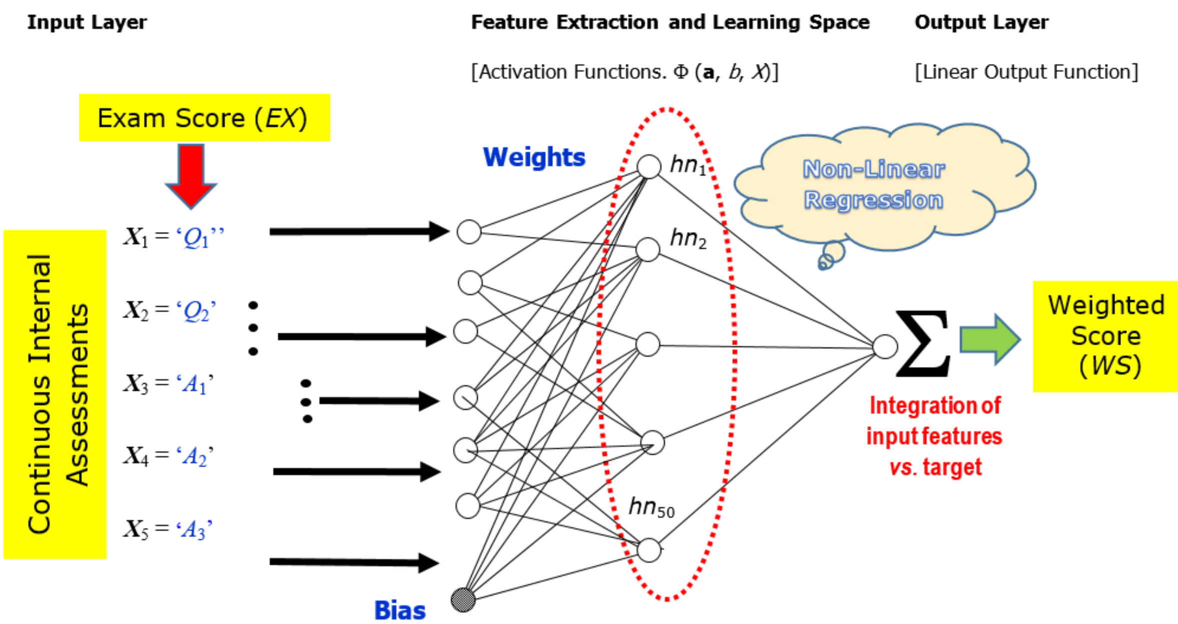


FIGURE 1. Schematic view of an extreme learning machine (ELM) model designed to predict weighted scores using continuous internal assessment and exam scores.

In Eq. (8) || || is the Frobenius (Euclidean) norm deduced as the sum square of absolute squares of the elements therein and **H** is the hidden layer output matrix [39]:

$$H = \begin{bmatrix} q(x_1) \\ \vdots \\ q(x_M) \end{bmatrix} = \begin{bmatrix} q_1(a_1x_1 + b) & \cdots & q_L(a_Lx_1 + b_L) \\ & \vdots & \\ q_1(a_Mx_M + b) & \cdots & q_L(a_Lx_M + b_L) \end{bmatrix} \quad (9)$$

The term **T** is a target matrix (i.e., predicted *EX* or *WS*) in the training dataset.

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & \cdots & t_{1m} \\ & \vdots & \\ q_{N1} & \cdots & t_{Nm} \end{bmatrix} \quad (10)$$

By solving a system of linear equations the ELM model generates an optimal solution for the output space [39]:

$$\omega = H^+T \quad (11)$$

where $\mathbf{H}^+$ is the matrix pseudoinverse or the Moore-Penrose generalized inverse function (+).

For details on the ELM model and its variants, readers can consult a recent review [39]. It is worth to mention, the input weights, output weights and bias of hidden neurons for the EX and WS, were tabulated in the Appendix A.

2) BASELINE MODEL 1: *RANDOM FOREST*

To evaluate the relative utility and statistical accuracy of ELM as a predictive modelling tool for student performance, a random forest (RF) algorithm that has the ability to generate accurate predictions with minimal overfitting, was developed. RF is a novel learning algorithm that relies on model aggregation principles [44], and is fundamentally different from a neuronal-based ELM system. It combines binary decision trees built with bootstrapped training samples from learning sample *D*, where a subset of explanatory variable ***X*** is chosen randomly at each node of a designated tree.

In an RF-based predictive model, the aggregation of model trees, typically exceeding 2,000, is facilitated where a training matrix is generated from bootstrapped samples on two thirds of the overall sample whereas one third of the training sample is left for validation purposes (or the 'out-of-bag' predictions). The branching of a random forest (or its trees) is performed on a randomized predictive framework with a mean outcome represented by an aggregated predictive model [45]. Since RF uses out of bag training samples to determine the model's error and is associated with independent observations to grow the decision tree, no cross-validation data are required [46]. The key steps are as follows:

i. Suppose *N* is a multi-dimensional training matrix with continuous assessments, (predictors) and a *WS* (target). To grow trees in random forest, a sample of these training cases is drawn with replacement.
ii. Depending on ζ predictors, RF considers $n < \zeta$ samples, which are drawn with replacement out of the ζ dataset and the best number of splits are adopted, fixing *n* as a constant as the forest evolves in size.
iii. The trees are split with oblique hyperplanes for better accuracy to enable them to grow without suffering from overtraining (by trees randomly restricted to be sensitive to only selected feature dimensions).

iv. The new (test) data are independently predicted by aggregating predictions of all trees where a mean is determined for a regression problem.

In RF the 'out of bag estimate' of generalization error is considerably low as long as a relatively large number of decision trees are grown [47].

3) BASELINE MODEL 2: *SECOND ORDER VOLTERRA*

The mathematical rule-based Volterra model, is a higher order extension of linear impulse response model built on Taylor series expansion for nonlinear, autonomous and causal systems [48]. If $X(i)$ is a target variable (*e.g.*, *WS*) where i represents the label of student performance data, the second order Volterra model is expressed as [49]:

$$Z(i) = \int_{\tau=0}^{\tau=i} k_1(\tau_1) X(\tau - \tau_1) d\tau_1 + \int_{\tau_2=0}^{\tau_2=i} \int_{\tau_1=0}^{\tau_1=i} k_2(\tau_1 \tau_2) X(\tau - \tau_1) X(\tau - \tau_2) d\tau_1 d\tau_2 \quad (12)$$

where, $k_1(\tau_1)$ and $k_2(\tau_1, \tau_2)$ are the Volterra kernel functions that can be condensed into the form:

$$Z(t) = K_1[x(i)] + K_2[x(i)] \quad (13)$$

where $K_1[x(i)]$ and $K_2[x(i)]$ are the 1st and 2nd order Volterra operators, respectively.

In this paper, the prediction of *WS* is driven by the multiple predictors drawn from internal assessment data. Hence, Volterra series expansion for multiple inputs *vs*. a single output (MISO approach) is written as:

$$Z(i) = \sum_{n=1}^{D} \sum_{\delta=1}^{E} k_1^{(n)} \delta x_n(t-\delta) + \sum_{n=1}^{D} \sum_{\kappa=1}^{E} \sum_{\delta=1}^{E} k_{2s}^{(n)}(\kappa, \delta) x_n(t-\kappa) x_n(t-\delta) + \sum_{n=1}^{D} \sum_{n2=1}^{n1-1} \sum_{\kappa=1}^{E} \sum_{\delta=1}^{E} k_{2\times}^{(n1.n2)}(\kappa, \delta) \times x_{n1}(t-\kappa) x_{n2}(t-\delta) \quad (14)$$

where, D *is* the number of predictor (input) variables, E is the memory length of each significant lagged input, $k_1^{(n)}$ are the first order kernels; $k_{2s}^{(n)}$ is the second order self kernel and $k_{2\times}^{(n1,n2)}$ is the second order cross-kernel.

In final step, the prediction of Volterra kernel is achieved by an orthogonal least squares (OLS) method that can handle collinearity amongst predictors [50], [51].

III. DEVELOPMENT OF THE MODELING FRAMEWORK

A. INPUT VARIABLE SENSITIVITY ASSESSMENT AND PREDICTIVE MODEL DEVELOPMENT

MATLAB (2017b) software running on Intel(R) Core i7-4770 CPU 3.4 GHz, Windows 10 platform was adopted in modelling student performance datasets. To predict *WS* and *EX* using an appropriate combination of input variables for engineering mathematics courses, the ELM model incorporated 6 predictors, $\boldsymbol{X} = [Q_1, A_3, A_2, Q_2, A_1, EX]$ for the case of ENM1600, and 5 predictor variables, $\boldsymbol{X} = [Q_2, Q_1, A_2, A_1, EX]$ for the case of ENM2600. *Note that for the latter course, it comprised of only two assignments and two quizzes as its internal assessment dataset*. In each case, the target variable was set to either *WS* or *EX*, and the performance was compared against RF and Volterra designed with the same predictors.

As a pivotal part of ELM model design the relationship between all possible predictors (*i.e.*, continuous internal assessment) *vs WS* was explored with cross-correlation between X_i and Y_i in the training data where Pearson correlation coefficient (r_{cross}) was recorded to measure the similarity and covariance. A bar graph of r_{cross} of each variable organized in order of their magnitudes on the principal axis (Figure 2) shows interesting patterns in terms of the relative strength of each internal assessment used to predict *WS*. This further indicates a clear distinction between two courses, ENM1600 and ENM2600. While, as expected, *EX* remains the most significant contributor towards *WS*, the importance of each variable in terms of its correlation with *WS* follows a different order for ENM2600. For instance, considering the ONC offer, the importance of Q_1 and Q_2 appears to be the weakest for ENM2600 (with $r_{cross} \approx 0.300$ & 0.205, respectively) whereas for ENM1600 the role of Q_2 appears quite significant ($r_{cross} \approx 0.557$) and that for Q_1 is 0.495, although they are not particularly high. In fact, the correlation between continuous internal assessments and *WS* seems to be more evenly distributed with r_{cross} between 0.495–0.569, and 0.547–0.613, whereas for ENM2600, r_{cross} lies between 0.205–0.547 and 0.382–0.604 for ONC and ONL offers, respectively, which again, remains much lower.

Based on the above, the most appropriate order of input variables was identified for ELM (and its counterpart) models, and input variables were added successively, to the predictor matrix from the lowest to the highest r_{cross} (Figure 2). This resulted in the ELM model's input orders as follows.

- ➢ $Q_1, A_3, A_2, Q_2, A_1, EX$ to predict *WS* for ONC, and
- ➢ $A_2, A_1, Q_1, A_3, Q_2, EX$ to predict *WS* for ONL for ENM1600.
- ➢ Q_2, Q_1, A_2, A_1, EX to predict *WS* for ONC, and
- ➢ Q_1, Q_2, A_2, A_1, EX to predict *WS* for ONL, for ENM2600

For more details, readers should consult Table 3(a–c).

While no specific rule exists for dividing the data for calibration and validation of an AI model, the majority of data are normally employed in the former with a reasonable amount of remainder data employed to validate and test the final results [52]. Accordingly, after data were normalized to be bounded by [0, 1], they were partitioned into training (60%), validation (20%) and testing (20%) sets with $n = 486, 162, 162$ (ONC) and $n = 767, 256, 255$ (ONL) for ENM1600, and $n = 455, 149, 149$ (ONC) and $n = 429, 143, 143$ (ONL) for ENM2600 (Table 2). Notably, the training data provided key features from the continuous internal assessments *vs*. the target variable; the validation set ensured that the optimal model was selected, and the testing data provided the independent data of training and validation to evaluate the selected model.

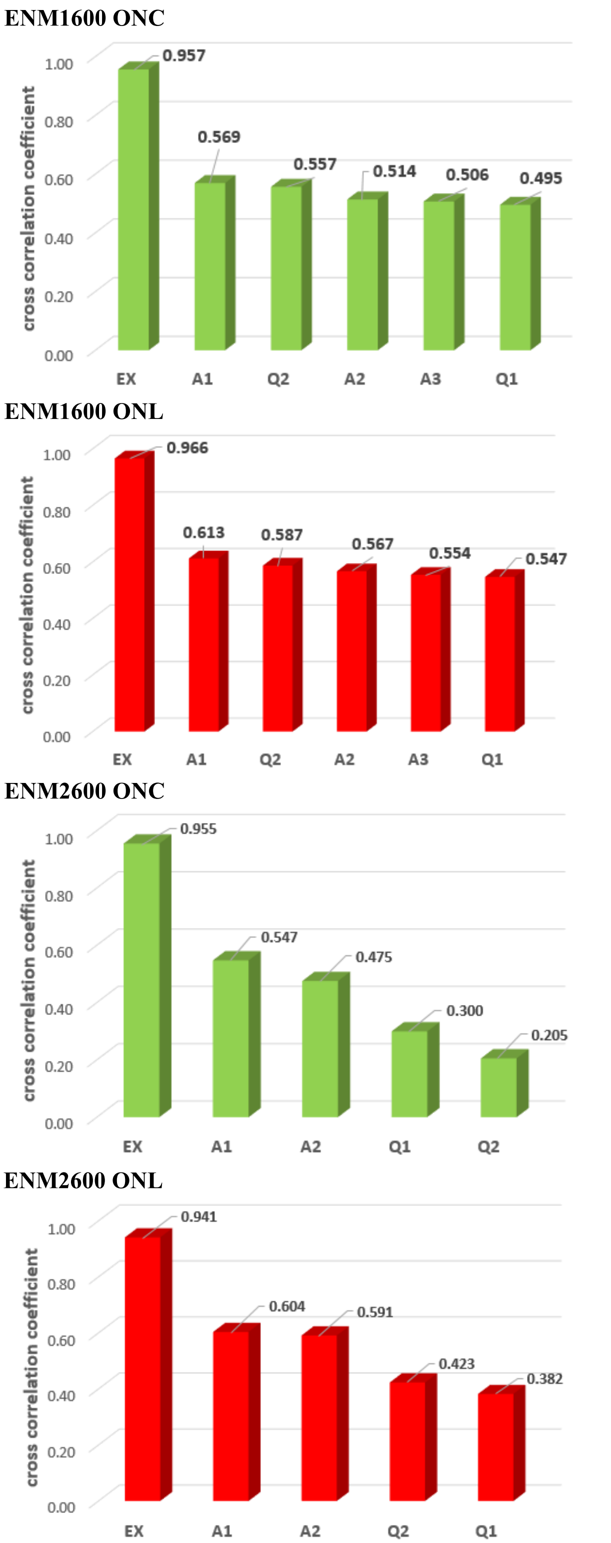


FIGURE 2. Cross correlation analysis of continuous internal assessment as a predictor variable for a weighted score (*WS*) in ENM1600 Engineering Mathematics and ENM2600 Advanced Engineering Mathematics on-campus (ONC) and the online course offers (ONL). Note that predictor variables are presented in order of their importance in respect to predicting *WS*. *Notations: EX = final exam mark, A1 = Assignment 1, A2 = Assignment 2, A3 = Assignment 3, Q1 = Quiz 1, Q2 = Quiz 2.*

In accordance with theoretical framework (Section 2.1) a three-layer neuronal network was constructed to define an input layer (where continuous internal assessment data were fed), a hidden layer (where model-extracted input features regressed against the *WS* or the *EX* for data mining) and the output layer (where the predicted *WS* or *EX* was generated). In the first step the ELM model was assigned hidden neurons following a rule 1 to $n + 1$ (increments of 1; $n =$ size of predictor data) with cross-validation optimisations used to deduce the optimal number of hidden neurons. To ensure optimal feature weights to be generated from predictor data, 5 different activation functions (Eq. 3–7) were tested. The objective criterion: mean square error (*MSE*) was monitored in each trial and each neuronal architecture was evaluated on a validation set (20% of the entire data). After the optimal number of hidden neurons were determined for each input combination (that was not surprisingly unique for every predictor variable), the ELM model was executed 1,000 times to generate an optimal neuronal layer, predictions with the smallest *MSE* generated for the validation set, and the final model runs on independent test set.

Table 3(a) displays the model's design parameters including the training and the validation errors attained by the optimal ELM model. To explore the credibility of the objective model (*i.e.*, ELM predictions), an RF model executed with the same inputs and target data was designed, where an ensemble decision tree was developed to regress the exploratory (*i.e.*, predictor) and the response (target) relationships. A sufficiently large number of ($T = 800$) trees, with leaf size 5 and Fboot set to 1 were used, out-of-bag (OOB) permuted change in error (E_DD) was monitored for each predictor regressed against target, and the model with the lowest error was selected. An ensemble result was recorded for the model applied in the testing phase. As a comparative tool a second order Volterra model was also constructed using orthogonal least squares approach. Table 3 (b–c) shows the optimal RF and the Volterra model's training and validation parameters.

B. MODEL PERFORMANCE CRITERIA

We adopted an appropriate combination of visual and descriptive statistics (*i.e.*, observed & predicted test data) to cross-check the discrepancies in terms of minimum, maximum, mean, variance, skewness and kurtosis, as well as the standardized performance metrics, were employed to comprehensively evaluate the credibility of ELM for the prediction of *WS* and *EX* in engineering mathematics courses. In this study we adopt metrics for the AI model evaluation that are recommended by the American Society for Civil Engineers [53], namely: the mean absolute error (*MAE*), mean absolute percentage error (*MAPE*; %), root mean square error (*RMSE*), relative root mean square error (*RRMSE*, %), correlation coefficient (r), Legate & McCabe's Index (*LM*), Nash Sutcliffe's Coefficient (*NS*), Willmott's Index (d), and

TABLE 2. Details of data used in designing artificial intelligence models for two engineering mathematics courses and their modes of offer at the University of Southern Queensland.

Course / Dataset	Course Name/Level	Data (after removal of incomplete or missing data)	Training (60%)	Validation (20%)	Testing (20%)
ENM1600 ONC (2013 to 2018)	Engineering Mathematics	810	486	162	162
ENM1600 ONL		1278	767	256	255
ENM2600 ONC (2014 to 2018)	Advanced Engineering Mathematics	742	445	149	148
ENM2600 ONL		715	429	143	143

the relative prediction error (%), given mathematically as follows [54], [55], (15)–(22), as shown at the bottom of the page, where WS_{Obs} and WS_{Pred} were the observed and the predicted i^{th} value of WS; $\overline{WS}_{Obs}$ and $\overline{WS}_{\text{Pred}}$ were the observed and forecasted mean WS in testing phase; and N was the number of datum points in the test set. Note that alternatively, the WS would revert to the EX (as a target) when examination scores were predicted using continuous internal assessments data as the predictor variable(s).

IV. MODELING RESULTS

In this section the results generated from AI models (*i.e.*, ELM & RF) and the second order Volterra model (a mathematical-based nonlinear extension of a linear convolution model) designed to predict engineering mathematics student performances at USQ, an Australian regional university, are appraised. In all predictive models, the official data from Examiner Returns for on-campus and online courses in ENM1600 and ENM2600, are modelled [56], [57]. In particular, the results are used to ascertain whether the optimized ELM model was able to accomplish an acceptable level of accuracy in predicting the WS and the EX, both of which are the key measures used to determine an overall passing grade and the grade point average (GPA) in the program of study. For effective feature extraction process drawn on historical student performance data, a total of five years (ENM2600) and six years (ENM1600), spanning back to a period when these courses were first introduced, were considered in terms of ELM model's training, model selection (validation), and final testing phases, performed through a rigorous approach (Section 3.2).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\left(WS_{pred,i} - WS_{Obs,i}\right)\right| \tag{15}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\left(WS_{pred,i} - WS_{Obs,i}\right)}{WS_{Obs,i}}\right| \times 100 \tag{16}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(WS_{pred,i} - WS_{Obs,i}\right)^2} \tag{17}$$

$$RRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(WS_{pred,i} - WS_{Obs,i}\right)^2}}{\frac{1}{N}\sum_{i=1}^{N}\left(WS_{Obs,i}\right)} \times 100 \tag{18}$$

$$r = \left(\frac{\sum_{i=1}^{N}\left(WS_{pred,i} - \overline{WS}_{Obs,i}\right)\left(WS_{pred,i} - \overline{WS}_{Obs,i}\right)}{\sqrt{\sum_{i=1}^{N}\left(WS_{pred,i} - \overline{WS}_{Obs,i}\right)^2}\sqrt{\sum_{i=1}^{N}\left(WS_{pred,i} - \overline{WS}_{Obs,i}\right)^2}}\right) \tag{19}$$

$$LM = 1 - \left[\frac{\sum_{i=1}^{N}\left|WS_{Obs,i} - WS_{pred,i}\right|}{\sum_{i=1}^{N}\left|WS_{Obs,i} - \overline{WS}_{Obs,i}\right|}\right], \quad 0 \leq LM \leq 1 \tag{20}$$

$$NS = 1 - \left[\frac{\sum_{i=1}^{N}\left(WS_{Obs,i} - WS_{pred,i}\right)^2}{\sum_{i=1}^{N}\left(WS_{Obs,i} - \overline{WS}_{Obs,i}\right)^2}\right], \quad -\infty \leq NS \leq 1 \tag{21}$$

$$d = 1 - \left[\frac{\sum_{i=1}^{N}\left(WS_{pred,i} - WS_{Obs,i}\right)^2}{\sum_{i=1}^{N}\left(\left|WS_{pred,i} - \overline{WS}_{Obs,i}\right| + \left|WS_{Obs,i} - \overline{WS}_{pred,i}\right|\right)^2}\right], \quad 0 \leq d \leq 1 \tag{22}$$

TABLE 3. (a) The optimal training parameters of extreme learning machine (ELM) model designed to predict the weighted scores (*WS*) and the exam scores (*EX*). Acronyms for model inputs are as per Table 1, and the input order was determined by cross correlation analysis of each predictor variable against *WS*.

Test Case	Model Parameters		Training Error			Validation Error		
	Model Property	Value	*r*	*RMSE* (%)	*MAE* (%)	*r*	*RMSE* (%)	*MAE* (%)
[ENM1600, ONC]	Hidden Layer	3	0.9999	0.2415	0.1846	0.9998	0.3760	0.2860
	Input Neurons	1, 2, …, 6						
	Inputs (*correct order*)	Q_1, A_3, A_2, Q_2, A_1, *EX*						
	Hidden Neurons	10, 20, …, 150						
	Output Neurons	1 (≡ predicted weighted score, *WS*)						
	Activation Function	***logarithmic sigmoid***						
	Architecture	6 – 11 – 1 (input-hidden-output)						
[ENM1600, ONL]		3	0.9999	0.2405	0.1687	0.9998	0.3451	0.2884
		1, 2, …, 6						
		A_2, A_1, Q_1, A_3, Q_2, *EX*						
		10, 20, …, 150						
		1 (≡ predicted weighted score, *WS*)						
		logarithmic sigmoid						
		6 – 12 – 1 (input-hidden-output)						
[ENM2600, ONC]		3	0.9999	0.2415	0.1846	0.9998	0.3760	0.2860
		1, 2, …, 5						
		Q_2, Q_1, A_2, A_1, *EX*						
		10, 20, …, 150						
		1 (≡ predicted weighted score, *WS*)						
		logarithmic sigmoid						
		5 – 8 – 1 (input-hidden-output)						
[ENM2600, ONL]		3	0.9999	0.2965	0.2126	0.9998	0.3358	0.2883
		1, 2, …, 5						
		Q_1, Q_2, A_2, A_1, *EX*						
		10, 20, …, 150						
		1 (≡ predicted weighted score, *WS*)						
		logarithmic sigmoid						
		5 – 8 – 1 (input-hidden-output)						

(b) Random Forest model

Test Case	Model Parameters			Training Error			Validation Error		
	Tree Bagger Property	Sym.	Value	*r*	*RMSE* (%)	*MAE* (%)	*r*	*RMSE* (%)	*MAE* (%)
[ENM1600, ONC]	Leaf	*L*	5	0.995	2.059	1.218	0.992	2.582	1.809
	Decision Trees	*T*	800						
	FBoot	*F*	1						
	Input (*correct order*)	*X*	Q_1, A_3, A_2, Q_2, A_1, *EX*						
	OOB Permuted Delta Error	E_DD	0.721, 0.780, 0.699, 0.995, 0.960, 3.406						
	No. Predictor Split	N_P	12799, 15295, 12829, 16485, 16478, 18329						
	Delta Criterion Decision	*C*	0.028, 0.035, 0.031, 0.038, 0.041, 0.131						
[ENM1600, ONL]		*L*	5	0.996	1.777	1.044	0.990	2.977	1.814
		T	800						
		F	1						
		X	A_2, A_1, Q_1, A_3, Q_2, *EX*						
		E_DD	0.701, 0.879, 0.793, 1.143, 1.110, 4.574						
		N_P	20483, 24125, 20208, 26570, 26294, 28276						
		C	0.026. 0.032, 0.032, 0.039, 0.041, 0.131						
[ENM2600, ONC]		*L*	5	0.996	1.870	1.106	0.986	3.381	2.142
		T	800						
		F	1						
		X	Q_2, Q_1, A_2, A_1, *EX*						
		E_DD	0.316, 0.613, 1.179, 1.235, 4.6132						
		N_P	9381, 15942, 17635, 17971, 19698						
		C	0.014, 0.023, 0.052, 0.054, 0.159						
[ENM2600, ONL]		*L*	5	0.997	1.855	1.171	0.994	2.109	1.459
		T	800						
		F	1						
		X	Q_1, Q_2, A_2, A_1, *EX*						
		E_D	0.309, 0.355, 0.976, 1.168, 5.096						
		N_P	7224, 10312, 17476, 18196, 20399						
		C	0.008, 0.010, 0.029, 0.043, 0.017						

TABLE 3. ***(Continued.)*** **(a) The optimal training parameters of extreme learning machine (ELM) model designed to predict the weighted scores (*WS*) and the exam scores (*EX*). Acronyms for model inputs are as per Table 1, and the input order was determined by cross correlation analysis of each predictor variable against *WS*.**

(c) Volterra Model

Test Case	Model Parameters			Training Error			Validation Error		
	Volterra Property	Sym.	Value	*r*	*RMSE* (%)	*MAE* (%)	*r*	*RMSE* (%)	*MAE* (%)
[ENM1600, ONC]	Order	*O*	2^{nd} Order	1.000	0.382	0.216	1.000	0.478	0.317
	Type	*T*	Orthogonal Least Squares (OLS)						
	Index Best Regressor	*Ind*	6, 4, 16, 1, 3, 5						
[ENM1600, ONL]		*O*	2^{nd} Order	1.000	0.440	0.271	1.000	0.465	0.362
		T	Orthogonal Least Squares (OLS)						
		Ind	6, 16, 4, 9, 24, 5						
[ENM2600, ONC]		*O*	2^{nd} Order	1.000	0.473	0.290	0.998	1.217	0.558
		T	Orthogonal Least Squares (OLS)						
		Ind	5, 16, 2, 8, 4						
[ENM2600, ONL]		*O*	2^{nd} Order	0.999	0.947	0.625	0.998	0.930	0.623
		T	Orthogonal Least Squares (OLS)						
		Ind	5, 16, 2, 1, 3						

Using performance metrics in Table 4 the three tested models' accuracies in predicting *WS* using single internal assessments, including Quiz 1 (Q_1), Quiz 2 (Q_2), Assignment 1 (A_1), Assignment 2 (A_2), or alternatively, Assignment 3 (A_3) data as the predictor variables were assessed. Except for ENM1600 for ONC (that had used A_1 or $Q1$ as the predictors), the metrics point out a better predictive capacity of the ELM than either the RF or the Volterra model when single input variables were used. For ENM1600 ONC-based model that used A_1 to predict the *WS*, the Nash-Sutcliffe and the Legates & McCabe's Index were the highest for the case of ELM (*i.e.*, 0.361 & 0.212, respectively) compared to the RF and the Volterra model, although the r-values, *RMSE*, and *MAE* indicated the RF to be the best model used to predict *WS* using a single predictor variable.

When single predictor variable based models for ENM1600, ONL were considered, ELM consistently outperformed the RF and the Volterra model, yielding the highest r values between the predicted and observed *WS* values in the testing phase, the smallest *RMSE*, *MAE*, and their relative percentage errors, as well as the largest Willmott's, Nash-Sutcliffe and Legate & McCabe's indices. Interesting patterns emerge when the model accuracies utilizing the Quiz and Assignment (as the model's predictors) were examined, where among the Quizzes, Q_1 led to a more accurate ELM model compared to Q_2, but among the Assignments, A_3 generated a more accurate model relative to A_1 as a single predictor variable.

The internal assessment Q_1, assigned to the students prior to Q_2 during the teaching semester seemed to provide a greater weight towards the model development whereas A_3, the last internal assessment prior to the examination period provided a greater weight towards modelling the *WS* relative to A_1 or A_2. While the exact reason for this not clear yet, this does indicate that an ELM model for student performance prediction is more likely to be influenced by Q_1 and A_3 than other internal assessment pieces (*i.e.*, Q_2, A_1, A_2). In terms of the overall contributory influence of Assignments and Quizzes in predicting the *WS* for ENM1600 ONL, the greatest contribution to the ELM model came from Q_1 (with $LM \approx 0.231$) and the smallest contribution from A_1 ($LM \approx 0.174$), while *EX* remains the most significant contributor to predict *WS* ($LM \approx 0.770$ and relative $RMSE \approx 5.915\%$). However, comparing the online and on-campus offers of engineering mathematics courses, ELM model registered a better performance metric for ENM1600 ONL than ENM1600 ONC offer (*i.e.*, $LM \approx 0.770$ *vs*. 0.732; relative RRMSE $\approx$ 5.915% *vs*. 8.51%).

Comparing student performance predictions for advanced engineering mathematics, significant differences among the models trained to predict the student performance using single predictor variables exemplify their individual contributory influence (Table 5). With respect to the influence of Quizzes on prediction of *WS*, Q_1 was observed to be a much better predictor of student performance for ENM2600 ONC, whereas Q_2 was a better predictor for its ONL equivalent offer. That is, for ENM2600 ONC, the ELM model generated an *RRMSE* of 27.79% *vs*. 28.68% for a model trained with Q_1 *vs*. Q_2, respectively, whereas for ENM2600 ONL, the corresponding *RRMSE* values were 24.73% *vs*. 24.11%, respectively. These also concurred with other performance metrics such as *LM, NS,* and d. In terms of the contributory

TABLE 4. Influence of each predictor (*i.e.* continuous assessment) incorporated to predict weighted score (*WS*) by ELM *vs.* RF and Volterra models in *ENM1600 Engineering Mathematics* in the testing phase. The optimal model is red/boldfaced. [A_1 = assignment 1; A_2 = assignment 2; Q_1 = Quiz 1; Q_2 = Quiz 2; *EX* = exam score, *r* = correlation coefficient, *RMSE* root mean square error, *MAE* mean absolute error, *d* Willmott's Index, *NS* Nash-Sutcliffe's coefficient, *LM* Legate & McCabe's Index].

Predictor Variable	Model	*r*	*RMSE*	*MAE*	*RRMSE, %*	*RMAE, %*	*d*	*NS*	*LM*
ENM1600 ONC									
A_1	ELM	0.638	14.59	11.75	21.85	19.13	0.988	**0.361**	**0.212**
	RF	**0.539**	**12.95**	**10.89**	**17.96**	**16.20**	0.992	0.267	0.114
	Volterra	0.611	14.86	11.99	22.24	19.67	0.988	0.337	0.196
A_2	**ELM**	**0.609**	**15.31**	**12.43**	**22.92**	**22.18**	**0.987**	**0.297**	**0.167**
	RF	0.505	13.43	11.20	18.63	16.59	0.991	0.211	0.089
	Volterra	0.603	15.83	13.00	23.70	21.93	0.985	0.248	0.129
A_3	**ELM**	**0.634**	**15.12**	**12.37**	**22.63**	**22.95**	**0.987**	**0.314**	**0.171**
	RF	0.474	13.61	11.23	18.88	16.71	0.991	0.189	0.086
	Volterra	0.639	16.76	13.71	25.09	21.92	0.983	0.156	0.081
Q_1	ELM	0.543	15.76	12.86	23.60	25.00	0.986	0.254	0.138
	RF	**0.605**	**12.42**	**10.30**	**17.22**	**15.12**	**0.992**	**0.325**	**0.162**
	Volterra	0.531	15.77	12.84	23.60	22.99	0.986	0.254	0.139
Q_2	**ELM**	**0.647**	**15.34**	**12.56**	**22.97**	**19.73**	**0.986**	**0.293**	**0.158**
	RF	0.527	13.15	11.10	18.24	16.61	0.992	0.243	0.097
	Volterra	0.664	16.47	13.39	24.66	20.26	0.984	0.185	0.102
EX	**ELM**	**0.966**	**5.68**	**4.00**	**8.51**	**9.78**	**0.998**	**0.903**	**0.732**
	RF	0.959	4.38	3.33	6.08	5.23	0.999	0.916	0.729
	Volterra	0.965	8.60	6.97	12.88	11.61	0.996	0.778	0.533
ENM1600 ONL									
A_1	ELM	**0.614**	**13.59**	**11.56**	**18.57**	**17.12**	**0.991**	**0.345**	**0.174**
	RF	0.609	13.62	11.56	18.61	17.08	0.991	0.342	0.174
	Volterra	0.589	13.83	11.75	18.90	17.56	0.991	0.321	0.160
A_2	**ELM**	**0.621**	**13.67**	**11.41**	**18.68**	**17.05**	**0.991**	**0.337**	**0.184**
	RF	0.594	13.91	11.67	19.01	17.36	0.991	0.313	0.166
	Volterra	0.609	13.70	11.44	18.72	17.06	0.991	0.334	0.182
A_3	**ELM**	**0.664**	**12.98**	**10.97**	**17.73**	**16.20**	**0.992**	**0.402**	**0.216**
	RF	0.626	13.60	11.29	18.58	16.58	0.991	0.344	0.193
	Volterra	0.654	16.46	13.63	22.50	20.92	0.987	0.038	0.026
Q_1	**ELM**	**0.649**	**13.03**	**10.78**	**17.80**	**16.36**	**0.992**	**0.398**	**0.230**
	RF	0.652	12.96	10.79	17.71	16.29	0.992	0.404	0.229
	Volterra	0.631	13.20	10.96	18.04	16.50	0.992	0.381	0.216
Q_2	**ELM**	**0.659**	**13.43**	**11.31**	**18.36**	**16.27**	**0.991**	**0.359**	**0.192**
	RF	0.603	14.21	11.75	19.42	16.82	0.990	0.283	0.160
	Volterra	0.658	13.48	11.22	18.41	15.92	0.991	0.355	0.198
EX	**ELM**	**0.969**	**4.328**	**3.213**	**5.915**	**5.290**	**0.999**	**0.933**	**0.770**
	RF	0.967	4.485	3.449	6.129	5.635	0.999	0.929	0.754
	Volterra	0.969	5.677	4.702	7.758	7.134	0.999	0.886	0.664

TABLE 5. Influence of each predictor (*i.e.* continuous assessment) incorporated to predict weighted score (*WS*) by ELM *vs.* RF and Volterra models in *ENM2600 Advanced Engineering Mathematics* in the testing phase. The optimal model is blue/boldfaced. [A_1 = assignment 1; A_2 = assignment 2; Q_1 = Quiz 1; Q_2 = Quiz 2; *EX* = exam score, *r* = correlation coefficient, *RMSE* root mean square error, *MAE* mean absolute error, *d* Willmott's Index, *NS* Nash-Sutcliffe's coefficient, *LM* Legate & McCabe's Index].

Input	Model	*r*	*RMSE*	*MAE*	*RRMSE, %*	*RMAE, %*	*d*	*NS*	*LM*
ENM2600 ONC									
A_1	ELM	**0.649**	**15.41**	**12.67**	**23.21**	**21.01**	**0.986**	**0.346**	**0.197**
	RF	0.587	0.19	0.15	30.54	65535	0.976	0.267	0.161
	Volterra	0.626	16.07	13.05	24.20	20.89	0.985	0.288	0.172
A_2	ELM	0.687	16.45	13.60	24.78	20.61	0.983	0.254	0.138
	RF	0.627	16.81	13.96	25.32	21.57	0.983	0.221	0.115
	Volterra	**0.691**	**16.38**	**13.55**	**24.67**	**20.05**	**0.983**	**0.261**	**0.141**
Q_1	ELM	0.399	**18.45**	**15.57**	**27.79**	**26.86**	**0.980**	**0.062**	**0.013**
	RF	0.394	0.21	0.18	34.53	65535	0.968	0.064	0.012
	Volterra	0.408	18.68	15.87	28.14	26.42	0.979	0.038	-0.007
Q_2	ELM	**0.344**	**19.04**	**15.89**	**28.68**	**28.44**	**0.978**	**0.001**	**-0.008**
	RF	0.303	19.01	15.82	28.64	27.27	0.978	0.004	-0.003
	Volterra	0.343	20.08	16.98	30.24	28.27	0.975	-0.111	-0.077
EX	ELM	**0.948**	**6.71**	**4.44**	**10.11**	**10.40**	**0.998**	**0.876**	**0.719**
	RF	0.946	6.85	4.59	10.32	10.58	0.998	0.871	0.709
	Volterra	0.949	9.42	7.87	14.19	13.63	0.996	0.756	0.501
ENM2600 ONL									
A_1	ELM	**0.517**	**14.44**	**12.01**	**22.41**	**20.22**	**0.987**	**0.250**	**0.134**
	RF	0.496	14.93	12.27	23.18	20.44	0.987	0.198	0.115
	Volterra	0.518	14.73	12.13	22.87	20.22	0.987	0.219	0.125
A_2	ELM	**0.587**	**14.12**	**11.83**	**21.92**	**19.31**	**0.988**	**0.283**	**0.147**
	RF	0.581	14.31	11.72	22.22	19.19	0.987	0.263	0.155
	Volterra	0.587	14.43	12.22	22.41	19.84	0.987	0.251	0.119
Q_1	ELM	0.388	**15.93**	**13.36**	**24.73**	**22.01**	**0.984**	**0.087**	**0.036**
	RF	0.361	15.99	13.34	24.82	22.12	0.984	0.081	0.038
	Volterra	0.409	16.68	13.97	25.89	22.59	0.982	0.000	-0.007
Q_2	ELM	**0.426**	**15.53**	**12.95**	**24.11**	**21.67**	**0.985**	**0.132**	**0.066**
	RF	0.354	16.31	13.58	25.32	22.93	0.983	0.044	0.020
	Volterra	0.421	16.73	13.87	25.97	22.83	0.982	-0.006	0.000
EX	ELM	**0.942**	**5.61**	**4.31**	**8.72**	**8.26**	**0.998**	**0.887**	**0.689**
	RF	0.942	5.64	4.46	8.76	8.25	0.998	0.886	0.678
	Volterra	0.943	8.48	6.92	13.17	12.18	0.996	0.741	0.501

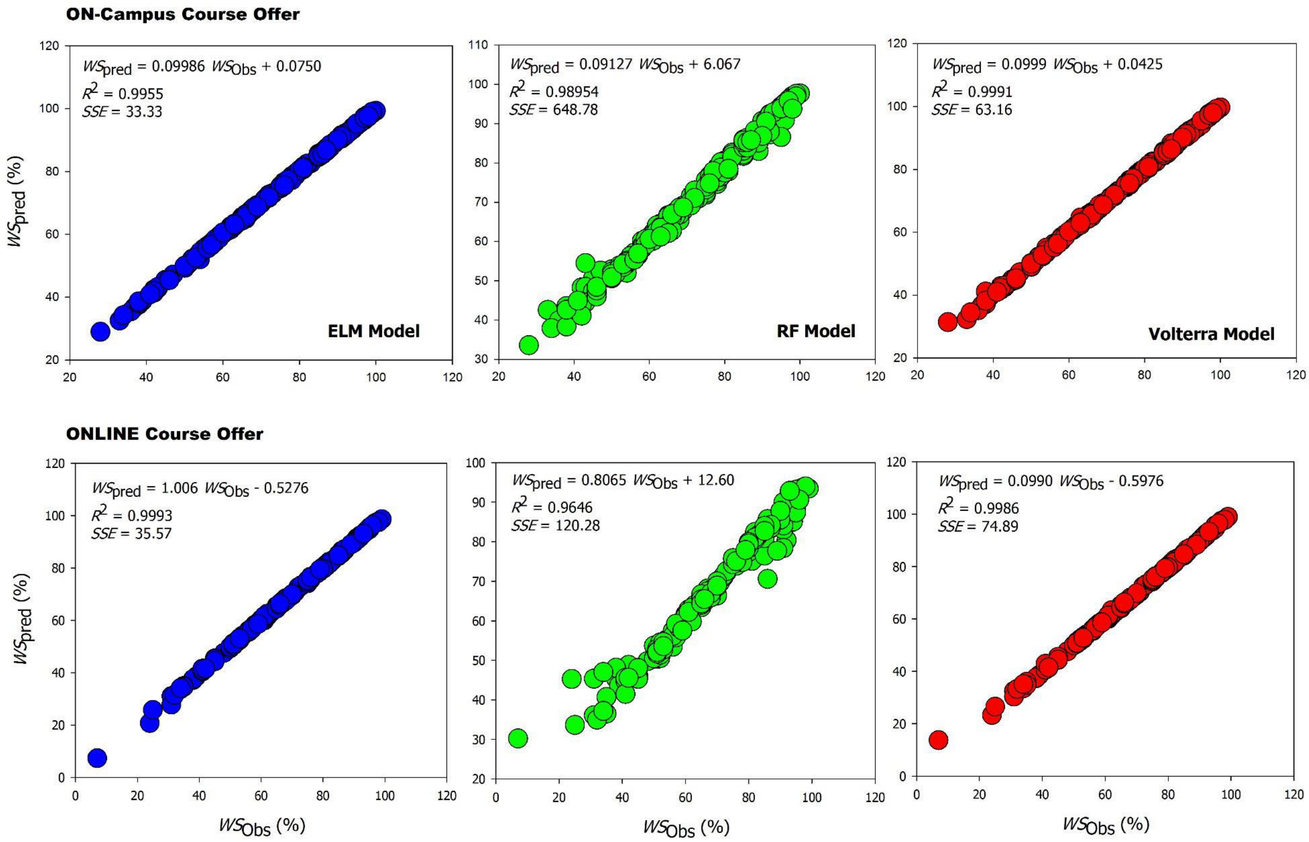


FIGURE 3. Scatterplots of the predicted *vs.* the observed weighted score (*WS*) generated by ELM model in the testing phase (with all predictor variables: Q_1, Q_2, A_1, A_2, A_3 and the exam score) relative to the RF and Volterra models for *ENM1600 Engineering Mathematics* course. Least square linear regression line with the coefficient of determination (R^2) and sum of square errors (*SSE*) is also included in each sub-panel.

influence of two Assignments in modelling *WS*, similar trends were evident where A_1 was the better predictor for ONC, but A_2 was the better predictor for ONL ($LM \approx 0.197$ *vs*. 0.141 for ONC & 0.134 vs. 0.147 for ONL).

In spite of the differences for different offers of the same course, for example, ONC & ONL, for both versions of the course the final examination remained the best predictor of weighted score, although the accuracy of ELM was superior for ENM2600 ONC relative to ENM2600 ONL. This concurred with the results obtained for ENM1600, where the predictive model for online offer of the course was more accurate. Importantly, the ELM model's accuracy in predicting *WS* (with examination results as a predictor variable) was greater for the lower level ENM1600 ($5.915\% \leq RRMSE \leq 8.910\%$) than for the higher level course ENM2600 ($8.72 \leq RRMSE \leq 10.11\%$) when examination scores are incorporated into the model for prediction of the weighted score. Clearly, this ascertains a fundamental difference between the two level of courses in terms of the artificial intelligence models' ability to predict a final grade.

For each case, the influence of multivariate predictors incorporated in the optimal ELM model is shown using test phase accuracy statistics (Tables 6 and 7). Equivalent statistics for the RF and Volterra model are also presented. The order of addition of model inputs was based on lowest to highest cross-correlation (and covariance) between the respective predictors and target variable. This approach was implemented to enable the study any improvements in the model's ability to generate *WS*, as the interactive influence of each predictor, from the weakest to the strongest, was considered.

With successive addition of internal assessment input data, the ELM model continued to improve for both engineering mathematics courses and for both offerings. However, the role of continuous internal assessments on the prediction of *WS* remained much greater for ENM1600 than ENM2600, *i.e.*, with Assignments and Quizzes as predictors (excluding the final examination). The ELM models developed for the on-campus and online versions of ENM1600 attained *LM* values of 0.295 and 0.397, respectively, whereas the same model attained *LM* values of 0.206 and 0.197, respectively, for ENM2600. This result was produced when the final examination mark was excluded as a predictor variable, generating an *RRMSE* of about 19.51% and 14.59% for ENM1600 ONC & ONL, respectively compared to 22.88% and 20.68% for ENM2600 ONC & ONL, respectively. This indicated that the final examination mark was expected to play a

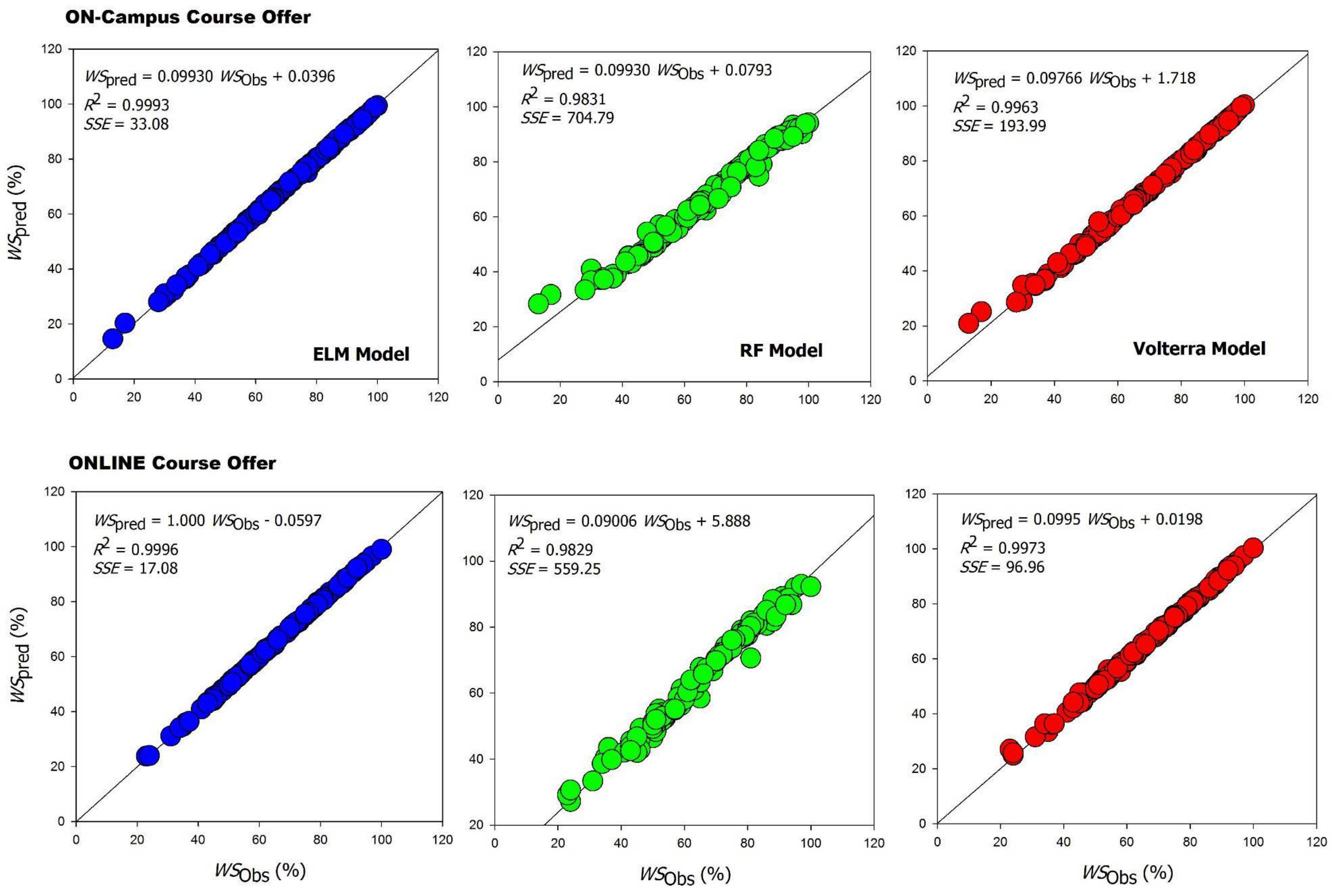


FIGURE 4. **Scatterplots of the predicted *vs.* the observed weighted score (*WS*) generated by ELM model in the testing phase (with all predictor variables: Q_1, Q_2, A_1, A_2, and the exam score) relative to the RF and Volterra models for *ENM2600 Advanced Engineering Mathematics* course.. Least square linear regression line with the coefficient of determination (R^2) and sum of square errors (*SSE*) is also included in each sub-panel.**

more significant role in modelling the *WS* for the higher-level (*i.e.*, ENM2600) compared to the lower level course (*i.e.*, ENM1600).

Similar deductions were corroborated when other performance measures (e.g., r) are checked, indicating the challenge faced by ELM in predicting student performance in advanced engineering mathematics data. Additionally, it is evident that for both courses investigated (including on-campus and online offers), the quality of performance of multivariate-based ELM model far exceeds that of either the RF or the Volterra model.

The scatterplots produced in the testing phase that represented the predicted versus the observed *WS* (*i.e.*, WS_{pred} *vs.* WS_{Obs}), including the sum of square errors (*SSE*), the coefficient of determination (R^2, a measure of estimated covariance) and a 1:1 line of best fit, were developed for all three predictive models. These drew upon all possible predictor variables, *i.e.*, Q_1, Q_2, A_1, A_2, A_3 including the examination score (for ENM1600; Figure 3), and Q_1, Q_2, A_1, A_2 and the examination score (for ENM2600; Figure 4). For both engineering mathematics courses and also for both modes of offer, the ELM and Volterra models showed the greatest accuracy in predicting *WS* (Tables 6–7). However, the ELM model showed lower *SSE* values for the ENM1600 ONC and ENM1600 ONL courses (33.33% and 35.57%, respectively), than did the Volterra model (63.16% and 74.89%, respectively). Similarly, for the ENM2600 ONC and ENM2600 ONL courses the *SSE* values for ELM (33.08% and 17.08%, respectively) showed better performance than for the Volterra model (193.99% and 96.96%, respectively). For both engineering mathematics courses and modes of the course offer, RF model showed a much greater degree of scatter between WS_{pred} and WS_{Obs} data in the testing phase that did the ELM or the Volterra model.

To explore the role of continuous assessments and the manner in which they lead to a *WS*, examination marks (*EX*) was excluded from predictor variables, thus modelling only the influence of Quizzes and Assignments on *WS* (Figures 5 and 6). Very interesting features emerged, especially between two levels of engineering mathematics courses. The extent of scattering between WS_{pred} and WS_{Obs} was much greater for ENM2600 than ENM1600, as was confirmed by the larger *SSE* and lower R^2 value for the latter course. All three predictive models failed to provide a

TABLE 6. Influence of multivariate predictor variables (*i.e.* internal assessment) adopted to predict weighted scores (*WS*) as the target variable based on ELM and the equivalent comparative counterpart models for *ENM1600 Engineering Mathematics* in the testing phase. The optimal model is red/boldfaced.

ENM1600 ONC

Model Designation	Multivariate Input Combinations	*r*	*RMSE*	*MAE*	*RRMSE*, %	*RMAE*, %	*d*	*NS*	*LM*
$M1_{ELM}$	Q_1	0.543	15.76	12.86	23.60	25.00	0.986	0.254	0.138
$M2_{ELM}$	$Q_1 + A_3$	0.697	14.15	11.41	21.18	21.89	0.989	0.399	0.235
$M3_{ELM}$	$Q_1 + A_3 + A_2$	0.749	13.29	10.70	19.89	18.12	0.990	0.470	0.283
$M4_{ELM}$	$Q_1 + A_3 + A_2 + Q_2$	0.730	13.69	11.16	20.49	18.84	0.989	0.437	0.252
$M5_{ELM}$	$Q_1 + A_3 + A_2 + Q_2 + A_1$	0.753	13.03	10.52	19.51	16.48	0.990	0.490	0.295
$M6_{ELM}$	**$Q_1 + A_3 + A_2 + Q_2 + A_1 + EX$**	**1.000**	**0.50**	**0.32**	**0.74**	**0.64**	**1.000**	**0.999**	**0.979**
$M6_{RF}$	$Q_1 + A_3 + A_2 + Q_2 + A_1 + EX$ [for optimal RF]	0.994	2.25	1.53	3.12	2.50	1.000	0.978	0.875
$M6_{Volterra}$	$Q_1 + A_3 + A_2 + Q_2 + A_1 + EX$ [for optimal Volterra]	0.999	0.71	0.40	1.06	1.24	1.000	0.999	0.973

ENM1600 ONL

Model Designation	Multivariate Input Combinations	*r*	*RMSE*	*MAE*	*RRMSE*, %	*RMAE*, %	*d*	*NS*	*LM*
$M1_{ELM}$	A_2	0.621	13.67	11.41	18.68	17.05	0.991	0.337	0.184
$M2_{ELM}$	$A_2 + A_1$	0.697	12.48	10.78	17.05	15.98	0.993	0.447	0.230
$M3_{ELM}$	$A_2 + A_1 + Q_1$	0.759	11.40	9.77	15.58	14.52	0.994	0.539	0.301
$M4_{ELM}$	$A_2 + A_1 + Q_1 + A_3$	0.800	10.53	8.89	14.38	13.00	0.995	0.607	0.365
$M5_{ELM}$	$A_2 + A_1 + Q_1 + A3 + Q_2$	0.799	10.68	9.00	14.59	12.97	0.995	0.595	0.357
$M6_{ELM}$	**$A_2 + A_1 + Q_1 + A_3 + Q_2 + EX$**	**1.000**	**0.38**	**0.32**	**0.51**	**0.46**	**1.000**	**0.999**	**0.977**
$M6_{RF}$	$A_2 + A_1 + Q_1 + A_3 + Q_2 + EX$ [for optimal RF]	0.994	2.23	1.52	3.05	2.36	1.000	0.982	0.891
$M6_{Volterra}$	$A_2 + A_1 + Q_1 + A_3 + Q_2 + EX$ [for optimal Volterra]	1.000	0.51	0.35	0.70	0.57	1.000	0.999	0.975

TABLE 7. Influence of multivariate predictor variables (*i.e.* internal assessment) adopted to predict weighted scores (*WS*) as the target variable based on ELM and the equivalent comparative counterpart models for *ENM2600 Advanced Engineering Mathematics* in the testing phase.

ENM2600 ONC

Model Designation	Multivariate Input Combinations	*r*	*RMSE*	*MAE*	*RRMSE*, %	*RMAE*, %	*d*	*NS*	*LM*
$M1_{ELM}$	Q_2	0.399	18.45	15.57	27.79	26.86	0.980	0.062	0.013
$M2_{ELM}$	$Q_2 + Q_1$	0.397	18.38	15.50	27.69	27.19	0.980	0.069	0.017
$M3_{ELM}$	$Q_2 + Q_1 + A_2$	0.705	16.13	13.56	24.29	20.91	0.984	0.283	0.140
$M4_{ELM}$	$Q_2 + Q_1 + A_2 + A_1$	0.685	15.19	12.53	22.88	19.43	0.986	0.364	0.206
$M5_{ELM}$	**$Q_2 + Q_1 + A_2 + A_1 + EX$**	**1.000**	**0.51**	**0.35**	**0.77**	**0.73**	**1.000**	**0.999**	**0.978**
$M6_{RF}$	$Q_2 + Q_1 + A_2 + A_1 + EX$ [for RF]	0.312	9.58	6.42	22.23	38.81	0.988	0.062	0.140
$M6_{Volterra}$	$Q_2 + Q_1 + A_2 + A_1 + EX$ [for Volterra]	0.998	1.24	0.62	1.87	1.71	1.000	0.996	0.961

ENM2600 ONL

Model Designation	Multivariate Input Combinations	*r*	*RMSE*	*MAE*	*RRMSE*, %	*RMAE*, %	*d*	*NS*	*LM*
$M1_{ELM}$	Q_2	0.426	15.53	12.95	24.11	21.67	0.985	0.132	0.066
$M2_{ELM}$	$Q_2 + Q_1$	0.490	15.10	12.71	23.44	20.91	0.986	0.180	0.083
$M3_{ELM}$	$Q_2 + Q_1 + A_2$	0.625	13.71	11.56	21.28	18.65	0.988	0.324	0.166
$M4_{ELM}$	$Q_2 + Q_1 + A_2 + A_1$	0.634	13.32	11.14	20.68	18.06	0.989	0.362	0.197
$M5_{ELM}$	**$Q_2 + Q_1 + A_2 + A_1 + EX$**	**1.000**	**0.35**	**0.28**	**0.54**	**0.48**	**1.000**	**1.000**	**0.979**
$M6_{RF}$	$Q_2 + Q_1 + A_2 + A_1 + EX$ [for RF]	0.992	2.63	1.80	4.08	3.23	1.000	0.975	0.870
$M6_{Volterra}$	$Q_2 + Q_1 + A_2 + A_1 + EX$ [for Volterra]	0.999	0.84	0.62	1.31	1.20	1.000	0.997	0.955

reasonable prediction of the *WS* for ENM2600 when considering only the continuous internal assessment data. This also indicates that the examination score is likely to influence the student outcome in a much greater degree in the advanced engineering mathematics course and to a lesser degree, for the mid-level engineering mathematics course (*i.e.*, ENM1600).

The plots of *RMSE* for the ELM model (Figure 7a–7b) illustrate the ELM's capability to generated two kinds of

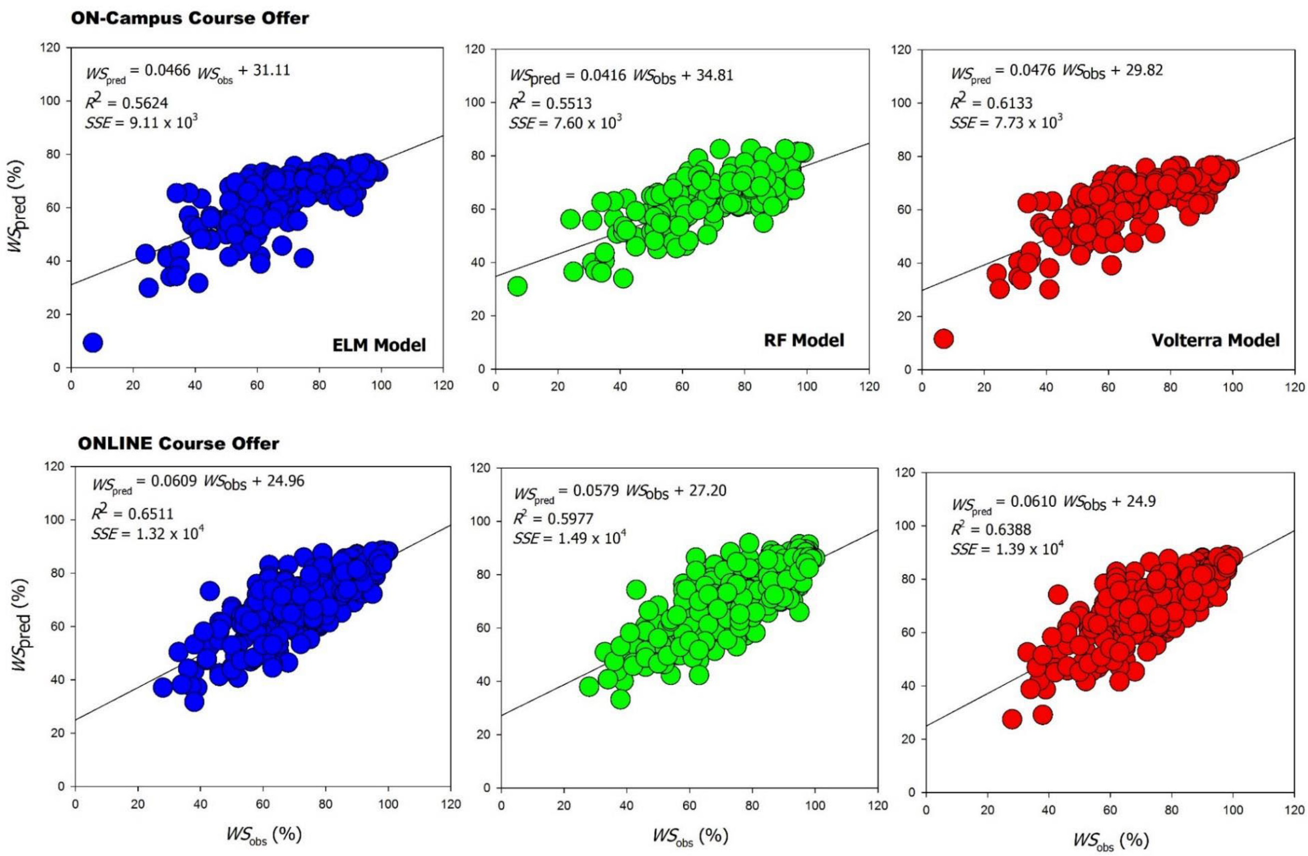


FIGURE 5. Prediction of weighted score (*WS*) based on continuous internal assessment (*i.e.* Q_1, Q_2, A_1, A_2 and A_3) as predictor variables for *ENM1600 Engineering Mathematics* course in the testing phase. Exam score has been excluded from predictor variables to check the influence of only continuous internal assessments on predicted *WS*.

model results: (*i*) prediction of the *WS* (as a target) and *EX* (as the single predictor variable) and (*ii*) prediction of *EX* (as a target) and continuous internal assessment data (as a suite of predictor variables). For both types of predictive model scenarios and the relevant course levels in engineering mathematics that were modelled, based on its lower RMSE the online version of the course results was more accurately predicted than those of the on-campus versions of the course. When the *WS* was modelled using the *EX*, the relative performance for both the ONC and ONL versions of ENM1600 was dramatically better than for ENM2600 (Figure 7a), concurring with earlier results (*e.g*., Figure 5-6; Table 6-7). A similar deduction could be made from Figure 7(b) where the *WS* was modelled from continuous internal assessment data (excluding the *EX*). However, for both courses and modes of offer, the ELM model showed significantly greater relative percentage errors when the *EX* was excluded from the matrix of predictor variables for both courses and all modes of offer.

A very important aspect of the predictive model evaluation process is to check the distribution of errors encountered in the testing phase. Figure 8 represents a boxplot with distribution of absolute value of prediction error generated in modelling *WS* where both the continuous internal assessments and examination scores for each course and the relevant modes of offer are incorporated as the predictor variables. There appears to be an undisputed quantitative evidence that the ELM model's performance accuracy far exceeds that of the RF in all modelling scenarios given the widely spread errors and large outliers indicating large error magnitudes. However, differences between ELM and Volterra models are less conspicuous, although the outliers in boxplots clearly depict extreme errors for Volterra that are encountered in prediction of student's *WS* values.

Since the edge of the boxplot denotes the upper and the lower quartile errors, and the central margin shows the median value of the error, Figure 8 reaffirms the relative success of the ELM *vs*. its two counterparts' models, as the former's quartiles and medians are significantly smaller. Consistent with the results presented earlier (*i.e*., Figures 3–7; Tables 6–7), the distribution of errors showed the ELM model to have a lower accuracy when predicting the *WS* for ENM2600 than for ENM1600; however, the trends for ONC and ONL versions of the course were quite different. For the ONC versions of the courses, the ENM1600 outliers showed a relatively smaller error value than those for ENM2600, confirming that model performance for student success predictions for the online version of the course

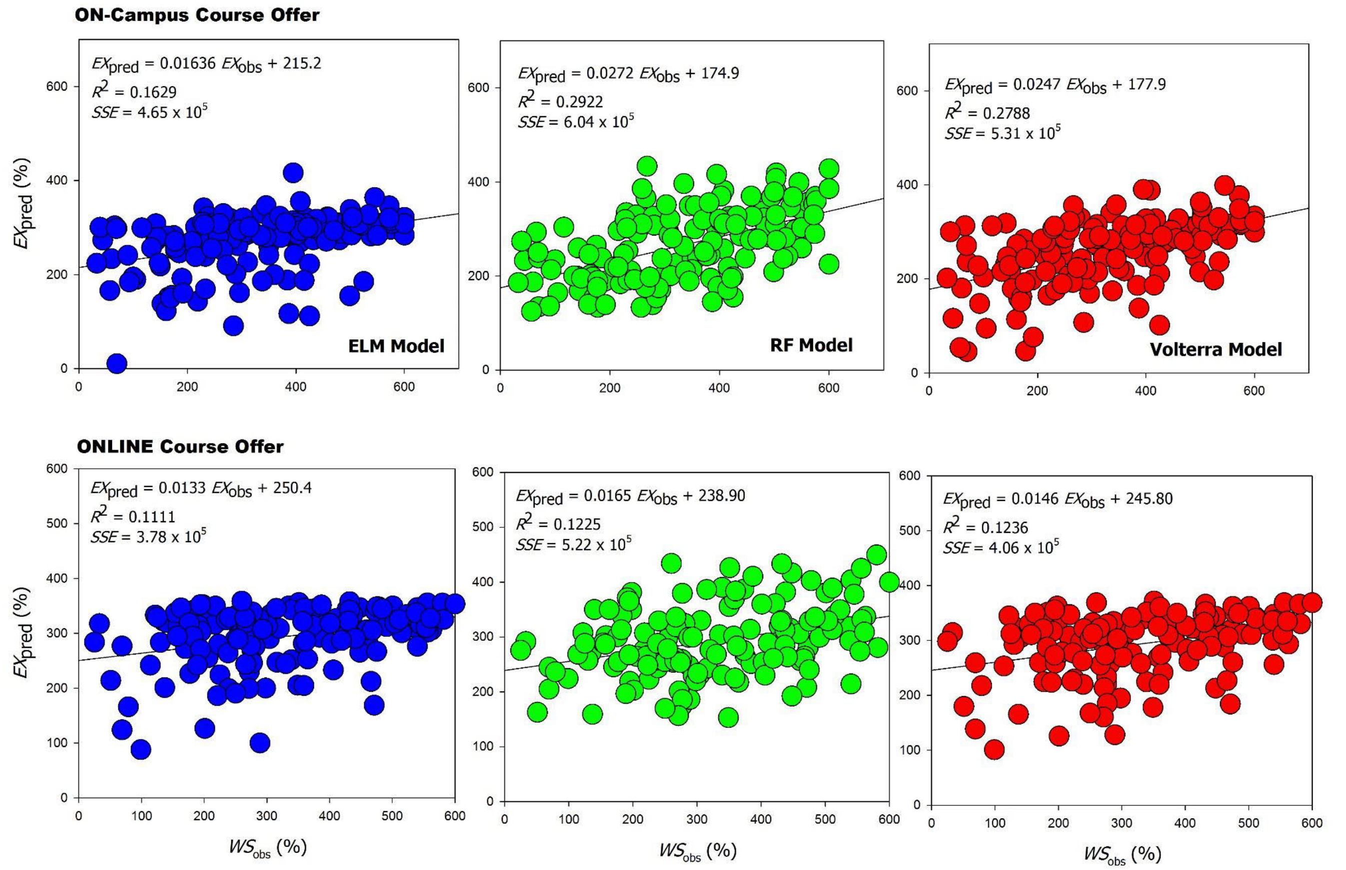


FIGURE 6. Prediction of weighted score (*WS*) based on continuous internal assessment (*i.e.* Q_1, Q_2, A_1, A_2) as predictor variables for *ENM2600 Advanced Engineering Mathematics* course in the testing phase. Exam score has been <u>*excluded*</u> from predictor variables to check the influence of only continuous internal assessments on predicted *WS*.

was more feasible for ENM2600 than ENM1600, at least with the current data and under the present the modelling context.

The mean prediction error of ELM, relative to the RF and the Volterra model, computed over the entire range of observed *WS* data used in the respective grade allocation processes are shown in Figure 9. Note that the allocation of grades follows the threshold: *Fail* (*F*), $WS < 50\%$; *C*, $50 \leq WS < 64$; *B*, $65 \leq WS < 74$; *A*, $75 \leq WS < 84$, *High Distinction*, $WS \geq 85$, so for each modelling scenario, the testing phase data were extracted within these categories and further analysed. In spite of somewhat mixed results for certain grade categories, the ELM models were generally more successful in matching letter grades than either the RF or Volterra models. In particular, the ELM model registers a much smaller error in predicting a *Fail* grade for ENM1600 ONC and ONL, and ENM2600 ONL test data. For ENM1600 ONL the error for the ELM was approximately 0.695% compared to 0.980% for the Volterra and 6.406% for the RF model, while for the ENM1600 ONC, the errors were 0.321% compared to 0.664% and 3.563%, and for ENM2600 ONL they were 0.297%, 1.159%, and 2.942%. For all other grades spanning C to HD, the ELM performance exceeded the comparative models for both engineering mathematics courses and modes of offer.

Of particular interest to this study was how the *WS* in the model's testing phase (considering all possible predictor variables) were distributed for the engineering mathematics course and different course offering modes when predicted results were allocated in their respective grading categories. A plot of observed *vs.* predicted grades generated by all three models using ENM1600 as an example (Figure 9) shows unambiguous evidence that the ELM model's predicted grade distribution more closely matches the real (observed) grade distribution than do the grade distributions predicted by the two benchmark models. This proved to be true for all five categories of allocated grades. While the difference between the ELM and Volterra model appear to be somewhat marginal in some cases, there is a clear distinction with respect to the RF model, which fails to generate a reasonable degree of accuracy.

The percentage frequency of predicted errors in *WS*, for the ENM2600 course pooled across the on-campus and online course offers (Figure 11), shows the greater efficacy of the ELM model compared to the other two models, where, for the ELM about 99% of all predictive errors are located

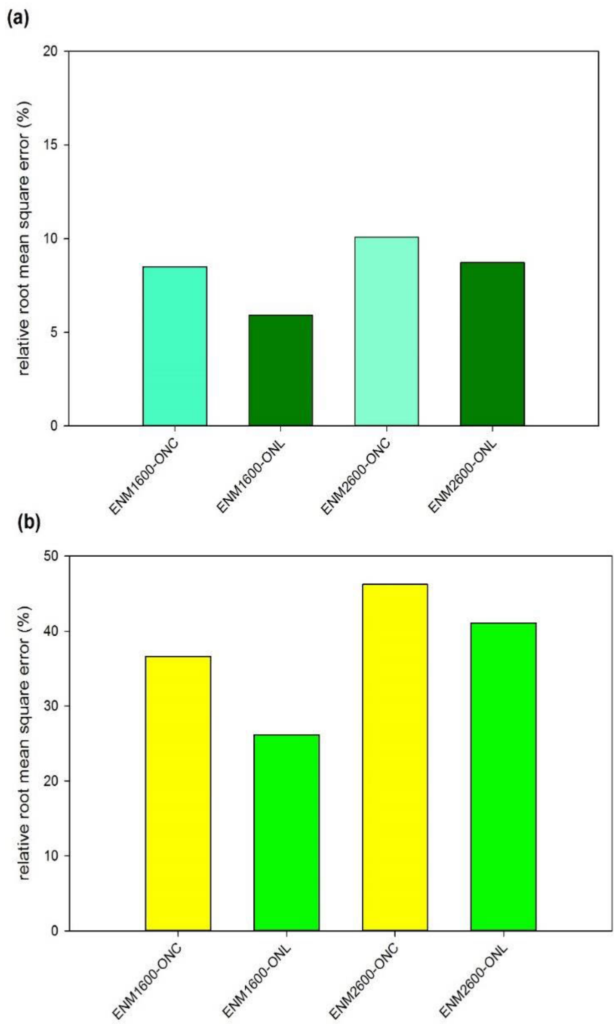


FIGURE 7. **The relative (%) root mean square error generated by the ELM model for ENM1600 and ENM2600 for on-campus (ONC) and online (ONL) offers. (a) Prediction of weighted score (*WS*) using the exam (*EX*) as a predictor variable. (b) Prediction of exam score (*EX*) using continuous internal assessments as predictor variables.**

in ± 1% range, compared to only 43% and 88% of all errors for the RF and Volterra models, respectively. Given the smaller proportion of errors in ±1% magnitude range for the RF and Volterra models compared to the ELM, the frequency of the former models' errors was distributed to a greater extent in ranges of greater error magnitude. For example, for the RF and Volterra models approximately 27% and 9%, respectively, of all errors were distributed in ±(1–2)% bracket, whereas for the ELM this bracket only included 1% of all errors. In fact, the frequency of errors exceeding ±2% for the ELM model was zero, whereas for the RF and Volterra models 27% and 4% were in exceedance, respectively).

For real-life and practical implementation of predictive models that enable important decisions in course management to be undertaken by examiners, a model's accuracy in predicting individual grades is critical and should be of primary interest. Figure 12 displays the modelled results for ENM2600 where on-campus and online datasets have been pooled, and the respective frequency of prediction errors in different error brackets are plotted across the 5 categories of grades.

For the entire category of grades, the ELM model was seen to outperform both the RF and Volterra models, especially for 'C' or '*F*' grades. That is, in modelling an '*F*' grade, approximately 78% of predictive errors were within a ±0.5% error margin for the ELM model compared to only 14% and 27% for RF and Volterra, respectively. This led to a redistribution of error into larger error brackets to which the ELM model contributed only 18%, whereas RF and Volterra models contributed 22% and 35%, respectively. Likewise, for modelling a '*C*' grade in ENM2600, approximately 93% of all errors were recorded within ±1% by the ELM model compared to only 25% and 59% by the RF and Volterra models, respectively, resulting in a redistribution of errors into the larger error brackets. While the capacity of the ELM model to predict '*HD*', '*A*' and '*B*' grades appeared to match that of the Volterra model, there were differences that indicate the ELM could be the preferred model for prediction of individual grades. Similar results were obtained for the case of ELM1600 (not shown here).

In accordance with the results for engineering mathematics performances modelled in this study, clear differences in model accuracy between ENM1600 and ENM2600 were evident. These are perhaps attributable to the nature of the courses rather than an influence of the model itself. In particular, ENM1600 provides a solid foundation in single variable Calculus, Matrix and Vector Algebra and is the prerequisite course for ENM2600, whereas Advanced Engineering Mathematics includes topics in Complex Numbers, Multivariable Calculus, Differential Equations and Eigenvalues and Eigenvectors [56], [57]. These topics, other than the Complex Numbers topic, are likely to require a firm understanding and proficiency of the topics in ENM1600. The lack of a firm grasp of these basic topics could possibly lead students to struggle in ENM2600, and would likely have a particular impact on students who only take ENM2600 as part of the Master of Engineering program without the prerequisite knowledge and skills from ENM1600.

An additional issue relevant to the present discussion may be that differences in course content and learning evaluation, e.g., the different structure of ENM1600 and ENM2600 quizzes, could explain differences in model performances. For example, in ENM1600 the Semester 1 quiz questions are randomly chosen whilst in ENM2600 they are identical for all students. Randomness in quizzes would likely decrease the sharing of answers and leading to a more independent attempt by each student. This is more likely to affect the results of the on-campus cohort than the online cohort given that sharing is less likely in the latter cohort. That said, the questions are predominantly multiple choice and so there is still the possibility of randomly chosen answers. As such, the quizzes alone may not be reliable predictors of student's understanding, but it rather depends

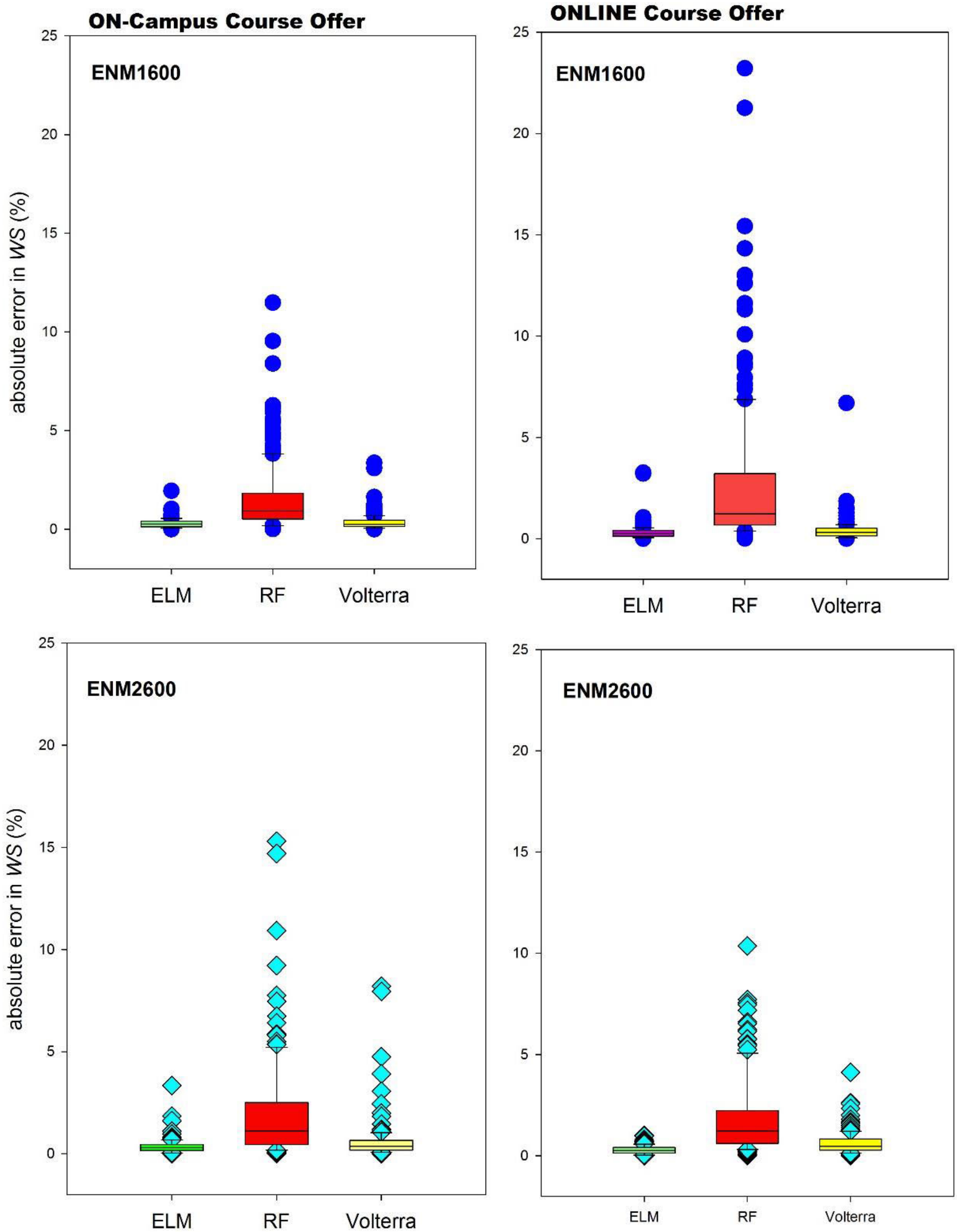


FIGURE 8. Boxplots showing the distribution of absolute prediction error in the weighted score (*WS*, %) generated by ELM relative to the RF and Volterra models in the testing phase. Here, the continuous internal assessments and exam score for each course/offer has been used the predictor variables.

upon how the options to the multiple-choice questions are structured. In ENM1600, the incorrect options are chosen based upon common errors made by students. This could explain some of the differences between the modelled results for ENM1600 and ENM2600 but a further study may be required to achieve more conclusive results.

V. DISCUSSION, LIMITATIONS AND OPPORTUNITY FOR FUTURE WORK

Supported by a diverse range of quantitative metrics and visual (predicted *vs*. observed) results, the greater accuracy of ELM (*vs*. RF and Volterra) models, was evident, although model performance was largely scaled according to the

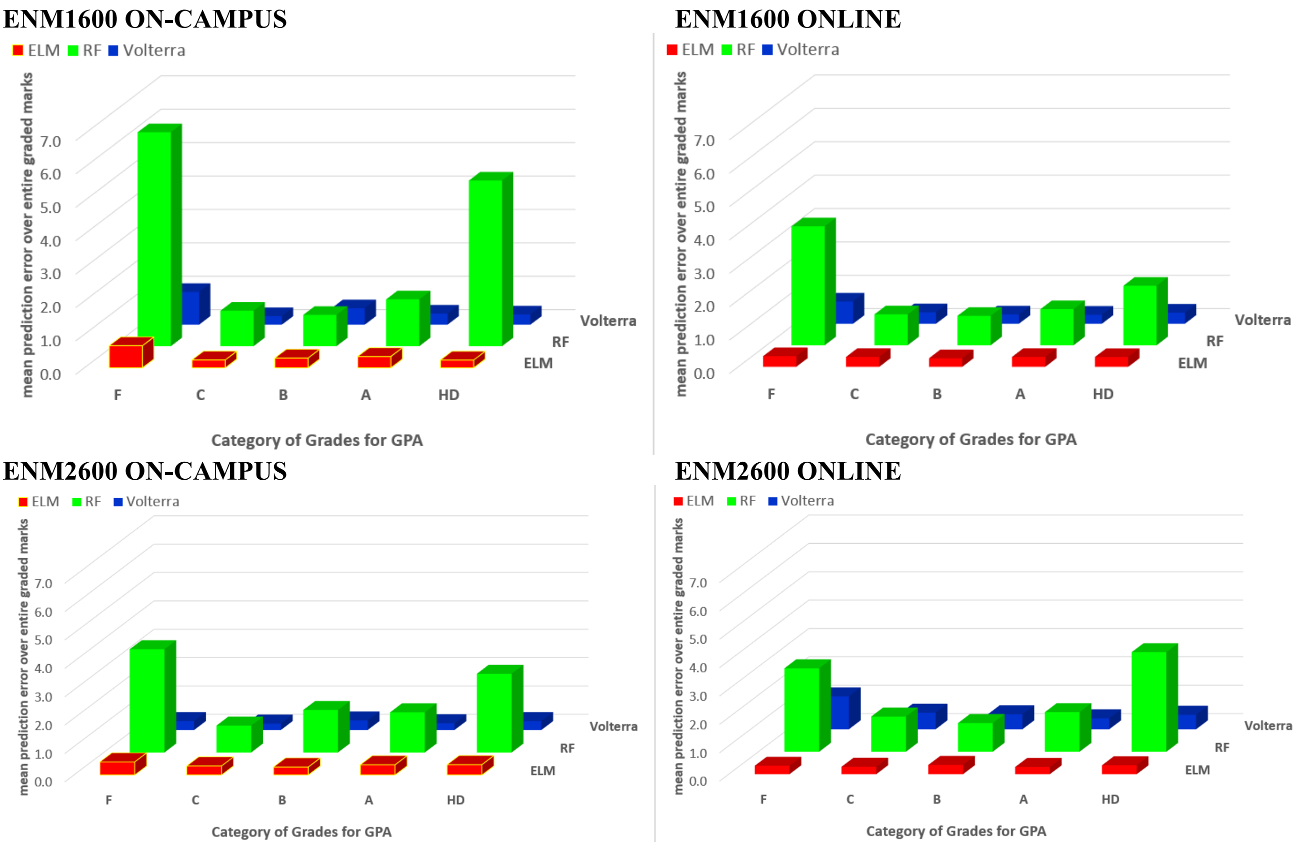


FIGURE 9. The mean prediction error (%) computed over entire range of observed weighted scores that are used in respective grade allocation process in testing phase of ELM (and its comparative counterpart) models.

specific engineering mathematics course level (i.e., whether it was ENM1600 or ENM2600) and whether the teaching mode was on-campus or online. Nonetheless, built on the success of the proposed methodology, which yielded an acceptable accuracy in the modelling of *WS* and *EX*, further exploration of the ELM-based model is encouraged to ascertain the viability of its practical adoption as a decision support tool in student performance prediction for the higher education sector. As a unique study employing ELM model for engineering mathematics performance evaluation, this investigation exemplifies the significant merits of the intelligent algorithm, such as a fast, accurate and efficient artificial intelligence platform where training, validation, and testing process are achieved in a relatively short model execution time compared to RF and Volterra models, without compromising the model's overall predictive accuracy. Practically, this is highly essential from the evaluators "lecturers" point of view to have a reliable technique that can assist the educational sector.

In spite of these merits, the ELM model does carry limitations, and therefore, this work sets a new pathway for follow-up research that could improve the model's versatility in predicting student performances. For example, in this study, the ELM was executed largely in a batch mode, and was designed with pre-defined training, validation, and testing data partitioned sequentially from 5-6 years of examiner returns. The model also had a validation set that enabled a selection of best trained models, while the testing set provided independent input data to simulate the *WS*. Taking this approach to further improve the methodology, one could enhance the present model's implementation by testing improved variants of this algorithm. One such algorithm is the online sequential ELM (OS-ELM) [58], [59]. With the added advantage of greater speed, especially with big datasets, the OS-ELM, based on recursive least-squares (RLS), can learn data one-by-one or a chunk-by-chunk (i.e., a block of data) approach, with a fixed or a varying chunk size [60]. This could be significantly advantageous compared to a baseline ELM, especially when a model needs to be implemented in a large, university-based course management system where new data arrives continuously (even if they are spaced with long temporal delays). In this respect, the OS-ELM model can offer a computationally less expensive modelling platform than the normal batch learning model in maintaining the model input up to date. Another variant of ELM, the variable complexity (VC-OSELM) algorithm can also be explored in a separate study, as that model can dynamically add or remove hidden neurons based on received data, allowing the model's structural complexity to evolve, and vary automatically as the online learning and modelling process proceeds.

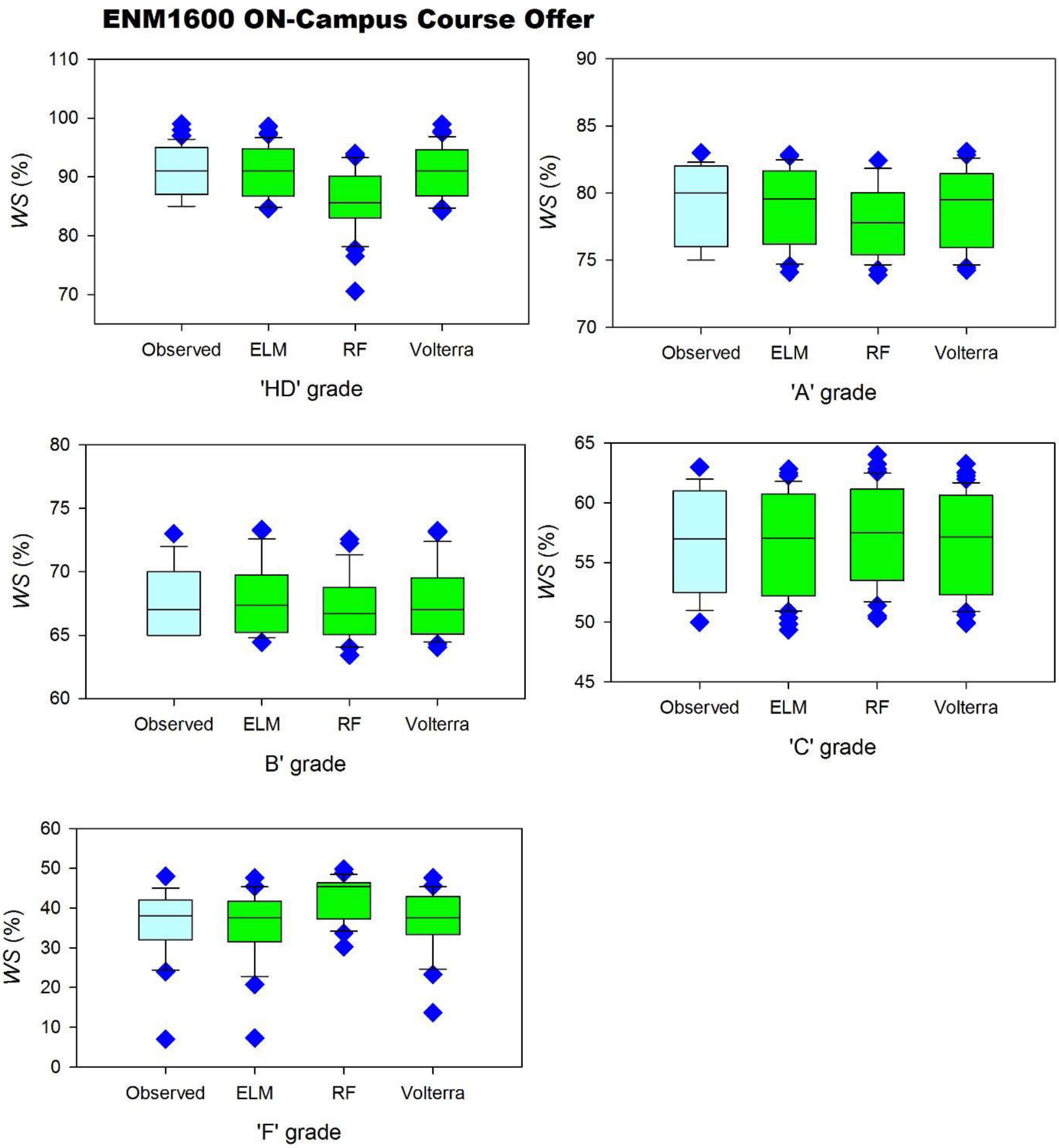


FIGURE 10. Boxplots exploring the ability of ELM model to predict weighted scores (*WS*) in different observed categories of allocated grades in the testing phase. Note that the official grade allocations typically follow: HD = High Distinction (≥ 85%), A (75% ≤ *WS* < 85%), B (65% ≤ *WS* < 74%), C (50% ≤ *WS* < 65%), F (<50%).

In this study, the most recent and relatively lengthy student performance records from two levels of engineering mathematics courses (ENM1600 2013–2018; ENM2600 2014–2018), given under both online and on-campus modes of offer, were incorporated to design an ELM model. While continuous internal assessment data (described by quizzes and assignments modelled in respect to possible *WS* and grade) are likely to provide ongoing evaluation of student learning and how the teaching approach may influence a student's grade, several other variables may also influence the learning process. Such variables must be investigated in a more rigorous, follow-up modelling study. For example, data derived from informal study desk activities (e.g., viewing of video snippets by students, spending sufficient time on study desk and regularly engaging with recorded lectures and tutorials posted online) could serve as ELM model inputs. Such a model could help assess the influence of regularly monitoring student participation patterns and the manner in which such external predictor variables affect a student's learning journey, leading to a successful grade.

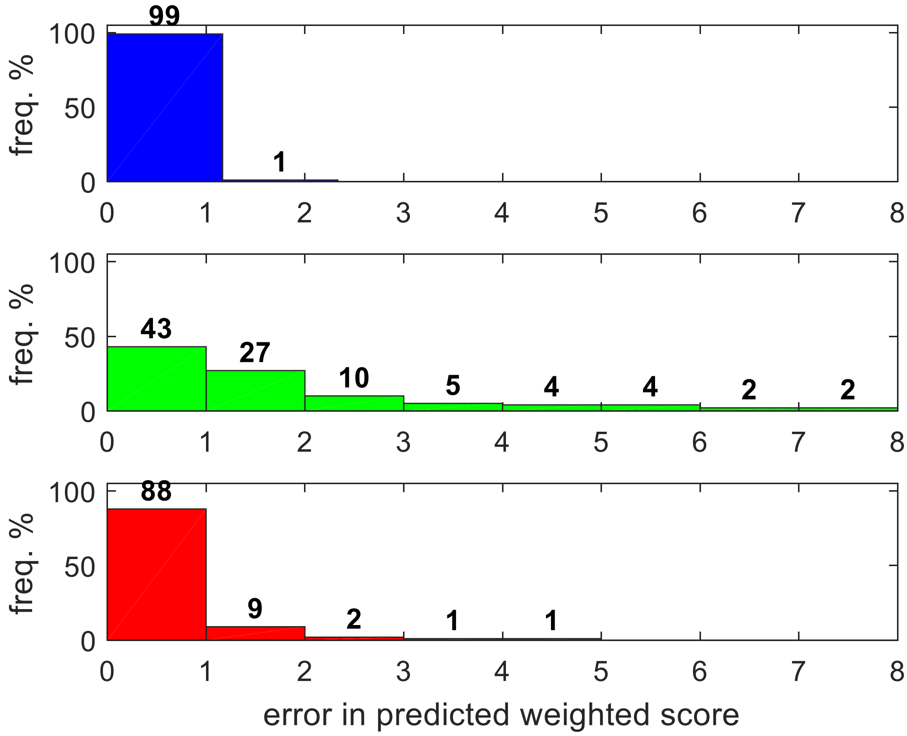


FIGURE 11. **Percentage frequency of predicted error in weighted score (*WS*) for *ENM2600 Advanced Engineering Mathematics* with all tested data pooled for on-campus and online course offers. Each bracket spanning a predictive error of ± 1% shows the respective occurrence frequency in the tested data.**

In particular, there is strong consensus in the existing literature that under-preparedness in mathematical content, particularly at pre-tertiary experience, has a significant influence on students' abilities to make a successful transition to a tertiary level mathematics course [61]. Furthermore, the amount of time spent on the Learning Management System (LMS) appears to also be correlated with learning outcomes [62]. For courses that are delivered in 'online only mode' with no face-to-face or synchronous lectures and tutorials, the incorporation of time spent on study desk as a possible predictor for student performance modelling is very important. To improve the existing approach, one could also consider a revised ELM algorithm by incorporating such potentially influential data within a global course management and a decision-support system. Such a versatile model based on these additional datasets could help generate an effective guide for course instructors in identifying their student's learning needs and in scaffolding the entire learning process [63]–[65]. While these are interesting insights to improve the existing ELM model, they were beyond the scope of the present investigation and therefore, must await an independent research study.

While a number of continuous internal assessments data were considered to model student performances, this study has not considered the influence of other external and inter-related factors. Some of these include the student's gender, age (i.e., whether mature aged), marital and school leaver status, socio-economic status, and the proper prerequisite knowledge to learn university mathematics, when modelling the *WS* and the grade. There is significant indication that these factors are related to student participation, access, retention, and overall success [66], [67]. Recent studies are showing relevance of such causal factors with respect to a successful attainment of knowledge and grade at the university level [68].

Contini *et al.*, considered gender differences in STEM discipline to investigate gaps in mathematics scores, showing that girls systematically underperformed boys [69]. Insights can be drawn from Devlin and McKay where academic success for students from low socioeconomic background at regional universities was considered, showing that such students lack confidence and self-esteem [70]. This can affect their overall sense of belongingness to higher education sector. Bonneville-Roussy *et al.*, investigated the effect of gender differences using two sets of multi-groups for prospective studies, particularly studying motivation and coping skills with stresses of assessment [71]. Importantly, their results showed that strong gender differences exist between coping and academic outcomes for university students.

Based on exploratory studies that indicate the important role played by many exogenous variables on the overall learning journey of students in higher education, future research could consider an improved ELM model where data such as the gender, age, socio-economic status and the pre-requisite knowledge are also incorporated to identify and develop

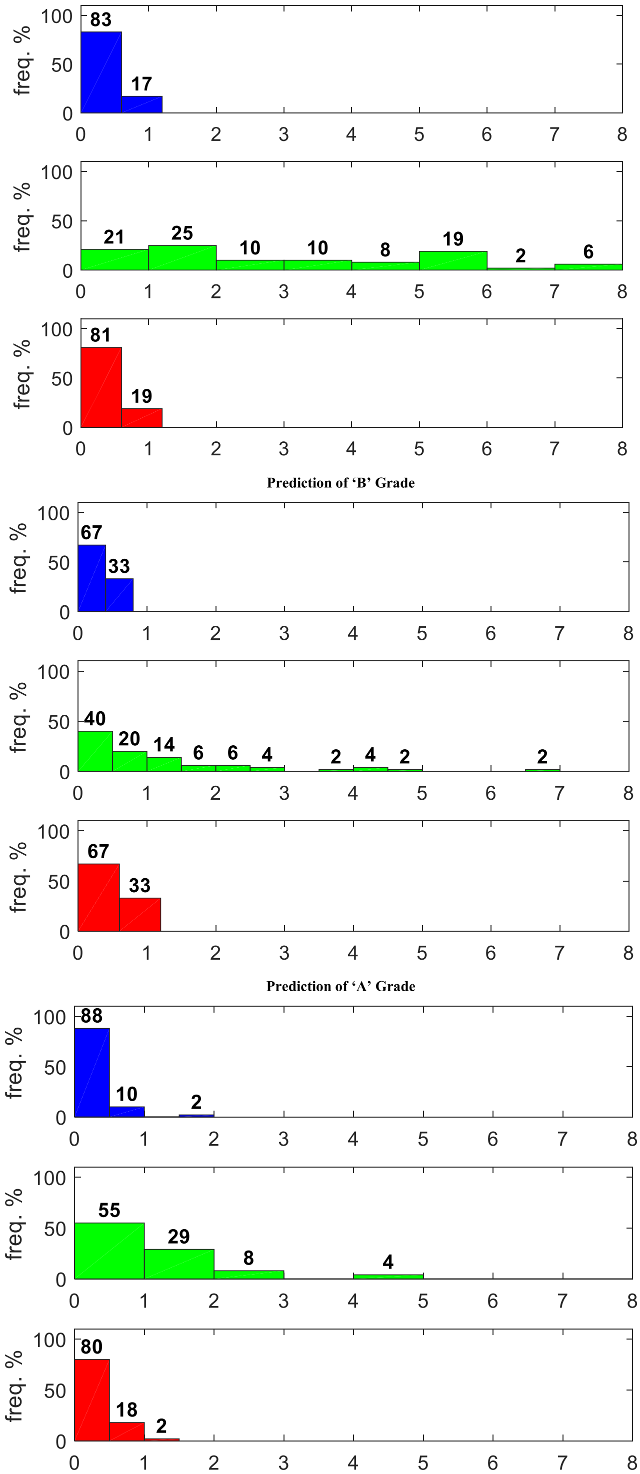


FIGURE 12. Exploring the capability of the ELM model in predicting different grades in *ENM2600 Advanced Engineering Mathematics*. Here, all tested data have been pooled for the on-campus and online offers of this course to determine the overall ability of ELM against RF and Volterra models.

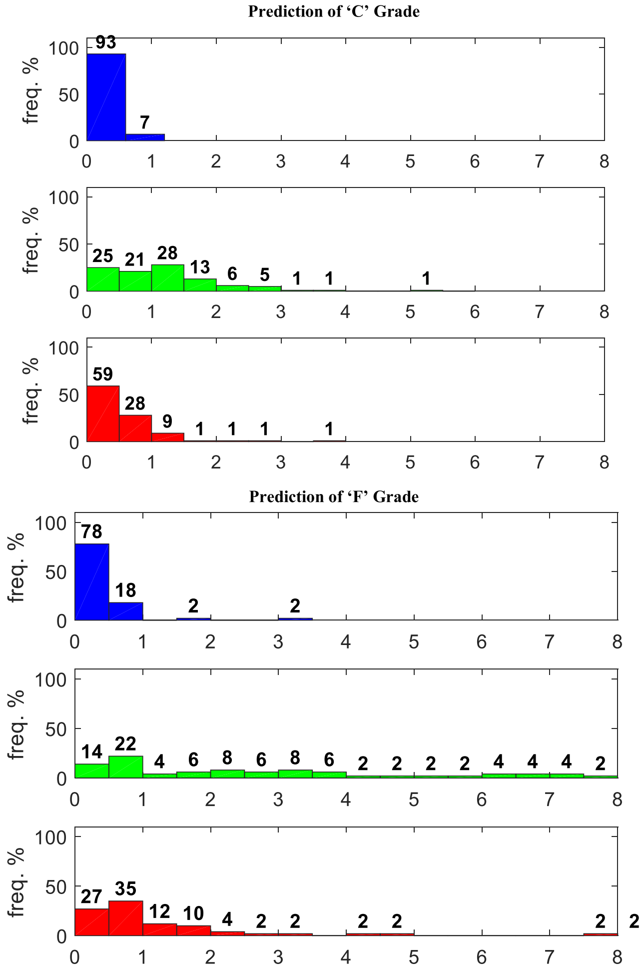


FIGURE 12. *(Contiued.)* Exploring the capability of the ELM model in predicting different grades in *ENM2600 Advanced Engineering Mathematics*. Here, all tested data have been pooled for the on-campus and online offers of this course to determine the overall ability of ELM against RF and Volterra models.

several modelling scenarios particularly in 1st year courses. These scenarios must consider the influence of these data as predictors for student's *WS* and *EX*. While data such as socio-economic status could be challenging to accumulate and even to properly authenticate their relevance to student's performance predictions, they nevertheless may enrich the conclusions drawn from this research study. More importantly, such exogenous data can also help performance educational performance modelers to segregate, identify and incorporate important influences in modelling their student's grades based on a much diverse range of predictors for student success for on-campus and online modes of courses in engineering mathematics, or other subject areas.

VI. SYNTHESIS AND CONCLUSIONS

Modern educational decision support systems adopted for student performance prediction and academic program quality assurance implementation by means of practical, responsive, user- and learner-friendly course management platforms can promote a successful student learning journey including student retention and progression, by embracing evidence-based models, preferably through a scholarship driven approach.

TABLE 8. **(Case 1: ENM1600 ONC). ELM Model Design Parameters in terms of the optimal input weights, optimal output weights and the bias value of hidden neurons.**

	Input Weights for Each Input Variable						Output Weights of Target Variable	Bias of Hidden Neurons
Neurons	*Q1,*	*A3,*	*A2,*	*Q2,*	*A1,*	*EX*	*WS*	*WS*
1	0.992	-0.269	0.968	-0.003	0.887	-0.560	-0.143	0.209
2	0.318	-0.452	-0.544	-0.343	0.810	0.223	-0.027	0.595
3	-0.608	-0.766	0.179	-0.264	-0.113	0.695	0.141	0.624
4	-0.804	-0.769	0.175	0.608	-0.806	0.890	-0.077	0.668
5	0.886	0.905	0.935	-0.235	-0.586	-0.420	0.178	0.173
6	0.890	0.617	0.315	0.540	-0.457	0.454	0.542	0.899
7	0.243	-0.670	0.170	-0.119	-0.032	-0.970	-1.786	0.621
8	-0.966	-0.586	0.038	0.688	-0.323	0.758	0.062	0.044
9	-0.549	0.311	0.529	-0.848	0.548	-0.872	-0.990	0.684
10	0.603	0.529	-0.788	-0.038	-0.048	0.467	0.630	0.196
11	0.751	0.621	-0.996	-0.066	0.741	0.989	-0.306	0.027
12	-0.092	-0.673	0.905	-0.471	0.992	0.002	1.632	0.551

TABLE 9. **Case 2: ENM1600 ONL.**

	Optimal Input Weights For Each Input Variable						Output Weights of Target Variable	Bias of Hidden Neurons
Neurons	A2,	A1,	Q1,	A3,	Q2,	EX	*WS*	*WS*
1	0.992	-0.092	0.621	-0.788	-0.848	-0.323	-0.637	0.015
2	0.318	-0.269	-0.673	-0.996	-0.038	0.548	-0.079	0.879
3	-0.608	-0.452	0.968	0.905	-0.066	-0.048	-2.468	0.064
4	-0.804	-0.766	-0.544	-0.003	-0.471	0.741	-2.414	0.733
5	0.886	-0.769	0.179	-0.343	0.887	0.992	1.865	0.995
6	0.890	0.905	0.175	-0.264	0.810	-0.560	-0.676	0.501
7	0.243	0.617	0.935	0.608	-0.113	0.223	4.353	0.209
8	-0.966	-0.670	0.315	-0.235	-0.806	0.695	-0.231	0.595
9	-0.549	-0.586	0.170	0.540	-0.586	0.890	3.259	0.624
10	0.603	0.311	0.038	-0.119	-0.457	-0.420	-0.540	0.668
11	0.751	0.529	0.529	0.688	-0.032	0.454	-3.446	0.173

Learning analytics, a rapidly evolving field in the higher education sector, can be used to design student performance management and modelling systems, and these types of decision systems can be supported by emerging digitalized technologies coupled with big data techniques. These tools can employ historical evidence of student performance based on their attainments in key learning tasks, to help examiners in exploring possible drivers of, or hindrance to, student success in a course. Such tools can also help determine possible causes of attrition and learning challenges faced by students in a teaching semester, and how the continuous internal assessments and other learning activities may influence a student's overall satisfaction in a course or program of study.

This research, for the first time, employed artificial intelligence models: Extreme Learning Machine and random forest, together with mathematically-based second order Volterra model, to investigate possible influence of continuous assessments on *WS*, leading to a successful grade in a first and second year engineering mathematics course at USQ, a global leader in both on-campus and online and distance education. To explore whether the mode of course offering registers a different pattern of accuracy, the optimal ELM model was also designed and evaluated with predictor datasets for both on-campus and online modes of course offer.

By drawing relevant statistical and visual evidenced from the prescribed data-driven model utilizing multivariate, continuous internal assessment data from 2013–2018 and including quizzes and assignments that were modelled against a *WS*, the ELM model was shown to be the most accurate predictive model. This yielded a relative error of 0.74%, 0.51% (ENM1600 ONC & ONL), and 0.77%, 0.54% (ENM2600 ONC & ONL) in the testing phase. In terms of the possibility of adopting ELM to simulate a *WS* and its respective grade allocation, the frequency of errors attained revealed significant benefits, such as yielding a much larger proportion of tested data that fell within the smallest error bracket. Importantly, the capability of the ELM model to correctly generate a '*Fail*' grade *WS* was clearly evident as was its ability to model other grades with a significantly lower prediction error for levels of engineering mathematics and modes of course offerings, as confirmed by boxplots of the distribution of errors in the testing phase.

TABLE 10. (Case 1: ENM2600 ONC). ELM Model Design Parameters in terms of the optimal input weights, optimal output weights and the bias value of hidden neurons.

	Optimal Input Weights for Each Input Variable					Output Weights of Target Variable	Bias of Hidden Neurons
Neurons	Q2,	Q1,	A2,	A1,	EX	*WS*	*WS*
1	-0.579	0.819	0.894	0.019	0.312	1.391	0.209
2	-0.158	-0.913	0.235	-0.406	-0.189	4.274	0.595
3	-0.564	0.414	-0.262	0.901	-0.485	1.655	0.624
4	0.692	-0.032	0.224	0.632	-0.835	0.040	0.668
5	-0.087	-0.112	-0.588	-0.354	-0.473	-2.920	0.173
6	-0.440	-0.927	-0.670	0.944	-0.457	-1.000	0.899
7	0.866	-0.919	-0.276	0.975	-0.203	-0.045	0.621
8	-0.371	-0.334	0.727	-0.183	-0.630	-3.618	0.044

TABLE 11. Case 2: ENM2600 ONL.

	Optimal Input Weights for Each Input Variable					Output Weights of Target Variable	Bias of Hidden Neurons
Neurons	*Q1,*	*Q2,*	*A2,*	*A1,*	*EX*	*WS*	*WS*
1	-0.579	0.819	0.894	0.019	0.312	1.391	0.954
2	-0.158	-0.913	0.235	-0.406	-0.189	4.274	0.103
3	-0.564	0.414	-0.262	0.901	-0.485	1.655	0.625
4	0.692	-0.032	0.224	0.632	-0.835	0.040	0.442
5	-0.087	-0.112	-0.588	-0.354	-0.473	-2.920	0.424
6	-0.440	-0.927	-0.670	0.944	-0.457	-1.000	0.372
7	0.866	-0.919	-0.276	0.975	-0.203	-0.045	0.868
8	-0.371	-0.334	0.727	-0.183	-0.630	-3.618	0.280

While this study has set a clear foundation for educational designers, course examiners, and higher educational institutions to explore the utility of artificial intelligence models in learning analytics for engineering mathematics performance evaluation (and quite possibly, other courses in the higher education sector), additional factors that can influence a student's success in a course should also be considered. These factors could include gaps in a student's educational history, age, gender, socio-economic status, student's dedicated time and engagement on digital platforms such as the LMS and informal learning activities. If included in a future learning analytics model, these factors could enhance the capability of artificial intelligence algorithms employed in extracting patterns in such data that relate to a grade, and therefore, may assist institutions to perform suitable course health checks, early intervention strategies and modify teaching and learning practices to promote quality education and desired graduate attributes.

CONFLICT OF INTEREST

The authors declare no conflict of interest to any party.

APPENDIX A

See Tables 8–11.

ACKNOWLEDGMENT

The authors would like to thank Prof. Y. Stepanyants for providing examiner return data and contributions to the project proposal and Dr. A. Murphy (Manager Learning Analytics) for providing the editing and other professional contributions. Our appreciation is extended to the respected reviewers and editors for their constructive comments.

REFERENCES

[1] J. Murumba and E. Micheni, "Big data analytics in higher education: A review," *Int. J. Eng. Sci.*, vol. 6, no. 06, pp. 14–21, 2017.

[2] R. Vinuesa, H. Azizpour, I. Leite, M. Balaam, V. Dignum, S. Domisch, A. Felländer, S. D. Langhans, M. Tegmark, and F. F. Nerini, "The role of artificial intelligence in achieving the Sustainable Development Goals," *Nature Commun.*, vol. 11, no. 1, pp. 1–10, 2020.

[3] C. Popa, "Adoption of artificial intelligence in agriculture," *Bull. Univ. Agricult. Sci. Vet. Med. Cluj-Napoca. Agricult.*, vol. 68, no. 1, 2011.

[4] Z. Allam and Z. A. Dhunny, "On big data, artificial intelligence and smart cities," *Cities*, vol. 89, pp. 80–91, Jun. 2019.

[5] K. Kumar and G. S. M. Thakur, "Advanced applications of neural networks and artificial intelligence: A review," *Int. J. Inf. Technol. Comput. Sci.*, vol. 4, no. 6, p. 57, 2012.

[6] R. Bajaj and V. Sharma, "Smart education with artificial intelligence based determination of learning styles," *Procedia Comput. Sci.*, vol. 132, pp. 834–842, 2018.

[7] G. Samigulina and A. Shayakhmetova, "The information system of distance learning for people with impaired vision on the basis of artificial intelligence approaches," in *Smart Education and Smart e-Learning*. Cham, Switzerland: Springer, 2015, pp. 255–263.

[8] W. Xing, R. Guo, E. Petakovic, and S. Goggins, "Participation-based student final performance prediction model through interpretable genetic programming: Integrating learning analytics, educational data mining and theory," *Comput. Hum. Behav.*, vol. 47, pp. 168–181, Jun. 2015.

[9] W. Xing and S. Goggins, "Learning analytics in outer space: A Hidden Naïve Bayes model for automatic student off-task behavior detection," in *Proc. 5th Int. Conf. Learn. Anal. Knowl.*, 2015, pp. 176–183.

[10] K. D. Strang, "Can online student performance be forecasted by learning analytics?" *Int. J. Technol. Enhanced Learn.*, vol. 8, no. 1, p. 26, 2016.

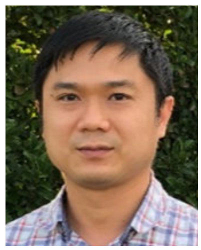
[11] C. J. Villagrá-Arnedo, F. J. Gallego-Durán, F. Llorens-Largo, P. Compañ-Rosique, R. Satorre-Cuerda, and R. Molina-Carmona, "Improving the expressiveness of black-box models for predicting student performance," *Comput. Hum. Behav.*, vol. 72, pp. 621–631, Jul. 2017.
[12] D. M. Olivé, D. Q. Huynh, M. Reynolds, M. Dougiamas, and D. Wiese, "A supervised learning framework for learning management systems," in *Proc. 1st Int. Conf. Data Sci., E-Learn. Inf. Syst. (DATA)*, 2018.
[13] R. L. Ulloa-Cazarez, C. López-Martín, A. Abran, and C. Yáñez-Márquez, "Prediction of online students performance by means of genetic programming," *Appl. Artif. Intell.*, vol. 32, nos. 9–10, pp. 858–881, Nov. 2018.
[14] K. Nuutila, H. Tuominen, A. Tapola, M.-P. Vainikainen, and M. Niemivirta, "Consistency, longitudinal stability, and predictions of elementary school students' task interest, success expectancy, and performance in mathematics," *Learn. Instruct.*, vol. 56, pp. 73–83, Aug. 2018.
[15] A. J. Gabriele, E. Joram, and K. H. Park, "Elementary mathematics teachers' judgment accuracy and calibration accuracy: Do they predict students' mathematics achievement outcomes?" *Learn. Instr.*, vol. 45, pp. 49–60, Oct. 2016.
[16] *The Higher Education Reform Package*, Austral. Government, Australia, 2017.
[17] K. Nelson and T. Creagh. (2013). A Good Practice Guide: Safeguarding Student Learning Engagement. Brisbane: Queensland University of Technology. Jun. 8, 2015. [Online]. Available: http://eprints.qut.edu.au/59189/1/LTU_Good
[18] G. Siemens, S. Dawson, and G. Lynch, "Improving the quality and productivity of the higher education sector," Policy Strategy Syst.-Level Deployment Learn. Anal., Soc. Learn. Anal. Res. Austral. Office Learn. Teach., Canberra, ACT, Australia, Tech. Rep., 2013.
[19] G. Siemens and P. Long, "Penetrating the fog: Analytics in learning and education.," *Educ. Rev.*, vol. 46, no. 5, p. 30, 2011.
[20] R. K. Yin and G. B. Moore, "The use of advanced technologies in special education: Prospects from robotics, artificial intelligence, and computer simulation," *J. Learn. Disabilities*, vol. 20, no. 1, pp. 60–63, Jan. 1987.
[21] I. Becerra-Fernandez, "The role of artificial intelligence technologies in the implementation of people-finder knowledge management systems," *Knowl.-Based Syst.*, vol. 13, no. 5, pp. 315–320, Oct. 2000.
[22] J. Campbell, P. DeBlois, and D. Oblinger, "Academic analytics: A new tool for a new era," *Educause Rev.*, vol. 42, no. 4, p. 40, Jul. 2007.
[23] C. E. Cook, M. Wright, and C. O'Neal, "8: Action research for instructional improvement: Using data to enhance student learning at your institution," *Improv. Acad.*, vol. 25, no. 1, pp. 123–138, 2007.
[24] G. Gokmen, T. Ç. Akinci, M. Tektaş, N. Onat, G. Kocyigit, and N. Tektaş, "Evaluation of student performance in laboratory applications using fuzzy logic," *Procedia-Social Behav. Sci.*, vol. 2, no. 2, pp. 902–909, 2010.
[25] R. S. Yadav and V. P. Singh, "Modeling academic performance evaluation using soft computing techniques: A fuzzy logic approach," *Int. J. Comput. Sci. Eng.*, vol. 3, no. 2, pp. 676–686, 2011.
[26] R. Sripan and B. Suksawat, "Propose of fuzzy logic-based students' learning assessment," in *Proc. ICCAS*, 2010, pp. 414–417.
[27] R. Bhatt and D. Bhatt, "Fuzzy logic based student performance evaluation model for practical components of engineering institutions subjects," *Int. J. Technol. Eng. Educ.*, vol. 8, no. 1, pp. 1–7, 2011.
[28] K.-A. Hwang and C.-H. Yang, "Attentiveness assessment in learning based on fuzzy logic analysis," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6261–6265, Apr. 2009.
[29] S. A. Naser, I. Zaqout, M. A. Ghosh, R. Atallah, and E. Alajrami, "Predicting student performance using artificial neural network: In the faculty of engineering and information technology," *Int. J. Hybrid Inf. Technol.*, vol. 8, no. 2, pp. 221–228, Feb. 2015.
[30] O. C. Asogwa and A. V. Oladugba, "Of students academic performance rates using artificial neural networks (ANNs)," *Amer. J. Appl. Math. Stat.*, vol. 3, no. 4, pp. 151–155, 2015.
[31] B. Naik and S. Ragothaman, "Using neural networks to predict MBA student success," *Coll. Stud. J.*, vol. 38, no. 1, pp. 143–150, 2004.
[32] W. L. Gorr, D. Nagin, and J. Szczypula, "Comparative study of artificial neural network and statistical models for predicting student grade point averages," *Int. J. Forecasting*, vol. 10, no. 1, pp. 17–34, Jun. 1994.
[33] S. Huang and N. Fang, "Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models," *Comput. Edu.*, vol. 61, pp. 133–145, Feb. 2013.
[34] R. Stevens, J. Ikeda, A. Casillas, J. Palacio-Cayetano, and S. Clyman, "Artificial neural network-based performance assessments," *Comput. Human Behav.*, vol. 15, nos. 3–4, pp. 295–313, 1999.
[35] M. D. Calvo-Flores, E. G. Galindo, M. C. P. Jiménez, and O. P. Pineiro, "Predicting students' marks from Moodle logs using neural network models," *Current Develop. Technol. Edu.*, vol. 1, no. 2, pp. 586–590, 2006.
[36] A. M. Shahiri, W. Husain, and N. A. Rashid, "A review on predicting Student's performance using data mining techniques," *Procedia Comput. Sci.*, vol. 72, pp. 414–422, 2015.
[37] M. Asiah, K. N. Zulkarnaen, D. Safaai, M. Y. N. N. Hafzan, M. M. Saberi, and S. S. Syuhaida, "A review on predictive modeling technique for student academic performance monitoring," in *Proc. MATEC Web Conf.*, vol. 255, 2019, p. 3004.
[38] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.
[39] G. Huang, G.-B. Huang, S. Song, and K. You, "Trends in extreme learning machines: A review," *Neural Netw.*, vol. 61, pp. 32–48, Jan. 2015.
[40] G.-B. Huang, "What are extreme learning machines? Filling the gap between frank Rosenblatt's dream and John von Neumann's puzzle," *Cognit. Comput.*, vol. 7, no. 3, pp. 263–278, Jun. 2015.
[41] G.-B. Huang, "An insight into extreme learning machines: Random neurons, random features and kernels," *Cognit. Comput.*, vol. 6, no. 3, pp. 376–390, Sep. 2014.
[42] G.-B. Huang, L. Chen, and C.-K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Trans. Neural Netw.*, vol. 17, no. 4, pp. 879–892, Jul. 2006.
[43] C. Deng, G. Huang, J. Xu, and J. Tang, "Extreme learning machines: New trends and applications," *Sci. China Inf. Sci.*, vol. 58, no. 2, pp. 1–16, Feb. 2015.
[44] L. Breiman, "Random Forrests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[45] R. Genuer, J.-M. Poggi, C. Tuleau-Malot, and N. Villa-Vialaneix, "Random forests for big data," *Big Data Res.*, vol. 9, pp. 28–46, Sep. 2017.
[46] K. Fawagreh, M. M. Gaber, and E. Elyan, "Random forests: From early developments to recent advancements," *Syst. Sci. Control Eng.*, vol. 2, no. 1, pp. 602–609, Dec. 2014.
[47] M. R. Segal, "Machine learning benchmarks and random forest regression," Center Bioinf. Mol. Biostatistics, Univ. California, San Francisco, CA, USA, Tech. Rep., 2003.
[48] L. Heng-Chao, Z. Jia-Shu, and X. Xian-Ci, "Neural Volterra filter for chaotic time series prediction," *Chin. Phys.*, vol. 14, no. 11, p. 2181, 2005.
[49] R. Prasad, R. C. Deo, Y. Li, and T. Maraseni, "Ensemble committee-based data intelligent approach for generating soil moisture forecasts with multivariate hydro-meteorological predictors," *Soil Tillage Res.*, vol. 181, pp. 63–81, Sep. 2018.
[50] S. A. Billings, M. J. Korenberg, and S. Chen, "Identification of non-linear output-affine systems using an orthogonal least-squares algorithm," *Int. J. Syst. Sci.*, vol. 19, no. 8, pp. 1559–1568, Jan. 1988.
[51] H. L. Wei and S. A. Billings, "A unified wavelet-based modelling framework for non-linear system identification: The WANARX model structure," *Int. J. Control*, vol. 77, no. 4, pp. 351–366, Mar. 2004.
[52] M. A. Shahin, H. R. Maier, and M. B. Jaksa, "Data division for developing neural networks applied to geotechnical engineering," *J. Comput. Civil Eng.*, vol. 18, no. 2, pp. 105–114, Apr. 2004.
[53] ASCE Task Committee, "Criteria for evaluation of watershed models," *J. Irrigation Drainage Eng.*, vol. 119, no. 3, pp. 429–442, May 1993.
[54] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature," *Geoscientific Model Develop.*, vol. 7, no. 3, pp. 1247–1250, Jun. 2014.
[55] P. H. Farquhar, "Applications of utility theory in artificial intelligence research," in *Toward Interactive and Intelligent Decision Support Systems*. Springer, 1987, pp. 155–161.
[56] *Study Book (ENM1600 Engineering Mathematics)*, USQ, Toowoomba, QLD, Australia, 2019.
[57] *Study Book (ENM2600 Engineering Mathematics*, USQ, Toowoomba, QLD, Australia, 2019.
[58] X. Wang and M. Han, "Online sequential extreme learning machine with kernels for nonstationary time series prediction," *Neurocomputing*, vol. 145, pp. 90–97, Dec. 2014.
[59] S. Scardapane, D. Comminiello, M. Scarpiniti, and A. Uncini, "Online sequential extreme learning machine with kernels," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 9, pp. 2214–2220, Sep. 2015.
[60] N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "A fast and accurate online sequential learning algorithm for feedforward networks," *IEEE Trans. Neural Netw.*, vol. 17, no. 6, pp. 1411–1423, Nov. 2006.
[61] M. Hourigan and J. O'Donoghue, "Mathematical under-preparedness: The influence of the pre-tertiary mathematics experience on students' ability to make a successful transition to tertiary level mathematics courses in ireland," *Int. J. Math. Edu. Sci. Technol.*, vol. 38, no. 4, pp. 461–476, Jun. 2007.

[62] I. Ryabov, "The effect of time online on grades in online sociology courses," *MERLOT J. Online Learn. Teach.*, vol. 8, no. 1, pp. 13–23, 2012.
[63] S. Vonderwell and S. Zachariah, "Factors that influence participation in online learning," *J. Res. Technol. Edu.*, vol. 38, no. 2, pp. 213–230, Dec. 2005.
[64] D. D. Curtis and M. J. Lawson, "Exploring collaborative online learning," *Online Learn.*, vol. 5, no. 1, 2019.
[65] S. R. Palmer and D. M. Holt, "Examining student satisfaction with wholly online learning," *J. Comput. Assist. Learn.*, vol. 25, no. 2, pp. 101–113, Mar. 2009.
[66] M. Devlin and J. McKay, "Reframing 'the problem': Students from low socioeconomic status backgrounds transitioning to University," 2014.
[67] J. McKay and M. Devlin, "Widening participation in Australia: Lessons on equity, standards, and institutional leadership," in *Widening Higher Education Participation*. Amsterdam, The Netherlands: Elsevier, 2016, pp. 161–179.
[68] K. P. Mongeon, S. W. Ulrick, and M. P. Giannetto, "Explaining University course grade gaps," *Empirical Econ.*, vol. 52, no. 1, pp. 411–446, Feb. 2017.
[69] D. Contini, M. L. D. Tommaso, and S. Mendolia, "The gender gap in mathematics achievement: Evidence from Italian data," *Econ. Edu. Rev.*, vol. 58, pp. 32–42, Jun. 2017.
[70] M. Devlin and J. McKay, "Facilitating success for students from low socioeconomic status backgrounds at regional universities," Federation Univ., Mount Helen, VIC, Australia, Tech. Rep., 2017.
[71] A. Bonneville-Roussy, P. Evans, J. Verner-Filion, R. J. Vallerand, and T. Bouffard, "Motivation and coping with the stress of assessment: Gender differences in outcomes for University students," *Contemp. Educ. Psychol.*, vol. 48, pp. 28–42, Jan. 2017.

RAVINESH C. DEO (Senior Member, IEEE) is Associate Professor (Research Leader – Artificial Intelligence) at University of Southern Queensland with international expertise in decision support systems, deep learning and metaheuristic algorithms. He was awarded internationally competitive fellowships such as Queensland US Smithsonian, Australia–China Young Scientist, JSPS, Chinese Academy of Science President's, Australia-India and Australia Endeavour Fellowship. He has supervised over eighteen research degrees including industry mentorship projects through externally-funded grants, held visiting professor roles internationally and is professionally affiliated under numerous scientific bodies. He serves on Editorial Board of Stochastic Environmental Research & Risk Assessment, Remote Sensing & Energies journals including the American Society for Civil Engineer's Journal of Hydrologic Engineering, contributing to teaching, research and service in data science, mathematics and environmental modelling. He published over 190 papers including 150 journals (mostly Scopus Quartile 1), six Books under Elsevier, Springer Nature and IGI Global and ten Book Chapters that have cumulative citation exceeding 4700 with H-Index of 37.

ZAHER MUNDHER YASEEN received the master's and Ph.D. degrees from the National University of Malaysia (UKM), Malaysia, in 2012 and 2017, respectively. He is currently a Senior Lecturer and a Senior Researcher of civil engineering. He has published over 150 articles in international journals, with a Google Scholar H-Index of 27 and a total of 2400 citations. His major interests are hydrology, water resources engineering, hydrological process modeling, environmental engineering, and the climate. In addition, his interests include machine learning and advanced data analytics.

NADHIR AL-ANSARI received the B.Sc. and M.Sc. degrees from the University of Baghdad, in 1968 and 1972, respectively, and the Ph.D. degree in water resources engineering from Dundee University, in 1976. He is currently a Professor with the Department of Civil, Environmental and Natural Resources Engineering, Lulea Technical University, Sweden. Previously, he worked at Baghdad University, from 1976 to 1995, and then at Al-Bayt University, Jordan, from 1995 to 2007. He has served several academic administrative post (the Dean and the Head of Department). His publications include more than 424 articles in international/national journals, chapters in books and 13 books. He executed more than 60 major research projects in Iraq, Jordan, and U.K. He has one patent on Physical methods for the separation of iron oxides. He supervised more than 66 postgraduate students at Iraq, Jordan, U.K., and Australia universities. His research interests include geology, water resources, and environment. He is a member of several scientific societies, e.g., International Association of Hydrological Sciences, Chartered Institution of Water and Environment Management, and Network of Iraqi Scientists Abroad, and the Founder and the President of the Iraqi Scientific Society for Water Resources, and so on. He awarded several scientific and educational awards, among them is the British Council on its 70th Anniversary awarded him top five scientists in Cultural Relations. He is a member of the editorial board of ten international journals.



THONG NGUYEN-HUY received the Ph.D. degree in applied statistics and mathematics from the University of Southern Queensland (USQ), Australia, in 2019. He is currently a postdoctoral research fellow at the Centre for Applied Climate Sciences (USQ). His research interests include statistical and data-driven modeling, climate extremes and adaptations, agricultural systems, optimization, GIS and remote sensing.

TREVOR ASHLEY MCPHERSON LANDLANDS is the Senior Lecturer in Mathematics and Discipline Leader for Mathematics at University of Southern Queensland, Australia. He received his PhD degree in applied mathematics from the Queensland University Technology (QUT), Australia in 2001. He is currently a Senior Lecturer in Mathematics at the University of Southern Queensland (USQ) since 2009 and previously at the University of New South Wales (UNSW) Australia from 2003–2009. His research interests include anomalous Subdiffusion, fractional calculus, and industrial modelling. Trevor is a member of the Australian Mathematical Society (AustMS) and Member of the Australia and New Zealand Industrial and Applied Mathematics (ANZIAM), and teaches operations research, engineering mathematics and other areas in applied mathematics. He has 1290 citations from his works in the area of applied mathematics modelling and computational maths and has a H index of 14 (Scopus).

LINDA GALLIGAN is the Head of School of Sciences at University of Southern Queensland, Australia. She is a Senior Executive within School of Sciences and conducts research in embedding numeracy in nursing, language and Mathematics, cross-cultural differences in mathematics education, adult/academic numeracy and technology in mathematics education. She is affiliated with the Queensland College of Teachers and has industry membership of the Mathematics Education Group of Australasia and Member of Adult Learning Mathematics.