Computers and Education: Artificial Intelligence 7 (2024) 100331

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence

Explainable artificial intelligence-machine learning models to estimate overall scores in tertiary preparatory general science course

Sujan Ghimire [a], Shahab Abdulla [b], Lionel P. Joseph [a], Salvin Prasad [e], Angela Murphy [c], Aruna Devi [d], Prabal Datta Barua [a], Ravinesh C. Deo [a,*], Rajendra Acharya [a], Zaher M. Yaseen [f]

[a] School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, QLD, 4300, Australia
[b] UniSQ College, University of Southern Queensland, Toowoomba, QLD, 4300, Australia
[c] Academic Portfolio, University of Southern Queensland, Toowoomba, QLD, 4350, Australia
[d] School of Education and Tertiary Access, The University of the Sunshine Coast, Moreton Bay Campus, Petrie, QLD 4502, Australia
[e] Fiji National University, Suva, Fiji
[f] Civil and Environmental Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran, 31261, Saudi Arabia

ARTICLE INFO

Keywords:
Tertiary preparatory program
Student performance prediction
Students at risk
Education models
Deep learning
Machine learning
Support vector regression

ABSTRACT

Educational data mining is valuable for uncovering latent relationships in educational settings, particularly for predicting students' academic performance. This study introduces an interpretable hybrid model, optimised through Tree-structured Parzen Estimation (TPE) and Support Vector Regression (SVR), to predict overall scores (OT) utilising five assignments and one examination mark as predictors. Neural Network-based, Tree-Based, Ensemble-Based, and Boosting-based methods are evaluated against the hybrid TPE-optimised SVR model for forecasting final examination grades among 492 students enrolled in the TPP7155 (General Science) course at the University of Southern Queensland, Australia, during the 2020-2021 academic year. Additionally, Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive explanations (SHAP) techniques are employed to elucidate the inner workings of these prediction models. The findings highlight the superior performance of the proposed model, exhibiting the lowest Root Mean Squared Error ($RMSE$) and Relative Root Mean Squared Error ($RRMSE$), as well as the highest Willmott's index (WI), Legates–McCabe index (LM), and Nash–Sutcliffe Efficiency (NS). With assignment and examination marks identified as pivotal predictors of OT. SHAP and LIME analyses reveal the examination score (ET) as the most influential feature, impacting predicted OT by an average of ±4.93. Conversely, Assignment 1 emerges as the least informative feature, contributing merely ±0.64 to OT predictions. This research underscores the efficacy of the proposed interpretable hybrid TPE-optimised SVR model in discerning relationships among continuous learning variables, thereby empowering educators with early intervention capabilities and enhancing their ability to anticipate student performance prior to course completion.

1. Background

This study develops explainable artificial intelligence (XAI)-based machine learning models to estimate overall student performance in a tertiary preparatory general science course. Given the high attrition rates and academic struggles of students from low socioeconomic (SES) backgrounds, early and targeted academic interventions are essential. The Australian Government's *Higher Education Standards Framework* (HESF) sets out minimum requirements for Australian universities to ensure high-quality education and equitable access across student demographics. It mandates that institutions monitor and improve student outcomes, with particular emphasis on identifying and supporting at-risk groups, such as students from low socioeconomic backgrounds. In the context of this study, the HESF's guidelines underscore the importance of implementing data-driven tools to identify students at risk of academic failure. (Alyahyan & Düştegör, 2020).

* Corresponding author.
E-mail addresses: sujan.ghimire@unisq.edu.au (S. Ghimire), shahab.abdulla@unisq.edu.au (S. Abdulla), joseph.lionel73@gmail.com (L.P. Joseph), psdshalvin@gmail.com (S. Prasad), angela.murphy@unisq.edu.au (A. Murphy), adevi@usc.edu.au (A. Devi), prabal.barua@unisq.edu.au (P.D. Barua), ravinesh.deo@unisq.edu.au, ravinesh.deo@unisq.edu.au (R.C. Deo), rajendra.acharya@unisq.edu.au (R. Acharya), Zaheryaseen88@gmzail.com, z.yaseen@kfupm.edu.sa (Z.M. Yaseen).

https://doi.org/10.1016/j.caeai.2024.100331
Received 9 July 2024; Received in revised form 13 November 2024; Accepted 14 November 2024

Available online 20 November 2024

Academic underperformance is a major contributor to attrition. Previous models have explored factors such as financial and personal challenges (Li & Jackson, 2024; Zając et al., 2024), but often lack transparency or practical applicability in university contexts. Additionally, many models overlook preparatory pathways such as enabling programs for students from low SES and other under-represented groups. This gap is significant as preparatory students face unique challenges, often struggling with academic preparedness and high attrition rates upon university entry, with attrition rates in preparatory programs reaching up to 50% in some cases (Chesters et al., 2018; Bookallil & Harreveld, 2017; Hodges et al., 2013).

Our study employs explainable machine learning models to address this gap, offering interpretable, data-driven insights that enable targeted academic interventions tailored to the specific needs of students in enabling programs. EML models, such as decision trees or interpretable neural networks, play a crucial role in educational settings by allowing educators to understand and trust the model's predictions and recommendations. Unlike "black-box" models, which offer high accuracy but limited interpretability (Djurisic et al., 2020), explainable models provide transparency regarding which factors most influence a student's predicted outcome (Jang et al., 2022). This informed, data driven approach empowers educators to make decisions that are not only based on statistical evidence but are also aligned with students' real-world needs. By using explainable AI (*x*AI) models, educational institutions can promote accountability and equity in their support strategies, ensuring that interventions are fairly distributed and that all students receive assistance based on transparent, evidence-based criteria (Khosravi et al., 2022). This approach not only enhances the effectiveness of support programs by offering clear, interpretable explanations, but also strengthens the ethical foundation upon which these data-driven decisions are made (Hu et al., 2021).

2. Literature review

2.1. Student performance prediction models in enabling programs

Enabling programs, funded by the Australian government, aim to increase participation from low SES backgrounds and other under-represented groups, promoting equity, economic opportunity, and social justice (Bradley et al., 2008). By 2013, 27 Australian universities offered over 35 such programs (Hodges et al., 2013), with offerings and enrolments steadily increasing to meet demand (Lisciandro & Gibbs, 2016). At the University of Southern Queensland (UniSQ), the Tertiary Preparatory Program (TPP) equips students lacking conventional entry credentials with essential academic skills and the confidence needed to succeed in university environments. TPP students face vulnerabilities (Chesters & Watson, 2013; Edwards & McMillan, 2015), often struggling with socioeconomic and personal barriers, including limited academic preparedness.

Attrition risk is heightened by factors such as family responsibilities, financial pressures (e.g., student debt), limited support services, and academic challenges, especially poor performance (Li & Jackson, 2024; Zając et al., 2024). Research indicates that students from low SES backgrounds experience lower completion rates (Cooper et al., 2000; Guenther & Johnson, 2010). Poor academic performance can lead to demotivation and depression (Kocsis & Molnár, 2024), reinforcing attrition risks (Arias et al., 2024). Unlike studies focused on general university cohorts, this research develops explainable machine learning models tailored for preparatory programs. These models predict academic outcomes and provide actionable insights to inform targeted interventions, enhancing retention and academic success in enabling programs (Onyema et al., 2020).

Higher education institutions are increasingly using empirical data to improve student performance (Nguyen-Huy et al., 2022). Universities have access to comprehensive data, including continuous assessment data, academic performance metrics, attendance records, course evaluations, and demographic information. Analysing these data sources can provide valuable insights into students' learning patterns, helping to identify strengths, weaknesses, and factors contributing to academic success. This analysis is crucial for monitoring TPP cohorts and facilitating early intervention when necessary. By leveraging learning analytics methods to predict student outcomes, educators can align their teaching strategies with evidence-based pedagogies that have been shown to improve student engagement and retention. For example, predictive models may suggest that students identified as at-risk benefit from peer mentorship or collaborative learning, guiding educators to implement evidence-based interventions that are specifically tailored to student needs (Paolucci et al., 2024).

2.2. Educational Data Mining (EDM) and learning analytics in higher education

Educational data mining (EDM) has gained significant interest for its potential to improve educational support (Yağcı, 2022). EDM applies various analytical methods to student data, enabling personalised educational support by identifying learning patterns and providing frameworks for the systematic analysis and interpretation of educational data (Buenaño-Fernández et al., 2019). EDM leverages statistical analysis and machine learning algorithms to extract actionable insights from complex educational datasets. Both qualitative and quantitative data, such as assessment scores, interviews and student feedback, are used to address educational inequities and optimise resource allocation (Hashim et al., 2020). EDM also contributes to achieving Sustainable Development Goal 4 (SDG 4), aiming to provide quality education globally (Saini et al., 2023). By analysing student data, EDM enables the development of personalised learning experiences (McKenney & Mor, 2015) improving learning outcomes, course effectiveness and student satisfaction (Injadat et al., 2020). Various studies have focused on predicting student performance, utilising assessment data and other continuous learning activities as key indicators (Bertolini et al., 2022; Priyambada et al., 2023; Santos & Henriques, 2023). Continuous assessment, such as quizzes, assignments or projects, provides essential feedback that enhances teaching activities and helps educators identify areas for improvement (Yadav & Singh, 2011). The marks obtained from continuous assessments and final examinations are quantifiable and, therefore can be represented mathematically, making them valuable inputs for predicting the overall score (OT) which determines the final grades and the grade point average (GPA) (Ahmed et al., 2022).

2.3. Machine learning for predicting academic outcomes in preparatory programs

In recent years, machine learning (ML) techniques have gained prominence in educational research, particularly for predicting student success and enhancing academic support. ML methods, including decision trees, random forests, and neural networks, are well-suited to handle complex datasets often encountered in higher education. Many studies have used tree-based models, neural networks, and kernel-based machine learning techniques for performance prediction. These models are effective in capturing complex relationships in student data and are particularly advantageous in settings such as preparatory programs, where students may face diverse challenges related to their socio-economic backgrounds, prior academic performance, and personal circumstances. The use of ML in educational settings supports inclusive pedagogy by providing early identification of students who may face barriers to success, such as financial hardship or lack of prior academic preparation (Dubey, 2024). By offering targeted interventions, such as financial aid or additional academic resources, machine learning models ensure equitable access to education for all students, promoting greater inclusion and diversity in preparatory programs

2.3.1. Decision trees and random forests

Decision trees are among the most widely used ML techniques in educational contexts, especially for predicting student performance based on a set of input features (e.g., prior academic records, socio-economic status, attendance). In preparatory programs, where students often come from non-traditional backgrounds and may lack academic preparedness, decision trees can identify which factors, such as lack of prior study skills or limited access to educational resources, most influence academic success. Decision trees operate by splitting the data into subsets based on feature values, which allows for easy interpretation and visualisation. Random forests, an ensemble of decision trees, can improve the robustness and accuracy of predictions. By aggregating the predictions of multiple trees, random forests reduce overfitting and enhance model reliability, particularly when dealing with noisy or imbalanced data, common in preparatory program cohorts. For example, random forests could provide a more reliable prediction of a student's likelihood of completing a preparatory program, factoring in the complex interplay of personal, academic, and financial challenges that these students face.

To explore the effectiveness of tree-structured models, Hussain and Khan (2023) developed a decision tree (DT) model to predict grades and forecast exam marks, achieving better accuracy than the k-nearest neighbour (KNN) model. Rai et al. (2021) applied a random forest (RF) model to predict student grades, showing that RF models can identify areas for improvement, demonstrating that the RF model effectively captured the complex relationships between the input features and student performance. The use of RF was particularly beneficial for students in the poor and average performance categories, as it provided insights that helped identify areas for improvement and tailor educational interventions. Cheng et al. (2024) used extreme gradient boosting (XGBoost) to predict student performance, achieving an accuracy of 80.4%, outperforming other models like RF, DT, and KNN. Tree-based models (e.g., DT, RF, and XGBoost) are interpretable and offer better predictive accuracy than statistical models. This is mainly due to their capacity to capture complex, nonlinear relationships within the data. However, when applied to large datasets, a single tree can grow excessively large, resulting in a high number of nodes (Joseph et al., 2023; Hakkal & Lahcen, 2024). This increase in model complexity often leads to over-fitting, where the model performs well on training data but poorly on unseen test data.

2.3.2. Neural networks and deep learning

Neural networks, particularly deep learning models, have shown great promise in handling large and complex datasets, often with higher accuracy than traditional models. In the context of preparatory programs, these models can capture non-linear relationships between a variety of factors, such as mental health status, socio-economic background, and prior educational attainment, that influence student outcomes.

Neural networks, such as artificial neural networks (ANN), utilise regularisation techniques and early stopping to address the overfitting issues often associated with tree-based models (Srivastava et al., 2014). Studies using ANN models such as Chavez et al. (2023) have achieved high predictive accuracy in predicting student outcomes, surpassing traditional models like RF and Naïve Bayes (NB). Similarly, Arsad et al. (2014) demonstrated that ANN models can predict the overall performance of students in a Bachelor of Electrical Engineering programme, utilising grade points (GP) from fundamental core subjects as input parameters, outperforming a benchmark linear regression model. Tomasevic et al. (2020) employed an ANN model to predict final examination scores and identify students at high risk of dropping out utilising demographic data, student engagement metrics, and historical performance records. The model outperformed benchmark models such as KNN, decision trees, and linear regression. An adaptive neuro-fuzzy inference system (ANFIS) model was trained using six input variables derived from assessments and examinations (Taylan & Karagözoğlu, 2009), achieving very low error in its predictions of student performance.

Unlike decision trees, which focus on explicit splits in data, neural networks learn intricate patterns from vast datasets, allowing for more nuanced predictions. Additionally, deep learning models can be trained to assess the effectiveness of different interventions, enabling preparatory programs to adapt and personalise their support strategies. While neural network-based models have demonstrated strong performance in the literature, it is important to note that they can sometimes get stuck in local minima during gradient descent optimisation (Pascanu et al., 2013), which may prevent the model from converging to the global optimum, leading to sub-optimal performance.

2.4. Support vector machines and kernel-based models

Kernel-based models, such as support vector machines (SVM), address some limitations associated with neural networks, particularly in avoiding local minima during training. Unlike neural networks, SVMs are grounded in the principle of margin maximisation, which aims to find a hyperplane that optimally separates data points while minimising classification error (Cao & Tay, 2003). Regularisation techniques, including adjustments to the regularisation parameter (C) and kernel coefficient (gamma) (Basak et al., 2007), help SVM models avoid overfitting and improve generalisation by controlling model complexity.

Studies have illustrated SVM's predictive potential in educational contexts. For instance, Pang et al. (2017) applied SVM with simulated annealing to predict student graduation, incorporating features like demographics and psychological data, achieving high predictive accuracy. Samsudin et al. (2022), applied support vector regression (SVR), a regression-based variant of SVM, to predict student academic performance during the Covid-19 pandemic by focusing on cumulative grade point averages (CGPA). Dewi and Widiastuti (2020) demonstrated the effectiveness of SVR for predicting CGPA in Indonesian students, using a radial basis function (RBF) kernel to yield an RMSE of 0.1861, outperforming standard linear regression.

Recent optimisations, such as particle swarm optimisation (PSO), have further refined SVR's performance. Apriyadi et al. (2023) developed a PSO-SVR model based on a variety of demographic and academic features, achieving superior accuracy over models like decision trees and neural networks. However, while promising, effective SVR implementation requires careful hyperparameter tuning, as seen with the use of Bayesian optimisation through Tree-structured Parzen Estimator (TPE), which outperformed PSO in other domains by achieving superior hyperparameter selection at lower computational costs (Vasanthanageswari, 2022). Without proper tuning, the model may suffer from poor accuracy and unreliable predictions, as indicated in several studies (Apriyadi et al., 2023; Pang et al., 2017). To address this, several researchers (Jiang et al., 2024; Omotehinwa et al., 2023; Tao et al., 2024) have identified the tree-structured Parzen estimator (TPE) as a highly effective method for hyperparameter optimisation. TPE employs a Bayesian optimisation approach, balancing exploration (evaluating diverse hyperparameter configurations) and exploitation (focusing on the most promising configurations) to efficiently search the hyperparameter space while minimising computational costs (Bergstra et al., 2011).

The TPE method's effectiveness has been demonstrated in various applications, such as optimising SVM for crop yield prediction, where it outperformed PSO by selecting superior hyperparameters (Vasanthanageswari, 2022). TPE has also been used to fine-tune SVM for electron energy loss spectroscopy (EELS) spectra classification (del Pozo-Bueno et al., 2023), significantly enhancing SVM's performance compared to random search (RS) optimisation. Furthermore, TPE has been applied to optimise SVR in predicting the remaining useful life of lithium-ion batteries, achieving an 89.84% reduction in RMSE compared to standalone SVR models (Deng et al., 2023).

Among the most widely used *x*AI techniques are SHapley Additive exPlanations (SHAP) (Lundberg et al., 2020) and Local Interpretable Model-agnostic Explanations (LIME) (Ribeiro et al., 2016) methods. LIME has been effectively used in previous studies to interpret tree-based models, such as XGBoost and DT, in the context of student performance prediction (Vultureanu-Albişi & Bădică, 2021). In these cases, LIME effectively provided insights into individual predictions. However, SHAP's ability to provide global insights into feature importance, derived from Shapley values in cooperative game theory, makes it especially valuable for understanding overall model behaviour.

Our study applies TPE-optimised SVR to predict academic success in enabling programs. The use of TPE helps identify high-potential configurations for SVR's hyperparameters, maximising prediction accuracy while maintaining computational efficiency (Bergstra et al., 2011). Furthermore, as SVR's "black-box" nature can limit interpretability, which is crucial for practical educational applications, we enhance interpretability by employing model-agnostic explanation methods to make SVR predictions more transparent for educational stakeholders. By demystifying the SVR model's predictions, educators and administrators can more effectively understand factors contributing to student success and attrition, thereby guiding targeted interventions in enabling programs.

3. Context and study objectives

This research reports the findings of the UniSQ Technology Demonstrator project, which uses an SVR model as the primary algorithm to predict student performance in the Tertiary Preparatory (TPP115) General Science course. Building on earlier studies on student performance models (Deo et al., 2020; Nguyen-Huy et al., 2022; Ahmed et al., 2022), this work aims to refine predictive methods for educational outcomes. The TPP7155 course, taught at UniSQ College under the Tertiary Preparatory Program, covers topics such as scientific methods, measurement in science, matter, antibiotic resistance, climate change, and genetics. The course aims to enhance scientific literacy, particularly in interpreting, analysing, and evaluating scientific data. With the increasing emphasis on STEM fields due to recent government initiatives, students in the course are provided with foundational concepts to pursue undergraduate degrees in nursing, teacher education, general science, and engineering.

In this research, the weighted scores of five independent assignments (W1 = 5%, W2 = 15%, W3 = 10%, W4 = 20%, W5 = 5%) and a final examination (ET = 45%) are used to determine the Overall Mark (OT = 100%) for students in the TPP7155 course. The non-linear relationships between these components present challenges for conventional statistical models, such as autoregressive integrated moving average (ARIMA), linear regression, and partial and ordinary differential equations (Deo et al., 2020), which fail to capture the complex dependencies in the data. In contrast, AI-based machine learning models are better equipped to handle these non-linear patterns and interactions.

The primary objective of this research is to develop an explainable hybrid SVR model, optimised with the tree-structured Parzen estimator (TPE), to predict students' Overall Marks (OT) in the TPP7155 General Science course. The model is designed to serve as a practical tool for academics, providing insights into how continuous assessments throughout the term influence students' final grades. This can help educators better understand the non-linear relationships between assessments, ultimately improving student performance. The key contributions of this research study include:

- Development of a robust SVR model to predict student performance in TPP7155 using empirical data from both assignments and exams.
- Optimisation of the SVR model through efficient hyperparameter tuning using the TPE algorithm, with rigorous benchmarking against other machine learning models.
- Application of SHAP (SHapley Additive exPlanations) for global interpretability, enhancing the transparency of the SVR model and providing actionable insights for educators.

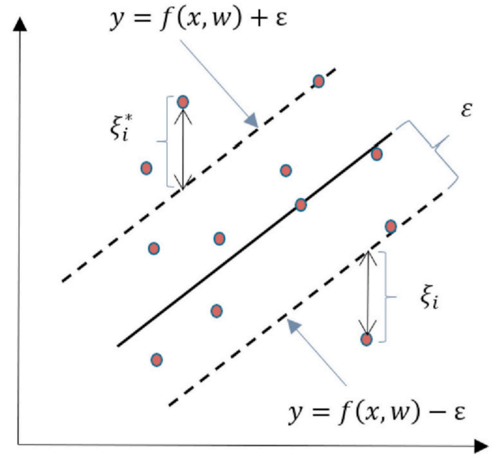

Fig. 1. Schematic representation of the SVR algorithm showing the slack variable (ξ). The mathematical symbols are outlined in Section 4.1 along with the SVR model equations.

This research offers a novel predictive model for higher education, enabling accurate and interpretable predictions of OT. By supporting early identification of at-risk students, the model allows for timely interventions, enhancing student success and retention.

4. Theoretical overview and methodology

4.1. Proposed Support Vector Regression (SVR) model

This study utilises Support Vector Regression (SVR) as the primary tool to predict OT of students in TPP7155 (General Science), a course that typically has students from diverse backgrounds with little experience in scientific related endeavours. The SVR model has been selected due to its efficacy in scenarios with limited training samples and computational resources (Ma'sum, 2022). SVR, grounded in statistical learning theory, excels in high-dimensional regression problems by offering robust data generalisation and achieving a global optimal solution, thus circumventing the "curse of dimensionality" (Ghimire et al., 2022). The superior performance of SVR in prediction tasks, compared to other non-linear models, has contributed to its extensive application in various fields. For a training data set $\{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_n, y_n) | \mathbf{x}_i \in \mathbb{R}^D, y_i \in \mathbb{R}\}$, where $\mathbb{R}^D$ is a D-dimensional real input vector, $y_i \in \mathbb{R}$ is the corresponding target value, and n is the total number of data patterns, the regression function of the SVR model is expressed as follows (Zhang & O'Donnell, 2020):

$$f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b \tag{1}$$

where $\mathbf{w} \in \mathbb{R}^D$ is a weight vector, T stands for the transpose operator. The term b is a bias, $\phi(\cdot)$ is a nonlinear transfer function mapping the input vectors into a high dimensional feature space (Fig. 1).

The slack variables ξ_i and ξ_i^* are defined to address infeasible constraints. The SVR algorithm's optimisation problem can be expressed using Equation (2):

$$\text{Subject to} \begin{cases} f(\mathbf{x}_i) - y_i \le \epsilon + \xi_i \\ y_i - f(\mathbf{x}_i) \le \epsilon + \xi_i^* \\ \xi_i \ge 0 \quad i = 1, 2, \cdots, n \\ \xi_i^* \ge 0 \quad i = 1, 2, \cdots, n \end{cases} \tag{2}$$

Again, transform the objective function into the unconstrained Lagrange objective function as Equation (3):

$$L(\mathbf{w}, b, \boldsymbol{\alpha}, \boldsymbol{\alpha}^*, \boldsymbol{\xi}, \boldsymbol{\xi}^*, \boldsymbol{\nu}, \boldsymbol{\nu}^*) = \frac{1}{2}\mathbf{w}^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) - \sum_{i=1}^{n}\mu_i\xi_i - \sum_{i=1}^{n}\mu_i^*\xi_i^* + \sum_{i=1}^{n} a_i(f(\mathbf{x}_i) - y_i - \epsilon - \xi_i) + \sum_{i=1}^{n} a_i^*(y_i - f(\mathbf{x}_i) - \epsilon - \xi_i^* \tag{3}$$

where the Lagrange multipliers are $a_i \geq 0$, $a_i^* \geq 0$, $\mu_i \geq 0$, and $\mu_i^* \geq 0$. C is the punishment factor for the SVR.

Let the partial derivative of Equation (3) be 0 and introduce the solution back into Equation (3); the dual SVR algorithm problem can be expressed by using Equation (6).

$$\max_{\boldsymbol{\alpha},\boldsymbol{\alpha}^*}\left(\sum_{i=1}^{n} y_i(a_i^* - a_i) - \epsilon(a_i^* + a_i) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(a_i^* - a_i)(a_j^* - a_j)\mathbf{x}_i^T\mathbf{x}_j\right) \tag{4}$$

$$\text{s.t}\begin{cases}\sum_{i=1}^{n}(a_i^* - a_i) = 0\\ 0 \leq a_i, a_i^* \leq C\end{cases} \tag{5}$$

The dual problem should satisfy the Karush–Kuhn–Tucker condition as follows (Ghimire et al., 2019):

$$\text{s.t}\begin{cases}a_i(f(\mathbf{x}_i) - y_i - \epsilon - \xi_i) = 0\\ a_i^*(y_i - f(\mathbf{x}_i) - \epsilon - \xi_i^*) = 0\\ (C - a_i)\xi_i = 0, (C - a_i^*)\xi_i^* = 0\\ a_i a_i^* = 0\end{cases} \tag{6}$$

Finally, the SVR solution can be expressed using Equation (7),

$$f(\mathbf{x}) = \sum_{i=1}^{n}(a_i^* - a_i)\mathbf{x}_i^T\mathbf{x} + b. \tag{7}$$

The inner product $\mathbf{x}_i^T\mathbf{x}$ can be replaced by the so-called kernel function $K(x_i, x)$ under Mercer's condition. Therefore, the final form of SVR function can be expressed using Equation (8).

$$f(\mathbf{x}) = \sum_{i=1}^{n}(a_i^* - a_i)K(\mathbf{x}_i, \mathbf{x}) + b. \tag{8}$$

The kernel function, as can be seen from Equation (8), plays a critical role in the SVR algorithm. In SVR model, polynomial, sigmoid, linear and radial basis function (RBF) can be used as kernel function (Ramedani et al., 2014). In this study, RBF kernel function was chosen because of it's *a)* capability of modelling nonlinear relationships by mapping data points from the input space into high dimensional feature space in a nonlinear fashion, *b)* Compared to polynomial and sigmoid kernels, RBF needs less customisable parameters, making it straightforward and functional and *c)* RBF's superior performance has been demonstrated in several literature (Halde, 2016; Ghimire et al., 2023a).

The kernel function RBF is expressed as Equation (9):

$$K(\mathbf{x}_i, \mathbf{x}) = e^{-\frac{-\|x_i - x\|^2}{2\sigma^2}} \tag{9}$$

where σ is variance and $\|\mathbf{x}_i - \mathbf{x}\|$ is the Euclidean distance (L_2-norm) between two points $\mathbf{x}_i$ and $\mathbf{x}$.

$$K(\mathbf{x}_i, \mathbf{x}) = e^{-\gamma\|\mathbf{x}_i - \mathbf{x}\|^2} \tag{10}$$

The RBF Kernel Support Vector Machines has two hyperparameters associated with it, C for SVR and γ for the RBF Kernel. Here, γ is inversely proportional to σ and can be expressed as below (Ghimire et al., 2023b).

$$\gamma = \frac{1}{2\sigma^2} \tag{11}$$

4.2. Benchmark models

4.2.1. Neural network-based methods

To fully appraise the performance of the proposed SVR model for prediction of TPP7155 overall scores, this study utilises three neural network methodologies: the Multilayer Perceptron (MLP) (Deo et al., 2018), Extreme Learning Machine (ELM) (Deo et al., 2020) and Deep Neural Network (DNN) (Zhang et al., 2016).

The MLP model is a fundamental architecture with an input layer, one or more hidden layers, and an output layer, facilitating the acquisition of intricate patterns through backpropagation. This architecture is widely applicable across various tasks, including regression and classification (Borghi et al., 2021). ELM distinguishes itself with its swift training pace and commendable generalisation performance, achieved by randomly assigning weights between input and hidden layers, keeping these weights fixed, and analytically deriving output weights, which makes it particularly suited for scenarios requiring rapid training (Ghimire et al., 2022). DNNs, which extend the MLP framework by integrating multiple hidden layers, enable the acquisition of more nuanced hierarchical data representations, thereby excelling in intricate tasks like image and speech recognition, natural language processing, and various time-series prediction tasks (Elsayed et al., 2021). While MLPs offer simplicity and ELMs provide speed advantages, DNNs generally outperform both in handling high-dimensional data and addressing complex problems due to their deep architecture and sophisticated training algorithms (Kouassi & Moodley, 2020).

4.2.2. Tree-based methods

The second benchmark model comprises of three tree-based regression models: Decision Trees Regression (DTR) (Jayasinghe et al., 2022), Random Forests Regression (RFR) (Deo et al., 2020), and Extra Trees Regression (ETR) (Li et al., 2020). Decision Trees construct tree-like structures where each node corresponds to a feature and each leaf node signifies an outcome, providing interpretability yet results in susceptibility to overfitting (Malakouti, 2023). Random Forests, composed of decision trees trained on random data subsets, mitigate variance by averaging predictions and yield insights into feature importance, benefiting from parallelisation during training (Li et al., 2020). Extra Trees introduce added randomness by selecting random split thresholds, aiming to further diminish variance and potentially enhance generalisation and computational efficiency compared to Random Forests, albeit at the expense of some interpretability (Junaid et al., 2023).

4.2.3. Ensemble-based methods

The third benchmark model involves two ensemble models: Light Gradient Boosting Machine (LGB) (Xu et al., 2021) and Extreme Gradient Boosting Machine (XGB) (Wahyuningsih et al., 2024). These methods, LGB and XGB, are formidable ensemble learning techniques acclaimed for their prowess. LGB, a creation of Microsoft, stands out for its efficiency and speed, especially adept with sizeable datasets, employing histogram-based algorithms and parallelisation. In contrast, XGB, developed by Chen and Guestrin (2016), garners recognition for its precision and competitive edge in predictive modelling competitions, employing a depth-wise tree growth strategy and regularisation methods to counter overfitting (Chen et al., 2015). Both approaches are extensively utilised to attain superior outcomes in predictive modelling endeavours, owing to their efficacy, adaptability, scalability, and resilience.

4.2.4. Boosting-based methods

The Support Vector Regression (SVR) model is further compared against boosting-based models: Adaboost Regression (ADBR) (Liu & Bai, 2023), Gradient Boosting Regression (GBR) (Tekgöz et al., 2022), and Bagging Regression (BGR) (Evangelista, 2023). Adaboost iteratively merges weak learners into a robust learner by targeting previously misclassified instances. Gradient Boosting constructs a strong learner by progressively rectifying errors using decision trees, thereby minimising

Table 1
Descriptive statistics of TPP7155 (General Science) student performance 2020–2021 data. The predictors (inputs) are: W1 = Assignment 1, W2 = Assignment 2, W3 = Assignment 3,W4 = Assignment 4, W5 = Assignment 5 and ET = Examination Score with the target OT = Overall mark. The cross-correlation coefficient (r) of each predictor with OT is shown in the last row. Note that a raw mark for each assessment has a different total with a particular percentage contribution towards the final grade.

Statistica properties	Predictors						Target
	Assignment 1	Assignment 2	Assignment 3	Assignment 4	Assignment 5	Exam	Overall Mark
	5% W1	15% W2	10% W3	20% W4	5% W5	45% ET	OT
Mean	8.086	25.759	25.44	32.661	10.842	45.772	79.701
Median	8.5	26.5	26	34	11.5	46.5	81
Standard Deviation	1.559	3.47	3.317	5.799	3.331	7.997	10.463
Minimum	0	8.5	7	7.5	0	17	37
Maximum	10	30	30	40	15	60	98
Skewness	-1.284	-1.231	-1.301	-1.049	-1.015	-0.559	-0.776
Flatness	2.319	1.913	3.042	1.26	0.702	0.097	0.825
Correlation with Target	0.593	0.663	0.625	0.721	0.666	0.886	

the overall loss function (Sharafati et al., 2020). Bagging trains multiple models on diverse bootstrap samples and aggregates their predictions to diminish variance and enhance model stability and generalisation (de Oliveira et al., 2022). These methodologies augment predictive accuracy by harnessing the capabilities of multiple learners.

5. Project design context and SVR model performance criteria

5.1. Student course data and ethical procedure

This research paper reports the findings of a project: *"Artificial intelligence as predictive analytics framework for learning outcomes, assessment and student success: UinSQ Technology Demonstrator"*. This study aims to design and evaluate a Support Vector Regression (SVR) (with several competing benchmark) models to predict student success in the TPP7155 (General Science) Course.

To construct an SVR model, the dataset consisted of continuous assessments and weighted scores of students for the period 2020 to 2021. Following a rigorous data-cleansing phase that removed any rows with missing values for student assessments, the TPP7155 dataset used in model design included 492 student records (from a pool of 727). Student performance was evaluated through a combination of five assignments, Assignment 1 (weight = 5%), Assignment 2 (weight = 15%), Assignment 3 (weight = 10%), Assignment 4 (weight = 20%), and Assignment 5 (weighted = 5%), alongside the final examination score (weight = 45%). These assessments collectively contributed to an overall score (OT) expressed as a percentage. This OT score was then utilised to determine a pass or fail grade for each student.

There was no direct recruitment of the participants in this study but instead, they were drawn from a TPP7155 course, which provides a strong foundation in General Science. This course is valuable for students' pursuing careers in Science, Technology, Engineering, and Mathematics (STEM) as well as for those intending to study early childhood and primary education. Through structured learning opportunities, students in this course develop a broad understanding of scientific topics, building analytical and problem-solving skills that boost confidence in introductory sciences and enable informed decisions about further undergraduate studies. Using self-paced instruction and principles of adult learning, participants are guided through a carefully sequenced series of topics designed to develop the scientific and mathematical literacy required for undergraduate education. Students study scientific methods, measurement, cellular biology, chemistry, climate change, and sustainable energy. Throughout the course, they are taught to interpret, analyse, and evaluate scientific data and to communicate findings effectively. In this way, they engage with scientific thought processes and content that remain relevant both now and in the future.

Ethics approval for this study [H18REA236] was granted by the University of Southern Queensland Human Ethics Committee. Given the low-risk nature of the project, expedited ethical approval was granted, and all students details were anonymised before processing performance data. Therefore, this research did not collect or use student's personally identifiable information, such as names, student identification numbers, gender, or socioeconomic status. To prevent bias in the proposed SVR model, any incomplete data records (e.g., students who did not submit assessments or did not take the examination) was removed during data pre-processing. Despite the exclusion of some records, we retained complete data for 498 students, including all assignment and examination scores, ensuring minimal impact on the SVR model's ability to predict overall scores. As a result, only records containing all relevant data points for each student, along with an OT value, were included in the SVR model's training phase.

5.2. Predictive model development

In Table 1, we show an overview of the statistical properties of course assessments W1, W2, W3, W4, W5, ET (Examination Mark) and the OT (Overall Score) in TPP7155. Additionally, the last row of Table 1 showcases the Pearson's Correlation Coefficient (r) computed between each assignment W1, W2, W3, W4 and W5 and the examination (ET) against the OT. Details of these data reveal that all assignments and examination marks exhibited negative skewness (Skewness ≤ 0) which indicates a particular distribution where the tail on the left side is somewhat longer or fatter than the right side. Moreover, in terms of kurtosis (or flatness) factor, the distribution of assignments W1, W2, W3 and W4 is leptokurtic, implying heavier tails and sharper peaks compared with a normal distribution. Conversely, the distribution of W5 and ET is slightly platykurtic and mesokurtic, respectively, which indicates lighter tails and flatter peaks compared with a normal distribution.

A closer examination shows that the degree of association between assignments, examination marks and the OT demonstrates significant variability, which could also indicate non-linearity between these variables that determine overall student scores. Notably, there is a positive correlation between all continuous assessment marks and the OT value; however, the strength of this correlation with ET is notably higher with $r \approx 0.886$ followed that of W4 ($r \approx 0.721$), W5 ($r \approx 0.666$), W2 ($r \approx 0.663$), W3 ($r \approx 0.625$) and W1 ($r \approx 0.593$). This relationship can be further validated by the Kendall's correlation heatmap shown in Fig. 2 where the ET value emerges to be highly positively correlated with the OT followed by W4, W5, W2, W3 and W1.

In terms of physical interpretation, this signifies that an increase in any assignment or examination marks tends to coincide with an increase in the OT, while a decrease in one tends to result in a corresponding decrease in the OT. This variance in correlation strength underscores the varying impacts of each assessment component on the overall student outcome and it also underscores the differing impacts of each assessment component on the overall score. Using the correlation plot as a

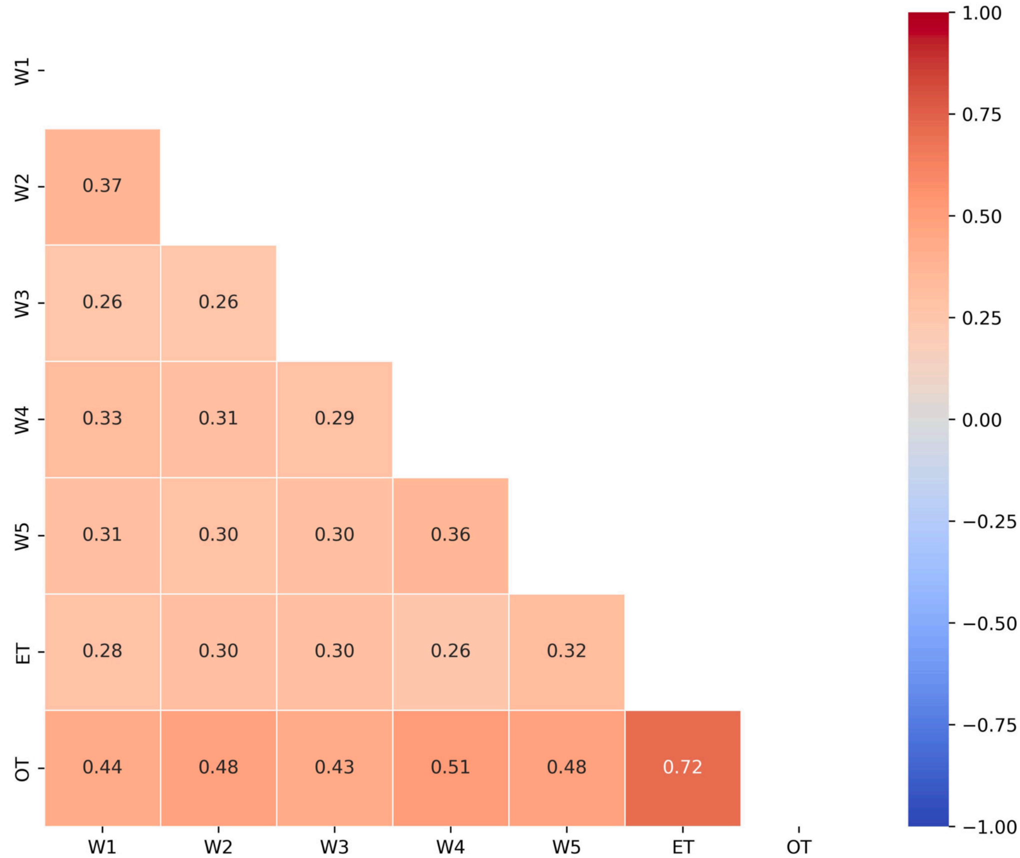

Fig. 2. The Kendall's correlation heat map used to explore the student performance data in TPP7155 (General Science).

Table 2
The input combinations based on TPP7155 (General Science) course student performance data used to construct the proposed SVR model.

Designated model	Input combinations (using predictors from Table 1)	Period	Total data	Training (64%)	Validation (16%)	Testing (20%)
M1	OT = f(W1)					
M2	OT = f(W1,W2)					
M3	OT = f(W1,W2,W3)					
M4	OT = f(W1,W2,W3,W4)					
M5	OT = f(W1,W2,W3,W4,W5)					
M6	OT = f(W1,W2,W3,W4,W5, ET)	2020-2021	492	314	79	99
M7	OT = f(ET)					
M8	OT = f(ET,W1)					
M9	OT = f(ET,W1,W2)					
M10	OT = f(ET,W1,W2,W3)					
M11	OT = f(ET,W1,W2,W3,W4)					

guiding principle, we structured the input sequences of the proposed SVR model to clearly capture the progressively increasing significance of each predictor variable. Subsequently, we conducted a comprehensive analysis of each predictor's individual contribution to the prediction of the overall score. To achieve this objective, we introduced two distinct categories of predictive modelling schemes within this study.

The first modelling framework developed models that featured only W1 as an input variable. Subsequent iterations of the proposed SVR model incorporated W1, W2, W3, W5, and ET sequentially, resulting in the creation of six distinct SVR models designated as Models M1, M2, M3, M4, M5, and M6. In contrast, the second model framework, designated as Models M7–M11, placed exclusive emphasis on Examination Score (ET) as a primary predictor variable with subsequent addition of each assignment mark to determine the influence of this sequence of inputs on the SVR model. This category began with ET as the sole input for Model M7 (i.e., $OT = f(ET)$) gradually incorporating additional predictors W1, W2, W3, and W4 into successive SVR models (i.e., $OT = f(ET, W1)$, $OT = f(ET, W1, W2)$, $OT = f(ET, W1, W2, W3)$, and $OT = f(ET, W1, W2, W3, W4)$). The specific configurations of each SVR model and the respective input combinations are detailed in Table 2. This systematic modelling approach aims to elucidate the nuanced relationships between the predictor variables and the OT, ultimately enhancing both the predictive accuracy and interpretability of the developed SVR and benchmark models.

To construct the proposed SVR and the benchmark models, an initial step involves normalising all original data. This normalisation alleviates the influence of different data magnitudes, ensuring that each variable contributes to the model training process proportionately (Ghimire et al., 2023b). Following normalisation, the sample data were split into an 80:20 ratio, with 80% allocated to training and validation of the SVR model, and the remaining 20% reserved for testing. Within the training subset, an additional partition set aside 20% specifically for validation. Given that the total dataset comprised 492 observations, 314 data points were designated for model training, 79 data points for validation, and 99 instances were earmarked as an independent test set to cross-validate the performance of the proposed SVR and all benchmark models.

For clarity, a normalised mathematical representation is delineated in Equation (12), where 'x' symbolises the original value of the variable, x_{norm} is the normalised value of x, $\max(x)$ and $\min(x)$ denote the highest and lowest values within the dataset, respectively.

$$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{12}$$

where x is the vector of values to be scaled, x_{scaled} is the normalised value of x, $\max(x)$ and $\min(x)$ are the maximum and minimum values of that vector, respectively.

Following normalisation, data were fed into various machine learning models. Optimising these model's hyperparameters is crucial for enhancing their performance, thus mitigating over-fitting and underfitting, improving generalisation capability and reducing computational time. To address these critical objectives, the present study employs Hyperopt, which is a Python-based tool as the hyperparameter optimisation (Hyperopt) tool (Komer et al., 2019). Hyperopt has a sophisticated framework for systematically tuning hyperparameters, ensuring that the models are finely calibrated to yield optimal results across diverse datasets and scenarios. By leveraging Hyperopt, this study aims to develop robust and efficient machine learning models capable of effectively generalising the unseen data while conserving computational resources (Bergstra et al., 2015).

For an outline of the hyperparameters and search range of hyperparameters for the various models, including the objective model (i.e., SVR) as well as neural network-based models, tree-based models, ensemble-based models and the boosting-based models, readers should refer to Table A.9 and Table A.10. The model training was undertaken on a high-performance computer with 32 GB of RAM and an Intel Core *i7* processor. All models were meticulously constructed using Keras framework version 2.2.4 (Ketkar & Ketkar, 2017), tapping into its wide range of functionalities for neural network development. The TensorFlow backend version 1.13.1 (TensorFlow, 2018) seamlessly integrated, leveraging its powerful computational capabilities for efficient model training. These frameworks collectively formed a robust environment, supporting various machine learning tasks. The entire model development process occurred within a Python 3.6 environment, leveraging the language's versatility and extensive library ecosystem (Table 3).

5.3. Performance evaluation

In this study, a meticulous comparison of the proposed SVR model against a suite of benchmark models is made to predict OT by employing a judicious blend of visual and comprehensive descriptive statistics of model performance. The model evaluation framework categorises these metrics into two distinct classes: Class A, where an ideal value should be 1 and Class B, where the ideal value should be 0. Class A metrics, including the Coefficient of Determination (R^2), Nash–Sutcliffe Efficiency (*NS*), Willmott's Index of Agreement (*WI*), Kling-Gupta Efficiency (*KGE*) and Legates and McCabe Index (*LM*), are utilised to evaluate the goodness-of-fit and predictive accuracy of the model.

Through these metrics, we aim to gain insights into the model's ability to capture the underlying patterns and variability in data, thereby facilitating informed decision-making. Conversely, Class B metrics focus on quantifying the disparities between predicted and actual values, providing valuable insights into the model's predictive precision. These metrics encompass the Mean Absolute Error (*MAE*), Root Mean Square Error (*RMSE*), and Absolute Percentage Bias (*APB*; %) Relative Mean Absolute Error (*RMAE*; %), and Relative Root Mean Square Error (*RRMSE*; %). The mathematical formulation of these metrics is shown below (Ghimire et al., 2023b):

$$R^2 = \frac{\sum_{i=1}^{n}(OT^a - \langle OT^a\rangle)(OT^p - \langle OT^p\rangle)}{\sqrt{\sum_{i=1}^{n}(OT^a - \langle OT^a\rangle)^2}\sqrt{\sum_{i=1}^{n}(OT^p - \langle OT^p\rangle)^2}} \tag{13}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(OT^p - OT^a)^2} \tag{14}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|OT^p - OT^a| \tag{15}$$

$$RRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(OT^p - OT^a)^2}}{\langle OT^a\rangle} \tag{16}$$

$$RMAE = \frac{1}{n}\sum_{i=1}^{n}\frac{|OT^p - OT^a|}{OT^p} \tag{17}$$

$$WI = 1 - \frac{\sum_{i=n}^{n}(OT^a - OT^p)^2}{\sum_{i=n}^{n}(|OT^p - \langle OT^a\rangle| + |OT^a - \langle OT^a\rangle|)^2} \tag{18}$$

$$NS = 1 - \frac{\sum_{i=1}^{n}(OT^a - OT^p)^2}{\sum_{i=1}^{n}(OT^a - \langle OT^a\rangle)^2} \tag{19}$$

$$LM = 1 - \frac{\sum_{i=1}^{n}|OT^a - OT^p|}{\sum_{i=1}^{n}|OT^a - \langle OT^a\rangle|} \tag{20}$$

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\langle OT^p\rangle}{\langle OT^a\rangle} - 1\right)^2 + \left(\frac{CV_p}{CV_a}\right)^2} \tag{21}$$

$$APB = \frac{\sum_{i=1}^{n}(OT^a - OT^p) * 100)}{\sum_{i=1}^{n} OT^a}, \tag{22}$$

where r is the correlation coefficient, CV is the coefficient of variation, OT^p refers to the predicted OT, OT^a is the actual OT, $\langle OT^a\rangle$ is the average value of the OT^a, $\langle OT^p\rangle$ is the average value of the OT^p and finally n is the number of actual values.

This study also adopts the Promoting Percentage (Deo et al., 2022) of KGE (Δ_{KGE}), and WI (Δ_{WI}) to further compare the various models completing models to validate the efficacy of the proposed SVR model used in OT prediction.

$$\Delta_{KGE} = \left|\frac{KGE_1 - KGE_2}{KGE_1}\right| \tag{23}$$

$$\Delta_{WI} = \left|\frac{WI_1 - WI_2}{WI_1}\right| \tag{24}$$

where, KGE_1, and WI_1 refers to the objective model (*i.e.*, SVR) performance metrics and KGE_2, and WI_2 refers to the benchmark model performance metrics.

5.4. Explainability of the proposed SVR model

In earlier studies on student performance predictions in the USQ Technology Demonstrator project (Deo et al., 2020; Nguyen-Huy et al., 2022; Ahmed et al., 2022), explainability of the machine learning model was not explored. Therefore in this study, we have adopted SVR model explainability techniques: SHAP (SHapley Additive exPlanations) (Ramon et al., 2020; Ghimire et al., 2024) and LIME (Local Interpretable Model-agnostic Explanations) (Dieber & Kirrane, 2020) to provide valuable insights into the inner workings of the complex machine learning model.

The SHAP method offers a global understanding of the importance of features by attributing the contribution of each feature to the proposed SVR model's output. This method can facilitate a better understanding of the SVR model's overall behaviour (Man & Chan, 2021). In contrast, the LIME method offers local interpretability of the SVR model by approximating it's decision boundaries around specific tested data instances, thus providing human-understandable explanations for the individual model predictions. The present study leverages both the

Table 3
The architecture of the Ensemble-based and the Boosting-based models developed to predict TPP7155 (General Science) course students' overall scores.

	Model	Hyperparameter	Seach range	Optimal value			
				M1	M4	M6	M11
Ensemble-Based Methods	**Extreme Gradient Boosting (XGB)**	Number of Estimators	hp.choice('n_estimators', range(50,500,2)),	167.000	173.000	214.000	149.000
		Learning rate	hp.uniform('learning_rate', 0.01, 0.1),	0.085	0.034	0.050	0.059
		Maximum Depth	hp.choice('max_depth', range(2,10,1)),	4.000	6.000	7.000	5.000
		min_child_ weight	hp.choice('min_child_weight', range(1,50,1)),	23.000	10.000	12.000	4.000
		subsample	hp.uniform('subsample',0.5, 1.0),	0.582	0.750	0.604	0.500
		colsample_bytree	hp.uniform('colsample_ bytree', 0.6, 1.0)	0.641	0.626	0.811	0.762
		L2 regularization term	hp.uniform('reg_alpha',;0, 1.0),	0.643	0.116	0.649	0.360
	Light Gradient Boosting (LGB)	Number of Estimators	hp.choice('n_estimators', range(50,500,2)),	94.000	38.000	209.000	142.000
		min_child_ weight	hp.uniform('min_child_ weight', 0.001, 0.2),	0.032	0.163	0.103	0.187
		min_child_ samples	hp.choice('min_child_samples', range(5,51,5)),	1.000	1.000	2.000	1.000
		lgb_colsample_ bytree	hp.uniform('lgb_colsample_ bytree', 0.6;1.0),	0.960	0.601	0.743	0.818
		subsample	hp.uniform('subsample';0.5, 1.0),	0.832	0.999	0.988	0.730
		Learning rate	hp.uniform('learning_rate', 0.01, 0.3),	0.050	0.066	0.107	0.157
		Maximum Depth	hp.choice('max_depth', range(2,10,1)),	5.000	2.000	7.000	5.000
		number of Leaves	hp.choice('num_leaves', range(2, 50, 1)),	0.000	42.000	3.000	30.000
		L2 regularization term	hp.uniform('reg_alpha'0, 1.0)	0.643	0.765	0.297	0.815
Boosting-Based Method	**AdaBoost Regressor (ADBR)**	Learning rate	hp.uniform('learning_rate', 0.0001, 0.3),	0.105	0.166	0.235	0.195
		Loss	hp.choice ('loss', ['linear', 'square' ,'exponential']),	square	square	exponential	linear
		Number of Estimators	hp.choice('n_estimators', range(5,800,2))	10.000	375.000	92.000	20.000
	Bagging Regressor (BGR)	Number of Estimators	hp.choice('n_estimators', range(40, 800,20))	240.000	160.000	60.000	80.000
		bootstrap	True	True	True	True	True
		Estimator	Decision Tree Regressor	DTR	DTR	DTR	DTR
	Gradient Boosting Regressor(GBR)	Number of Estimators	hp.choice('n_estimators', range(5,800,2)),	71.000	347.000	713.000	351.000
		Learning rate	hp.uniform('learning_rate', 0.0001, 0.3),	0.045	0.060	0.089	0.222
		Maximum Depth	hp.choice('max_depth', range(1,110,1))	1.000	4.000	3.000	1.000

SHAP and the LIME approaches to achieve a comprehensive understanding of the proposed SVR model predictions, which facilitates informed decision-making and enhances the trust in such AI systems used within the education industry (Muhamedyev et al., 2020).

In summary, the SHAP-based Kernel Explainer is employed for global explanation, highlighting the effects of the respective predictors (i.e., W1, W2, W3, W4, W5, ET) on the entire model performance in estimating OT through the SHAP summary and feature importance (i.e., visual) plots. Subsequently, the LIME approach explains every instance of the tested dataset, providing local explanations of the model. For discussion purposes, however, only the explanations for the six instances are presented including the 1st, 28th, 57th, 72nd, 85th, and 99th instances of the test datasets.

6. Results and comparative evaluation of SVR model performance

In this section, the results generated by the proposed SVR model, along with comparative counterpart models, are presented in respect to their capability to predict overall course score (OT) for grading purposes in TPP7155 (General Science) course taught in the Tertiary Preparatory Program at the University of Southern Queensland, Australia.

6.1. Model performance with single assignment marks

We now compare the proposed SVR model with four machine learning models in the category of Neural Network-Based (ELM), Tree-based (ETR), Ensemble-based (XGB) and Boosting-based (GBR) methods. This comparison aims to evaluate the models' accuracy in predicting OT using any single assignment (i.e., Assignment 1: W1, Assignment 2: W2, Assignment 3: W3, Assignment 4: W4 or Assignment 5: W5), or the Examination Mark (ET) as predictor variables. The model validation is then conducted by utilising R^2, *RMSE*, *APB*, and *KGE* metrics, which are shown in Table 1.

The results indicate a superior predictive capacity of the proposed SVR model compared with ELM, XGB, ETR and GBR models when single input variables are employed. Specifically, when W1 data are utilised to predict OT, we note that the SVR model exhibits the highest R^2 and the *KGE* values whereas it attains the lowest *RMSE* and *APB* values. These values for SVR model are $RMSE \approx 8.061$, $APB \approx 8.215$, $R^2 \approx 0.467$ and $KGE \approx 0.483$. Regarding the overall contributory influence of the assignments and the examination marks in predicting OT for TPP7155 course, the proposed SVR model demonstrates that W5 results in the most significant contribution ($RMSE \approx 7.7833$, $APB \approx 7.642$, $R^2 \approx 0.503$, and $KGE \approx 0.363$) with the smallest contribution originat-

Table 4
Influence of each predictor variable used to predict the overall score (OT) using the proposed SVR model *vs.* ELM, XGB, ETR and GBR for TPP7155 General Science course in the model's testing phase. The optimal model is blue and boldfaced. [W1 = assignment 1; W2 = assignment 2; W3 = assignment 3; W4 = assignment 4; W5 = assignment 5; ET = exam score.

Predictor variable	Model	*RMSE*	R^2	*APB*	*KGE*
W1	**SVR**	**8.061**	**0.467**	**8.215**	**0.483**
	ELM	8.393	0.422	8.753	-0.019
	XGB	8.532	0.403	8.793	0.130
	ETR	8.658	0.385	9.104	-0.211
	GBR	8.662	0.384	9.041	-0.014
W2	**SVR**	**7.837**	**0.496**	**8.459**	**0.712**
	ELM	24.284	-3.839	12.202	-0.245
	XGB	8.245	0.442	9.036	0.254
	ETR	8.514	0.405	9.418	0.107
	GBR	8.363	0.426	9.213	0.371
W3	**SVR**	**8.118**	**0.459**	**8.620**	**0.434**
	ELM	8.737	0.374	9.561	0.369
	XGB	9.125	0.317	9.890	-0.601
	ETR	9.038	0.330	9.914	-0.147
	GBR	8.744	0.373	9.555	0.257
W4	**SVR**	**7.389**	**0.552**	**6.988**	**0.213**
	ELM	7.550	0.532	7.014	0.255
	XGB	7.451	0.544	7.071	0.285
	ETR	7.503	0.538	7.087	0.177
	GBR	7.470	0.542	7.090	0.184
W5	**SVR**	**7.783**	**0.503**	**7.642**	**0.363**
	ELM	7.969	0.479	8.039	0.275
	XGB	8.145	0.456	8.336	0.057
	ETR	8.273	0.438	8.446	0.025
	GBR	8.109	0.461	8.067	0.315
ET	**SVR**	**4.588**	**0.827**	**4.535**	**0.913**
	ELM	4.846	0.807	4.878	0.715
	XGB	4.908	0.802	5.101	0.656
	ETR	5.136	0.784	5.208	0.648
	GBR	5.004	0.795	5.151	0.763

ing from W3 ($RMSE \approx 8.118$, $APB \approx 8.620$, $R^2 \approx 0.459$, and $KGE \approx 0.434$).

6.2. Performance with examination mark as predictor

The present results demonstrate that the ET value remains the most significant contributor towards predicting OT, yielding an $RMSE \approx 4.588$, $APB \approx 4.535$, $R^2 \approx 0.827$, and $KGE \approx 0.913$. To provide deeper insights, the individual interaction of examination scores on assignment marks has been analysed using partial dependence plots (PDP). The PDP illustrate how the predicted outcome of an SVR model changes as a function of the two input features while averaging out the effects of all other features, as shown in Fig. 3.

It is noteworthy that the PDP plots show the causal relationships between ET and the five different assignment scores (W1 to W5) with respect to the model's predictions for OT. Here, the $x - axis$ represents the values of ET whereas the $y - axis$ represents the individual assignment mark and the $z - axis$ or the Partial Dependence shows the predicted outcome (i.e., OT) from the proposed SVR model. In terms of its physical explanation, the upward slope in PDP plots indicates a positive relationship between the input features and the predicted outcome from an SVR model. It is not surprising to state that the higher examination scores and assignment marks are associated with a higher predicted OT while a lower mark corresponds to a lower OT (Table 4).

6.3. Effect of combined inputs on predictive accuracy

After analysing the interactions of individual assignment marks and examination scores on the OT predictions, we now evaluate the proposed SVR model's performance based on input combinations listed in Table 2, designated as Model M1 to M11. In this analysis, the individual predictor variables such as the assignment marks or the examination scores have purposely been incrementally added to the proposed SVR model to predict the overall score.

Fig. 4 is a comparison of the actual and the predicted OT, represented in terms of the coefficient of determination (R^2) for different input combinations for Models M1 to M11. Evidently, the proposed SVR model (specifically designated as M6 with input features of: W1, W2, W3, W4, W5 and ET, as well as M11 with input features of: ET, W1, W2, W3 and W4) exhibits the highest accuracy of all the machine learning models. Notably, the performance of the proposed SVR models designated as Model **M6 and M11** significantly surpasses that of the Neural Network-based, the Tree-based, Ensemble-based and the Boosting-based models in terms of the R^2 values between the actual and predicted overall scores.

To explore this result further, let us consider R^2 value for the proposed SVR model with Model M6 input combination, which registered an $R^2 \approx 0.999$ compared with ≈ 0.467, ≈ 0.646, ≈ 0.715, ≈ 0.809, ≈ 0.861, ≈ 0.827, ≈ 0.877, ≈ 0.930, ≈ 0.945, and ≈ 0.993 for Models M1, M2, M3, M4, M5, M7, M8, M9, M10, and M11, respectively. It is therefore evident that with the addition of each assignment mark as an input variable (in Models M1 to M2 and so forth), the R^2 value increases steadily reaching its maximum at Model M6 input combination. Furthermore, the comparison models (Neural Network-based) i.e., ELM and DNN with M6 input combination achieved an R^2 value of ≈ 0.999 followed by the comparison model MLP (≈ 0.994), Boosting-based Model (GBR; ≈ 0.992), Ensemble-based model (XGB; ≈ 0.992), Tree-based model (ETR; ≈ 0.979), and the BGR model (≈ 0.878).

Fig. 5 illustrates the *RMSE* plot comparing the proposed SVR model with the Neural Network-based, Tree-based, Ensemble-based, and Boosting-based models. Evidently, all models with M6 input combination exhibit the lowest *RMSE*. Among these, the proposed SVR model recorded the lowest *RMSE* magnitude with M6 input combination. For instance, when comparing Tree-based model with SVR for the M6 input combination, the *RMSE* value for SVR was ≈ 0.305 whereas the models DTR, ETR and RFR (Tree-based Models) had *RMSE* values of ≈ 2.699, ≈ 1.961, and ≈ 1.588, respectively. Similarly, when compared with the Ensemble-based model, the *RMSE* of ADBR, GBR, and BGR are ≈ 3.051, ≈ 0.987, and ≈ 3.861, respectively. Additionally, for the Boosting-based model, the *RMSE* is ≈ 1.391 and ≈ 0.994 for LGB and XGB, respectively. The Neural Network-based model also generated low *RMSE* values of ≈ 0.401, ≈ 0.395, and ≈ 0.862 for ELM, DNN, and MLP, respectively. Similarly, when comparing the SVR model with different input combinations, the *RMSE* values were ≈ 8.061, ≈ 6.569, ≈ 5.899, ≈ 4.824, ≈ 4.112, ≈ 0.305, ≈ 4.588, ≈ 3.866, ≈ 2.928, ≈ 2.588, and ≈ 0.952 for M1 through M11, respectively. Notably, the *RMSE* value decreased by 332% from M5 ($OT = f(W1, W2, W3, W4, W5)$); *RMSE* = 4.112) to M11 ($OT = f(ET, W1, W2, W3, W4)$); *RMSE* ≈ 0.952) input combination for SVR. Similarly, from M5 (with all assignment marks) to the M6 ($OT = f(W1, W2, W3, W4, W5, ET)$) input combination, the *RMSE* decreased by $\approx 93\%$.

In accordance with these findings, we conclude that the inclusion of the examination mark plays a vital role in predicting OT in TPP7155. Based on R^2 and *RMSE* values, the proposed SVR model with the M6 input combination exhibited the highest accuracy in predicting the overall score.

6.4. Error analysis: RMSE and RRMSE

Table 5 shows the performance of all models for input combinations M1 to M11 using Relative *RMSE* (*RRMSE*). Notably, the proposed SVR model has achieved the lowest *RRMSE* of 0.37% for the input combination M6 compared with 9.91%, 8.20%, 7.36%, 5.87%, 5.04%, 5.67%, 4.76%, 3.63%, 3.21%, and 1.16% for input combinations M1, M2, M3, M4, M5, M7, M8, M9, M10, and M11, respectively. When considering only the M6 input combination, the BGR model performed the worst with the highest *RRMSE*(4.72%) followed by ADBR (3.79%) and DTR (3.29%). Additionally, the inclusion of the examination mark (ET) as

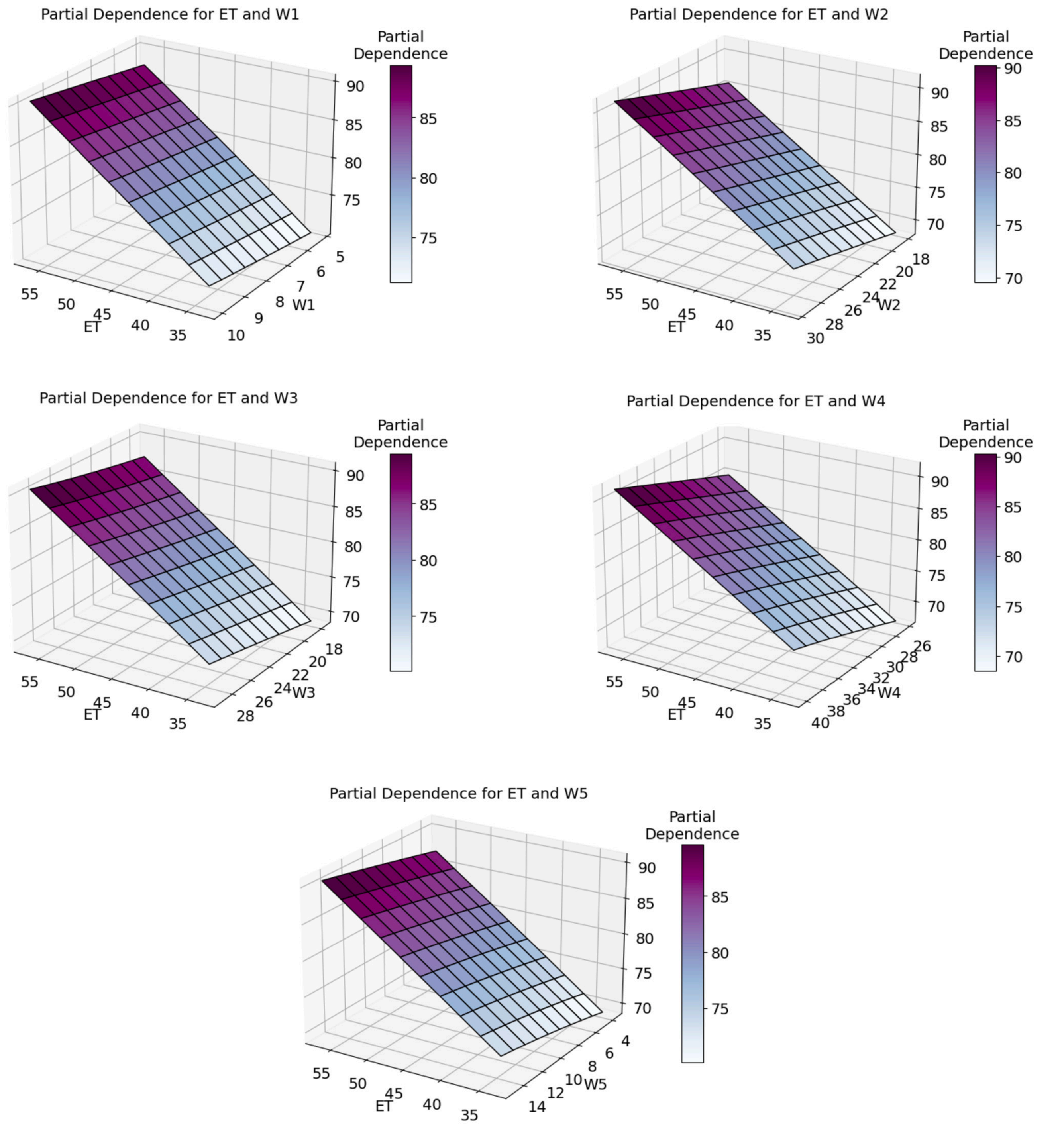

Fig. 3. Interactive effects of the examination (ET) and assignment marks (W1, W2, W3, W4, W5) features on the proposed SVR model prediction of overall scores (OT) in TPP7155 (General Science) Course.

an input along with assignment marks W1, W2, W3, W4, and W5 in Model M6 significantly improved the *RRMSE* by 95% compared to M5 ($OT = f(W1, W2, W3, W4, W5)$). When M6 is compared with M7 (OT = f(ET)), the *RRMSE* shows an increase of 1420% (Fig. 6(e).

It is therefore conclusive that including the assignment marks results in a gradual decrease in *RRMSE* values: M8 ($OT = f(ET, W1)$) shows a 1176% increase, M9 ($OT = f(ET, W1, W2)$) an 870% increase, M10 ($OT = f(ET, W1, W2, W3)$) a 760% increase, and M11 ($OT = f(ET, W1, W2, W3, W4)$) a 210% increase. In summary, considering the *RMSE* and R^2 values, the low *RRMSE* for M6 indicates that the examination score remains the most significant predictor of overall scores. However, the contributions of assignments are also important and should not be overlooked.

6.5. Evaluation using MAE and RMAE

Table 6 shows *MAE* and *RMAE* to evaluate the predictive performance of all machine learning models where *MAE* is seen to offer a clear indication of the average magnitude of the model errors. Unlike the $RMSE$, the *MAE* values do not square the errors, meaning that it does not excessively penalise the larger errors, thus making it more robust to the outliers. Conversely, the *RMAE* normalises the error by a scale of the actual values, which facilitates easier comparison of the

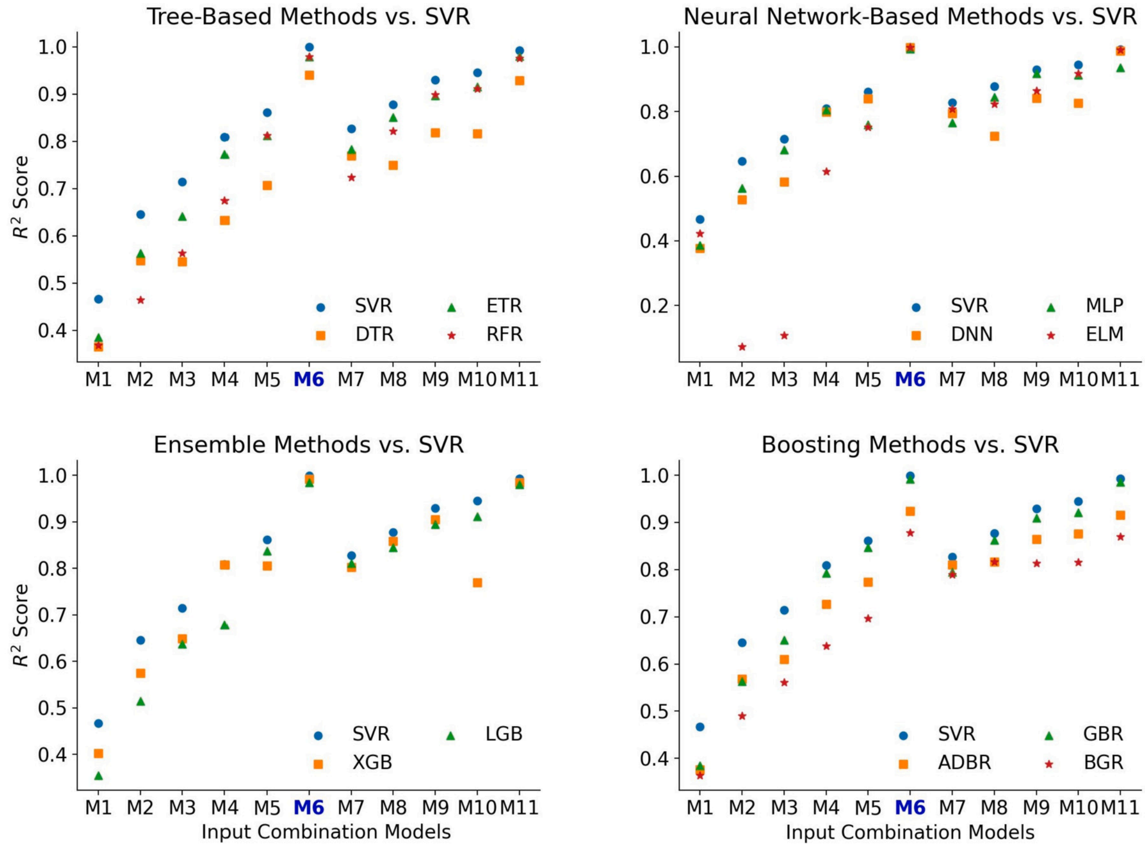

Fig. 4. Using the correlation Coefficient of Determination (R^2) for a statistical evaluation of the proposed SVR model *vs.* the Tree-based, Neural Network-based, Ensemble-based and Boosting-based algorithms in the testing phase.

Table 5
Comparison of the predictive skill of the proposed SVR model for student's OT prediction *vs.* benchmark methods in terms of the Relative Root Mean Square Error (*RRMSE %*).

	SVR	ELM	XGB	MLP	ETR	GBR	DNN	RFR	DTR	**LGB**	**ADBR**
M6	**0.37%**	**0.49%**	**1.22%**	**1.06%**	**1.96%**	**1.21%**	**0.48%**	**1.94%**	**3.29%**	**1.70%**	**3.79%**
M1	9.91%	10.40%	10.63%	10.81%	10.79%	10.76%	10.88%	10.90%	10.96%	11.02%	10.76%
M2	8.20%	13.60%	9.05%	9.25%	9.21%	9.18%	9.69%	10.20%	9.32%	9.68%	9.10%
M3	7.36%	13.30%	8.21%	7.84%	8.35%	8.24%	9.19%	9.17%	9.41%	8.40%	8.71%
M4	5.87%	8.40%	5.94%	5.95%	6.48%	6.16%	6.09%	7.87%	8.22%	7.75%	7.16%
M5	5.04%	6.77%	6.00%	6.94%	5.91%	5.28%	5.53%	5.91%	7.44%	5.46%	6.52%
M7	5.67%	6.03%	6.13%	6.75%	6.40%	6.24%	6.29%	7.25%	6.62%	5.97%	5.99%
M8	4.76%	5.73%	5.16%	5.38%	5.29%	5.08%	7.23%	5.77%	6.89%	5.37%	5.92%
M9	3.63%	5.07%	4.25%	3.95%	4.42%	4.15%	5.52%	4.36%	5.87%	4.49%	5.08%
M10	3.21%	3.95%	6.64%	4.06%	4.00%	3.88%	5.86%	4.11%	5.89%	4.11%	4.88%
M11	1.16%	1.25%	1.70%	3.43%	1.91%	1.61%	1.52%	2.11%	3.57%	1.93%	3.96%

model performance across different datasets. Likewise, the *RMAE* provides a scale-invariant measure, which ensures that the metric is meaningful regardless of the magnitude of the data values. Both *MAE* and *RMAE* values close to 0 therefore indicate excellent prediction accuracy.

In Table 6, the proposed SVR model is compared against best-performing models in terms of *MAE* and *RMAE* from each category: neural network-based (ELM, DNN), tree-based (ETR), ensemble-based (XGB), and boosting-based (GBR). The results reveal that the SVR model designated as M6 ($OT = f(W1, W2, W3, W4, W5, ET)$), achieves the lowest *MAE* and *RMAE* values ($MAE \approx 0.268$, $RMAE \approx 0.33$). With the successive addition of assignment marks as an input, the SVR model continues to improve in terms of *MAE* and *RMAE*. For instance, comparing the SVR with input combinations M1 (OT = f(W1)), M2 ($OT = f(W1, W2)$), M3 ($OT = f(W1, W2, W3)$), M4 ($OT = f(W1, W2, W3, W4)$), M5 ($OT = f(W1, W2, W3, W4, W5)$), and M6 ($OT = f(W1, W2, W3, W4, W5, ET)$), we note that the *MAE* gets reduced by 22% (M1 to M2), 12% (M2 to M3), 20% (M3 to M4), 14% (M4 to M5), and 92% (M5 to M6). Furthermore, when comparing the SVR model with M6 input combination to M7 ($OT = f(ET)$), M8 ($OT = f(ET, W1)$), M9 ($OT = f(ET, W1, W2)$), M10 ($OT = f(ET, W1, W2, W3)$), and M11 ($OT = f(ET, W1, W2, W3, W4)$), the *MAE* increased by 13% (M7 to M8), 25% (M8 to M9), 12% (M9 to M10), and 67% (M10 to M11). Consistent with earlier results, Table 6 further confirms that the proposed SVR model with M6 input combination had the most exceptional predictive power compared with the other models used currently in predicting the OT in TPP7155.

6.6. Further model validation: LM and NS indices

In the next part of SVR model evaluation, we now adopt the Legates and McCabe Index (*LM*) that measures its performance by comparing the average magnitude of absolute errors to the mean OT value of actual data. Note that *LM* is bounded by [0, 1] with 1 indicating a perfect prediction and 0 indicating a mediocre SVR model. In conjunction with *LM*, we adopt the Nash–Sutcliffe Efficiency (*NS*) to benchmark the pro-

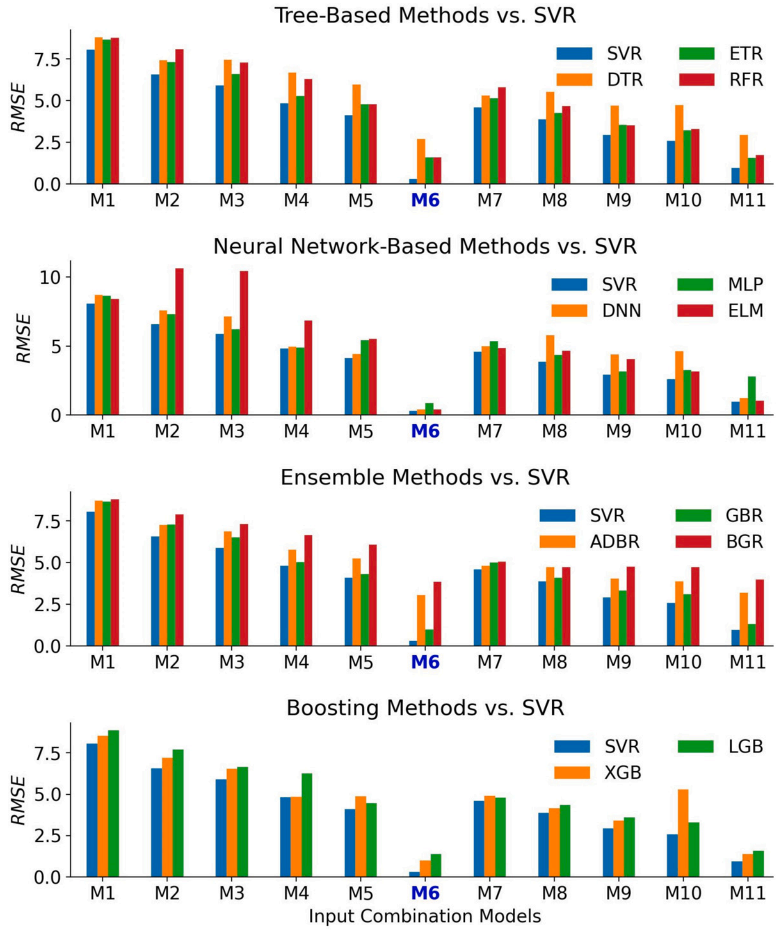

Fig. 5. Statistical evaluation of the proposed SVR model for student's OT prediction against Tree-based, Neural Network-based, Ensemble-based, and Boosting-based algorithms in the test phase using Root Mean Square Error ($RMSE$).

Table 6
Mean Absolute Error (*MAE*) and Relative MAE (*RMAE %*) computed between the observed and the predicted OT generated by the proposed SVR model compared with the ELM, DNN, XGB, ETR and GBR models.

	SVR		ELM		DNN		XGB		ETR		GBR	
	MAE	*RMAE*	*MAE*	*RMAE*	*MAE*	*RMAE*	*MAE*	*RMAE*	*MAE*	*RMAE*	*MAE*	*RMAE*
M6	0.268	0.33%	0.31	0.40%	0.311	0.40%	0.706	0.93%	0.973	1.29%	0.643	0.89%
M1	6.683	8.40%	7.063	8.92%	6.928	8.92%	7.057	8.95%	7.305	9.25%	7.278	9.23%
M2	5.418	6.95%	9.161	11.82%	6.328	11.82%	5.976	7.64%	6.248	7.96%	6.194	7.92%
M3	4.77	5.95%	6.374	17.99%	6.101	17.99%	5.57	7.05%	5.676	7.21%	5.519	7.13%
M4	3.832	4.73%	4.491	5.66%	3.982	5.66%	4.037	5.06%	4.366	5.40%	4.077	5.12%
M5	3.297	4.17%	3.839	5.12%	3.714	5.12%	3.996	5.03%	3.816	4.72%	3.516	4.42%
M7	3.668	4.66%	3.921	5.00%	4.087	5.00%	4.083	5.16%	4.176	5.32%	4.128	5.37%
M8	3.198	4.09%	3.511	4.44%	3.906	4.44%	3.387	4.40%	3.62	4.64%	3.393	4.42%
M9	2.387	3.14%	2.891	3.95%	3.177	3.95%	2.734	3.57%	2.839	3.66%	2.751	3.62%
M10	2.109	2.75%	2.466	3.27%	3.678	3.27%	4.295	5.41%	2.68	3.48%	2.584	3.35%
M11	0.701	0.89%	0.765	1.00%	0.972	1.00%	0.981	1.29%	1.087	1.44%	0.982	1.29%

posed SVR model against the other models across a range of model input combinations (i.e., M1–M11). We have selected *NS* index as an improvement over *LM* as it penalises the larger errors more severely by squaring the differences between the predicted and the observed values, making it more sensitive to significant deviations and highlighting the areas where the model may underperform. In contrast, the *LM* index, which uses absolute differences, is less sensitive to outliers and may underemphasise large errors.

As shown in Table 7 for the testing phase, the *LM* and *NS* are closely aligned, with the highest *LM* (≈ 1) and *NS* (≈ 0.97) achieved by the proposed SVR model, particularly for Model M6 (with all predictors), compared to M7 (*LM*≈ 0.49, *NS*≈ 0.45), M8 (*LM*≈ 0.58, *NS*≈ 0.55), M9 (*LM*≈ 0.72, *NS*≈ 0.67), M10 (*LM*≈ 0.76, *NS*≈ 0.72), and M11 (*LM*≈ 0.92, *NS*≈ 0.91).

Fig. 6(a), Fig. 6(b), Fig. 6(c), and Fig. 6(d) show scatterplots comparing the predicted and observed overall scores generated by the proposed SVR model and the benchmark models: ELM, ETR, GBR, and XGB. Each plot includes the least-squares fit represented by $y = mx + C$, where m = gradient and C = y-intercept. A gradient (m) close to 1 and a low $y - intercept$ (C) indicate a strong linear relationship between y and x.

The proposed SVR model (with $y = 1.00x + 0.15$) using all predictor variables for M6 demonstrates superior performance as its slope is within the proximity of 1 and with the smallest y-intercept. This indicates a more accurate and stronger linear relationship with the actual

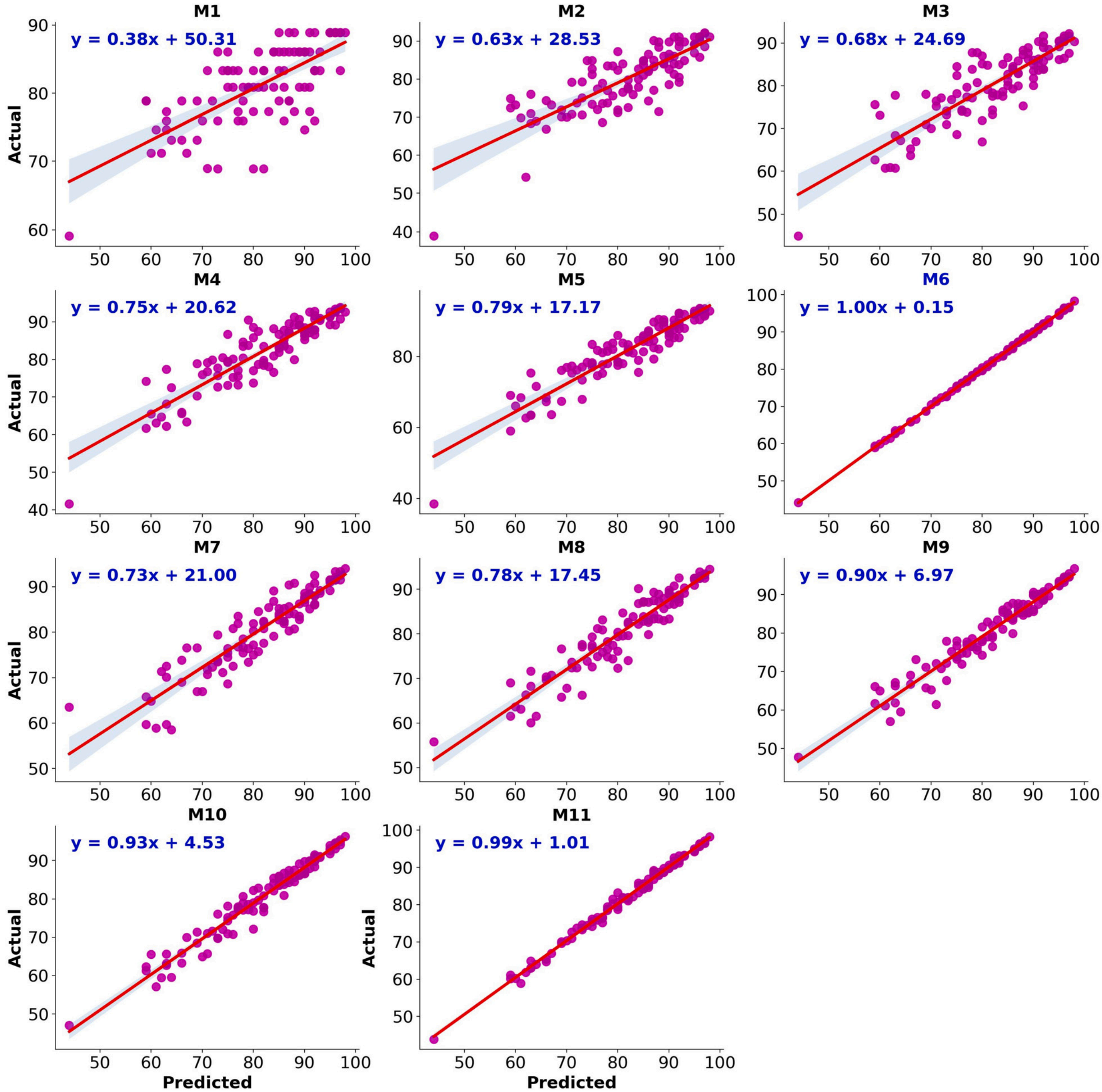

Fig. 6(a). Scatter plot of the predicted and the actual OT in the testing phase for all machine learning models (M1, M2,...., M11). The least square regression line is shown in each sub-panel.

Table 7
Statistical evaluation of the proposed SVR model against the benchmark models for OT prediction in the testing phase using the Nash–Sutcliffe coefficient (*NS*) and the Legates and McCabe's Index (*LM*).

	SVR		ELM		DNN		XGB		ETR		GBR	
	LM	*NS*	*LM*	*NS*	*LM*	*NS*	*LM*	*NS*	*LM*	*NS*	*LM*	*NS*
M6	0.999	0.966	0.978	0.921	0.887	0.928	0.969	0.949	0.929	0.951	0.976	0.952
M1	-0.394	-0.545	-0.158	-0.374	-0.765	-0.52	-0.832	-1.236	-0.171	-0.935	-1.881	-1.339
M2	0.19	-1.384	0.159	0.028	-0.109	-0.037	0.404	-3.464	0.387	0.032	-0.069	-0.001
M3	0.312	0.252	0.21	0.036	0.11	0.23	0.54	0.376	0.507	0.154	0.351	0.491
M4	0.473	0.435	0.521	0.446	0.337	0.478	0.725	0.589	0.789	0.722	0.599	0.736
M5	0.537	0.506	0.505	0.355	0.411	0.551	0.806	0.725	0.793	0.592	0.67	0.801
M7	0.494	0.449	0.568	0.401	0.373	0.436	0.726	0.688	0.803	0.648	0.611	0.696
M8	0.578	0.548	0.529	0.574	0.505	0.573	0.819	0.752	0.703	0.814	0.762	0.816
M9	0.721	0.666	0.659	0.673	0.65	0.677	0.919	0.845	0.855	0.882	0.86	0.893
M10	0.757	0.717	0.593	0.278	0.669	0.703	0.939	0.909	0.84	0.401	0.89	0.914
M11	0.922	0.914	0.893	0.887	0.872	0.89	0.992	0.991	0.988	0.983	0.977	0.985

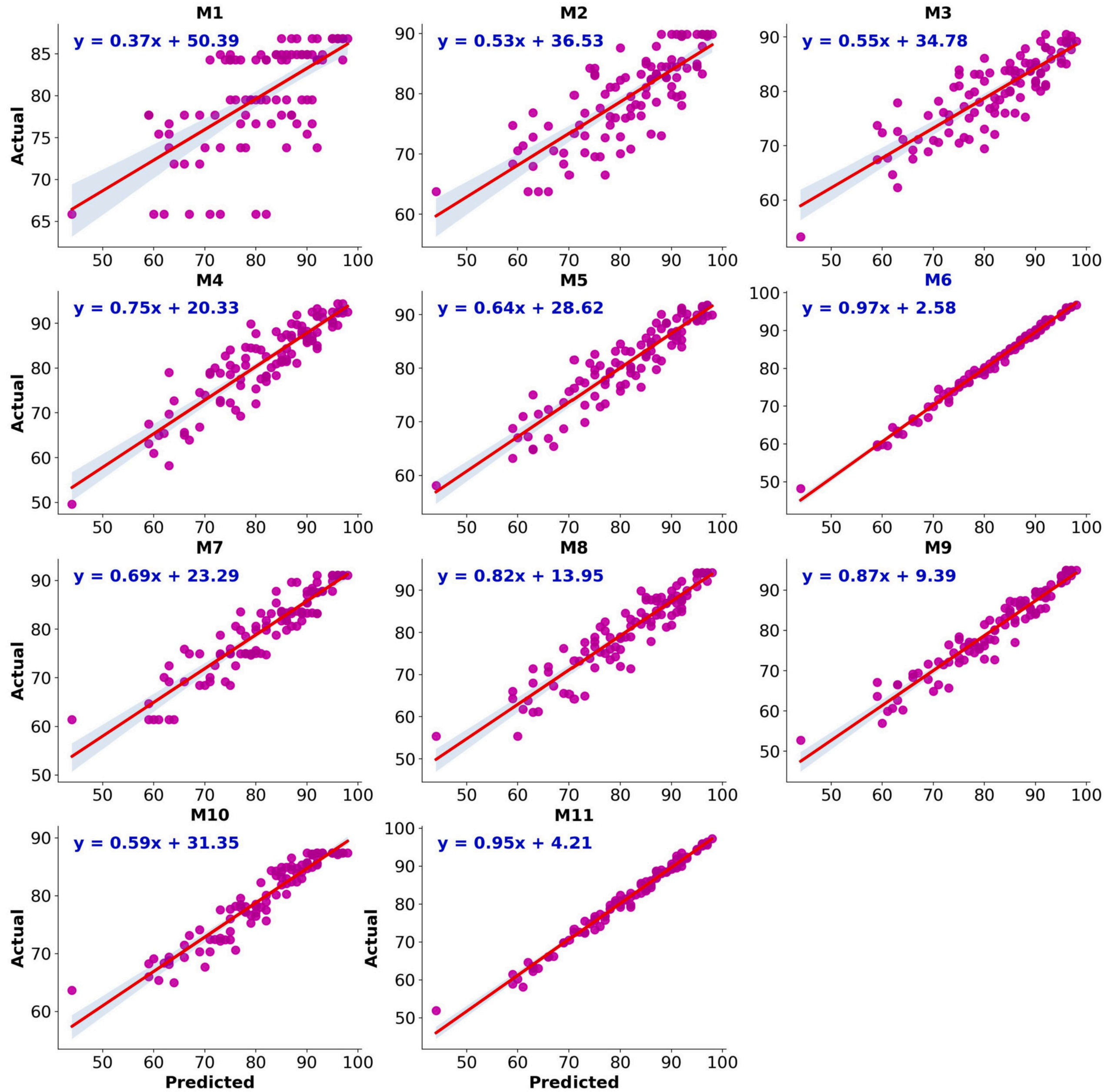

Fig. 6(b). Scatter plot of the predicted and actual OT in the testing phase for all the models (M1, M2,...., M11) using the Neural Network-based algorithm (ELM). The least square regression line is shown in each sub-panel.

OT data. Among the benchmark models, the GBR model (with equation $y = 0.97x + 2.033$) ranks first, followed by the ELM model (with equation $y = 0.97x + 2.58$), the XGB model (with equation $y = 0.96x + 3.43$), and the ETR model (with equation $y = 0.93x + 6.25$), all using all predictors (M6).

When comparing SVR model with input combinations M6 ($OT = f(W1, W2, W3, W4, W5, ET)$) and M11 ($OT = f(ET, W1, W2, W3, W4)$), the gradients are nearly identical with a value of ≈ 1. However, the $y-intercept$ increases from 0.15 to 1.01, emphasising the significance of including all student evaluation components in predicting the OT. Therefore, the proposed SVR model with the M6 input combination is well-suited for accurately predicting the OT of TP7155 students.

In subsequent analysis, the frequency distribution of prediction errors ($|PE|$) produced by SVR model with M6 ($OT = f(W1, W2, W3, W4, W5, ET)$), M11, and M5 input combinations is compared to benchmark methods, as illustrated in Fig. 7(a), Fig. 7(b), and Fig. 7(c), respectively.

Each error bin, displaying the percentage of all tested points at the top, has a size of 0.5 ($0 \le |PE| \le 0.5$) for the M6 and M11 input combinations and 2.0 ($0 \le |PE| \le 2.0$) for the M5 input combination. Notably, the OT predictions using the SVR model with the M6 input combination (Fig. 7(a)) exhibited the highest frequency of errors within the smallest error bracket ($0 \le |PE| \le 0.5$), covering 96% of the test data.

In comparison, the ETR, ELM, DNN, GBR, and XGB models recorded 33%, 88%, 75%, 53%, and 46%, respectively. Moreover, the frequency of errors exceeding $0 \le |PE| \le 1.0$ for the SVR model was zero, while the ELM, GBR, XGB, and ETR models showed exceedances of 2%, 25%, 23%, and 34%, respectively. Similarly, the SVR model with the M11 input combination demonstrated superior performance with 46% of the test data falling within the $0 \le |PE| \le 0.5$ error bracket, followed by the ELM, GBR, DNN, XGB, and ETR models.

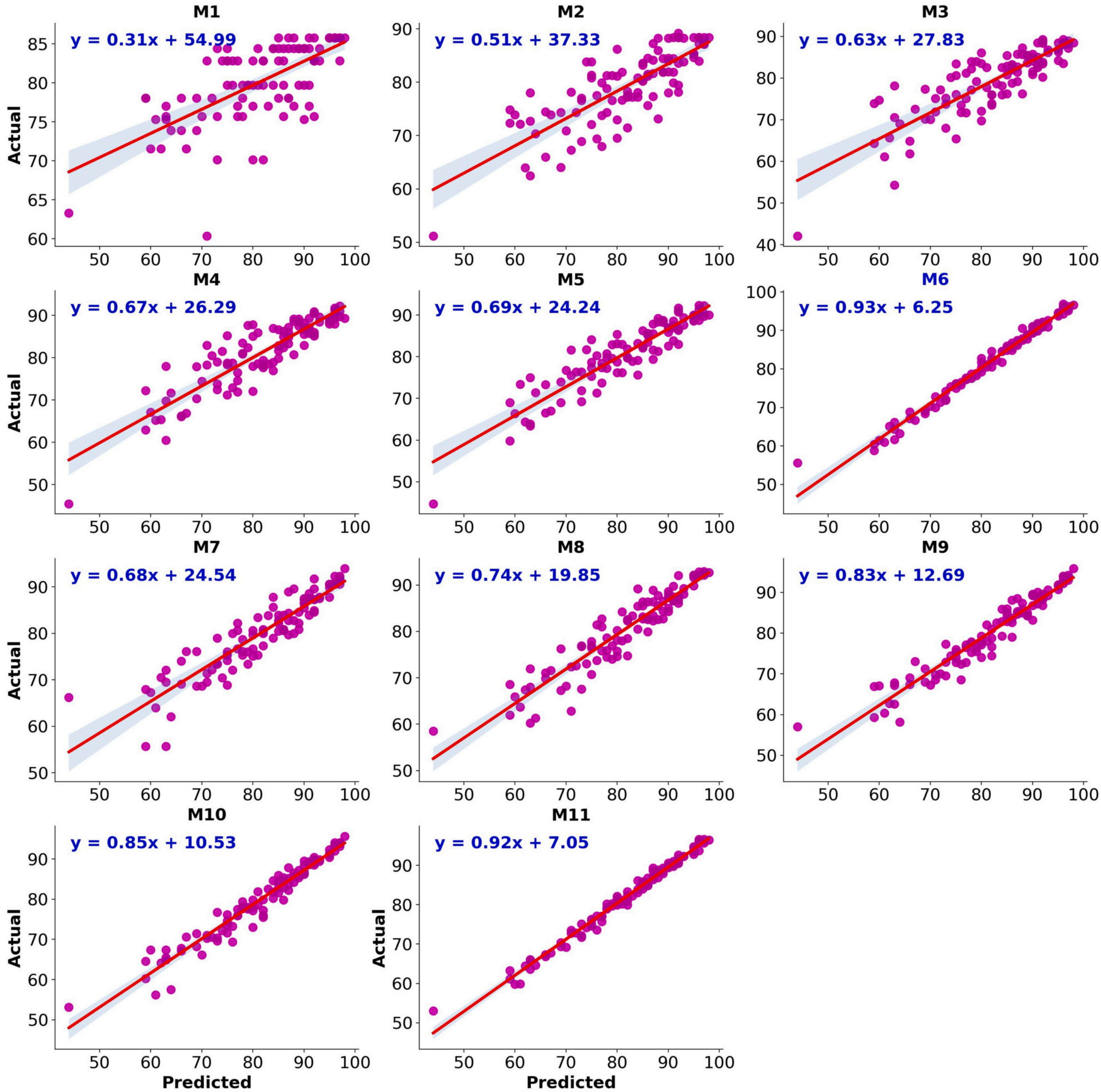

Fig. 6(c). Scatter plot of the predicted and actual OT in the testing phase for all the models (M1, M2,...., M11) using the Neural Tree-based algorithm (ETR). The least square regression line is shown in each sub-panel.

For the SVR model with the M5 input combination, 35% of the errors were within the $0 \leq |PE| \leq 2.0$ range, compared to 34% for the ELM and GBR models, 32% for the ETR and DNN models, and 30% for the XGB model. As evidenced by Table 6, Table 7 and Fig. 7(a), Fig. 7(b), and Fig. 7(c), the SVR model with the M6 input combination generates most of its prediction errors within the lowest magnitude band, making it the most accurate in predicting the OT.

To further assess the predictive efficacy of the proposed SVR model, promoting percentage based on the Willmott's Index (Δ_{WI}) and the Kling Gupta's Efficiency (Δ_{KGE}) are utilised. Table 8 presents a comparative examination between the SVR model employing the M6 input combination and alternative input combinations. For instance, within the M11 input combination, in comparison to the SVR model, the Willmott's Index (*WI*) and Kling Gupta's Efficiency (*KGE*) exhibit reductions of 2.17% and 2.64%, respectively. Notably, all promotion percentage errors are non-negative, indicating that the SVR model utilising the M6 input combination ($OT = f(W1, W2, W3, W4, W5, ET)$) is deemed well-suited for predicting the overall scores of TPP7155 students.

6.7. Model interpretability

In this phase of our study, we delve into eXplainable Artificial Intelligence (*x*AI) techniques, employing the SHAP (SHapley Additive exPlanations) values analysis to shed light on the predictions derived from our SVR Model. In our research, the SVR model emerges as the most robust tool for forecasting overall scores in the TPP7155 course. This comprehensive analysis allows us to delve into the intricate details of individual features and their impact on the decision-making process of the model, offering profound insights into the determinants shaping overall scores.

Through the utilisation of SHAP's KernelExplainer, our analysis deciphers the predictions made by the SVR model by calculating SHAP values for the test dataset, thus revealing insights into feature impor-

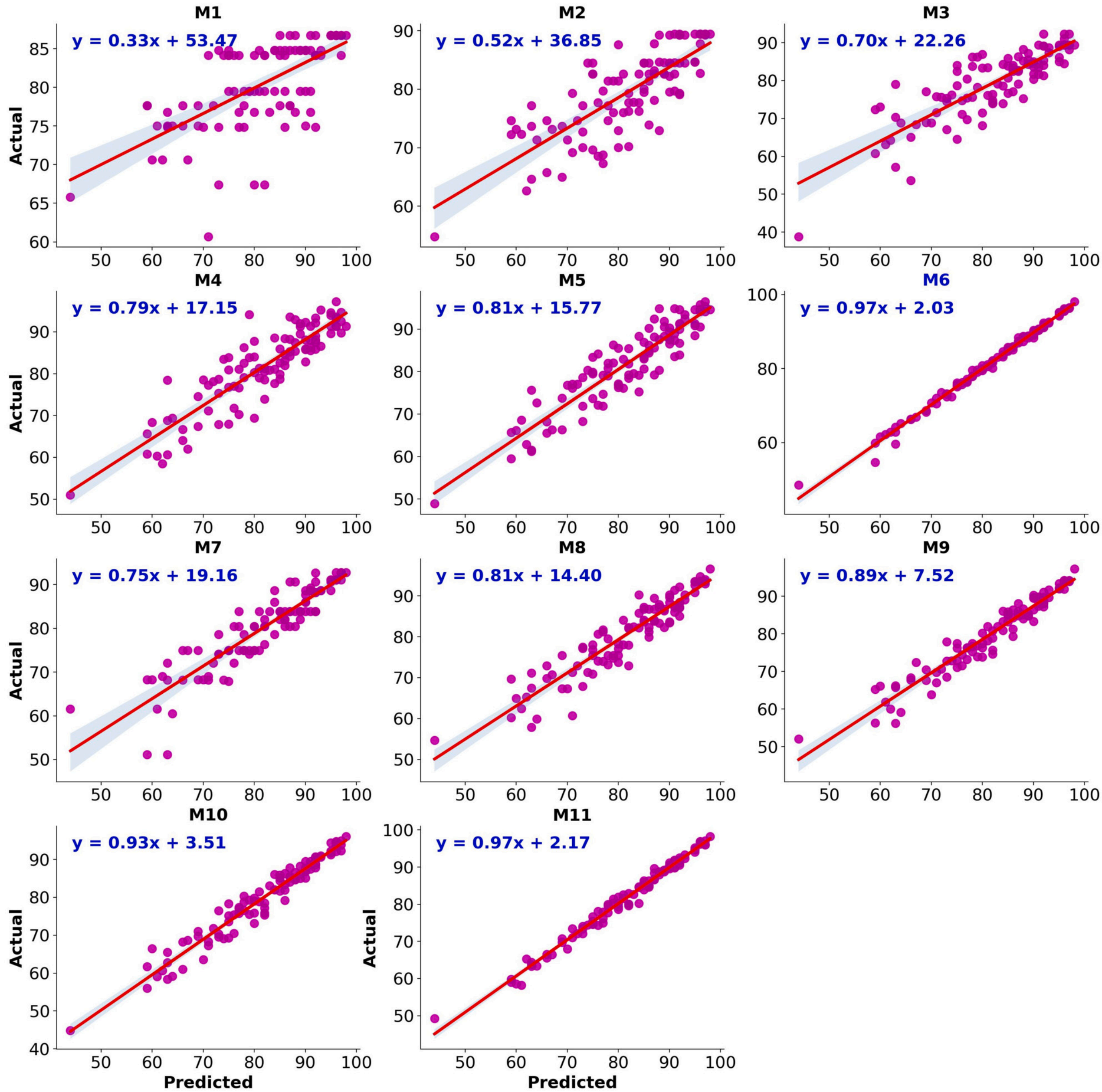

Fig. 6(d). Scatter plot of the predicted and actual OT in the testing phase for all the models (M1, M2,...., M11) using the Neural Ensemble-based algorithm (GBR). The least square regression line is shown in each sub-panel.

Table 8
The promoting percentage change in Willmott's Index (Δ_{WI}) and Kling Gupta's Efficiency (Δ_{KGE}) calculated with respect to the Model M6 for OT prediction.

	SVR		ELM		DNN		XGB		ETR		GBR	
	Δ_{WI}	Δ_{KGE}	Δ_{WI}	Δ_{KGE}	Δ_{WI}	Δ_{KGE}	Δ_{WI}	Δ_{KGE}	Δ_{WI}	Δ_{KGE}	Δ_{WI}	Δ_{KGE}
M11	2.17%	2.64%	2.19%	2.36%	2.28%	1.09%	2.21%	1.59%	1.03%	1.31%	0.16%	1.30%
M10	3.44%	5.84%	4.15%	7.40%	6.14%	8.60%	9.64%	6.23%	3.85%	8.05%	3.86%	3.35%
M9	3.89%	8.15%	3.65%	9.19%	3.83%	7.25%	2.47%	9.19%	2.47%	10.76%	2.28%	7.68%
M8	3.76%	19.42%	5.30%	13.77%	7.49%	14.69%	3.95%	43.42%	4.22%	28.47%	3.88%	13.96%
M5	4.17%	21.95%	6.75%	20.31%	4.58%	13.14%	6.81%	13.27%	5.74%	21.59%	4.30%	14.78%
M7	5.53%	22.17%	6.21%	20.05%	5.13%	9.89%	6.43%	20.01%	6.81%	29.47%	6.23%	14.93%
M4	5.87%	27.02%	10.50%	28.46%	5.33%	8.10%	5.78%	32.30%	7.17%	30.23%	5.91%	21.90%
M3	9.33%	30.25%	21.06%	36.03%	11.45%	17.36%	13.27%	55.82%	11.75%	33.42%	10.71%	24.69%
M2	11.90%	34.81%	47.73%	132.68%	13.99%	24.57%	16.04%	53.49%	16.51%	56.78%	16.56%	54.59%
M1	24.30%	90.35%	27.28%	101.88%	23.33%	51.43%	26.58%	86.61%	30.54%	122.67%	29.03%	101.39%

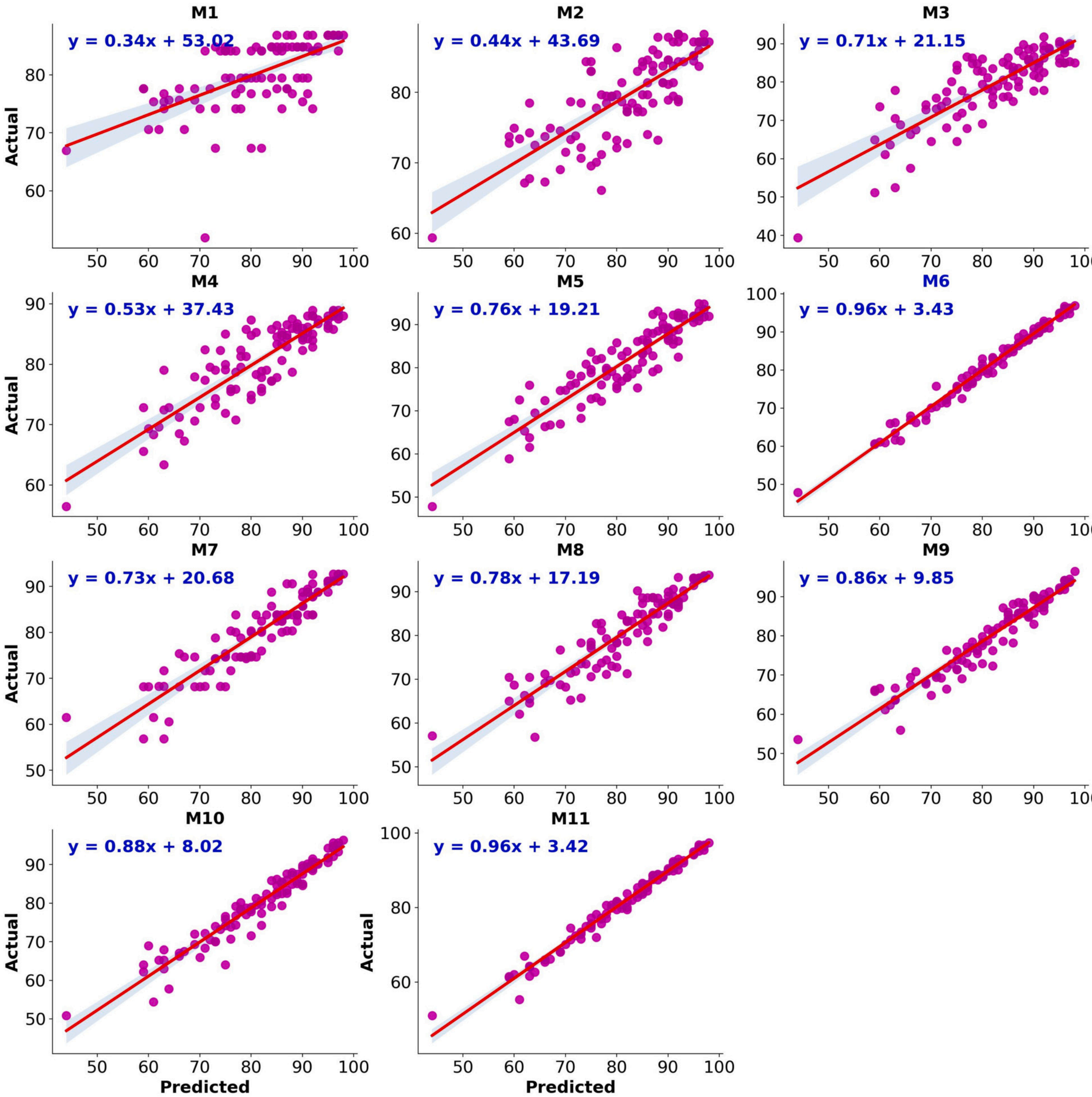

Fig. 6(e). Scatter plot of the predicted and actual OT in the testing phase for all the models (M1, M2,...., M11) using the Boosting-based algorithm (XGB). The least square regression line is shown in each sub-panel.

tance and individual predictions on previously unseen data instances. These SHAP values serve as crucial metrics of feature importance, illustrating the degree to which each feature contributes, positively or negatively, to each individual prediction. This quantification enables a nuanced understanding of how each feature influences the model's predictive outcomes.

We present two visually generated outputs in Fig. 8 and Fig. 9 to aid in the interpretation of model predictions through SHAP values analysis. This approach not only enhances the transparency and interpretability of the SVR model but also furnishes valuable insights into the underlying factors propelling OT predictions within the context of the TPP7155 course.

The beeswarm plot in Fig. 8 depicts the distribution of SHAP values for each feature across the dataset. The x-axis position reflects the feature's impact on the model's output, with features increasing the prediction shown on the right and those decreasing it on the left. The red and blue colours in the plot represent higher and lower values of the OT prediction, respectively. This distribution highlights the variability in each feature's impact.

Fig. 8 clearly shows that ET (Examination Score) is the most significant feature, as higher ET values (indicated in red) correspond to higher OT (overall score) predictions, suggesting a positive relationship between ET and OT. Although the beeswarm plots indicate a positive association between OT and assignment marks (W1, W2, W3, W4, W5), this correlation might introduce redundancy among these features, reducing their importance in the feature ranking.

The global summary plot in Fig. 9 displays each feature's mean absolute SHAP values, which measure their overall importance. A higher value signifies a greater impact on the model's predictions. The bars are colour-coded and show the mean absolute SHAP value for each feature, with a sign ($\pm$) indicating the direction of the impact. From Fig. 9, ET is identified as the most significant feature, affecting the predicted OT by an average of ± 4.93. In contrast, W1 is the least informative feature, contributing only ± 0.64 to the OT prediction.

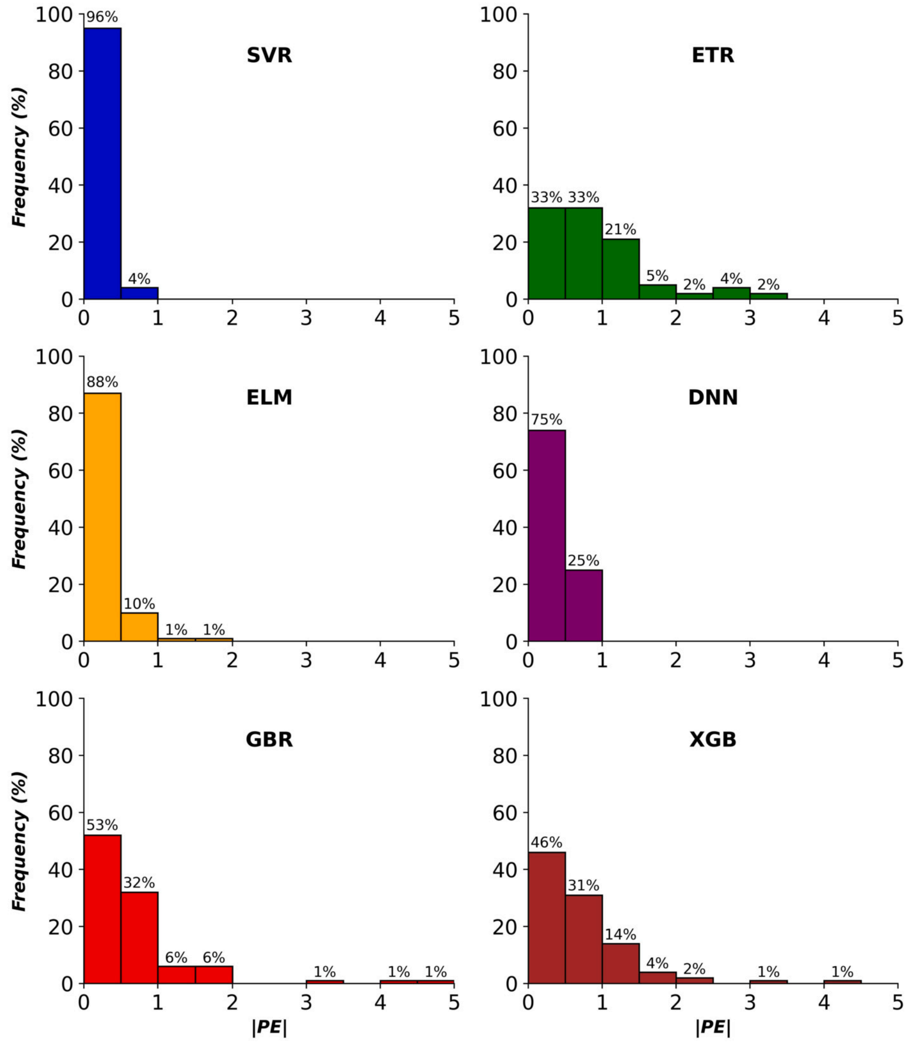

Fig. 7(a). Histogram illustrating the frequency (%) of absolute Prediction errors ($|PE|$) of the best performing algorithm during the test period compared with benchmarked models for the prediction of the student OT of the M6 model.

For explanability of the proposed SVR model, we now show the results of the LIME method to provide an interpretation of the predicted instances of the tested datasets at a local level. These results are provided for six specific test case instances: $i = 1$, 28, 57, 72, 85, and 99. Fig. 10 illustrates the bar graphs that depict the contributions of six predictors to the model predictions for each i^{th} instance of tested OT value. In these graphs, the red bars indicate negative LIME values, suggesting a lower OT prediction, while the green bars indicate positive LIME values that suggest a higher OT prediction. The summary for each instance is also detailed below.

a. $i = 1$

  i. ET: When $ET \leq 40.50$, this condition has a positive influence on the SVR model prediction as indicated by the green bar.
  ii. In respect to W1 to W5: The conditions for W1, W2, W3, W4, and W5 similarly show how these assignment scores influence the proposed SVR predictions under the specified conditions. For example, $W2 \leq 24.00$ and $33.00 < W4 \leq 36.00$ also contribute positively to influence the prediction of the OT whereas the condition $24.00 < W3 \leq 26.00$ appears to have a slight negative influence.

b. $i = 28$

  i. ET: For the condition $46.00 < ET \leq 52.00$, the examination score appears to have a negative influence on the SVR model predictions, shown by the red bar.
  ii. In respect to W1 to W5: $W4 \geq 36.00$ and $26.00 < W3 \leq 28.00$ contribute negatively to the prediction. On the contrary, $8.50 < W1 \leq 9.00$ has a very slight positive influence.

c. $i = 57$

  i. ET: The condition $46.00 < ET \leq 52.00$ negatively impacts the prediction, indicating that higher examination scores within this range reduce the prediction value.
  ii. In respect to W1 to W5: $W2 \leq 28.50$, $W1 \geq 9.00$, and $24.00 < W3 \leq 26.00$ have negative impacts, while $28.50 < W4 \leq 33.00$ shows a slight positive influence.

d. $i = 72$

  i. ET: $ET \leq 40.50$ has a positive impact on the prediction, contributing significantly.
  ii. In respect to W1 to W5: $28.50 < W4 \leq 33.00$ and $W5 \leq 9.00$ also positively influence the prediction, while $24.00 < W2 \leq 26.50$ and $W1 \leq 7.00$ show negative influence.

e. $i = 85$

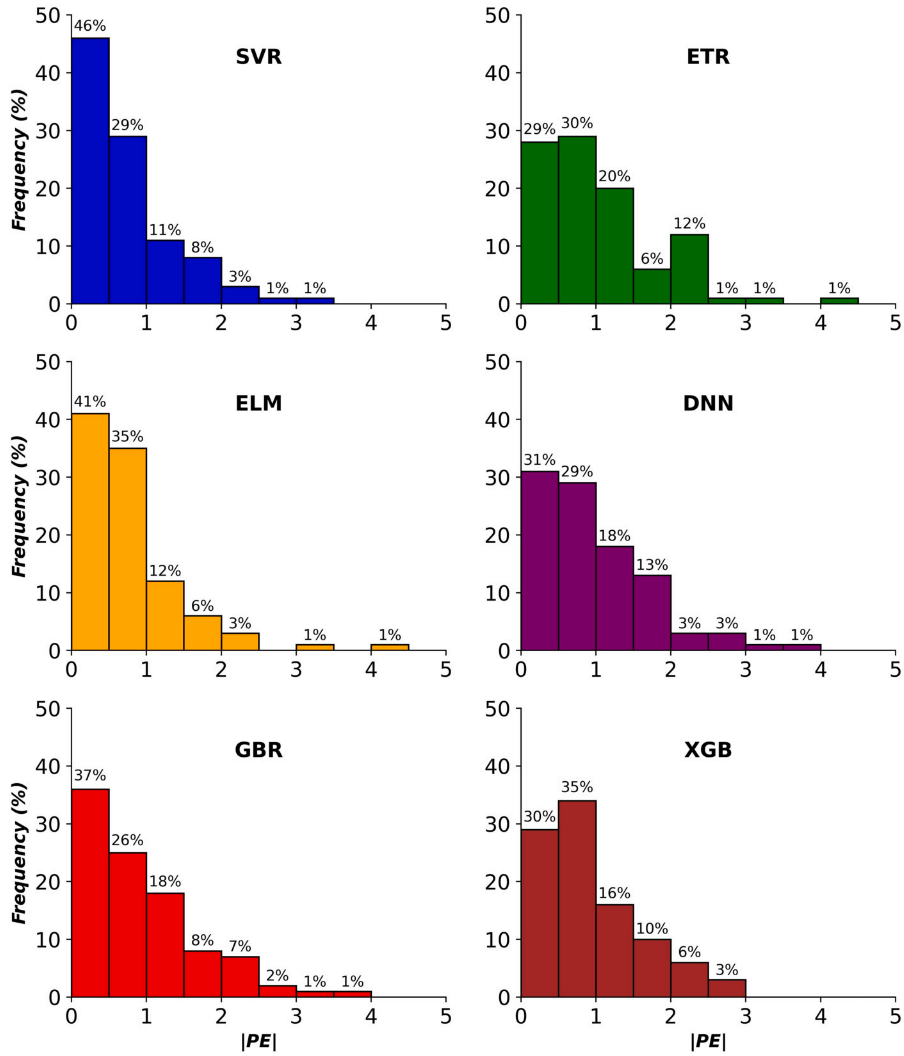

Fig. 7(b). Histogram illustrating the % frequency of absolute prediction errors ($|PE|$) of the best performing model during the testing period compared with benchmarked models for the prediction of OT of the M11 model.

i. ET: $ET \leq 40.50$ again shows a strong positive contribution to the prediction.
ii. In respect to W1 to W5: All the assignments scores (W1, W2, W3, W4, W5) are positively impacting the prediction when they meet the specified conditions.

f. $i = 99$

i. ET: The condition $ET \geq 52.00$ significantly negatively impacts the prediction, as indicated by the red bar.
ii. In respect to W1 to W5: $W2 \leq 24.00$ and $W1 \geq 9.00$ have positive impacts, whereas $11.50 < W5 \leq 13.00$ and $24.00 < W3 \leq 26.00$ show slight negative contributions.

The results presented underscore the growing necessity for explainability in AI-based models across various domains, driven by regulatory demands for capabilities such as traceability, transparency, and interpretability (Holzinger, 2021; Holzinger et al., 2022). In practice, however, powerful machine learning frameworks often exhibit sensitivity to minor variations in input data, making predictive outcomes challenging to explain. This sensitivity can introduce significant variability in results, especially in critical domains that frequently contend with low-quality datasets due to non-ideal, non-independent, and non-identically distributed (i.i.d.) data (Holzinger, 2021). Moreover, ensuring AI models are legally and ethically compliant is essential to foster trust in AI systems moving forward (Holzinger et al., 2022).

In this context, enhancing explainability and performance robustness in AI models bolsters confidence and reliability, empowering human experts with greater control over the AI pipeline. Addressing these critical concerns, this research integrates trustworthy AI principles into the OT prediction framework using powerful model-agnostic *x*AI tools. The newly developed interpretable hybrid SVR model created through EDM and *x*AI tools, equips higher education institutions with consistent, transparent, and more accurate predictions of OT. This trustworthy and high-performing system holds significant promise as a decision-support tool, potentially driving strategic improvements in student retention and success rates in the near future.

7. Interpretation of findings and educational implications

This study aimed to assess the predictive accuracy of a hybrid TPE-optimised Support Vector Regression (SVR) model to predict final academic performance using data from the TPP7155 (General Science). Data included the Overall Mark (OT) and scores from written assignments (W1, W2, W3, W4, and W5) along with the final examination mark (ET). Through statistical and visual analyses of the pre-

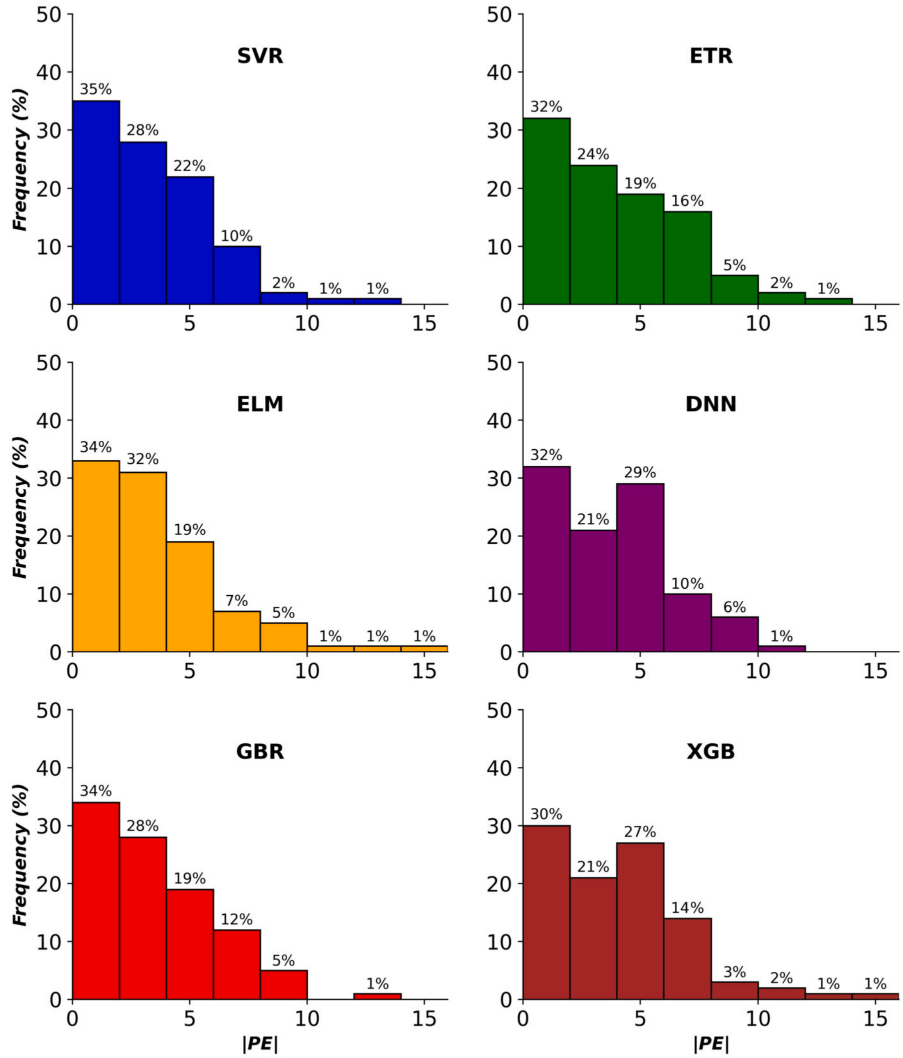

Fig. 7(c). Histogram illustrating the frequency (%) of absolute Prediction errors ($|PE|$) of the best performing algorithm during the test period compared with benchmarked models for the prediction of the OT of the M5 model.

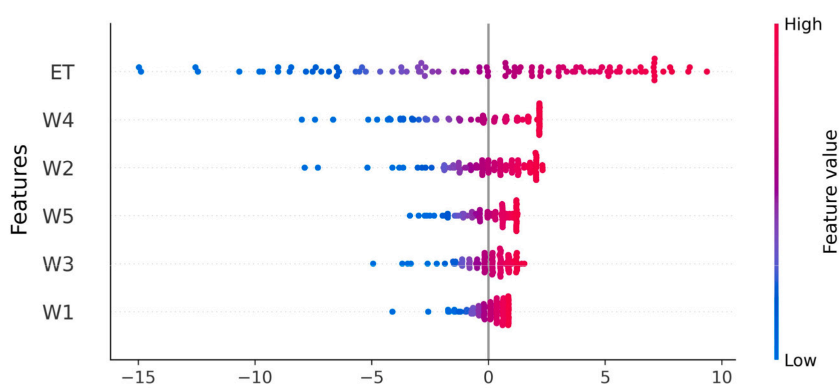

Fig. 8. SHAP beeswarm plot displaying the combined view of feature importance and feature impacts on the OT prediction in the dataset with SVR algorithm using M6 input combination. Each point on $x-axis$ represents a Shapley value for a feature, while $y-axis$ displays six features arranged by their mean absolute Shapley values. The colour gradient from blue to red indicates the feature values, with blue representing low values and red representing high values.

dicted and actual OT values, a hybrid model integrating Tree-structured Parzen Estimation (TPE) and Support Vector Regression (SVR) effectively captured the interdependence between predictors and the target variable. The developed model was meticulously tailored to fit the specific datasets sourced from the University of Southern Queensland, Australia, showcasing its adaptability and effectiveness in educational predictive analytics.

The findings indicated that the hybrid SVR model significantly outperformed traditional predictive models such as linear regression and simple machine learning approaches, yielding a high degree of accuracy

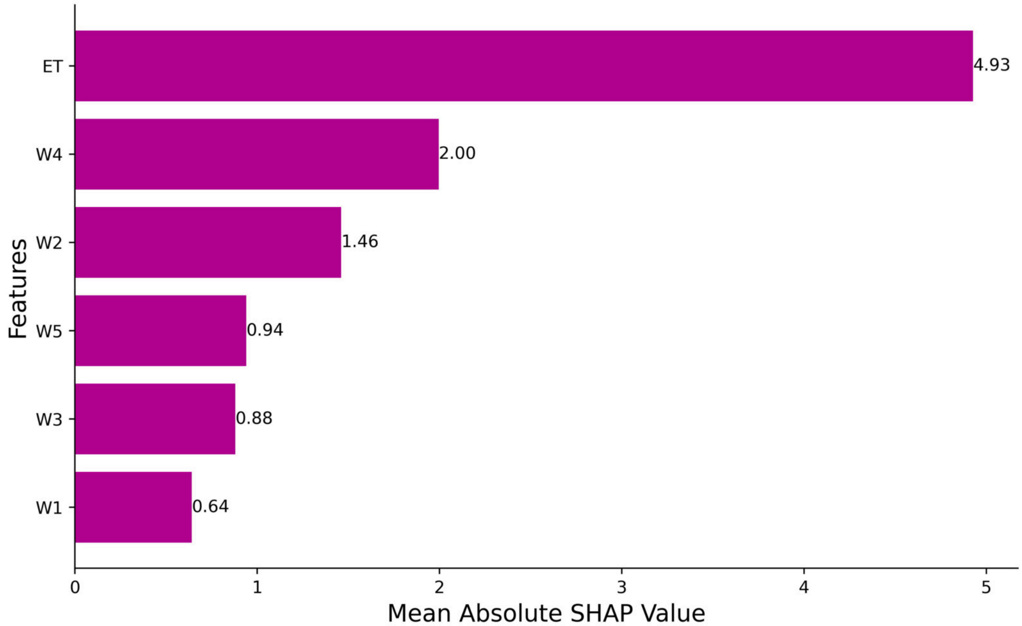

Fig. 9. The SHAP Global Summary plot shows the mean impact of each input feature (i.e., assessment marks) on the proposed SVR model's capability for OT prediction using the M6 input combination. Features are ranked by their importance. The length of the bar represents the mean absolute SHAP value for each feature, indicating its importance. Positive SHAP values indicate a feature's increasing impact on the model's prediction, while negative values indicate a decreasing impact.

in predicting final grades. Specifically, the model showed a Mean Absolute Error (MAE) and Relative Mean Absolute Error ($RMAE$) values ($MAE \approx 0.268$, $RMAE \approx 0.33$), which was lower than most comparable predictive models found in the literature. Examination of prediction error distributions highlighted the model's capability to yield a higher frequency of errors within a narrower error range ($0 \leq |PE| \leq 0.5$), covering 96% of the test data compared to alternative models. The model also exhibited a higher Legate's & McCabe's Index, a measure of its robustness and reliability. These results are in comparison to previous research by Ahmed et al. (2022) that recorded a Legate's & McCabe's Index of less than 0.80 when testing a multivariate regression splines (MARS) model. These results validate the utility of combining hyperparameter optimisation (via TPE) with SVR, which allowed for improved precision in predicting academic outcomes.

The model's root-mean-square error (RMSE) and coefficient of determination (R^2) highlighted a strong predictive relationship between assessment marks (including examination scores and continuous assessments) and the final grade. The simulations revealed that both the examination and assignments contributed significantly to the OT, with each assignment exerting a discernible influence. Examination scores, however, emerged as the most influential predictor. These findings demonstrate the practical utility of the proposed hybrid TPE-optimised SVR model in educational contexts and aligns with existing research, which suggests that early evaluations in higher education significantly impact final academic success (Nguyen-Huy et al., 2022).

7.1. Comparison with previous studies

The results of this study align with and extend the work of other scholars in the field of academic performance prediction, particularly the use of assessments/examinations but with different statistical and machine learning models. For instance, Deo et al. (2020) demonstrated the value of extreme machine learning models, specifically Support Vector Machines (SVM), in predicting academic outcomes. While their study utilised a simpler SVM model without optimisation techniques, the present research builds on this by introducing TPE optimisation, a method that fine-tunes the hyperparameters of the model to improve its overall predictive performance. The improvements achieved by the hybrid SVR model in this study show that optimisation techniques like TPE can effectively enhance model accuracy compared to traditional methods.

When compared to Nguyen-Huy et al. (2022), who used regression models in educational forecasting, our approach adds value by incorporating advanced machine learning algorithms. SVR has shown a remarkable capacity to capture non-linear relationships, which is crucial in the context of predicting academic performance, as it allows the model to handle complex, multifaceted data. Unlike traditional regression models, SVR is particularly adept at learning from small datasets, such as the one used in this study, which may not have been possible using other methods.

In respect to the study of Nguyen-Huy et al. (2022), a direct comparison with the current study is impossible as that study has used a purely statistical/copula-based model with different performance metrics. However, their copulas-based model does provide a significant advantage over the proposed SVR model by generating a joint distribution model or a conditional probability plot showing the probability of an examination score, for example, being less than or equal to a threshold mark, conditional upon a specific assignment being less than or equal to a threshold mark value. Such a model can provide significant insights into key decision-making regarding how each assessment impacts a probability-based outcome of overall mark in the course. Therefore, our study not only extends previous methodologies but also provides evidence of the long-term potential of optimised machine learning models in the context of educational analytics.

7.2. Interpretation of results

To gain insight into the factors influencing the Overall Mark (OT), both global and local interpretations of the model were examined using Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP). This analysis clearly showed that Examination Score (ET) is the primary factor, with higher ET values correlating with increased OT predictions, suggesting a positive relationship between ET and OT. Specifically, ET was identified as the most influential feature, impacting the predicted OT by an average of ± 4.93, whereas Assignment 1 (W1) was found to be the least informative, contributing only ± 0.64 to the OT prediction.

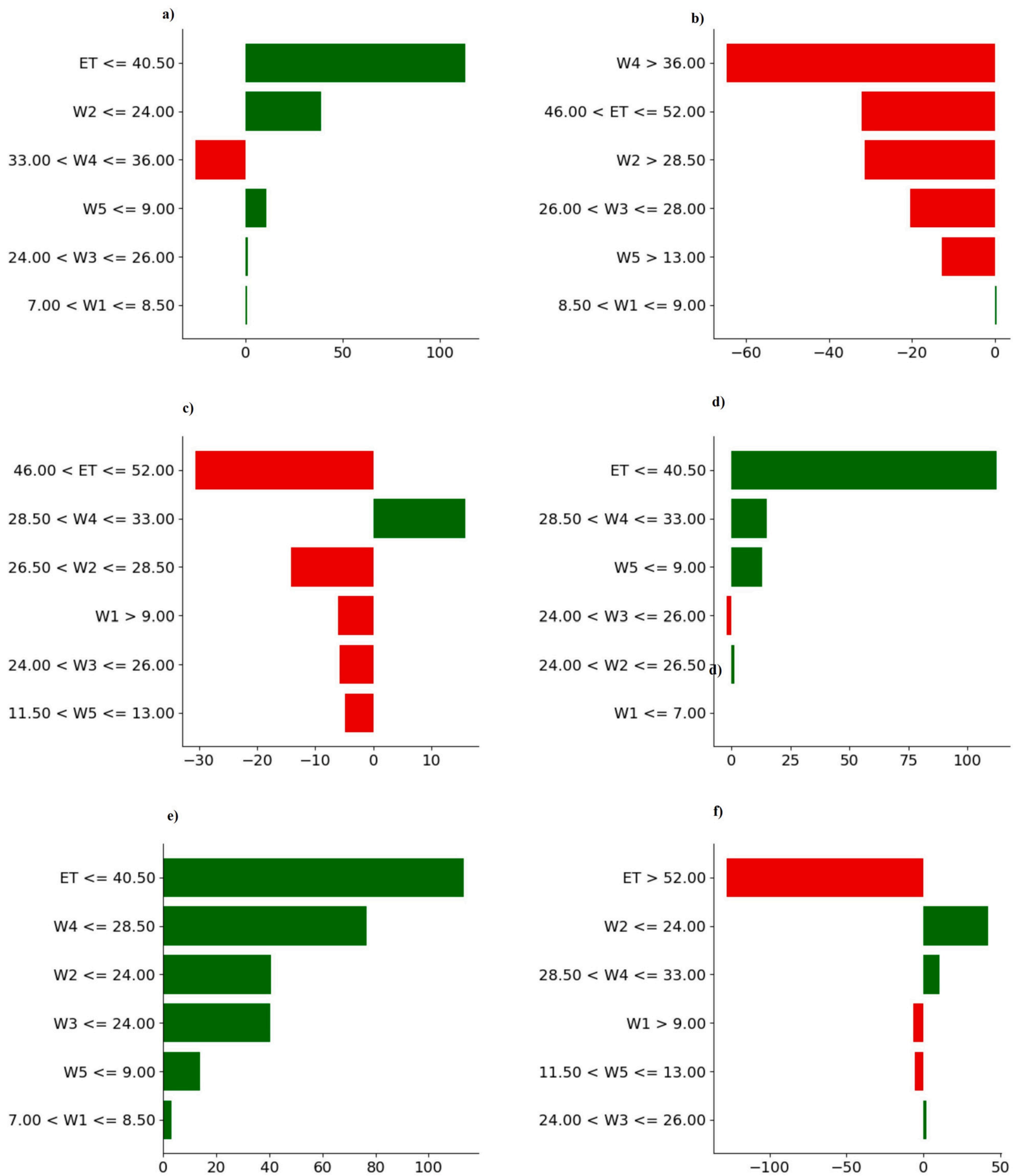

Fig. 10. Bar plots representing the LIME scores for the local instances (a) 1 (i.e., first data point in the test phase), (b) 28, (c) 57, (d) 72, (e) 85, and (f) 99 (i.e., last or 100 percentile data point) with SVR algorithm for M6. Note: The green bar (i.e., positive LIME score) indicates that the predictor favours a higher predicted OT and the red bar is when the predictor favours a lower predicted overall score.

The hybrid model's high performance is indicative of the predictive power of summative assessments, particularly examinations, in determining overall academic outcomes. The model's sensitivity to exam scores underscores the importance of these evaluations in shaping student trajectories. Findings suggest that while formative assessments are crucial for providing ongoing feedback and guiding instructional adaptations, it is the performance in summative assessments like final exams that ultimately correlates most strongly with final academic success in this context.

SHAP and LIME analyses of the hybrid model further support these findings by highlighting the importance of continuous assessments such as assignments, quizzes, and participation. These variables appeared to be the second most significant predictors of final academic performance, suggesting that ongoing student engagement with course material plays a key role in success. SHAP values provided a more granular understanding of how each individual assessment factor influenced the final grade, reinforcing the importance of early feedback in identifying students who may need additional support.

Additionally, the model's strong predictive power calls attention to the need for early intervention programs in higher education. Early identification of at-risk students, based on early assessments, can trigger timely academic support and remediation efforts. This approach not only enhances individual student outcomes but also contributes to institutional success by reducing drop-out rates and fostering more personalised learning experiences.

7.3. Practical implications for education

The implications of these findings for educational practice, particularly within preparatory programs serving potentially underprepared or previously disadvantaged students, are profound. By implementing predictive models such as the hybrid TPE-optimised SVR, universities and educational institutions can identify factors influencing student success and intervene early, long before final grades are assigned. For students in enabling programs, who may lack prior academic experience or face additional barriers, this proactive approach could be transformative. Early in the semester, institutions could use predictive insights to offer targeted support to students at risk of low academic performance, such as supplementary tutoring, access to specialised learning resources, or mentorship programs designed to build foundational skills and enhance engagement.

The ability to predict final grades accurately also brings a more data-driven approach to academic advising. Advisers could use model insights to guide their interactions with students, prioritising those flagged as needing additional support. Rather than relying solely on anecdotal or intuitive assessments, advisers could offer customised learning strategies and discuss intervention plans that reflect individual student needs and potential risk factors. For students with historically limited access to academic guidance, this structured, data-informed approach could be especially impactful, creating a safety net that fosters their academic confidence and persistence.

Beyond individual support, the findings offer a path to more strategic resource allocation within educational institutions. By identifying students most likely to benefit from additional resources, institutions could optimise staff time and materials, directing them where they will have the greatest impact. This ensures that resources are used effectively and equitably, supporting students who might otherwise struggle to access essential academic assistance. In the context of preparatory programs, such resource allocation may be crucial in building a more inclusive educational environment that proactively bridges gaps in academic preparedness and promotes long-term student success. Moreover, predictive models enable real-time, dynamic monitoring of student performance, a significant improvement over static, end-of-term assessments. Educators can use these real-time insights to continually adjust teaching strategies, lesson plans, and curricula, thereby fostering a responsive teaching environment that reflects students' evolving needs.

7.4. Implications for pedagogy and educational theories

The integration of predictive modelling into educational settings, particularly within preparatory programs, complements established pedagogical theories aimed at enhancing student-centred learning. Vygotsky's Zone of Proximal Development (ZPD) suggests that students learn best when engaged in tasks that challenge them at an appropriate level of competence (Cai et al., 2024). Predictive models can identify students who are within or outside their ZPD, empowering educators to provide personalised support that aligns with each student's learning needs. For instance, students identified as struggling with early assessments could receive customised guidance and resources to target specific areas of weakness, potentially bridging gaps in foundational knowledge.

In addition to personalising support, these predictive insights can support the creation of adaptive learning environments, tailoring the educational experience to meet diverse student needs. For students in preparatory programs, who may face unique challenges due to underpreparedness, adaptive systems based on models like the SVR can help ensure equitable learning opportunities. By analysing various factors, such as engagement and prior performance, educators can adjust learning activities, support, and assessments to create more responsive, inclusive learning environments. Early identification of at-risk students empowers instructors to intervene proactively, supporting a more equitable distribution of learning resources and minimising the risk of disengagement.

In line with constructivist pedagogy, which values timely, personalised feedback (Huang et al., 2024), predictive models enable a formative assessment approach by flagging students who might benefit from targeted feedback before critical summative assessments. This approach can improve engagement and encourage active participation, fostering a learning environment where students build on prior knowledge and receive support when they need it most. The SVR model's ability to analyse engagement metrics (e.g., interaction frequency with course materials) could enable instructors to monitor motivation and intervene with engagement-centred strategies when needed (Xia et al., 2022). This data-driven approach enables the design of motivational interventions that are tailored to individual engagement patterns, helping educators reduce disengagement.

The SVR model also aligns with mastery learning principles (Alsafadi et al., 2023), offering data-driven insights that allow educators to gauge when students are ready to progress to more advanced topics. Such mastery-based interventions are particularly beneficial for students in enabling programs, as they provide ample opportunities to master foundational skills before moving forward, fostering a more solid and self-paced learning foundation. Additionally, models like the SVR can inform curriculum design by analysing student data to suggest adjustments that minimise cognitive overload. Curriculum elements could be structured based on common areas where students struggle, ensuring content sequences that support comprehension.

From a policy perspective, the SVR model also allows institutions to assess the effectiveness of specific interventions. By comparing predicted outcomes with actual results post-intervention, educators and administrators can refine their approaches and evidence-based practices. In summary, integrating SVR and similar predictive models within preparatory programs could transform educational practice, helping to close achievement gaps and ensure that all students, especially those from disadvantaged backgrounds, receive the targeted support they need to thrive academically.

8. Limitations and future directions

While the proposed explainable hybrid TPE-optimised Support Vector Regression (SVR) model demonstrated promising results in predicting final marks for the TPP7155 course, there are several limitations that should be considered when interpreting the findings.

8.1. Model limitations

The TPE method's tendency to struggle with global exploration can lead to premature convergence to suboptimal solutions, which could hinder its overall effectiveness, especially in larger and more complex datasets. The use of Hyperopt's Tree-structured Parzen Estimator (TPE) algorithm for hyperparameter optimisation, although effective in improving model accuracy, has inherent drawbacks. The TPE algorithm converges more slowly compared to other optimisation techniques such as Bayesian optimisation, leading to longer processing times, particularly when handling large datasets. Additionally, TPE's tendency to converge prematurely to local optima may hinder its ability to fully explore the solution space, potentially limiting the model's performance.

In terms of *x*AI, both LIME and SHAP were used to explain the model's predictions. While LIME provides local explanations, it is sensitive to kernel width parameters, potentially producing inconsistent results. SHAP offers global feature importance but is computationally expensive and assumes feature independence, which may not always hold. Both methods have limitations that must be considered when interpreting model predictions.

Support Vector Regression (SVR) is well-suited for smaller datasets but can become less efficient as the dataset size and dimensionality increase. The optimisation process, requiring numerous iterations to fine-tune accuracy, can be time-intensive and computationally expensive. Future research could explore alternative machine learning algorithms or hybrid approaches, such as deep learning or ensemble methods, that scale more efficiently with larger and more complex datasets. These approaches could better capture the intricate patterns of student behaviour and learning, providing more nuanced predictions of academic success.

8.2. Data limitations and generalisability

A significant limitation of the current study lies in the scope of the dataset. The model was developed using data from a single course (TPP7155) within a specific institution (University of Southern Queensland, Australia) for the 2020-2021 period. While the results are promising, this narrow focus on one course and academic year restricts the generalisability of the findings and the predictive model may perform differently across different academic disciplines, institutions, or cultural contexts. Therefore, future research should aim to validate the model across a wider range of educational contexts, using data from multiple academic years and diverse student populations. Key student attributes, such as gender, socio-economic status, and family background, which are known to influence academic performance, particularly in preparatory learning environments, were excluded. Including such variables would provide a more comprehensive understanding of student outcomes and improve the predictive accuracy of the model.

Another important limitation is the exclusive reliance on assignments and examination marks as the sole predictors of student performance. We had five fixed assignments, Assignment 1 (weight = 5%), Assignment 2 (weight = 15%), Assignment 3 (weight = 10%), Assignment 4 (weight = 20%), and Assignment 5 (weighted = 5%), besides the final examination score (weight = 45%). These assessments collectively contributed to an overall score expressed as a percentage. While these assessments provide valuable information, they fail to capture the dynamic nature of student engagement throughout the semester. The model does not account for changes in student learning behaviours or engagement, which may evolve over time and may not adapt effectively without retraining. To address this limitation, future research could incorporate real-time learning analytics, such as time spent on the Online Learning Management System (LMS), participation in group activities, or engagement with multimedia resources. Ouyang et al. (2023) highlights a course design involving online lectures, group discussions, and collaborative writing, which could be incorporated into prediction models. By integrating continuous learning metrics, such as video watch time or interaction with discussion forums, the model could better reflect ongoing student engagement and its influence on academic success. Further research is needed to evaluate how these additional factors beyond fixed, quantitative assessments, could improve model accuracy and adaptability.

8.3. Model scalability and future directions

Further studies might examine multiple datasets and experiment with deep learning models, such as recurrent neural networks, to capitalise on advancements in technology and access to larger datasets. While this study focused on fixed, quantitative assessments, future research could incorporate diverse data types such as qualitative feedback from both instructors and students, including course satisfaction, student interactions, such as chat logs, and self-reported data (e.g., study habits). A modified methodology, using deep learning and natural language processing (NLP) models, such as those proposed by Dann et al. (2022), could effectively leverage labelled student feedback before each assessment. This approach would not only demonstrate the value of AI in analysing student feedback but could also offer insights into how engagement influences academic outcomes. Human-in-the-loop evaluation methods, wherein instructors and students assess the model's predictions throughout the term, could enhance the model's practical relevance and improve accuracy. Research that combines both quantitative and qualitative data will be essential for developing models that are not only statistically accurate but also reflect the complexities of real-world teaching and learning environments.

The current study focuses solely on a single course, but future research could validate the model across a range of educational contexts. For example, adapting the model for undergraduate or postgraduate programs, or even diverse disciplines, would test its scalability and robustness. Exploring the application of the model in various educational settings could provide valuable insights into the broader applicability of predictive modelling for student success. Additionally, exploring machine learning techniques such as recurrent neural networks (RNNs) and deep learning could improve predictions for outcomes like student retention or dropout. These techniques could also be useful for developing personalised recommendations for students, such as suggesting remedial courses or additional support services based on predictive insights.

8.4. Broader applications of the model

While the current study focuses on academic performance, the predictive model developed could have applications beyond the educational sector where data mining methods are required for pattern recognition and predictive modelling. Among the applications are the humanities, social sciences, health, psychology, cognitive science, and human behaviour, which provide practitioners with tools to analyse trends, forecast outcomes, and make decision-making (such as Liao et al. (2012); Mia et al. (2024); Qin (2024); Lin and Marques (2024)). Due to large datasets compiled from books, articles, and social media, the proposed model may also be used to predict cultural shifts in order to anticipate changes in culture, linguistics, or social behaviour over time, or to analyse historical events in order to predict political or social revolutions by identifying patterns in past events, understanding their causes, and modelling future events.

Similar models have been successfully applied in fields such as healthcare, where early indicators (such as symptoms or medical tests) are used to predict patient outcomes. By adapting this model to different contexts, it could be used to predict the health trajectories of individuals based on early diagnostic indicators. The hybrid SVR model could be used in social services to predict outcomes such as employment success or social reintegration based on early engagement with support programs. Using the proposed predictive model in the social mobility space, we could determine how social policies, economic conditions, and educational opportunities impact social mobility, especially providing a framework for reducing inequality and improving opportunities for disadvantaged groups in policymaking. The potential for this model to be generalised to other domains underscores its versatility and adaptability, positioning it as a tool that can be used for a variety of predictive tasks.

9. Conclusion

This research underscores the promise of advanced predictive modelling, specifically the TPE-optimised SVR approach, for predicting student success. By accurately capturing the impact of key assessment variables, this model offers practical, actionable insights for educational institutions aiming to support student success, particularly within preparatory programs where students often face significant challenges in academic readiness and retention. The findings highlight both the potential of such models to enhance student support and the need for continuous refinement in model generalisability and interpretability to meet the demands of increasingly complex educational data.

The findings of this research highlight several practical implications. First, by identifying specific areas where students struggle, educational institutions can refine preparatory programs to address students' unique needs, reducing their risk of early withdrawal. Furthermore, the positive outcomes observed in this intervention emphasise the value of regular, data-driven evaluations of support programs. Such evaluations enable continuous refinement, ensuring that support remains relevant and responsive to changing student profiles and needs.

Considering the unique challenges facing preparatory students, future research should continue to explore and evaluate interventions aimed at enhancing academic readiness. These efforts are vital in building a supportive framework that can alleviate the barriers to success faced by preparatory students, ultimately improving retention rates and contributing to their long-term academic success. Future research expanding on these methods across diverse contexts could further advance our understanding of the relationship between predictive analytics and educational outcomes, facilitating a more equitable and responsive learning environment.

CRediT authorship contribution statement

Sujan Ghimire: Writing – original draft, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Shahab Abdulla:** Writing – review & editing, Supervision, Resources, Project administration, Conceptualization. **Lionel P. Joseph:** Writing – original draft, Visualization, Validation, Formal analysis. **Salvin Prasad:** Writing – review & editing, Supervision, Software, Resources. **Angela Murphy:** Writing – review & editing. **Aruna Devi:** Writing – review & editing. **Prabal Datta Barua:** Writing – review & editing. **Ravinesh C. Deo:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **Rajendra Acharya:** Writing – review & editing. **Zaher M. Yaseen:** Writing – review & editing, Validation, Software.

Statements on open data and ethics

Student performance data from the TPP7155 course was released under Ethics Application [H18REA236], as approved by the University of Southern Queensland Human Ethics Committee. All data are fully anonymised, with no personal identifiers reported. Participants were informed of the data's use for research purposes, and their privacy rights were strictly observed. Requests for access to the data can be directed to the corresponding author

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

Tables A.9 and A.10 show the hyperparameters and search range of hyperparameters for various models, including the objective model (i.e., SVR) as well as neural network-based models, tree-based models, ensemble-based models and the boosting-based models. It is important to note that, for brevity, only the optimal parameters for Models M1, M4, M6, and M11 are presented. These tables also serve as a reference for understanding the hyperparameter optimisation and the specific configurations used for each model in the study.

References

Ahmed, A. A. M., Deo, R. C., Ghimire, S., Downs, N. J., Devi, A., Barua, P. D., & Yaseen, Z. M. (2022). Introductory engineering mathematics students' weighted score predictions utilising a novel multivariate adaptive regression spline model. *Sustainability*, *14*(17), Article 11070.

Al-safadi, H. A., Shgair, M. S. A., & Al Qatawnih, K. S. (2023). The effectiveness of designing e-learning environment based on mastery learning and artificial intelligence on developing English speaking skills among tenth graders in Palestine. *IUG Journal of Educational & Psychological Studies*, *31*(1).

Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: Literature review and best practices. *International Journal of Educational Technology in Higher Education*, *17*(1), 3.

Apriyadi, M. R., Rini, D. P., et al. (2023). Hyperparameter optimization of support vector regression algorithm using metaheuristic algorithm for student performance prediction. *International Journal of Advanced Computer Science and Applications*, *14*(2).

Arias, A., Linares-Vásquez, M., & Héndez-Puerto, N. R. (2024). Undergraduate dropout in Colombia: A systematic literature review of causes and solutions. *Journal of Latinos and Education*, *23*(2), 612–627.

Arsad, P. M., Buniyamin, N., et al. (2014). Neural network and linear regression methods for prediction of students' academic achievement. In *2014 IEEE global engineering education conference (EDUCON)* (pp. 916–921). IEEE.

Basak, D., Pal, S., Patranabis, D. C., et al. (2007). Support vector regression. *Neural Information Processing-Letters and Reviews*, *11*(10), 203–224.

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. *Advances in Neural Information Processing Systems*, *24*.

Bergstra, J., Komer, B., Eliasmith, C., Yamins, D., & Cox, D. D. (2015). Hyperopt: A python library for model selection and hyperparameter optimization. *Computational Science and Discovery*, *8*(1), Article 014008.

Bertolini, R., Finch, S. J., & Nehm, R. H. (2022). Quantifying variability in predictions of student performance: Examining the impact of bootstrap resampling in data pipelines. *Computers and Education: Artificial Intelligence*, *3*, Article 100067.

Bookallil, C., & Harreveld, B. (2017). Insights into attrition from university-based enabling programs. *Australian Journal of Adult Learning*, *57*(1), 58–81.

Borghi, P. H., Zakordonets, O., & Teixeira, J. P. (2021). A covid-19 time series forecasting model based on mlp ann. *Procedia Computer Science*, *181*, 940–947.

Bradley, D., Noonan, P., Nugent, H., & Scales, B. (2008). *Review of Australian Higher Education*.

Buenaño-Fernández, D., Gil, D., & Luján-Mora, S. (2019). Application of machine learning in predicting performance for computer engineering students: A case study. *Sustainability*, *11*(10), 2833.

Cai, L., Msafiri, M. M., & Kangwa, D. (2024). Exploring the impact of integrating AI tools in higher education using the zone of proximal development. *Education and Information Technologies*, 1–74.

Cao, L.-J., & Tay, F. E. H. (2003). Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on Neural Networks*, *14*(6), 1506–1518.

Chavez, H., Chavez-Arias, B., Contreras-Rosas, S., Alvarez-Rodríguez, J. M., & Raymundo, C. (2023). *Artificial neural network model to predict student performance using nonpersonal information. Frontiers in education: Vol. 8*. Frontiers (p. 1106679).

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., et al. (2015). Xgboost: Extreme gradient boosting. R package version 0.4-2, *1*(4), 1–4.

Cheng, B., Liu, Y., & Jia, Y. (2024). Evaluation of students' performance during the academic period using the xg-boost classifier-enhanced aeo hybrid model. *Expert Systems with Applications*, *238*, Article 122136.

Table A.9
The architecture of the proposed SVR model, including the Neural Network and the tree-based models developed for predicting TPP7155 General Science course overall score. Optimal hyperparameter for 4 models with different input combinations ((M1, M4, M6 and M11) are also shown. *Note:* $ReLU$ = *Rectified Linear Units;* $Adam$ = *Adaptive Moment Estimation,* $gbdt$ = *traditional Gradient Boosting Decision Tree,* rbf = *Radial Basis Function,* $logistic$ = *Logistic Sigmoid Function,* $tanh$ = *Hyperbolic Tangent Activation Function,* $sigmoid$ = *sigmoid Activation Function,* $hp.choice$ = *Selects a value from a list of discrete options,* $hp.uniform$ = *Samples a continuous value uniformly between two bounds,* $hp.quniform$ = *Samples a continuous value uniformly between two bounds and then quantizes it to a discrete step size,* $hp.loguniform$ = *Samples a value uniformly in log-space (useful for parameters spanning several orders of magnitude,* $hp.randint$ = *Samples an integer uniformly between two bounds (inclusive lower bound, exclusive upper bound,* $hp.lognormal$ = *Samples a value from a log-normal distribution and* $hp.scope$ = *Defines parameter expressions using Python's standard scope, allowing for complex parameter transformations and dependencies.*

	Model	Hyperparameter	Search range	Optimal value			
				M1	M4	M6	M11
Objective Model	**Support Vector Regression (SVR)**	Kernel	hp.choice('kernel', ['rbf']),	**rbf**	**rbf**	**rbf**	**rbf**
		C	hp.loguniform ('C', -40,10),	**551.745**	**321.103**	**84485.910**	**3229.949**
		gamma	hp.loguniform('gamma', -40,10),	**4.223**	**0.148**	**0.000**	**0.004**
		epsilon	hp.quniform('epsilon', 1e-5,1e-1,1e-1),	**0.100**	**0.100**	**0.100**	0.100
Neural Network-Based Models	**Multi-layer Perceptron (MLP)**	Hiddden Layer Size	50 + hp.randint('hidden_layer_sizes', 100),	**37.000**	**99.000**	**41.000**	**26.000**
		Activation	hp.choice('activation', ['tanh', 'relu','logistic']),	**relu**	**relu**	**relu**	**relu**
		Solver	hp.choice('solver', ['adam']),	**Adam**	**Adam**	**Adam**	**Adam**
		L2 regularization term	hp.uniform('alpha', 0.05, 1.0),	**0.129**	**0.648**	**0.920**	**0.133**
		Learning rate	hp.uniform('learning_rate_init', 0.01, 0.1)	**0.074**	**0.051**	**0.023**	**0.021**
	Extreme Learning Machine (ELM)	Hiddden Layer Size	hp.choice('num_hidden_units', np.arange(20,500,2, dtype=int)	**20.000**	**22.000**	**54.000**	**20.000**
		Activation	hp.choice('num_input_nodes', ['sigmoid','hardlimit','fourier']	**Sigmoid**	**Sigmoid**	**Sigmoid**	**Sigmoid**
	Deep Neural Network (DNN)	Dropout rate	hp.uniform('rate',0.0001,0.3),	**0.042**	**0.172**	**0.101**	**0.190**
		Dense Unit	scope.int(hp.quniform ('units',10,100,5)),	**20.000**	**90.000**	**95.000**	**80.000**
		Batch Size	scope.int(hp.quniform ('batch_size',20,200,20)),	**20.000**	**20.000**	**40.000**	**80.000**
		Number of Layers	scope.int(hp.quniform ('layers',1,3,1))	**2.000**	**3.000**	**3.000**	**3.000**
Tree- Based Models	**Decision Tree (DT)**	Maximum Depth	hp.choice('max_depth', range(1,200,1))	**2.000**	**6.000**	**192.000**	**138.000**
	Extra Tree Regressor (ETR)	Number of Estimators	hp.choice('n_estimators', range(5,800,2)),	**34.000**	**134.000**	**22.000**	**92.000**
		Maximum Depth	hp.choice('max_depth', range(1,110,1))	**3.000**	**7.000**	**16.000**	**22.000**
	Random Forest Regressor (RFR)	Nmber of Estimators	hp.choice('n_estimators', range(5,800,2)),	**150.000**	**3.000**	**23.000**	**108.000**
		Maximum Depth	hp.choice('max_depth', range(1,110,1)),	**104.000**	**38.000**	**19.000**	**100.000**
		Minimm Sample Leaf	hp.choice(' min_samples_leaf', range(1,100,1)),	**21.000**	**2.000**	**0.000**	**0.000**

Chesters, J., & Watson, L. (2013). Understanding the persistence of inequality in higher education: Evidence from Australia. *Journal of Education Policy*, *28*(2), 198–215.

Chesters, J., Rutter, K., Nelson, K., & Watson, L. (2018). Alternative pathways into university: Are tertiary preparation programs a viable option?. *The Australian Universities' Review*, *60*(1), 35–44.

Cooper, N., Ellis, B., & Sawyer, J. (2000). Expanded future opportunities provided by a bridging course at a regional university campus. Ph.D. thesis. Queensland University of Technology.

Dann, C., Redmond, P., Fanshawe, M., Brown, A., Getenet, S., Shaik, T., Tao, X., Galligan, L., & Li, Y. (2022). Making sense of student feedback and engagement using artificial intelligence. *Australasian Journal of Educational Technology*.

de Oliveira, R. S., Reis Jr, A. S., & Sperandio Nascimento, E. G. (2022). Predicting the number of days in court cases using artificial intelligence. *PLoS ONE*, *17*(5), Article e0269008.

del Pozo-Bueno, D., Kepaptsoglou, D., Peiró, F., & Estradé, S. (2023). Comparative of machine learning classification strategies for electron energy loss spectroscopy: Support vector machines and artificial neural networks. *Ultramicroscopy*, *253*, Article 113828.

Deng, Y., Zhu, Z., Shen, Z., Wang, Y., & Xu, Y. (2023). An online forecasting method of remaining useful life of lithium-ion batteries based on tpe-svr. *Applied Mathematics, Modeling and Computer Simulation*, 100–109.

Deo, R. C., Ghimire, S., Downs, N. J., & Raj, N. (2018). Optimization of windspeed prediction using an artificial neural network compared with a genetic programming model. In *Handbook of research on predictive modeling and optimization methods in science and engineering* (pp. 328–359). IGI Global.

Deo, R. C., Yaseen, Z. M., Al-Ansari, N., Nguyen-Huy, T., Langlands, T. A. M., & Galligan, L. (2020). Modern artificial intelligence model development for undergraduate student performance prediction: An investigation on engineering mathematics courses. *IEEE Access*, *8*, 136697–136724.

Deo, R. C., Grant, R. H., Webb, A., Ghimire, S., Igoe, D. P., Downs, N. J., Al-Musaylh, M. S., Parisi, A. V., & Soar, J. (2022). Forecasting solar photosynthetic photon flux density under cloud cover effects: Novel predictive model using convolutional neural network integrated with long short-term memory network. *Stochastic Environmental Research and Risk Assessment*, *36*(10), 3183–3220.

Table A.10
The architecture of the Ensemble-based and the Boosting-based models developed to predict TPP7155 (General Science) course students' overall scores.

	Model	Hyperparameter	Seach Range	Optimal Value			
				M1	M4	M6	M11
Ensemble-Based Methods	**Extreme Gradient Boosting (XGB)**	Number of Estimators	hp.choice('n_estimators', range(50,500,2)),	**167.000**	**173.000**	**214.000**	**149.000**
		Learning rate	hp.uniform('learning_rate', 0.01, 0.1),	**0.085**	**0.034**	**0.050**	**0.059**
		Maximum Depth	hp.choice('max_depth', range(2,10,1)),	**4.000**	**6.000**	**7.000**	**5.000**
		min_child_ weight	hp.choice('min_child_weight', range(1,50,1)),	**23.000**	**10.000**	**12.000**	4.000
		subsample	hp.uniform('subsample',0.5, 1.0),	**0.582**	**0.750**	**0.604**	**0.500**
		colsample_bytree	hp.uniform('colsample_ bytree', 0.6, 1.0)	**0.641**	**0.626**	**0.811**	**0.762**
		L2 regularization term	hp.uniform('reg_alpha',;0, 1.0),	**0.643**	**0.116**	**0.649**	**0.360**
	Light Gradient Boosting (LGB)	Number of Estimators	hp.choice('n_estimators', range(50,500,2)),	**94.000**	**38.000**	**209.000**	**142.000**
		min_child_ weight	hp.uniform('min_child_ weight', 0.001, 0.2),	**0.032**	**0.163**	**0.103**	**0.187**
		min_child_ samples	hp.choice('min_child_samples', range(5,51,5)),	**1.000**	**1.000**	**2.000**	**1.000**
		lgb_colsample_ bytree	hp.uniform('lgb_colsample_ bytree', 0.6;1.0),	**0.960**	**0.601**	**0.743**	**0.818**
		subsample	hp.uniform('subsample';0.5, 1.0),	**0.832**	**0.999**	**0.988**	**0.730**
		Learning rate	hp.uniform('learning_rate', 0.01, 0.3),	**0.050**	**0.066**	**0.107**	**0.157**
		Maximum Depth	hp.choice('max_depth', range(2,10,1)),	**5.000**	**2.000**	**7.000**	**5.000**
		number of Leaves	hp.choice('num_leaves', range(2, 50, 1)),	**0.000**	**42.000**	**3.000**	**30.000**
		L2 regularization term	hp.uniform('reg_alpha'0, 1.0)	**0.643**	**0.765**	**0.297**	**0.815**
Boosting-Based Method	**AdaBoost Regressor (ADBR)**	Learning rate	hp.uniform('learning_rate', 0.0001, 0.3),	**0.105**	**0.166**	**0.235**	**0.195**
		Loss	hp.choice ('loss', ['linear', 'square' ,'exponential']),	**square**	**square**	**exponential**	**linear**
		Number of Estimators	hp.choice('n_estimators', range(5,800,2))	**10.000**	**375.000**	**92.000**	**20.000**
	Bagging Regressor (BGR)	Number of Estimators	hp.choice('n_estimators', range(40, 800,20))	**240.000**	**160.000**	**60.000**	**80.000**
		bootstrap	True	**True**	**True**	**True**	**True**
		Estimator	Decision Tree Regressor	**DTR**	**DTR**	**DTR**	**DTR**
	Gradient Boosting Regressor(GBR)	Number of Estimators	hp.choice('n_estimators', range(5,800,2)),	**71.000**	**347.000**	**713.000**	**351.000**
		Learning rate	hp.uniform('learning_rate', 0.0001, 0.3),	**0.045**	**0.060**	**0.089**	**0.222**
		Maximum Depth	hp.choice('max_depth', range(1,110,1))	**1.000**	**4.000**	**3.000**	**1.000**

Dewi, K., & Widiastuti, N. (2020). *Support vector regression for gpa prediction. IOP conference series: Materials science and engineering: Vol. 879*. IOP Publishing (p. 012112).

Dieber, J., & Kirrane, S. (2020). Why model why? Assessing the strengths and limitations of lime. arXiv preprint. arXiv:2012.00093.

Djurisic, V., Kascelan, L., Rogic, S., & Melovic, B. (2020). Bank crm optimization using predictive classification based on the support vector machine method. *Applied Artificial Intelligence*, *34*(12), 941–955.

Dubey, C. (2024). Exploring the role of artificial intelligence in inclusive education (p. 287).

Edwards, D., & McMillan, J. (2015). Completing university in a growing sector: Is equity an issue?

Elsayed, S., Thyssens, D., Rashed, A., Jomaa, H. S., & Schmidt-Thieme, L. (2021). Do we really need deep learning models for time series forecasting? arXiv preprint. arXiv: 2101.02118.

Evangelista, E. (2023). An optimized bagging ensemble learning approach using bestrees for predicting students' performance. *International Journal of Emerging Technologies in Learning (iJET)*, *18*(10), 150.

Ghimire, S., Deo, R. C., Raj, N., & Mi, J. (2019). Wavelet-based 3-phase hybrid svr model trained with satellite-derived predictors, particle swarm optimization and maximum overlap discrete wavelet transform for solar radiation prediction. *Renewable & Sustainable Energy Reviews*, *113*, Article 109247.

Ghimire, S., Bhandari, B., Casillas-Perez, D., Deo, R. C., & Salcedo-Sanz, S. (2022). Hybrid deep cnn-svr algorithm for solar radiation prediction problems in Queensland, Australia. *Engineering Applications of Artificial Intelligence*, *112*, Article 104860.

Ghimire, S., Deo, R. C., Casillas-Perez, D., & Salcedo-Sanz, S. (2022). Improved complete ensemble empirical mode decomposition with adaptive noise deep residual model for short-term multi-step solar radiation prediction. *Renewable Energy*, *190*, 408–424.

Ghimire, S., Nguyen-Huy, T., AL-Musaylh, M. S., Deo, R. C., Casillas-Pérez, D., & Salcedo-Sanz, S. (2023b). A novel approach based on integration of convolutional neural networks and echo state network for daily electricity demand prediction. *Energy*, *275*, Article 127430.

Ghimire, S., Nguyen-Huy, T., Prasad, R., Deo, R. C., Casillas-Pérez, D., Salcedo-Sanz, S., & Bhandari, B. (2023a). Hybrid convolutional neural network-multilayer perceptron model for solar radiation prediction. *Cognitive Computation*, *15*(2), 645–671.

Ghimire, S., Deo, R. C., Casillas-Pérez, D., Salcedo-Sanz, S., Pourmousavi, S. A., & Acharya, U. R. (2024). Probabilistic-based electricity demand forecasting with hybrid convolutional neural network-extreme learning machine model. *Engineering Applications of Artificial Intelligence*, *132*, Article 107918.

Guenther, J., & Johnson, P. (2010). Engaging pre-tertiary students with low English literacy using online technologies.

Hakkal, S., & Lahcen, A. A. (2024). Xgboost to enhance learner performance prediction. *Computers and Education: Artificial Intelligence*, Article 100254.

Halde, R. R. (2016). Application of machine learning algorithms for betterment in education system. In *2016 international conference on automatic control and dynamic optimization techniques (ICACDOT)* (pp. 1110–1114). IEEE.

Hashim, A. S., Awadh, W. A., & Hamoud, A. K. (2020). *Student performance prediction model based on supervised machine learning algorithms. IOP conference series: Materials science and engineering: Vol. 928*. IOP Publishing (p. 032019).

Hodges, B., Bedford, T., Hartley, J., Klinger, C., Murray, N., O'Rourke, J., & Schofield, N. (2013). Enabling retention: Processes and strategies for improving student retention in university-based enabling programs. Final report 2013. Australian Government Office for Learning and Teaching.

Holzinger, A. (2021). The next frontier: AI we can really trust. In *Joint European conference on machine learning and knowledge discovery in databases* (pp. 427–440). Springer.

Holzinger, A., Dehmer, M., Emmert-Streib, F., Cucchiara, R., Augenstein, I., Del Ser, J., Samek, W., Jurisica, I., & Díaz-Rodríguez, N. (2022). Information fusion as an integrative cross-cutting enabler to achieve robust, explainable, and trustworthy medical artificial intelligence. *Information Fusion*, *79*, 263–278.

Hu, Y., Mello, R. F., & Gašević, D. (2021). Automatic analysis of cognitive presence in online discussions: An approach using deep learning and explainable artificial intelligence. *Computers and Education: Artificial Intelligence*, *2*, Article 100037.

Huang, Z., Mao, Y., & Zhang, J. (2024). The influence of artificial intelligence technology on college students' learning effectiveness from the perspective of constructivism—taking chatgpt as an example. *Journal of Education, Humanities and Social Sciences*, *30*, 40–46.

Hussain, S., & Khan, M. Q. (2023). Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. *Annals of Data Science*, *10*(3), 637–655.

Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. (2020). Systematic ensemble model selection approach for educational data mining. *Knowledge-Based Systems*, *200*, Article 105992.

Jang, J., Choi, S., Jung, H., & Kim, H. (2022). Practical early prediction of students' performance using machine learning and explainable AI. *Education and Information Technologies*, 12855–12889.

Jayasinghe, W. L. P., Deo, R. C., Ghahramani, A., Ghimire, S., & Raj, N. (2022). Development and evaluation of hybrid deep learning long short-term memory network model for pan evaporation estimation trained with satellite and ground-based data. *Journal of Hydrology*, *607*, Article 127534.

Jiang, B.-N., Zhang, Y.-Y., Zhang, Z.-Y., Yang, Y.-L., & Song, H.-L. (2024). Tree-structured parzen estimator optimized-automated machine learning assisted by meta–analysis for predicting biochar–driven n2o mitigation effect in constructed wetlands. *Journal of Environmental Management*, *354*, Article 120335.

Joseph, L. P., Deo, R. C., Prasad, R., Salcedo-Sanz, S., Raj, N., & Soar, J. (2023). Near real-time wind speed forecast model with bidirectional lstm networks. *Renewable Energy*, *204*, 39–58.

Junaid, M., Ali, S., Eid, F., El-Sappagh, S., & Abuhmed, T. (2023). Explainable machine learning models based on multimodal time-series data for the early detection of Parkinson's disease. *Computer Methods and Programs in Biomedicine*, *234*, Article 107495.

Ketkar, N., & Ketkar, N. (2017). Introduction to keras. *Deep learning with python: A hands-on introduction* (pp. 97–111).

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *3*, Article 100074.

Kocsis, Á., & Molnár, G. (2024). Factors influencing academic performance and dropout rates in higher education. *Oxford Review of Education*, 1–19.

Komer, B., Bergstra, J., & Eliasmith, C. (2019). Hyperopt-sklearn. *Automated machine learning: Methods, systems, challenges* (pp. 97–111).

Kouassi, K. H., & Moodley, D. (2020). An analysis of deep neural networks for predicting trends in time series data. In *Artificial intelligence research: First southern African conference for AI research* (pp. 119–140). Springer.

Li, H., Zhang, C., Zhang, S., & Atkinson, P. M. (2020). Crop classification from full-year fully-polarimetric l-band uavsar time-series using the random forest algorithm. *International Journal of Applied Earth Observation and Geoinformation*, *87*, Article 102032.

Li, I. W., & Jackson, D. (2024). Influence of entry pathway and equity group status on retention and the student experience in higher education. *Higher Education*, *87*(5), 1411–1431.

Li, Y., Bao, T., Gong, J., Shu, X., & Zhang, K. (2020). The prediction of dam displacement time series using stl, extra-trees, and stacked lstm neural network. *IEEE Access*, *8*, 94440–94452.

Liao, S.-H., Chu, P.-H., & Hsiao, P.-Y. (2012). Data mining techniques and applications–a decade review from 2000 to 2011. *Expert Systems with Applications*, *39*(12), 11303–11311.

Lin, C. Y., & Marques, J. A. L. (2024). Stock market prediction using artificial intelligence: A systematic review of systematic reviews. *Social Sciences & Humanities Open*, *9*, Article 100864.

Lisciandro, J. G., & Gibbs, G. (2016). 'OnTrack' to university: Understanding mechanisms of student retention in an Australian pre-university enabling program. *Australian Journal of Adult Learning*, *56*(2), 198–224.

Liu, L., & Bai, Q. (2023). Under the background of ideological and political education, the path optimization of college students' consumption outlook education based on adaboost model. *Applied Mathematics and Nonlinear Sciences*.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56–67.

Malakouti, S. M. (2023). Utilizing time series data from 1961 to 2019 recorded around the world and machine learning to create a global temperature change prediction model. *Case Studies in Chemical and Environmental Engineering*, *7*, Article 100312.

Man, X., & Chan, E. (2021). The best way to select features? Comparing mda, lime, and shap. *The Journal of Financial Data Science Winter*, *3*(1), 127–139.

Ma'sum, M. A. (2022). Predicting student achievement before final exam: A regression-based approach. In *2022 2nd international conference on information technology and education (ICIT&E)* (pp. 424–428). IEEE.

McKenney, S., & Mor, Y. (2015). Supporting teachers in data-informed educational design. *British Journal of Educational Technology*, *46*(2), 265–279.

Mia, M. T., Ferdus, M. Z., Rahat, M. A. R., Anjum, N., Siddiqua, C. U., & Raju, M. A. H. (2024). A comprehensive review of text mining approaches for predicting human behavior using deep learning method. *Journal of Computer Science and Technology Studies*, *6*(1), 170–178.

Muhamedyev, R., Yakunin, K., Kuchin, Y., Symagulov, A., Buldybayev, T., Murzakhmetov, S., & Abdurazakov, A. (2020). The use of machine learning "black boxes" explanation systems to improve the quality of school education. *Cogent Engineering*, *7*(1), Article 1769349.

Nguyen-Huy, T., Deo, R. C., Khan, S., Devi, A., Adeyinka, A. A., Apan, A. A., & Yaseen, Z. M. (2022). Student performance predictions for advanced engineering mathematics course with new multivariate copula models. *IEEE Access*, *10*, 45112–45136.

Omotehinwa, T. O., Oyewola, D. O., & Dada, E. G. (2023). A light gradient-boosting machine algorithm with tree-structured parzen estimator for breast cancer diagnosis. *Healthcare Analytics*, *4*, Article 100218.

Onyema, E. M., Eucheria, N. C., Obafemi, F. A., Sen, S., Atonye, F. G., Sharma, A., & Alsayed, A. O. (2020). Impact of coronavirus pandemic on education. *Journal of Education and Practice*, *11*(13), 108–121.

Ouyang, F., Wu, M., Zheng, L., Zhang, L., & Jiao, P. (2023). Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education*, *20*(1), 4.

Pang, Y., Judd, N., O'Brien, J., & Ben-Avie, M. (2017). Predicting students' graduation outcomes through support vector machines. In *2017 IEEE frontiers in education conference (FIE)* (pp. 1–8). IEEE.

Paolucci, C., Vancini, S., Bex II, R. T., Cavanaugh, C., Salama, C., & de Araujo, Z. (2024). A review of learning analytics opportunities and challenges for k-12 education. *Heliyon*, *10*(4), Article e25767.

Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the difficulty of training recurrent neural networks. In *International conference on machine learning, PMLR* (pp. 1310–1318).

Priyambada, S. A., Usagawa, T., & Mahendrawathi, E. (2023). Two-layer ensemble prediction of students' performance using learning behavior and domain knowledge. *Computers and Education: Artificial Intelligence*, *5*, Article 100149.

Qin, X. (2024). Intelligent technologies and methodological transformations in the social sciences. *Chinese Political Science Review*, *9*(1), 1–17.

Rai, S., Shastry, K. A., Pratap, S., Kishore, S., Mishra, P., & Sanjay, H. (2021). *Machine learning approach for student academic performance prediction. Evolution in computational intelligence: Frontiers in intelligent computing: Theory and applications (FICTA 2020): Vol. 1*. Springer (pp. 611–618).

Ramedani, Z., Omid, M., Keyhani, A., Shamshirband, S., & Khoshnevisan, B. (2014). Potential of radial basis function based support vector regression for global solar radiation prediction. *Renewable & Sustainable Energy Reviews*, *39*, 1005–1011.

Ramon, Y., Martens, D., Provost, F., & Evgeniou, T. (2020). A comparison of instance-level counterfactual explanation algorithms for behavioral and textual data: Sedc, lime-c and shap-c. *Advances in Data Analysis and Classification*, *14*, 801–819.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135–1144).

Saini, M., Sengupta, E., Singh, M., Singh, H., & Singh, J. (2023). Sustainable development goal for quality education (sdg 4): A study on sdg 4 to extract the pattern of association among the indicators of sdg 4 employing a genetic algorithm. *Education and Information Technologies*, *28*(2), 2031–2069.

Samsudin, N. A. M., Shaharudin, S. M., Sulaiman, N. A. F., Ismail, S., Mohamed, N. S., & Husin, N. H. M. (2022). Prediction of student's academic performance during online learning based on regression in support vector machine. *International Journal of Information and Education Technology*, *12*(12), 1431–1435.

Santos, R. M., & Henriques, R. (2023). Accurate, timely, and portable: Course-agnostic early prediction of student performance from lms logs. *Computers and Education: Artificial Intelligence*, *5*, Article 100175.

Sharafati, A., Haji Seyed Asadollah, S. B., Motta, D., & Yaseen, Z. M. (2020). Application of newly developed ensemble machine learning models for daily suspended sediment load prediction and related uncertainty analysis. *Hydrological Sciences Journal*, *65*(12), 2022–2042.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(1), 1929–1958.

Tao, S., Peng, P., Li, Y., Sun, H., Li, Q., & Wang, H. (2024). Supervised contrastive representation learning with tree-structured parzen estimator Bayesian optimization for imbalanced tabular data. *Expert Systems with Applications*, *237*, Article 121294.

Taylan, O., & Karagözoğlu, B. (2009). An adaptive neuro-fuzzy model for prediction of student's academic performance. *Computers & Industrial Engineering*, *57*(3), 732–741.

Tekgöz, H., Omurca, S. I., & Koc, K. Y. (2022). Estimation of remaining useful life based on time series analysis. In *2022 7th international conference on computer science and engineering (UBMK)* (pp. 273–277). IEEE.

TensorFlow, D. (2018). Tensorflow, site oficial.

Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers and Education*, *143*, Article 103676.

Vasanthanageswari, P. (2022). Improving svm classifier model using tree structured parzen estimator optimization for crop prediction. *Journal of Theoretical and Applied Information Technology*, *100*(22), 6808–6818.

Vultureanu-Albişi, A., & Bădică, C. (2021). Improving students' performance by interpretable explanations using ensemble tree-based approaches. In *2021 IEEE 15th international symposium on applied computational intelligence and informatics (SACI)* (pp. 215–220). IEEE.

Wahyuningsih, T., Iriani, A., Purnomo, H. D., & Sembiring, I. (2024). Predicting students' success level in an examination using advanced linear regression and extreme gradient boosting. *Computer Science and Information Technologies*, *5*(1), 29–37.

Xia, Q., Chiu, T. K., Lee, M., Sanusi, I. T., Dai, Y., & Chai, C. S. (2022). A self-determination theory (sdt) design approach for inclusive and diverse artificial intelligence (AI) education. *Computers and Education*, *189*, Article 104582.

Xu, S., Han, C., & Ran, C. (2021). A time series combined forecasting model based on prophet-lgbm. In *2021 2nd international conference on artificial intelligence and information systems* (pp. 1–6).

Yadav, R. S., & Singh, V. P. (2011). Modeling academic performance evaluation using soft computing techniques: A fuzzy logic approach. *International Journal on Computer Science and Engineering*, *3*(2), 676–686.

Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 11.

Zając, T., Perales, F., Tomaszewski, W., Xiang, N., & Zubrick, S. R. (2024). Student mental health and dropout from higher education: An analysis of Australian administrative data. *Higher Education*, *87*(2), 325–343.

Zhang, F., & O'Donnell, L. J. (2020). Support vector regression. In *Machine learning* (pp. 123–140). Elsevier.

Zhang, J., Zheng, Y., Qi, D., Li, R., & Yi, X. (2016). DNN-based prediction model for spatio-temporal data. In *Proceedings of the 24th ACM SIGSPATIAL international conference on advances in geographic information systems* (pp. 1–4).