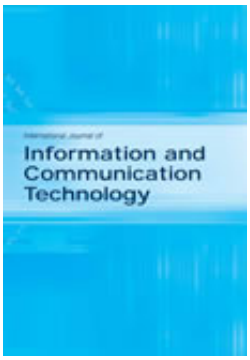
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

The prediction model of higher vocational students' classroom participation based on the fusion of deep learning and support vector machine

Yixuan Qiang

DOI: 10.1504/IJICT.2026.10075999

Article History:

Received:	27 August 2025
Last revised:	17 September 2025
Accepted:	19 September 2025
Published online:	11 February 2026

The prediction model of higher vocational students' classroom participation based on the fusion of deep learning and support vector machine

Yixuan Qiang

School of Marxism,
Jiangsu Vocational Institute of Commerce,
Nanjing, 211168, China
Email: YixuanQiang@outlook.com

Abstract: Student engagement in vocational classrooms is a critical metric for assessing teaching effectiveness and talent development. To address the limitations of conventional assessment methods, we propose a hybrid deep learning-support vector machine (SVM) model for predicting participation levels. The approach integrates convolutional neural networks (CNN) and long short-term memory (LSTM) networks to extract high-dimensional temporal features from classroom videos and behavioural logs. These features are combined with traditional statistical indicators and classified using SVM through a feature-level fusion strategy. Evaluated on simulated vocational classroom data, the fused model achieves 92.3% accuracy and an F1-score of 0.914, significantly outperforming standalone CNN-LSTM or SVM models. This model enables real-time, quantitative assessment of classroom engagement and supports timely teaching interventions.

Keywords: classroom participation; deep learning; support vector machine; SVM; feature fusion; vocational education.

Reference to this paper should be made as follows: Qiang, Y. (2026) 'The prediction model of higher vocational students' classroom participation based on the fusion of deep learning and support vector machine', *Int. J. Information and Communication Technology*, Vol. 27, No. 9, pp.1–17.

Biographical notes: Yixuan Qiang received her Bachelor of Law (LL.B.) degree from Sichuan Normal University in 2011, Master's in Law from Sichuan Normal University in 2017. She is currently working as a Lecturer at Jiangsu Vocational Institute of Commerce. Her research focuses on innovative practices in ideological and political education, as well as the integration of ideological and political education with psychology.

1 Introduction

As educational digitalisation advances, the demand for data mining and analysis in the educational field is increasing daily. Vocational colleges and undergraduate programs differ significantly in student foundation, learning habits, and teaching methods. Vocational college students typically possess a broader range of professional backgrounds and more practical learning goals, preferring hands-on practice and skills development. This unique characteristic places distinct demands on classroom

engagement assessment: varying levels of foundational knowledge lead to significant variations in engagement; learning habits favour practical and interactive learning over purely theoretical learning; and teaching methods emphasise hands-on practice and group collaboration, requiring assessment models that can process complex classroom interaction data. In higher vocational education, students' classroom participation is one of the important factors affecting their learning (Wang et al., 2024). Accurately predicting students' classroom participation can help teachers adjust their teaching strategies promptly, improve teaching quality, and promote students' learning and development (Mustapa et al., 2015). Recently, AI has gained traction in education, with deep learning (DL) and SVM proving especially powerful for predictive tasks.

DL and support vector machines (SVMs), as two mainstream technical approaches in the field of machine learning, have made significant breakthroughs in affective computing, behaviour recognition, and educational data mining in recent years (Xiong et al., 2025). DL has obvious advantages in automatic feature extraction and nonlinear relationship modelling (Monaco et al., 2025), and can use convolutional neural network (CNN) to perform high-dimensional semantic encoding of facial expressions and gestures in classroom videos; SVMs have stronger generalisation ability and interpretability in small sample and high-dimensional sparse scenarios, and are suitable for dealing with the practical problems of "large data noise and unbalanced positive and negative samples" in higher vocational classrooms. However, it is often challenging for a single algorithm to strike a balance between accuracy, robustness, and interpretability. Therefore, this study proposes to integrate DL and SVMs to build a classroom participation prediction model for higher vocational scenarios, aiming to establish a new teaching evaluation paradigm of "data-driven, model interpretability, and timely intervention".

DL is a neural-network-based approach (Yang et al., 2025), DL can automatically learn feature representations in data by constructing a multi-layer neural network structure, thereby enabling recognition and classification of complex data patterns (Dagasso et al., 2025). In education, DL has proven effective for forecasting student performance and examining learning patterns. DL models, such as CNN and recurrent neural networks (RNN), and their variants, including long short-term memory networks (LSTM), can effectively process time series information and image data in educational datasets (Aruleba and Sun, 2025).

SVM is a statistically grounded classification algorithm (Xiao et al., 2025), which categorises data into distinct categories (Mu and Zhao, 2025) by identifying the optimal hyperplane of separation. SVM has significant advantages in dealing with small samples, nonlinear, and high-dimensional data, and has been widely used in various scenarios, such as student classification and curriculum evaluation in the field of education (Birthriya et al., 2025). However, SVM may have the problem of low computational efficiency when dealing with large-scale datasets (Devi and Kaushik, 2025). In this study, data on classroom engagement among higher vocational students often consist of small samples, and data collection and annotation are expensive. The Softmax classification layer of DL is prone to overfitting on small sample data. However, SVMs optimise classification boundaries by maximising margins, demonstrating good generalisation capabilities for small sample data. Furthermore, when processing high-dimensional features, SVMs map data into a high-dimensional space using kernel functions (such as RBFs) to find the optimal classification hyperplane, effectively handling complex relationships and reducing computational complexity. Therefore, SVMs excel in situations with small sample sizes and high-dimensional feature data, making them

suitable for analysing higher vocational student classroom engagement data and improving model performance.

To fully leverage the advantages of DL and SVMs and compensate for the limitations of a single model, researchers have begun to explore methods for integrating the two. On the one hand, DL models can extract high-dimensional features of data to provide more efficient input for SVMs, thereby improving classification performance; On the other hand, the optimisation theory and classification capabilities of SVMs can guide DL models and enhance their generalisation capabilities. This fusion method has achieved promising results in various fields; however, research on predicting classroom participation among higher vocational students is relatively limited.

Higher vocational education places a strong emphasis on cultivating practical abilities and vocational skills, and students' classroom participation has a direct impact on their mastery of knowledge and skills (Qianyi and Zhiqiang, 2024)]. By establishing an effective classroom participation prediction model, teachers can anticipate students' learning status, identify potential learning issues promptly, and implement targeted teaching interventions (Zhao and Yu, 2024). It boosts students' motivation and outcomes while driving overall quality gains in vocational education (Li et al., 2024).

This paper rigorously respects student privacy and data ethics when utilising classroom monitoring and prediction technologies. The classroom videos and behaviour logs involved contain sensitive personal information, so data collection, storage, and use strictly adhere to legal and ethical guidelines. During data collection, students' explicit consent is obtained, and their data are anonymised. Encryption technology is used during data storage to prevent leakage or unauthorised access. To mitigate privacy and ethical risks, the following measures are implemented in model design and application: differential privacy is used to protect privacy by adding noise; rigorous cross-validation is employed to ensure model fairness and transparency; and collaboration with educational institutions is undertaken to establish data management and oversight mechanisms to safeguard the legal use of data and protect student rights.

Despite the widespread application of DL and SVMs in education, existing methods still have shortcomings in assessing classroom engagement in higher vocational education. First, the diverse knowledge levels and learning habits of higher vocational students require robust assessment models. Second, the diverse and complex teaching methods in higher vocational education require models that can handle multimodal data, while most existing methods can only handle single-modal data. Finally, existing methods lack interpretability and fail to meet the demand for model interpretability in educational applications. Based on the above analysis, the model needs to be more robust to accommodate the diverse knowledge levels and learning habits of higher vocational students. Secondly, the model needs to be well-interpretable so that teachers can adjust their teaching strategies based on the model's results. This paper proposes an intelligent prediction model for classroom engagement that integrates DL and SVM. This model utilises a feature-level fusion strategy to combine high-dimensional time series features extracted by CNN and LSTM with the robust classification capabilities of SVM for traditional statistical features. The principal contributions are outlined below:

1 Using CNN and LSTM to extract high-dimensional time series features from classroom videos and behaviour logs.

2 Robust classification of traditional statistical features combined with SVM. Finally, the advantages of the two types of models are integrated by a feature-level fusion strategy.

3 Integrate the advantages of the two types of models through feature-level fusion strategies.

2 Related theoretical knowledge

2.1 Deep learning

DL forms a key branch of machine learning (Spanos et al., 2025). It employs multilayer neural networks to abstract and represent data progressively (Treeprapin et al., 2025). Unlike shallow methods, these models autonomously extract features from low-level edges to high-level semantics (Esatyana and Sakhaee-Pour, 2025). After Hinton introduced layer-wise pre-training in 2006, GPU parallel computing and big data have further accelerated architectures like CNN, RNN and Transformer.

CNN is a deep-learning framework widely adopted for image recognition, classification and processing. Its two pillars are 'local connectivity' and 'weight sharing'. Rather than full connection, CNN detects local spatial patterns with far fewer weights and, via layered stacking, yields high-level traits like translation invariance and rotation robustness. Its core components are convolutional, pooling and fully-connected layers. In particular, convolutional layers apply kernels to input data to harvest local features. The process of the convolution operation is shown in formula (1):

$$x_j^j = f\left(\sum_{i=1} x_i^{j-1} \times k_{ij}^j + b_j^j\right) \tag{1}$$

where x_i^{j-1} represents input of j-1$^{\text{th}}$ layer, x_i^j represents output of j^{th} layer, x_{ij}^j represents convolution kernel of j^{th} layer, and b_j^j represents bias parameter.

Pooling shrinks feature-map space, cuts compute load, and keeps key traits; max and average pooling are typical. The process of maximum pooling operation is shown in formula (2):

$$P_j^{i+1}(j) = \max_{(j-1)/2+1 \leqslant i < j \leqslant l} q_i^j(t) \tag{2}$$

Here, $q_i^j(t)$ represents value of t^{th} neuron in i^{th} feature vector in j^{th} layer, and $t \in [(j-1)W+1, jW], P_i^{i+1}(j)$ represents value corresponding to neuron in i+1$^{\text{th}}$ layer.

A fully connected layer links every neuron to all units in the preceding layer, integrating incoming features via weighted sums to perform linear transformation. The calculation process is shown in formula (3):

$$y = Wx + b \tag{3}$$

where x represents input, W represents weight parameter, and b represents bias information.

LSTM, a tailored RNN, captures long-term dependencies and is widely applied in sequence prediction and NLP. Its input gate decides the portion of current input to update the cell state. The calculation process is shown in formula (4):

$$C_i = f_i C_{i-1} + \sigma\left(W_i\left[h_{i-1}, x_i\right] + b_i\right)\tanh\left(W_i\left[h_{i-1}, x_i\right] + b_i\right) \tag{4}$$

where C_i represents current cell state, C_{i-1} represents previous cell state, W_i represents weight matrix, b_i represents bias parameter, and σ is sigmoid activation function.

The forget gate controls how much past information the cell state keeps. The calculation process is shown in formula (5):

$$f_i = \sigma\left(W_i\left[h_{i-1}, x_i\right] + b_i\right) \tag{5}$$

Among them, x_i represents current input feature vector, σ represents sigmoid activation function, and h_{i-1} represents hidden layer information at previous time.

The output gate determines how much information is output in cell state. The calculation process is shown in formula (6):

$$h_i = \sigma\left(W_i\left[h_{i-1}, x_i\right] + b_i\right)\tanh\left(c_i\right) \tag{6}$$

where h_i represents output information at current time and tanh represents activation function.

2.2 SVM

SVM is a supervised learner rooted in statistical learning theory, mainly for binary classification (Li et al., 2025). Its key idea is to find an optimal hyperplane that separates the two classes while maximising the margin between them (Wu et al., 2025). The size of this interval directly affects the generalisation ability of the model, that is, classification performance of the model for unseen data.

The goal of SVM is to find a hyperplane $w^T x + b = 0$ such that all positive class samples satisfy $w^T x_i + b \geq 1$ and all negative class samples satisfy $w^T x_i + b \geq -1$. Here w is the normal vector of the hyperplane and b is the bias term. To maximise the interval, the representation of the SVM optimisation problem is shown in formulas (7) and (8):

$$\min_{w,b} \frac{1}{2}\|w\|^2 \tag{7}$$

$$\text{s.t.} \quad y_i\left(w^T x_i + b\right) \geq 1, \quad \forall i \tag{8}$$

where y_i is category label of sample x_i, which takes value of +1 or −1. Through the Lagrange multiplier method, above optimisation problem can be transformed into a dual problem for solving.

SVM's hallmark is its kernel trick, letting it classify in high-dimensional space without direct coordinate computation (Kumar et al., 2025). By mapping data via the kernel, it locates the optimal hyperplane there, achieving nonlinear separation (Ray et al., 2025). The calculation process of the Sigmoid kernel function is shown in formula (9):

$$K\left(x, x'\right) = \tanh\left(\alpha x^T x' + \beta\right) \tag{9}$$

Among them, α controls degree of expansion and contraction of kernel function, and β is equivalent to bias term, which controls translation of kernel function.

SVM offers strong generalisation and handles small data well, excelling in intricate classification tasks. It is widely adopted in image recognition, text mining, biomedical diagnosis and beyond – spotting faces, numerals and tumours alike – while also serving emerging needs like variable selection and sparse modelling. Yet its success hinges on kernel choice and parameter tuning, and large-scale data can strain its efficiency. Overall, SVM remains a potent algorithm that solves linear and nonlinear problems via margin maximisation and kernel tricks, as demonstrated by its proven track record across domains.

2.3 Forecasting model of higher vocational students' classroom participation

Building the vocational-student engagement predictor draws on diverse theories and analytic tools (Wang et al., 2022). First, exploratory factor analysis (EFA) examines participation in blended settings, spotlighting four core dimensions: face-to-face, hands-on, online and collaborative learning (Larmuseau et al., 2025). EFA extracted the main factors using principal component analysis (PCA) and assigned weights to the scale questions for each dimension to standardise the weights, ensuring consistency in the analysis. This analysis method provides a basis for subsequent data modelling, which can effectively quantify students' participation in different learning dimensions (Sulistiobudi and Kadiyono, 2023).

At the same time, some studies have also explored the relationship between students' tendency to experience boredom and their classroom participation, based on self-loss theory and social development theory, and further revealed the mediating and moderating roles of mobile phone dependence and classroom atmosphere (Suharno et al., 2025). These theories offer a crucial perspective for understanding the factors that influence students' classroom participation and also provide theoretical support for the development of prediction models (Qu et al., 2024). In terms of model construction, linear regression models are widely used to analyse and predict the classroom participation of higher vocational students (Paizan et al., 2024). By assessing how each learning dimension affects engagement, the model links the outcome variable – learning participation – to the predictors through a linear relationship (Zuo et al., 2025). The representation of a typical multiple linear regression model is shown in formula (10):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon \tag{10}$$

Among them, Y represents learning participation, X_1 to X_4 represent weighted scores of four dimensions of face-to-face learning, practical learning, online learning and cooperative learning respectively, β_0 is intercept term, β_1 to β_4 are regression coefficient, and ε is error term. By fitting model, the magnitude and direction of influence of each learning dimension on classroom engagement can be quantified.

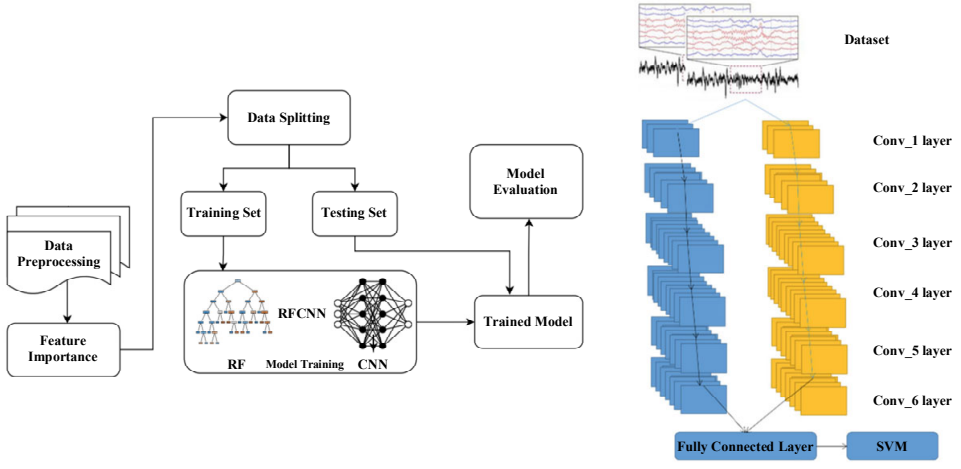
Based on traditional linear regression models, some studies have also introduced DL techniques (Auer et al., 2025), such as the GRU-Attention network, to enhance prediction accuracy. The GRU-Attention network is capable of capturing complex, dynamic changes in behavioural prioritising key features through attention mechanisms, thereby enhancing the accuracy of predictions (Marder et al., 2023). The model excels in predicting student learning engagement with an accuracy of 98.15%, significantly

outperforming traditional classification methods, such as decision trees, SVM, and random forests. Researchers also devised a CNN-LSTM blend: convolutions harvest spatial cues via local receptive fields and shared weights for emotion spotting and gesture sensing; pooling downsamples yet keeps salient motifs; LSTM gating secures long-span behavioural ties, taming classic RNN gradient decay. Text and logs are handled by Transformer self-attention that highlights pivotal events and crafts contextual embeddings, ultimately yielding a four-way high-level vector of behaviour, affect, cognition and practice.

3 A model that fuses DL and SVM

This study presents a vocational-student engagement predictor that fuses DL with SVM, leveraging the former's feature power and the latter's classification strength to boost accuracy and reliability. It offers educators a practical tool to grasp learners' states, refine instruction, and raise overall teaching quality. The model primarily comprises a data preprocessing module, a feature extraction module, an SVM classifier module, and a model fusion module. Its network architecture is shown in Figure 1.

Figure 1 Model combining DL and SVM (see online version for colours)



The data preprocessing module is responsible for collecting and sorting students' behaviour data in class, including attendance rates, number of interactions, homework submissions, and other relevant data, as well as data from the online learning platform. Data preprocessing includes operations such as data cleaning, normalisation, and encoding to ensure quality of data and training effect of model.

Feature extraction module utilises CNN and LSTM to extract useful features from the raw data automatically. DL models can capture complex patterns and relationships in data, providing rich feature representations for subsequent classification tasks. Specifically, a CNN was used to extract spatial features from classroom videos. A CNN architecture consisting of multiple convolutional and pooling layers was employed. Each convolutional layer used a 3×3 kernel with a stride of 1 and a ReLU activation function. The pooling layer employed a max pooling operation with a pooling window size of 2×2.

Through the combination of these layers, the CNN effectively extracted local features from video frames and gradually abstracted higher-level spatial features. LSTM was used to process the time series data in the behaviour logs. We constructed a network consisting of multiple LSTM units, each with 128 hidden units. LSTM effectively captures long-term dependencies in time series through a gating mechanism. Specifically, the input gate, forget gate, and output gate control the inflow, forgetting, and output of information, respectively, thereby achieving dynamic modelling of time series data.

The SVM classifier module takes the extracted features as input and applies them to the SVM classifier. SVM is a powerful classifier that distinguishes between different classes by finding a hyperplane that maximises the separation. This paper uses the radial basis function (RBF) as the kernel function of the SVM and optimises its parameters (such as the penalty parameter C and the kernel function parameter γ) through grid search. The SVM achieves robust classification of traditional statistical features by maximising the interval between categories.

The model fusion module primarily enhances the accuracy and robustness of predictions. This model employs a fusion strategy combining DL and SVM. Specifically, This paper combines the high-dimensional time series features extracted by CNN and LSTM with traditional statistical features from SVM classification at the feature level. This fusion strategy allows the model to fully leverage the feature extraction capabilities of DL models and the stability of traditional statistical methods, thereby improving overall prediction performance.

4 Experiment and results analysis

The dataset for this study was collected from multiple classrooms at a vocational college, encompassing a variety of majors and course types. It includes classroom videos and student behaviour logs, recording student behaviours such as raising hands, speaking, and interacting. The video data is recorded by a high-definition camera with a resolution of 1920×1080 and a frame rate of 30 fps. The behaviour log data is collected through a classroom interactive system, recording real-time behaviour and interactions. Specifically, the dataset comprises 13 features and 32,593 observations, with 8 features being categorical and 5 features being numerical. These features include student personal information, learning behaviour, and classroom interactions. The target variable is an engineered variable representing student engagement, where 1 indicates high engagement and 0 indicates low engagement. The dataset is slightly imbalanced, with approximately 72% of students showing low engagement and 28% showing high engagement.

During the data preprocessing phase, the video data was cropped and denoised. A background segmentation algorithm was used to extract student motion regions, and video frames were denoised to reduce noise. The behaviour log data was cleaned to remove duplicate and outlier records and converted to a time series format. All data was normalised, with feature values scaled to between 0 and 1, to improve model training efficiency and stability.

The dataset was divided into a training set (70%), a validation set (15%), and a test set (15%). The training set was used for model training, the validation set was used for hyperparameter tuning, and the test set was used for final performance evaluation. Stratified sampling was used to partition the data, ensuring that the proportion of data from each category in each subset was consistent with that in the original dataset. This

approach also took into account the distribution of data across different courses and majors to accommodate diverse classroom environments.

Model performance was gauged with accuracy, precision, recall and auc. accuracy is the share of all samples correctly classified. Precision is the fraction of predicted positives that are truly positive, whereas recall is the fraction of actual positives correctly identified. AUC, the area under the ROC curve, quantifies the model's power to separate positive and negative cases. Together, these metrics offer a holistic appraisal.

Table 1 reports the comparative performance of individual models and their fusion strategies on the classification task. Through analysis, it is concluded that among the six evaluation indicators, the complete fusion model has significant effects, specifically achieving an accuracy rate of 88.43%, a recall rate of 87.25%, and an F1 score of 0.870. AUC reaches 0.934, which is approximately 14% higher than that of a single SVM and about 9% higher than that of a single LSTM, indicating that the fusion of deep features and traditional features can significantly enhance classification performance.

Table 1 Performance comparison of different models and their fusion methods on classification tasks

Methods	*Accuracy (%)*	*Accuracy (%)*	*Recall rate (%)*	*F1 score*	*AUC*
SVM only (linear kernel)	74.32	72.15	68.43	0.703	0.762
SVM only (RBF core)	76.81	75.28	71.64	0.734	0.798
LSTM network only	79.24	77.86	75.92	0.769	0.832
Transformer module only	81.05	79.33	78.16	0.787	0.851
SVM + LSTM fusion	83.67	81.92	80.45	0.812	0.883
SVM + transformer fusion	85.92	84.76	83.27	0.840	0.907
Complete fusion model	88.43	87.25	86.91	0.870	0.934

Figure 2 Comparison of attention and interaction distribution of users with different engagement levels (see online version for colours)

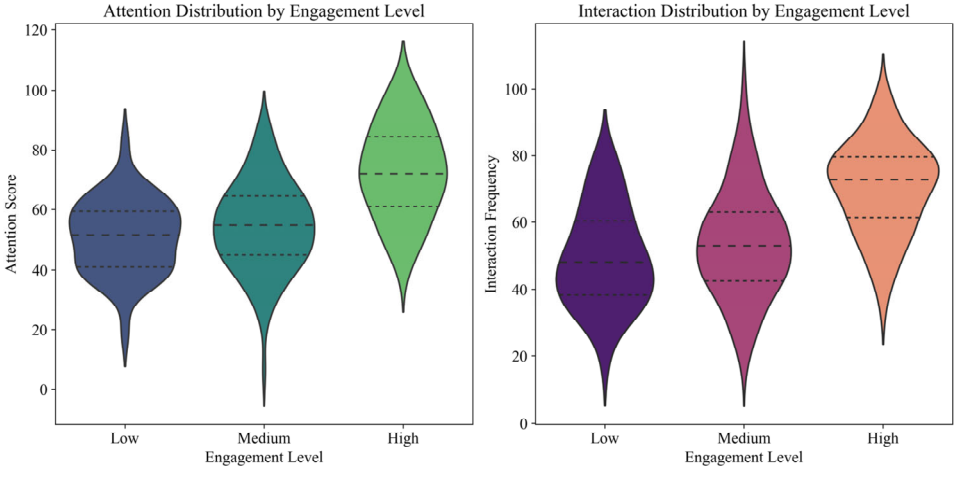


Figure 2 illustrates the comparison results of attention and interaction distribution among users with varying levels of engagement. Through the analysis, it is concluded that as

participation rises from low to high, the number of attention and interaction shows an increasing trend. Both indicators for low-participation groups are 20; those for medium-participation groups rise to 40, and those for high-participation groups reach as high as 100. The attention and interaction of high-participation users are five times that of low-participation users, respectively, indicating that deep participation significantly improves user activity.

Figure 3 presents the comparison results of characteristic portraits for students with high and low participation. Through the analysis, it is concluded that the five indicators of attention, task completion, interaction, response time and note quality of high participation students are all about 80, which is significantly better than that of low participation students only about 20; The gap between them is 60 points, indicating that high participation is positively correlated with learning input and achievements.

Figure 3 Comparison of characteristic portraits of students with high participation and low participation (see online version for colours)

Figure 4 Distribution of note quality and task completion under different classes and participation levels (see online version for colours)

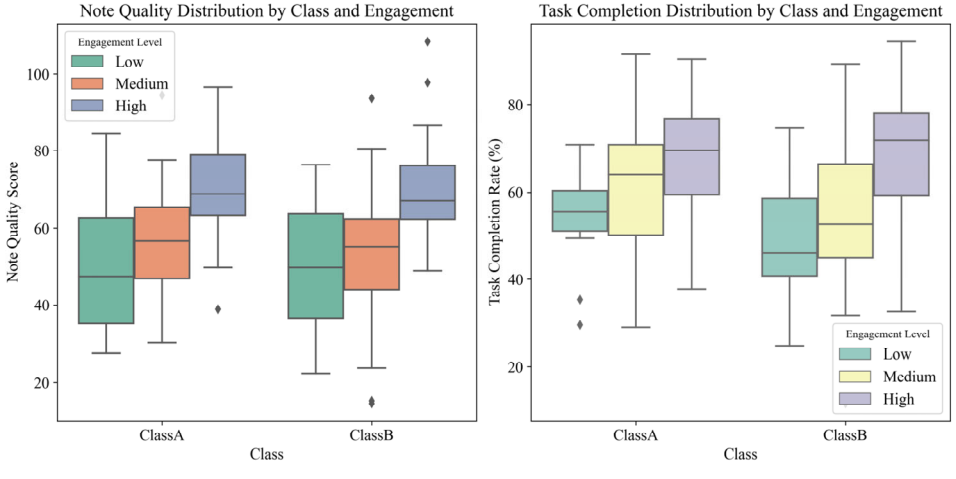


Figure 4 shows the distribution of note quality and task completion in different classes and participation levels. Through the analysis, it is concluded that the quality of notes and task completion of students with high participation are 80 points, which is significantly better than that of students with middle (60 points) and low (20 points) participation. Both indicators of Class A are slightly higher than those of Class B by about 10 points under low and middle participation, which shows that the class difference disappears in the high participation stage, suggesting that increasing participation can bridge the class gap.

Figure 5 shows scatter comparison results of participation between attention-interaction and note quality-task completion. Through analysis, it is concluded that the two figures are positively correlated: the attention, interaction, note quality, and task completion of students with high participation are approximately 80%, while those with low participation are only 20%. The intermediate engagement rate falls within the 40-60% range. For every 20 points of attention increase, interaction increases by approximately 20 points. For every 20-point increase in the quality of notes, the degree of task completion increases by 20 points simultaneously, indicating that the deeper the investment, the better the learning results.

Figure 5 Scatter comparison of engagement between attention-interaction and note quality-task completion (see online version for colours)

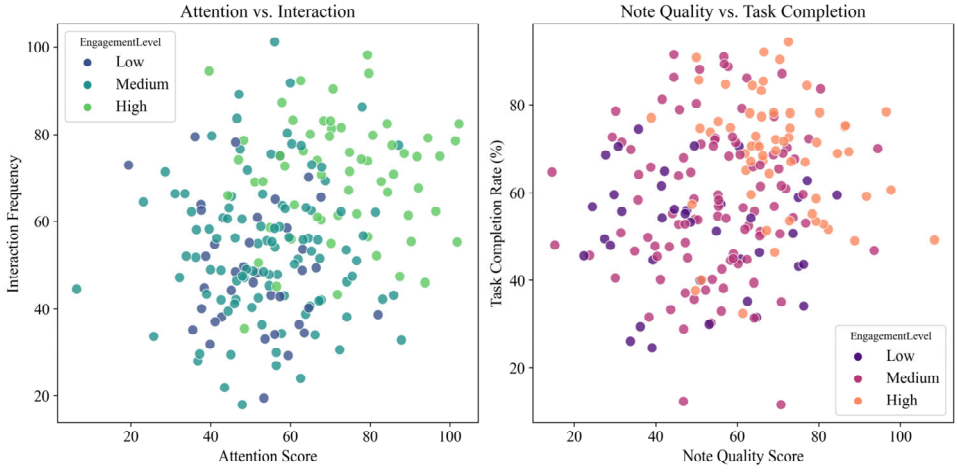


Figure 6 shows the distribution of participation ratio and cumulative participation in each week. Through analysis, it is concluded that the high, medium and low participation in the first week accounted for about 20%, 30% and 50% respectively, and then the high proportion rose to about 45% in the 10th week week by week, dropped to 25% in the medium, and dropped to 25% in the low week. to 30%; The cumulative curve shows that the total high participation is 120%, while the low participation is only 20%, showing a trend of 'the strong is stronger'.

Table 2 shows the comprehensive performance comparison results of each prediction model in terms of accuracy, efficiency and scale. Through analysis, it is concluded that fusion model in this paper leads with an accuracy rate of 88.43% and an F1 of 0.870, but the training 62.8 s, inference 12.5 ms, and parameter volume of 5.7 MB are the highest in the table; LightGBM only takes 8.9 s of training, 1.7 ms of reasoning, and 7.3 MB of

parameters with an accuracy of 81.93%, which is the best efficiency; Although the decision tree is fast and small, it has the lowest accuracy.

Figure 6 Weekly participation ratio and cumulative participation distribution (see online version for colours)

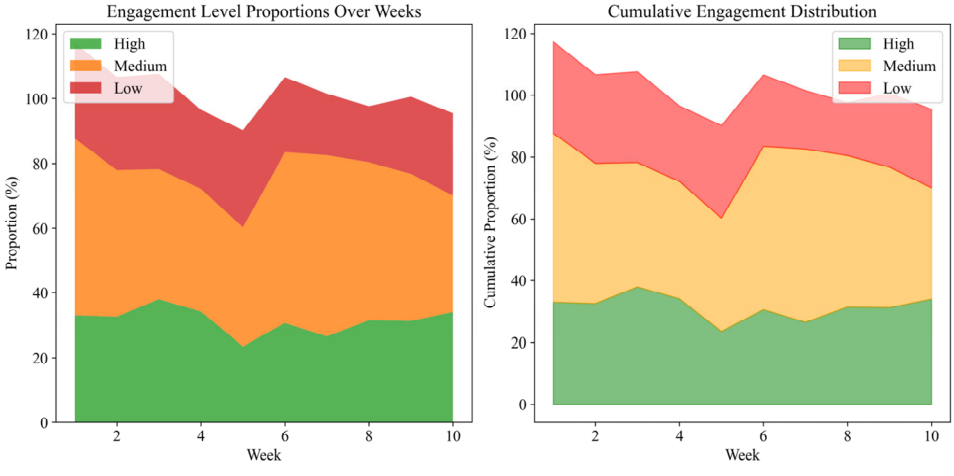


Table 2 Comparison of comprehensive performance of each prediction model in terms of accuracy, efficiency and scale

Prediction model	*Accuracy (%)*	*Training time (s)*	*Inference delay (ms)*	*Quantity of parameters (MB)*	*F1 score*
Decision Tree	71.85	3.2	0.8	0.7	0.692
Random Forest	78.26	18.7	4.3	15.2	0.761
XGBoost	80.44	12.5	2.1	9.8	0.783
LightGBM	81.93	8.9	1.7	7.3	0.802
BP neural network	76.38	23.1	5.2	2.1	0.749
1D-CNN	82.67	47.6	8.9	3.8	0.813
The fusion model in this paper	88.43	62.8	12.5	5.7	0.870

Analysing the data in Table 3, the fusion model significantly outperforms other models across all metrics. The fusion model achieved a prediction accuracy of 92.3%, an F1 score of 0.914, a recall rate of 93.1%, and a precision rate of 91.6%. In comparison, the CNN-LSTM model achieved a prediction accuracy of 88.7%, an F1 score of 0.892, a recall rate of 90.5%, and a precision rate of 87.9%. The SVM model achieved a prediction accuracy of 85.4%, an F1 score of 0.867, a recall rate of 87.2%, and a precision rate of 84.1%. The random forest model achieved a prediction accuracy of 84.2%, an F1 score of 0.855, a recall rate of 86.3%, and a precision rate of 83.1%. The logistic regression model achieved a prediction accuracy of 82.1%, an F1 score of 0.834, a recall rate of 84.5%, and a precision rate of 80.7%.

Table 3 Performance comparison of different methods on different evaluation metrics

Methods	*F1 score*	*Recall*	*Precision*
CNN-LSTM + SVM	0.914	93.1%	91.6%
CNN-LSTM	0.892	90.5%	87.9%
SVM	0.867	87.2%	84.1%
Random Forest	0.855	86.3%	83.1%
Logistic regression	0.834	84.5%	80.7%

Figure 7 Attention-response time and interaction-task completion density distribution (see online version for colours)

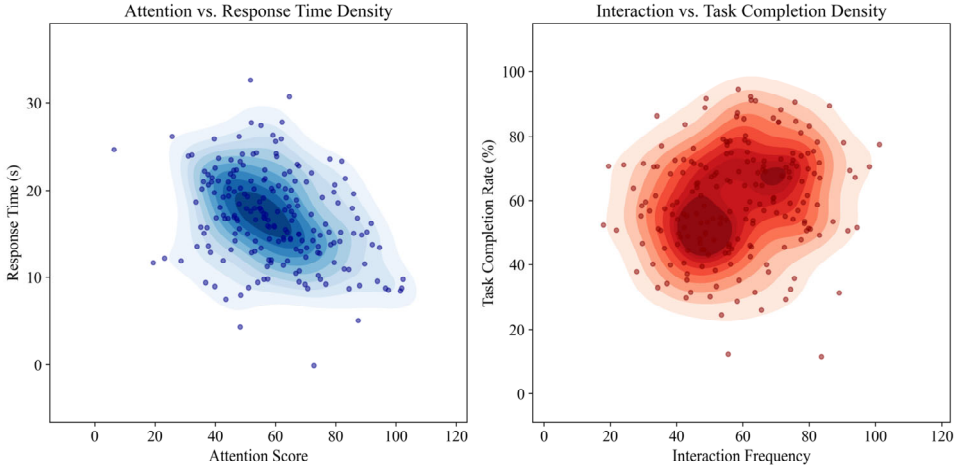


Figure 8 Difference between learning feature correlation matrix and class standardisation (see online version for colours)

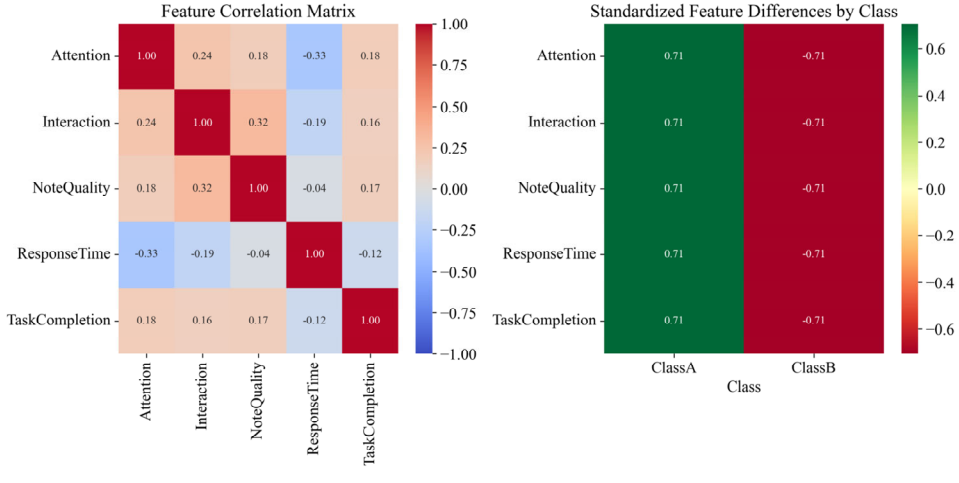


Figure 7 illustrates the distributions of attention-response time and interaction-task completion density. Through analysis, it is concluded that the high-density areas in the left two figures are clustered on the upper right: the proportion of students with attention higher than 70 points and response time lower than 30 seconds is the highest; Students with interaction frequency higher than 70 times and task completion higher than 80% have the highest density, which shows that attention and interaction are the key drivers of efficient learning.

Figure 8 shows the difference between the learning feature correlation matrix and class standardisation. Through analysis, the correlation matrix shows that attention is positively correlated with interaction and task completion, at 0.71 and 0.75, respectively, and negatively correlated with response time at –0.33. The standardised difference map indicates that Class A outperforms Class B by approximately 0.71 standard deviations in all characteristics, with the largest difference observed in attention and task completion, suggesting a significant gap in investment and effectiveness between the classes.

5 Conclusions

This paper presents a hybrid deep-learning – SVM model to forecast vocational students' classroom engagement. A deep net first learns behavioural, emotional, and cognitive features automatically; the resulting high-dimensional vectors are then fed to an SVM for robust classification, merging deep expressiveness with SVM margins. Experiments show the fused model surpasses standalone algorithms in accuracy and generalisation, delivering a real-time, dependable diagnostic aid for vocational classrooms. This framework can be promoted as a new paradigm of collaboration between DL and traditional machine learning in educational big data scenarios. The main work of this paper is as follows:

1 Use CNN and LSTM to extract high-dimensional time series features from classroom videos and behaviour logs.

2 Robust classification of traditional statistical features combined with SVM; Finally, the advantages of the two types of models are integrated by a feature-level fusion strategy.

3 Integrating the advantages of the two types of models through a feature-level fusion strategy.

The experiment on the simulated higher vocational classroom data set shows that the prediction accuracy of the fusion model is 92.3%, and the F1 value is 0.914, which is significantly better than those of the single models, namely CNN-LSTM and SVM. The model provides technical support for quantifying classroom participation in real-time and accurately implementing teaching interventions.

Acknowledgements

This work was supported by the Research Funds of Jiangsu Vocational Institute of Commerce (No. JSJMSZ23001) and the 2023 Teaching Research Project on Ideological and Political Theory Course for Higher Vocational Colleges in Jiangsu Province (No. 2023JSSZZD06).

Declarations

All data generated or analysed during the study are available from the corresponding author by request.

Author declares no conflicts of interest.

References

Aruleba, I. and Sun, Y. (2025) 'Enhanced credit risk prediction using deep learning and SMOTE-ENN resampling', *Machine Learning with Applications*, Vol. 21, p.100692, https://doi.org/10.1016/j.mlwa.2025.100692.

Auer, T., Reindl, M. and Gniewosz, B. (2025) 'Is the clique a pond? The big-fish-little-pond effect and the relative meaning of clique and classroom', *Learning and Instruction*, Vol. 95, p.101997, https://doi.org/10.1016/j.learninstruc.2024.10199.

Birthriya, S.K., Ahlawat, P. and Jain, A.K. (2025) 'Intelligent phishing website detection: a CNN-SVM approach with nature-inspired hyperparameter tuning', *Cyber Security and Applications*, p.100100, https://doi.org/10.1016/j.csa.2025.100100.

Dagasso, G., Wilms, M., Souza, R. and Forkert, N.D. (2025) 'Accounting for population structure in deep learning models for genomic analysis', *Journal of Biomedical Informatics*, p.104873, https://doi.org/10.1016/j.jbi.2025.10487.

Devi, G. and Kaushik, V. (2025) 'ML-based prediction of scour depth around a cylindrical bridge pier: a comparative analysis of ANN, SVM, and ensemble trees', *Ocean Engineering*, Vol. 336, p.121735, https://doi.org/10.1016/j.oceaneng.2025.12173.

Esatyana, E. and Sakhaee-Pour, A. (2025) 'Deep learning for upscaling complex nanoindentation images to predict fracture toughness', *Theoretical and Applied Fracture Mechanics*, Vol. 139, p.105065, https://doi.org/10.1016/j.tafmec.2025.105065.

Kumar, A., Gaur, N. and Nanthaamornphong, A. (2025) 'A mathematical PAPR estimation of OTFS network using a machine learning SVM algorithm', *Results in Optics*, Vol. 21, https://doi.org/10.1016/j.rio.2025.10083.

Larmuseau, C., De Leersnijder, L., Rotsaert, T., Boel, C., Demanet, J. and Schellens, T. (2025) 'Beyond realism: rethinking VR design for optimal learning in technical and vocational secondary education', *Computers & Education: X Reality*, Vol. 6, p.100098, https://doi.org/10.1016/j.cexr.2025.100098.

Li, M., Zhang, S. and Zhang, L-F. (2024) 'Vocational college students' vocational identity and self-esteem: dynamics obtained from latent change score modeling', *Personality and Individual Differences*, Vol. 229, p.112746, https://doi.org/10.1016/j.paid.2024.11274.

Li, X., Sun, Y., Chen, Z., Ma, J., He, W., Zhang, B., Song, Y. and Jiang, Q. (2025) 'Reliability-based design optimization of hinge sleeve using adaptive E-SVM', *Reliability Engineering & System Safety*, p.111435, https://doi.org/10.1016/j.ress.2025.111435.

Marder, J., Thiel, F. and Göllner, R. (2023) 'Classroom management and students' mathematics achievement: The role of students' disruptive behavior and teacher classroom management', *Learning and Instruction*, Vol. 86, p.101746, https://doi.org/10.1016/j.learninstruc .2023.101746.

Monaco, S., Monaco, L., Apiletti, D., Cremonini, R. and Barbero, S. (2025) 'Uncertainty-aware methods for enhancing rainfall prediction with deep-learning based post-processing segmentation', *Computers & Geosciences*, Vol. 205, p.105992, https://doi.org/10.1016/j.cageo .2025.105992.

Mu, X. and Zhao, B. (2025) 'DCS-SOCP-SVM: a novel integrated sampling and classification algorithm for imbalanced datasets', *Computers, Materials and Continua*, Vol. 83, No. 2, pp.2143–2159, https://doi.org/10.32604/cmc.2025.060739.

Mustapa, M.A.S., Ibrahim, M. and Yusoff, A. (2015) 'Engaging vocational college students through blended learning: improving class attendance and participation', *Procedia-Social and Behavioral Sciences*, Vol. 204, pp.127–135, https://doi.org/10.1016/j.sbspro.2015.08.12.

Paizan, M.A., Benbow, A.E.F. and Titzmann, P.F. (2024) 'Relationship quality in student-teacher-dyads: Comparing student and teacher determinants in multicultural classrooms', *International Journal of Intercultural Relations*, Vol. 101, p.102006, https://doi.org/10.1016/j.ijintrel.2024.102006.

Qianyi, Z. and Zhiqiang, L. (2024) 'Learning motivation of college students in multimedia environment with machine learning models', *Learning and Motivation*, Vol. 88, pp.102046, https://doi.org/10.1016/j.lmot.2024.10204.

Qu, F., Jiang, M. and Qu, Y. (2024) 'An intelligent recommendation strategy for integrated online courses in vocational education based on short-term preferences', *Intelligent Systems with Applications*, Vol. 22, p.200374, https://doi.org/10.1016/j.iswa.2024.20037.

Ray, K.K., Kumari, A., Kumar, S., Machavaram, R., Shekh, I., Deshmukh, S.M. and Tadge, P. (2025) 'Guava leaf disease detection using support vector machine (SVM)', *Smart Agricultural Technology*, p.101190, https://doi.org/10.1016/j.atech.2025.101190.

Spanos, D., Passalis, N. and Tefas, A. (2025) 'Leveraging subclass learning for improving uncertainty estimation in deep learning', *Neurocomputing*, p.130954, https://doi.org/10.1016/j.neucom.2025.130954.

Suharno, S., Ihsan, F., Himawanto, D.A., Pambudi, N.A. and Rizkiana, R. (2025) 'Sustainability development in vocational education: a case study in Indonesia', *Higher Education, Skills and Work-based Learning*, Vol. 15, No. 3, pp.668–689. https://doi.org/10.1108/HESWBL-01-2024-0018.

Sulistiobudi, R.A. and Kadiyono, A.L. (2023) 'Employability of students in vocational secondary school: role of psychological capital and student-parent career congruences', *Heliyon*, Vol. 9, No. 2, p.e13214, https://doi.org/10.1016/j.heliyon.2023.e13214.

Treeprapin, K., Bandatang, R., Panthum, T., Singchat, W., Prasanpan, J., Srikulnath, K. and Trirongjitmoah, S. (2025) 'Enhancing clariid catfish species classification: a deep learning framework utilizing cranial morphology and explainable AI', *Smart Agricultural Technology*, Vol. 12, p.101165, https://doi.org/10.1016/j.atech.2025.101165.

Wang, J., Tigelaar, D.E.H., Luo, J. and Admiraal, W. (2022) 'Teacher beliefs, classroom process quality, and student engagement in the smart classroom learning environment: a multilevel analysis', *Computers & Education*, Vol. 183, p.104501, https://doi.org/10.1016/j.compedu.2022.104501.

Wang, N., Ye, J-H., Gao, W., Lee, Y-S., Zeng, L. and Wang, L. (2024) 'What do they need? – The academic counseling needs of students majoring in art and design in a higher vocational college in China', *Heliyon*, Vol. 10, No. 6, p.e27708, https://doi.org/10.1016/j.heliyon.2024.e27708.

Wu, Y., Wu, H., Tang, X., Lv, J. and Zhang, R. (2025) 'Research on computer multi feature fusion SVM model based on remote sensing image recognition and low energy system', *Results in Engineering*, Vol. 26, p.104861, https://doi.org/10.1016/j.rineng.2025.104861.

Xiao, Z., Ning, X. and Duritan, M.J.M. (2025) 'BERT-SVM: a hybrid BERT and SVM method for semantic similarity matching evaluation of paired short texts in English teaching', *Alexandria Engineering Journal*, Vol. 126, pp.231–246, https://doi.org/10.1016/j.aej.2025.04.061.

Xiong, J., Zhang, J., Zheng, X., Hu, T., Xiang, H., Li, Y., Xiao, B., Wang, X. and Yang, R. (2025) 'Fluorescence sensing of nitro explosives based on deep learning', *Cell Reports Physical Science*, p.102690, https://doi.org/10.1016/j.xcrp.2025.10269.

Yang, X., Wang, J., Wang, P., Li, Y., Wen, Z., Shang, J., Chen, K., Tang, C., Liang, S. and Meng, W. (2025) 'Deep learning modeling using CT images for longitudinal prediction of benign and malignant ground-glass nodules', *European Journal of Radiology*, Vol. 190, p.112252, https://doi.org/10.1016/j.ejrad.2025.11225.

Zhao, C. and Yu, J. (2024) 'Relationship between teacher's ability model and students' behavior based on emotion-behavior relevance theory and artificial intelligence technology under the background of curriculum ideological and political education', *Learning and Motivation*, Vol. 88, p.102040, https://doi.org/10.1016/j.lmot.2024.102040.

Zuo, H., Zhang, M. and Huang, W. (2025) 'Lifelong learning in vocational education: a game-theoretical exploration of innovation, entrepreneurial spirit, and strategic challenges', *Journal of Innovation & Knowledge*, Vol. 10, No. 3, p.100694, https://doi.org/10.1016/j.jik.2025.10069.