International Journal of Advanced Technology and Engineering Exploration, Vol 11(112)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2023.10102002

Research Article

Learning analytics with correlation-based SAN-LSTM mechanism for formative evaluation and improved online learning

Valliammal Narayan* **and Sudhamathy Ganapathisamy**
Assistant Professor, Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women and Coimbatore, India

Received: 20-July-2023; Revised: 16-March-2024; Accepted: 19-March-2024

Abstract
Learning analytics (LA) is the measuring, gathering, analyzing, and reporting of data on learners and their environments. This data is used to analyze and improve e-learning. LA has traditionally been used for a variety of purposes, including the prediction of student academic progress and more specifically, the identification of students who are in danger of failing a course or quitting their studies. However, the majority of the existing schemes have the issue of accurately predicting the students' performance in online courses. In this paper, correlation-based self-attention network-long short-term memory (SAN-LSTM) is proposed to predict students' outcomes, along with the effectiveness of the teaching experience, as well as the assessment methods. Initially, the data is collected from three datasets namely, WorldUC, Liru, and Junyi to evaluate the performance of the proposed approach. The min-max normalization is employed to improve the performance of the approach. The correlation-based feature selection (CFS) is employed to select appropriate features from the pre-processed data. Finally, the correlation-based SAN-LSTM is established to forecast the effectiveness of fine-grained learning. Three real-world datasets gathered from various e-learning empirically validated that the proposed model improves prediction outcomes and provides useful data for formative evaluation. The existing methods such as adaptive sparse self-attention network (AS-SAN), Bangor engagement metric (BEM), and deep belief network learning style (DBNLS) are used for comparison to justify the effectiveness of the correlation-based SAN-LSTM method. The proposed correlation-based SAN-LSTM achieves better results of 98% of accuracy and 93% of precision. The proposed method achieves 98% accuracy which is higher when compared to those of AS-SAN, BEM, and DBNLS.

Keywords
Correlation-based self-attention network-long short-term memory, E-learning, Feedback, Fine-grained performance prediction, Long-term feature development.

1.Introduction
Massive open online courses (MOOC) have seen a considerable rise in the number of students, resources, and services because of the improvement of learning methods in computer-assisted and technology communications. Altogether, a better predictive model not only accurately predicts students' precise performance at the beginning of every stage of learning in an online course, but also efficiently represents the essential elements of the process of learning that have an inferred or explicit impact on performance [1].

When teachers and students develop learning processes using e-learning along with learning technologies, conventional and regular assessments are still in use to show students' performance and understanding. E-learning makes it simple to assess both the faculty's style of instruction and the degree of student satisfaction through self-evaluation techniques [2]. Technologies and tools for learning analytics (LA) are being developed to support this effort [3]. Formative assessment with feedback enhances the students' ability for self-learning and also assists the teachers in modifying their instructional technique [4]. The resources of learning are classified into three groups: hybrid filtering, collaborative filtering (CF), and content-based filtering [5].

*Author for correspondence

In any case, the challenge of creating training materials that are appropriate for learners is not limited only to the audience of employees/adults but also applies to other entities, such as educational institutions [6]. Tutors prefer to be able to quickly determine if students are feeling positive or negative emotions throughout a session [7]. While some students prefer to do their coursework during the day, others prefer to learn at night [8]. A e-learning is a primary factor in the learning context, because of the huge number of benefits and possibilities which it entails in various academic environments [9]. The Punjab education sector reform program (PESRP), a study on school enrolment that was released by the Punjab government in 2020-2021, highlights the value of enhancing student enrolment and the causes of young children falling out of school [10]. Universities and other educational institutions instruct a very large number of individuals, irrespective of their location, for both academic and lifelong learning, due to their size and accessibility [11]. Personalized learning systems focus on delivering content based on learner suitability, preferences, knowledge, and cognition level [12]. Data warehouses (DW) have been used in certain research to examine the ways students make decisions while utilizing the learning management system (LMS) to further decide based on their assessment [13, 14]. Using online tools and resources like LMS, students obtain a degree through distant learning after passing the examination [15]. Students benefit from methods of instruction in online learning, and teachers assist students in restoring interest for studying in a variety of ways, such as by providing educational resources [16]. Self-regulation is essentially needed for promising online learning sessions that enable students to handle their learning without the guidance of instructors [17]. The LA ensures that various stakeholders influence the invisible data and enhances the experiences for student learning depending on data analytics [18] [19]. Student-teacher, student-student, and student-content correlation have a significant place in the approach of community of inquiry (CoI) [20]. LA aggregates data that is established to the end user in the learning analytics dashboard (LAD) form [21]. To overcome the issue of inaccurate predictions of the students' performances in online courses, this paper proposes a correlation-based self-attention network-long short-term memory (SAN-LSTM) to predict student outcomes, teaching experience, and gauge the assessment methods. In the following, the primary contribution of the paper is summarized below:

- The min-max normalization is employed to improve the performance of the approach. The correlation-based feature selection (CFS) is employed to select the appropriate features from the pre-processed data.
- The correlation-based SAN-LSTM is established to forecast the effectiveness of fine-grained learning.
- Student outcome prediction, teaching experience prediction, and analysing of assessment methods, carried out by the proposed model, are evaluated based on accuracy, precision.

The rest of the paper is organized as follows: The literature survey is explained in section 2. Section 3 explains the proposed methodology and section 4 elucidates the results. Section 5 describes the discussion and section 6 presents the conclusion.

2.Literature survey

Wang et al. [22] implemented an adaptive sparse self-attention network (AS-SAN) for predicting the performance of fine-grained learning. A feature matrix representation enabled selection of position-wise features. It was easier to identify the most associated elements from the previous learning stage which resulted in an embedding of the original features as well as the spatial relationship between them. AS-SAN had a dependence on long-term and achieved good results and significant gain in its performances. However, the length of the sequences considerably lowered the model's effectiveness.

Tawafak et al. [23] presented a university communication model (UCOM) to highlight the usage of technologies on the internet in educational institutions and categorize them as a crucial tool for getting enhanced feedback on student assessment. The method illustrated the value of e-learning technologies as an evaluation strategy for the UCOM model, including the application form of Google, online chat, and modified MOODLE platform. UCOM method increased the student satisfaction levels, thereby giving enhanced feedback, by creating new methods and a balance between teachers and students. This led to improvement in students' academic performance. But the development and completion of the e-learning program had technological problems, right from connectivity challenges to learners' failure of remembering their login details.

Gray and Perkins [24] implemented a descriptive metric – the Bangor engagement metric (BEM) to generate student outcome predictions and this metric was initially used to improve Bangor University's current systems of student data. Machine learning was applied to solve the issue of student retention and classifier, and measurement combination was selected to fulfil the aim of early identification through a series of trials. The model achieved a high level of accuracy and reduced the amount of non-essential interventions and the number of wrongly predicted outcomes for the students. However, this implemented approach has over-fitting problems.

Rajabalee and Santally [25] presented a mixed-method approach to examine student comments and present the results of a study on the relationships between the contentment level of students and their participation in an online course with their comprehensive execution. The primary method was quantitative data collection and examination using association degree measures between the variables. These methods assisted the students in learning and achieving beneficial outcomes. However, this method did not generalize the research findings for the items having student comments.

Ez-zaouia et al. [26] implemented an emotional monitoring and observation dashboard for online learning (EMODASH), an interactive dashboard that helps tutors be more conscious of learners' emotions in the environment of video-conferencing. It aided the tutors to know about technological difficulties and the lack of emotional awareness brought on by asynchronous learning interactions. Tutors were able to focus on carrying out the learning activity correctly and aiming for pedagogical goals. The ability to visualize students' feelings helped tutors become more self-aware. However, this method was less concerned with accurately identifying emotions at any given time because it focused on the effect of retrospective emotion awareness on the tutor's feedback.

Zhang et al. [27] presented a learning style classification method based on the deep belief network learning style (DBNLS) to identify and categorize students' learning styles in online education on a large scale. First, a model of student learning style was created based on the existing method Felder Silver man model, and within the sessions of the individual the relationships between behaviors of network learning were examined to learn the style then deep learning method was applied to learn those learning styles' aspects. The learning style classification method achieved good performance and accuracy. However, the predictions for understanding the learning style dimensions were not entirely perfect.

Ma et al. [28] developed a deep temporal convolutional network for knowledge tracing (DTKT) to forecast the performances of student's future learning by demonstrating the student's historical exercise records. The data processing approach was employed to prevent the loss of input information. DTKT executed the parallel data processing with convolutional network features which significantly increased the power of computing. However, this approach struggled to capture the dependency of long-term in student learning because of the fixed convolutional layers.

Zhao et al. [29] introduced a transition-aware multi-activity knowledge tracing (TAMKOT) for the transition of student among assessed and non-assessed materials of learning. This approach was expressed as the learning model of deep recurrent multi-activities that determined knowledge transmission by learning and activating a knowledge transfer set metrics for transition among the activities of students. This approach achieved the student knowledge and forecasted their performance accurately. However, this approach suffered from overfitting risks.

Ke et al. [30] implemented a hierarchical transformer approach for session-aware knowledge tracing (HiTSKT) to model the sessional data for the issue of knowledge tracing (KT). This approach captured the forgetting behavior of student via the mechanism of power-law-decay attention scale in the consolidation and acquisition element. The implemented approach achieved better performance in terms of the correctness of future response prediction. However, the HiTSKT created additional complexity in the session of modeling data hence it prevented the effective capture and description of complex relationship among various learning sessions.

Yang and Ogata [31] presented a personalized LA intervention technique for increasing student learning achievement and the behavioral engagement in blended learning (BL). This approach generated actionable feedback for students with respect to personalized remedial actions to assist them to engage in the system of ebook and prevent academic failures while evaluating the BL. This approach increased the achievement of student learning and behavioral engagement in BL. However, this approach was not suitable for students in higher education from other institutions.

Hilliger et al. [32] developed the structure of two-cycle building-testing to determine the usability of the curriculum analytics (CA) tool. The first cycle included constructing the tool's initial version and analyzing its usage via the case study. Then, the next cycle included reconstructing the tool with respect to the lessons learnt from the first cycle and analyzing its usage via workshops. The results showed that the CA tool assisted teachers to gather a large number and variety of proof regarding students' competencies and achievements. However, this approach needed managers to determine the relationship among courses and program competencies for appropriate functioning.

Hadyaoui and Cheniti-belcadhi [33] introduced an ontology-based approach for the analytics of group assessment which examined the impact of intra-group interactions inside project-based collaborative learning (PBCL). It also forecasted the performance of learners depending on their interactions. This approach achieves robustness to evaluate the performance of group and provide the outcomes for student learning effectively. However, this approach contained a smaller sample size which needed to be expanded to increase the ability of generalization.

Jääskelä et al. [34] presented LA to analyze a higher education-based student agency. The information was evaluated with LA approach based on unsupervised robust clustering techniques. The entire procedure of analytics was performed automatically by utilizing the analytics process in two university courses. This approach enhanced the academic advising and teachers' pedagogical knowledge. However, this approach lacked efficacy due to small sample size.

Mubarak et al. [35] introduced an input-output hidden markov model (IOHMM) to predict students at the early dropout stage. This model was helpful to enable effectiveness in the course instruction and timely interventions with the students. The prediction of dropout was employed for an issue of sequence labeling. It achieved a better result in terms of accuracy. However, the model had smaller sample size which led to inaccurate findings.

Ouyang et al. [36] implemented an integrated artificial intelligence (AI) approach with LA to enhance student learning in the course of online engineering. The Quasi research was established to evaluate the variation of collaborative learning's effect on the student, with and without the integrated AI and LA techniques. This approach enhanced the student engagement, increased the performance of collaborative learning, and also the satisfaction of students about learning. However, combining AI with LA suffered difficulties in safeguarding student information.

Bhattacharya et al. [37] presented Manas Chakshu-a real-time information communication technology (ICT)-based interactive LA for huge blended classrooms. The system of visual LA included two interactive levels known by the names of overview and overview+ details levels. These two levels were employed to optimize the usage of screen-area from the obtainable display. This approach achieved high usability for efficiency, teacher's satisfaction, and the ratings of learnability. However, this approach did not pay attention to the problem of capturing the mental state.

Yan et al. [38] implemented learning design-analytic (LDA) and LA to enable the collection of data and pedagogical association. The analytics of learning design was established in which course learning and LA supported each other to maximize the success of learning. This technique enhanced the learning awareness, effectively recognized struggling students, and generated timely academic intervention. However, this technique was appropriate only for STEM disciplines.

Cole et al. [39] developed an online learning climate scale (OLCS) to predict the student engagement in online course. This model achieved evidence of the predictive potential for online student engagement (OSE). However, the model's sample was not random and did not provide generalizability.

Fu et al. [40] presented a visual learning analytics dashboard for the online judge system (VisOJ) that contains two kinds of user interface: student and teacher. The interface of teacher provided ranking trends and learning status that assisted teachers in monitoring and presenting feedback on students' online activities. The student interface generates views like error type analysis and evaluation which promotes students' self-regulation and reflection. This method achieved the system of self-adaptation. However, the VisOJ did not display the visual interface of more than two students or classes simultaneously.

El and Al [41] implemented flipped anatomy classroom that resulted in an enhancement in the students' performances and perceptions. It was established to provide the teacher more time and opportunity for assisting the students in analyzing, applying, and determining the materials of anatomy. It determined effectively the performance of students. However, this approach did not consider the various awareness sessions of a flipped classroom.

Albó et al. [42] introduced a knowledge-based design analytics for authorizing courses. This approach was determined in the programming course of higher education context. The within-subject users were performed to evaluate the design tool usage with and without visualization. This approach enhanced the entire quality of learning design and assisted teachers to avoid performing design error. However, this approach did not enhance the visualization of the introduced design analytics.

Kaliisa and Dolonen [43] developed a canvas discussion analytics dashboard (CADA) to manage teachers' roles, the perceptions and participation of students, as well as discourse patterns in the online environment. The dashboard measured alignment with appropriate theoretical construction, which enabled teachers to manage the design of learning. This approach has two iterations with key stakeholders which increased the value and uptake of LA system. However, the CADA mainly depended only on the user interviews and observations.

Sridharan and Akilashri [44] presented a hybrid attention network (HANet)-based analytic approach for assessing student behavior data by using enhanced capuchin search technique with multimodal data. By utilizing convolutional neural network (CNN), the deep features from the multi-modal behavior data were retrieved. The autoencoder and transformer net were employed to extract the features and then HANet approach was executed. This presented HANet model achieved increased prediction accuracy. However, the applicability of practical approaches was not employed in the presented model.

Cope et al. [45] implemented an AI approach for education that provided a certain tentative answer and then practically overviewed the outcomes of several experimental implementations. However, this approach had limited connections to other subjects because of the reinforcement learning and the structure of the curriculum.

Preuveneers et al. [46] developed a hybrid cloud and edge-based service orchestration model for the analysis of multimodal engagement. The solution of edge-based browser was applied for the evaluation of various modality behavior by the aggregation of cross-user data and secure computation of multi-party. This resolved the privacy concerns in the online LA. However, the extension of web browser in the presented approach was limited for the online learning engagement.

There are some limitations of LA that are mentioned in the literatures such as technological problems, overfitting, connectivity challenges, and length of the sequences significantly lowering the model's effectiveness. To overcome this issue, the correlation-based SAN-LSTM is proposed for LA to forecast the effectiveness of fine-grained learning.

3.Methods

Practically, a self-attention network presents an attention mechanism associating with various positions in a single sequence across overall sequence and it shows high potential for a variety of activities. However, n components are required for calculating the weights n times, which leads to a $O\ (N^2)$ time complexity that minimizes the model efficiency while the sequence is too long. Therefore, the correlation-based SAN-LSTM is established which effectively minimizes the computation amount by collecting information on local features with unique factorized attention heads. The overview of the LA with correlation-based SAN-LSTM mechanism is shown in *Figure 1*.

In this segment, the implementation of a correlation-based SAN-LSTM, used for accurate student performance prediction, is being explained. To effectively capture the interaction between the learning behaviors of eigenvalues, the proposed method uses an adaptive feature selection strategy in place of the standard mechanism of self-attention that is generally used in a model's prediction. Given that the learning characteristics of a numerical range (duration, engagement, etc.) vary greatly, there might be an indirect relationship between these characteristics. Self-attention uses historical information to record the hidden connection of every attribute in the course semantic framework using a decoder-only design. To acquire the distribution of features of the present location in the feature matrix, this must be compared with contributive elements. Following that, a maximum potential distribution value is used to anticipate a vacant position.

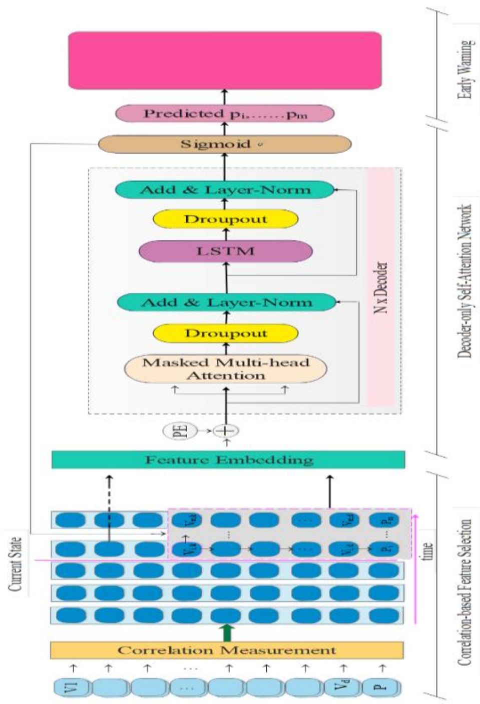

Figure 1 The implemented correlation-based SAN-LSTM model detailed architecture

This section is divided into two portions:

(i) A pair-wise correlation feature estimation approach is presented for the selection of local features before self-attention computing.

(ii) Autoregressive learning feature creation using correlation-based SAN-LSTM is used to forecast lesson-level performance in embedding layer. Finally, every aspect of the system is being discussed in great depth, in the upcoming sections.

3.1Model architecture

The model is made up of three parts: an early warning layer, a self-attention network decoder-only, and a CFS layer. The learning features are produced one at a time until the matrix learning feature is finished, using historical information from previously completed courses, as the input in the present course. *Figure 1* illustrates the AS-SAN model's architecture. The next subsections provide further information on the structure.

3.2Pre-processing

Min-Max normalization [47] is used to increase model performance and for normalizing raw data. It reduces data's outlier's impact and scales acquired data among 0 to 1 range that is shown in Equation 1.

$$p_{i,n} = \frac{p_{i,n} - min(p_i)}{\max(p_i) - \min(p_i)}(nMax - nMin) + nMin \quad (1)$$

Where max and min indicates i^{th} attribute's maximum and minimum values. By employing $nMax$ and $nMin$, obtained is rescaled by upper and lower boundaries. The final pre-processed data is passed via feature selection process.

3.3CFS

After pre-processing, CFS is employed for feature selection. CFS helps in identifying and maintaining features that have a powerful relationship with the variable of target, to improve predictive accuracy. In CFS, the features are selected depending on the correlation with target variables. The high correlation features are exhibited to generate more appropriate data for predictive modeling. This approach assists in enhancing the performance of the model by choosing most significant features while decreasing dimensionality. CFS has dynamic features, hence cannot separate the output or constrain the feature selection. Finding the items from the initial list of learning features that contribute the most to the local self-attention input is the aim of the feature selection layer. Memory blocks (MB) are utilized as the location arrays of the chosen attributes to succinctly define the feature selection range. The creation of the MB is separated into two parts based on the traits of the matrix feature learning:

Step 1: The static features remain unchanged via entire course that provides pairwise computation between target features and static features are unnecessary. Therefore, feature static sets are mentioned by $S_{sta} = \{v_{i,1}, v_{i,2,} \dots, v_{i,s}\}$.

Step 2: Every matrix of the column represents a time sequence of the dynamic properties. In terms of times series, the historical condition $\{v_{1,k}, v_{2,k,} \dots, v_{i-1,k}\}$ is directly connected to $v_{i,k}$ however, it is unclear how separate rows are connected. This concept, takes into account three techniques.

3.3.1 Stride feature selection

A threshold $\gamma \in [d+1, (i-1)x\,(d+1)]$ is allowed as a principle contestant for attributes adjoining. Hence, γ nearest neighbour features is expressed in Equation 2.

$$S(v_{i,k}, \gamma) = \{\, v_{i-\frac{\gamma}{d+1},k}, \dots, v_{i-\frac{\gamma}{d+1},d+1}, \dots, v_{i,k-1}\} \quad (2)$$

Where $\gamma \in [d+1, (i-1)x\,(d+1)]$ is the threshold.

By considering historical state prior target feature v_{ik} as candidate region, stride MB $MB_{stride}(v_{i,k})$ is described in Equation 3.

$$MB_{stride}(v_{i,k}) = S\,(v_{i,k}, \gamma) \cup \{v_{i,1}, v_{i,2}, \dots, v_{i,k-1}\} \quad (3)$$

Where $v_{i,k}$ is the target feature.

3.3.2 Similarity-based Feature Selection:

With two steps the similarity-based feature selection operation is reviewed: dynamic feature selecting, and similarity measuring. First, the various student types depending on the reliability of the static characteristics are identified and then the resemblance of the attributes of the historical information is computed. The similarity feature $r(\mathcal{F}^{(a)}, \mathcal{F}^{(b)}) \in [0,1]$, is computed by providing $\mathcal{F}^{(a)}, \mathcal{F}^{(b)}$ as learning features which is shown in below Equation 4.

$$r(\mathcal{F}^{(a)}, \mathcal{F}^{(b)}) = \frac{1}{n}\sum_{j=1}^{n} |\rho(\mathcal{F}_j^{(a)}, \mathcal{F}_j^{(b)})| \qquad (4)$$

Here, the online dataset course of the number of historical information is denoted as n. $\rho(\mathcal{F}_j^{(a)}, \mathcal{F}_j^{(b)}) \in [-1,1]$ is the correlation Pearson among $\mathcal{F}^{(a)}$ and $\mathcal{F}^{(b)}$ which is expressed as Equation 5.

$$\rho(\mathcal{F}^{(a)}, \mathcal{F}^{(b)}) = \frac{cov\ (\mathcal{F}^{(a)}, \mathcal{F}^{(b)})}{\sigma_{\mathcal{F}^{(a)}}\ \sigma_{\mathcal{F}^{(b)}}} = \frac{E\ \left[\left(\mathcal{F}^{(a)} - \mu_{\mathcal{F}^{(a)}}\right)\left(\mathcal{F}^{(b)} - \mu_{\mathcal{F}^{(b)}}\right)\right]}{\sqrt{\sum_{k=1}^{d+1}\left(\mathcal{F}^{(a)} - \mu_{\mathcal{F}^{(a)}}\right)^2}\ \sqrt{\sum_{k=1}^{d+1}\left(\mathcal{F}^{(b)} - \mu_{\mathcal{F}^{(b)}}\right)^2}} \qquad (5)$$

Where, covariance among $\mathcal{F}^{(a)}$ and $\mathcal{F}^{(b)}$ is $cov\ (\mathcal{F}^{(a)}, \mathcal{F}^{(b)})$ and the standard deviation of $\mathcal{F}^{(a)}$ and $\mathcal{F}^{(b)}$ are $\sigma_{\mathcal{F}^{(a)}}, \sigma_{\mathcal{F}^{(b)}}$, respectively. After this, value of $r \in [0,1]$ is employed for the features of the corresponding filters in Equation (6). For the features of k^{th} column,

$$S(v_{i,k}, k') = \begin{cases} \{v_{i-(i-1)r,k'}, \dots v_{i,k'}\}, if\ k' < k \\ \{v_{i-ir,k'}, \dots, v_{i-1,k'}\}, if\ k' \geq k \end{cases} \qquad (6)$$

$v_{i,k}$ is the feature target and $MB_{sim}(v_{i,k})$ is the whole similarity-based MB as indicated in Equation 7.

$$MB_{sim}(v_{i,k}) = \cup_{k'=1}^{d+1} S\ (v_{i,k}, k') \qquad (7)$$

3.3.3 Joint feature selection:

The model has more comprehensive horizontal characteristics in the stride feature selection, while the corresponding temporal features are obtained more successfully in the similarity-based feature selection. To integrate these two approaches, the properties of the neighboring sequence and the similarity-based approach is also taken into account. The $MB_{joint}\ (v_{i,k})$ is defined in Equation 8.

$$MB_{joint}\ (v_{i,k}) = S\ (v_{i,k}, \gamma) \cup \cup_{k'=1}^{d+1} S\ (v_{i,k}, k') \qquad (8)$$

3.4Adaptive spare self-attention network

Correlation based SAN-LSTM is implemented as a method to forecast the effectiveness of fine-grained learning, in this section. For the implementation of predicting the performance of students, a decoder-only multi-layer transformer is employed. The output layer, decoder layer, and feature embedding layer, result in creation of this model.

3.4.1Embedding layer

Due to the varying placements of the anticipated outcomes for an autoregressive task, the input features of the length vary. Discrete variables must be transformed into continuous vector representations of constant length along the embedding layer. The significance of every feature for the anticipated target is also determined by its location in the matrix learning feature. For that reason, the embedding of features usually decides a map $f_{fe}()$ from the discrete feature index to a D dimensional vector. Likewise installing of position decides further $f_{pe}()$ map from a discrete position index to a D dimensional vector. The input feature of the final embedding pos^{th} is expressed in Equation 9.

$$f\ (x, pos) = f_{fe}(x) + f_{pe}\ (pos) \qquad (9)$$

An extra positional embedding is often employed to represent the geometrical connections of information with regard to the embedding of the input in AS-SN. In the list of initial inputs $\in \mathbb{R}^{n' x\ d_{input}}$, d_{input} is the dimension of every input element and n' is the length sequence, for the preceding feature selection correlation based on the adaptive method. A position embedding $PE \in \mathbb{R}^{n' x\ d_{input}}$ is utilized to describe the absolute location of the matrix feature so that it takes advantage of the subsequent information in the starting list. The matrix embedding for the model's input is determined by in Equation 10.

$$E = emb(X, PE) = (x_{i'}W_e + pe_i^r W_r + pe_{i'}^c W_c)_{x_{i'} \in X, pe_{i'} \in PE'} \qquad (10)$$

Where, $emb()$ represents the function embedding along with feature embedding $f_{fe}()$ and position embedding $f_{pe}()$, the input data is represented as the parameter matrix, row and column are denoted as r and c of every input element as attention embeddings, while pe denotes a position embedding of two-dimensional. W_c, and W_r stand for the parameter matrices of every position's column and row. W_e represents input data's parameter matrix. Here W_c, W_r and W_e are learned through data training.

3.4.2Decoder layer

A decoder layer in a correlation-based SAN-LSTM combines both the memory cells of LSTM and self-attention approach. It employs correlations among memory states and input tokens to specifically utilize appropriate data. This fusion helps to capture both long-range correlations and sequential dependencies which improves the ability of the model to provide coherent output than other methods. The LSTM units process sequential data from prior layers. Correlation-based attention approach calculates relevance score among input tokens and LSTM states. The decoder layer contains three sub-layers which includes cross-attention, self-attention, and feed forward network (FFN). In decoders, the inner-relationships information is captured by a stack of N identical decoders. A LSTM and a multi-head attention (MHA) layer are present in every decoder. A highly optimized matrix multiplication algorithm is used to accomplish the complete self-attention procedure, which is then carried out concurrently over all channels of the features. The following are the primary variations between correlation-based SAN-LSTM and standard self-attention: (i) A typical network of self-attention, ensures that every input element $I_{i'} = \{I_1, I_2, \dots, I_{n'}\}$ contact all previous positions. (ii) Based on factorized attention, AS-SAN uses h to represent E in Equation 11.

$$A_{j'}^{(h)} \subset E, A^{(h)} \in \mathbb{R}^{j' \, x \, h}, \tag{11}$$

Where $j' \, (j' < i' - h)$- is every subset length $A^{(h')}, h' \in [1, h]$.

3.4.2.1 Masked MHA

The h sets of matrices $W_v^{(h)}, W_k^{(h)}, W_q^{(h)}$ denote values, keys and queries respectively, for the MHA, according to transformers. Before determining the attention, a masking function $M()$ is supplied to choose the top k contributing components in which ($k \in \{4, 8, 16, 32\}$). Concatenating threshold ς with k^{th} (k = 8) greatest rate of Z yields the following result in Equation 12.

$$M_\varsigma(Z) = \begin{cases} Z, if \ Z \geq \varsigma \\ -\infty, otherwise \end{cases} \tag{12}$$

The function of attention is assumed as follows in Equation 13 and 14.

$$head_h\left(E, A^{(h)}\right) = softmax\left(M_\varsigma\left(\frac{W_q^{(h)} e_{i'}\left(\left(A^{(h)} \, W_k^{(h)}\right)^T\right)}{\sqrt{d}}\right)\right) A^{(h)} W_v^{(h)}, \tag{13}$$

$$MH_h\left(E, A^{(h)}\right) = concat\left(head_1\left(E, A^{(h)}\right), \dots, head_h\left(E, A^{(h)}\right)\right) W^0, \tag{14}$$

Where, project matrix is denoted as $W^0 \in \mathbb{R}^{hn' x \, d_{input}}$, d denotes dimensionality of queries and keys, e_i' represents i^{th} element in E. The total values had their weights adjusted for the similarity dot-product scaled between the queries and keys.

3.4.2.2LSTM

The LSTM technique [48, 49] generates numerous outcomes like healthcare diagnosis, image processing, and natural language processing (NLP). LSTM is a recurrent neural network (RNN) type which has capability for learning long-term dependencies between input features. LSTM has following layers called hidden layers, input layers, output layers and LSTM layers. The input layer receives the data sequences and LSTM layers have various LSTM cells that process the information over time. The hidden layer contains an additional layer of dropout or dense for better performances. The output layer generates classifications depending on the processed data. The LSTM has three gates: input gate, forget gate, and output gate. The cell and hidden states are the two states makes LSTM. At a time, step t, LSTM evaluates data derived from cell state. At t, the x_t and h_t are the input and the hidden states respectively. Unlike RNN technique, it has c_t cell state, and three gates which are f_t forget gate, i_t input gate, o_t output gate. The f_t forget gate evaluates how many rates it manages from previous cell state value c_{t-1} at the time t.

The forget gate is represented in Equation 15.

$$f_t = \sigma(U_f x_t + W_f h_{t-1} + b_f) \tag{15}$$

Where σ is the sigmoid function, W_f and U_f are the weight values and b_f is the bias value. The sigmoid function is used as a neural network's activation function by hyperbolic tangent function. At time t, the i_t input gate evaluates how much x_t reflects to c_t output processing. The input gate i_t is indicated in Equation 16.

$$i_t = \sigma(U_i x_t + W_i h_{t-1} + b_i) \tag{16}$$

Where U_i and W_i are the weight values and b_i is the bias value. At t, the o_t output gate adopts output of a stored value in c_t and it is formulated in equation 17. The c_t at t is represented in Equation 18.

$$o_t = \sigma(U_o x_t + W_o h_{t-1} + b_o) \tag{17}$$

$$c_t = i_t \circ a_t + f_t \circ c_{t-1} \tag{18}$$

Where a_t and $\circ$ are new cell state at t and element-wise product. The a_t is indicated in Equation 19.At last, h_t hidden state is evaluated in Equation 20.

$$a_t = tanh(U_c x_t + W_c h_{t-1} + b_c) \tag{19}$$

$$h_t = o_t \circ tanh(c_t) \tag{20}$$

The output size of every gate is similar to a number of hidden units that is 256. The LA achieves better performance by using this LSTM approach.

3.5 Dropout

Dropout is a regularization approach for minimizing overfitting in neural networks that avoids the intricate co-adaptation of training data. This procedure results in the neural network's decreased units. The premise of this approach is to randomly drop units from the network during the training. This approach is effective in establishing model averaging with the neural networks. Each hidden neuron's fully-connected layer output is established to zero with a 0.5 probability. In this manner, the neurons are "dropped out" which does not contribute to the backpropagation (BP). The fully-connected layers of neurons are employed during training however, their outcomes are multiplied by 0.5.

4.Results

In this paper, the proposed correlation-based SAN-LSTM is simulated using MATLAB 2018 environments with the system requirements: Intel Core i7, 16 GB of RAM, GPU: 22GB, Libraries: OpenCV and TensorFlow, and the Windows 10 (64-bit) operating system having a runtime of 300 m/s for 100 iterations which is shown in *Table 1*. The performance of the correlation-based SAN-LSTM model is examined by means of accuracy, precision, mean absolute error (MAE), and root-mean squared error (RMSE).

Table 1 System configurations

System configurations	
Software Tool	MATLAB 2018
Processor	Intel i7
RAM	16 GB
GPU	22 GB
Libraries	OpenCV and TensorFlow
Operating System	Windows 10
Runtime	300 m/s for 100 iterations

4.1Dataset description

WorldUC

WorldUC dataset [50] contains 10,523 student entries with 10 continuous courses in a three-year course online. Eight input factors are considered for each student in a class, which contain three static characteristics (gender, expectation score, and age), five dynamic factors (liveliness, duration time, engagement, sentiment, and the importance of resources in learning and material), and one outcome factor (score of assessment). The dataset is divided into training, testing, and validation sets in the ratio of 70:20:10. Its characteristics contains specific-lesson engagement data, detailed student performance metrics, and longitudinal insights into the behavior of online learning.

Liru

The Liru dataset [51] is sourced from the online course e-learning platform. Its characteristics surrounds the metrics of student engagement, course enrolment information, and interactions of online learning temporal sequence. The dataset is divided into training, testing, and validation sets in the ratio of 70:20:10. It includes information on 18 input consecutive classes of 1046 students. The remaining dynamic five characteristics and the outcome characteristics are alike to the database of WorldUC unless for the static characteristics (grade, college, gender, and class) and two features of dynamic (assignment score and knowledge level).

Junyi

Junyi dataset [52] is a public dataset from the Academy of Junyi, wherein data on 2063 students are available on e-learning platforms such as the Khan Academy, with 29 input characteristics divided into 6 categorizations, including user-taken in exercises (3 features), accuracy answering (6 features), student modeling (4 features), exercise-related problem (4 features), answer duration of time (6 features, and orders of the user answering (6 features). Six features of the answering accuracy are merged into one output feature, the four student modelling features are used as static features, and the remaining features are dynamic. The dataset is divided into training, testing, and validation sets in the ratio of 70:20:10. This dataset is basically sourced from the online educational platforms and it includes student interactions with exercises. Its characteristics contains real-time learning behavior data and extensive student action's temporal sequences. *Table 2* indicates the three-dataset description.

Table 2 Three dataset description

	WorldUC	LIRU	JUNYI
Input features	8	18	29
Students	10,523	1046	2063

4.2Evaluation metrics

- MAE – It is determined by the absolute difference between the value predicted and actual values. MAE does not concern about the direction of errors it measures their magnitude as expressed in Equation 21.

$$MAE = \frac{1}{K}\sum_{i=1}^{K}|\hat{y_i} - y_i| \quad (21)$$

- RMSE – It is evaluated by the average size of the error and is concerned with the variations from the actual value. RMSE is expressed in Equation 22.

$$RMSE = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(\widehat{y_i} - y_i)^2} \quad (22)$$

- Accuracy: It is a proportion of accurate predictions to every input sample which is represented in Equation 23.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (23)$$

- Precision-It evaluates a actual data records versus expected data records. Which is expressed in Equation 24.

$$Precision = \frac{TP}{TP+FP} \quad (24)$$

4.3Experimental results

The MAE and RMSE are employed to measure deviation in the prediction performance of the suggested model, in three datasets: WorldUC, Junyi, and Luri with a fixed value *b*, to illustrate the variation for the baseline and proposed techniques on various phases of prediction performance. Accuracy generates an overall measure of accurate predictions which reflects the general system performance. Precision represents the true positive proportion between all positive predictions which ensures reliable and appropriate outcomes. The error rates MAE and RMSE provide insights into incorrect predictions which helps to identify areas for enhancement and refining learning interventions. By using these metrics, an evaluation is provided on the proposed approach to see how effectively it assesses the e-learning models to ensure efficacy and quality in educational outcomes. The employed hyperparameters are, grid search with a batch size of 32, Adam optimizer, initial learning rate of 0.001, 0.01, momentum of 0.0, 0.2, and decay rate of 0.001 respectively. *Table 3* shows when compared with the several techniques correlation-based SAN-LSTM technique offers the best results. When SAN and LSTM outcomes are compared, SAN has more advantages for long-term feature development since SAN collects global features with more data for prediction. However, Correlation-based SAN-LSTM performs better in terms of short-term feature generation because filtering away unimportant features reduces the noise in the data. For the testing dataset, various mask settings of 20%, 50%, and 75% are explored to access the outcomes of all the baselines and the methods respectively. In *Table 4*, for all three datasets, MAE and RMSE are used to measure their deviation for performance prediction. A sequence of features that is extremely long will damage the bidirectional -LSTM'S ability to predict outcomes, even though LSTM is a structure of the bi-directional network and gating algorithm. However, the results show that the correlation-based SAN-LSTM is significantly more efficient than other methods.

Table 3 The performance on datasets with b=3 mask ratios

Dataset	Metric	SAN	LSTM	SAN-LSTM	Feature selection LSTM	Correlation-based SAN-LSTM
WorldUC (b=3)	MAE	3.97	3.45	2.85	2.02	1.50
	RMSE	7.84	6.98	5.91	5.09	4.35
Liru (b=3)	MAE	6.88	5.98	4.74	3.26	2.19
	RMSE	8.23	7.65	6.92	5.32	4.02
Junyi (b=3)	MAE	9.33	8.86	7.21	6.50	4.21
	RMSE	11.69	9.56	7.87	6.24	5.78

Table 4 The performance on datasets with b=5 mask ratios

Dataset	Metric	SAN	LSTM	SAN-LSTM	Feature selection LSTM	Correlation-based SAN-LSTM
WorldUC (b=5)	MAE	4.38	4.26	4.09	3.82	3.45
	RMSE	9.49	9.14	8.98	7.66	6.45
Liru (b=5)	MAE	7.66	6.02	5.76	4.04	3.28
	RMSE	8.6	7.2	6.3	5.24	4.98
Junyi (b=5)	MAE	9.24	9.01	8.54	7.68	6.35
	RMSE	12.43	11.20	10.35	9.06	8.54

In comparison to the other baselines for the generation of long sequences, the results of correlation-based SAN-LSTM show significant performance gains in learning feature generation. Concerning the formative assessment and personalized feedback on the process of online learning, empirical validations on three real-world datasets: WorldUC, Luri, and Junyi are gathered from various e-learning platforms that demonstrate that the proposed method produces better prediction results as per *Table 5*. The advantage of correlation-based SAN-LSTM is reduced overall computing complexity for each layer and maximized number of computations that are done in parallel, as measured by the minimal number of sequential operations. This demonstrates that predictions made later in a course will generally be slightly more accurate than those made early. *Table 6* shows the performance evaluation of computation training time. When compared with the existing methods SAN, LSTM, SAN-LSTM, and feature selection LSTM, the proposed Correlation-based SAN-LSTM achieves lesser computational training time of 300 ms.

Table 5 The performance on datasets with b=9 mask ratios

Dataset	Metric	SAN	LSTM	SAN-LSTM	Feature selection LSTM	Correlation-based SAN-LSTM
Liru (b=9)	MAE	10.06	8.31	7.45	6.02	4.25
	RMSE	14.05	13.96	11.76	9.45	7.82
Junyi (b=9)	MAE	10.98	9.54	7.23	5.67	3.78
	RMSE	15.07	14.25	11.13	9.87	6.19

Table 6 Performance evaluation of computation training time

Methods	Computation training time (ms)
SAN	700
LSTM	650
SAN-LSTM	600
Feature selection LSTM	500
Correlation-based SAN-LSTM	300

Table 7 represents the performance evaluation of k-fold values for the correlation-based SAN-LSTM approach. k-fold validation divides the dataset into k subsets or folds for the purpose of evaluating predictive models. The model is trained and tested k times with each fold serving as the validation set, with performance metrics from each fold collected to estimate the model's generalization performance. This method facilitates model evaluation and selection by providing a more reliable measure of a model's effectiveness. Each fold set is executed perfectly once to avoid overfitting. To achieve k-fold validation, the dataset is divided into three sections: training, testing, and validation. When the dataset is divided into 5 folds and the training and testing processes are done, k=5 results in the model's highest values. *Table 8* shows performance evaluation of interpretability of model prediction. In this table, two reviews are evaluated with actual and predicted label. If both the actual and predicted label are identical, then it means that the model has provided accurate prediction in the LA. *Table 9* represents the performance evaluation of inaccurate prediction. In this table, two reviews are analyzed with actual and predicted label. If actual and predicted labels are different, it means that the model has generated inaccurate prediction in LA. A complete list of abbreviations is summarized in *Appendix I*.

Table 7 Performance evaluation of k-fold values

k-fold values	Accuracy (%)	Precision (%)
k=3	95	92
k=5	98	93
k=8	96	91
K=10	92	90

Table 8 Performance evaluation of interpretability of model prediction

Review	Actual Label	Predicted Label
The system of LA efficiently tracks the progress of student which helps in timely interventions.	Positive	Positive
The dashboard of LA provides a user-friendly which facilitate data-driven decision making.	Negative	Negative

Table 9 Performance evaluation of inaccurate prediction

Review	Actual Label	Predicted Label
The system of LA efficiently tracks the progress of student which helps in timely interventions.	Positive	Negative
The dashboard of LA provides a user-friendly which facilitate data-driven decision making.	Negative	Positive

4.4 Comparative analysis

Table 10 shows the comparative analysis of existing methods with the proposed correlation-based SAN-LSTM. The existing AS-SAN [22], BEM [24], and DBNLS [27] are compared with the proposed correlation-based SAN-LSTM. The proposed correlation-based SAN-LSTM achieves better accuracy of 98% and precision of 93% respectively compared to existing techniques. The correlation-based SAN-LSTM approach improves the ability of model to capture complex relationships in sequential data for different tasks of LA. This approach enhanced accuracy and interpretability in the application of LA by efficiently using both sequential contexts and dependencies. Therefore, the proposed correlation-based SAN-LSTM outperforms other approaches.

Table 10 Comparative analysis with existing methods

Authors	Methods	Accuracy (%)	Precision (%)
Wang et al. [22]	AS-SAN	96.2	89
Gray and Perkins [24]	BEM	97	87.8
Zhang et al. [27]	DBNLS	84	89
Proposed method	Correlation-based SAN-LSTM	98	93

5. Discussion

This section discusses the limitations of existing approaches and advantages of proposed approach. The existing approaches have some limitations such as AS-SAN [22] has lower length of the sequences data which significantly decreases the model's effectiveness. BEM [24] approach suffers from the over fitting. DBNLS [27] has predictions for the dimensions of understanding learning style but are not entirely perfect. The proposed correlation-based SAN-LSTM overcomes the existing approaches' limitations. Integrating LSTM with correlation-based approach enables the model to efficiently capture and analyze dependencies among longer sequences. This approach makes more accurate predictions in the tasks of LA which leads to better decision-making and insights. The proposed approach captures complex relationship among input tokens which increases feature representation. The results are evaluated for correlation-based SAN-LSTM for LA. The acquired outcomes indicate that proposed approach achieves better accuracy of 98% and precision of 93% respectively. The comparative results represent that correlation-based SAN-LSTM achieves better accuracy of 98% and 93% which is comparatively greater than AS-SAN (96.2% accuracy and 89% precision), BEM (97% accuracy and 87.8% precision), and DBNLS (84% accuracy and 89% precision) respectively. Hence, the overall results indicate the effectiveness of proposed approach for overall metrics which leads to increased insights into the performance of students. The correlation-based SAN-LSTM effectively captures key indicator which provides more personalized intervention and informed decision-making that increases the effectiveness of student outcomes and educational strategies.

Limitations

In this study, actual data is gathered and evaluated for a module that was not intended to be presented in an experimental context. This proposed method is less concerned with accurately identifying emotions due to more focus on the effect of respective awareness of the tutor's feedback.

6. Conclusion and future work

In this paper, the correlation-based SAN-LSTM method was proposed to accurately predict student outcome performance. The empirical validations of the method on real-world three datasets: WorldUC, Liru, and Junyi obtained from various platforms on e-learning show that the method improves forecast outcomes. The min-max normalization is utilized to enhance model performance. The CFS is employed to select the appropriate features from the pre-processed data. The Correlation based SAN-LSTM is established to forecast the effectiveness of fine-grained learning. When compared with the existing methods AS-SAN, BEM, and DBNLS, the correlation-based SAN-LSTM achieves better accuracy of 98%. In the future, the proposed Correlation-based SAN-LSTM method may further be improved to accommodate new application scenarios.

Acknowledgment
None.

Conflicts of interest
The authors have no conflicts of interest to declare.

Data availability
The datasets utilized in this study are publicly accessible. The Liru dataset is available at https://github.com/DeepReSTProject/The-LiruNET-2018-dataset/?tab=readme-ov-file, and the Junyi dataset can be accessed at https://www.kaggle.com/datasets/junyiacademy/learning-activity-public-dataset-by-junyi-academy. Additionally, the WorldUC dataset is accessible at https://doi.org/10.3390/su14137654.

Author's contributions statement
Valliammal Narayan: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing. **Sudhamathy Ganapathisamy**: Supervision, project administration, writing—review and editing.

References
[1] Maher Y, Moussa SM, Khalifa ME. Learners on focus: visualizing analytics through an integrated model for learning analytics in adaptive gamified e-learning. IEEE Access. 2020; 8:197597-616.

[2] Kawamura R, Shirai S, Takemura N, Alizadeh M, Cukurova M, Takemura H, et al. Detecting drowsy learners at the wheel of e-learning platforms with multimodal learning analytics. IEEE Access. 2021; 9:115165-74.

[3] Paneque M, Del MRM, García-nieto J. e-LION: data integration semantic model to enhance predictive analytics in e-Learning. Expert Systems with Applications. 2023; 213:118892.

[4] Wang D, Han H. Applying learning analytics dashboards based on process- oriented feedback to improve students' learning effectiveness. Journal of Computer Assisted Learning. 2021; 37(2):487-99.

[5] Shahbazi Z, Byun YC. Agent-based recommendation in E-learning environment using knowledge discovery and machine learning approaches. Mathematics. 2022; 10(7):1-19.

[6] Díaz RRP, Caeiro RM, López EJJ, Fernández VA. Integrating micro-learning content in traditional e-learning platforms. Multimedia Tools and Applications. 2021; 80(2):3121-51.

[7] Distante D, Villa M, Sansone N, Faralli S. MILA: a SCORM-compliant interactive learning analytics tool for moodle. In 20th international conference on advanced learning technologies 2020 (pp. 169-71). IEEE.

[8] Hasan R, Palaniappan S, Mahmood S, Abbas A, Sarker KU, Sattar MU. Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. Applied Sciences. 2020; 10(11):1-20.

[9] Muñoz S, Sánchez E, Iglesias CA. An emotion-aware learning analytics system based on semantic task automation. Electronics. 2020; 9(8):1-24.

[10] Abideen ZU, Mazhar T, Razzaq A, Haq I, Ullah I, Alasmary H, et al. Analysis of enrollment criteria in secondary schools using machine learning and data mining approach. Electronics. 2023; 12(3):1-25.

[11] Hmedna B, El MA, Baz O. A predictive model for the identification of learning styles in MOOC environments. Cluster Computing. 2020; 23(2):1303-28.

[12] Pal S, Pramanik PK, Choudhury P. Enhanced metadata modelling and extraction methods to acquire contextual pedagogical information from e-learning contents for personalised learning systems. Multimedia Tools and Applications. 2021; 80(16):25309-66.

[13] Li Q, Duffy P, Zhang Z. A novel multi-dimensional analysis approach to teaching and learning analytics in higher education. Systems. 2022; 10(4):1-18.

[14] Khalil M, Topali P, Ortega-arranz A, Er E, Akçapınar G, Belokrys G. Video analytics in digital learning environments: exploring student behaviour across different learning contexts. Technology, Knowledge and Learning. 2023:1-29.

[15] Ozdamli F, Aljarrah A, Karagozlu D, Ababneh M. Facial recognition system to detect student emotions and cheating in distance learning. Sustainability. 2022; 14(20):1-19.

[16] Ma Y, Wei Y, Shi Y, Li X, Tian Y, Zhao Z. Online learning engagement recognition using bidirectional long-term recurrent convolutional networks. Sustainability. 2022; 15(1):1-14.

[17] Han I, Obeid I, Greco D. Multimodal learning analytics and neurofeedback for optimizing online learners' self-regulation. Technology, Knowledge and Learning. 2023; 28(4):1937-43.

[18] Wong BT, Li KC, Cheung SK. An analysis of learning analytics in personalised learning. Journal of Computing in Higher Education. 2023; 35(3):371-90.

[19] Li L, Farias HL, Liang L, Law N. An outcome-oriented pattern-based model to support teaching as a design science. Instructional Science. 2022:1-32.

[20] Banihashem SK, Farrokhnia M, Badali M, Noroozi O. The impacts of constructivist learning design and learning analytics on students' engagement and self-regulation. Innovations in Education and Teaching International. 2022; 59(4):442-52.

[21] Rets I, Herodotou C, Bayer V, Hlosta M, Rienties B. Exploring critical factors of the perceived usefulness

of a learning analytics dashboard for distance university students. International Journal of Educational Technology in Higher Education. 2021; 18:1-23.
[22] Wang X, Mei X, Huang Q, Han Z, Huang C. Fine-grained learning performance prediction via adaptive sparse self-attention networks. Information Sciences. 2021; 545:223-40.
[23] Tawafak RM, Romli AB, Alsinani M. E-learning system of UCOM for improving student assessment feedback in Oman higher education. Education and Information Technologies. 2019; 24:1311-35.
[24] Gray CC, Perkins D. Utilizing early engagement and machine learning to predict student outcomes. Computers & Education. 2019; 131:22-32.
[25] Rajabalee YB, Santally MI. Learner satisfaction, engagement and performances in an online module: implications for institutional e-learning policy. Education and Information Technologies. 2021; 26(3):2623-56.
[26] Ez-zaouia M, Tabard A, Lavoué E. EMODASH: a dashboard supporting retrospective awareness of emotions in online learning. International Journal of Human-Computer Studies. 2020; 139:102411.
[27] Zhang H, Huang T, Liu S, Yin H, Li J, Yang H, et al. A learning style classification approach based on deep belief network for large-scale online education. Journal of Cloud Computing. 2020; 9:1-7.
[28] Ma R, Zhang L, Li J, Mei B, Ma Y, Zhang H. Dtkt: an improved deep temporal convolutional network for knowledge tracing. In 16th international conference on computer science & education 2021 (pp. 794-9). IEEE.
[29] Zhao S, Wang C, Sahebi S. Transition-aware multi-activity knowledge tracing. In IEEE international conference on big data (Big Data) 2022 (pp. 1760-9). IEEE.
[30] Ke F, Wang W, Tan W, Du L, Jin Y, Huang Y, et al. HiTSKT: a hierarchical transformer model for session-aware knowledge tracing. Knowledge-Based Systems. 2024; 284:111300.
[31] Yang CC, Ogata H. Personalized learning analytics intervention approach for enhancing student learning achievement and behavioral engagement in blended learning. Education and Information Technologies. 2023; 28(3):2509-28.
[32] Hilliger I, Aguirre C, Miranda C, Celis S, Pérez-sanagustín M. Lessons learned from designing a curriculum analytics tool for improving student learning and program quality. Journal of Computing in Higher Education. 2022; 34(3):633-57.
[33] Hadyaoui A, Cheniti-belcadhi L. Ontology-based group assessment analytics framework for performances prediction in project-based collaborative learning. Smart Learning Environments. 2023; 10(1):1-27.
[34] Jääskelä P, Heilala V, Kärkkäinen T, Häkkinen P. Student agency analytics: learning analytics as a tool for analysing student agency in higher education. Behaviour & Information Technology. 2021; 40(8):790-808.
[35] Mubarak AA, Cao H, Zhang W. Prediction of students' early dropout based on their interaction logs in online learning environment. Interactive Learning Environments. 2022; 30(8):1414-33.
[36] Ouyang F, Wu M, Zheng L, Zhang L, Jiao P. 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. 2023; 20(1):1-23.
[37] Bhattacharya S, Biswas U, Damkondwar S, Yadav B. Real-time ICT-based interactive learning analytics to facilitate blended classrooms. Education and Information Technologies. 2023:1-31.
[38] Yan H, Lin F, Kinshuk. Including learning analytics in the loop of self-paced online course learning design. International Journal of Artificial Intelligence in Education. 2021; 31(4):878-95.
[39] Cole AW, Lennon L, Weber NL. Student perceptions of online active learning practices and online learning climate predict online course engagement. Interactive Learning Environments. 2021; 29(5):866-80.
[40] Fu Q, Bai X, Zheng Y, Du R, Wang D, Zhang T. VisOJ: real-time visual learning analytics dashboard for online programming judge. The Visual Computer. 2023; 39(6):2393-405.
[41] El SA, Al AW. Improvement in student performance and perceptions through a flipped anatomy classroom: shifting from passive traditional to active blended learning. Anatomical Sciences Education. 2021; 14(4):482-90.
[42] Albó L, Barria-pineda J, Brusilovsky P, Hernández-leo D. Knowledge-based design analytics for authoring courses with smart learning content. International Journal of Artificial Intelligence in Education. 2022; 32(1):4-27.
[43] Kaliisa R, Dolonen JA. CADA: a teacher-facing learning analytics dashboard to foster teachers' awareness of students' participation and discourse patterns in online discussions. Technology, Knowledge and Learning. 2023; 28(3):937-58.
[44] Sridharan TB, Akilashri PS. Hybrid attention network-based students behavior data analytics framework with enhanced capuchin search algorithm using multimodal data. Social Network Analysis and Mining. 2023; 13(1):145.
[45] Cope B, Kalantzis M, Searsmith D. Artificial intelligence for education: knowledge and its assessment in AI-enabled learning ecologies. Educational Philosophy and Theory. 2021; 53(12):1229-45.
[46] Preuveneers D, Garofalo G, Joosen W. Cloud and edge based data analytics for privacy-preserving multi-modal engagement monitoring in the classroom. Information Systems Frontiers. 2021; 23(1):151-64.
[47] Rani M, Gagandeep. Effective network intrusion detection by addressing class imbalance with deep neural networks multimedia tools and applications.

Multimedia Tools and Applications. 2022; 81(6):8499-518.
[48] Park J, Jeong J, Park Y. Ship trajectory prediction based on bi-LSTM using spectral-clustered AIS data. Journal of Marine Science and Engineering. 2021; 9(9):1-22.
[49] Arunkumar KE, Kalaga DV, Kumar CM, Kawaji M, Brenza TM. Forecasting of COVID-19 using deep layer recurrent neural networks (RNNs) with gated recurrent units (GRUs) and long short-term memory (LSTM) cells. Chaos, Solitons & Fractals. 2021; 146:110861.
[50] Wang X, Zhang L, He T. Learning performance prediction-based personalized feedback in online learning via machine learning. Sustainability. 2022; 14(13):1-16.
[51] https://github.com/DeepReSTProject/The-LiruNET-2018-dataset/?tab=readme-ov-file. Accessed 26 February 2024.
[52] https://www.kaggle.com/datasets/junyiacademy/learning-activity-public-dataset-by-junyi-academy. Accessed 26 February 2024.

Valliammal Narayan is working as Assistant Professor (SG) for Avinashilingam Institute for Home Science and Higher Education for Women from 2007 to till date. Her years of experience get added with prior 13 yrs in teaching since 1995 as Assistant Professor in the teaching profession. To her credit she has published 6 Book Chapters and one chapter in Online Special Volume and also has compiled 4 Hand Books for Ancillary course offered to all the majors at University level. She has presented 15 papers in International Conferences. She has also published 25 papers in International Journals. She has also been appreciated as outstanding contributor for introducing ICT based teaching and learning methodology at University level certified by Spoken Tutorial IIT Bombay, 2021.
Email: valliammal_cs@avinuty.ac.in

Sudhamathy Ganapathisamy is currently working as Assistant professor (SS) in the Department of Computer Science in Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore. She has a rich industrial exposure of around 11 years working in various multinational Information Technology companies like Cognizant Technologies Solutions, L&T Infotech, etc. She has worked with International clients and has worked in London for a year. She also has around 13 years of academic and research experience. Her research interests are in Web Mining and Applications of Data Mining using R. She has published a Book "R Programming - An Approach to Data Analytics". She has also published 15 papers in International Journals and have presented several papers in International Conferences.
Email: Sudhamathy_cs@avinuty.ac.in

Appendix I

S. No.	Abbreviation	Description
1	AI	Artificial Intelligence
2	AS-SAN	Adaptive Sparse Self-Attention Network
3	BEM	Bangor Engagement Metric
4	BL	Blended Learning
5	BP	Backpropagation
6	CA	Curriculum Analytics
7	CADA	Canvas Discussion Analytics Dashboard
8	CF	Collaborative Filtering
9	CFS	Correlation-Based Feature Selection
10	CNN	Convolutional Neural Network
11	CoI	Community of Inquiry
12	DBNLS	Deep Belief Network Learning Style
13	DTKT	Deep Temporal Convolutional Network for Knowledge Tracing
14	DW	Data Warehouses
15	EMODASH	Emotional Monitoring and Observation Dashboard for Online Learning
16	FFN	Feed Forward Network
17	HANet	Hybrid Attention Network
18	HiTSKT	Hierarchical Transformer Approach for Session-Aware Knowledge Tracing
19	ICT	Information Communication Technology
20	IOHMM	Input-Output Hidden Markov Model
21	KT	Knowledge Tracing
22	LA	Learning Analytics
23	LAD	Learning Analytics Dashboard
24	LDA	Learning Design-Analytic
25	LMS	Learning Management System
26	MAE	Mean Absolute Error
27	MB	Memory Block
28	MHA	Multi-Head Attention
29	MOOC	Massive Open Online Courses
30	NLP	Natural Language Processing
31	OLCS	Online Learning Climate Scale
32	OSE	Online Student Engagement
33	PBCL	Project-Based Collaborative Learning
34	PESRP	Punjab Education Sector Reform Program
35	RMSE	Root-Mean Squared Error
36	RNN	Recurrent Neural Network
37	SAN-LSTM	Self-Attention Network-Long Short-Term Memory
38	TAMKOT	Transition-Aware Multi-Activity Knowledge Tracing
39	UCOM	University Communication Model
40	VisOJ	Visual Learning Analytics Dashboard For The Online Judge System