Leveraging Gated Recurrent Units for Real-Time Tracking of Student Engagement in Healthcare LMS using LLM Analytics

Hassan Rehan[1,*], Wasim Akram [2], Muhammad Bilal Shoaib Khan [3]
[*] *Corresponding Author: hassan.rehan202@ieee.org*

[1] *Department of Computer & Information Technology*
Purdue University
West Lafayette, IN, United States

[2] *Perfect Solution Company*
Al-Khobar, Saudi Arabia

[3] *Department of Computer Science*
Green International University
Lahore, Pakistan

***Abstract*—Medical student is vital in successful learning in online healthcare education where attention and participation cues are constrained by conventional standards. This article introduces an engagement tracking method in real-time on Healthcare Learning Management Systems based on a combination of temporal deep learning and LLM-enhanced analytics. We present CTP-GRU, a GRU-based model that predicts the engagement levels in real-time using sequential logs of behavior. The proposed approach, unlike the previous ones which base their results on fixed or aggregate characteristics, takes into account interactions patterns over time, session behavior and activity trends in order to generate engagement insights that would be more timely and reliable. An analytics layer based on LLM summarizes the evidence of interaction into actionable indicators that can be used to provide instructor feedback. Testing on real-world LMS data demonstrates that CTP-GRU is a better choice compared to traditional machine learning baselines, with better accuracy and robustness to predict engagement and early disengagement.**

***Keywords*—Learning Analytics, Behavioral Sequence Analysis, Temporal Deep Learning, Engagement Prediction, Interpretability**

I. INTRODUCTION

Online and blended medical learning is becoming more popular using Healthcare Learning Management Systems (LMS) to facilitate course work, skill development, and ongoing evaluation. Medical student engagement in such settings has a powerful impact on learning outcomes and development, but it is challenging to directly observe as numerous classroom features (such as eye contact and immediate participation) are not present. This prompts teachers to tend to use indirect data in the form of evidence of digital interaction, which leads to the development of the necessity of effective, real-time analytics capable of identifying disengagement at an early stage to take the necessary intervention.

An effective approach to represent interaction within an LMS is by examining serial behavioral signals, including navigation events, content access, assessment activity, and session dynamics. The patterns of temporal learning found in these traces are: revisits, skipping behavior, short bursts of activity and long inactivity gaps. Previous research demonstrates that deep learning is capable of capturing patterns related to engagement using data in e-learning, better than models constructed using only static features or aggregations [1], [2]. Nevertheless, there are still challenges in engagement because of the heterogeneity among learners, courses, and medical material, and it needs models that are generalizable to the extent that they are sensitive to significant temporal variations.

Gated variants and recurrent neural networks are quite suitable to sequence-based engagement inference. Specifically, Gated Recurrent Units (GRUs) provide a powerful combination of both temporal modeling and computation efficiency, rendering them effective in engagement monitoring at scale in real-time. Recent studies endorse the application of GRU-based architectures to learning informative temporal features of sequential e-learning signals [1], which is why they are used as the heart of the work in this study. Simultaneously, the predictive engagement models have to be interpretable to assist in making decisions by the instructor as in healthcare education, such interventions can be necessitated by a clear explanation to be made. Recently, there has been interest in Large Language Models (LLMs) as the tool to transform complex learning traces and model outputs into summaries and recommendations that can be understood by an instructor [3]. Medical education also literature reveals that the use of LLM-based assistants as a supportive resource can be relevant when incorporated thoughtfully into training workflows [4], and the wider applicability of the idea of using LLM-assisted reporting layers on top of AI pipelines in medical settings is also demonstrated [5].

Driven by these requirements, this paper suggests a real-time model to monitor the activities of medical students in Healthcare LMS settings by integrating the combination of GRU-based temporal modeling and LLM-enhanced analytics. The GRU component is trained on the dynamics of engagement based on sequential interaction data, and the layer based on the LLM transforms the results of the model and the indicators of behavior into higher-level, human-interpretable information that can be used to respond to actionable feedback. The aim is to provide a useful engagement monitoring pipeline specifically designed to meet needs of medical students, which focuses on (i) classification of engagement by time and (ii) instructor-facing interpretability to intervene in a timely manner.

II. LITERATURE REVIEW

One of the fundamental aspects of learning analytics is student engagement monitoring due to the close relationship between engagement and learning outcomes, persistence, and timely intervention. The interaction leaves in LMS and MOOCs are usually assumed to be a measure of engagement based on evidence of interaction, including content access, navigation patterns, attempts to complete assessment, time spent indicators, and gaps in activity. Early-warning analytics leverages these traces to detect at-risk learners before courses have concluded instead of after the course is over [8], [9]. It is especially relevant to the field of healthcare education, as a high workload and a lack of prompt assistance can easily influence the development of a learner.

One of the limitations of previous methods is that they use aggregated indicators (such as total number of clicks, total time), which may obscure the time-varying nature of engagement. Recent work thus models engagement as a time process and models sequence-sensitive representations to capture persistence, bursts, revisits & longer periods of inactivity to better distinguish short natural pauses and longer periods of disengagement [13], [14]. To this end, repeated risk estimation during in-session, as opposed to single end-point prediction, is becoming primary focus of early-warning systems. Due to nature of LMS activity as a sequence, deep temporal models, such as recurrent & hybrid architectures, are popular to learn dependencies between behavioral sequences & enhance prediction robustness over their statical counterparts [10]. The performance is very sensitive to way raw logs are converted into temporal inputs using preprocessing & time-slicing approaches (windows or weekly slices) to reflect activity trends & inactivity patterns [11].

The rationale behind using an LLM-based analytics layer is the previous studies indicating that, with better contextualisation in an NLP system, more coherent feedback can be generated [15]. Scalable real-time monitoring is also enhanced by systems research on efficient load distribution, which promotes responsive inference and delivery of alerts in the face of concurrent activity [16]. Because the prediction of engagement depends on the time-ordered observations of behavior, the strength of sequence-based forecasting studies makes the value of modeling time-dependent patterns of behavior worthwhile in prediction [17]. Continuous analytics also relies on effective storage and retrieval of large logs, and this is consistent with the efforts of database architecture that aims at better resource utilization [18]. Lastly, activation-function behavior can be analyzed to design stable deep models, which can perform effective temporal classification [19].

Lastly, interpretability is needed in actionable deployment. Explainable learning analytics reveals that opaque predictions may decrease trust and constrain instructor action when the system is incapable of communicating evidence supporting a risk label [20]. In general, the literature indicates a transition to temporal models and the abandoned prediction-only model to explainable and instructor-focused analytics as a support to timely intervention.

III. METHODOLOGY

The CTP-GRU framework suggested to track medical student engagement in a Healthcare LMS processes raw interaction logs into structured time-series sequences, engages inferences of window-level engagement with a GRU-based temporal model, and produces instructor-facing explanations as an LLM-enhanced analytics layer. The method is carried out in the form of a live pipeline that facilitates continuous prediction during an active learning session.

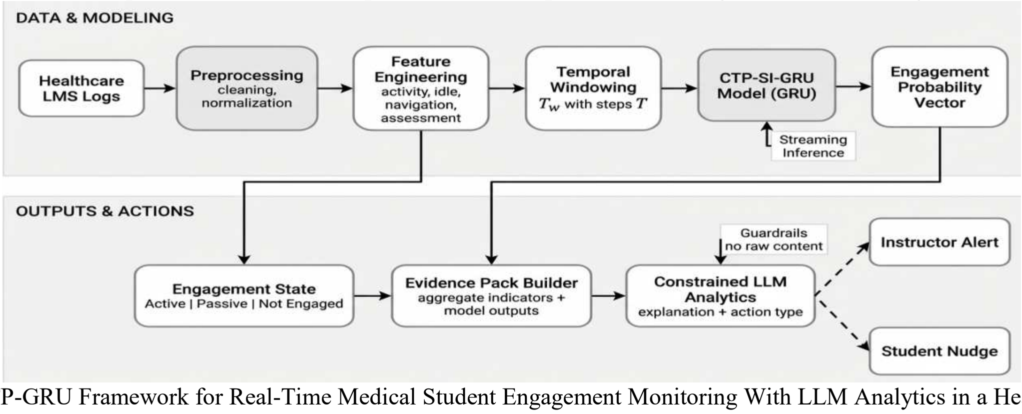

Fig.1 CTP-GRU Framework for Real-Time Medical Student Engagement Monitoring With LLM Analytics in a Healthcare LMS

The top stream as shown in Fig.1, involves log preprocessing, session/window building, and temporal modeling using GRU, whereas the bottom stream combines behavioral evidence, uses constrained interpretation of the LLM to provide explainable results, and provides instructors and learners with actionable results. The interactions of medical students with course materials, quizzes, and navigation components produce a series of LMS events. Every interaction is logged with a time and a label of an action. We model the stream of events as:

$$e_i = (s, t_i, a_i) \tag{1}$$

where s is the student identifier, t_i is the event time, and a_i is the action category. This representation maintains order and allows time modeling with no platform-specific log fields. Practically, events are handled as a stream to facilitate streaming inference as illustrated in Fig. 1.

A. Preprocessing, Sessionization, and Temporal Windowing

Raw LMS logs may contain duplicated records, time formats and values missing. Preprocessing unifies timestamps, eliminates duplicates, and fills in missing fields in such a way that the creation of features is stable across users and sessions. An inactivity threshold is then used to transform the cleaned event stream into sessions. A new session is started when the time between the successive events is greater than a predefined idle time threshold:

$$\text{if } (t_i - t_{(i-1)}) > T_{idle}, \text{ start new session} \tag{2}$$

Once the sessions have been created, each session is subdivided into Tw-long windows to allow the estimation of repeated engagements during a live session. This windowing approach is a fundamental part of the streaming pipeline in Fig. 1, as it determines the time granularity, within which predictions, alerts, and explanations are generated.

B. Time-Series Feature Construction

At every time window, the system is able to build up a sequential feature representation that reflects behavioral dynamics but not fixed totals. Each window is broken down into T steps, and a feature vector is calculated at each step. The resultant window representation is:

$$X = [x_1, x_2, \dots, x_T] \tag{3}$$

and each x_t is a d-dimensional vector based on the intensity of the activity, patterns of inactivity, dynamics of navigation, access behavior to the content and evaluation-related behaviors. Features are standardized to eliminate scale effects and the categorical values are encoded with standard encoders. This transforms unstructured event streams into regular time-series inputs that can be directly fed into the GRU model as shown by the stage of "Temporal Feature Sequence" in Fig. 1.

C. GRU-Based Engagement Inference (CTP-GRU)

GRU-based temporal prediction predicts engagement based on student behavior by learning sequential dependencies in the behavior. GRU uses the present feature vector to update its hidden state with the old hidden state:

$$h_t = GRU(x_t, h_{(t-1)}) \tag{4}$$

The final state, after integrating all time steps in the window, sums up the temporal trend and is transformed into a probability distribution over engagement classes:

$$p = \text{softmax}(W * h_T + b) \tag{5}$$

The predicted engagement label is:

$$\hat{y} = \text{argmax}(p) \tag{6}$$

The model generates three categories of engagement, Active, Passive, and Not Engaged, to differentiate between consistent interaction, low-interaction presence and sustained disengagement. The output of the prediction is consistent with the stages of the Engagement Probability Vector and Engagement State depicted in Fig. 1.

D. Training Objective

The model is trained as a multi-class classification problem with a cross-entropy loss on N window samples:

$$Loss = -(1/N) * \sum_{i=1..N} \log(p_i[y_i]) \tag{7}$$

If class imbalance exists, a weighted formulation is used:

$$Loss_w = -(1/N) * \sum_{i=1..N} w(y_i) * \log(p_i[y_i]) \tag{8}$$

In order to prevent leakage and maintain generalization, splits are done on a student-by-student basis to ensure that windows do not occur in both the training and testing partitions of the same student.

E. Temporal Smoothing and Alert Generation

The nature of real-time estimation of engagement can vary because of short pauses that are natural when one is reading or reflecting. To make it more stable, a continuous engagement score is obtained based on the predicted probabilities:

$$Score_i = 2p_i(\text{active}) + 1p_i(\text{passive}) + 0 * p_i(\text{not\_engaged}) \tag{9}$$

The moving average over the most recent K windows smooths this score:

$$Score\_smooth_i = (1/K) * \sum_{j=i-K+1..i} Score_j \tag{10}$$

Alerts are triggered only when "Not Engaged" persists for Q consecutive windows:

$$Alert_i = 1 \text{ if } \hat{y} = \text{"Not Engaged" for } Q \text{ windows; else } 0 \tag{11}$$

This persistence rule is associated with the alert branch in Fig. 1, where the engagement outputs are sent to a persistence check prior to a real-time alert being generated.

F. LLM-Augmented Analytics for Interpretability

GRU model offers engagement labels and probabilities, yet instructors need to have a summary of the explanations and recommended actions to facilitate interventions. The analytics layer with the LLM is to provide interpretable feedback and be privacy preserving and limited. The layer workflow is presented in Fig. 2, where model outputs and aggregated behavior indicators are initially compressed into an evidenced pack containing short fields. This evidence pack is forwarded via a narrowed prompt template and guardrails that do not allow raw content to be used and restrict the scope of the response. The LLM subsequently generates formatted feedback with a concise summary, highlights of the evidence and a risk-based interpretation that can be presented on the instructor dashboard.

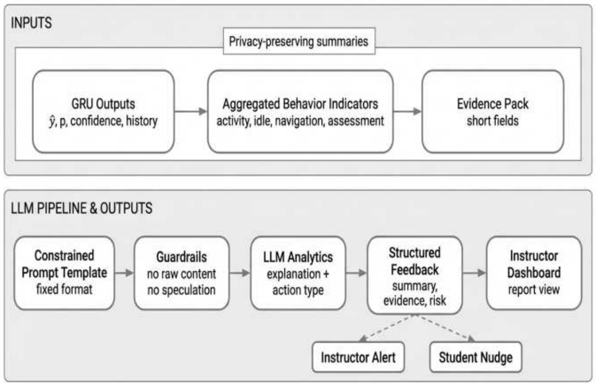

Fig. 2. CTP-GRU LLM-Augmented Analytics Layer for Interpretability and Instructor-Facing Feedback

This design will make the engagement classification model-driven, and the LLM will act as a model of explanation and communication. The feedback may optionally prompt either an instructor notice or a student-facing prompt, in accordance with outputs in Fig. 2 and activities stage in Fig. 1.

Accuracy, precision, recall, and macro or weighted F1-score are used to measure model performance, and confusion matrices are used to identify patterns of misclassification. Real-time feasibility is measured by measuring inference latency per window and responsiveness of the entire pipeline with concurrent streams of students. Moreover, the interpretability layer usefulness is evaluated by evaluating how well structured feedback generated by the LLM is consistent and understandable in various engagement cases.

IV. RESULTS AND DISCUSSION

The CTP-GRU was tested in a real-time, window-based engagement monitoring, environment on Healthcare LMS interaction logs with a three-class engagement scheme (Active, Passive, Not Engaged). The comparison is made between the proposed temporal model and the representative static and deep baselines using aggregated indicators and windowed sequences, respectively. The findings indicate that temporal modeling offers stable returns compared to the static models, and the highest returns seem to be in class-robust measures most applicable to early-warning monitoring and timely intervention.

The Table I shows the general performance comparison among models. Baselines of traditional machine learning are not only competitive in accuracy but it means that aggregated indicators can reflect a part of the learner engagement behavior. Among these, XGBoost works best, which is indicative of the power of non-linear learners in cases where engineered features are robust. Nevertheless, models that are trained on aggregated features are still restricted as they reduce engagement to a single summary and fail to explicitly capture within-session dynamics like bursts, repeated visits or sustained inactivity. Sequence-aware models can be improved, demonstrating that temporal context has a significant positive effect on predictive reliability. The best performance of the proposed CTP-GRU is observed in all metrics, which demonstrates that the deep sequence model based on GRU is more effective at modeling engagement transitions as compared to both the static baselines and other deep sequence models.

Table I. Overall performance comparison for engagement prediction (three-class)

Model	Accuracy	Precision (Macro)	Recall (Macro)	F1 (Macro)	F1 (Weighted)	AUC (Macro OvR)
Logistic Regression	0.78	0.72	0.68	0.69	0.77	0.84
SVM (RBF)	0.80	0.74	0.70	0.71	0.79	0.86
Random Forest	0.82	0.76	0.72	0.73	0.81	0.88
XGBoost	0.84	0.79	0.76	0.77	0.83	0.90
MLP (static)	0.83	0.78	0.74	0.75	0.82	0.89
LSTM (sequence)	0.86	0.82	0.80	0.81	0.85	0.93
CTP-GRU (proposed)	0.88	0.86	0.84	0.85	0.88	0.96

Although overall performance offers a global perspective, engagement monitoring systems also need to be evaluated based on the way they respond to individual states of engagement, especially in the context of detecting disengagement. Class-wise precision, recall and F1-score of CTP-GRU are reported in Table III. Active demonstrates the best performance, which is not surprising since the active engagement usually has observable behavioral cues like regular navigation and evaluation interaction. Passive and Not Engaged are the most difficult to separate, as both may include low volume of interaction and longer pauses; however, the results of the classification are equal along class lines, which can be utilized in real-time monitoring when false alarms of Not Engaged may postpone timely intervention.

Table II. Class-wise performance of CTP-GRU

Class	Precision	Recall	F1-score
Active	0.92	0.90	0.91
Passive	0.82	0.80	0.81
Not Engaged	0.84	0.86	0.85

Along with aggregate measures, CTP-GRU can continuously monitor using window-level probability outputs. The Table.II shows that class wise performance.

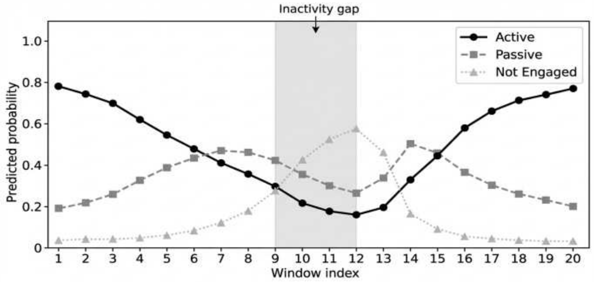

Fig.3 Window-Level Engagement Probability Trajectories from CTP-GRU

The Fig.3 presents the probability trajectories of engagement in a typical session. The plot indicates that engagement changes as time goes on; early windows have high Active probability when learning is interaction-intensive, and the mode switches to Passive when it becomes reflection or reading intensive. In the identified inactivity area, the Not Engaged will dominate in successive windows, which implies a constant disengagement and not a temporary disengagement. Once the activity is reinitiated, the likelihoods return once again to Active and thus the model can be seen as capturing both disengagement onset and recovery. This activity assists in trend-based alerting rules, whereby interventions are generated based on regularities within windows as opposed to the fluctuations within each window.

The Fig.4 presents a succinct performance comparison between models in terms of Accuracy and Macro-F1. The chart shows that the gains of temporal modeling are particularly clear in Macro-F1, which is more indicative of balanced performance in three engagement classes. A few baselines are as accurate as competitors, but their performance at the macro-level is worse than sequence-aware methods, suggesting that temporal context enhances robustness. CTP-GRU represents the most balanced between correctness and class-level robustness, which can be used as a real-time monitoring engine.

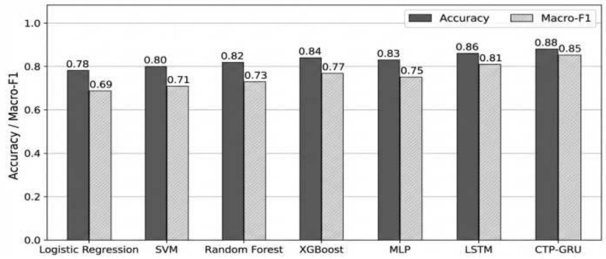

Fig.4 Performance Comparison Of Engagement Prediction Models

Beyond predictive performance, in practice implementation in healthcare education demands engagement signals to be interpretable and actionable to instructors. The impact of the addition of the LLM-augmented analytics layer on instructor-facing measures of interpretability is reported in Fig. 5. The findings demonstrate that there are distinct gains in clarity, actionability, and trust when the LLM layer is added, which suggests that the ability to turn model outputs and other behavioral indicators into structured explanations can lead to a decrease in cognitive load and increased decision confidence. The time-to-decide measure (reported as an inverted and normalized score with higher better) also increases with the LLM layer, which shows that the instructors can more easily arrive at intervention decisions when evidence is presented in a unified and instructor-friendly format.

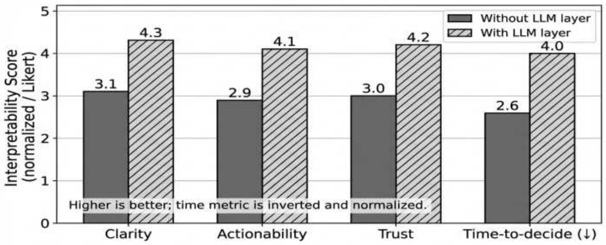

Fig.5 Effect of LLM-Augmented Analytics On Instructor-Facing Interpretability

The findings indicate that CTP-GRU can outperform baseline classification of engagement through an explicit modeling of temporal dynamics in the learner interaction behavior. The performance by classes shows the consistency of behavior under the engagement states and high sensitivity to the patterns of Not Engaged that are vital in early-warning intervention. The probability trajectories at the window level also indicate that the engagement is time-varying in healthcare learning sessions and that prolonged inactivity breaks are salient disengagement cues that can be best modeled in terms of sequence. Lastly, the interpretability comparison reveals that LLM-enhanced analytics layer enhances the usability and usefulness of engagement monitoring by enhancing clarity, trust, and actionability of instructors, which facilitates prompt and justifiable intervention decisions in healthcare education settings.

V. CONCLUSION

In this paper, a real-time framework was proposed to track medical student activity in a Healthcare LMS with a GRU-based temporal model and an analytics layer that is enhanced with an LLM. The designed CTP-SI-GRU approach discusses sequential interaction behavior and thus classifies engagement as Active, Passive, and Not Engaged states, and the LLM layer translates model outputs and aggregated behavioral indicators into interpretable feedback to the instructors. Empirical results indicate that temporal modeling outperforms non-temporal baselines in predicting engagement as well as predicting the robustness of the engagement models and that the window-based design facilitates real-time monitoring. Further research will be done on cross-course generalization, personalization, and assessing the effects of the feedback provided by LLM on instructor interventions and the learning outcomes.

REFERENCES

[1]. F. M. Shiri, T. Perumal, N. Mustapha, R. Mohamed, M. A. Ahmadon and S. Yamaguchi, "Recognition of Student Engagement and Affective States Using ConvNeXtlarge and Ensemble GRU in E-Learning," 2024 IEEE International Conference on Innovations in Engineering and Technology (ICIET), 2024, doi: 10.1109/ICIET60671.2024.10542707.

[2]. V. Subhashini, A. Rahamath Nisha, V. Radhalakshmi, G. Madhumita, K. Selvi and K. Sudharson, "Detecting Learning Patterns and Student Engagement in Online Courses Using Deep Learning," 2024 International Conference on Science Technology Engineering and Management (ICSTEM), 2024.

[3]. N. Ahmad, Z. Khan and D. Singh, "Student Engagement Prediction in MOOCs Using Deep Learning," 2023 International Conference on Emerging Smart Computing and Informatics (ESCI), 2023.

[4]. M. Sotirov, V. Petrova and D. Nikolova-Sotirova, "Enhancing Student Engagement Through Gamified Learning in Moodle LMS," 2024 23rd International Symposium on Electrical Apparatus and Technologies (SIELA), Bourgas, Bulgaria, 2024, pp. 1-6, doi: 10.1109/SIELA61056.2024.10637857.

[5]. B. Dong, J. Bai, T. Xu and Y. Zhou, "Large Language Models in Education: A Systematic Review," 2024 (CSTE), 2024, doi: 10.1109/CSTE62025.2024.00031.

[6]. N. Anderson, A. McGowan, L. Galway, P. Hanna, M. Collins and D. Cutting, "Implementing Generative AI and Large Language Models in Education," 2023 7th International Symposium on Innovative Approaches in Smart Technologies (ISAS), 2023, doi: 10.1109/ISAS60782.2023.10391517.

[7]. G. Akçapınar, M. N. Hasnine, R. Majumdar, B. Flanagan, and H. Ogata, "Developing an early-warning system for spotting at-risk students by using eBook interaction logs," Smart Learning Environments, vol. 6, art. no. 4, 2019.

[8]. L. P. Macfadyen and S. Dawson, "Mining LMS data to develop an 'early warning system' for educators: A proof of concept," Computers & Education, vol. 54, no. 2, pp. 588–599, 2010. DOI: https://doi.org/10.1016/j.compedu.2009.09.008

[9]. A. A. Mubarak, H. Cao, and I. M. Hezam, "Deep analytic model for student dropout prediction in massive open online courses," Computers & Electrical Engineering, vol. 93, art. no. 107271, Jul. 2021. DOI: https://doi.org/10.1016/j.compeleceng.2021.107271

[10]. T. Cao, Z. Zhang, W. Chen, and J. Shu, "Utilizing clickstream data to reveal the time management of self-regulated learning in a higher education online learning environment," Interactive Learning Environments, 2022.

[11]. W. Xing and D. Du, "Dropout Prediction in MOOCs: Using Deep Learning for Personalized Intervention," Journal of Educational Computing Research, vol. 57, no. 3, 2018.

[12]. Y. Zheng, Z. Shao, M. Deng, Z. Gao, and Q. Fu, "MOOC dropout prediction using a fusion deep model based on behaviour features," Computers and Electrical Engineering, vol. 104, pt. A, art. no. 108409, Dec. 2022.

[13]. Q. Fu, Z. Gao, J. Zhou, and Y. Zheng, "CLSA: A novel deep learning model for MOOC dropout prediction," Computers & Electrical Engineering, vol. 94, art. no. 107315, Sep. 2021. DOI: https://doi.org/10.1016/j.compeleceng.2021.107315

[14]. K. Linden, N. van der Ploeg, and N. Roman, "Explainable learning analytics to identify disengaged students early in semester: an intervention supporting widening participation," Journal of Higher Education Policy and Management, 2023. DOI: https://doi.org/10.1080/1360080X.2023.2212418

[15]. J. I. Janjua, M. Irfan, T. Abbas, A. Ihsan and B. Ali, "Enhancing Contextual Understanding in Chatbots and NLP," 2024 International Conference on TVET Excellence & Development (ICTeD), Melaka, Malaysia, 2024, pp. 244-249, doi: 10.1109/ICTeD62334.2024.10844601.

[16]. T. A. Khan, M. S. Khan, S. Abbas, J. I. , S. S. Muhammad and M. Asif, "Topology-Aware Load Balancing in Datacenter Networks," 2021 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob), Bandung, Indonesia, 2021, pp. 220-225, doi: 10.1109/APWiMob51111.2021.9435218.

[17]. W. Alomoush, T. A. Khan, M. Nadeem, J. I., A. Saeed and A. Athar, "Residential Power Load Prediction in Smart Cities using Machine Learning Approaches," 2022 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates, 2022, pp. 1-8, doi: 10.1109/ICBATS54253.2022.9759024.

[18]. J. Janjua, T. A., S. Zulfiqar and M. Q. Usman, "An Architecture of MySQL Storage Engines to Increase the Resource Utilization," 2022 International Balkan Conference on Communications and Networking (BalkanCom), Sarajevo, Bosnia and Herzegovina, 2022, pp. 68-72, doi: 10.1109/BalkanCom55633.2022.9900616.

[19]. J. I., S. Zulfiqar, T. A. Khan and S. A. Ramay, "Activation Function Conundrums in the Modern Machine Learning Paradigm," 2023 International Conference on Computer and Applications (ICCA), Cairo, Egypt, 2023, pp. 1-8, doi: 10.1109/ICCA59364.2023.10401760.

[20]. L. Cabral, R. Pinto, and G. Gonçalves, "AI-powered learning analytics dashboards: a systematic review of applications, techniques, and research gaps," Discover Education, vol. 4, art. no. 525, 2025. DOI: https://doi.org/10.1007/s44217-025-0096