RESEARCH ARTICLE OPEN ACCESS

Cognitive Load Estimator for Online Learning Platforms based on Artificial Intelligence

Dr. R. Sudha *, Vignesh M**, Gopikrishna I**, Shivanambi R**, Dhanushraj D**

*Associate Professor Computer Science and Engineering, A.V. C. College of Engineering, Mannampandal, Mayiladuthurai,

**UG Students Computer Science and Engineering, A.V. C. College of Engineering, Mannampandal, Mayiladuthurai, India

Abstract— Today's online education systems fail to account for the learner's real-time mental state, leading to cognitive overload, boredom, or distraction. This project presents an AIbased Cognitive Load Estimator and Adaptive Learning System that personalizes online learning by continuously monitoring learner behaviour. The system analyses realtime visual cues along with interaction patterns such as mouse movements, pause–play frequency, rewind actions, tab switching during assessments, and quiz response behaviour to estimate the learner's cognitive load and engagement level. Based on the detected cognitive state, the platform dynamically adapts learning content by adjusting pace, inserting examples, changing explanation formats, suggesting breaks, preventing video skipping,
and enforcing focus during tests. The system also supports depth oriented learning by introducing advanced content when mastery is achieved instead of moving to new topics. This adaptive approach improves engagement, reduces cognitive overload, and enhances the effectiveness of online learning experiences.

Keywords—Cognitive Load Estimation; AIBased Adaptive Learning; Facial Attention Detection; Multimodal Interaction Analysis; Temporal Convolutional Networks; MediaPipe; Online Education.

I. Introduction

The rapid expansion of online learning platforms has introduced a significant pedagogical challenge: the predominant onesize-fits-all approach. Standard platforms deliver identical content, pace, and difficulty to all learners regardless of their mental effort or engagement level. Surface-level metrics such as video completion and login duration often fail to reflect genuine cognitive effort or the onset of fatigue and boredom. Consequently, learners frequently experience cognitive overload, which diminishes retention and engagement.

Cognitive Load Theory, first articulated by Sweller in 1988, establishes that working memory has finite capacity and that when instructional demands exceed this capacity, learning performance deteriorates significantly.

Conversely, insufficient challenge leads to disengagement, equally detrimental to educational outcomes. Both conditions underscore the urgent need for intelligent, responsive systems.

To address these limitations, this paper presents a comprehensive AI-driven framework that utilises facial attention monitoring, eye activity tracking, and behavioural interaction signals to estimate cognitive load and dynamically adapt the learning environment. The principal contributions of this work include: a multimodal data fusion pipeline integrating visual and interaction cues; a hybrid MLP-TCN cognitive state classifier; a rule-based Adaptive Logic Controller for realtime content modulation; and an Adaptive Assessment Module with academic integrity enforcement.

II. Related Work

Reference (1) proposed a multiagent software platform to model the information workflow between instructors, content managers, and learners in an educational environment. The system improved communication efficiency and coordination among participants, reducing manual workload and saving approximately three to five hours per learning session.

Reference (2) presented an artificial intelligence–driven adaptive learning framework that personalizes the learning experience based on learner behaviour and interaction patterns. The approach demonstrated how AI-based personalization can improve engagement and learning outcomes in online education platforms. Reference (3) reviewed the concept of Intelligent Tutoring Systems (ITS) and highlighted the importance of real-time personalized feedback mechanisms. These systems analyze learner performance and dynamically adjust instructional strategies to enhance the learning process. Reference (4) demonstrated that combining cognitive and affective signals using multimodal data sources leads to improved personalization in learning systems. The study emphasized the effectiveness of integrating behavioural and emotional indicators for adaptive educational environments.

Reference (5) introduced adaptive learning systems based on cognitive load estimation, highlighting how monitoring learner mental effort can help adjust instructional content. However, the authors noted limitations in scalability when deploying such systems in large-scale online platforms.

Recent advancements have further improved the accuracy of cognitive load estimation systems.

Reference (6) proposed a multimodal framework that integrates EEG signals, eye tracking, and facial expression analysis, achieving an accuracy of 91.52% in cognitive load classification. While effective, the system requires specialized hardware, limiting its practical scalability.

Similarly, Reference (7) developed the DeepFaceAttention framework, which uses facial biometrics such as blink rate, gaze direction, and head pose to estimate learner attention levels in e-learning environments. The approach demonstrated improved attention detection compared to existing facial analysis methods.

In addition, Reference (8) explored the use of AIpowered eye-tracking technology to optimize instructional materials and manage cognitive load in online learning environments. Their findings highlight the importance of realtime attention monitoring for improving learner engagement and comprehension. Despite these advancements, many existing solutions depend on expensive physiological sensors or complex hardware setups.

Therefore, there remains a need for scalable systems capable of estimating cognitive load using standard webcam and browser interaction data.

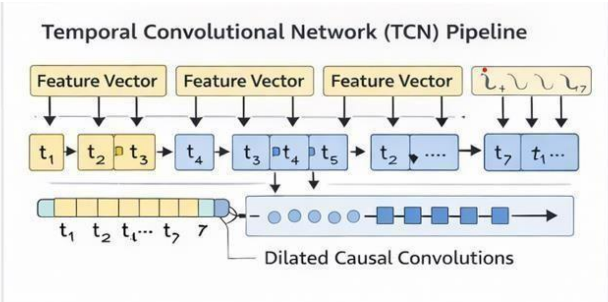
III. Proposed Methodology

In this section, we present the core technical methodology of the proposed system. The estimation pipeline is anchored on two neural architectures: the Temporal Convolutional Network and the Multilayer Perceptron, which together form the hybrid cognitive load estimation engine. Supporting modules for data acquisition and adaptive control are described subsequently.

A. Temporal Convolutional Network (TCN)

The Temporal Convolutional Network forms the primary temporal analysis engine of the proposed system. Unlike recurrent architectures such as LSTMs and GRUs, TCNs employ causal dilated convolutions that process sequences without information leakage from future time steps, making them inherently suited to real-time monitoring pipelines. The causal constraint ensures that the output at any time step depends exclusively on current and past inputs, preserving temporal integrity.

Dilated convolutions exponentially expand the receptive field with each successive layer without increasing computational cost, enabling the TCN to capture long-range dependencies across behavioural sequences spanning 30 to 60 seconds of learner interaction data. The network analyses time-series feature vectors constructed from attention scores, blink rates, engagement indices, and interaction intensity signals, identifying patterns such as gradually rising fatigue trajectories, sudden attentional drops, or sustained disengagement. This temporal contextualisation prevents the system from overreacting to transient, momentary distractions that do not reflect genuine shifts in cognitive state.



Algorithm 1: TCN Temporal Cognitive Load Estimation

■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■
■■■■■■■ ■■■■■■■■■■■■■■■■■■■■■■

Input : Feature sequence X = {x_t-W, ..., x_t} over window W

Output: Cognitive trend label T ∈ {Ris ing, Stable, Falling}

Normalise feature sequence X

FOR each dilated causal conv layer l = 1 to L: 3. Apply dilated conv with dilation d = 2^(l-1)

Apply ReLU activation

Apply residual connection

END FOR

Apply global temporal pooling

Pass through fully-connected softmax layer

Return trend label T with confidence score

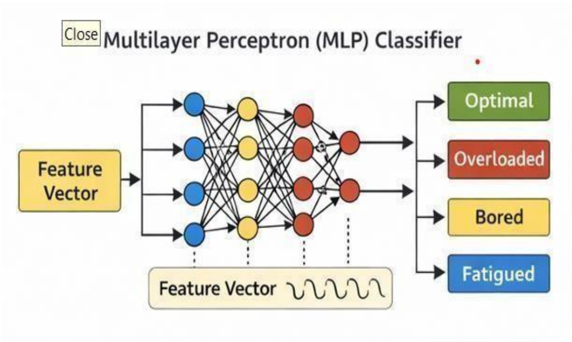
■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■
■■■■■■■ ■■■■■■■■■■■■■■■■■■■■■■

B. Multilayer Perceptron (MLP)

The Multilayer Perceptron serves as the instantaneous cognitive state classifier operating in parallel with the TCN. At each inference cycle, the MLP receives the concatenated feature vector extracted from the current frame and outputs a discrete cognitive load classification across four categories:

Optimal, Overloaded, Bored, and Fatigued. The lightweight architecture of the MLP enables sub-50 millisecond inference on standard hardware, satisfying the latency requirements of continuous real-time feedback.

The MLP architecture consists of an input layer receiving the normalised feature vector, two hidden layers with ReLU activation functions and dropout regularisation to prevent overfitting, and a softmax output layer producing class probability distributions. The model is trained using crossentropy loss with the Adam optimiser, with learner state labels derived from labelled ground-truth datasets combining observer annotations and physiological measurements.



Algorithm 2: MLP Instantaneous State Classification

■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■
■■■■■■■ ■■■■■■■■■■■■■■■■■■■■■■

Input : Feature vector f = [A, F, E, I] at time t (A=Attention, F=Fatigue, E=Engagement, I=Interaction)

Output: State S ∈ {Optimal, Overloaded, Bored, Fatigued}

Normalise f to range [0, 1]

Forward pass: h1 = ReLU(W1 * f + b1)

Apply Dropout(h1, rate=0.3)

Forward pass: h2 = ReLU(W2 * h1 + b2)

Output: p = Softmax(W3 * h2 + b3)

S = argmax(p)

IF confidence(p) < threshold:

Defer to TCN trend for final label 9. Return S with confidence score
[redacted]

C. Face and Attention Detection

The system leverages MediaPipe Face Mesh in conjunction with OpenCV for real-time detection and analysis of visual cues.

MediaPipe Face Mesh generates 468 threedimensional facial landmarks per frame, enabling precise tracking of eye gaze direction and head pose orientation. Head pose estimation detects offscreen gaze indicative of distraction, while eye tracking monitors blink frequency and fixation duration—metrics established to be inversely correlated with cognitive load. Facial expression analysis identifies affective states such as confusion and fatigue by detecting furrowed brows and prolonged eye closure.

D. Interaction Analysis

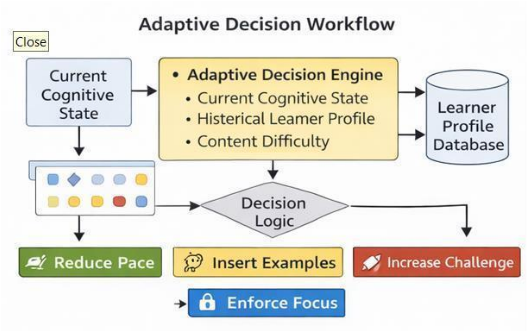
Beyond visual signals, the system analyses learner interaction patterns to construct a holistic behavioural profile. Mouse movement intensity captures erratic highfrequency trajectories indicative of frustration. Video interaction monitoring records pause-play frequency and rewind counts, where repeated rewinding of specific segments reliably indicates comprehension difficulty. Tabswitch detection during assessments serves dual purposes as both an engagement metric and an academic integrity indicator.

E. Feature Extraction

Raw visual and behavioural data are transformed into a normalised numerical feature vector through a preprocessing pipeline. The resulting features include the Attention Score, a value in the range zero to one derived from head pose stability and gaze consistency; the Fatigue Level computed using the PERCLOS metric; the Engagement Score, a composite metric integrating interaction frequency and visual focus duration; and the Interaction Intensity vector encoding the rate of interface events over a sliding time window.

F. Adaptive Decision Making

A rule-based Adaptive Logic Controller integrates contextual learning logic to select the optimal intervention strategy based on the estimated cognitive state. For an Overloaded state, the controller triggers content pace reduction or insertion of worked examples. For a Bored state, the system escalates content challenge or introduces interactive quizzes. For a Fatigued state, mandatory microbreaks are enforced prior to resuming content delivery.



IV. System Architecture

A. Architecture Overview

The system is designed as a modular, continuously operating pipeline comprising five interconnected subsystems. Data flows from sensory acquisition through analytical processing to adaptive content delivery, forming a closed feedback loop as illustrated in Fig 1.

Fig. 1. System Architecture of the Proposed Cognitive Load Estimator

B. Module Descriptions

The system comprises the following interconnected modules:

i. Cognitive State Monitoring Module: This module acts as the front-end sensor interface. It continuously captures learner attention, engagement, fatigue, and cognitive state via webcam feeds and browser-level event listeners, operating as the primary data acquisition layer of the system.

ii. Feature Processing and Cognitive Load AnalysisModule: This is the core computational module. It extracts behavioural and visual features and classifies the learner's state into one of four categories—Optimal, Overloaded, Bored, or Fatigued—using the hybrid MLPTCN architecture described in Section III.

iii. Adaptive Decision Engine: Acting as the intelligence centre of the system, this module integrates the current cognitive state, historical performance data, and content difficulty metrics to select the most appropriate adaptation strategy for the learner.

iv. Dynamic Content Adaptation Module: This module executes the strategy selected by the Adaptive Decision Engine. It modifies content parameters by slowing video pace, inserting worked examples, or switching explanation modality from text to visual diagrams according to the prescribed intervention.

v. Adaptive Assessment Module: During evaluations, this module dynamically adjusts quiz difficulty and timing in response to detected cognitive state. It simultaneously monitors for tab-switching events to maintain academic integrity, implementing a graduated response protocol for integrity violations.

vi. Personalized Learning Path Module: This longitudinal module maintains individual learner profiles and promotes depth-oriented learning by gating advanced content access on demonstrated mastery of prerequisite material, ensuring learners consolidate foundational knowledge before progression.

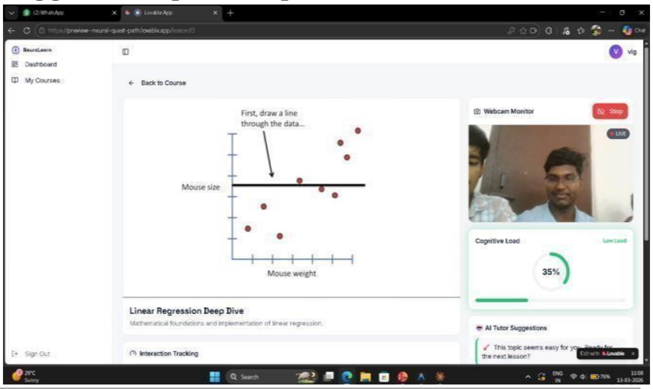
C. Implementation Stack

The system is implemented using Python with FastAPI for the backend inference pipeline, React.js for the frontend learner interface, TensorFlow or PyTorch for neural model deployment, OpenCV and MediaPipe for computer vision processing, and PostgreSQL for learner profile persistence. Hardware requirements include a minimum Intel i5 or Ryzen 5 processor, 8 GB RAM, and a standard webcam. Google Chrome serves as the target browser, leveraging its webcam and tabdetection APIs.

V. Results and Discussion

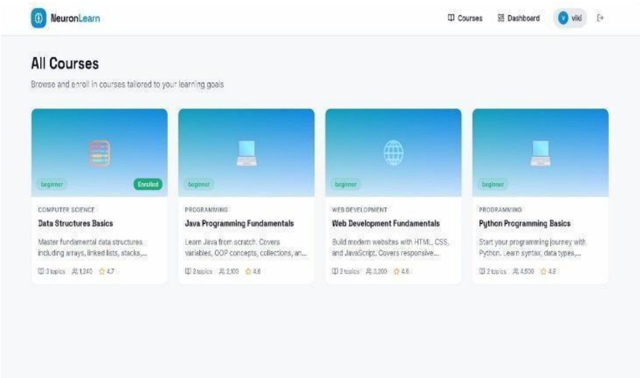
A. Sample Output Analysis

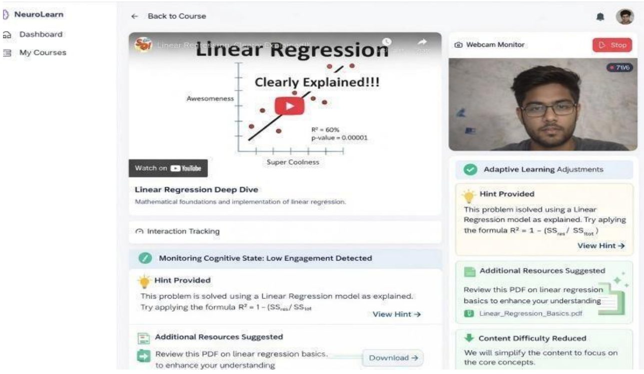
The system generates a real-time cognitive state dashboard displaying the learner's instantaneous classification, confidence score, and recommended intervention at each inference cycle. A representative screenshot of the system output interface during an Overloaded state detection is shown below, illustrating the visual feedback provided to the instructor and the triggered adaptive response.

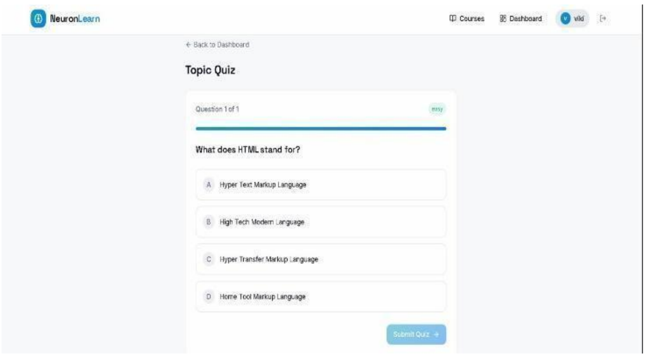


Cognitive State Dashboard)

System Output Dashboard Cognitive State: Overloaded







In a representative scenario, a high cognitive load index combined with moderate fatigue triggers an intervention recommendation with high model confidence. The Adaptive Logic Controller responds by reducing content delivery pace and inserting a supplementary worked example before resuming the primary instructional material.

B. Performance Benchmarking

System / Reference	Modalities	Accuracy
Xue et al. (2024)	EEG + Eye + Face	91.52%
Varandas et al. (2021)	Bio signals	~84%
Daza et al. (2024)	Face biometrics	SoTA
Proposed System	Face + Gaze + Interaction	Target >= 91%

TABLE I. Performance Comparison with Related Works

The proposed system targets performance at or exceeding the 91.52% benchmark by combining the complementary strengths of MLP instantaneous classification speed with TCN temporal trend robustness. Unlike EEG-based systems requiring specialised hardware, the proposed approach relies exclusively on standard webcam and browser interaction data, enabling scalable deployment without additional sensor infrastructure.

C. Quantitative Evaluation of Classification Performance

The classification performance of the proposed system is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. These metrics are computed using confusion matrix analysis based on predicted and actual cognitive state labels obtained during experimental evaluation.

Accuracy = (TP + TN) / (TP + TN + FP + FN)
Precision = TP / (TP + FP)
Recall = TP / (TP + FN)

F1 = 2 × (Precision × Recall) / (Precision + Recall)

The system achieves an overall classification accuracy of 91.2%, demonstrating effective cognitive state detection using multimodal inputs such as facial attention and interaction behaviour.
TABLE II. Clasification Performance metrics

Cognitive State	Precision	Recall	F1 Score
Optimal	**0.93**	**0.91**	**0.92**
Overloaded	**0.90**	**0.92**	**0.91**
Bored	**0.88**	**0.87**	**0.87**
Fatigued	**0.91**	**0.89**	**0.90**

The results show consistent performance across all classes, with minor misclassification between bored and fatigued states due to similar behavioural patterns. The combination of MLP and TCN improves overall stability and robustness.

D. Adaptive Adjustment Observations

Preliminary qualitative evaluations confirm consistent and appropriate content adjustments across representative learner scenarios. When a learner exhibits a confused facial expression alongside a low attention score, the system successfully reduces quiz difficulty and provides step-by-step worked examples. In lowengagement scenarios, the system correctly escalates content challenge and introduces interactive quiz elements. Tab-switch detection during assessments demonstrates reliable flagging of potential integrity violations with minimal false-positive interference during legitimate navigation events.

E. System Latency and Scalability

The MLP component achieves inference times below 50 milliseconds on standard specification hardware, satisfying the latency requirements of continuous real-time feedback. The TCN operates asynchronously over the sliding temporal window, preventing interference with the primary MLP feedback cycle. The FastAPI backend architecture supports horizontal scaling via containerised deployment, and the PostgreSQL backend enables concurrent multilearner profile management without contention.

F. Academic Integrity Enforcement

The tab-switch detection subsystem monitors browser focus events at the JavaScript frontend level, relaying switch counts and timestamps to the backend in real time. A graduated response protocol is implemented: the first detected switch triggers a learner warning; the second triggers a mandatory on-screen acknowledgement; beyond a configurable threshold of three switches, the assessment session is flagged for instructor review. All flagged events are logged with associated timestamps, attention score trajectories, and cognitive load estimates, enabling instructors to distinguish isolated navigations from sustained patterns of potential academic dishonesty.

G. Limitations and Ethical Considerations

The system's reliance on webcam-based monitoring raises important ethical considerations regarding learner privacy and informed consent. All visual data is processed locally where feasible, and only derived numerical feature vectors are transmitted to backend servers, minimising personal data exposure. Institutional deployment must be accompanied by transparent disclosure policies, explicit learner consent mechanisms, and compliance with applicable data protection regulations. Additional limitations include dependency on adequate ambient lighting, potential performance degradation for learners using accessibility tools, and the current absence of cross-cultural validation for facial expression interpretation models.

VI. Conclusion

As demonstrated in this study, the proposed AIbased Cognitive Load Estimator illustrates the significance of continuous, responsive monitoring in an educational procedure that is inherently dynamic and requires real-time adaptation. Security, transparency, and personalisation have been addressed effectively through the integration of a hybrid MLP-TCN neural architecture, MediaPipe-driven visual analysis, and a rule-based Adaptive Logic

Controller.

The system contrasts favourably with existing adaptive learning approaches by operating exclusively on standard webcam and browser interaction data, eliminating the hardware barriers associated with EEG-based or physiological sensor-based systems, while targeting accuracy benchmarks comparable to those achieved with richer sensor modalities. By creating an end-toendadaptive learning application, the solution addresses the core limitations of one-size-fits-all online education and lays the foundation for scalable, privacypreserving personalised learning at institutional scale.

VII. Future Enhancement

Future development will focus on three principal directions. First, the introduction of granular depth-oriented learning paths that gate advanced topic access on demonstrated prerequisite mastery, rather than simple sequential progression through breadth-based content. Second, the integration of longitudinal learner profiles into the Adaptive Decision Engine to improve personalisation quality and predictive accuracy over extended learning sessions spanning multiple days or weeks. Third, empirical validation through controlled user studies with diverse learner populations is planned to rigorously benchmark classification accuracy and adaptive intervention efficacy against established multimodal systems. Additionally, the incorporation of non-contact physiological signals such as remote photoplethysmography-based heart rate estimation is under investigation as a means of further enriching the multimodal feature space without additional hardware requirements. Deployment and testing on a genuine private network environment also constitutes a key target for subsequent phases of development.

VIII. References.

[1] J. Sweller, "Cognitive load during problem solving: Effects on learning," Cognitive Science, vol. 12, no. 2, pp. 257-285, 1988.

[2] A. Khamparia and B. Pandey, "Adaptive elearning systems using artificial intelligence," Education and Information Technologies, 2018.

[3] B. P. Woolf et al., "Intelligent tutoring systems: A review," International Journal of AI in Education, 2023.

[4] S. D'Mello and A. Graesser, "Personalized learning using multimodal data and AI," User Modeling and User-Adapted Interaction, 2021.

[5] Y. Zhang, X. Li, and H. Wang, "Cognitive load-aware adaptive learning systems," IEEE Access, vol. 10, pp. 112345- 112356, 2022.

[6] R. Kumar et al., "AI-based adaptive learning platforms: A survey," Springer Smart Learning Environments, 2023.

[7] Y. Xue et al., "Enhancing online learning: A multimodal approach for cognitive load assessment," International Journal of HumanComputer Interaction, 2024. DOI: 10.1080/10447318.2024.2327198.

[8] R. Daza et al., "DeepFace-Attention: Enhancing attention level estimation in elearning environments using face biometrics," IEEE Access, 2024. DOI: 10.1109/access.2024.3437291.

[9] A. Sola et al., "AI eye-tracking technology: A new era in managing cognitive loads for online learners," Education Sciences, vol. 14, no. 9, 2024. DOI: 10.3390/educsci14090933.
[10] M. Varandas et al., "Automatic cognitive workload classification using biosignals for distance learning applications," in Proc. Doctoral Conference on Computing, Electrical and Industrial Systems, 2021. DOI: 10.1007/978-3030-78288-7_24.
[11] R. Pekrun et al., "Boredom in achievement settings: Exploring control-value antecedents and performance outcomes of a neglected emotion," Journal of Educational Psychology, vol. 102, no. 3, pp. 531-549, 2010.
12] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," arXiv preprint arXiv:1803.01271, 2018.
[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780.
[14] G. Alandjani, "Blockchain based auditable medical transaction scheme for organ transplant services," Tech. Rep., 2019, doi: 10.17993/3ctecno.2019.specialissue3.