Cognitive Load Optimization in Digital (ESL) Learning: A Hybrid BERT and FNN Approach for Adaptive Content Personalization

Dr. Komminni Ramesh[1], Dr Christine Ann Thomas[2], Dr Joel Osei-Asiamah[3], Dr. Bhuvaneswari Pagidipati[4], Elangovan Muniyandy[5], B.V.Suresh Reddy[6], Prof. Ts. Dr. Yousef A.Baker El-Ebiary[7]

Assistant Professor of English, Chairperson, BoS, Anurag Engineering College, Kodad, Suryapet District, Telangana, India[1]
Assistant Professor, Department of English and Cultural Studies, Christ University, Bengaluru, India[2]
Graduate Research Fellow, Department of Science and Technology Education, University of South Africa (Unisa) Pretoria, Gauteng Province, South Africa[3]
Associate Professor of English (Ratified by JNTU K), Dept. of English and Foreign Languages, Sagi Rama Krishnam Raju Engineering College (A), Bhimavaram – 534204, West Godavari Dt, Andhra Pradesh, India[4]
Department of Biosciences-Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India[5]
Applied Science Research Center, Applied Science Private University, Amman, Jordan[5]
Assistant Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India[6]
Faculty of Informatics and Computing, UniSZA University, Malaysia[7]

***Abstract*—Traditional English as a Secondary Language (ESL) learning platform rely on static content delivery, often failing to adapt to individual learners' cognitive capacities, leading to inefficient comprehension and increased cognitive load. A novel hybrid Feedforward Neural Network and Bidirectional Encoder Representation Transformer (FNN-BERT) framework stands as our solution because it performs dynamic content personalization through predictions of real-time cognitive load. The proposed approach incorporates Feedforward Neural Networks (FNN) alongside Bidirectional Encoder Representations from Transformers (BERT) to process behavioral analytics for optimized content complexity adjustment and adaptive and scalable learning delivery. Real-time adaptability, scalability and high computational needs of current models reduce their effectiveness in personalized learning environments. Through the application of Test of English for International Communication (TOEIC), International English Language Testing System (IELTS) and Test of English as a Foreign Language (TOEFL) datasets, our methodology uses Feedforward Neural Network (FNN) to forecast cognitive load based on student engagement behaviors and application errors then Bidirectional Encoders Representations from Transformer (BERT) processes content difficulty adjustments automatically. The proposed model delivers a 95.3% accuracy rate, 96.22% precision level, 96.1% recall capability and 97.2% F1-score which surpasses conventional Artificial Intelligence-based English as a Secondary Language (ESL) learning systems. The system makes use of Python for its implementation to improve understanding as well as student focus and mental processing speed. Personalized content presentation methods lead to lower cognitive strain which simultaneously advances student achievement numbers. The research adds value to smart educational frameworks through its introduction of a scalable framework that allows adaptable learning systems for English as a second language (ESL). The following research steps include simplifying system complexity while adding multimodal learning signals including eye monitoring and speech recognition and further developing the model across various educational subject areas. The research works as a promising foundation which propels AI real-time adaptive education systems for students from various backgrounds.**

Keywords*—Cognitive load management; artificial intelligence-based English as a secondary language learning; adaptive content personalization

I. Introduction

English as a Secondary Language (ESL) education serves an important purpose in developing the language proficiency of foreign speakers to communicate effectively with specific goals in pursuing academia, working life, or personal interest [1], [2]. Many ESL programs are traditional and use the static content delivery method based on a rule-following approach that does not consider the actual cognitive needs of individual learners[3]. Cognitive load is a key element that determines a student's success in learning ESL [4]. When cognitive load exceeds a learner's capacity, frustration, disengagement, and reduced comprehension can result. On the contrary, when managed optimally, learners are then able to pay attention to tasks of importance within the language while avoiding being occupied by its demands [5]. Here, the process of optimizing ESL learning through effective management of cognitive load is central to developing learning experiences that will be more efficient and effective [6]. Therefore, ESL sites should not use a normal approach but instead use adaptive and personalized systems that can determine and change learning material to fit the learner's cognitive ability, thereby enhancing understanding and remembering [7].

Several studies have investigated techniques to minimize cognitive load in ESL learning, but many approaches that have been developed have limitations. Rule-based simplification of content for traditional methods are useful in specific contexts but ignore the complexity of the language learning process and the various cognitive needs of different learners [8]. Moreover, static adaptive systems do not offer flexibility in changing the content dynamically according to the learner's behavior, progress, and changing cognitive load [9]. These systems have inherent difficulties in providing adaptive learning experiences where the experience is continually evolving with a changing learner [10]. Although these methods are adept at temporarily enhancing comprehension [11], they miss the sense of continuous, individualized nature of the learning experience. This study addresses these problems by using BERT as a more complex, data-centric approach to learning experience generation. Moreover, the study utilizes a FNN in order to predict cognitive load through the analysis of behavioral data such as task duration, error patterns, and engagement metrics.

The research has contributed by optimizing the cognitive load during ESL learning with the help of BERT and Feedforward Neural Networks. Bidirectional architecture by BERT aids in the increase of contextual understanding, hence it leads to proper representations of processes involving language such as reading comprehension, vocabulary acquisition, and sentence structure. Through application of BERT to analyze learner interaction data including quiz performance, time taken for the completion of the task, and engagement metrics, the framework analyzes cognitive load and modifies learning content based on such load. Simultaneously, the FNN analyzes behavioral data like duration and error patterns due to the multi-layered architecture that enables prediction of cognitive load. BERT and FNN thus modify content difficulty dynamically to align with learner capacity without either overloading or under loading. This is contrary to the conventional methods because the bidirectional understanding of the context of BERT and the predictive power of FNN makes for a more efficient system in processing and interpreting learner data. Combining BERT with auxiliary neural networks such as FNNs in this personalized, scalable, and adaptive ESL learning framework will ensure effective comprehension, retention, and reduced cognitive overload. The proposed system advances existing AI-based ESL learning models since it employs deep learning approaches to process real-time behavioral information. Real-time cognitive fluctuations become the centerpiece of personalized and scalable learning through the BERT model for contextual content adaptation and FNN model for cognitive load prediction. This surpasses previous ESL tutoring models because they lack real-time cognitive fluctuation analytics.

The key contribution of the research is as follows:

- Implemented a hybrid FNN-BERT framework that dynamically adjusts content complexity based on real-time cognitive load predictions, enhancing personalized learning experiences for ESL learners.

- Developed a cognitive load estimation model using FNN, analyzing behavioral metrics like task completion time, engagement levels, and error patterns for adaptive content delivery.

- Integrated BERT-based content personalization, enabling context-aware adjustments to learning materials based on learner comprehension, improving adaptability over rule-based and static models.

- Enhanced ESL learning through real-time cognitive load management, reducing cognitive overload while maintaining an optimal balance between content complexity and learner capacity.

- Validated the proposed framework using standard ESL datasets, demonstrating improved learning efficiency, comprehension, and engagement compared to traditional and AI-based adaptive learning models.

The remaining of the section is structured as follows: Section II delves into existing research on enhancing English Learning skills through mobile application interventions. Section III outlines the specific challenges addressed by the proposed framework. Section IV provides a detailed explanation of the components and methodology of the proposed framework. Following this, Section V presents the results obtained from implementing the framework and includes a comprehensive discussion of the findings. Finally, Section VI concludes the study.

II. Related Works

Feng [12], focuses on the application of AI-based language learning strategies, which emphasize personalized feedback, adaptive learning systems, and speech recognition technology with interactive exercises. The core innovation of this study is the combination of these strategies to optimize the process of language acquisition by reducing cognitive load. AI-supported methods are focused on delivering personalized learning with respect to different students, where content is drawn upon accordingly to personalize it and set it up to their proficiency. Overall, the students engaging in AI-assisted language learning showed considerably enhanced language skills, especially in cases of English as a Foreign Language (EFL) students. The readers showed improved cognitive loads since the items were placed in a way that suits the reader's actual understanding. However, the limitation of this study is that it is based on a single cohort of 484 EFL students, which might limit the generalization of the findings to other student populations or to language learners from various cultural or educational backgrounds. It also didn't consider the hypothetical technological challenges that may come into play in varying learning contexts, such as limits on resources or inequality in access to AI-driven tools.

Ding et al. [13], proposed the Gaze Reader method that uses a webcam and transformer-based machine learning models to detect unknown words in ESL learners. The innovation of this method is accessibility, as it does not require expensive and specialized eye-tracking devices. Instead of using an expensive high-end camera, the system utilizes a standard webcam to track the learners' gaze while detecting the attention towards unfamiliar or challenging words. Utilizing transformer-based models, the method allows for real-time feedback and enables learners to identify the unfamiliar vocabulary on which they should focus more. Results from the study shows that, the Gaze Reader method measured an impressive accuracy of 98.09% while recording an F1-score of 75.73%, showing its effectiveness towards ESL learners. This is particularly valuable for language learners because the system is able to identify unknown words as they appear in context, which tends to help language learners build their vocabulary in a way that's organic and contextual. A limitation of this study, however, is the sole use of one dataset from which this method will be applied, which might limit generalization of ESL learners' range. The applicability of the method can be verified and proved only after being used in various contexts, dialects, and language settings.

Vasu et al. [14], investigate how self-assessment and indirect teacher feedback promote the use of self-regulated learning (SRL) strategies for ESL students. The study's uniqueness is that it focuses on the practice of self-assessment as well as indirect teacher feedback to encourage more responsibility in the ESL learner's process. One of the most important parts of language learning is self-regulation, allowing students to self-monitor their performance, set personal goals for learning, and modify their learning strategies. The results revealed that students with self-assessment were able to develop their self-regulation abilities better and enhanced their language performance. Indirect teacher feedback was also observed to improve student motivation and overall performance as it gives students a chance to reflect on their own learning without explicit instructions from teachers. The strategies combined apparently, significantly contribute to SRL, but the study is limited due to its narrow scope that concentrates only on a particular group of students. This group may not represent the diversity of ESL learners across different educational contexts, cultures, and language backgrounds. In this regard, findings may not totally capture the effectiveness of self-regulation strategies for a more heterogeneous student population.

Brown et al. [15], explore the few-shot learning capabilities of GPT-3, which is a state-of-the-art autoregressive language model. The main innovation in GPT-3 is that it can accomplish a wide variety of NLP tasks without requiring any task-specific fine-tuning. Unlike other predecessors, GPT-3 can adapt to different NLP tasks by providing just a few minimal examples or prompts, and thereby it performs excellently across an extremely wide spectrum of applications that include translation, question answering, summarization, etc. The work demonstrated that Generative Pre-Trained Transformer (GPT)-3 performed competitively on several benchmark Natural Language Processing (NLPs) that are extensively used. This reduces the task-specific model training, which, in general, has to be undertaken for traditional NLP models. While it impressively performs its tasks, this study acknowledges a limitation in the capabilities of GPT-3 towards specialized or domain-specific tasks, where performance cannot reach the threshold of models undergoing task-specific fine-tuning. More critically, large size and high computation requirements may severely limit scalability and accessibility of the GPT-3 model, further restricting its adoption in resource-poor environments. Future work would focus on those problems and the enhanced performance of the model on particular tasks.

Yang et al. [16], introduces the novel autoregressive pretraining method known as Extra Long Network (XLNet). It extends from the constraints through enabling bidirectional context learning. What makes XLNet unique is that it bases training on permutation instead of random masking, and its model is trained to learn contexts from every possible direction rather than only through BERT's masked language modeling approach. XLNet learns from all permutations in the sequence for more robust and comprehensive contextual understanding. In the experiment, the authors had shown that XLNet performed better in all kinds of NLP tasks such as question answering, sentiment analysis, and text classification, with an average improvement across twenty different benchmarks. These improvements were said to be a result of the dependency and nuances that XLNet captures better than the other models. However, this study also presents a significant limitation of XLNet: its computational complexity. Permutation-based training is computationally intensive, meaning XLNet would require more computing power than most models designed for real-time or large-scale applications. This problem might limit XLNet's applicability in practical, resource-limited, or time-sensitive settings. Future work may focus on improving the efficiency of the model while preserving its enhanced contextual learning ability.

Šola, Qureshi, and Khawaja [17], discussed using AI-driven eye-tracking technology for the evaluation of cognitive load within online learning settings. This innovative approach introduced eye-tracking technology into the assessment using AI-powered prediction software, enabling real-time observations of the level of students' attention and concentration during a task. This is a novel research in the monitoring and analysis of cognitive load where students are focused on where to and for how long on certain parts of a learning material. Moreover, with integration into AI, this process improves prediction and interpretation of cognitive loads in the direction of enhancing the ability of instructors towards better understanding their students' mental states as they go through different tasks. The study reveals that eye-tracking systems powered by AI significantly enhance the learning experience. These systems help to identify the level of cognitive load and generate actionable data that improve instructional design. According to the findings, knowledge of cognitive load may help in optimizing the pace of content as well as methods of instruction. The main limitation of this study is its dependency on one type of eye-tracking software, which might not represent the complete gamut of cognitive load. Other technologies or methods might provide more subtle data, which may make this approach too simplistic for complex learning tasks. The findings may also not be generalizable because of the specific software used.

Sujatha and Rajasekaran [18], investigate the blended model to teach listening in language learning which is based on Cognitive Load Theory (CLT). The primary objective of the study is the improvement of processing efficiency of the auditory information by making use of the top-down approach, which can help students use contextual background knowledge to process the information. What the study does in fact is combine CLT with a structured approach where it focuses more on reducing extraneous cognitive load while promoting deep learning. The experimental results showed the comparison of listening comprehension and information prediction from the control group to the experimental group exposed to the blended model. Student improvement was, therefore, seen in the listening skills of the language learners. However, this study will be limited in that it has only a small sample size, so it is not generalizable toward other larger population ranges with various learning needs. The findings may not be generalizable to all learner groups, particularly in diverse educational settings or when the proficiency level is different. Future studies should include a more extensive and heterogeneous sample to establish the validity of the results across different contexts and understand the effectiveness of the model better.

The current research on adaptive ESL learning faces three main limitations because, the studies employ limited scalability and depend on small datasets as well as struggle to update learning in real time. Gaze Reader and self-regulated learning methods have improved learning engagement but they do not track dynamic cognitive changes. Researchers developed a hybrid BERT-FNN framework to provide both real-time personalization capabilities and increase scalability in the system.

III. Problem Statement

Existing method based on AI-based ESL learning methods face scalability and applicability issues. Most of the existing approaches focus on personalized feedback and adaptive learning systems however, their findings are usually limited only to small cohorts and fail to generalize [19]. Moreover, some of the techniques depend on a single dataset, while they don't take into consideration different dialects or language setups, and self-reporting [20] and implicit feedback-based strategies often do not consider the complexity of ESL. GPT-3 and XLNet have computationally intensive costs, thus unable to be applied in real time especially in resource-limited settings. Using BERT-the transformer-based model, with a Feedforward Neural Network, the proposed approach builds up scalar and adaptive scalability with respect to overcoming these kinds of limitations. BERT utilizes bidirectional learning. It enhances the ability of a learner to be more contextually aware in processing language. FNN then analyzes behavioral data, such as task duration, error patterns, engagement metrics, etc., in order to accurately predict cognitive load. This framework is dynamic in real-time, balancing the complexity of content with cognitive capacity, ensuring learners manage mental effort effectively while doing tasks. The use of these models ensures scalability, adaptability, and efficiency with a personalized, optimized learning experience for ESL learners.

IV. Proposed Hybrid FNN- BERT for Cognitive Load Management for Enhancing ESL Learners

The proposed framework starts with data collection that involves collecting learner interaction data in all its forms, including quiz performance, task completion time, and engagement metrics from the English Test Prep Data: Test of English for International Communication (TOEIC), International English Language Testing System (IELTS), Test of English as a Foreign Language (TOEFL) dataset. This data is cleaned, handled for missing values, and transformed into a numerical format for model compatibility in the Data Preprocessing block. The subsequent block is Cognitive Load Estimation, where FNN analyze behavioral data like time spent on tasks and error patterns to predict the learner's cognitive load. This estimation is used to determine instances where the learner is overloaded, which is critical for the framework's next step: content adaptation. The BERT Model Integration block is then followed, in which the pre-trained BERT model assesses the learner's understanding based on quiz responses. This adaptive system takes into account the learner's performance, and with an assessment of the comprehension gaps, it will either simplify or rephrase the content dynamically in line with the learner's cognitive capacity, avoid making the material either too complex or too simplistic.

The Adaptive Feedback Generation block takes over, generating real-time personalized feedback in relation to the learner's needs based on their cognitive load and comprehension analysis. This is intended to channel the learner into important learning objectives with deeper learning, aimed at filling specific gaps. Dynamic Content Delivery dynamically adjusts material complexity in a real-time and performance-based-cognitive load dependent manner. It will present relatively easier vocabulary words or examples, for example, if the learner makes mistakes, it will introduce difficult content once more for accurate comprehension results. The Evaluation and Outcome Measurement block entails measuring performance using various comprehension scores, assessments of cognitive loads, and indications of learner engagement for effective evaluation and outcomes. Subsequently, based on these indicators, the optimized system approach towards delivering content in such a way as to allow it to progressively get better regarding maximized outcomes from learning results is ensured. The whole system was implemented in Python and utilizes the deep learning libraries, TensorFlow, and Keras. This helps to train and deploy the model effectively. The last Optimization block allows the system to adjust based on the feedbacks it receives from learners and also from the learners' performance as they learn to deliver content that reduces cognitive loads with optimal performance as shown in Fig. 1.

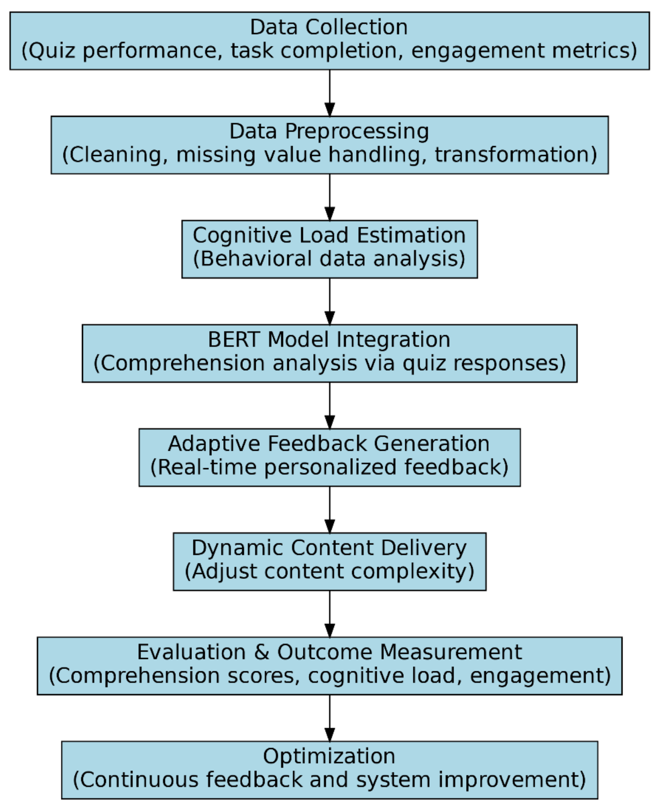

Fig. 1. Overall workflow.

A. Dataset Description

The proposed framework uses a dataset [21] of TOEIC, IELTS and TOEFL practice exams with detailed reports of learner test achievements and interactions. A variety of learner performance data points appear in the dataset to support cognitive load research and optimization efforts for ESL learners. The key attributes of the dataset are shown in Table I.

TABLE I. ATTRIBUTES OF THE DATASET

Attribute	Description
Learner ID	A unique identifier for each learner.
Quiz Performance	Data on the learner's performance in different quizzes, including correct answers, incorrect answers, and time spent on each quiz.
Task Completion Times	The time taken by the learner to complete various tasks or exercises within the platform.
Engagement Metrics	Metrics such as the frequency of interactions with the platform, time spent on learning materials, and response time during quizzes.
Content Interaction Data	Information about how learners interact with different types of content, such as vocabulary, grammar exercises, or reading comprehension passages.
Behavioral Data	Data on how learners respond to specific tasks, including error patterns, patterns of skipped questions, and engagement with different content types.

These data points will allow for a personalized learning experience by analyzing how individual learners engage with the platform and adjusting the content accordingly.

B. Data Preprocessing

The implementation of data preprocessing methods proves essential to ready learner interaction data for use in ML applications. The following reflects the essential procedures inside the data preprocessing framework as shown in Fig. 2.

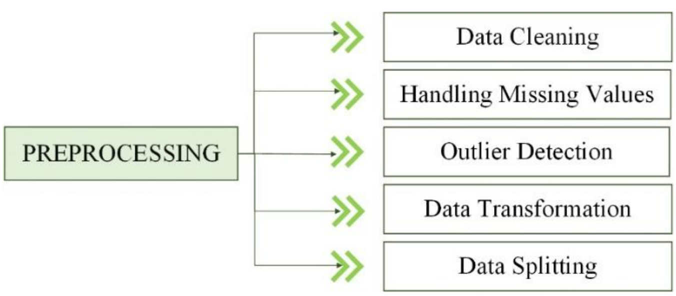

Fig. 2. Steps in data preprocessing.

1) Data cleaning

2) Handling missing values: It arises from incomplete learner interaction or system malfunction, through methods such as imputation or dropping rows/columns which have high significant missing values.

3) Outlier detection: Task completion time outliers and quiz performance data outliers are detected and then corrected. When the outliers were extreme, there was removal of such outliers for ensuring that there is no decrease in model's performance.

4) Data transformation

a) Encoding categorical data: One-hot encoding and label encoding is used to change categorical data into numerical format-for example, one-hot encoding on learner IDs, and the same for the content types.

b) Feature scaling: Continuous data such as the time taken to complete tasks or engagement metrics are normalized or standardized so that the features are in the same scale and do not influence the model disproportionately.

c) Processing of time series data: To capture time dependency between interactions in time spent on tasks, the preprocessing is applied as sequence-based.

5) Data transformation for model input: After the data cleaning process, organize the data in a machine learning format for the BERT model. This typically includes learning interaction sequences along with corresponding performance measures, ensuring that all input points contain relevant features such as quiz scores, task time, and cognitive load.

6) Data splitting: Divide the pre-processed data into the train, validation, and test sets to measure the performance of the model and overall generalization. Training dataset is provided for training the model; on the other hand, validation and test datasets are kept for performance estimation of the model and its generalization.

These preprocessing steps ensure that the dataset is clean, well-structured, and ready for use in the proposed machine learning model, which will drive the dynamic content adjustment and cognitive load optimization for ESL learners.

C. Task Completion and Engagement Using Feedforward Neural Network

First, gather behavioral data about learners. Such metrics are: time to complete tasks, patterns of errors, and completion or engagement with the tasks. These give an indication when a learner is undergoing high cognitive load and tell one how to adapt the learning content. Greater duration might be symptomatic of it being a cognitively taxing activity or for that matter difficult or the content the learner might not be absorbing very well leading to cognitive overload. The task time as an analytical function in terms of learner's characteristics and task difficulties is given in Eq. (1):

$$T_{task} = f(Learner\ Difficulty, Task\ Complexity) \quad (1)$$

where, T_{task} is the time spent completing a task; Learner Difficulty can be inferred from past performance; Task Complexity can be quantified through the task's intrinsic difficulty. The signal to adjust the content may be provided through a higher T_{task}. For example, if {task} surpasses the {threshold}, then the system will sometimes intervene using hints or task simplifications as in Eq. (2):

$$Intervention = \begin{matrix} Hints/Support & if\ T_{task} > T_{threshold} \\ No\ Action & otherwise \end{matrix} \quad (2)$$

Frequent errors might indicate that the learner has not mastered the content properly, thus experiencing high cognitive load. Let {errors} be the number of errors committed by a learner while performing a certain task as defined in Eq. (3):

$$E_{errors} = \sum_{i=1}^{n} Error_i \quad (3)$$

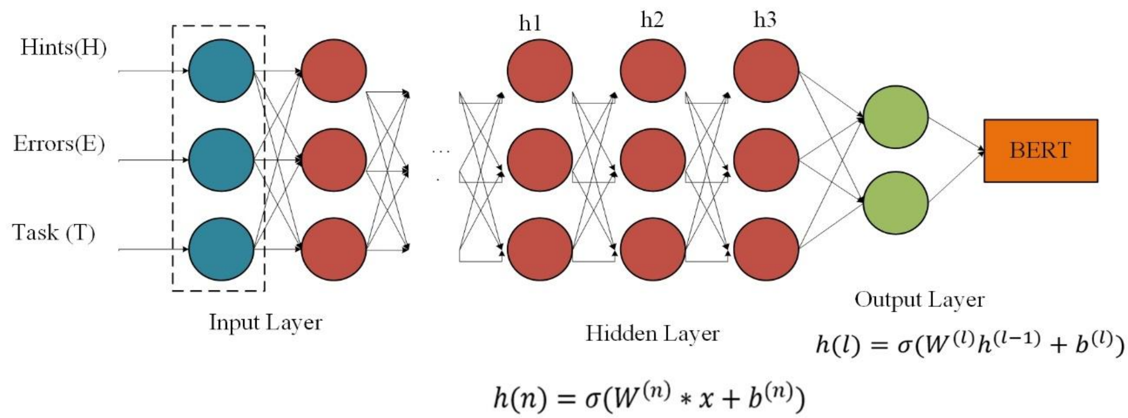

Fig. 3. FNN Architecture.

where, n is the number of steps or sub-tasks involved in a task, and $Error_i$ is the binary indicator set at 1 in case of a learner's mistake and 0 otherwise. Errors repeated multiple times indicate cognitive overload by the learner, prompting a response from the system in simplifying instructions or other support.

Task Completion and Engagement: Students who abandon tasks or spend a long time to complete high cognitive loads. Let completion be a dummy variable indicating if the task has been completed; it is set to 1 if the task was completed and to 0 otherwise. The engagement measure engage could be defined as the time used on the task divided by expected time to finish as in Eq. (4),

$$E_{engage} = \frac{T_{task}}{T_{expected}} \tag{4}$$

where, $T_{expected}$ is the time that a learner should ideally take to complete a task. If E_{engage} is too low or $T_{expected}$ = 0, it suggests the learner is disengaged, and the system should intervene by providing support or simplifying content. Once the behavioral data is collected, we use FNNs to predict cognitive load. FNNs are very appropriate for this purpose because they can learn non-linear relationships between input features such as time on task, errors, and engagement, and the output variable, which is cognitive load.

Input Features: The FNN will take various behavioral metrics as input as in Eq. (5), including:

- Time on Task T_{task}
- Number of Errors {errors}
- Number of Hints Requested {hints}

These features are transformed into numerical vectors, which are input into the FNN for cognitive load prediction. Let the input vector be denoted as {x}:

$$x = [T_{task}, E_{errors}, H_{hints}] \quad 5)$$

Feedforward Neural Network: The FNN consists of multiple layers, with each layer performing linear transformations followed by a non-linear activation function σ as in Eq. (6):

$$h(l) = \sigma(W^{(l)}h^{(l-1)} + b^{(l)}) \tag{6}$$

where, $h(l)$ is the output of the l th hidden layer, $h^{(l-1)}$ is the weight matrix for the l th layer; $b^{(l)}$ is the bias term; σ is a non-linear activation function, typically ReLU or Sigmoid as shown in the Fig. 3. The final layer produces a cognitive load score {load}, which predicts whether a learner is experiencing a low, a medium, or high cognitive load. Where L is the total number of layers. The predicted level of cognitive load {load} can then be forced into discrete categories, for example, low, medium, and high. To train the FNN, we leverage historical data collected from previous learners. The loss that guides the training aims to minimize the difference between what the model is predicting for the cognitive load and the true values labeled {true} for the labels. The above model is optimized using a mean squared error as in Eq. (7):

$$L = \frac{1}{N}\sum_{I=1}^{N}(y^\wedge load(i) - ytrue(i))^2 \tag{7}$$

where, N is the number of training samples. After training, the FNN will predict cognitive load in real time as learners are interacting with the system, hence guiding adjustments to task complexity and content delivery. Combining all these techniques will ensure that the system continuously analyses learner behavior and predicts cognitive load to personalize learning. The model adapts in real time to adjust the difficulty of tasks based on the predicted cognitive load, such that learners neither feel overwhelmed nor under-challenged.

D. Personalization Using BERT Model in the Proposed Frameworks

The core idea of personalization in this framework revolves to fine-tune the pre-trained model of BERT on dynamic adjustments, according to changing levels of a learner's cognitive load, performance, and task type. Ensuring BERT Fine-Tuning adapts it to the specifics of ESL learning, so as to process different learner inputs with the model able to provide recommendations accordingly. The input features relevant for BERT include learner performance (such as quiz scores and task completion), type of task (easy vs. difficult), and the cognitive load prediction by the auxiliary neural network. Other behavioral measures such as time on task, engagement, and errors can be encoded as in Eq. (8):

$$X_{input} = [P_{learner}, T_{task}, \hat{y}_{load}, E_{engage}, E_{errors}] \tag{8}$$

where, $P_{learner}$ represents the learner's performance data; T_{task} represents the difficulty level of the task; $\hat{y}_{load}$ is the predicted cognitive load from the auxiliary neural network; E_{engage} represents the engagement level (calculated as discussed earlier); E_{errors} captures the number of errors made during the task. BERT inputs are then transformed through the multi-layer attention mechanisms of BERT into context-aware representations of the learner's current state, upon which personalized recommendations are generated or predictions of what is next in terms of action, for example, what content best serves the learner or if reinforcement is needed in weaker areas or new challenging material is best presented as in Fig. 4.

Task Complexity Adjustment Based on Cognitive Load BERT has processed the learner's input and made its predictions, the system uses its cognitive load predictions to dynamically adapt the task difficulty {load} predicted cognitive load is low, it presents more difficult content to engage the learner. Let's call the action it takes when its cognitive load is low as in Eq. (9):

$$Action_{low\_load} = \begin{cases} Increase\ Difficulty & if\ \hat{y}_{load} < L_{threshold} \\ No\ Change & otherwise \end{cases} \quad (9)$$

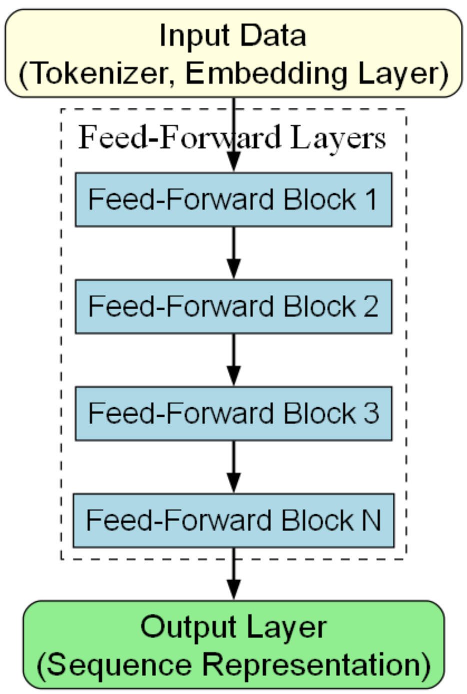

Fig. 4. BERT Architecture.

where, {threshold} is a pre-defined threshold below which the learner is considered to have low cognitive load. For example, if a learner successfully completes multiple tasks with a low cognitive load, the system might increase the complexity of subsequent tasks or introduce new challenges, such as advanced exercises or new content that builds on previously learned concepts. This ensures that the learner is constantly engaged and not under-challenged, which helps to maintain motivation.

Conversely, if cognitive {load} is high, the system reduces task complexity or offers support to prevent learner frustration and cognitive overload as in Eq. (10):

$$Action_{high\_load} = \begin{cases} Decrease\ Difficulty & if\ \hat{y}_{load} < H_{threshold} \\ Provide\ Support\ (hints) & if\ \hat{y}_{load} < H_{threshold} \end{cases} \quad (10)$$

where, $H_{threshold}$ is a pre-defined threshold above which the learner is considered to be in a high cognitive load state. In this case, the system would make complex tasks ahead less effective, give hints, provide simpler exercises, or simplify the overall task by breaking it into simpler smaller-sized sub-tasks. This scaffolding approach ensures that a learner is not overwhelmed and can move on to mastering major concepts in an acceptable manner.

Dynamic Content Delivery Based on Real-Time Cognitive Load: The dynamic content delivery mechanism is at the heart of the proposed framework, which dynamically adjusts the learning path based on real-time predictions of cognitive load. After every task or interaction, the system evaluates the learner's cognitive load using the auxiliary neural network. To finally predict the corresponding cognitive load given the learner performance data P and engagement metrics feeds into the learned BERT which processes this in order to dynamically update the LC and task difficulty aa in Eq. (11):

$$X_{input}^{new} = [P_{learner}, T_{task}, \hat{y}_{load}{}_{load}^{new}, E_{engage}^{new}, E_{errors}^{new}] \quad (11)$$

Based on this revised input, BERT can come up with another set of tasks or suggestions. If a learner is unable to get a right answer several times {errors}, the system might provide them with easier forms of the same content or supplement (such as hints or examples). If, on the other hand, a learner is successful, BERT might challenge a learner to higher-order content by gradually making tasks harder or by giving a learner some new challenges, depending on a learner's background.

For instance, if a learner has successfully completed a set of tasks with a low cognitive load and the system predicts that they are capable of handling more complex content, the system might offer a more challenging exercise as in Eq. (12):

$$Next\ Task = f(\hat{y}_{load}, Engagement) = \begin{cases} Advanced\ Content & if\ \hat{y}_{load} < L_{threshold} \\ Simplified\ Content & if\ \hat{y}_{load}] > H_{threshold} \end{cases} \quad (12)$$

This ensures continuous challenge in the correct degree, neither overburdened nor under-stimulated for maximum engagement and learning. A personalized learning framework builds individual optimal learning trajectories for learners. The system controls assignment difficulty according to accurate ongoing mental workload predictions which protects learners from information overload while sustaining their peak ability level. Using this method produces maximum student involvement and motivation together with overload prevention. Through a continuous learning performance and cognitive load prediction cycle the system maintains active adaptation.

V. Result and Discussion

The results section of this study evaluates the effectiveness of the proposed deep learning-based framework for optimizing cognitive load in ESL learning environments by implementing it on a python software tool. The performance of the framework is assessed through a variety of metrics, such as learner comprehension scores, cognitive load assessments, and engagement indicators. These metrics demonstrate the ability of the system to dynamically adapt to the needs of individual learners, thereby improving their learning experience. The results are compared with traditional, static content delivery methods to determine the potential of the framework in enhancing learning efficiency, learner engagement, and overall comprehension. The following subsections provide a detailed analysis of the results obtained from the implementation of the Feedforward Neural Network (FNN) and Transformer-based BERT model, along with a discussion of the implications for future ESL learning platforms.

A. *Analysis of the English Test Prep Dataset*

The result section displays the comprehensive analysis of the data set by visualizing the key trends and distributions across competency categories and test levels through graphical representation. The findings are highlighted based on these visual representations, indicating the prevalence of certain language skills such as listening, speaking, reading, and writing skills across various testing frameworks such as Test of English for International Communication (TOEIC), International English Language Testing System (IELTS), and Test of English as a Foreign Language (TOEFL). Furthermore, an examination of how the different competency categories and corresponding test difficulties were aligned may inform about the process of design and organization that occurred. In doing so, it becomes not only the constitution of this dataset but may even shed more light on shortcomings or potential points for improving on methods used within the process of assessing languages. The next parts shows what such figures might elucidate.

1) Language competency distribution analysis: Fig. 5 shows the percentage distribution of language competency categories: Listening, Speaking, Reading, and Writing, based on the attributes of the dataset. The graph shows that the "Other" category is the most dominant, making up 51% of the total competencies. This probably includes tests that cover more than one competency or are not classified. There is a parity observed between Listening and Speaking, each at 24% of the dataset. Both are crucially necessary in language understanding and communication. These skills are given balanced attention. Reading and Writing are put separately in a category labeled "Other," which indicates that they appear to constitute a reduced portion or less of the listed tests in the dataset. This visualization highlights oral skills (Listening and Speaking) as being underlined in assessments of language competence, reflecting a strong presence of these skills in real-world language usage. It further underlines potential underrepresentation in Writing and Reading as separate competencies, which therefore deserve further elaboration. The chart provides an overview of the organization of the data set, depicting the major focus areas in language testing, along with relative proportions.

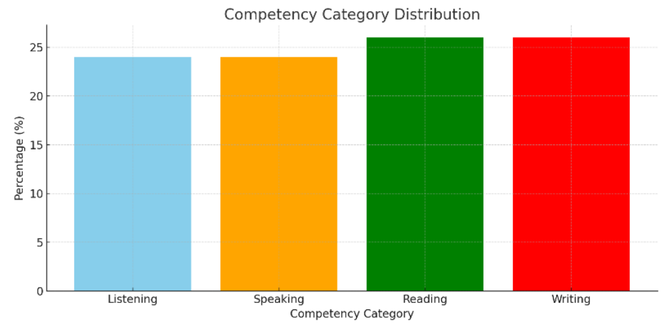

Fig. 5. Competency category distribution.

2) Comparison of test designs at different levels of competency: Fig. 6 illustrates the distribution of tests in terms of proficiency levels that range from A1 to B2 on the Common European Framework of Reference for Languages (CEFR) and standardized exams like TOEIC, IELTS, and TOEFL. The chart shows that assessments based on grammar are prominent in A1 to B2 levels as it is a starting point for learning language. For Listening and Reading competencies, TOEIC and IELTS tests are spread across well-defined score ranges, such as 110 to 495 for TOEIC and 4.0 to 9.0 for IELTS, offering clear gradations of proficiency. Speaking and Writing tests follow similar trends but feature fewer levels, reflecting their emphasis on qualitative assessment. TOEFL tests, on the other hand, have fewer but highly focused levels, with Listening ranging from scores of 9 to 30. This distribution shows the diversity of testing frameworks and their different focus areas. It also shows the ability of the dataset to meet the needs of learners with different levels of proficiency and test requirements, providing insight into the balance of grammar, listening, speaking, reading, and writing tests across different proficiency frameworks.

B. *Evaluation of Cognitive Load Prediction*

Behavioral data is fed to the FNN, which predicts the cognitive load. The FNN processes time on task, error patterns, and engagement metrics and generates a score for cognitive load. This score is used to classify the cognitive state of the learner as either high or low.

This bar chart in the Fig. 7 represents the cognitive load scores calculated for ten different tasks in terms of time spent, error patterns, and engagement levels. Each task is plotted along the x-axis, and the score on the y-axis represents the cognitive load. A red dashed line is drawn at a score of twenty to signify crossing over from high to low cognitive loads. All tasks in this dataset score below this limit, which classifies them as "Low Cognitive Load". Scores are highly variable between tasks, with a peak of around 13.84 for Task 9, and troughs around 7.06 for Task 7. Score variation describes these differences in the difficulty and user performance across the tasks under study. The color gradient of the bars, being based on the "viridis" palette, emphasizes these differences visually. This Fig. 7 is all-inclusive and gives an overview of the cognitive demands in respect to the task, thereby providing a comparison tool and indicating which areas could possibly be improved on for performance. The chart further helps visually distinguish outliers or anomalous cases within cognitive performance by representing scores graphically. Generally, low scores across all tasks indicate that these are manageable cognitive demands, thereby fitting the target group or setting for the activity. However, these results might further be influenced by other factors, such as user fatigue or task sequencing.

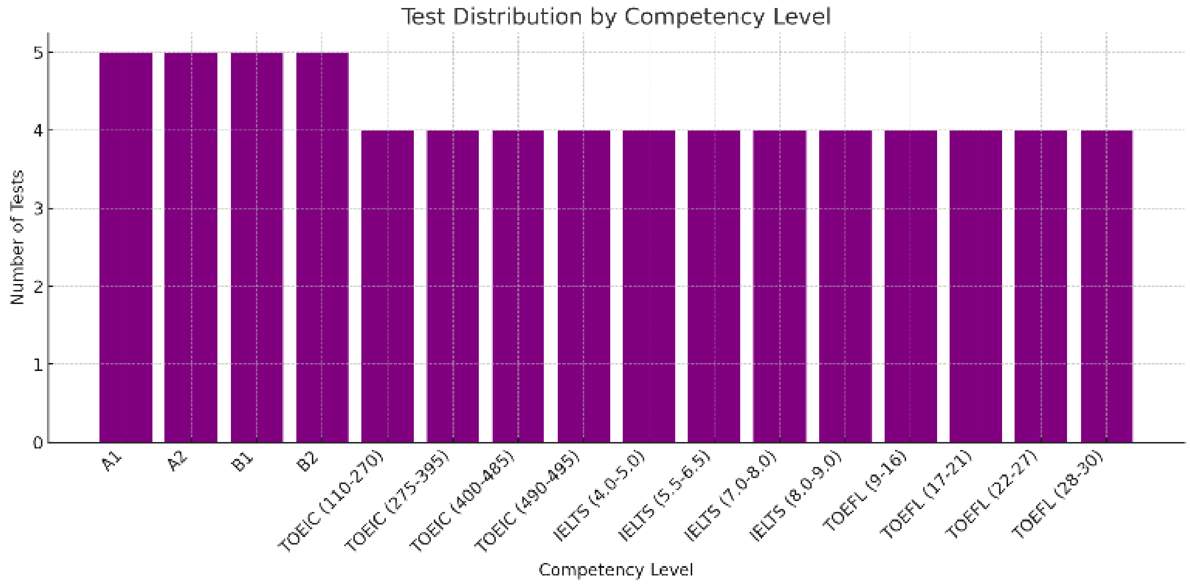

Fig. 6. Distribution of tests across levels of competence.

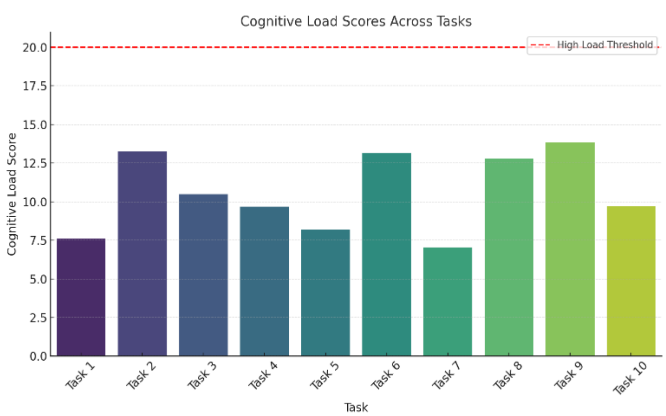

Fig. 7. Cognitive load scores across tasks.

The Table II below illustrates how task complexity should be modified for different language proficiency tests depending on the cognitive load of various skill levels. Each competency listed in the table is associated with a specific skill, such as listening or reading, and the corresponding cognitive load level: Low, Moderate, or High. For tasks with a Low Cognitive Load, the recommendation is to increase the task complexity. This could be the presentation of more complex tasks or increased speed to push the learner even further to his or her full potential. For instance, in the "Listening Test in TOEIC for Level 110 to 270," task complexity is introduced by the use of more challenging listening tasks or content with faster speed. Similarly, in the "Reading Test in TOEIC for Level 115 to 270," the increase in task complexity is through the introduction of more complex texts or comprehension questions.

For Moderate Cognitive Load, the task complexity is adjusted in order to keep the challenge balanced. Rather than increasing the difficulty of the task, the solution would instead be to provide the learner with more practice materials or exercises with the same level of difficulty. For example, the "Listening Test in TOEIC for Level 400 to 485" adjusts task complexity through providing extra practice content that matches the difficulty level against which the learner currently operates. Similarly, in the "Reading Test in TOEIC for Level 385 to 450," complexity is maintained by introducing additional exercises of similar difficulty.

For tasks characterized by High Cognitive Load, the system simplifies tasks to help the learner from getting overwhelmed by the task itself. Simplification may include: breaking down big tasks into tiny components, hinting, adjusting the structure of the task itself to reduce its mental effort to be executed by the learner. For instance, in "Listening Test in TOEIC for Level 490 to 495, it is advisable to simplify tasks by giving out hints or making tasks smaller or more divided portions. For instance, for "Reading Test in TOEIC for Level 385 to 450", a task reduction approach guarantees not to let the learning student overwhelmed due to complexity about the subject itself.

C. Analysis of Content Personalization with BERT

The fine-tuned BERT model takes in the prediction of cognitive load of the learner, the task performance, and engagement data and then uses that information to customize the delivery of content. Thus, based on the predicted cognitive capacity, the system provides appropriate content. Using real-time cognitive load prediction, the BERT model adjusts content complexity in a dynamic way, thereby making the challenge for the learner appropriate at each step.

TABLE II. TASK COMPLEXITY ADJUSTMENT FOR LANGUAGE SKILLS BASED ON COGNITIVE LOAD

Competency Name	Skill	Cognitive Load	Task Complexity Adjustment
Listening Test in TOEIC for Level 110-270	Listening	Low	Increase task complexity (e.g., add more difficult listening tasks or faster-paced content).
Listening Test in TOEIC for Level 275-395	Listening	Low	Increase task complexity (e.g., add more difficult listening tasks or faster-paced content).
Listening Test in TOEIC for Level 400-485	Listening	Moderate	Adjust task complexity (e.g., provide additional practice content of similar difficulty).
Listening Test in TOEIC for Level 490-495	Listening	High	Simplify tasks, provide hints, or break tasks into smaller components to reduce mental effort.
Reading Test in TOEIC for Level 115-270	Reading	Low	Increase task complexity (e.g., introduce more complex texts or comprehension questions).
Reading Test in TOEIC for Level 275-380	Reading	Low	Increase task complexity (e.g., introduce more complex texts or comprehension questions).
Reading Test in TOEIC for Level 385-450	Reading	Moderate	Adjust task complexity (e.g., provide additional exercises with the same complexity).

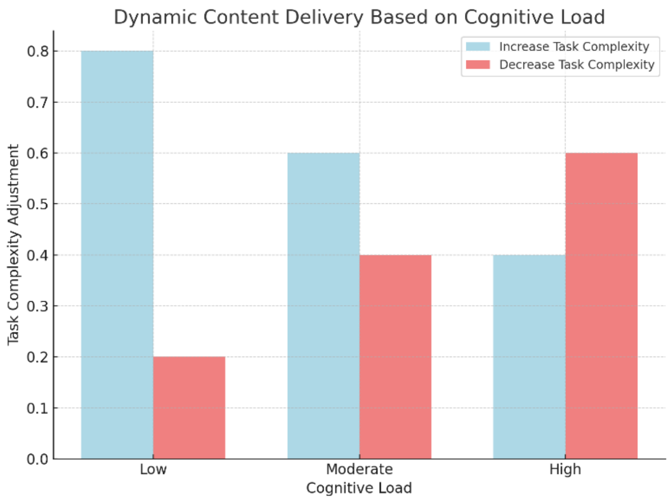

Fig. 8. Dynamic content delivery based on cognitive load performance evaluation.

This Fig. 8 looks at the way dynamic adjustments to task complexity depend on changing levels of cognitive load. The x-axis classifies the cognitive load as Low, Moderate, and High, while the y-axis expresses the changes in task complexity, in both increments and decrements. There are two sets of bars for each level of cognitive load. The light blue represents the increase in task complexity, and the light coral represents the corresponding decrease.

For "Low Cognitive Load," task complexity increases with a large effect size (0.8), while decreasing minimally (0.2). This means that when the user's cognitive demand is low, he can withstand a huge rise in content complexity without a bad effect. The increase in task complexity drops to 0.6 and the decrease to 0.4 when the cognitive load moves to "Moderate." It means that, when the cognitive load becomes moderate, there will be a better balance in presenting the content. For "High Cognitive Load," the trend reverses, with a small increase in complexity (0.4) and a substantial decrease (0.6). This reflects the need to reduce task difficulty significantly to accommodate users experiencing high cognitive demands.

The chart makes the principle of adaptive content delivery visually clear. It does neither overwhelm nor underchallenge the users. The adaptive model of varying task complexity and cognitive load would ensure optimal learning and performance. The width of the bars along with the distinction in color enables better readability, and the overlapping positioning of the bars for each of the cognitive loads allows for immediate comparison. Overall, this Fig. 8 provides for an intuitive representation of how content complexity adjustments align with the user cognitive states, and it forms a valuable tool for educators, designers, and researchers seeking to optimize task performance and engagement.

D. *Performance Metrices*

1) Accuracy: Accuracy gives the ratio of the correctly classified instances to the total instances. Here from, the proposed framework achieved a collective training accuracy. Accuracy is computed by the following Eq. (11).

$$Accuracy = \frac{PN + PP}{IP + PN + IN} \tag{11}$$

2) Precision: It measures the ratio of correctly identified positive cases by the model out of all the cases which the model predicted to be positive. Indeed, the proposed framework achieved impressive precision in the accuracy across various segments including; High spenders and young professionals. Precision is calculated by the help of the Eq. (12).

$$Precision = TP/TP + FP \tag{12}$$

This shows that in Practice segments, the model is able to minimize these false positives, and correctly identify the positive cases to ensure that most cases that are classified as positive are indeed positive.

3) Recall: Recall measures the ratio of true positive instances with reference to the total actual positive instances. This is a testament of this proposed frameworks good recall which would imply its ability to recollect or recognize most of the 'real' outputs such as the Low Spenders and the Value Seekers. The F1-score for each gene set is computed on the basis of the following Eq. (13).

$$Recall = TP/TP + FN \tag{13}$$

This high recall ensures that the true positives were identified by the model without omitting many of them, as it established an all-round understanding of each customer segment.

4) F1 Score: The F1 score is defined as the harmonic mean of precision and recall therefore is balanced between the two measures. The proposed framework closely attained forefront F1 vector, confirming its good precision-recall balance for different sorts of customer. The F1-score is given by Eq. (14).

$$F1\,Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{14}$$

This metric therefore validates the effectiveness of the framework to classify the different customers as described earlier of achieving a trade-off between false positive and false negative detection.

TABLE III. PERFORMANCE METRICS OF FNN-BERT MODEL

Metrics	Values (%)
Accuracy	95.3
Precision	96.22
Recall	96.1
F1 score	97.2

Table III presents the performance metrics of the proposed FNN-BERT model, evaluating its effectiveness in cognitive load-based ESL learning. The model achieves 95.3% accuracy, ensuring reliable classification. It records 96.22% precision, minimizing false positives, while 96.1% recall indicates strong sensitivity to relevant cases. The 97.2 F1-score confirms a balanced precision-recall tradeoff, highlighting its robust performance.

TABLE IV. PERFORMANCE COMPARISON OF OF FNN-BERT MODEL WITH EXISTING MODEL

Methods	Accuracy	Precision	Recall	F1 score
PT-GRU **[22]**	78.85	75.90	77.33	76.71
SVC (R) **[23]**	94.8	92.56	95.87	96.3
Logistic Regression**[24]**	89	88	90	93
FNN-BERT (proposed)	95.3	96.22	96.1	97.2

Table IV compares the proposed FNN-BERT model with PT-GRU and SVC (R). FNN-BERT surpasses PT-GRU (78.85% accuracy) and SVC (R) (94.8% accuracy), achieving 95.3% accuracy. It also leads in precision (96.22%), recall (96.1%), and F1-score (97.2%), demonstrating superior effectiveness in cognitive load-based ESL learning.

Fig. 9 illustrates the four models—PT-GRU, SVC (R), Logistic Regression, and the proposed FNN-BERT—are compared on the basis of four significant performance indicators: F1 Score, Accuracy, Precision, and Recall. In online ESL learning systems, the FNN-BERT model consistently outperforms the others in all categories, illustrating its remarkable ability for adaptive content personalization. Logistic Regression shows decent efficiency, though SVC (R) is competitive, particularly in Recall and F1 Score. With the poorest performance on every criterion, PT-GRU shows how optimally the hybrid FNN-BERT approach maximizes cognitive load.

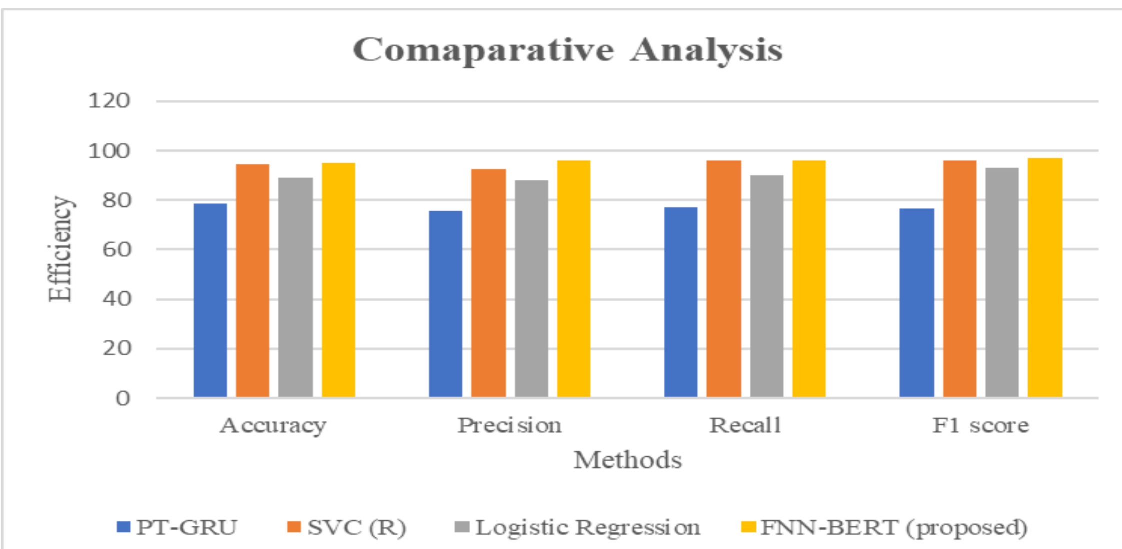

Fig. 9. ESL Model performance comparison.

E. Discussion

The FNN-BERT framework achieves effective content adaptation by demonstrating superior performance with 95.3% accuracy and precision of 96.22% and recall of 96.1% and an F1-score of 97.2%. This system serves digital ESL learning platforms where it uses cognitive load measurements to adjust content difficulty levels for each learner. The system can apply to language tutoring platforms and e-learning tools and educational AI assistants to enhance both student understanding and involvement. The educational benefits provided by this technology include live adjustments, ability to scale and better learning effectiveness through personalized content distribution that reduces mental stress without losing student focus. FNN and BERT together boost behavioral data analysis and contextual understanding thus delivering superior outcomes than rule-based static AI models. High computational needs stand as a major disadvantage for deployment since low-resource environments struggle with these requirements. The optimal fine-tuning process requires numerous labeled datasets which represent an obstacle. Future advancements in this model should prioritize minimalizing its complexity while adding various learning indicators including eye tracking alongside speech analysis and expanding its useable applications to benefit educational processes beyond ESL education.

With 95.3% accuracy and an F1-score of 97.2%, the proposed FNN-BERT model works well; however, these results are based on a specific dataset and controlled conditions. Verifying the effectiveness of the model across various student populations, language proficiency levels, and online learning environments is essential to ensuring its strength, usability, and applicability. To truly assess the model's scalability and flexibility, future studies will focus on applying the evaluation to larger and more varied datasets and real-world ESL learning contexts.

The extensive computational requirements and availability of AI models pose serious challenges, particularly in contexts with limited resources such as schools. In fact, these factors can render it even more challenging for AI systems to be utilized broadly in some contexts. It will be important to investigate further alternate remedies, such as the design of lighter weight, more efficient models and strategies for optimizing computational processes, to solve this issue. In an effort to make the proposed systems more easily deployable within resource-limited environments and facilitate greater practical application and use in schools, there will need to be an exploration of methods like model compression, quantization, and other resource-conserving tactics.

VI. Conclusion and Future Works

The proposed FNN-BERT framework successfully improves ESL learning by modeling content adjustment according to cognitive load which demonstrates 95.3% accuracy and 96.22% precision and 96.1% recall as well as 97.2% F1-score. Through the integration of FNN for behavioral analysis with BERT for contextual adaptation the model delivers superior results to existing methods which guarantees both personal learning experiences and higher student engagement. The presented research introduces advancements to adaptive learning systems powered by AI which both enhance student understanding and diminish cognitive stress factors. The system requires additional attention to meet two main barriers including heavy computational needs with substantial data labeling requirements. The future research direction emphasizes model speed improvement and combines eye-tracking and speech analysis data alongside the development of new applications between the proposed framework and STEM education and vocational training fields. This research will test the deployment of this system within digital education platforms to determine practical implementations that promote widespread accessibility and effect within intelligent educational systems.

References

[1] M. Nivedita, “The Role of Effective Oral Communication Skills in Globalized English Education _,” 2023, Accessed: Jan. 24, 2025. [Online]. Available: https://www.rjoe.org.in/Files/v8i4/30.Dr.NIVEDITA(259-269).pdf

[2] Z. N. Ghafar, “English for specific purposes in English language teaching: design, development, and environment-related challenges: an overview,” Canadian Journal of Language and Literature Studies, vol. 2, no. 6, pp. 32–42, 2022.

[3] M. Ramzan, R. Bibi, and N. Khunsa, “Unraveling the Link between Social Media Usage and Academic Achievement among ESL Learners: A Quantitative Analysis,” Global. Educational Studies Review, VIII, pp. 407–421, 2023.

[4] P. Evans, M. Vansteenkiste, P. Parker, A. Kingsford-Smith, and S. Zhou, “Cognitive Load Theory and Its Relationships with Motivation: a Self-Determination Theory Perspective,” Educ Psychol Rev, vol. 36, no. 1, p. 7, Mar. 2024, doi: 10.1007/s10648-023-09841-2.

[5] J. Sweller, “The Development of Cognitive Load Theory: Replication Crises and Incorporation of Other Theories Can Lead to Theory Expansion,” Educ Psychol Rev, vol. 35, no. 4, p. 95, Dec. 2023, doi: 10.1007/s10648-023-09817-2.

[6] D. Apostolou and G. Linardatos, “Cognitive load approach to digital comics creation: A student-centered learning case,” Applied Sciences, vol. 13, no. 13, p. 7896, 2023.

[7] B. Maraza-Quispe, O. M. Alejandro-Oviedo, W. C. Fernandez-Gambarini, L. E. Cuadros-Paz, W. Choquehuanca-Quispe, and E. Rodriguez-Zayra, “Analysis of the cognitive load produced by the use of subtitles in multimedia educational material and its relationship with learning,” International Journal of Information and Education Technology, vol. 12, no. 8, pp. 732–740, 2022.

[8] M. Suryani et al., “Role, Methodology, and Measurement of Cognitive Load in Computer Science and Information Systems Research,” IEEE Access, 2024.

[9] A. Bahari, “Challenges and Affordances of Cognitive Load Management in Technology-Assisted Language Learning: A Systematic Review,” International Journal of Human–Computer Interaction, vol. 39, no. 1, pp. 85–100, Jan. 2023, doi: 10.1080/10447318.2021.2019957.

[10] A. Ezzaim, A. Dahbi, A. Aqqal, and A. Haidine, “AI-based learning style detection in adaptive learning systems: a systematic literature review,” J. Comput. Educ., Jun. 2024, doi: 10.1007/s40692-024-00328-9.

[11] J. Zare and K. Aqajani Delavar, “A data-driven learning focus on form approach to academic English lecture comprehension,” Applied Linguistics, vol. 44, no. 3, pp. 485–504, 2023.

[12] L. Feng, “Investigating the Effects of Artificial Intelligence-Assisted Language Learning Strategies on Cognitive Load and Learning Outcomes: A Comparative Study,” Journal of Educational Computing Research, vol. 62, no. 8, pp. 1961–1994, Jan. 2025, doi: 10.1177/07356331241268349.

[13] J. Ding, B. Zhao, Y. Huang, Y. Wang, and Y. Shi, “GazeReader: Detecting Unknown Word Using Webcam for English as a Second Language (ESL) Learners,” Mar. 18, 2023, arXiv: arXiv:2303.10443. doi: 10.48550/arXiv.2303.10443.

[14] K. A. Vasu, Y. Mei Fung, V. Nimehchisalem, and S. Md Rashid, “Self-Regulated Learning Development in Undergraduate ESL Writing Classrooms: Teacher Feedback Versus Self-Assessment,” RELC Journal, vol. 53, no. 3, pp. 612–626, Dec. 2022, doi: 10.1177/0033688220957782.

[15] T. B. Brown et al., “Language Models are Few-Shot Learners,” arXiv.org. Accessed: Jan. 24, 2025. [Online]. Available: https://arxiv.org/abs/2005.14165v4

[16] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, “XLNet: Generalized Autoregressive Pretraining for Language Understanding,” Jan. 02, 2020, arXiv: arXiv:1906.08237. doi: 10.48550/arXiv.1906.08237.

[17] H. M. Šola, F. H. Qureshi, and S. Khawaja, “AI Eye-Tracking Technology: A New Era in Managing Cognitive Loads for Online Learners,” Education Sciences, vol. 14, no. 9, Art. no. 9, Sep. 2024, doi: 10.3390/educsci14090933.

[18] U. Sujatha and V. Rajasekaran, “Optimising listening skills: Analysing the effectiveness of a blended model with a top-down approach through cognitive load theory,” MethodsX, vol. 12, p. 102630, Jun. 2024, doi: 10.1016/j.mex.2024.102630.

[19] J. C. Lawrance, P. Sambath, C. Shiny, M. Vazhangal, S. Prema, and B. K. Bala, “Developing an AI-Assisted Multilingual Adaptive Learning System for Personalized English Language Teaching,” in 2024 10th

International Conference on Advanced Computing and Communication Systems (ICACCS), IEEE, 2024, pp. 428–434. Accessed: Jan. 24, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10716887/

[20] Y. Xia, S.-Y. Shin, and J.-C. Kim, "Cross-cultural intelligent language learning system (cils): Leveraging ai to facilitate language learning strategies in cross-cultural communication," Applied Sciences, vol. 14, no. 13, p. 5651, 2024.

[21] "English Test Prep Data: TOEIC, IELTS, TOEFL." Accessed: Jan. 24, 2025. [Online]. Available: https://www.kaggle.com/datasets/duongduong123/english-test-prep-data-toeic-ielts-toefl

[22] X. Wang, L. Zhang, and T. He, "Learning performance prediction-based personalized feedback in online learning via machine learning," Sustainability, vol. 14, no. 13, p. 7654, 2022.

[23] F. Qiu et al., "Predicting students' performance in e-learning using learning process and behaviour data," Scientific Reports, vol. 12, no. 1, p. 453, 2022.

[24] V. M. Jayakumar et al., "Advancing Automated and Adaptive Educational Resources Through Semantic Analysis with BERT and GRU in English Language Learning," International Journal of Advanced Computer Science & Applications, vol. 15, no. 4, 2024.