TECHNIUM
SOCIAL SCIENCES JOURNAL

Vol. 80/2026

A New Decade for Social Changes

Technium Social Sciences Journal
Vol. 80, 35-55, February, 2026
ISSN: 2668-7798
www.techniumscience.com

Construction of AI Learning Support Model in Response to Conversation Content and Emotional Changes by EQ

Kento Yasuda[1*], **Hiromitsu Shimakawa**[2], **Fumiko Harada**[3]

[1, 2, 3] Ritsumeikan University

*Corresponding author: Kento Yasuda, is0584ih@ed.ritsumei.ac.jp

Abstract. Instructors generally find it difficult to intuitively grasp learners' emotions and level of understanding. Therefore, there is an increasing need for individualized instruction that provides support tailored to the student's emotional state and level of understanding. To objectively assess learner states, AI-based learning support grounded in physiological indicators that interpret emotional changes is indispensable. This study estimates learner states in data science education based on emotional changes detected through conversation content and EQ. After state estimation, the study proposes a method that constructs an AI-based learning support model according to learner states. It identifies appropriate instructional strategies for instructors. Instructor shortages and the difficulty of providing individually optimized support have become apparent. This study collects conversational data and electrodermal activity to estimate learner states. A hidden Markov model (HMM) estimates learners' internal states from the conversational behaviors of instructors and learners. This study conducts one-on-one tutoring sessions between an instructor and a learner to evaluate the effectiveness of the proposed method. The experimental results reveal three states: a trial and error state, a state of searching for understanding, and an initial state of reaching understanding. The above results indicate that learner states can generally be estimated from conversation content. The study also constructs a Random Forest model based on the estimated learner state and conversation content. The F1 score is 0.824, enabling the identification of key features strongly associated with each learning state. Furthermore, the study fine-tunes a BERT model using utterance-level dialogue data annotated with state labels to classify learner states. The F1 score of 0.9015 indicates high accuracy, demonstrating the model's ability to accurately estimate the learner's state based on conversation content. Analyzing the state transitions, the study can help instructors decide on effective teaching methods. The findings obtained in the study hold promise for enhancing the quality of individualized instruction and as a model for autonomous learning support through AI agents.

Keywords. EQ, Cognitive Load, Data Science, One-on-One Tutoring, EDA, BERT

1. Introduction

In most fields of modern industry, it is essential to analyze data to improve work efficiency. As a result, the demand for data scientists has increased rapidly. The Ministry of Education, Culture, Sports, Science, and Technology in Japan launched the Mathematics, Data Science, and AI Education Program Accreditation System in September 2025 [1]. These initiatives demonstrate the importance of data science.

Data scientists must possess data analysis abilities grounded in IT skills. Nevertheless, a survey by the Ministry of Economy, Trade and Industry warns Japan will face a shortage of approximately 545,000 IT professionals by 2030 [2]. Due to it, the shortage of IT personnel extends beyond the IT industry to affect other sectors, including manufacturing, logistics, and healthcare.

High-level education is necessary to bring up personnel equipped with IT skills and data analysis abilities. Such education should provide instruction tailored to each learner's emotional state. The instruction can only be achieved through individually optimized learning support.

On the other hand, the education sector has faced a severe shortage of teachers in recent years. According to a survey by the Japanese government, teacher shortage rates in Japan ranged from 0.10% to 0.33% across elementary to high schools in fiscal year 2021 [3]. Furthermore, the training system for teachers remains insufficient. The shortage of teachers and the insufficiency of the teacher training system make it difficult to provide optimal individual support to each learner.

Instruction tailored to each learner's emotional state should be achieved to overcome the shortage of data scientists. AI agents would be one promising solution for such instruction. AI agents would be one promising solution for such instruction. Previous studies have reported that AI-based adaptive learning systems improve learners' motivation and academic performance [4]. However, instructions that consider learners' emotional states have not yet been sufficiently established.

The study focuses on emotional intelligence (EQ). EQ refers to the ability to properly understand and appropriately regulate emotions. Previous studies have reported that learners with higher EQ tend to achieve higher academic performance [5]. The findings suggest that EQ may contribute to academic achievement. However, existing research has not yet established AI-based learning support models that incorporate EQ. Methods to utilize EQ remain a challenge.

As a method utilizing EQ, the study proposes a learning support model based on conversation content and physiological indicators. AI agent would identify learners' emotions through dialogues with them to provide personalized information and learning plans. Learners' emotional changes are known to manifest as physiological responses [6]. Physiological indicators enable us to grasp emotional changes. The proposed method identifies effective teaching strategies for instructors according to learners' states, which means the method utilizes EQ.

Instructors' verbal behaviors are not uniform. Instructors differ in their amount of verbal output, with some instructors speaking frequently and others speaking less. Question rates from instructors reflect learners' levels of understanding. They decrease for learners whose understanding has progressed, whereas they increase for learners whose learning has stagnated. Based on these hypotheses, the study analyzes conversation content and physiological indicators. The analysis extracts effective teaching strategies for each learner state.

The proposed method first classifies learners' states using an HMM based on physiological indicators. The state classification identifies differences among learner states. Next, a random forest model is constructed with the HMM-derived states as the target variable and conversational features as explanatory variables. Furthermore, a BERT model is fine-tuned using conversation utterances labeled based on the HMM. These models identify effective instructional behaviors for instructors in each learner state. Eventually, the analysis of state transitions clarifies the actions that instructors and learners should take for each learner state.

Experimental results show that the HMM has classified learner states into three states. Each state is semantically interpretable. The F1-score is 0.824 for the random forest model while 0.9015 for the BERT model. The results indicate that accurate state discrimination is achievable. The study demonstrates the effectiveness of a learning support model that considers EQ. The model enables the extraction of appropriate instructional strategies for each learner state. The remaining parts of the paper are organized as follows.

- Section 2 describes the underlying techniques, reviewing related works on learner state estimation.
- Section 3 proposes a state estimation method based on physiological responses and conversation data.
- After describing the experimental setup, Section 4 presents the experimental results.
- Based on the experimental results, Section 5 discusses learner-state-specific characteristics and instructional effects.
- Section 6 concludes the paper as well as outlines future research directions.

2. Fundamental Techniques for Learner State Estimation

2.1. Cognitive Load

Cognitive load is the mental strain placed on working memory when a person performs a task. Working memory refers to the mechanisms required to keep things in mind when performing complex tasks such as reasoning, comprehension, and learning. Working memory holds a small amount of information for a short period of time for use in performing cognitive tasks. On the other hand, long-term memory is a vast accumulation of knowledge and a record of past events. It can essentially hold an infinite amount of information. Once information in working memory is stored in long-term memory, it is organized to be readily available. Cognitive load theory identifies three different types of cognitive load.

The first is task-intrinsic load that comes from the inherent difficulty of the learning task itself. It depends on the complexity of relationships that learners must process simultaneously to understand a task. Ayres describes task-intrinsic load as something that is “innate and fixed to the task” [7]. Sweller et al. state that task-intrinsic load is a load that educators cannot directly control [8]. Haler et al. also point out that the intrinsic load remains constant [9]. Task-intrinsic load is considered to be cognitive loads unrelated to learning. Instructional materials should be structured progressively from simple to complex content, which enables learners to gradually adapt to increasing task complexity [10]. Furthermore, instructional approaches may influence learners’ perceived task-intrinsic load [11].

The second is task-external load. Task-extrinsic load is the cognitive load caused by the teaching materials and the learning environment. This cognitive load does not directly contribute to comprehension of the learning content or knowledge construction. Van Merrienboer et al. define task-external load as cognitive load that is unnecessary for learning [12]. Much of the research on cognitive load theory has focused on task-external load as a factor that inhibits learning effectiveness, and aimed to reduce it [13][14]. Minimizing task-related external load is a key requirement for effective instructional design.

The third is learning-related load. It is also referred to as germane cognitive load. Learning-related cognitive load arises during the process of integrating information into long-term memory. The main purpose of education is to enable learners to store important information in their long-term memory. Instructional design should guide learners to actively engage in meaningful processing and knowledge elaboration, while ensuring that germane cognitive load increases appropriately. Ericsson et al. point out that expertise, which relies on information held in long-term memory, improves the ability to process information in working memory [15]. The cognitive load that enhances learning effectiveness is learning-related load. Since overloading working memory leads to poor performance, the other cognitive load should be suppressed.

2.2. Estimation of Internal States

Learner states change in response to multiple factors, such as levels of understanding and emotional conditions. Accurately capturing these factors is essential for learning support and for improving instructional practices by teachers. Previous studies have primarily estimated learner states using behavioral records such as test scores and task accuracy rates [16][17]. These approaches enable quantitative evaluation of learning outcomes.

However, prior research has pointed out that emotional changes during the learning process vary across individuals, which makes it difficult to directly capture learners' internal states [18]. In recent years, challenges to estimate internal states utilize diverse data such as physiological indicators and conversational information [19].

2.3. The Relationship between Physiological Indicators and Cognitive Load

Physiological indicators enable us to understand states related to emotion and stress. One of the widely used physiological indicators is electrodermal activity (EDA). It refers to the change in the electrical properties of the skin in response to the secretion of sweat on the skin. EDA responds strongly to psychological stimuli such as stress. It is also affected by emotional changes. It is caused because changes in stress or emotion would prompt sweating, which lessens skin resistance to increase electrical conductivity.

In individual instruction, the instructor guides the learner face-to-face. During the conversation, both parties change their emotions. Therefore, in individual instruction, both cognitive load-based stimuli and emotion-based stimuli may be present. Nourbakhsh et al. suggest that skin conductance activity increases under cognitive load [20]. Several studies have confirmed a positive correlation between electrodermal activity, a physiological indicator, and cognitive load [21][22]. Electrodermal activity can significantly and objectively distinguish different cognitive load levels, even under emotional changes. Therefore, by analyzing electrodermal activity, it is possible to estimate the learner's learning state.

2.4. Conversation-based Features and Learning State

Conversational information is an important source of information that externally expresses a learner's understanding and attitude. Conversation-based features are used to estimate the learner's learning state. In individualized instruction, there is continuous dialogue between the learner and the instructor. Changes in the content and amount of speech reflect the learner's understanding and cognitive state.

The number of characters spoken is an indicator of the amount of speaking activity of a learner, and has been used in dialogue-based learning state estimation [23]. The question rate indicates the proportion of question expressions in the entire utterance. It serves as an indicator reflecting the state in which learners' cognitive activities and processes of examining understanding are made explicit. Filler rate indicates the frequency of filler expressions such as "um" and "uh." It is associated with trial hesitation and uncertainty.

In general, as learners' understanding improves, the structure of the task becomes clearer, reducing cognitive load and confusion [8]. The improvement of the level of comprehension may decrease the rate at which learners ask questions and fill in the text. On the other hand, as content comprehension deepens, learners engage in more appropriate dialogue. Like physiological

TECHNIUM SOCIAL SCIENCES JOURNAL

Technium Social Sciences Journal
Vol. 80, 35-55, February, 2026
ISSN: 2668-7798
www.techniumscience.com

indicators, conversation-based features serve as effective indicators for estimating learners' comprehension status and learning progress.

3. State Estimation Based on Physiological Responses and Conversation

3.1. State Transitions through Individual Instruction

The study aims to construct an AI learning support model based on conversation content and physiological indicators, with the goal of extracting teaching methods tailored to each learner's state. Figure 3.1 shows an overview of the proposed method. The method enables us to extract the teaching methods that instructors should take according to each learner's state.

The proposed method estimates the learner's state using electrodermal activity, a physiological indicator. Learner feature comparisons verify differences in conversation feature states for each learner and instructor. The feature extraction process identifies important features from conversational data. Subsequently, the classification of large language models (LLMs) enables the classification of learner states from conversational sentences in utterance level. Finally, the calculation of state transitions clarifies the instructor's instruction content required for each learner state.

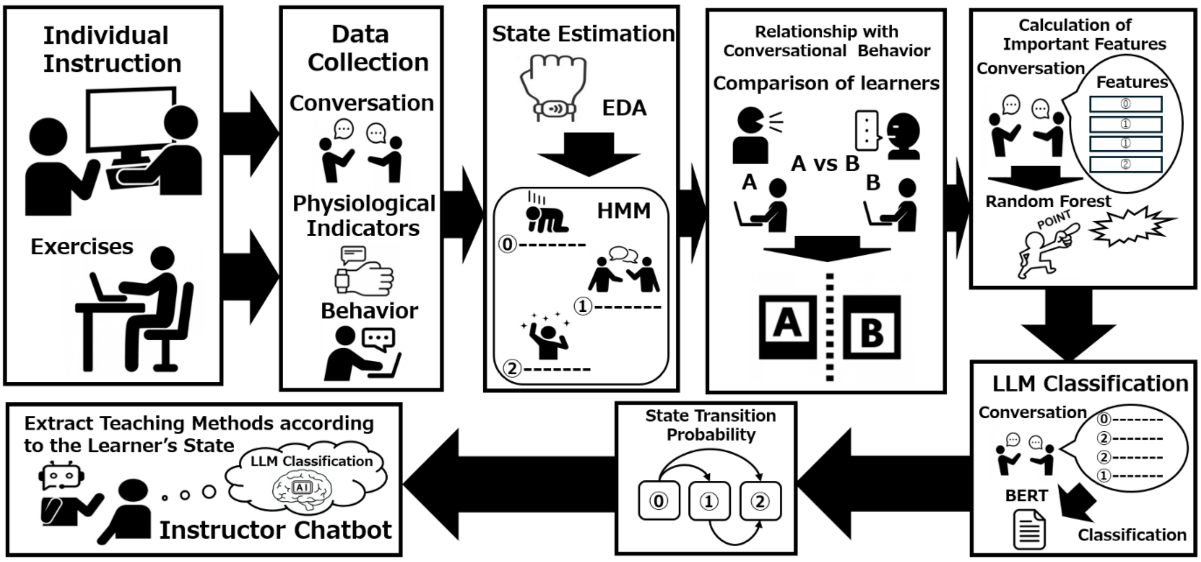

Figure 3.1: Method Overview

The study takes data science as an example of a specialized field that learners must master through individual instruction. However, the proposed method is not specific to the field. The proposed method assigns a data science task to collect conversational data, electrodermal activity, audio recordings, and performance scores.

The model construction uses multiple machine learning methods to extract effective teaching strategies for each learner state. First, a hidden Markov model (HMM) estimates learners' latent states from electrodermal activity to infer their internal states. In the feature analysis for each learner state, the method compares high-performing learners with low-performing ones. High-performing learners tend to exhibit lower question rates and filler rates, whereas low-performing learners tend to show higher question rates and filler rates. After state estimation, a random forest model extracts key conversational features for each learner state. Subsequently, BERT is fine-tuned to construct a model that classifies learner states at the utterance level. Finally, the study identifies the effective teaching methods for each learner state.

3.2. Student State Estimation Based on Latent States

The study collects EDA as a physiological time-series signal. For state estimation, the proposed method computes raw EDA, differential EDA, and change-point scores using ChangeFinder. Raw EDA represents the directly measured electrodermal activity during learning. It serves as an indicator that reflects changes in learners' emotional states and arousal levels. The level of raw EDA changes depending on the degree of stress and attention. Differential EDA captures differences in raw EDA across consecutive time intervals. This index captures sudden changes and temporary responses that are difficult to capture using the absolute value of EDA. By using differential EDA, transient emotional changes that occur during learning can be detected more clearly.

ChangeFinder builds an AR model of the actual measurements locally for the window width. It then performs smoothing using the constructed model. Based on the smoothed predictions, ChangeFinder locally learns a predictive autoregressive (AR) model. ChangeFinder treats the difference in log-likelihood between the predictive model and the observed-value model as a loss and defines this value as the change-point score.

Human cognitive load fluctuates in response to external stimuli, and such changes occur intermittently over time. Consequently, data observed at short time intervals may contain transient fluctuations. To reduce the influence of momentary variations and to capture stable trends in cognitive load, this study applies one-minute averaging. In addition, this study segments approximately 30 minutes of one-on-one tutoring conversations into one-minute intervals for analysis. By aligning with this time division, changes in skin electrical activity can be detected as changes in the learning process.

The study feeds raw EDA, differential EDA, and change-point scores to a hidden Markov model (HMM) to estimate state transition. Multiple features reduce the influence of changes that are likely to be treated as noise in a model using a single feature. The HMM estimates learners' internal states by considering the state transition most plausible to a sequence of observed data, which is never achieved by models addressing no sequence. The study determines the appropriate number of states in the HMM, using the log-likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC).

3.3. Learner Status Definition

To estimate learners' states, the study records conversations. During the experiment, the study records video of one-on-one tutoring conversations between an instructor and a learner using Temas. At the same time, Temas transcribes the speech of both participants using its transcription function. The automated transcription results contain errors. After the experiment, the study reviews the recorded videos to manually correct errors in the transcribed text.

The study quantifies instructors' behaviors toward learners based on the conversation content. Specifically, the study represents features that can be expressed as quantities, such as the number of characters in instructors' utterances and the number of questions posed by instructors, using measured values. On the other hand, items that require subjective judgment, such as whether the instructor guides the learner toward the correct answer or the instructor asks probing questions, are recorded manually by reviewing the video to note whether such behaviors occur for each learner state. Each behavior is represented as 1 when present, while 0 otherwise.

With AIC and BIC, the study determines the number of latent states estimated by the hidden Markov model as three. The following states turn out through manual analysis. These states are interpreted based on physiological indicators calculated for each state and conversational features. The first is the state of trial and error. Here, the learners do not fully understand the task content. They are engaged in trial and error. The second is the understanding search state. The learners begin their consideration, but have not understood yet. The third is the initial state of understanding. Although the understanding is not complete, it is the stage where the learners begin to grasp the clues to understanding.

3.4. Learner State Classification Using the Classification Model

To extract important conversational features, the study employs a random forest model due to the following reasons.

- The predictions using a random forest model are easy to discuss.
- The random forest model outperforms linear regression in prediction.
- Since the random forest model selects explanatory variables, it is useful to identify important explanatory variables.

For learner state classification, the study computes 18 types of conversational features from utterances based on instructors' behaviors during one-on-one tutoring. Conversation features that can be represented numerically for each state include: the number of characters in learner utterances, the number of characters in instructor utterances, the number of questions asked by the learner, the number of questions asked by the instructor, the learner filler rate, the instructor filler rate, the learner question rate, instructor question rate, the learner speaking time, and the instructor speaking time. Non-quantifiable explanatory variables for specific instructor behaviors include: affirming, guiding to the answer, providing the answer immediately, using concrete examples in explanations, asking probing questions, connecting to prior knowledge, showing teaching materials, and encouraging. The explanatory variables above are represented using one-hot encoding.

The study analyzes spectrograms and cepstrograms, which are acoustic features. A spectrogram visualizes the distribution of frequency components in each time segment by dividing an audio signal along the time axis based on the short-time Fourier transform (STFT). The cepstrogram is obtained by applying the inverse Fourier transform to the logarithmic power spectrum. However, no differences between states can be observed in the spectrum and cepstrum. Therefore, the study excludes speech features from the analysis.

Furthermore, to incorporate temporal context, the study augments conversational features with derived features, including the immediately preceding state, the state two steps earlier, state-wise changes, and state-wise mean values. The augmentation contributes to extracting features that are particularly important for classifying learner states into three categories. In addition, in case of extracting features using a random forest model, the objective variables are the following three binary ones: state 0 and any other state, state 1 and any other state, and state 2 and any other state. They also reveal important features for each state.

In constructing the classification model, the study uses the state labels estimated by the hidden Markov model as training data. In addition, BERT classifies learner states from utterance-based conversation sentences. Since it can simultaneously consider the contextual relationships of words within a sentence, it can grasp the meaning and intent of utterances based on context.

Preprocessing involves removing unnecessary line breaks and extra spaces in the dialogue to normalize the text into a format easy for the model to learn. Each utterance is treated as one sample. The input length is standardized to 256 tokens to reduce the impact of excessively long utterances on learning. After preprocessing, the text is divided into subwords using a tokenizer provided by the Japanese version of DeBERTa [24]. It allows us to deal with unknown words and compound words in Japanese. Each utterance is fed to the model as a sequence of tokens and information indicating the valid range of the tokens. By aligning all input sequences to the same length, the model construction achieves stable mini-batch learning. The labels are three-state labels estimated by the hidden Markov model. Each state reflects differences in the learner's understanding and behavior. To address class imbalance, the study introduces weighting based on the frequency of state occurrence into the loss function to improve classification performance. This method allows us to identify conversations that are effective for instruction in each state.

3.5. State Transition

In the analysis of learner state transitions, the study compares the states of all participants. After the occurrence frequency of the previous and subsequent states is figured out, the state transition probability is calculated by normalizing each previous state. In addition, the study uses the top 10 conversation features that clearly show state transition probabilities. The top 10 conversation features are: instructor's concrete examples, instructor's encouragement, instructor's affirmation, instructor's questions, instructor's proportion of speech, instructor's amount of speech, learner's questions, learner's filler rate, learner's proportion of speech, and learner's amount of speech. To evaluate changes in learners’ attitudes and instructors’ instructional features, the study uses Cohen’s d. It is the difference in the mean values of two groups divided by the standard deviation, which plays the role of an indicator of the extent of the difference in the mean values. Its negative value indicates that the corresponding feature has decreased after the state transition. The features themselves are non-negative. The sign is determined by the difference before and after the transition and the direction of the effect size. It allows us to extract instructional features that contribute to a transition to a good learner state based on physiological responses and conversational features.

4. Experiments and Results

4.1. Experiment Overview

The study conducts experiments in individual tutoring to examine the effectiveness of the proposed method. The task learners work on is a data analysis assignment on non-negative matrix factorization (NMF). The experiment involves dialogue between the learner and the instructor. The study verifies whether effective teaching methods can be extracted for each learner state. The one-on-one tutoring session follows the procedure described below.

- A preparatory lecture is conducted to deepen learners’ understanding of the pre-specified algorithm.
- Over approximately 30 minutes, learners work on a practice assignment with support from the instructor.
- Learners take a confirmation test to assess their understanding of the learning content.

The participants consist of 12 instructors and 12 learners. The instructors are fourth-year undergraduate students or graduate students, while learners are third-year undergraduate students. The instructors have studied data analysis for more than one year. They have experience in analysis using non-negative matrix factorization in lectures. All instructors participate in the experiment only after they have attained sufficient abilities to teach the assignment. Before the experiment begins, instructors are provided with the assignment and sample solutions. All instructors participate in the experiment only after they have prepared enough to teach the assignment. There is no specific guideline on instructional methods. The instructors have told to teach as they normally do. Students can see the assignment before the experiment begins. They cannot prepare in advance. After the assignment, students take a test

Technium Social Sciences Journal
Vol. 80, 35-55, February, 2026
ISSN: 2668-7798
www.techniumscience.com

to confirm their understanding of NMF. There are 20 questions, consisting of 16 true/false and 4 written ones. Each of them is worth 2 points; 40 points in total.

To measure learners' cognitive load, the study measures electrodermal activity (EDA) with an Embrace PLUS. Embrace PLUS can acquire skin potential activity values at 4Hz when wrapped around the wrist. Before attachment, the study checks the condition of the participant's wrist to ensure that it is not wet with sweat or moisture. After that, the learner wears the device on either their left or right wrist before starting the experiment. After wearing the device, students receive individual instruction in a natural posture.

4.2. Physiological Responses According to Learner State

Analysis of physiological responses examines data from highly talkative instructors and high-achieving students to identify clear classifications of student states. The reason is that when instructors talk to students more, it increases their cognitive load, resulting in a clear difference. For state classification using a hidden Markov model, this analysis uses three features: raw EDA, differential EDA, and change-point scores derived from ChangeFinder. In the experiment, ChangeFinder picks up learner states during one-on-one tutoring at all time points where the change-point score exceeds 50. Figure 4.1 shows the raw EDA and change point scores. The yellow dashed line is the change point score, and the blue line is the raw EDA. The horizontal axis shows the time spent in individual instruction (minutes). The left side of the vertical axis shows the EDA value (μs), while the right side shows the change point score.

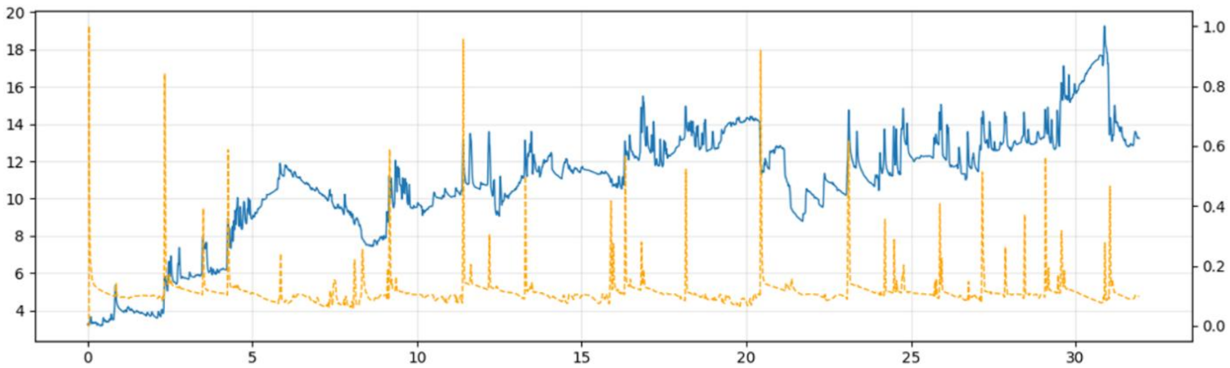

Figure 4.1: Raw EDA and Change Point Scores

To examine learner state classification, the study categorizes the latent states of the learning process using a hidden Markov model. Figure 4.2 shows the state transitions of a learner paired with an instructor who speaks a lot and a learner who has high grades. The horizontal axis shows the time (minutes) during individual instruction, and the vertical axis shows the learner's state estimated by the hidden Markov model. There are three learner states: 0, 1, and 2. The study computes the log-likelihood, AIC, and BIC to validate the appropriateness of the three-state classification. Table 4.1 shows the basis for determining that the number of states of a hidden Markov model is three. Figure 4.2 clearly classifies the learner's state into three distinct states. We can see that the states transition gradually over time. Table 4.1 shows that the log-likelihood is the largest, while the AIC and BIC are the lowest values. It means the model of three states adequately explains the observed data without overfitting.

Technium Social Sciences Journal
Vol. 80, 35-55, February, 2026
ISSN: 2668-7798
www.techniumscience.com

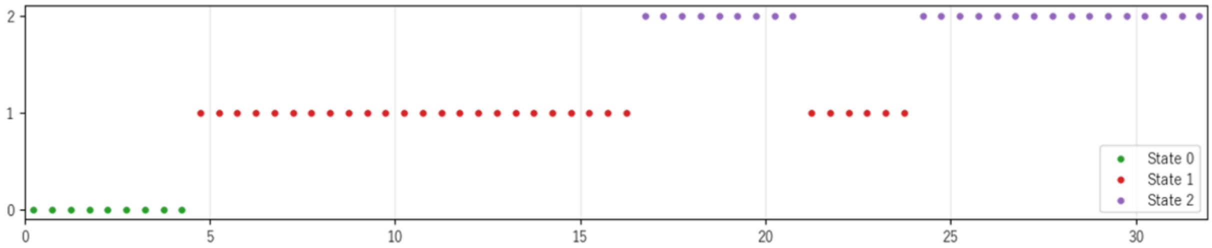

Figure 4.2: State Transitions for a High-Performing Learner

Table 4.1: Evaluation Index for Number of States

Number of States	Log-likelihood	AIC	BIC
2	-252.422	555.711	588.095
3	-179.435	531.422	587.553
4	-220.175	533.363	617.559
5	-209.003	549.513	666.093
6	-202.635	568.171	721.451

To clarify the meaning of each state, it is necessary to confirm the differences among states. The study focuses on the range of variation in raw EDA. It examines differences among states by confirming the variance within each time window. The time series of raw EDA is segmented into one-minute intervals to compute the variance within each interval. Figure 4.3 shows the evolution of raw EDA dispersion over one minute. The horizontal axis shows the elapsed time (minutes) during individual instruction, and the vertical axis shows the variance of raw EDA in each 1-minute interval (μS²). Figure 4.3 shows that a high variance zone is observed in State 1 around 20 to 25 minutes. It suggests that physiological responses fluctuate significantly in this section, which is potentially caused by high cognitive load for learners. On the other hand, the variance decreases in the subsequent intervals, showing that the physiological response is relatively stable. The change indicates that learners have attained a certain level of understanding of the task content. The learner may transition to a different state due to changes in cognitive load. During this period, the instructor's explanation comes to an end. It may reduce the cognitive load. It is also possible that the instructor's utterances resolve the learner's uncertainty to relax the learner.

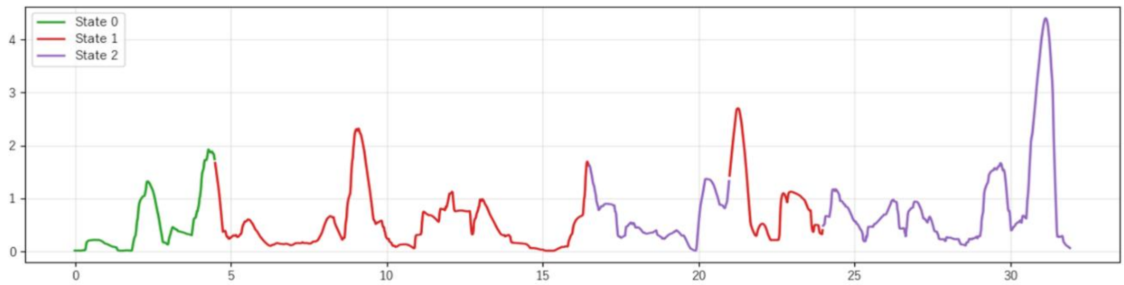

Figure 4.3: Variance of Raw EDA per Minute

The learner's state may change in response to cognitive load. Figure 4.4 shows the mean and variance of raw EDA for each learner state. The left and the right figure shows the mean of the raw EDA (μs), and the variance of the raw EDA (μS²), respectively. Figure 4.4 shows that the

Technium Social Sciences Journal
Vol. 80, 35-55, February, 2026
ISSN: 2668-7798
www.techniumscience.com

average value of raw EDA is highest in state 2, followed by state 1, and then state 0. The dispersion of native EDA shows low values in state 1 while high values in states 0 and 2.

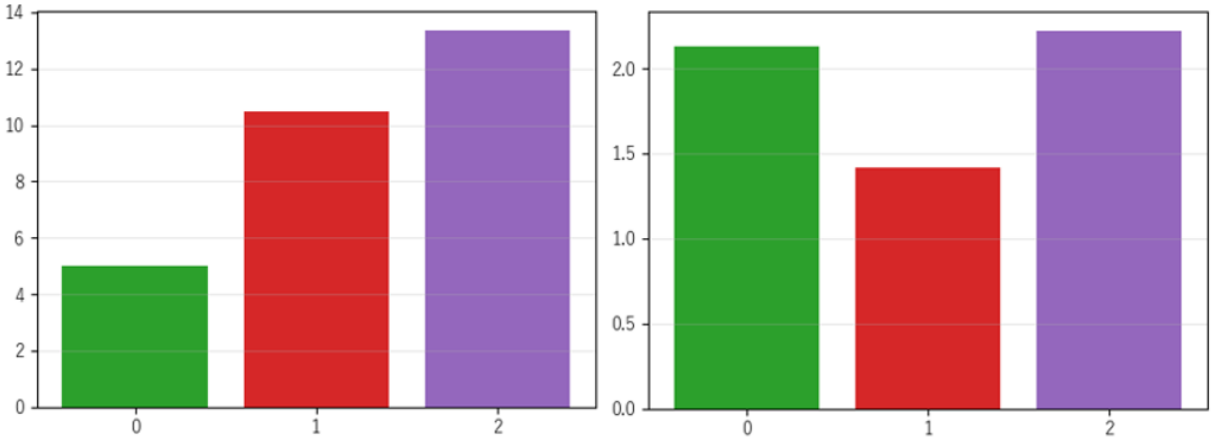

Figure 4.4: Mean and Variance of Raw EDA

To capture characteristics specific to each learner state, this analysis computes conversational features. Table 4.2 shows conversation features of instructor and learner speaking time, filler rate, and question rate. From Table 4.2, let us read the conversation features using the learner state defined in Section 3.3.

Table 4.2: Conversation Features by Learner State (High-Performing Learner)

State	Speaking time (Instructor)	Speaking time (Learner)	Filler Rate (Instructor)	Filler Rate (Learner)	Question Rate (Instructors)	Question Rate (Learner)
0	226 s	43 s	10.4 %	15.4 %	22.9 %	23.1 %
1	776 s	116 s	13.1 %	19.2 %	12.4 %	15.4 %
2	515 s	179 s	7.1 %	8.8 %	14.3 %	2.9 %

First, in state 0, the learner's speaking time is short. Both the question rate and filler rate show relatively high values. The learners stay in a stage where they cannot fully grasp the task content. They speak while organizing their own thoughts. As another feature, the instructor spends a lot of time talking. The learners are in the state of trial and error, relying more on explanations and assistance from the instructor than on the students themselves.

In state 1, the learner's speaking time increases while the rate of questions remains relatively stable. It indicates that although learners are actively thinking about the task, they have not yet reached an understanding. Accordingly, the stage is interpreted as one in which learners produce many utterances accompanied by uncertainty. We can also say it is the understanding search state where the instructor speaks the most, supporting learners' thinking while promoting understanding.

Finally, in state 2, learners exhibit the longest speaking duration, while both the question rate and filler rate decrease substantially. The learner has grasped the key points of the task content to articulate their own thoughts in a stable manner. The instructor's speaking time and filler rate have decreased. It represents that the state has shifted to learner-centered dialogue with excessive instructor intervention decreased.

Technium Social Sciences Journal
Vol. 80, 35-55, February, 2026
ISSN: 2668-7798
www.techniumscience.com

Based on the above, differences in the mean and variance of physiological indicators such as raw EDA are likely to reflect state changes corresponding to learners' stages of understanding. The results suggest that the learner states estimated by the hidden Markov model appropriately capture changes in cognitive states during the learning process.

4.3. Learning Situations Due to Instructor Behaviors

Differences in instructors affect the learning situation. Let us compare instructors who talk a lot with instructors who talk less. Generally, in the experiments, instructors who speak more advance the understanding of the learner. On the other hand, instructors who speak less tend to have students with lower grades. The state transitions for learners with low grades are shown in Figure 4.5. In Figure 4.5, State 1, observed immediately after the start of the experiment, is a conversation in the preparatory stage before the start of individual instruction, because the section includes utterances by third parties other than the learner and the instructor. It is excluded from the analysis. Table 4.3 shows the speaking time, filler rate, and question rate in a session of a low-achieving learner and an instructor with low speech output.

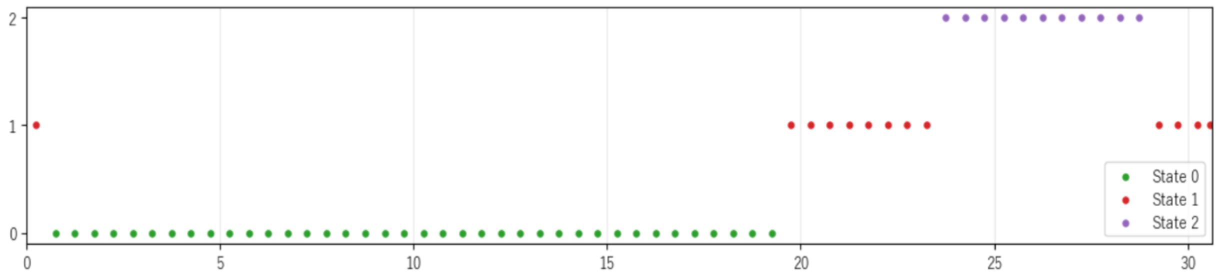

Figure 4.5: State Transitions for a Low-Performing Learner

Table 4.3: Conversation Features by Learner State (Low-Performing Learner)

State	Speaking time (Instructor)	Speaking time (Learner)	Filler Rate (Instructor)	Filler Rate (Learner)	Question Rate (Instructors)	Question Rate (Learner)
0	457 s	728 s	13.4 %	11.4 %	2.4 %	11.0 %
1	168 s	191 s	20.5 %	11.1 %	2.6 %	11.1 %
2	160 s	174 s	9.8 %	10.0 %	5.9 %	8.3 %

From, conversational features are read using the learner state defined in Section 3.3. In state 0 of Table 4.3, the learner's speaking time is long, and the instructor's speaking time is also relatively long. Learners' filler and question rates are high. In the state, the learner has not fully understood the task content, which means the individual is currently in a trial-and-error phase. The learner is working on the tasks, but cannot formulate clear questions.

In state 1, the instructor's filler rate is particularly high. The learners' filler and question rates are also high. This shows that both of them express hesitation and uncertainty during their speech. The state corresponds to the “understanding search state,” where one is progressing in their thinking about the task but has not yet reached understanding. Although learners are actively asking questions, their understanding is not yet stable.

Finally, in state 2, learners exhibit lower question and filler rates. The low rate of learners' questions suggests that their uncertainty about the task content is reduced. This state corresponds to the initial state of understanding, where one has begun to grasp the thread of understanding, though it is not yet complete. Learners speak based on their own understanding, with minimal instructor intervention.

As shown above, just as with high-performing learners, distinct differences in conversational features are evident for each learner state. Furthermore, as shown in Figure 4.2, high-performing learners ultimately reach the initial stage of understanding. In contrast, as shown in Figure 4.5, low-performing learners remain in the understanding search state. The one-on-one tutoring session ends before they reach sufficient understanding. The difference indicates that there is a clear distinction between high-performing learners and low-performing learners in the state transitions of the learning process.

4.4. Teaching Methods for Each Student's Condition

In classifying learner states, the study identifies key teaching methods based on conversational features. Let us examine the acoustic features of the conversations. Figure 4.6 shows the spectrogram and cepstrogram as speech features. The horizontal axis represents time (minutes). In the spectrogram on the left of Figure 4.6, the vertical axis represents frequency (Hz), while in the ceptragram on the right, it represents quefrequency (ms). The shade of the color is expressed as the amplitude (dB). Figure 4.6 shows the results for conversations between instructors with high speech volume and high-achieving learners, using the same subjects as in Section 4.2. The brightness of the colors in the figure represents the amplitude. The brighter the color, the larger the amplitudes. The calculation conditions are a sampling frequency of 16 kHz, an FFT length of 1024, and a hop length of 256. The amplitude is converted to a dB scale based on the maximum value. On the other hand, the brighter the cepstrogram, the stronger the periodicity corresponding to the pitch of the voice. The analysis extracts the quefrency range from 2 ms to 12 ms corresponding to the fundamental frequency. Figure 4.6 shows that consistent patterns across learner states are not clearly discernible in either the spectrogram or cepstrogram. The study does not use acoustic features as explanatory variables for learner state classification.

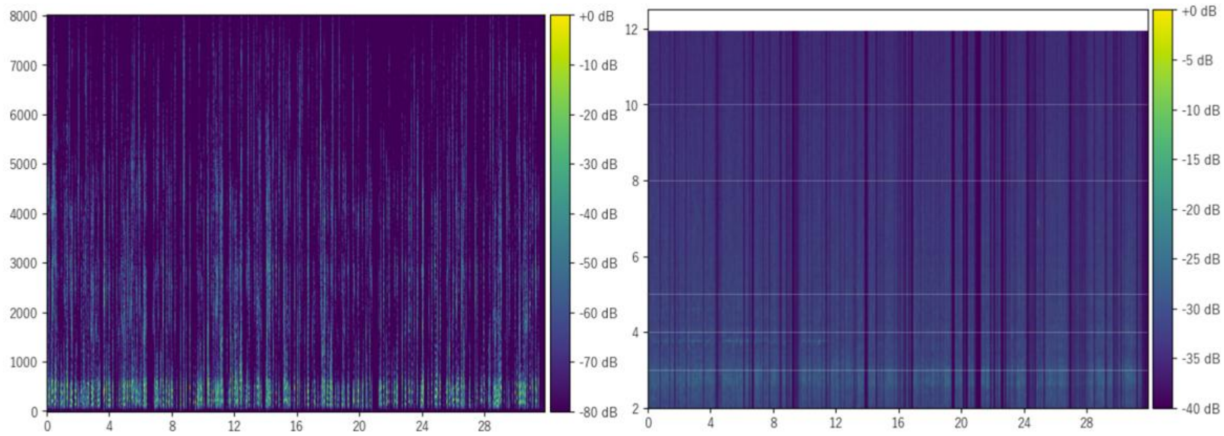

Figure 4.6: Spectrogram and Cepstrogram

In classifying learner states, the study applies a random forest model to identify appropriate instructional methods. This study optimizes hyperparameters using grid search. As a result, the F1-score of the model is 0.824. To identify key conversational features influencing the learner's three states, Figure 4.7 shows the top five variables by importance as determined by the Random Forest model. Based on experiments involving instructor–learner pairs, the analysis aggregates

Technium Social Sciences Journal
Vol. 80, 35-55, February, 2026
ISSN: 2668-7798
www.techniumscience.com

67 samples from 12 pairs. It splits the data into 75% and 25% for training and testing, respectively.

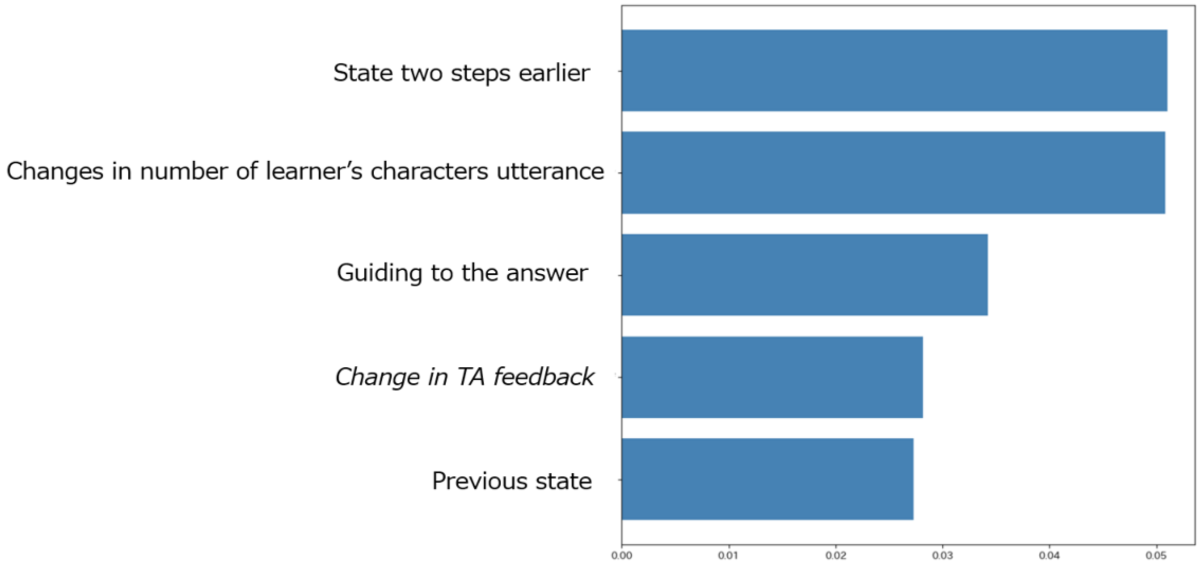

Figure 4.7: Variable Importance in a Random Forest Model

Figure 4.7 shows that the variables with the highest importance are, in order: "State two steps earlier", "Changes in number of learner's characters utterance ", and "Guiding to the answer". This result indicates that the learner's current state is strongly influenced not only by the immediately preceding state but also by the state prior to that. Furthermore, changes in the amount of speech of learners are important. Increases or decreases in speech behavior are an effective indicator for capturing changes in a learner's state. In addition, the instructor's instructional behavior of "guiding students to the answer" is important. It suggests that support designed to encourage learners to think may contribute to improving learning progress.

4.5. State Estimation Accuracy Based on Conversation Content

The study trains BERT to classify the states of conversational sentences using the labels of learner states estimated by a hidden Markov model as training data. Table 4.4 shows the classification performance evaluation metrics for each state. The total number of correct answers is 619, and the F1-Score for each state is 0.88 or higher. In particular, the average F1-Score for states 0 and 2 is 0.9015. We can accurately estimate learner states using only electrodermal activity and conversational data.

Table 4.4: Classification Performance

Number of States	Precision	Recall	F1-Score	Number of Samples
0	0.9177	0.9021	0.9099	235
1	0.8902	0.8848	0.8875	165
2	0.8929	0.9132	0.9029	219

Technium Social Sciences Journal
Vol. 80, 35-55, February, 2026
ISSN: 2668-7798
www.techniumscience.com

Figure 4.8 shows the confusion matrix for state classification using BERT, along with the trends in Training Loss and Validation Loss during the learning process. The blue line shows the loss on the training data (train_loss), and the orange line shows the loss on the validation data (eval_loss). Figure 4.8 shows that both the loss on the training data and the loss on the validation data consistently decrease as training progresses. Moreover, the validation loss does not show any significant increase. These results indicate that the proposed model is trained stably without causing overfitting.

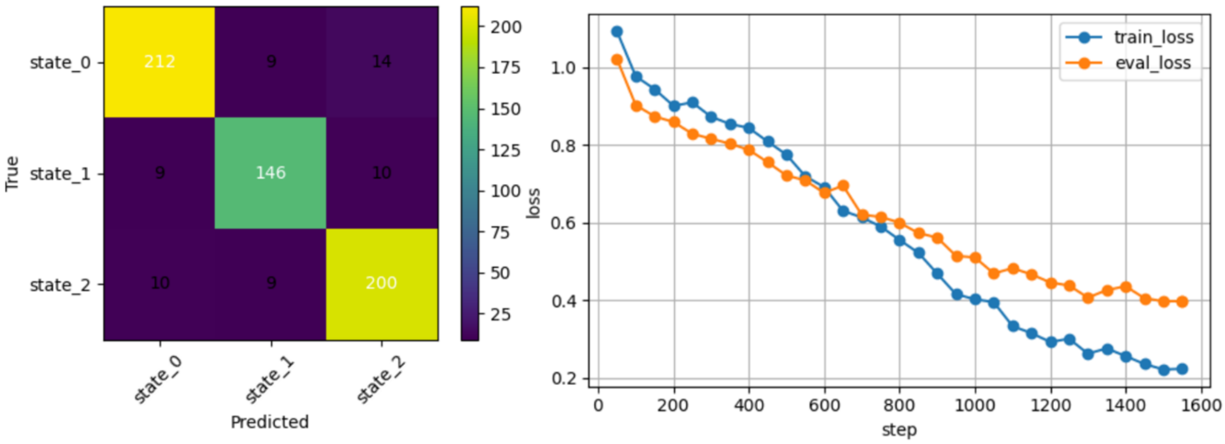

Figure 4.8: Confusion Matrix and Loss Function

4.6. Predicting the Optimal Teaching Method for Each Learner

This analysis calculates the frequency of occurrence of the pre-state and post-state for all participants. The analysis normalizes the calculations for each state to extract features representing learner attitudes and instructor instructional behaviors. Cohen's d is the difference between the means of two groups divided by the standard deviation. It is an indicator of how large the difference in the means is. Figure 4.9 shows the results of Cohen's d for the state transition probability.

Figure 4.9: State Transition Probability (Cohen's d)

Technium Social Sciences Journal
Vol. 80, 35-55, February, 2026
ISSN: 2668-7798
www.techniumscience.com

Figure 4.9 shows clear differences in the changes in conversation characteristics for each state transition. First, in the transition from state 0 to state 1, "instructor’s concrete example" and "instructor’s amount of speech" show positive values. It is a transition in which the learner lacks confidence. Their understanding begins to improve when the instructor provides concrete examples. Next, in the transition from state 0 to state 2, both “instructor's questions” and “learner's questions” indicate positive values. It suggests that the instructor's questions may have stimulated students' thinking. Their understanding may have progressed to a new level through dialogue. In the transition from state 1 to state 2, the “learner's proportion of speech and “learner’s amount of speech” show positive values, while the “instructor's amount of speech” tends to be suppressed. It means that instructors should avoid excessive explanations when students begin to think independently. The transition deepens understanding by encouraging learners to explain concepts in their own words. Finally, in the transition from state 2 to state 1, the “instructor's amount of speech” shows a positive value, while the “learner's amount of speech” and the “learner's proportion of speech” show negative values. It causes learners to become passive and reduces their speech output. On the other hand, although instructors provide support increasing explanations, it does not sufficiently help learners regain their confidence or maintain their understanding. In other words, it is a transition from the "initial state of understanding" back to the "understanding search state".

From the above, it is clear that the instructor's responses and manner of speech, tailored to the learner's state, are crucial in the process of initiating the learner's understanding. In particular, it is necessary for instructors to encourage learners' independent thinking not through one-sided explanations, but through posing questions and adjusting the amount of speech. The experiment results demonstrate that the proposed model can accurately identify instructional strategies that instructors should adopt for each learner state.

5. Characteristics of Learner State and Instructional Effectiveness

5.1. Effective Teaching Methods for Each Learner State

To identifying effective teaching methods for each learner state, let us clarify the key conversational features for each state. The random forest model performs binary classification for each of the three states. For example, one of the objective variables indicates state 0 or other states. The same applies to states 1 and 2. On the other hand, the explanatory variables utilize conversational behavior features. For each classification, the random forest model calculates the importance of features. Figure 5.1 shows the top five variables by importance for each state.

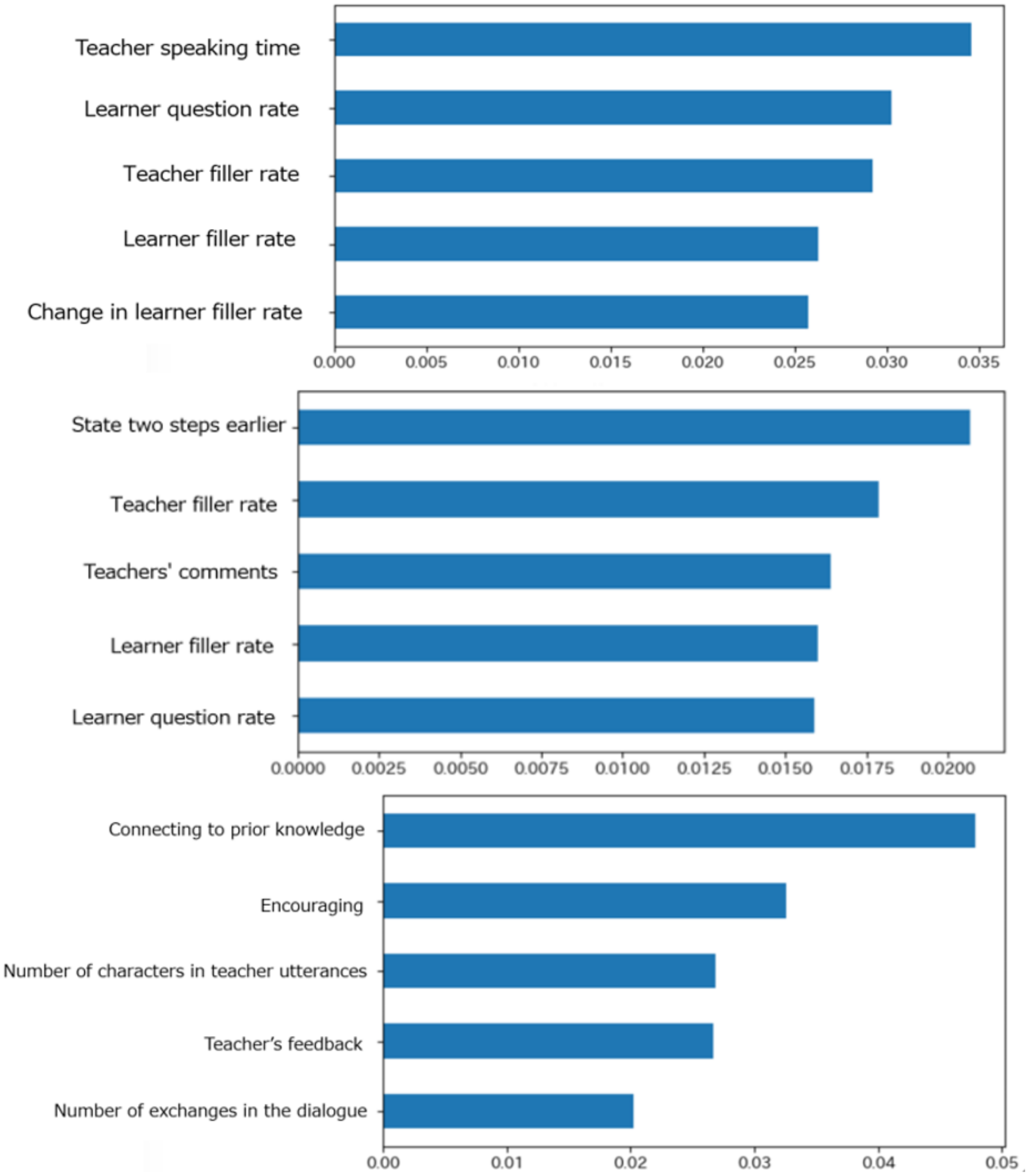

Figure 5.1: Identifying Key Features (State 0, State 1, State 2)

Figure 5.1 shows that in identifying state 0, the variables exhibit high importance in the following order: “state two steps prior,” “instructor's filler rate,” and “instructor's speech proportion.” State 0 is strongly influenced not only by the immediately preceding state but also by the learner's state before that. Instructors are also frequently hesitant and making adjustments when speaking. In other words, when learners are trying and failing, instructors also need to seek out appropriate intervention methods. In other words, when learners are trying and failing, instructors also need to seek out appropriate intervention methods.

In identifying state 1, the most important variables are "instructor speaking time," "student question rate," and "instructor filler rate," in that order. It involves an increase in explanations and questions posed by the instructor, along with numerous attempts to facilitate understanding for struggling learners. Furthermore, because instructors' filler usage rates are high, instructors themselves adjust their speech content while checking the students' reactions. In state 1, the learner is thinking towards understanding. On the other hand, it is clear that learners still need support from instructors.

When identifying state 2, the most important variables are "connecting past knowledge," "encouragement," and "number of characters spoken by the instructor," in that order. This result indicates that learners are in the process of deepening their understanding by connecting new content to existing knowledge. At this stage, it is important for instructors to organize knowledge and provide positive feedback. The number of characters spoken by the instructor is also important. In state 2, it is the well-organized explanation of content rather than the sheer volume of speech that supports learners' understanding.

5.2. Comparison with Time Series Analysis

The study compares its results with those of Prophet forecasting model, which is recently noted [25]. Prophet has the advantage of being able to extract trends even in the presence of noise. It also provides a smooth approximation as a continuous value. The original data features the state, time, and conversation sentences estimated using HMM. The objective variable is the current state. A continuous regression model predicts whether the learner is approaching State 2 using a linear function. The model is constructed for each individual. Figure 5.2 shows the prediction results for HMM and Prophet. The horizontal axis indicates the time (minutes : seconds) during individual instruction, while the vertical axis indicates the learner's state. The chart shows the results for high-performing learners and low-performing learners. The blue line shows the prediction results from the HMM, while the orange line shows the prediction results from Prophet.

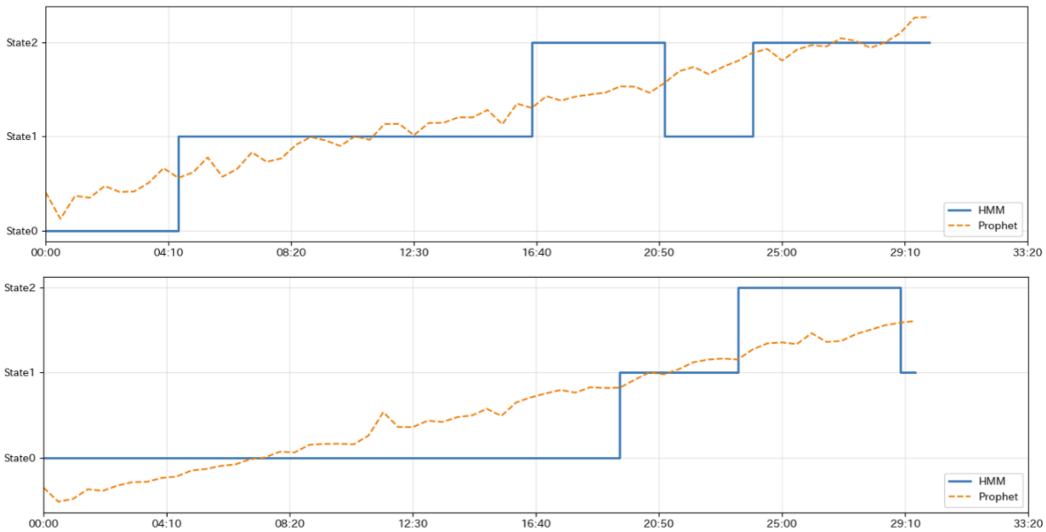

Figure 5.2: Comparison of HMM and Prophet (High- and Low-Performing Learner)

The graphs for HMM and Prophet are similar, but it is more difficult to see the changes than with HMM. For the data of 12 learners, there are many differences between the results of HMM and Prophet. To classify the learner's state, it is clear that is necessary to use HMM as the optimal teaching method for each learner state.

5.3. Limitations of this Method

The method used in this study focuses on analyzing cases in which high performance is confirmed. This method may not be applicable to all learning situations. Especially, it is necessary to verify whether the method is effective for cases where learning outcomes are insufficiently achieved. The study employs EDA as a physiological indicator for estimating learner states. However, these measurements require specialized equipment. There is a constraint regarding introduction costs. The operation of the proposed method needs careful handling so that the sensor works properly.

6. Conclusion

The study proposes a learning support model necessary for identifying the instructional methods instructors should adopt for each learner state. The proposed method classifies the learner's internal state based on the content of the conversation and physiological indicators to extract effective teaching actions.

An experiment to verify the effectiveness of the proposed method confirms that learner states can be classified into three semantically interpretable states. Based on three types of features calculated from skin potential activity, the hidden Markov model estimates the learner's state into three states. When estimating the three states, the F1 score is 0.824. It allows us to identify important instructional actions for each learning state. When estimating three states based on sentence units using BERT, the F1 score is 0.9015. The proposed method can estimate the learner's state with high accuracy. It has sufficient performance in determining the learner's state.

The state transition analysis reveals that instructional behaviors play different roles depending on the changes in learner states. In particular, in the transitions that develop understanding, it is effective for instructors to avoid excessive explanations, instead, to encourage students to think.

From the above, the study can extract the teaching methods that instructors should take depending on the learner's state. It clarifies the effectiveness of a learning support model that corresponds to EQ. The proposed method is applicable not only to instructor support in one-on-one instruction but also has the potential for future application to autonomous learning support by AI agents.

This study relies on a dataset limited to specific instructor–learner combinations, which constrains its scale. Other issues include the fact that the study relies on the measurement of physiological indicators. Future work will collect larger-scale data involving diverse combinations of instructors and learners.

Acknowledgments

We would like to thank many people for their guidance and encouragement in the preparation of this paper. We would like to thank everyone who willingly participated in the experiments during the execution of this paper. Finally, we would like to thank the members of the Data Engineering Laboratory for their invaluable advice and cooperation.

Conflict of interest

The authors declare no conflicts of interest in relation to the publication of this article.

References

[1] Ministry of Education, Culture, Sports, Science and Technology (MEXT), Higher Education Bureau, “Accreditation system for mathematics, data science, and AI education programs,” Sep. 25, 2025 (in Japanese).
[2] Ministry of Economy, Trade and Industry (METI), “Survey on IT human resource supply

and demand," commissioned research report (Mizuho Information & Research Institute), Mar. 2019 (in Japanese).
[3] Ministry of Education, Culture, Sports, Science and Technology (MEXT), "Fact-finding survey on teacher shortages," pp. 1–13, 2022 (in Japanese).
[4] O. Onesi-Ozigagun, O. O. Adebayo, and S. A. Ajayi, "Revolutionizing education through AI: A comprehensive review of enhancing learning experiences," International Journal of Applied Research in Social Sciences, vol. 6, no. 4, pp. 589–607, 2024, https://doi.org/10.51594/ijarss.v6i4.1011.
[5] P. Qualter, A. Gardner, C. Pope, and K. Whiteley, "The role of emotional intelligence in college students' success," Journal of Educational Psychology, vol. 113, no. 3, pp. 501–513, 2021, https://doi.org/10.1037/edu0000466.
[6] R. Pekrun, "The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice," Educational Psychology Review, vol. 18, no. 4, pp. 315–341, 2006, https://doi.org/10.1007/s10648-006-9029-9.
[7] P. Ayres, "Using subjective measures to detect variations of intrinsic cognitive load within problems," Learning and Instruction, vol. 16, no. 5, pp. 389–400, 2006, https://doi.org/10.1016/j.learninstruc.2006.09.001.
[8] J. Sweller, J. J. G. van Merriënboer, F. G. W. C. Paas, "Cognitive architecture and instructional design," Educational Psychology Review, vol. 10, no. 3, pp. 251–296, 1998, https://doi.org/10.1023/A:1022193728205.
[9] B. S. Hasler, B. Kersten, J. Sweller, "Learner control, cognitive load and instructional animation," Applied Cognitive Psychology, vol. 21, no. 6, pp. 713–729, 2007, https://doi.org/10.1002/acp.1345.
[10] J. J. G. van Merriënboer, P. A. Kirschner, L. Kester, "Taking the load off a learner's mind: Instructional design for complex learning," Educational Psychologist, vol. 38, no. 1, pp. 5–14, 2003, https://doi.org/10.1207/S15326985EP3801_2.
[11] P. Gerjets, K. Scheiter, R. Catrambone, "Designing instructional examples to reduce intrinsic cognitive load: Molar versus modular presentation of solution procedures," Instructional Science, vol. 32, no. 1–2, pp. 33–58, 2004, https://doi.org/10.1023/B:TRUC.0000021809.10236.71.
[12] J. J. G. van Merriënboer, J. Sweller, "Cognitive load theory and complex learning: Recent developments and future directions," Educational Psychology Review, vol. 17, no. 2, pp. 147–177, 2005, https://doi.org/10.1007/s10648-005-3951-0.
[13] R. E. Mayer, P. Chandler, "When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages?," Journal of Educational Psychology, vol. 93, no. 2, pp. 390–397, 2001, https://doi.org/10.1037/0022-0663.93.2.390.
[14] R. E. Mayer, G. T. Dow, S. Mayer, "Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds," Journal of Educational Psychology, vol. 95, no. 4, pp. 806–813, 2003, https://doi.org/10.1037/0022-0663.95.4.806.
[15] K. A. Ericsson, "Long-term working memory," Psychological Review, vol. 102, no. 2, pp. 211–245, 1995, https://doi.org/10.1037/0033-295X.102.2.211.
[16] R. S. Baker, P. S. Inventado, "Educational data mining and learning analytics, "Learning Analytics, pp. 61–75, Springer, 2014, https://doi.org/10.1007/978-1-4614-3305-7_4.
[17] C. Romero, S. Ventura, "Educational data mining: A review of the state of the art," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 6, pp. 601–618, 2010, https://doi.org/10.1109/TSMCC.2010.2053532.
[18] S. D'Mello, A. Graesser, "Dynamics of affective states during complex learning,"

Learning and Instruction, vol. 22, no. 2, pp. 145–157, 2012, https://doi.org/10.1016/j.learninstruc.2011.10.001.
[19] P. Blikstein, M. Worsley, "Multimodal learning analytics and education data mining," Proceedings of the 3rd International Conference on Learning Analytics and Knowledge, pp. 102–106, 2013, https://doi.org/10.1145/2460296.2460316.
[20] N. Nourbakhsh, Y. Wang, F. Chen, R. A. Calvo, "Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks," in Proceedings of the 24th Australian Computer-Human Interaction Conference, pp. 420–423, 2012, https://doi.org/10.1145/2414536.2414602.
[21] P. Ayres, J. Y. Lee, F. Paas, J. J. G. van Merriënboer, "The validity of physiological measures to identify differences in intrinsic cognitive load," Frontiers in Psychology, vol. 12, 2021, https://doi.org/10.3389/fpsyg.2021.702538.
[22] B. Mehler, B. Reimer, J. F. Coughlin, "Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: An on-road study across three age groups," Human Factors, vol. 54, no. 3, pp. 396–412, 2012, https://doi.org/10.1177/0018720812442086.
[23] S. Bibauw, W. Van den Noortgate, T. François, and P. Desmet, "Dialogue systems for language learning: A meta-analysis," Language Learning & Technology, vol. 26, no. 1, pp. 1–24, 2022, https://doi.org/10.64152/10125/73488.
[24] P. He, X. Liu, J. Gao, W. Chen, "DeBERTa: Decoding-enhanced BERT with disentangled attention," in Proceedings of the International Conference on Learning Representations (ICLR), 2021, https://openreview.net/forum?id=XPZIaotutsD.
[25] S. J. Taylor, B. Letham, "Forecasting at scale," The American Statistician, vol. 72, no. 1, pp. 37–45, 2018, https://doi.org/10.1080/00031305.2017.1380080.