ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

The Deep Integration of Modern Educational Technology in Chinese Language Classroom Instruction

Xin Qi[1,*]

[1] School of Chinese Language and Literature, Jiaozuo Normal College, Jiaozuo, Hanan, 454000, China

SUMMARY: *Present age education technology, which is a product of current era, is obtaining more and more attention from experts and scholars. It changes abstract concepts into visible actual uses, crosses time and space restrictions, arouses students' study passion, solves teaching hard problems, and expands study visual fields. This article puts forward the SPOC blended learning method to construct an intelligent teaching framework for Chinese language courses, focusing on four aspects: prior-stage assessment, resource arrangement, teaching activity planning, and evaluation systems. Modern education technique utilizes speech recognition calculation methods to promote the correctness and speed of classroom interactive activities. A comparison has been done among the classification effect of many different models, that is the Gaussian Mixture Model (GMM), the Event-Based Frequency Model (EV-FM), and the i-vector. Actual experiments prove that the EV-FM can acquire more superior outcomes on the testing corpus. Afterwards, experience-based investigation which is built on picked case studies shows that the SPOC mixed teaching pattern for Chinese language intelligent classrooms, that is established on the EV-FM, therefore has an obviously active influence on students' participation in deep study and their ability in study methods. Furthermore, this model can promote the learning of students and the interaction inside classroom.*

KEYWORDS: *SPOC teaching model; EV-FM algorithm; speech recognition; Chinese language smart classroom*

1 Introduction

Modern epoch education technologies, which include multimedia-dependent teaching, electronic learning platforms, and virtual reality simulation experiments, have been broadly carried out in intelligent teaching work [1, 2].The combination of multi-media technology, which is featured by its attractive visual charm, life-like depiction, and mutual-action functions, has caused a great change in traditional teaching methods [3, 4]. The network study platforms break through the restrictions of time and space. This gives students the capability to get study materials at any time and from any place. Therefore, they have the contribution to the cultivation of the capabilities of students' self-regulated learning[5, 6]. Through the rebuilding of real world situations, virtual simulation experiments let students get practical abilities inside a safe and no-risk surrounding. This method thus effectively conquers the limits that connect with traditional experiment teaching[7, 8]. The modern education technologies not merely change the old teaching frameworks thoroughly, but also promote teaching effect and good quality [9].

*1295015010@jzsz.edu.cn
https://doi.org/10.65102/is2026622 Published: 30 April, 2026

Multimedia programs and online education platforms make the learning materials in domains such as language, science, technology become tangible objects. Therefore, they make the learning process become simpler, and thus can catch students' attention in a more effective way[10].Already in the year 2006, Reinhard and his work colleagues [11] many research works have indicated that the learning which is helped by multiple medias can greatly promote students' achievement when they handle complicated time-space events. The students who are in the multimedia-based learning group required clearly less time for finishing learning (105±24 minutes) when compared with those in the text-only learning group, which took 122±30 minutes. In 2012, Jian-Hua [12] one web-based multimedia course framework has been gotten developed by us. This framework with great skill has put together multimedia technology and the teaching ability of education workers. Through this way of doing, it comprehensively used the advantages of multimedia to promote the efficiency of physics teaching in higher education. Subsequently, GebreYohannes et al.[13] research already discovered that instruction which is based on multimedia was more successful in the enhancement of students' long-term memory for concepts of science, when compared to teaching methods which are conventional. In the educational experiment that they carried out, the ninth-grade students displayed very notable academic advancement in the science subject. It is worth pointing out that, not even one student was remained in the extremely low achievement level. De et al. [14] demonstrated through instructional case studies that diverse multimedia combinations positively impact academic performance, effectively addressing the unique challenges of social sciences and enhancing student learning outcomes. To improve resource integration, Zhao et al. [15] we have already established an online learning platform, it is pushed forward by the search engine of Google and contains a very great number of education materials. By means of experiments, it has been proved that this platform made classroom interactions more active and provided important extra help for teaching methods. Lin et al. [16] this research investigated the influences of two kinds of network studying approaches, that is video-dependent courses and web practice question working-out, on learners' degree of concentration and study achievement. Therefore, the findings which this research has gotten display effects that are different from each other. More specifically speaking, students that have participated in the video lecture group have exhibited a higher level of attention when compared with those in the online exercise group.

With the advent of the information age, virtual reality (VR) technology has seen increasingly widespread application in education [17]. Schirmer et al. [18] introduced a neurosurgical VR simulator in neurosurgical training. Results indicated the simulator improved trainees' written test scores and demonstrated significant progress in understanding relevant anatomical structures and reducing procedure completion times during simulated experiments. Chen et al.[19] evaluated the effectiveness of virtual reality in fox education. Results indicated that VR instruction was more effective than comparison methods in enhancing knowledge acquisition. However, it did not demonstrate greater advantages in skills, satisfaction, confidence, or operational time.

The above practical cases concretely illustrate the application and effects of modern educational technology in teaching. It is believed that with the continuous development and refinement of modern educational technology, it will play an increasingly vital role in education in the future.

SPOC represents a blended learning model integrating traditional and online education, effectively leveraging high-quality MOOC resources. Research builds upon this foundation to construct intelligent Chinese language classrooms. Based on this foundation, therefore, the research has the objective of establishing intelligent Chinese language classrooms. In addition, through combining speech data identification with speech strengthening networks, one voiceprint recognition algorithm which takes feature compensation and inherent sound

adaptation as the core is brought forward for speaker identification. The experiment design gives explanation to the algorithm's basic work parameters and the choice of the feature group. In the end, one teaching experiment has been conducted by us. By using the pre-test and post-test methods, together with the improved Flanders Interactive Analysis System (IFLAS), the digitalized analysis has been done by us on the interaction of language and behavior between teachers and students that is recorded in classroom videos. This research has explored the actual meanings that come from the complete merging of modern educational technology into the teaching of Chinese language in the classroom.

2 Building an Interactive Teaching Model for Chinese Language Smart Classrooms

2.1 Curriculum Framework Development

2.1.1 Course Nature and Objectives

To confirm the essence and objectives of a curriculum begins with dividing and drawing its basic viewpoint. As a basic public lesson for new first-year students, the Chinese language class, based on teaching experience, displays the following characteristics in students' language ability: Firstly, their foundational knowledge of the Chinese language is generally sufficient. Secondly, they frequently look on Chinese only as one group of knowledge and skills, and their literary appreciation abilities are comparatively weak.Thirdly, the literary understanding of them is scattered, they do not have a practical knowledge structure. Fourthly, their writing usually remains on the surface and is filled with hollow slogan-type words. It cannot attain the deep emotional level which literature works ought to possess or the reasonable quality which is required by argumentation articles.

The educational objectives of the Chinese language curriculum should need to embody three basic characteristics. At the first step, it is necessary that this content has comprehensiveness. That is to say, it must bring together knowledge, values, and comprehensive expression teaching, emphasizing the combination of knowledge from many dimensions. This helps cultivate students' capability to actively construct knowledge, find out core questions for investigation, strengthen their comprehension of texts, and grasp the quintessence of literature reading. Secondly, actual putting into practice: Students should foster judgment ability, analysis skills, and critical thinking ability to carry out systematic arrangement of the topics that are being discussed. This method facilitates active, self-reflection learning, helps the effective transmission of knowledge, and therefore solves real-world difficulties. Thirdly, on the aspect of advanced cognition: the curriculum’s objectives, beyond simple memorization, are to cultivate the higher-order cognitive abilities. Students ought to grasp to make use of critical thinking and other advanced brain activities, deliberately taking part in goal-directed, self-leading study.

2.1.2 Course Content and Implementation

To leverage the distinct advantages of both MOOC online instruction and traditional classroom teaching, a blended learning model combining MOOC with traditional classroom instruction—referred to as the SPOC model—has been designed. Under this model, beyond the scheduled class hours outlined in the course description, the core content is divided into two main components:

First, the reading of classic works. The primary educational objectives of the classic works reading component are to broaden students' horizons, integrate existing knowledge systems,

enhance and refine students' knowledge structures across textual, literary, and cultural dimensions, strengthen their language proficiency, provide examples of close reading, and improve students' insight, comprehension, and critical thinking regarding texts.

Second, training in verbal expression skills. This component is primarily conducted in physical classrooms, comprising total class hours, accounting for of the total course hours. It primarily employs the flipped classroom teaching model, centered around two major projects: reading report activities and project research activities. Using a task-driven approach, it promotes reading through writing and thinking through writing. Through writing and expression training related to course content, it cultivates students' deep learning abilities.

2.1.3 Course Instruction Evaluation

A sound teaching evaluation system should comprehensively and accurately reflect students' learning levels. The evaluation system based on the blended teaching model for Chinese language integrates multiple assessment methods:

First, MOOC platform assessment. This evaluating work covers the whole scope of the Massive Open Online Course (MOOC), including the quizzes which locate at the end of every teaching unit. The MOOC platform provides data analysis which depends on students' answer responses, hence permitting teachers to rapidly make modification or promotion to their instruction strategies. Furthermore, indicators of formative assessment contain learning progress, click-through rates, and participation in discussion boards.

Secondly, we carry out the evaluation work on blended classroom teaching. In the entity classrooms, when teachers use the flipped classroom method, they provide continuous evaluation and quick response in each stage of a project. This this includes the full whole process, it begins from the making of project project plans, goes through talks, composing, and ends with report reports. The objective is to correct any deviations which happen when the project is being carried out. At the same time, the Large-scale Open Network Curriculum (MOOC) platform lets students upload their homework and project outcomes. This method allows net friend mutual marking, which can, in certain degree, remedy the deficiencies that exist in the teacher's grading.

2.2 Establishing the Basic Framework

2.2.1 Front-End Analysis

Front-end analysis mainly includes a series of execution components, which contain the expected learners, course teaching objectives, study materials, and study platforms, as it is displayed in Figure 1. The first phase is a comprehensive assessment of the objective learners. This requires that we have a grasping of their basic line abilities, whole characteristics, study inclinations, and their proficiency degree in using the college's net teaching platform or the Learning Pass app to carry on independent study.Then, based on the course's teaching objectives—covering knowledge and skills, processes and methods, as well as attitudes and values—we not only establish the overall course objectives but also refine them into modular goals. We meticulously design a fragmented knowledge system suitable for online self-study and reconstruct teaching content tailored for face-to-face interaction and practical exercises between teachers and students in offline settings. By selecting high-quality online teaching platforms that are widely understood and proficiently used by students, we ensure the smooth implementation of SPOC blended learning activities.

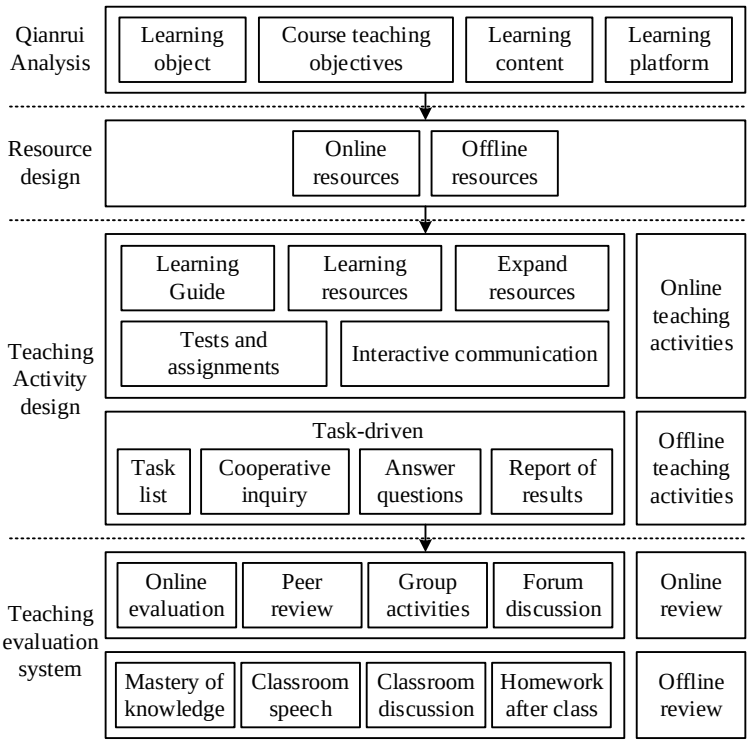


Figure 1: SPOC blended teaching model

2.2.2 Resource Design

Developing a series of SPOC course teaching resource libraries that align with the university's style and possess distinctive characteristics is a crucial step in actively advancing the smooth implementation of blended teaching models. Currently, numerous open MOOC videos, high-quality university courses, and video open lectures can be selected to provide high-quality resources that meet teaching needs for application in SPOC blended learning, achieving resource sharing. Furthermore, course resources independently developed through university-industry collaboration are highly valued. Tailored to specific course characteristics and student learning styles, these resources better align with teaching objectives and more effectively enhance instructional outcomes. Therefore, the SPOC blended learning resource repository for the university's Chinese language courses is constructed by integrating and upgrading resources developed by faculty alongside high-quality shared online materials.

The SPOC mixed study method has the characteristic of having a large-scoped and plentiful collection of teaching materials. This mainly includes short video segments, speech projection pages, question storage banks, discussion forums, and electronic literature resources. The contents of online course are constituted by short video segments, presentation documents, question compilations, question-answer talks, and extra learning materials. These are formulated on the basis of the course’s special characteristics, the situations of the learners, and the actual demands of actual work posts. These core knowledge points and skills are divided by these resources into pieces that can be handled. These are then uploaded to internet platforms, thus it lets learners carry out learning in broken time periods.

2.2.3 Instructional Activity Design

The foundation of offering online lessons is that people must complete the establishment or the selection of a digital study platform. In current time, the university's network platform, which is done co-development together with Chaoxing, makes possible both web-type browsing and

study through the Learning Pass application. Education workers can upload teaching materials to this platform for making standardized courses or utilize the already existing resources on the platform to assist teaching activities. By this method, it is able to fulfill the teaching demands that come from both teachers and students.

The SPOC mixed learning method requires that online and offline parts must be smoothly combined, so that teaching activities can be carried out through joint efforts. Network study mostly relies on students' own-direction study. Because the fact that students' online study gives direct effect to the effect of offline teaching is widely known, hence it has very important meaning to cause students' interest and ensure the successful finishing of online learning. Therefore, the below group of teaching activities has been designed: giving out study guides before class, providing carefully chosen study materials, giving related quizzes and assignments, promoting mutual discussions, and providing extra resources.

2.2.4 Teaching Evaluation System

The appraisal flow combines many different checking methods, including marks from internet examinations, the times of companion evaluations, a target of board discussion participation, and joining in team actions. To each of these measures, we assign different weights, for the calculation of a total online score. Offline academic scores under the line mainly are constituted by the marks which are for the finish of tasks. When students do their tasks, the degree of their participation in the classroom and the degree of their comprehension of online learning ideas are measured. Based on their concrete real behaving situation, thus extra bonus points are added by us to their task result scores.

3 Voice-Interactive Chinese Language Classroom

3.1 Processing of Language Data in Chinese Language Classrooms

3.1.1 Classroom Speech Data Processing

The interactive speech materials that are exchanged between teachers and students inside the classroom are one crucial component of speech materials. It contains explanations from teachers, questions from teachers, answers from teachers, together with answers from students and thinkings by students. These data show the always changing character of classroom mutual actions and contain a large amount of information, such as teaching materials, emotional situations, and mutual action modes. The technology of speech recognition is the main method that gets this data. It carries out the conversion of audio signals into the corresponding character strings of text. In common situations, the speech recognition utilizes the combination that consists of acoustic models and language models. It finishes the transformation from speech to text through the maximizing of the posterior probability.Its mathematical expression is:

$$\hat{W} = \arg\max_{W} P(W \mid O) = \arg\max_{W} P(O \mid W) P(W) \tag{1}$$

Among these, $\hat{W}$ represents the optimal recognized word sequence, $P(W \mid O)$ denotes the acoustic model, and $P(W)$ signifies the language model. Analysis methods for speech data encompass three primary directions: speech sentiment analysis, discourse analysis, and interaction pattern analysis. Speech sentiment analysis identifies the speaker's emotional state by extracting features such as intonation, volume, and speech rate. Sentiment analysis typically employs either traditional feature engineering-based methods or end-to-end deep learning

approaches. The first method makes extraction of acoustic characteristics that are named Mel-frequency cepstral coefficients (MFCC). By opposite way, the second method directly deals with original speech signals through using convolution nerve networks (CNNs) or cycle nerve networks (RNNs). Sentiment analysis possesses the capability to excavate a teacher's emotional diffusion and the degrees of student participation.

In the domain of discourse analysis, artificial intelligence methods make use of natural language processing to identify and give feedback about teaching materials[20]. The text generation methods which are established upon sequence-to-sequence (Seq2Seq) models have the ability to inspect the logical structure and content completeness of explanations that are given by teachers. The objective function which belongs to these approaches is:

$$L(\theta) = -\sum_{t=1}^{T} \log P(y_t \mid y_{<t}, x; \theta) \tag{2}$$

Here, x represents the input sequence, y_t denotes the generated word, and θ signifies the model parameters. Through the carrying out of a discourse analysis, this system has the capability of accurately finding the key important parts inside the teacher's explanatory statements. It then can carry out measurement on how good these aspects match the objectives of the course. At the same time, it can also carry out appraisal on the quality of students' replies.

3.1.2 Classroom Language Enhancement

The promotion of time-domain signals is generally completed by means of a convolutional encoder-decoder framework.

In this framework, an encoder carries out the transformation of the speech signal into one feature vector, and thereafter a decoder does the reconstruction of the speech signal from this vector. The core of this technique is to obtain a mapping function between the starting noisy speech and the handled clean speech, that is:

$$\bar{s}(m) = f_{\alpha}[y(m)] \tag{3}$$

In the equation: $\bar{s}(m)$ represents the enhanced speech signal, $y(m)$ denotes the original noisy signal, and $f_{\alpha}(\cdot)$ is the mapping function.

To accurately obtain this mapping function, the loss between the enhanced signal and the clean signal must be minimized. The enhanced signal $\bar{s}(m)$ is obtained by minimizing the loss function, expressed as:

$$E = \frac{1}{M} \sum_{m=1}^{M} \| \tilde{s}(m) - s(m) \|_2^2 \tag{4}$$

In the equation: E is the loss function, M is the number of training samples, and $s(m)$ is the pure speech signal.

3.2 Speech Recognition Algorithm for Chinese Language Smart Classroom

3.2.1 Speaker Feature Compensation

This paper proposes an EV-FM speaker recognition algorithm based on feature compensation and intrinsic voice adaptation. Utilizing a small amount of unknown speaker data (adaptive data), under the maximum likelihood or maximum a posteriori probability principle to generate speaker factors containing maximum speaker information. It can adjust speaker-independent (SI) models to speaker-dependent (SD) models, achieving good recognition performance with limited training data while eliminating interference from environmental mismatches in the model domain. The main workflow is illustrated in Figure 2.

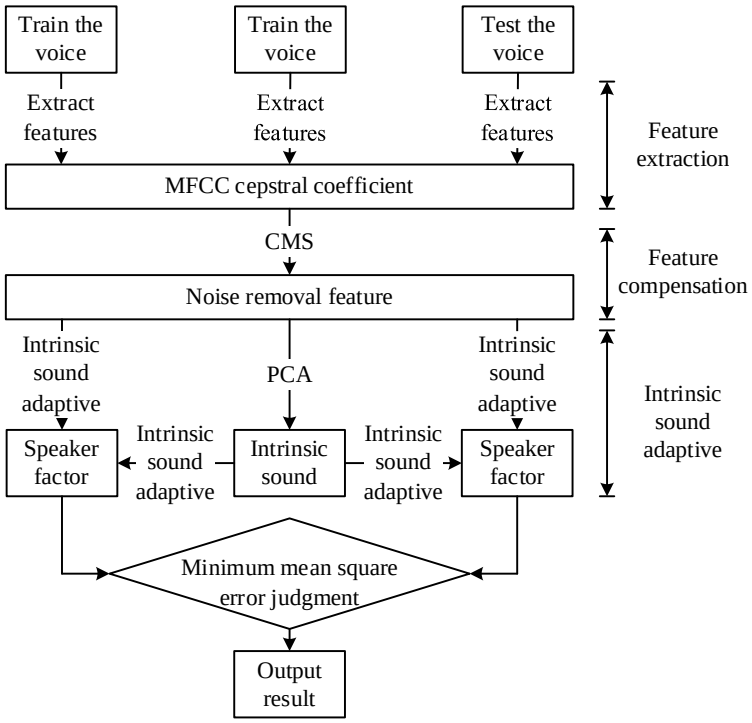


Figure 2: Algorithm flow chart

Feature extraction is the first step in voiceprint recognition. This paper employs Mel-frequency cepstral coefficients (MFCC) as the speech feature parameters. The entire process includes pre-emphasis, segmentation, endpoint detection, windowing, Fourier transform, and triangular bandpass filtering. The spectral features are calculated using Equation (5).

$$B(f) = 1125\ln(1 + f / 700) \tag{5}$$

The CMS algorithm is based on the acoustic model assumption of the vocal tract, abstracting vocalization behavior as a nonlinear process where the excitation signal convolves with the vocal tract model and vocal tract function. It consists of three parts:

The first part, $DFT*[\cdot]$, converts the convolved signal into an additive signal;

The second part performs linear processing on $\hat{s}(n)$ to yield $\hat{y}(n)$; The third part, $DFT^{-1}*[\cdot]$, performs an inverse transform on $\hat{y}(n) = \hat{y}_1(n) + \hat{y}_2(n)$, yielding the output $y(n)$ as the spectroacoustic features of speech.

Assuming the signal $y(n)$, performing short-time analysis on the signal using frame extraction and calculating its cepstral coefficients yields T cepstral coefficients:

$$Y = \{y_1, y_2, \cdots, y_T\} \tag{6}$$

The mean values of these inverted spectrum features are as follows:

$$\overline{y} = \frac{1}{T}\sum_{t=1}^{T} y_t \tag{7}$$

Each y_t vector is subtracted by the deficit to obtain the normalized inverse spectrum vector $\hat{y}_t$:

$$\hat{y}_t = y_t - \overline{y} \tag{8}$$

For a known speech spectrogram sequence, divide its spectrum into background frames and speech frames. Utilize the maximum energy value $E_{\max}$ within the sequence as the basis for generating a discrimination function, defined as:

$$Dis(t) = \begin{cases} 1, & E_t < a \cdot E_{\max} \\ 0, & \text{others} \end{cases} \quad t = 1, \cdots, T \tag{9}$$

Let $Y = \{y_1, y_2, ..., y_T\}$ is the cepstrum sequence of a speech frame, and C_e and C_b denote the cepstrum averages of the background frame and speech frame respectively, then the cepstrum mean subtraction (CMS) can be expressed as:

$$\hat{y}_t = \begin{cases} y_t - C_c, & \text{Dis}(t) = 1 \\ y_t - C_b, & \text{others} \end{cases} \tag{10}$$

3.2.2 Eigenfrequency Adaptation

The principle of eigenphones assumes that all speaker parameter models can be mapped into a low-dimensional space. By applying principal component analysis, a sequence of eigenvalues is obtained from largest to smallest. The top k largest eigenvalues are then selected, and these represent the eigenphones. Within the feature space generated by eigenphones, the maximum likelihood criterion is used to estimate each speaker's coordinate coefficients. Since eigenphone basis vectors exhibit the maximum variance of the speaker model, they reflect the prior probability of speaker model parameters, thereby enabling speaker adaptation. The eigenphone algorithm requires estimating fewer feature parameters, making it particularly suitable for fast speaker recognition with limited speech data.

Assume the speaker-independent Gaussian mean vector c is μ_c with covariance $\sum c$, and the speaker-dependent Gaussian mean vector c for Speaker S is μ_c^S. A discrimination function value of 0 indicates glottal features fall within the central interval. Define the Gaussian mean vector s for speaker-dependent speakers as:

$$\mu(s) = \left[\mu_1^T(s), \right] \mu_2^T(s), \cdots, \mu_C^T(s) \tag{11}$$

The dimension of the above equation is $D*C$. The speaker vector can be defined as $M=\{\mu(s), s=1,2,\cdots,S\}$. Assuming all μ fall within the same subspace, performing principal component analysis (PCA) [21] on M yields S basis vectors, denoted as $e(1\cdots k\cdots s)$, where $e(k)$ represents the k th eigenvoice.

For a speaker-dependent vector $\mu(s')$, it can be expressed as follows:

$$\mu(s')=\bar{\mu}+x_1(s')e(1)+x_2(s')e(2)+\cdots+x_K(s')e(K) \tag{12}$$

Here, μ represents the mean vector of the training speaker, and $x(s')$ denotes the coordinate coefficient corresponding to the K th eigenvoice. The speaker adaptation process essentially involves obtaining the coordinates of the speaker vector $\mu(s')$ in the K -dimensional speaker interference space. These coordinates are typically referred to as speaker factors and can be solved using the maximum likelihood criterion and maximum expectation algorithm. The adaptation process is equivalent to solving an optimization problem. Assuming the adaptation data is $O=\{o_1,o_2,\cdots,o_T\}$, the formula for solving the coefficient speaker factor $x(s')$ is as follows:

$$\begin{aligned} x(s') &= \arg\max_x \sum_{t=1}^{T}\sum_{n=1}^{N}\lambda_n(t)\log p(o(t)\mid\mu_n(s)) \\ &= \arg\max_x \{-\frac{1}{2}\sum_{t=1}^{T}\sum_{n=1}^{N}\lambda_n(t)[o(t)-\mu_n(s)]^T \\ &\cdot\sum_{n}^{-1}[o(t)-\mu_n(s)]\} \end{aligned} \tag{13}$$

where $\lambda_n(t)$ is the posterior probability that the N th eigenvector belongs to the n th Gaussian component in the SI model. Differentiating the objective function above with respect to x and setting the derivative to zero yields the maximum likelihood estimate of the speaker vector:

$$\hat{x}(s')=\left\{\sum_{n=1}^{N}\left[\sum_{t=1}^{T}\lambda_n(t)\right]E_n^T\sum_{n}^{-1}E_n\right\}^{-1}\cdot\sum_{n=1}^{N}E_t^T\sum_{n}^{-1}\left\{\sum_{t=1}^{T}\lambda_n(t)[o_t-\bar{\mu}_m]\right\} \tag{14}$$

The above equation represents the eigenestimation of the speaker factor without channel mismatch.

3.3 Classroom Experiments and Application Analysis

3.3.1 Training Data Collection and Labeling

This study utilized two data sources: one derived from surveillance footage of classrooms in a Shenzhen primary and secondary school, and the other from publicly available online resources, primarily audio-sharing platforms like Himalaya and video platforms like Sohu Video. One batch of high-grade classroom teaching examples which have already been published and spread were downloaded. Audio data was extracted out from video files through using FFmpeg, which is an open-source software bag that is made for recording, transforming, and flowing digital audio and video. This software provides an all-sided method with respect to catching,

changing, and sending audio and video materials. The extraction parameters were all set in the uniform way as below: single channel (mono), the sampling rate of 16000Hz, and 16-bit WAV file format.

The course audio was categorized. During the actual classification process, five main categories covered most audio scenarios. For the few scenarios not included, an extended set of classification criteria was defined, as shown in Table 1.

Table 1: Criteria extension for Scene Classification

Scenario category	Key words sorting	Scenario note
Read with music in the background	M	Students read or sing with background music
Pure recording playback	P	The playback of recorded audio is common in Chinese and English classes
As recorded in the audio	F	The recording is played and the students read along
Clap one's hands	C	The students clapped together
Cough	CO	The teacher or the student coughed loudly
Table and chairs moved	DC	The sound of desks and chairs moving against the ground
Blackboard writing	B	The sound of chalk writing on the blackboard when a teacher is writing

3.3.2 Experimental Design

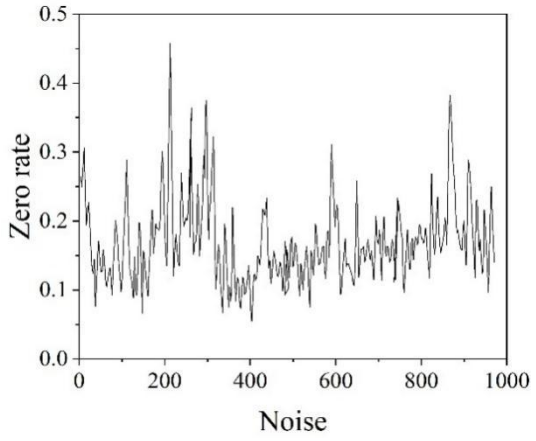
This paper's experiments are primarily divided into two categories: the first is Voice Activity Detection (VAD), and the second is Speaker-Scene Recognition. The training and validation sets are composed of a mixture of classroom speech and internet audio data. The speech portion consists of a blend of four data types: S, T, R, and D, totaling 250M in size. These were randomly allocated to the training and test sets at a 3:1 ratio.

For the speaker-scenario recognition experiments, single-layer SVM, multi-layer SVM, and GM methods were employed to investigate scenario classification. For the EV-FM and i-vector models, the constructed training and validation sets were randomly allocated at a 3:1 ratio, while the test set utilized actual classroom audio data (processed through VAD).

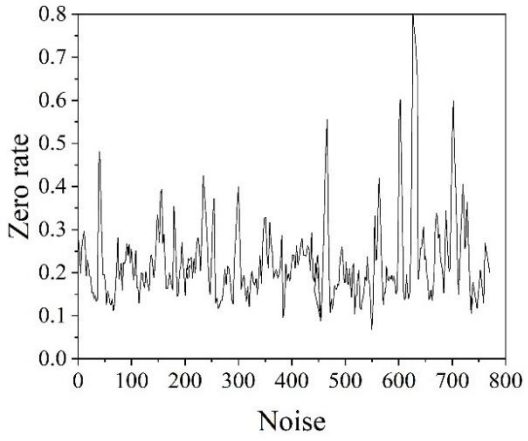
3.3.3 Test Results from Authentic Classroom Corpus

1. Real-world VAD Results

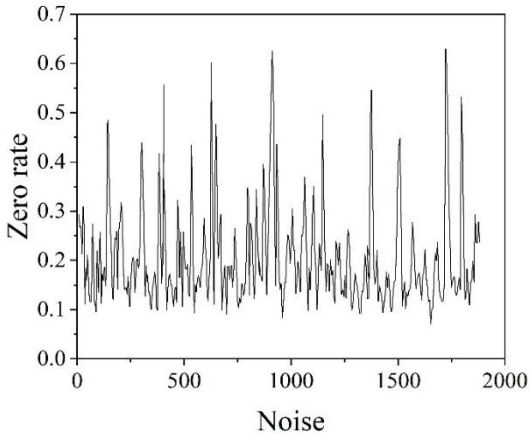
Figure 3 gives the zero-crossing frequency values of N, T, D speech segments. In the real classroom Voice Activity Detection (VAD) experiments, the two-stage method brought about not-good-enough outcomes. Many researches have shown that too much environmental noise will cause increased zero-crossing frequencies in the time when no sound exists, which therefore causes that it is very difficult to make a distinction between the silence and the speech. Under calm environments, the zero-crossing frequency values generally lie inside the interval of 0.1 to 0.5, hence there exist obvious peak values. These changing tendencies did not present any difference that can be clearly separated from the speech of teachers or the group reading aloud of texts.



(a) loudspeaker mute



(b) Teachers speak



(c) Group reading

Figure 3: Three types of scene speech fragments over zero rate

One method which depends on models has been selected for the detection of speech activity. We have utilized the EV-FM model for constructing separate models under speaking and silent environments. The results of our experiments are showed in Table 2. When we make contrast among different course kinds, it is very clear that the recognition effect for already published courses is much better than that of courses which are recorded. The recording courses have extremely low speech remembering rate and silent correctness, but they have high speech correctness and silent remembering rate, this situation shows a quite obvious unbalance. This point thus indicates that the speech loudness in recorded courses is smaller when compared with published courses. This also points out that there exists a difference between the training data and real classroom circumstances, therefore it leads to a deviation in classification.

Nevertheless, among already printed courses, speech accuracy is generally similar to recall, hence only the recall of quiet time intervals exceeds accuracy. This kind of phenomenon can be partly attributed to that in the data annotation work, short stop between sentences had not been comprehensively marked as silence. Therefore, the number of the predicted silent intervals is larger than the number of the annotated silent intervals. This is displayed in the results as a little decreased correctness for quiet intervals and a little lowered recollection for voice, which is accordance with actual world observations.

Table 2: EV-FM test results for VAD classification in classroom

curriculum	Speech accuracy rate	Voice recall rate	Silent accuracy	Silent recall rate	F1 value
Publishing Chinese-05	0.96	0.84	0.57	0.95	0.87
Publishing Chinese-25	0.92	0.97	0.77	0.67	0.91
Publishing Chinese-52	0.99	0.74	0.51	0.98	0.77
Record fourth grade English lessons	1.00	0.58	0.52	1.00	0.64
Record a Chinese lesson for grade 7	1.00	0.43	0.35	1.00	0.53
Record the class of second grade mathematics	1.00	0.41	0.56	1.00	0.67
Record the second grade physics class	1.00	0.47	0.56	1.00	0.60

2. Real-world Environment Scene Classification Results

During the initial speech activity detection phase, recorded courses suffered from poor audio quality and subpar recording standards, resulting in an average F1 score of only around 60%. This significantly impaired subsequent scene classification performance. In contrast, published courses achieved an overall F1 score 20 percentage points higher than recorded courses, enabling relatively accurate speech activity differentiation. Therefore, all experiments in this section utilize published courses.

By comparing model accuracies during validation, we propose a multi-layer differential model for scene recognition: First, apply a GMM model to classify classroom speech into group speech (D, R) or single-person speech (T, S); If classified as T or S, the EV-FM model further distinguishes between teacher (T) and student (S) speech. If classified as D or R, the EV-FM model further distinguishes between background discussion (D) and group discussion (R), as shown in Table 3.

Experimental results indicate that validation on test corpora achieved relatively satisfactory outcomes. However, recognition rates declined significantly in real-world scenario data, with particularly sharp drops in collected corpora and slightly better performance in internet-based corpora.

Table 3: Different classroom environment classification

GMM test results for D,R, T,S classification in Classroom					
Curriculum	D,R precision	D,R recall	T,S precision	T,S recall	F1 value
Publishing Chinese-05	0.38	0.97	1.00	0.63	0.74
Publishing Chinese-25	0.51	0.76	0.86	0.67	0.71
Publishing Chinese-52	0.62	0.97	0.96	0.58	0.77
I-vector test results for T, S classification in Classroom					
Curriculum	T precision	T recall	S precision	S recall	F1 value
Publishing Chinese-05	0.75	0.91	0.79	0.67	0.75
Publishing Chinese-25	0.47	0.83	0.92	0.73	0.79
Publishing Chinese-52	0.85	0.69	0.65	0.90	0.72
EV-FM test results for D, R classification in Classroom					
Curriculum	T precision	T recall	S precision	S recall	F1 value
Publishing Chinese-05	0.66	0.96	0.96	0.68	0.79
Publishing Chinese-25	0.15	0.86	0.99	0.67	0.77
Publishing Chinese-52	0.95	0.74	0.65	0.89	0.84

4 The Effectiveness and Impact of Comprehensive Chinese Language Learning Practices

4.1 Experimental Design and Data Collection

4.1.1 Experimental Design

To test whether the EV-FM-based SPOC blended teaching model for Chinese language smart classrooms can enhance students' deep learning abilities and academic performance, this experiment employs a pre-post test design for the experimental group and a pre-mid-post test design for the academic performance assessment across both experimental and control groups. The subjects are seventh-grade students from a Shenzhen middle school, with Class L (experimental group) comprising 52 students and Class Y (control group) comprising 53 students. In the whole process of Chinese language study, the experiment class used a high-level mutual activity teaching mode. By comparison, the contrast class used the old-type teaching approach.

4.1.2 Data Collection

Data collection during the model application process primarily drew from questionnaires, classroom videos, and exam scores. The specific timing, content, and analysis methods are detailed in Table 4.

Table 4: Data collection and analysis

Data content	Collection methods	Collect time	Analytic procedure
Deep learning capabilities	Scale	2025.5.7	Pairs sample t-test
First examination	Harmonized examinations	2025.5.15	Independent samples t-test
Classroom teaching interaction	Classroom video	2025.5.26	IFLAS interactive analysis system
Classroom teaching interaction	Classroom video	2025.6.6	IFLAS interactive analysis system
Second examination	Harmonized examinations	2025.6.11	Independent samples t-test
Classroom teaching interaction	Classroom video	2025.6.17	IFLAS interactive analysis system
Deep learning capabilities	Scale	2025.6.23	Independent samples t-test
Third examination	Harmonized examinations	2025.6.8	Pairs sample t-test

Based on the characteristics of comprehensive learning in junior high school Chinese language education using the flipped classroom approach, the 16 codes are further explained as shown in Table 5.

Table 5: Coding interpretation of the improved Flanders interactive analysis system

Code	Explain
1	Teachers accept the emotions or opinions expressed in students 'movements, expressions and words, such as listening patiently to students' speeches
2	Teachers 'reinforcement behaviors for students' excellent performance, such as "very good" and "right"
3	Teachers affirm students 'opinions, such as using students' opinions to explore deeply in class
4.1	The teacher's question gives the students room to play freely
4.2	Teachers ask questions that limit student responses or vote on a decision
5	Teachers impart knowledge to students or read out students' work through spoken language
6	The teacher asks the students to perform a specific action or language in class
7	Teachers educate students about certain undesirable behaviors, such as maintaining classroom discipline
8	Students passively answer questions raised by teachers
9.1	Students actively answer or respond to teachers and classmates
9.2	Students host activities, take the initiative to ask questions of others, express their own opinions or report their own results
10	Students discuss learning-related topics with peers
11	There is confusion and noise in the classroom that is not conducive to learning, such as students discussing with their peers about things that are not related to learning
13	Teachers use multimedia or other teaching tools
14	Students use multimedia and other teaching tools or use videos to present

4.2 Experimental Data Analysis

4.2.1 Impact on Students' Deep Learning Competency Levels

A paired samples t-test was conducted on students' two-time depth learning ability scales, with results shown in Tables 6–8. The pretest mean score for depth learning motivation (DJ) was 3.137, while the posttest mean score was 3.258, yielding a correlation coefficient of 0.359. The paired samples t-test value was -1.214, with a significance level of 0.236. The p-value exceeded 0.05, indicating non-significant results. The pretest mean score for deep learning outcomes (JG) was 2.957, while the posttest mean score was 3.016, yielding a correlation coefficient of 0.692. The paired samples t-test value was -0.992, with a significance level of 0.334. The p-value exceeded 0.05, indicating non-significant results. The above data indicates that the EV-FM-based SPOC blended teaching model for Chinese language smart classrooms positively influences students' deep learning motivation levels and deep learning outcomes, but the effect is not statistically significant.

The correlation coefficient for deep learning engagement (TR) was 0.577, with significance at 0.000 ($p<0.01$), indicating extremely significant results. The correlation coefficient for deep learning strategy (CL) was 0.784, with significance at 0.018 ($p<0.05$), indicating significant results.

The above data indicates: First, the EV-FM-based Chinese Smart Classroom SPOC blended teaching model exerts a positive influence on students' deep learning motivation levels and deep learning outcomes, though this influence is not statistically significant. Second, the EV-FM-based Chinese Smart Classroom SPOC blended teaching model significantly enhances students' deep learning engagement and strategy levels, demonstrating a positive effect.

Table 6: Paired samples statistics

		Mean	N	SD
Group 1	DJ1	3.137	50	0.095
	DJ2	3.258	50	0.063
Group 2	TR1	2.714	50	0.082
	TR2	3.217	50	0.067
Group 3	CL1	2.968	50	0.079
	CL2	3.087	50	0.064
Group 4	JG1	2.957	50	0.060
	JG2	3.016	50	0.088

Table 7: Correlation coefficient of paired samples

		N	Suez Canal	Sig.
Group 1	DJ1&DJ2	50	0.359	0.013
Group 2	TR1&TR2	50	0.577	0.000
Group 3	CL1&CL2	50	0.784	0.000
Group 4	JG1&JG2	50	0.692	0.000

Table 8: Sample test for paired samples

		t	df	Sig.(double-tailed)
Group 1	DJ1&DJ2	-1.214	50	0.236
Group 2	TR1&TR2	-7.512	50	0.000**
Group 3	CL1&CL2	-2.573	50	0.018*
Group 4	JG1&JG2	-0.992	50	0.334

*P<0.05**P<0.01

4.2.2 Effects on the Level of Classroom Speech Interaction Between Teachers and Students

Through the iFIAS analysis system, three coding statistical tables of classroom interaction behaviors were generated for the EV-FM-based SPOC blended teaching model in Chinese language smart classrooms. The results are shown in Table 9. The findings are as follows:

(1) Regarding the proportion of teacher versus student speech: - Case 2 exhibited a higher teacher speech ratio than Cases 1 and 3. - Student speech ratios were 43.99% for Case 1, 52.09% for Case 2, and 85.88% for Case 3. - Student speech proportion gradually increased from Case 1 to Case 3, while teacher speech showed an initial rise followed by a decline.

(2) Regarding the ratio of indirect versus direct influence in teacher language: Case 1 showed 19.94%, Case 2 31.16%, and Case 3 56.45%. Similarly, the ratio of positive reinforcement to negative reinforcement increased progressively from Case 1 to Case 3.

(3) The proportion of questions in teacher language was 0.87% in Case 1, 2.24% in Case 2, and 3.09% in Case 3. The proportion of teacher questions remained relatively low across all three cases.

(4) Regarding student-initiated responses, both the proportion of student-initiated responses within total student speech and the proportion of responses within total student contributions were high across all three cases, with Case 3 slightly exceeding Cases 1 and 2. The proportion of student-initiated questions within total student speech was 85.23% in Case 1, 0.79% in Case 2, and 81.38% in Case 3, with Case 2's proportion significantly lower than Cases 1 and 3.

Table 9: Statistical table of the coding of interactive behavior in classroom teaching of case class

Statistical items		Case 1	Case 2	Case 3
Language ratio	Teachers	23.07%	46.17%	14.67%
	Students	43.99%	52.09%	85.88%
What teachers say to students	Ratio of indirect to direct influence	19.94%	31.16%	564.51%
	Positive reinforcement versus negative reinforcement	68.94%	324.17%	1746%
Information technology applications	Teacher manipulation techniques comparison	27.1%	30.26%	0%
	Student manipulation techniques comparison	73.44%	70.68%	100%
Students respond proactively	Compare students' active speaking	15.31%	14.76%	19.34%
	Student response ratio	85.23%	0.79%	81.38%
The proportion of students who take the initiative to ask questions compared with those who take the initiative to speak		100%	93.27%	100%
Beneficial teaching silence ratio		0.71%	0%	0.62%
Technology application comparison		31.16%	3.18%	2.93%
The proportion of questions in teachers' language		0.87%	2.24%	3.09%
Student language students actively speak more than		95.98%	71.47%	94.93%
The proportion of open questions in teachers' questions		100	86.18%	100%

4.2.3 Impact on Academic Achievement in Language Arts

To investigate the impact of the EV-FM-based SPOC blended teaching model on Chinese language academic performance, an independent samples t-test was conducted on the three examination scores of the experimental and control classes. The results are presented in Table 10. The mean scores for Exam 1 in the experimental and control classes were 74.183 ± 14.783 and 74.207 ± 14.216, respectively. Levene's test for homogeneity of variances did not reach statistical significance ($F=0.147$, $P=0.992>0.05$). This indicates no statistically significant difference in Exam 1 scores between the experimental and control classes.

The Levene test for equality of variances in Exam 2 for the experimental class did not reach significance ($F=1.328$, $P=0.843>0.05$). The Levene test for equality of variances in Exam 3 for the experimental class did not reach significance ($F=0.370$, $P=0.275>0.05$).

The results indicate that the average score in the EV-FM-based SPOC blended teaching model for Chinese language smart classrooms was slightly higher than that of the control class. However, overall, it did not have a statistically significant impact on students' Chinese language academic performance.

Table 10: Independent sample t-test of experimental class and control class

Examination	Class	N	Mean	SD	F	P
1	Experimental class	52	74.183	14.783	0.147	0.992
	Control class	53	74.207	14.216		
2	Experimental class	52	75.612	12.465	1.328	0.843
	Control class	53	74.825	14.824		
3	Experimental class	52	79.074	12.577	0.379	0.275
	Control class	53	76.549	11.368		

5 Conclusion

This study proposes a SPOC-based blended teaching model for Chinese language instruction, utilizing the EV-FM algorithm to design and implement a speech analysis system for Chinese classroom teaching. Subsequently, the model underwent training, validation, and testing using data, with comparative analysis of classification performance across different scenarios. The EV-FM algorithm demonstrated relatively satisfactory validation results for the test corpus. When applied to the SPOC-based Chinese language blended teaching model, teaching effectiveness was evaluated using a combination of qualitative and quantitative methods. The results revealed that this teaching model significantly influenced both the level of classroom interaction between teachers and students and the level of students' deep learning, with a significance level of 0.000 and $p<0.01$. Furthermore, the EV-FM-based Chinese language smart classroom SPOC blended teaching model demonstrated a certain impact on improving Chinese course grades.

About the Author

Xin Qi, graduated from the Ningxia University in 2015. Currently working at School of Chinese Language and Literature, Jiaozuo Normal College. Her research interests include Chinese language teaching.

References

[1] Moshinski, V., Pozniakovska, N., Mikluha, O., & Voitko, M. (2021). Modern education technologies: 21st century trends and challenges. In SHS Web of Conferences (Vol. 104, p. 03009). EDP Sciences.

[2] Gruzdeva, M. L., Vaganova, O. I., Kaznacheeva, S. N., Bystrova, N. V., & Chanchina, A. V. (2019). Modern educational technologies in professional education. In Growth poles of the global economy: Emergence, changes and future perspectives (pp. 1097-1103). Cham: Springer International Publishing.

[3] Abdulrahaman, M. D., Faruk, N., Oloyede, A. A., Surajudeen-Bakinde, N. T., Olawoyin, L. A., Mejabi, O. V., ... & Azeez, A. L. (2020). Multimedia tools in the teaching and learning processes: A systematic review. Heliyon, 6(11).

[4] Riza, L. S., Hasanah, L. N., Putri, A. H., Budiman, B., Safitri, F., Putri, L. A., ... & Samah, K. A. F. A. (2023). Educational technology using multimedia in science learning: A systematic review. Bulletin of Social Informatics Theory and Application, 7(2), 163-181.

[5] Liu, Z. Y., Lomovtseva, N., & Korobeynikova, E. (2020). Online learning platforms: Reconstructing modern higher education. International Journal of Emerging Technologies in Learning (iJET), 15(13), 4-21.

[6] Mashau, P., & Nyawo, J. C. (2021). The use of an online learning platform: A step towards e-learning. South African Journal of Higher Education, 35(2), 123-143.

[7] Wu, Q., Wang, Y., Lu, L., Chen, Y., Long, H., & Wang, J. (2022). Virtual simulation in undergraduate medical education: a scoping review of recent practice. Frontiers in

medicine, 9, 855403.

[8] Cant, R., Cooper, S., & Ryan, C. (2022). Using virtual simulation to teach evidence-based practice in nursing curricula: A rapid review. Worldviews on Evidence-Based Nursing, 19(5), 415-422.

[9] Makarenya, T. A., Stash, S. V., & Nikashina, P. O. (2020, November). Modern educational technologies in the context of distance learning. In Journal of Physics: Conference Series (Vol. 1691, No. 1, p. 012117). IOP Publishing.

[10] Kapi, A. Y., Osman, N., Ramli, R. Z., & Taib, J. M. (2017). Multimedia education tools for effective teaching and learning. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 9(2-8), 143-146.

[11] Friedl, R., Höppler, H., Ecard, K., Scholz, W., Hannekum, A., Öchsner, W., & Stracke, S. (2006). Multimedia-driven teaching significantly improves students' performance when compared with a print medium. The Annals of thoracic surgery, 81(5), 1760-1766.

[12] Jian-Hua, S. (2012). Explore the effective use of multimedia technology in college physics teaching. Energy Procedia, 17, 1897-1900.

[13] GebreYohannes, H. M., Bhatti, A. H., & Hasan, R. (2016). Impact of multimedia in Teaching Mathematics. International Journal of Mathematics Trends and Technology-IJMTT, 39.

[14] De Sousa, L., Richter, B., & Nel, C. (2017). The effect of multimedia use on the teaching and learning of Social Sciences at tertiary level: a case study. Yesterday and Today, (17), 1-22.

[15] Zhao, K., Yang, Q., & Ma, X. (2017). Exploration of an Open Online Learning Platform Based on Google Cloud Computing. International Journal of Emerging Technologies in Learning, 12(7).

[16] Lin, C. H., Wu, W. H., & Lee, T. N. (2022). Using an online learning platform to show students' achievements and attention in the video lecture and online practice learning environments. Educational Technology & Society, 25(1), 155-165.

[17] Martirosov, S., & Kopecek, P. (2017). Virtual reality and its influence on training and education-literature review. Annals of DAAAM & Proceedings, 28.

[18] Schirmer, C. M., Elder, J. B., Roitberg, B., & Lobel, D. A. (2013). Virtual reality–based simulation training for ventriculostomy: an evidence-based approach. Neurosurgery, 73, S66-S73.

[19] Chen, F. Q., Leng, Y. F., Ge, J. F., Wang, D. W., Li, C., Chen, B., & Sun, Z. L. (2020). Effectiveness of virtual reality in nursing education: meta-analysis. Journal of medical Internet research, 22(9), e18290.

[20] Guoqiang Sun, Yang Zhao & Xiaoyan Qi. (2025). Sequence to sequence architecture based on hybrid LSTM global and local encoders approach for meteorological factors forecasting. Scientific Reports,15(1),22753-22753.

[21] Kaixin Gao, Zheng Hai Huang & Yang Xu. (2025). Tensor Robust Principal Component Analysis Based on a Two-Layer Tucker Rank Minimization Model. SIAM Journal on Imaging Sciences,18(2),1522-1561.