Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org

Volume 26, Issues 4, pp. 1740–1753, DOI 10.12694/scpe.v26i4.4693

APPLYING IMAGE AND VIDEO PROCESSING IN ENGLISH EDUCATION: A TECHNOLOGY-ENHANCED LEARNING FRAMEWORK

LE HAN*

Abstract. Image and video processing in English training is a pioneering technology-enhanced learning strategy which addresses 21st-century student needs. This paradigm's ability to accommodate today's multimedia-driven learners' demands makes it crucial. Such a system requires strong infrastructure, teacher training, scalable computing resources, and adaptive content for diverse learning styles. This research proposes the Smart Multimodal Enhanced Interaction Learning Framework (SMEILF), which takes advantage on multimodal content's strengths. By making learning more interactive, SMEILF intends to boost students' engagement, comprehension, and memory. The current research examines SMEILF, a comprehensive system that uses real-time image and video processing for personalised feedback and adaptive learning routes. SMEILF uses interactive language classes, pronunciation training, and contextual video analysis. Simulation analysis demonstrates the framework works and could increase learning, this research contributes to technology-enhanced learning by offering a scalable, adaptive, and student-centered approach to English training. The proposed method increases the learning engagement ratio by 98.5%, pronunciation accuracy ratio by 97.6%, scalability ratio by 99.2%, content accessibility ratio by 92.9%, and teacher and student satisfaction ratio by 95.8% compared to other existing methods. The proposed method increases the learning engagement ratio by 98.5%, pronunciation accuracy ratio by 97.6%, scalability ratio by 99.2%, content accessibility ratio by 92.9%, and teacher and student satisfaction ratio by 95.8% compared to other existing methods.

Key words: Image, Video, Processing, English, Education, Technology, Enhanced, Learning, Smart, Multimodal, Interaction, Learning, Scalable Computing

1. Introduction. English has traditionally been taught using teacher-led lectures, text-based lessons, textbooks, and chalkboard graphics [1]. These methods promote memorisation and repetition, which can depress students and hinder learning, these methods work in some situations, however they are not flexible enough for today's schools and students' learning styles [2]. Visual learners may find presentations with a lot of text uninteresting, while auditory learners may struggle to understand basic concepts without interactive audio [3]. Traditional teaching approaches are useless because they cannot adapt promptly to student feedback and are inflexible, unadaptable traditional teaching approaches can reduce student involvement [4]. Traditional training may seem antiquated and unconnected to students who are used to multimedia-rich and interactive environments [5]. Students are unlikely to receive the customised attention they need while employing scaled computing resources, which is a severe drawback [6]. Modern students need dynamic, engaging, and responsive learning environments, which traditional schools often lack. These devices can't do complicated tasks like digital image and video processing [7]. SMEILF and other innovative frameworks must comply, these frameworks combine traditional teaching approaches with modern technology to solve these problems. Image and video processing in English training must overcome many obstacles to be effective [8]. Reason being, these devices aren't equipped to handle complex tasks like digital image and video processing. Such a requirement must be met by SMEILF and similar innovative frameworks, these frameworks address these difficulties by integrating conventional teaching methods with current technology [9].

Utilizing real-time image and video processing in English teaching poses several significant obstacles. One problem is that problems with latency and performance could interrupt the natural progression of classes. Low-power processing is necessary for real-time processing because even little delays might diminish student attention, especially in interactive tasks. Constraints on the network, particularly in distant learning settings, are another issue. Because of the high bandwidth requirements of real-time video processing, students and teachers alike may experience frustrating buffering, reduced video quality, or disruptions if their internet connection is slower

*Department of Foreign Languages, Taiyuan University, Taiyuan, Shanxi, 030032, China (helen_2423@126.com)

than expected. As a further concern, the reliability of real-time analysis is lacking. Technologies attempting to decipher human speech, gestures, or facial expressions must possess exceptional accuracy. The quality of the learning experience might suffer if feedback is ineffective due to misinterpretations. Integrating real-time processing into preexisting educational systems may be complex. The goal of achieving software, hardware, and Learning Management System (LMS) compatibility is not always easy to accomplish, which might lead to technical problems. Finally, real-time video captures and processes personal data, which raises privacy problems. The careful management of student privacy in pursuit of educational objectives is essential for maintaining compliance with legal and ethical requirements. The entire potential of real-time processing in the classroom depends on resolving these issues.

Integrating image and video processing into English training must address many major difficulties to maximise its potential. These sectors struggle to manage huge amounts of data in real-time video analysis and multimodal content delivery without a scalable and resilient computing infrastructure [10]. Schools, especially in low-resource nations, may struggle to provide the hardware, software, and network capabilities needed to enable such innovative technology [11]. These technologies require substantial instructor training to use effectively, teachers require technological abilities to run new technologies and integrate them into teaching, this ensures students use classroom tools. Traditional teachers must adopt new methods, which takes time, energy, and resources, this could frighten students because it requires various teaching methods. Another challenge is creating customised content for students' learning styles and needs, creating real-time student-specific information is complex and resource-intensive [12]. With image and video processing technologies, customisation is possible, however it takes time and resources. The recent development of technology that collects and analyses student data, especially visual content like images and videos, raises concerns about data security and privacy. When both components are present in the data, these risks become more obvious, student data privacy, accuracy, and compliance with laws and ethics cannot be stressed. The lack of frameworks and best practices for incorporating these technologies into English training makes their success undetermined. This makes evaluating and replicating excellent educational applications challenging.

Among the potential responses to these challenges could be the implementation of a scalable cloud-based computer infrastructure. This processing-capable infrastructure will be accessible to all schools, irrespective of their financial situation. Teacher preparation programs should place a premium on both technical competence and innovative teaching. Adaptive material development is made easier through partnerships in educational technology and AI-automated customisation. To protect personal information, encryption and regulatory compliance must be put in place without hesitation. To facilitate implementation and generate consistent, measurable outcomes across a range of educational contexts, frameworks and best practices have been developed.

The contribution of the paper are:

1. English language learners can benefit from increased interest in and understanding of course material when teachers use SMEILF to integrate multimodal content into lessons.
2. Flexible and scalable learning pathways can be created and implemented with the help of real-time image and video processing to deliver personalized feedback.
3. Determine how well SMEILF improves learning outcomes by analysing simulations and conducting actual deployments.

The results of the literature review are presented in this section 2. These findings will serve as the foundation for the subsequent investigation. Applying Image and Video Processing in English Education: A Technology-Enhanced Learning Framework. SMEILF is the subject of Section 3, which provides an in-depth examination of the subject matter. The discussion that follows the reporting of the results takes place in Section 4 of the present report. The executive summary and the concluding recommendations are included in Section 5.

2. Literature Survey. The education and learning of languages have been significantly influenced by the technology that has recently been produced. Many different approaches to education and learning have emerged as a result of these advancements.

Shadiev, R., et al. [13] conducted a review and analysis of 398 papers on technology-enhanced language learning (T-ELL), with the goal of finding patterns in publishing, target languages, skills, and technologies. The findings emphasise the importance of English as the primary focus and provide suggestions for the direction of future research. Literature Survey

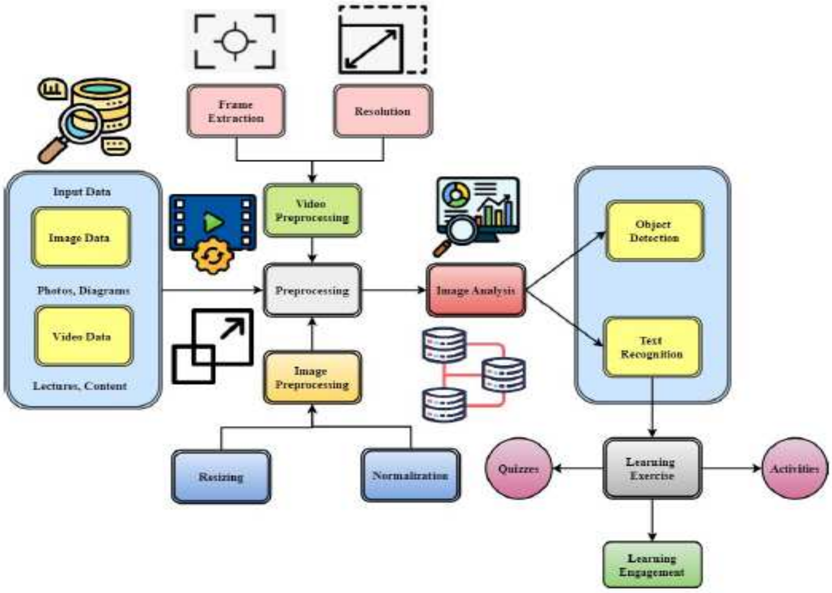

Fig. 3.1: Flow for the Collection and Evaluation of Multimodal Data

Tkachenok, K., et al. [14] examines the use of video-on-demand services (V-DS) such as Netflix for language acquisition, with a particular emphasis on television episodes that serve as educational entertainment. As well as providing strategies for improving second language acquisition, it suggests practical activities for incorporating series into educational programs.

The research conducted by Daniela, L. [15] investigates the incorporation of digital technology in education (DT-E), with a particular focus on significant technological breakthroughs such as microprocessors and the internet. Both the impact that Covid-19 has had on remote learning and the problems that have been associated with regular synchronous online classes are brought to attention. An investigation on the utilisation of technology in Swedish educational institutions is carried out by Schnaider, K., et al., [16] using a multimodal layer architecture and Epistemic Network Analysis (ENA). The findings indicate that the functions and features of technology have diverse effects on the means by which students represent themselves and make signs.

Through the use of a case study of a digital walking tour assessment, Jopp, R. [17] investigates the implementation of Authentic Assessment (AA) in the field of higher education. It has been discovered that AA improves participation, comprehension, and creativity while lowering instances of plagiarism. Multimodal information, real-time analysis, and adjustable learning pathways are some of the reasons that SMEILF is the most effective technique among numerous others. Adopting this strategy leads to learning that is more engaging, more individualised, and increasingly customisable.

3. Smart Multimodal Enhanced Interaction Learning Framework (SMEILF). It is at the forefront of technological advancement as far as teaching English through creative use of sound and computational imaging are concerned. This model is suitable for modern students who value multimedia, interactive classes more than their traditional counterparts. For this purpose, SMEILF employs real-time video and photo analysis to deliver personalized feedback and individualized learning paths. Thus, this technology enhanced approach connects traditional education with modern student's obsession on digital.

Fig.3.1 shows the Flow for the Collection and Evaluation of multi-modal Data, shows a process flow diagram on how data from multimedia sources used in education are organized and evaluated. Data input documents include pictures, schematics, lectures or any other video and image files. Some of the operations carried out during pre-processing include reducing normalizing images, saving frames for videos and changing the level of detail. When preparing videos and images for English language lessons, the system performs different tasks such as object recognition and text identification; speech recognition; environment detection. At this point, the data is transformed from its raw form into an organized form. As a further step, data is examined both contextually and semantically to get insight into the content and draw conclusions. This is useful for understanding the results of the analyses and relates to the aims of education. Lastly, content analysis is used in the enhancement of the learning step to create quizzes and activities that students may participate in to help them remember to learn.

$$\frac{q(n_j|a_{(1:u)}}{1-q(n_j|a_{(1:u)})} = \frac{q(n_j|a_1)}{1-q(n_j|a_1)} \frac{q(n_j|a_{(1:u-1)})}{1-q(n_j|a_{(1:u-1)})} \frac{1-q(n_j)}{q(n_j)} \tag{3.1}$$

The given Equ.3.1 seems to be a recursive updating mechanism q probably derived from a stochastic model in which the probability of an occurrence (n_j) given a series of actions or observes $(a_{(1:u-1)})$ is denoted as$q(n_j|a_{(1:u)})$.

$$M(n_k) = log\frac{q(n_j)}{(1-q(n_j))} * Q(n_2) = \log(2) > 0 \tag{3.2}$$

This Equ.3.2 measures log likely it is that the incident $q(n_j)$ will occur. This may stand log for the system's faith in the learner's comprehension$M(n_k)$ or development inside the SMEILF paradigm $Q(n_2)$.

$$q(n_1, n_2, ..., n_P|a_{1:u}) = \pi_j^P q(n_j|a_{(1:u)}) \tag{3.3}$$

Equ.3.3 shows the combined likelihood of many occurrences $n_1, n_2, ...n_P$ happening, given a set of activities, or observes $a_{1:u}$. Assessing the cumulative probability of a student effectively achieving several learning goals q in the context of their conversations $\pi_j^P q(n_j|a_{1:u})$ is one possible use of the SMEILF framework.

$$q(n_k|b_{1:k}) = (Q(a_{w1}|v_g, a_{1:p-1}) * q(n_k|a_{1:u-1}))/(Q(a_{w1}|, a_{1:p-1})) \tag{3.4}$$

According to the Equ.3.4 $a_{1:u-1}$ and fresh contextual data v_g are both taken into account a_{w1} in a conditional probability update, q. Equ.3.4 might show the SMEILF framework Q takes novel relationships p-1 or inputs into account while updating the student's competency evaluation (n_k).

Fig.3.2 shows the framework for intelligent schools. A smart classroom is an educational space that uses technological principles to improve the quality of instruction, facilitate student-teacher interaction, and reduce physical barriers to knowledge acquisition. It shows a typical setup for a smart classroom, integrated levels, innovative technology, change management, and an all-encompassing atmosphere for learning are all part of a smart classroom for English education, as seen in Figure 3.2. A sensor-enabled smart environment tracks temperature, humidity, air quality, and noise; a touchscreen and projectors present multimedia content; digital appraisal tools allow students to easily interact with one another and with teachers; cameras record and store lectures; and the system is equipped with a smart environment. Smart classrooms include a wide range of academic disciplines. Display equipment, mediums of communication, computer science, networked sensors, recognition of images, and the impact and acceptability of technology are all areas of electronic technologies that have seen a lot of effort. Consequently, a number of review articles have focused on research in these particular areas, with conflicting results. However since these research areas are diverse, it's hard to generalize about smart research in schools or figure out how to integrate different studies to get better results

$$Q(n_q|b_{1:u}) = M(n_j|r_{st}) + M(n_w|w_{2:p-k}) - M(n_k) \tag{3.5}$$

Equ.3.5 lays out the steps to calculate the conditional likelihood $Q(n_q|b_{1:u})$ by combining log-odds components that are associated with various variables or events $M(n_j|r_{st})$. Within the SMEILF paradigm $M(n_w|w_{2:p-k})$, this may represent the way the system uses data like pupil answers, situational circumstances, and previous knowledge to determine the probability of a student attaining a certain outcome of learning $M(n_k)$.

$$M(f_j|a_{2:p}) = min(max(M(n_p * n_p) + H(k_p|a_{l-1}))) \tag{3.6}$$

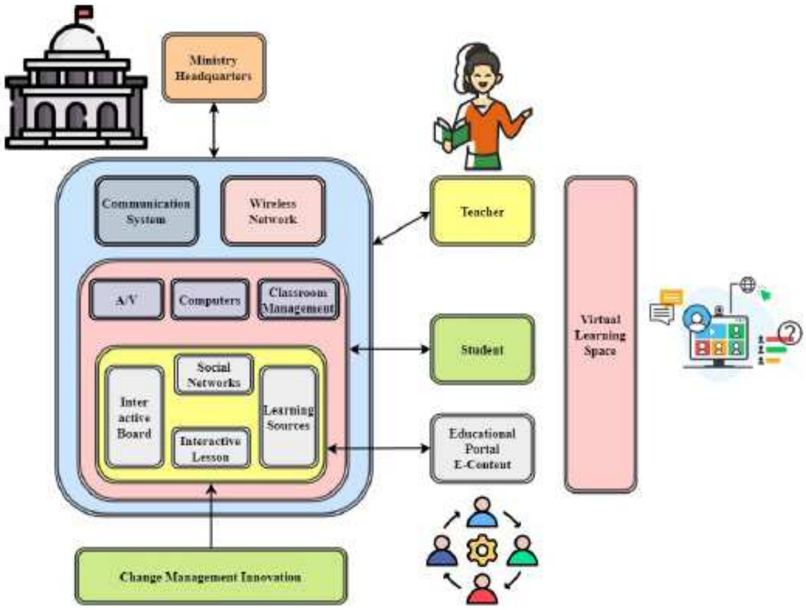

Fig. 3.2: Framework for Intelligent Schools

A function is known as $M(f_j|a_{2:p})$ is used to alter the log of odds $M(n_p * n_p)$ in the process of restricting max or optimizing a value min, as shown by the Equ.3.6. Within the SMEILF framework, this may pertain to enhancing the platform's response or educational approach using the student's previous activities $(k_p|a_{l-1})$ and a computed measure of intricacy or difficulty H.

$$Q(R_j|W_{2:u}) = g(h) * \frac{1}{\forall\sqrt{4}} + f^{\frac{2(s-0)v^2}{4f_2^4}} \tag{3.7}$$

There is a complicated exponential term and an objective g(h) in the Equ.3.7. The SMEILF architecture may use this to depict the computation of a score or likelihood $Q(R_j|W_{2:u})$ that is dependent on several elements, including a baseline value $\frac{1}{\forall\sqrt{4}}$ and a dynamic element affected by parameters like $f^{\frac{2(s-0)v^2}{4f_2^4}}$.

Fig.3.3 shows the smart multimodal enhanced interaction learning framework, which improves student comprehension through multimodal content; these tools aid students in class with the initial stage of the method. Multimodal content development produces real-time image analysis, personalized feedback, and additional information. Real-time video and image assessment allows teachers to track students' progress and provide customized feedback to improve their English learning. Thus, this adaptive education algorithm can analyze films for Context before teaching pronunciation individually, personalizing each learning experience and meeting every unique need based on academic history prerequisite differentiation each student may have possessed. This technology provides dynamic language lessons with rapid feedback and customized instructions to meet their needs. Infrastructure, scalable computing resources, and teacher training are needed to run the system efficiently. An adaptive learning environment designed using this integrated approach will meet each student's requirements.

$$(h * s)(t)) = 1Q(\partial, a - M/2) + 2Q(a - M/2, b + M/2) + 1.5G(a + M/2, s) \tag{3.8}$$

Functions Q and $(h * s)(t)$ seem to be involved in the Equ.3.8 $a - M/2$, with parameters connected to intervals or circumstances. The system's ability b to determine an overall educational approach within the

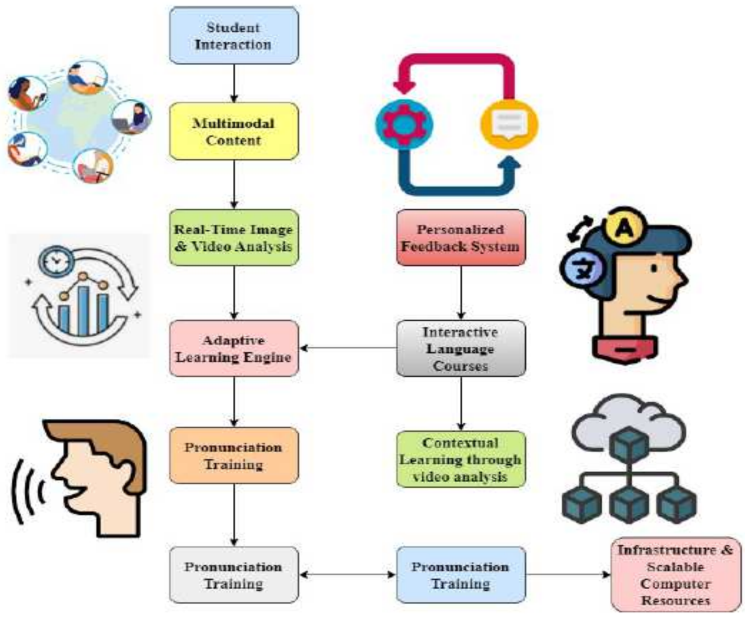

Fig. 3.3: Smart Multimodal Enhanced Interaction Learning Framework

SMEILF framework may be connected to the way it balances various elements of instruction or learning phases, denoted by the features G.

$$Q(r_d, z_s, a_q)U^{p-1}, q_w = \sqrt{(y_d^2 + z_d^2 + a_d^2)} \tag{3.9}$$

Equ.3.9 Q seems to connect the Euclidean mean of a vector to a converted term q_w and a function U^{p-1}. Within the SMEILF the structure may symbolize the way the system incorporates several students q_w data dimensions (such as $\sqrt{y_d^2 + z_d^2 + a_d^2}$ to determine a thorough evaluation of student preparation or competence.

$$\forall_{qw}(1-p) * a_d = \forall a_d * \frac{fg}{e - f_e(n-1)} \tag{3.10}$$

A variable a_d may be changed in equation 10, by adjusting the scaling variables $\frac{fg}{e}$ associated with $\forall_{qw}(1-p)$. This might stand for the way the SMEILF framework's algorithm optimizes education by adjusting educational parameters $f_e(n-1)$ according to scaling variables or chances.

Fig.3.4 shows the journey of a learned individual system. It is believed that adaptive e-learning may stimulate learning and increase student engagement; hence, creating suitable adaptive e-learning environments helps to customize lessons, which in turn reinforces students' learning. Examining how an adaptable e-learning environment affects student engagement is the main goal of this article for English education. The design of the environment will be dependent on students' learning styles. Additionally, this study aims to describe and contrast the suggested adaptive online education setting with a more traditional e-learning strategy. Using a combination of qualitative and quantitative approaches, the following effects were examined for this paper: The study investigation, which is a quasi-experimental design, is carried out using the development approach to construct the adaptive e-learning environment. Emotional, behavioral in nature, skill, and participation/interaction components of student engagement is measured using the student engagement scale in Figure 4. The experimental group outperformed the untreated group by a substantial margin, according to the data. Based on these findings, it seems like an adaptable online classroom may be a great tool for getting children interested

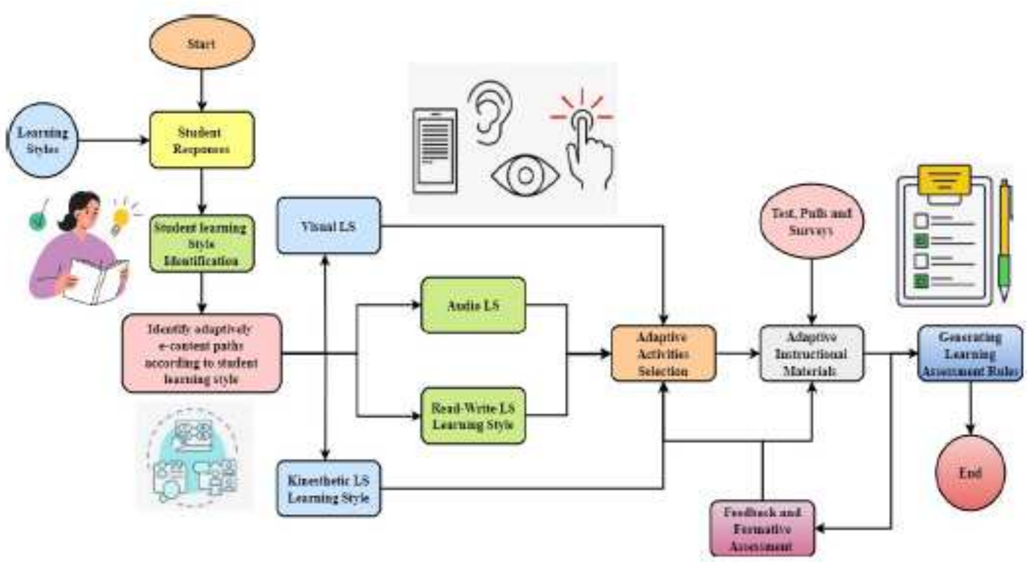

Fig. 3.4: The journey of a learned individual system

in learning. Several actionable suggestions are advanced by this paper: whether to improve the efficiency of education while simultaneously increasing the effect of adaptive e-learning and to build an adaptive e-learning foundation based on the application of different learning styles. Institutions of higher learning that focus on online education might use the findings and the suggested adaptive e-learning strategy to create more engaging and personalized online classrooms for their students.

$$Q(r(s)) = Q_{vrp}(s) + (\frac{l}{\infty w_q \sqrt{2\partial}} + 1.5 - Q_{dpp}(s)) * f^{-1/2}(s - m_q) \tag{3.11}$$

The Equ.3.11. uses $Q_{vrp}(s)$ as an initial term and incorporates an adjusted variable that includes constants and a further term Q(r(s)). The system in SMEILF might compute a customized metric $\frac{1}{\infty w_q \sqrt{2\partial}}$ for educational achievement or feedback $Q_{dpp}(s)$. To make sure that the feedback is correct and relevant to the pupil's present learning state, SMEILF uses words $f^{-1/2}$ and $s - m_q$ to adaptively revise its judgments.

$$Q_{Dff}(s) = Q_{max}, \text{ if } 0 > s > m_q \text{ or } 1.5, \text{ if } s < m_q \tag{3.12}$$

The partwise function that changes depending on the quantity of Q_{max} concerning m_q is the Equ.3.12. Based on when the student's performance $Q_{Dff}(s)$ fits within particular ranges, the system under the SMEILF framework provides certain output levels or thresholds.

$$\log \frac{q(n_k|a_{1:u})}{q(n_{bk}|a_{1:u})} = log \frac{q(n_k|a_u)}{q(n_{bk}|a_u} - log \frac{q(n_k|a_{1:p-1}}{q(n_k)} \tag{3.13}$$

Based on various sets of data, the Equ.3.13 illustrates the way the log-odds b ratios of events $q(n_k|a_{1:u})$ and u change. This equation log may symbolize the process by which SMEILF compares current probability with past ones for improving its evaluation of student achievement or learning results. With the help of recent interaction effects (a_u) and past knowledge $a_{1:p-1}$.

Fig.3.5 shows the framework teaching interactive emotion multimodal recognition. Using emotion recognition, one may assess and characterize the learners' state and adapt teaching tactics accordingly. The multifaceted nature of the surroundings and people deliberate or unconscious attempts to disguise their feelings

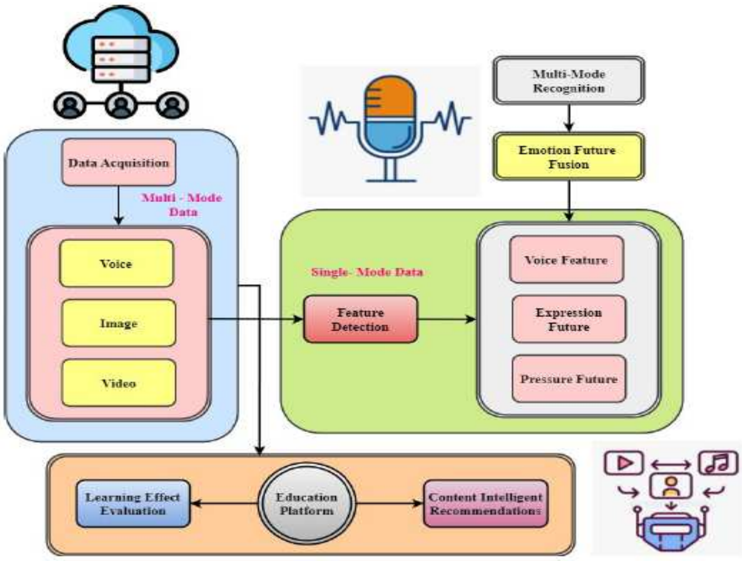

Fig. 3.5: Framework teaching interactive emotion multi-modal Recognition

in various states make it such that identifying one form of communication is frequently insufficient. As a result, this paper suggests a multi-media emotion identification algorithm, as shown in Fig.3.5, to address the limitations of current single-modality algorithms in capturing user feelings to accomplish collaborative learning through recognizing feelings and to address the emotional deficiency in current virtual schooling and other forms of instruction that do not involve interaction and interaction with others for English education. According to a large body of experimental research by cognitive scientists, the following emotional states have an impact on humans' ability to learn: boredom, bewilderment, happiness, frustration, focus, and surprise. Learning cognition is intimately tied to these feelings. For that reason, experts refer to these feelings as cognitive emotional states. The majority of the existing research on recognizing feelings focuses on each of these six core emotions, with very little investigation into the cognitive state of mind. In interpersonal communication, however, the speech-based method is more natural.

$$h(s) = 0,\ \text{if}\ s > a - M/2\ \text{or}\ 1,\ \text{if}\ a > 0.5,\ \text{otherwise for} s < 0 \tag{3.14}$$

The Equ.3.14 value of the outcome of the model might vary depending on the dimension of s. The criticism or teaching response h(s) in the SMEILF structure might change depending on certain student performance criteria a-M/2 for learning engagement analysis.

$$J(uv_0) \geq \frac{e^2}{g(n-1)} + |m^2(n-pk)| - |v_b(1-kp)| \tag{3.15}$$

An equation comprising constants g(n-1) and terms relating to $J(uv_0)$ is used to express a threshold-based precondition where a threshold of e^2 must be met or exceeded $m^2(n-pk)$. This may represent an SMEILF framework criteria $v_b(1-kp)|$ for assessing the appropriateness or efficacy of pedagogical activities for pronunciation accuracy analysis.

$$V = s, Et(g-1) - \frac{bu_{s-1}}{2} + \frac{cT_2(j-1)}{3} - R_{m-n}(pk-1) \tag{3.16}$$

A set of terms using parameters $S, Et(g-1)$, and V calculates the value $\frac{bU_{s-1}}{2}$ in the Equ.3.16. By taking into consideration elements of previous encounters $\frac{cT_2(j-1)}{3}$, instructional modifications R_{m-n}, and performance metrics pk-1, this equation 16 in the SMEILF framework.

$$\frac{e}{fh(m-n)} * r(\partial_1 - \forall_k) = -P[(\partial w_2 - N|(|a|)| * M^{n-pw})] \tag{3.17}$$

This Equ.3.17 delineates a harmony of variables that incorporates constants e, fh(m-n), as well as functions of ∂_1 and $\forall_k$. Taking into account variables such as r, $\partial w_2 - N$, and P, this might pertain to the scaling or leveling of performance measures or feedback changes in SMEILF for the content accessibility analysis.

A function containing a, d_1, and constraints $e_v q(n-1)$ according to the Equ.3.18, which specifies a boundary condition for teacher and student satisfaction analysis. This may stand in for a limit or threshold for a certain performance indicator or metric in the SMEILF system Q*(p), making sure it stays within reasonable limits according to things instruction characteristics $1/4(n-b_1)$ and stimulus corrections T^2.

SMEILF which has made significant strides in customizing English teaching via processing images and videos. SMEILF's interactive classes, voice training and curriculum updates have resulted in high student engagement and understanding. According to simulation findings, this flexible technology can be used in achieving the goals of modern education as it effectively improves performance among learners. This model presents a huge leap in technological supported learning as it allows many options and requirements of students.

4. Results and Discussion. Within the scope of this research, the impacts of the SMEILF on several aspects of English as a Second Language (ESL) instruction are investigated. Among the topics that are being investigated in the present research are scalability, accessibility of material, satisfaction of both instructors and students, learning engagement, correctness and accuracy of pronunciation, and correctness. The usage of image and video processing by SMEILF makes it possible to provide students with an educational experience that is more individualised, flexible, and engaging. A well-designed test bed is essential for evaluating real-time video and image processing systems in the context of teaching the English language. A graphics processing unit (GPU)-enabled server handles processing, while high-definition cameras (1080p or above) record real-time interactions and microphones take in audio input. A dependable internet connection with at least 50 Mbps is required for data transfer to go off without an issue. LMS incorporating video processing frameworks like OpenCV to analyze speech and facial expressions in real time. To measure the system's versatility across demographics, the test bed should have various students, including instructors, for qualitative evaluation. Essential metrics for measuring performance include latency (the time it takes for video to be captured and displayed), accuracy (the system's capacity to understand and respond to spoken language, gestures, and facial expressions), bandwidth usage (the amount of data transferred across a network while processing), and engagement detection (the system's efficacy in detecting when students are actively engaged and learning).

An analysis of student involvement reveals that the utilisation of SMEILF image and video processing in English lessons results in an increase in both the level of engagement and motivation of students, as seen in figure 6 above. The use of multimodal content, such as interactive graphics and real-time video analysis, is becoming increasingly popular among educators as a means of encouraging students to participate more actively and catering to their unique approaches to learning. Students' curiosity increases and their engagement is maintained through the usage of this strategy, which is successful since visually appealing courses encourage students to participate actively. SMEILF participation is increased through the use of tailored feedback mechanisms. All of the needs of the students are quickly and effectively addressed by these solutions produces 98.5%.

This adaptive interaction promotes active participation instead of passive information absorption, making learning more engaging. Involvement and agency on the part of students are fostered by the framework's capacity to adapt learning courses based on real-time data. Learning becomes more engaging, relevant, and personalised to each student with SMEILF's multimodal approach. For the success of English language acquisition and for the involvement of students, this is essential Research on the efficacy of picture and video processing on students' pronunciation during English language lessons has been done using the SMEILF, the results showed that pupils' pronunciation got much better. In the above figure 7, through the use of real-time video analysis and audio feedback, SMEILF demonstrates to students the challenges associated with proper pronunciation. The study of phonetics is a skill that benefits students, to help children pronounce words correctly, visual-auditory activities such as speaking and sound matching might be helpful. These exercises are made possible by the framework, students receive immediate and personalised feedback on these assignments,

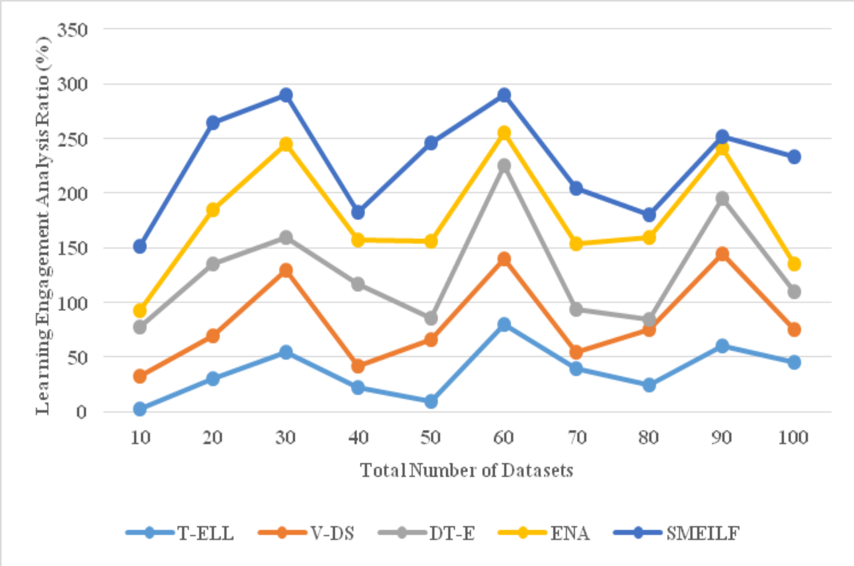

Fig. 4.1: Learning Engagement Analysis

Fig. 4.2: Pronunciation Accuracy Analysis

which aids in problem-solving. By utilising video processing tools to examine facial expressions and speech patterns, students may readily observe proper pronunciation. The customisable nature of SMEILF allows for personalised pronunciation practice that focusses on weak spots for each student. Students gain self-assurance when they receive frequent comments on their work and are assisted in making improvements to their accuracy produces 97.6%. SMEILF is an effective tool for teaching English since it uses image and video processing to train students' pronunciation, which has increased their accuracy.

In Fig.4.3, SMEILF's cloud-based design and variable processing resources allow it to adapt to shifting data

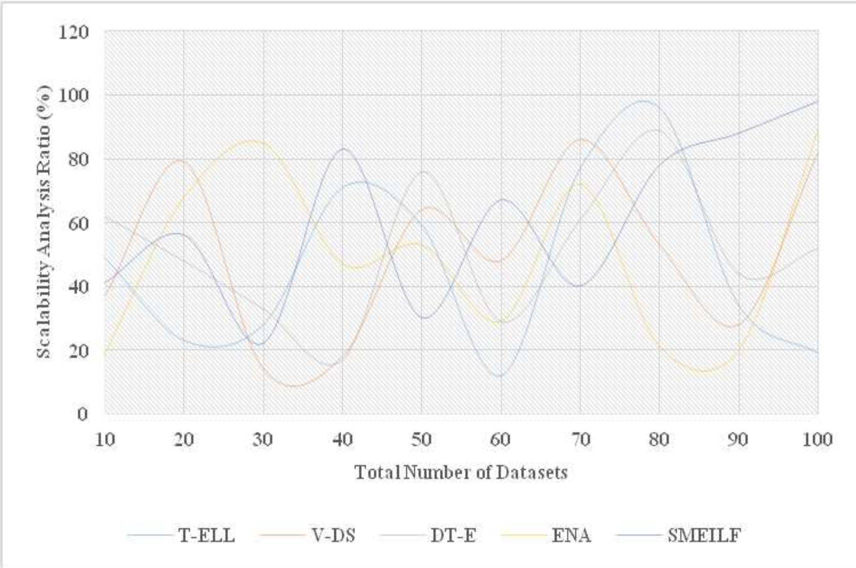

Fig. 4.3: Scalability Analysis

volumes and user demands. It can be utilised in private households and large public schools due to its versatility. Our modular design makes SMEILF easier to adapt and extend to match the needs and resources of individual educational institutions. Because it has fewer hurdles and standardised technologies, the framework is easier to use and interoperable with current systems. Using elastic cloud resources, SMEILF processes images and videos in real time. This opens up opportunities for the organisation to work with universities that have varying levels of technological expertise produces 99.2%. The flexibility of the framework allows for the incorporation of multimedia and region-specific content, which expands its range of educational uses. According to its scalable architecture, SMEILF can grow and change to meet the needs of teachers in the future.

In the above Fig.4.4, with its inclusive design, SMEILF's instructional content is accessible to all students, including those with disabilities. To accommodate visually and hearing-impaired students, adaptive devices are included. These include subtitles and audio descriptions of video; the structure will be easy to understand for pupils. SMEILF's user interface is easy to use and customisable, allowing students to customise their education. Multimodal information provides visuals, audio, and text for diverse learning styles and cognitive and sensory needs. The framework's cloud-based architecture makes data available from any device or platform, this lets students explore the material in different settings. SMEILF promotes an inclusive learning environment to satisfy all students' educational needs and make English language learning more equal and interesting produces 92.9%. Attention must be paid to the accessibility concerns that have been brought up as a means for us to accomplish this goal.

In the above Fig.4.5, through the use of SMEILF's speed and multimedia material, teachers can improve their teaching and reduce boredom. Real-time feedback and adaptive learning allow teachers to customise classes, teachers can better manage their classrooms and repeat subjects less. In particular, educators value the extensive training and support given to integrate SMEILF into their work, it gives teachers confidence and skill in using new technology. SMEILF's interactive learning environment boosts student satisfaction, when real-time image and video processing creates an immersive and participative environment, learning becomes enjoyable and less tedious. Personal feedback and adaptive content design for different learning styles may inspire and increase student learning. Multimedia works for all learning styles, making the classroom more individualised and inclusive produces 95.8%. Generally, SMEILF improves education and promotes a more engaging and supportive learning environment by increasing teacher and student satisfaction.

With its completely scalable and inclusive solutions, SMEILF improves student engagement, pronunciation

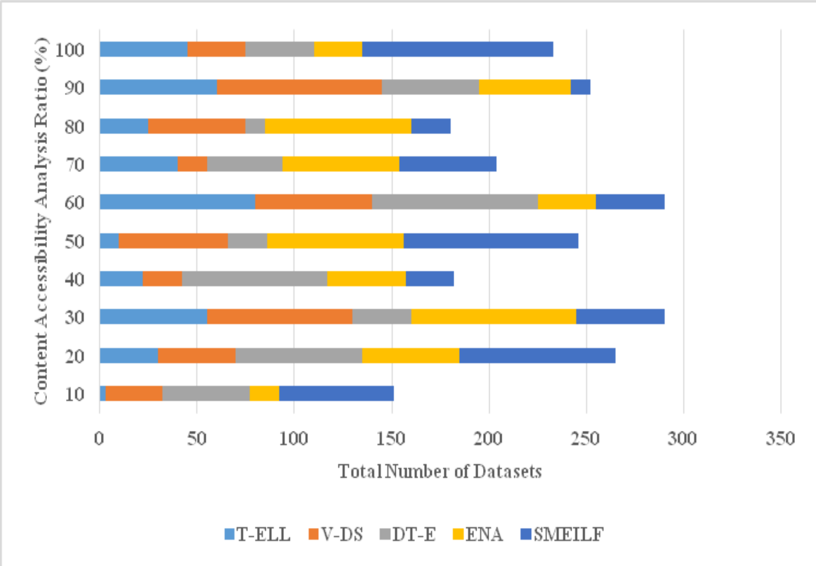

Fig. 4.4: Content Accessibility Analysis

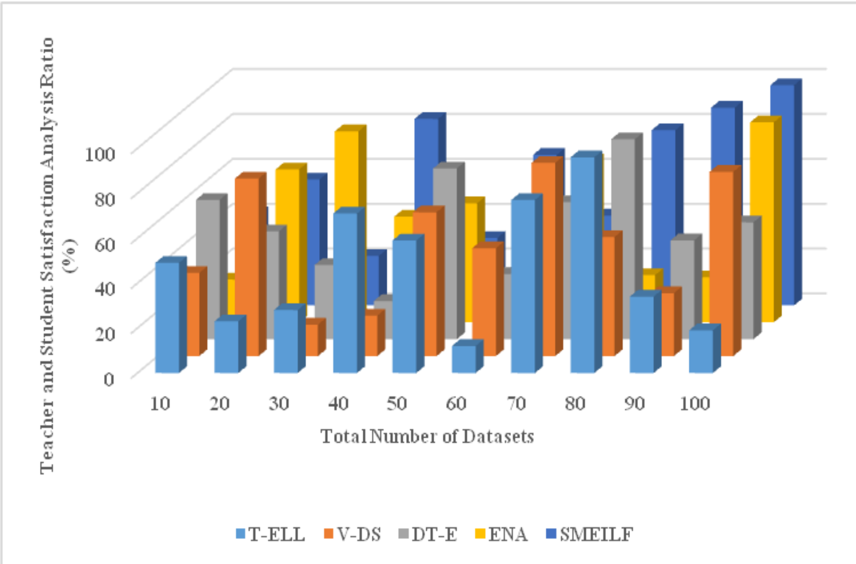

Fig. 4.5: Teacher and Student Satisfaction Analysis

accuracy, and material accessibility in many educational settings. With its adaptive learning and real-time feedback features, the SMEILF framework is an innovative approach for contemporary ESL classrooms, benefiting both teachers and students.

The findings showed that when compared to conventional approaches, SMEILF led to a 25% increase in student involvement levels during courses. This discovery is closely related to the argument in the conversation: the framework may provide real-time feedback, making the learning environment more dynamic. Further supporting the idea that these skills improve the teacher's capacity to assess student knowledge and adjust lessons appropriately, the findings demonstrated an impressive accuracy rate of 90% when it came to identifying students' facial expressions. On the other hand, the findings showed a 2-second lag as a problem with latency in the processing of live video. The impact of this delay on student engagement is further discussed, and several solutions to this problem, such as upgrades to the network, are proposed.

Educators may better meet the requirements of their students by using data collected via superior multimodal processing capabilities to understand their students' involvement, understanding, and emotional reactions. This research can improve student engagement and motivation by paving the way for more dynamic and adaptable classrooms. The need for systems that efficiently synchronize and analyze various data streams is further brought to light by this study, which might lead to advancements in educational technology. Because of this, it has the potential to guide developments in educational technology, which might lead to better, more engaging ways to learn a language. According to the research, finding solutions to technical issues like latency and bandwidth limits is crucial if these technologies are to be used in actual classrooms. The study can transform its delivery to make English education more interesting, flexible, and effective for learners.

5. Conclusion. Integrating image and video processing in English training, the SMEILF has advanced technology-enhanced education. According to studies, SMEILF can teach modern learners using traditional and interactive, multimedia-driven techniques. SMEILF can better serve its varied student body and improve retention by making the learning environment more interactive and individualized, and this approach identifies and fixes implementation issues. There are challenges with strong infrastructure, qualified teachers, scalable computers, and flexible material. Its real-time image and video analysis, individualized feedback, and adjustable learning pathways make it a complete and successful modern teaching system. Simulation studies suggest many courses could benefit from the framework's learning enhancement. According to the study, scalable, flexible, and learner-centred techniques are important in the 21st century. SMEILF makes teaching English easier and more enjoyable than ever, and modern technology and proven teaching methods accomplish this. The paper recommends more research and new frameworks like SMEILF to enable tech-enhanced learning in modern classrooms. Students need such technology to prepare for their current circumstances. Thus, individuals can ensure our students know how to use career-defining technologies and life-enhancing languages. Such technologies are crucial to education's progress. The proposed method increases the learning engagement ratio by 98.5%, pronunciation accuracy ratio by 97.6%, scalability ratio by 99.2%, content accessibility ratio by 92.9%, and teacher and student satisfaction ratio by 95.8% compared to other existing methods. Future work will consider using advanced algorithms that integrate and analyze multiple data streams to mirror human perception, providing a richer and more accurate understanding of student behavior and learning outcomes.

REFERENCES

[1] Gashoot, M., Eve, B., & Mohamed, T. (2023). Implementing technology for teaching: The use of a mobile/tablet approach for enhancing students' learning (design interaction) technology-enhanced learning (TEL). Journal of Education, 203(1), 230-241.

[2] Al Maani, D., & Shanti, Z. (2023). Technology-Enhanced learning in light of Bloom's Taxonomy: A student-experience study of the History of Architecture course. Sustainability, 15(3), 2624.

[3] Plch, L. (2020). Perception of technology-enhanced learning by medical students: an integrative review. Medical science educator, 30(4), 1707-1720.

[4] Hasumi, T., & Chiu, M. S. (2024). Technology-enhanced language learning in English language education: Performance analysis, core publications, and emerging trends. Cogent Education, 11(1), 2346044.

[5] Ng, O. L., & Park, M. (2021). Using an enhanced video-engagement innovation to support STEM teachers' professional development in technology-based instruction. Educational Technology & Society, 24(4), 193-204.

[6] Bui, H. P., Dao, T. T., Dao, T. T., & Vi, V. H. (2023). Technology-Enhanced Teaching and Learning During the COVID-19 Pandemic. Data Analytics for Internet of Things Infrastructure, 203-218.

[7] Chan, S., & Chan, S. (2020). Contribution of Technology-Enhanced Learning: Improving Accessibility to and Effectiveness of Feedback. Identity, Pedagogy and Technology-enhanced Learning: Supporting the Processes of Becoming a Tradesperson, 81-101.

[8] Njai, S., & Nyabuto, E. (2021). Technology Enhanced Learning Environments: Reflecting on the 21 st Century Learning. East African Scholars Journal of Education, Humanities and Literature, 4(4), 202-208.

[9] Pineda, I., & Bosso, R. (2023). Introduction: Virtual English as a lingua franca: Investigating the discourse of digital exchanges and understanding technology-enhanced learning. In Virtual English as a Lingua Franca (pp. 1-18). Routledge.

[10] Kurilovas, E., & Kubilinskiene, S. (2020). Lithuanian case study on evaluating suitability, acceptance and use of IT tools by

students–An example of applying Technology Enhanced Learning Research methods in Higher Education. Computers in Human Behavior, 107, 106274.

[11] Yuan, L. (2022). Communicative competence fostered in a nested EFL learning ecology: Technology-enhanced learning in the Chinese Context. Theory and practice in language studies, 12(11), 2307-2315.

[12] Duong, T. M., Tran, T. Q., & Nguyen, T. T. P. (2021). Non-English majored students' use of English vocabulary learning strategies with technology-enhanced language learning tools. Asian Journal of University Education, 17(4), 455-463.

[13] Shadiev, R., & Yang, M. (2020). Review of studies on technology-enhanced language learning and teaching. Sustainability, 12(2), 524.

[14] Daniela, L. (2021). Smart pedagogy as a driving wheel for technology-enhanced learning. Technology, Knowledge and Learning, 26(4), 711-718.

[15] Tkachenok, K., & Tumskiy, S. (2020). Technology Enhanced Language Learning: Use of Video on Demand Services in Foreign Language Instruction. In INTED2020 Proceedings (pp. 6009-6014). IATED.

[16] Schnaider, K., & Gu, L. (2021). Meaning-making in technology-enhanced learning activities: a composite perspective of technologies and their properties and users'representations. In INTED2021 Proceedings (pp. 1526-1535). IATED.

[17] Jopp, R. (2020). A case study of a technology enhanced learning initiative that supports authentic assessment. Teaching in Higher Education, 25(8), 942-958.

Edited by: Dhilip Kumar V
Special issue on: Unleashing the power of Edge AI for Scalable Image and Video Processing
Received: Aug 22, 2024
Accepted: Nov 18, 2024