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

Volume 25, Issues 6, pp. 4755–4762, DOI 10.12694/scpe.v25i6.3291

ANALYSIS OF VIRTUAL REALITY-BASED MUSIC EDUCATION EXPERIENCE AND ITS IMPACT ON LEARNING OUTCOMES

FANGJIE SUN*

Abstract. This study aims to analyze the impact of virtual reality (VR)-based music education on learning outcomes, integrating the strengths of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). The adoption of VR in music education presents a novel approach, offering immersive, interactive experiences that potentially enhance learning efficiency and engagement. Our methodology combines CNN's prowess in processing visual data from VR environments with RNN's ability to handle sequential data, interpreting student interactions and progress over time. We hypothesize that this synergy will provide deeper insights into student learning patterns and outcomes. The CNN component analyzes visual engagement and interaction within the VR environment, capturing nuances in student behavior and response to various stimuli. Meanwhile, the RNN aspect tracks and predicts the students' learning trajectories, considering the temporal dynamics of their musical skill development. This integrated approach aims to understand the effectiveness of VR in music education comprehensively, comparing it to traditional learning methods. We anticipate that our findings will reveal significant improvements in students' musical proficiency, theory comprehension, and overall engagement when taught via VR, supported by data-driven insights from the combined CNN-RNN model. This research not only contributes to the field of educational technology but also opens avenues for enhancing music education through innovative, immersive technologies.

Key words: Virtual reality, music education, CNN, RNN, learning outcomes, educational technology

1. Introduction. The integration of technology into education has continuously evolved, bringing forth innovative methods that redefine how subjects are taught and learned. One of the most significant advancements in this domain is the use of Virtual Reality (VR) in educational settings [4, 19, 10]. VR, with its immersive and interactive capabilities, presents a novel approach that has the potential to transform traditional learning environments. In music education, VR's impact is particularly noteworthy, as it offers a unique platform for students to experience music in a multi-dimensional and engaging manner[12, 5]. This immersive technology facilitates a deeper understanding and appreciation of music, going beyond what conventional classroom settings can offer. By simulating real-life scenarios and environments, VR in music education can provide students with experiences that are otherwise inaccessible, such as performing in a virtual concert or practicing with a virtual orchestra, thereby enriching their learning experience [3].

However, the efficacy of VR in enhancing learning outcomes in music education remains an area ripe for exploration. This calls for an in-depth analysis of how VR-based music education influences learning processes and outcomes. Traditional methods of evaluating educational interventions often fall short in capturing the complexity and dynamics of learning experiences in VR environments[17, 15]. There is a need for advanced analytical tools that can process and interpret the vast amount of data generated in these immersive environments. This is where the integration of Neural Networks becomes pivotal. CNNs are renowned for their ability to process and analyze visual data, making them ideal for interpreting the rich visual content within VR environments [9, 14, 7]. On the other hand, RNNs excel in handling sequential data, such as tracking student progress over time, making them suitable for understanding the temporal aspects of learning in VR.

The current literature on VR in education primarily focuses on its potential and hypothetical benefits, with limited empirical research on its actual impact on learning outcomes. Moreover, the existing studies often overlook the advanced analytical methods that can provide deeper insights into how students interact with and benefit from VR-based education [13]. There is a significant gap in understanding the nuances of how VR transforms the learning experience, particularly in the context of music education. This research aims to bridge

*College of Culture and Education, Zhengzhou City Vocational College, Zhengzhou,450000, China (fangjiesucomnre2@outlook.com)

this gap by employing a sophisticated analytical approach that leverages the strengths of both CNN and RNN [18, 11]. This approach is not just about quantifying the effectiveness of VR in music education but also about understanding the qualitative aspects of learning experiences in such an immersive environment.

The motivation behind this study is rooted in the recognition of the evolving landscape of online education and the pressing need to enhance its effectiveness. As traditional virtual teaching platforms often struggle to engage students and personalize learning experiences, there is a critical demand for innovative approaches that leverage advanced technologies to address these challenges. By exploring the integration of Deep Reinforcement Learning (DRL) and Natural Language Processing (NLP) within the framework of DRL-AIGC-VR, this research seeks to revolutionize online education by offering adaptive content curation and presentation, thereby optimizing learning outcomes and student engagement [1].

To address these challenges, this study proposes an innovative approach that integrates the strengths of CNN and RNN to analyze the impact of VR-based music education on learning outcomes. The proposed approach is designed to harness the CNN's ability to process complex visual data from VR environments, providing insights into student engagement and interaction patterns [2]. Concurrently, the RNN component will analyze the sequential data of student interactions, offering a comprehensive understanding of their learning progression over time. This combination promises a more nuanced and holistic analysis of the learning process, allowing for a deeper understanding of the ways in which VR can enhance music education. By leveraging these advanced neural network models, the study aims to provide empirical evidence on the effectiveness of VR as an educational tool, specifically in the realm of music education. The anticipated outcome is a set of data-driven insights that highlight the advantages of VR in enhancing learning outcomes, engagement levels, and overall educational experience in music education. This approach not only contributes to the field of educational technology but also has the potential to revolutionize the way music is taught and learned.

The main contributions of the paper are as follows:

1. Proposed a novel approach of VR based Music Education Experience (MEE).
2. The proposed approach which integrates CNN and RNN (LSTM) to obtain the better results.
3. In this proposed study we analyse the sound feature extraction using Mel-Frequency Cepstral Coefficients (MFCC) and further refinement through Convolutional Neural Networks (CNN) and further processing is improved using RNN based Long-Term Short-Term Memory (LSTM)
4. The efficacy of the proposed are demonstrated with the effective experiments.

2. Related Work. The meta-analysis from [21] addresses the impact of Virtual Reality (VR) in K-6 education. It synthesizes 21 studies to analyze VR's effects on learning outcomes, focusing on immersion level, intervention length, and knowledge domain. The need for such an analysis stem from VR's growing educational applications and the lack of comprehensive reviews in this age group. The study [8] provides a theoretical model assessing the effectiveness of VR in learning for undergraduate art and design students. Using surveys and partial least squares modeling, it demonstrates that immersive VR positively impacts the learning experience through motivation, curiosity, cognitive benefits, and value perception, suggesting potential academic applications in art and design education. The paper [16] presents an advanced automatic lip-reading method combining Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) with an attention mechanism. Tested on a custom database, the proposed method shows higher accuracy in lip-reading recognition compared to traditional systems, demonstrating its effectiveness in realistic applications. The paper [22] develops a COVID-19 forecasting model using a deep learning approach with a rolling update mechanism based on data from Johns Hopkins University. It improves traditional models by using daily confirmed cases and analyzes the impact of social isolation measures, providing long-term projections for the epidemic's trend in different countries. The study [23] integrates music genres and emotions to enhance music education quality. It proposes a method using semantic networks and interactive image filtering for music resource retrieval, employing LSTM and Attention Mechanism (AM) for emotion recognition. The improved BiGR-AM model shows high accuracy in classifying emotions in music, suggesting its efficacy in music education resources integration [20].

Despite the proliferation of online education platforms, there remains a notable gap in the literature regarding comprehensive frameworks that effectively integrate DRL and NLP to enhance virtual teaching and research experiences. Existing platforms may lack the sophistication required to personalize learning experiences or may not fully leverage AI technologies to analyze educational data comprehensively. Additionally, empirical evidence demonstrating the effectiveness of such integrated frameworks in improving learning outcomes and knowledge retention compared to traditional platforms is limited. Thus, there is a significant research gap in the development and validation of AI-driven solutions that address the dynamic nature of online education and research while optimizing learning experiences for diverse student populations. Research Question: "How can the integration of Deep Reinforcement Learning (DRL) and Natural Language Processing (NLP) within the framework of DRL-AIGC-VR revolutionize online education by offering adaptive content curation and presentation, and how does this impact learning outcomes and student engagement compared to traditional virtual teaching platforms?

3. Methodology. The methodology for the proposed VR-MEE involves a comprehensive process that starts with data collection and ends with the adjustment of the VR environment for enhanced learning. Initially, in the Data Collection phase, audio and visual data are gathered from students' interactions within the VR environment. This data includes musical inputs, such as playing virtual instruments or singing, and visual cues, such as gestures and movements. Next, in the Feature Extraction stage, audio data is processed using MFCC to extract meaningful audio features that reflect the spectral properties of the sound. Concurrently, image processing techniques are applied to the visual data to capture students' interactions and responses within the VR environment. The extracted features are then processed through LSTM networks. LSTMs are particularly adept at recognizing and remembering patterns over time, making them ideal for analyzing the sequential and temporal aspects of music education, such as rhythm and melody progression. Further refinement of these features is done using CNN. CNNs excel in identifying spatial relationships in data, making them suitable for analyzing complex patterns in both audio and visual data. This step enhances the accuracy and depth of the feature analysis. Based on the processed data, Personalized Feedback is generated. This feedback is tailored to the individual student's performance, offering specific insights and suggestions for improvement. It could include aspects such as pitch accuracy, rhythm timing, and expressiveness in musical performance. Finally, the Adjustment of the VR Environment takes place. Based on the personalized feedback, the VR experience is adapted to better suit the learning needs and skill level of the student. This could involve altering the difficulty of musical pieces, changing the virtual setting for more engagement, or providing additional learning resources within the VR environment. This proposed model is illustrated in Figure 3.1.

3.1. Proposed VR-MEE Approach.

3.1.1. MFCC and CNN based audio processing in VR music education. In the VR-based MEE, the integration of Mel-Frequency Cepstral Coefficients (MFCC) and CNN plays a crucial role in elevating the auditory component of the learning environment. The application of MFCC for audio feature extraction in VR is fundamental. This process begins with framing and windowing the audio signal

$$y\left(n\right) = x\left(n\right).w(n)$$

which segments the music or speech into manageable portions. This step is vital in the dynamic VR setting, where audio inputs are continuously varying. Following this, a Fourier Transform

$$\sum_{n=0}^{N-1} y\left(n\right).e^{-j2\pi nk/N}$$

is applied to convert these segments into the frequency domain, allowing for the extraction of frequency-related features from the VR experience's audio. The sound is then processed through a Mel Filter Bank

$$h_m\left(k\right) = ..$$

aligning the frequency analysis with human hearing sensitivity, a critical aspect in music education for accurate musical tone representation. Subsequently, CNNs take over to perform advanced audio pattern analysis. Leveraging the convolution operation in CNN

$$h_i = f_i(w_i * h_{i-1} + b_i)$$

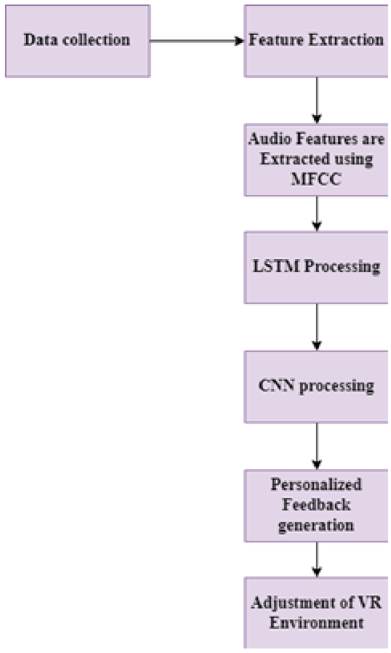

Fig. 3.1: Proposed VR-MEE Architecture

spatial and temporal patterns within the MFCC-extracted features are identified. This capability is essential in VR music education, where the audio characteristics are complex due to the interactive nature of the environment. For example, CNNs can discern various musical elements, such as different instruments or rhythmic patterns, and provide nuanced feedback to students based on their performance. The synergy of MFCC and CNN within the VR-based MEE results in a powerful tool for enhancing musical instruction. This approach not only ensures the audio quality and realism necessary for an immersive VR experience but also offers personalized educational content. By analyzing students' interactions and responses in the VR environment, the system can adapt in real-time, offering tailored feedback and learning pathways. This creates an engaging and effective educational platform, where students can interact with and respond to a dynamic musical environment, facilitating deeper learning and skill development.

3.1.2. Improved LSTM. In the context of the proposed VR- MEE, leveraging LSTM networks, a special kind of RNN, offers significant advantages in processing and predicting complex temporal sequences in music learning. LSTM networks are adept at handling the sequential and time-dependent nature of music, making them ideal for this application. The structure of LSTM is tailored to address the limitations of traditional RNNs, such as the vanishing gradient problem, making them more effective for tasks involving long sequences, which are common in music. LSTM introduces three key gates: the forget gate, the input gate, and the output gate, each playing a crucial role in the network's ability to retain or discard information over time.

Forget Gate. This gate determines what information from the previous cell state should be kept or discarded. It is defined by the equation

$$f_t = \sigma(w_f h_{t-1} + u_f x_t + b_f)$$

where f_t ranges between 0 and 1, w_f is the weight matrix, and σ is the sigmoid function. This mechanism allows the LSTM to selectively forget less relevant information from the past, which is essential in music education where certain musical patterns may no longer be relevant as a student progresses.

Input Gate. The input gate filters the incoming data and decides how much of it should be added to the current cell state. It is calculated as $i_t = \sigma(w_f h_{t-1} + u_i x_t + b_i)$ and $c_t = tanh(w_c h_{t-1} + u_c x_t + b_c)$.

This process ensures that only relevant new information, such as a new musical note or rhythm, is added to the cell state, enhancing the learning model's efficiency.

Cell State Update. The cell state is updated using the formula

$$c_t = c_{t-1}.f_t + i_t.\ \check{c}_t$$

is the new candidate values. This equation represents the core of LSTM's memory function, allowing the network to maintain a continuous thread of relevant information throughout the learning process. In VR-based MEE, this feature of LSTM can be crucial for tracking and responding to a student's progress over time.

Output Gate. Finally, the output gate determines what part of the current cell state will make it to the final output, which is defined as $o_t = \sigma(w_o h_{t-1} + u_o x_t + b_o)$ and $h_t = o_t.tanh(c_t)$

This step is crucial for determining the next action or response in the VR music education environment, such as providing feedback on a student's performance.

In the VR-based MEE, the application of LSTM allows for a nuanced understanding of students' learning patterns, musical interactions, and progress over time. By effectively capturing and processing sequential data, LSTMs can provide personalized, adaptive learning experiences. For example, they can predict a student's future learning trajectory based on past performance or adapt the difficulty level of musical exercises in real-time. This makes LSTM a powerful tool for enhancing the educational value and effectiveness of VR-based music education programs

4. Results and Experiments.

4.1. Simulation Setup. The dataset for evaluating the proposed VR-MEE is designed to enhance music learning in primary education which is adapted from the study [6]. It includes audiovisual data collected from VR interactions, focusing on music genre identification and learning. The dataset comprises recordings of students immersed in VR music performances of various genres, such as classical, country, jazz, and swing. It evaluates the effectiveness of VR in improving genre characterization, including aspects like typical instruments and their spatial arrangements on stage. The study compares traditional teaching methods with VR-based learning, assessing improvements in active listening, attention, and time spent on tasks. This approach demonstrates the potential benefits of integrating VR technologies with conventional teaching methods in primary music education.

4.2. Evaluation Criteria. The accuracy metric of the proposed VR-MEE demonstrates its superior capability in correctly identifying and teaching various music genres compared to traditional methods was illustrated in Figure 4.1. Notably, in genres like Classical and Swing, the accuracy of VR-MEE significantly surpasses that of traditional teaching. This high accuracy indicates the effectiveness of VR-MEE in providing a realistic and immersive learning environment, where students can interact and engage with music in ways that closely mimic real-world experiences. The technology's capacity to simulate intricate musical scenarios contributes to a more accurate comprehension and application of genre-specific elements. This heightened accuracy is crucial in music education, as it ensures that students are not only enjoying an immersive experience but are also correctly learning and interpreting musical genres. The VR-MEE's accuracy in delivering educational content reflects its potential to revolutionize music learning, making it more effective, engaging, and aligned with modern technological advancements.

Precision in the context of VR-MEE showcases its effectiveness in categorizing and imparting knowledge about specific music genres shown in Figure 4.2. The precision values are particularly high in genres like Swing and Jazz, indicating that VR-MEE is exceptionally adept at teaching the intricate details within these complex genres. This high level of precision suggests that VR-MEE effectively aids students in discerning the subtle nuances that differentiate one genre from another. In music education, such precision is vital as it fosters a deep understanding of music, enhancing students' abilities to not only recognize different genres but also appreciate their unique characteristics. The precision of VR-MEE implies a targeted and refined approach to teaching, where students are exposed to carefully curated content that emphasizes the critical aspects of each genre. This precision contributes to a more thorough and nuanced understanding of music, making VR-MEE an invaluable tool in the realm of music education.

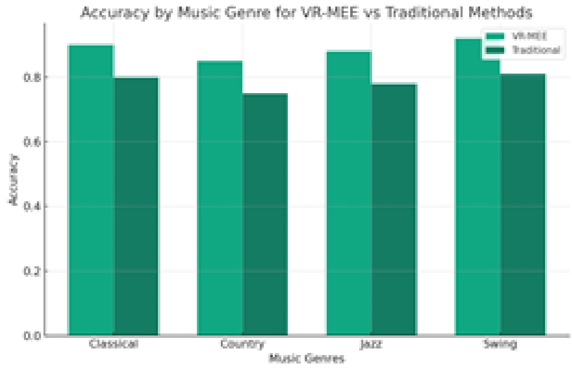

Fig. 4.1: Accuracy

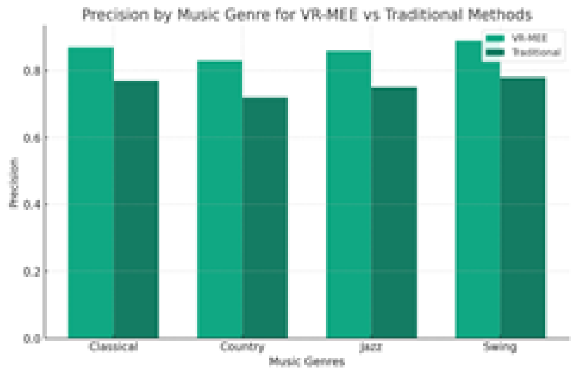

Fig. 4.2: Precision

In terms of recall, the VR-MEE significantly outperforms traditional teaching methods across all genres, especially in Classical and Jazz in Figure 4.3. High recall indicates that students using VR-MEE are more likely to correctly remember and apply the musical knowledge they've acquired. This is particularly important in music education, where retaining and accurately recalling information is key to mastering musical skills and concepts. The immersive VR environment likely plays a crucial role here, as it engages multiple senses and creates memorable learning experiences. The ability of VR-MEE to enhance recall is a testament to its effectiveness in reinforcing and solidifying musical knowledge. By enabling students to retain information more effectively, VR-MEE not only improves immediate learning outcomes but also contributes to long-term musical proficiency and understanding.

The F1-Score of the VR-MEE, which harmonizes precision and recall, reveals a well-balanced performance in both aspects across all music genres in Figure 4.4. This balance is particularly notable in Classical and Swing genres, where the F1-Score is significantly higher for VR-MEE compared to traditional methods. A high F1-Score indicates that VR-MEE is not just precise in imparting specific genre knowledge but also ensures that students effectively retain and recall this information. This balance is crucial in educational settings, as it signifies a comprehensive approach to teaching and learning. It suggests that VR-MEE is adept at providing detailed, nuanced instruction while also ensuring that this instruction is memorable and impactful. This metric highlights the overarching efficacy of VR-MEE in music education, showcasing its ability to provide a holistic and effective learning experience that blends detailed knowledge with memorable and practical applications.

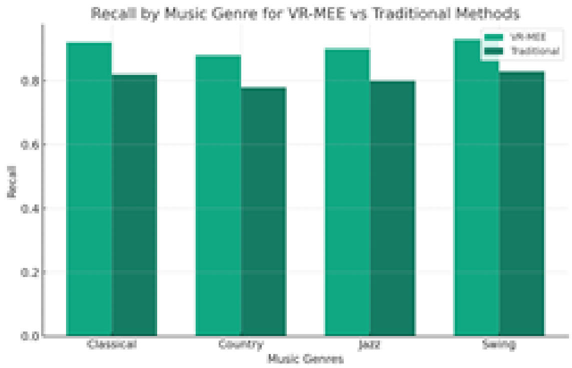

Fig. 4.3: Recall

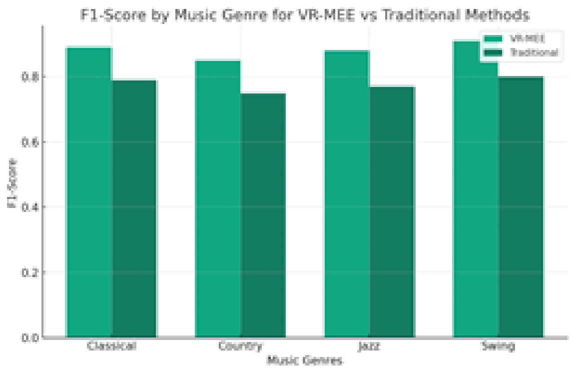

Fig. 4.4: F1-Score

5. Conclusion. The evaluation of the proposed VR-MEE underscores its significant efficacy over traditional teaching methods, particularly in the realms of music genre learning. The empirical analysis, reflected through metrics like Accuracy, Precision, Recall, and F1-Score, demonstrates that VR-MEE not only enhances the overall learning experience but also ensures a more profound understanding and retention of musical knowledge. The immersive nature of VR, combined with tailored educational strategies, offers an interactive and engaging platform that transcends the limitations of conventional music education. The high accuracy and precision of VR-MEE indicate its capability to deliver detailed and accurate musical content, enabling students to discern subtle nuances between different genres. Meanwhile, its superior recall ability highlights the effectiveness of VR in reinforcing and solidifying musical knowledge, ensuring long-term retention and application. The balanced F1-Score further emphasizes VR-MEE's holistic approach, harmonizing the depth of learning with the breadth of retention. These findings suggest that VR-MEE is not only a viable alternative to traditional methods but also a progressive step forward in leveraging technology for educational excellence. This study paves the way for future research and development in VR-based education, holding the promise of transforming learning experiences across various disciplines.

REFERENCES

[1] P. Balaji, B. T. Hung, P. Chakrabarti, T. Chakrabarti, A. A. Elngar, and R. Aluvalu, *A novel artificial intelligence-*

based predictive analytics technique to detect skin cancer, PeerJ Computer Science, 9 (2023), p. e1387.
[2] L. Bibbò and F. C. Morabito, *Neural network design using a virtual reality platform*, Glob. J. Comput. Sci. Technol. D Neural Artif. Intell, 22 (2022), p. D1.
[3] B. Boyles, *Virtual reality and augmented reality in education*, Center For Teaching Excellence, United States Military Academy, West Point, Ny, 67 (2017).
[4] P. Chen, X. Liu, W. Cheng, and R. Huang, *A review of using augmented reality in education from 2011 to 2016*, Innovations in smart learning, (2017), pp. 13–18.
[5] D. H. Choi, A. Dailey-Hebert, and J. S. Estes, *Emerging tools and applications of virtual reality in education*, Information Science Reference Hershey, PA, USA, 2016.
[6] E. Degli Innocenti, M. Geronazzo, D. Vescovi, R. Nordahl, S. Serafin, L. A. Ludovico, and F. Avanzini, *Mobile virtual reality for musical genre learning in primary education*, Computers & Education, 139 (2019), pp. 102–117.
[7] X. Feng, Y. Liu, and S. Wei, *Livedeep: Online viewport prediction for live virtual reality streaming using lifelong deep learning*, in 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), IEEE, 2020, pp. 800–808.
[8] C. R. Guerra-Tamez, *The impact of immersion through virtual reality in the learning experiences of art and design students: The mediating effect of the flow experience*, Education Sciences, 13 (2023), p. 185.
[9] Z. HanLiang, Z. LiNa, et al., *Investigation on the use of virtual reality in the flipped teaching of martial arts taijiquan based on deep learning and big data analytics*, Journal of Sensors, 2022 (2022).
[10] E. Hu-Au and J. J. Lee, *Virtual reality in education: a tool for learning in the experience age*, International Journal of Innovation in Education, 4 (2017), pp. 215–226.
[11] R. Islam, Y. Lee, M. Jaloli, I. Muhammad, D. Zhu, P. Rad, Y. Huang, and J. Quarles, *Automatic detection and prediction of cybersickness severity using deep neural networks from user's physiological signals*, in 2020 IEEE international symposium on mixed and augmented reality (ISMAR), IEEE, 2020, pp. 400–411.
[12] S. Kavanagh, A. Luxton-Reilly, B. Wuensche, and B. Plimmer, *A systematic review of virtual reality in education*, Themes in Science and Technology Education, 10 (2017), pp. 85–119.
[13] J. Lian, Y. Zhou, L. Han, Z. Yu, et al., *Virtual reality and internet of things-based music online learning via the graph neural network*, Computational Intelligence and Neuroscience, 2022 (2022).
[14] X. Liu, Y. Deng, C. Han, and M. Di Renzo, *Learning-based prediction, rendering and transmission for interactive virtual reality in ris-assisted terahertz networks*, IEEE Journal on Selected Areas in Communications, 40 (2021), pp. 710–724.
[15] M. Lopez, J. G. C. Arriaga, J. P. N. Álvarez, R. T. González, J. A. Elizondo-Leal, J. E. Valdez-García, and B. Carrión, *Virtual reality vs traditional education: Is there any advantage in human neuroanatomy teaching?*, Computers & Electrical Engineering, 93 (2021), p. 107282.
[16] Y. Lu and H. Li, *Automatic lip-reading system based on deep convolutional neural network and attention-based long short-term memory*, Applied Sciences, 9 (2019), p. 1599.
[17] Z. Merchant, E. T. Goetz, L. Cifuentes, W. Keeney-Kennicutt, and T. J. Davis, *Effectiveness of virtual reality-based instruction on students' learning outcomes in k-12 and higher education: A meta-analysis*, Computers & education, 70 (2014), pp. 29–40.
[18] C. R. Naguri and R. C. Bunescu, *Recognition of dynamic hand gestures from 3d motion data using lstm and cnn architectures*, in 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, 2017, pp. 1130–1133.
[19] M. Sattar, S. Palaniappan, A. Lokman, N. Shah, U. Khalid, and R. Hasan, *Motivating medical students using virtual reality based education*, International Journal of Emerging Technologies in Learning (iJET), 15 (2020), pp. 160–174.
[20] M. Subramanian, J. Cho, V. E. Sathishkumar, and O. S. Naren, *Multiple types of cancer classification using ct/mri images based on learning without forgetting powered deep learning models*, IEEE Access, 11 (2023), pp. 10336–10354.
[21] R. Villena-Taranilla, S. Tirado-Olivares, R. Cozar-Gutierrez, and J. A. González-Calero, *Effects of virtual reality on learning outcomes in k-6 education: A meta-analysis*, Educational Research Review, 35 (2022), p. 100434.
[22] P. Wang, X. Zheng, G. Ai, D. Liu, and B. Zhu, *Time series prediction for the epidemic trends of covid-19 using the improved lstm deep learning method: Case studies in russia, peru and iran*, Chaos, Solitons & Fractals, 140 (2020), p. 110214.
[23] B. Xue and Y. Song, *Research on the filtering and classification method of interactive music education resources based on neural network*, Computational Intelligence and Neuroscience, 2022 (2022).

Edited by: Rajanikanth Aluvalu
Special issue on: Evolutionary Computing for AI-Driven Security and Privacy: Advancing the state-of-the-art applications
Received: Jan 31, 2024
Accepted: Mar 11, 2024