

Article

Sustainable Educational Metaverse Content and System Based on Deep Learning for Enhancing Learner Immersion

Jaekyu Lee ¹ and Yeichang Kim ^{2,*}
¹ Department of Techno-Management Cooperation Course, Dongguk University, 123 Dongdae-ro, Gyeongju-si 38066, Gyeongsangbuk-do, Republic of Korea; duckjk89@daum.net

² Department of Information Management, Dongguk University, 123 Dongdae-ro, Gyeongju-si 38066, Gyeongsangbuk-do, Republic of Korea

* Correspondence: kimyc@dongguk.ac.kr; Tel.: +82-54-770-2345

Abstract: Social distancing has encouraged the use of various non-face-to-face services utilizing information and communication technology, especially in the education sector. Educators and learners are increasingly utilizing online technology to conduct non-face-to-face classes, which has resulted in an increased use of EduTech. Virtual education is expected to expand continuously. However, students involved in virtual education find it difficult to focus and participate in the classes. Hence, we propose a system that can improve learners' focus and immersion in metaverse-based education. In this paper, we propose a sustainable educational metaverse content and system based on deep learning that can enhance learners' immersion. We built an AI-based simulation that judges learning activities based on the learning behavior rather than on the learner's device and program events and allows the user to proceed to the next level of education. In the simulation implemented in this study, virtual reality educational contents were created for 12 educational activities, and the effectiveness of four learning models in assessing the learning effectiveness of learners was evaluated. From the four models, an ensemble model with boosting was adopted considering its accuracy, complexity, and efficiency. The F1-score and specificity of the adopted learning model were confirmed. This model was applied to the system in a simulation.

Keywords: metaverse; EduTech (education technology); immersion; virtual reality; deep learning



Citation: Lee, J.; Kim, Y. Sustainable Educational Metaverse Content and System Based on Deep Learning for Enhancing Learner Immersion.

Sustainability **2023**, *15*, 12663.

<https://doi.org/10.3390/su151612663>

su151612663

Received: 18 July 2023

Revised: 18 August 2023

Accepted: 18 August 2023

Published: 21 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The metaverse technology enables political, economic, social, and cultural activities that occur in the real world to be implemented in the virtual world. Metaverse services have expanded across various fields. In particular, in the field of education, the convergence of educational services with information technology has led to the emergence of the concept of EduTech to meet the diverse needs of learners. EduTech is the convergence of information and communication technology (ICT) and educational services, such as virtual reality (VR), augmented reality (AR), artificial intelligence (AI), and big data, to provide new learning methods [1].

During the COVID-19 pandemic, the participation of educators and learners in non-face-to-face education increased significantly. However, there have been negative reviews of non-face-to-face education, mainly related to the problem of reduced concentration of the learners [2,3]. Learners' concentration and immersion have a direct impact on the effectiveness and quality of learning. Learners with high concentration and immersion understand new information more quickly, retain it longer, and enhance their ability to apply that knowledge in practice. Here, immersion is a key element to increase a learner's concentration and participation [3]. Conversely, when concentration decreases, the effectiveness of learning significantly diminishes, and the time spent on learning increases. As a result, numerous prior studies were actively conducted to enhance concentration in

online classes [4,5]. In particular, metaverse-based education is being increasingly explored, leading to the emergence of metaverse campuses and metaverse educational contents [6–8].

The existing metaverse content was developed to recognize events based on objects and triggers between objects. In situations that require sophisticated manipulation, a high level of control skills is required to activate the triggers [9,10]. Especially in a VR educational content, users with poor control skills have difficulty progressing through the content because they do not have the same freedom of movement as in real life. Such users can become frustrated because they are unable to complete their learning owing to their lack of control skills, even if they are diligent learners [10]. For example, in the VR course on assembling a machine, if the user is not skilled enough to perform a precise assembly, the trigger for the lesson completion will not be activated. Although this is a perfectly acceptable situation in the real world, it is not for VR educational content, and the learner is unable to progress to later lessons, resulting in frustration among learners and reduced immersion in the learning content.

The primary aim of this study is to propose a system designed to enhance the waning concentration and immersion of learners in the prevailing remote virtual education paradigm. While virtual reality (VR) educational content can potentially augment focus and immersion compared to traditional educational modalities, it is not devoid of limitations. A significant challenge arises for learners who, due to geographical constraints or economic factors, have limited exposure to cutting-edge technologies, which will lead to a lack of proficiency in their operation. Such operational inefficiencies hinder these learners from completing VR educational tasks, precluding their advancement to subsequent educational phases. This impediment prevents learners from fully engaging with VR educational content, culminating in persistent disparities in accessing continuous VR education. In this paper, we introduce an artificial intelligence-driven system that discerns and assists the educational activities of learners, irrespective of their operational proficiency. This initiative seeks to develop a sustainable educational metaverse system predicated on deep learning, anticipated to bolster learners' immersion and concentration.

This paper discusses the effects of metaverse-based education and the enhancement of immersion in educational content using Artificial Intelligence (AI). Researchers have discussed the creation of VR educational content and methods to implement educational content with the assistance of AI. In the Introduction section, the effects of metaverse-based education are discussed. It delves into the limitations and problems of traditional educational methods and how metaverse-based education can enhance learners' concentration and participation. However, since there are limitations to the effects of metaverse education for students with inadequate operational skills, the paper discusses AI systems to address this. In the Related Work section, AI technologies to enhance the educational effects of VR content are discussed, focusing on AI technologies to solve the decreased immersion in education due to learners' lack of operational skills. The section System Design discusses the design of the system and learning models, implementing various learning models and comparing their performance. It includes descriptions of the models used to extract features from video datasets and their results. In the Implementation section, the performance of the selected model is measured, and learners' behaviors in dealing with VR educational content are analyzed, conducting experiments to grant students the authority to proceed to the next level of education. The section Conclusion discusses the limitations of the paper and directions for its future expansion.

2. Related Work

2.1. Metaverse-Based Educational VR Content to Increase Learners' Immersion

Jean Piaget [11] provided profound insights into how children's ways of thinking change as they grow by introducing the theory of cognitive development. This theory divides the developmental process humans experience as they grow into four stages: the Sensory–Motor Stage, the Pre-operational Stage, the Concrete Operational Stage, and the Formal Operational Stage.

The Sensory–Motor Stage corresponds to the age of 0–2 years. Infants in this stage understand the world through physical actions and grasp the concept of object permanence, which means objects continue to exist even when not visible. The Pre-operational Stage spans the 2–7 years of age. Children in this stage explore the world using symbolic thinking but often show characteristics like centration, i.e., they consider only one perspective, and struggle with the idea of conservation, which means that a specific quantity remains unchanged regardless of its shape or arrangement. The Concrete Operational Stage covers the age of 7–11 years. Children in this stage begin to think more abstractly and logically, starting to understand principles like conservation, classification, and seriation. The Formal Operational Stage regards the age of 12 years onwards. Adolescents develop higher levels of thinking about abstract concepts and problem solving, employing hypothetical and conditional thinking for problem resolution.

Piaget’s theory of cognitive development offers a fundamental understanding of the changes in the ways children and adolescents think and has influenced many aspects of modern education. This theory provides a framework explaining how children’s ways of thinking evolve.

Education based on the metaverse can facilitate learning that meets the criteria of constructivism, connectivism, and experiential learning theories [12]. The essence of constructivism is that learning is an individual and subjective process, where learners find meaning by assimilating or accommodating new information into their existing knowledge structures. In metaverse-based education, personalized learning paths can be provided tailored to the learner’s existing knowledge and experiences. Connectivism emphasizes that learning is not just about absorbing simple information but is about forming and managing connections with various information sources. In metaverse education, learners can be given an environment where they can connect with diverse information. The experiential learning theory posits that learning occurs through an individual’s experiences. In metaverse education, learners can experience historical sites and learn about history and nature beyond the constraints of time and space.

While metaverse-based education can offer an effective learning environment from various pedagogical perspectives, if learners find it uncomfortable and avoid it, the expected learning outcomes may not be achieved. To address this, this paper researched the limitations of metaverse-based education. Among these, the study focused on overcoming the challenges learners face in operating the metaverse, implementing a system aided by artificial intelligence.

The coronavirus pandemic has revitalized non-face-to-face video classes, leading to the increased commercialization of educational methods using ICT. However, the problem with non-face-to-face video-based educational methods is that they are limited to learning through observation, and the lack of supervision can lead to a loss of concentration [13]. To address this problem, research has been conducted on VR-based education using metaverse technologies [14,15]. Metaverse comprises four technologies: lifelogging, mirror worlds, AR, and VR. It provides an environment that seamlessly connects the real and the virtual worlds [16].

Traditional educational methods predominantly rely on textbooks and other instructional materials. In contrast, VR (Virtual Reality) education offers learners an immersive learning environment, allowing them to experience a variety of scenarios in a virtual world that mimics real-life settings [17]. Platforms such as Oculus Education, Google Expeditions, Labster, Nearpod, Unimersiv, ENGAGE, AltspaceVR, UroVerse [18,19], and PedUroVerse provide learners with a realistic learning experience, significantly enhancing the effectiveness and efficiency of education.

UroVerse and PedUroVerse offer surgical simulations in the medical field, enabling physicians to practice and refine their surgical techniques in various scenarios before operating on actual patients [20]. Labster provides science experiment simulations, allowing students to safely practice complex experiments, while Google Expeditions offers virtual trips to various places and eras [19]. Nearpod and Unimersiv deliver immersive learning

experiences across diverse topics and environments. ENGAGE and AltspaceVR facilitate interactive learning in virtual educational spaces where students and teachers can engage with one another.

Ryu et al. [21] studied the implementation and learning effects of history-learning content using the VR technology. This study found that the educational satisfaction of experiencing a historical time and place through VR content is higher than the educational satisfaction that learners receive from traditional educational media.

In a study conducted by Guanjie Zhao [17], the efficacy of VR-based education was compared to that of traditional educational methods within the realm of medical training. Their findings indicated that students who underwent VR-based training had a higher pass rate than those who received conventional education. This research underscores the potential of VR education to offer a more effective learning experience for students. Furthermore, a study by Bibo Ruan [22] employed VR technology to educate about environmental conservation. The results from this research highlighted that VR education enhanced students' learning efficiency and their awareness of environmental protection. This suggests that VR education can elevate student engagement and interest more than traditional educational methods. Previous studies showed that VR education provides a richer and more immersive learning experience compared to conventional educational methods. The utility of such VR educational content is also evident in its profitability. Globally, the VR education market is witnessing consistent growth, indicating an increasing investment in VR educational solutions by educational institutions and businesses. Such investments significantly contribute to the advancement of VR technology and the enhancement of educational quality [17].

The widespread adoption of education utilizing the metaverse and AI presents a new horizon for modern education. However, this proliferation is not without its challenges. One primary concern is the issue of technological accessibility. The feasibility of large-scale technology can pose significant challenges [20]. While advanced cities with high technological resources can easily access the metaverse, peripheral regions might face difficulties. Such disparities in accessibility can impact the effectiveness and efficiency of education [23].

Moreover, the issue of social inequality cannot be overlooked. Particularly vulnerable to exclusion from the advancements of the metaverse are poorer populations, less educated groups, and the elderly. Such digital divides can significantly influence the effectiveness and efficiency of education [24].

Lastly, with the expansion of metaverse and AI education, data security is becoming increasingly paramount. Concerns surrounding personal data protection policies, regulations, and cybersecurity can greatly affect users' trust and the efficacy of education [25,26]. Metaverse and AI-based educational systems collect and analyze a plethora of data, such as students' grades, behaviors, interactions, and learning habits. While these data play a crucial role in optimizing individual learning experiences, they also pose potential threats to students' personal information. Hence, a thorough review and regulation of data collection, storage, and utilization methods are imperative [26]. Given that educational data often contain sensitive information, their protection is of utmost importance. Hacking or security breaches can risk exposing students' personal information, potentially damaging the credibility of educational institutions [27].

While education systems utilizing the metaverse and AI hold immense potential to enhance the quality and efficiency of education, the introduction and application of these technologies bring forth several ethical concerns related to data privacy and security. Addressing these issues necessitates continuous collaboration and discussion among educational institutions, policymakers, and AI developers [28]. This paper addresses one of the limitations of the large-scale proliferation of VR education: the decline in realism due to differences in content manipulation proficiency stemming from social inequality.

In virtual reality educational content, a decline in presence can impede immersion, thereby compromising the potential for optimal educational outcomes. The sense of

presence makes the users feel as if they are actually in a virtual environment, not in the real world. When learners are immersed in the VR training content, their sense of presence is enhanced. Increased presence leads to a realistic experience and improves the educational effectiveness [10,23]. Presence depends on visual factors such as the quality, size, proportion, color, and viewing distance of the images presented in the content, as well as on auditory factors such as sound. It also depends on the media usage skills. The user skills relate to the user's familiarity with the control scheme and control. If learners have poor dexterity and are unable to act independently, they will not be able to become immerse in the content, even if the VR educational content provides high-quality visual and auditory elements [10,24]. This reduces the sense of presence and learning effectiveness.

Learners with low control skills have greater difficulty in completing the educational activities because they cannot accustom themselves to the VR experience. Consequently, they are unable to progress to the next lesson and cannot become immerse in the educational content. To increase the learner's presence, it is necessary to lower the barrier to entry regarding the control of the VR educational content, so that the learner can have an immersive experience.

This paper introduces a system designed to address a specific limitation: the challenges faced by users unfamiliar with VR educational content due to social inequalities, such as receiving inadequate education or having limited opportunities to engage with such systems. To overcome this limitation, we implemented a system aimed at enhancing learners' immersion by counteracting factors that reduce their presence. In this study, we implemented a VR educational content involving 12 educational activities to address the factors that hinder learners' immersion. The implemented VR educational content was monitored using AI. If the AI determines that the learner is performing an educational activity, it grants the learner permission to proceed to the next lesson, even if the learner is unable to complete the educational mission owing to low control skills. This simulation improved the sense of presence of unskilled learners by lowering the strict demand for control skills, which interfered with immersion. Learners could participate in virtual reality educational content sustainably, which will lead to an increase in educational effectiveness.

2.2. Convolutional Neural Network

AI is the study of how computers perceive, think, and learn like humans. AI is being applied not only to the natural sciences and computer science but also to humanities and social fields. AI is categorized as rule-based, neural network-based, and statistics-based AI. Neural-network-based AI is organized based on a connection model of neurons. In particular, it has made significant progress in pattern recognition using learning methods. It is primarily used for learning and recognizing fields with large amounts of data and high complexity, such as character, voice, and image recognition. Recently, the introduction of the deep learning technology into the neural network family has yielded satisfactory results in speech, image, and video recognition.

Convolutional neural networks (CNNs) are deep learning techniques used for handwriting character recognition; however, more recently, they have been used to mimic the manner in which the human optic nerve processes image information. CNNs scale low-dimensional information into higher dimensions to easily classify the information from the input images. A CNN consists of a convolutional layer, a pooling layer, and a classification layer, as shown in Figure 1. The convolutional layer separates the image into multiple dimensions, and the pooling layer contours them using filters. Subsequently, parts of the image are separated into multiple dimensions, and the contours are refined with filters. The convolutional and pooling layers are alternatively used to extract small features of the image in detail. The classification layer then classifies and separates the extracted feature images [29].

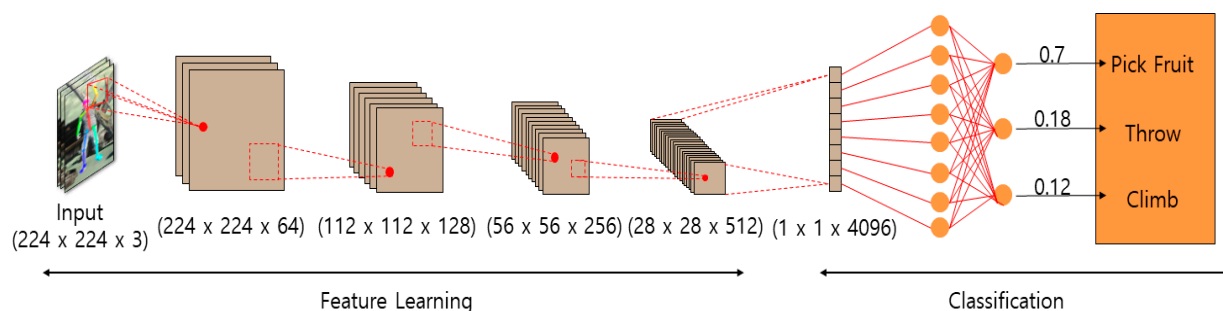


Figure 1. Structure of the CNN.

2.2.1. Object-Tracking Learning

Object-tracking learning is a technology that detects and tracks people or objects based on the object-tracking technology and identifies abnormal behaviors through changes in the movement or pose of the tracked object and the relationship between objects. It is popular because it can be used to easily extract objects from the background of stationary CCTV videos and track them. However, because the method is based on object tracking, it has limited effectiveness in crowded scenes with distractors that obscure the tracked objects [30] (see Figure 2).

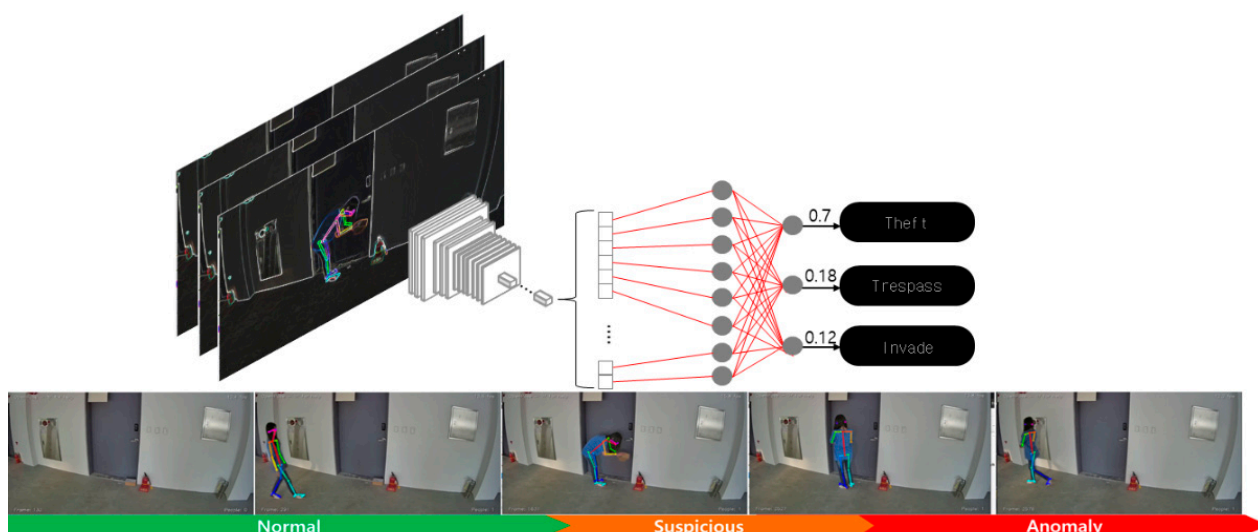


Figure 2. Object recognition in object-tracking learning.

2.2.2. Multiple-Instance Learning

Multiple-instance learning uses multiple instances to automatically detect and learn anomalous behaviors in a video. It learns from video data containing both normal and abnormal behaviors. However, a disadvantage of this learning method is that it requires a large amount of data to determine a single anomaly.

Figure 3 shows the method used for multiple-instance learning. This learning method separates the video data in which the object behaves abnormally from the normal image data in which the object exhibits a normal behavior. Two bag instances are created: one for the negative bag and one for the positive bag. The general image data are placed in the positive bag, and uncommon image data are stored in the negative bag. The image data divided into positive and negative forms were subjected to a convolutional process. The features of positive or negative behaviors extracted through the convolutional process were then classified [31].

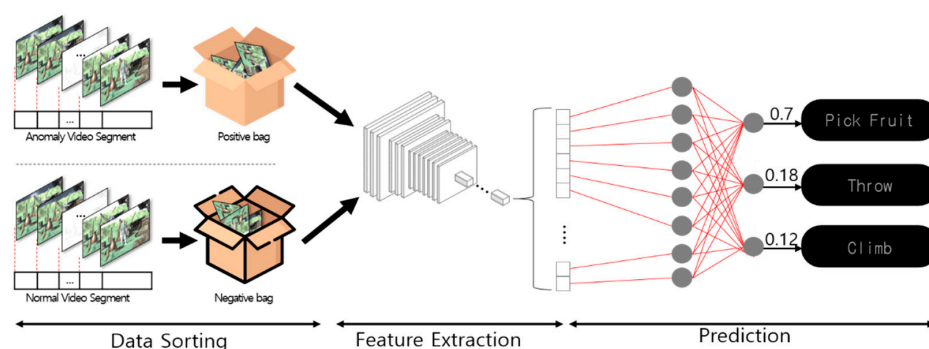


Figure 3. Image analysis in multiple-instance learning.

In this study, we used CNN-based object-tracking learning and multiple-instance learning during deep learning to analyze a video content. The anomalous behavior described in the learning method was transformed into a learning behavior, learned, and detected. Therefore, an anomaly or anomalous behavior was the learning behavior in this case, and the state of no learning was the normal situation. In other words, the learning behavior was detected as a method for detecting an anomalous behavior.

2.2.3. Ensemble Models Based on Deep Learning

Ensembling combines multiple learning models for a better predictive performance in AI. An ensemble is a strong classifier that identifies an action by combining the results of N weak classifiers to analyze a specific action in a video. Ensemble techniques include boosting and voting [32].

1. **Boosting:** Boosting is a serialized ensemble technique. Boosting works collaboratively with different models such that each data instance has a weight. When the first model makes a prediction, the data are weighted according to the prediction results. The next learning model learns by focusing on the error in the weight provided by the previous model. This method increases the predictability by focusing on misclassified data and repeatedly creating new rules [32]. For example, a dataset consisting of “+” and “−” was classified as shown in Figure 4. In D1, the “+” and “−” data were separated by the dividing line at the 2/5 point. However, the circled “+” in the top half of D1 and the two circled “−” at the bottom were misclassified. Misclassified data are assigned a higher weight, and well-classified data are assigned a lower weight. In D2, the size of the well-classified data in D1 decreased as the weight decreased, whereas the size of the misclassified data increased as the weight increased. The reason for assigning weights to misclassified data is to focus on the classification in the next model. In D2, the right three “−” classified as vertical lines were misclassified. Therefore, in D3, the weights of the three “−” increased. Because “+” and “−”, which were weighted in the first model, were classified well in D2, the weights were reduced again in D3. By combining the classifiers of D1, D2, and D3, the “+” and “−” were accurately distinguished in the final classifier.
2. **Voting:** Voting is a method for constructing k weak classifiers from different algorithms and selecting the final result by collecting the results of the weak classifiers. Two types of methods are used to derive the results: hard and soft. The hard type is determined by the majority vote, based on the results of the weak classifiers. The soft type determines the final result by averaging the probabilities, which are the resulting values of the weak classifiers (Figure 5). In this study, the hard type of method was used [33].

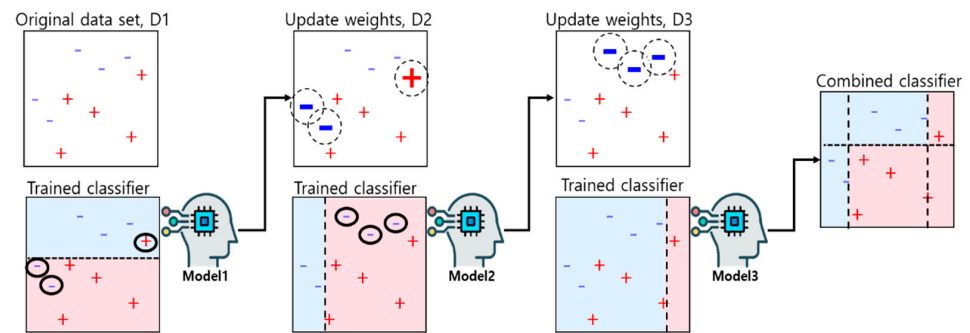


Figure 4. Boosting prediction method.

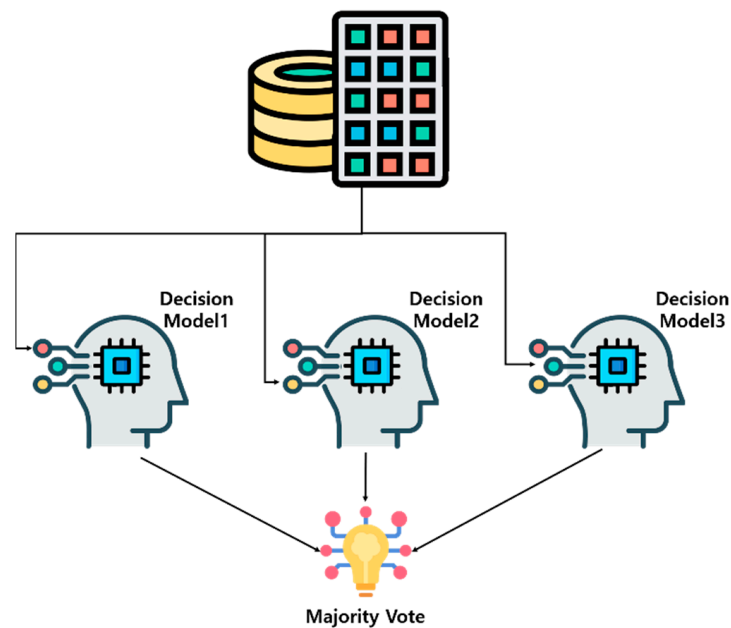


Figure 5. Voting prediction method.

In addition, the following models were used: an object-tracking learning method (Model_A), a multiple-instance learning method (Model_B), a boosting technique that ensembled Model_A and Model_B (Model_C), and a voting technique (Model_D) that ensembled Model_A, Model_B, and Model_C, which were compared and analyzed.

3. System Design

3.1. VR Educational Contents for Simulation

Figure 6 shows the content developed by AI to help learners analyze their learning behaviors. The learners explored nature, responded to emergencies, and accessed history education contents in VR. During the course of the educational content, the AI was designed to analyze whether the learners performed the educational activities.

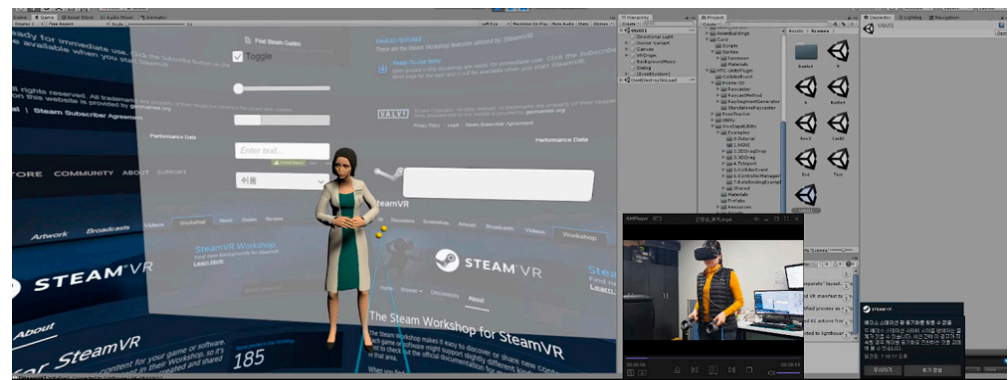


Figure 6. VR content for education.

3.2. Learning Model Design

In this study, we compared the performances of object-tracking learning, multi-instance learning, and the ensemble method of the two learning methods to analyze a learner's behavior while progressing through the educational content. The learning methods were classified as Model_A, Model_B, Model_C, and Model_D, as mentioned above.

3.2.1. Model_A: Object-Tracking Learning

Object-tracking learning is a highly advantageous model for intensive detection when the learning behavior to be extracted is clearly defined. However, it has the disadvantage of defining unlimited and ambiguous actions in detail and labeling them individually. This research team used the message-passing encoder–decoder recurrent neural network (MPED-RNN) model as Model_A. The MPED-RNN analyzes minute movements such as the deformations due to internal skeletal movements [34] (see Figure 7).



Figure 7. Example of a skeleton dataset for learning object tracking.

3.2.2. Model_B: Multiple-Instance Learning

Multi-instance learning is a model suitable for detecting unlimited and ambiguous behaviors because it requires less labeling of training data; however, it is difficult to target the behaviors to be detected. This research team used the C3D (convolutional 3D) model as the multi-instance learning method, Model_B. The C3D model is designed to process 3D video data by extending them to 3D from an existing 2D convolution architecture that recognizes only spatial features. The model uses 3D convolutional layers to capture the temporal features of a video. Figure 8 depicts the sample data used by Model_B.



Figure 8. Example dataset for learning by Model_B.

3.2.3. Model_C: Boosting

Model_C is the boosting of Model_A and Model_B. Model_C uses adaptive boosting (AdaBoost), a representative model for boosting (Figure 9). AdaBoost was selected because it has the advantage of robustness against imbalanced data environments. Algorithm 1 presents the pseudocode for the structural design of Model_C.

Algorithm 1 Model_C structure.

1. $X = \text{edu_data.data}$
 2. $y = \text{edu_data.target}$
 3. $X_{\text{train}}, X_{\text{test}}, y_{\text{train}}, y_{\text{test}} = \text{train_test_split}(X, y, \text{test_size} = 0.2)$
 4. $\text{model_C} = \text{AdaBoostClassifier}(\text{base_estimators} = [(\text{C3DWrapper}()),$
 5. $(\text{MPED_RNN_Wrapper}())])$
 6. $\text{model_C} = \text{model_C.fit}(X_{\text{train}}, y_{\text{train}})$
 7. $y_{\text{pred}} = \text{model_C.predict}(X_{\text{test}})$
 8. $\text{print}(\text{"Accuracy:"}, \text{metrics.accuracy_score}(y_{\text{test}}, y_{\text{pred}}))$
-

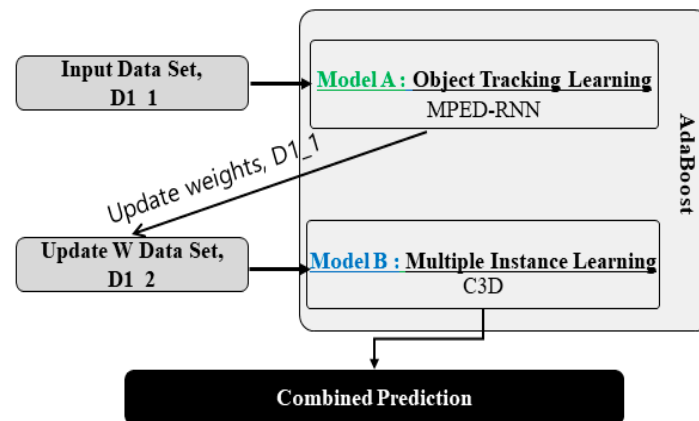
Model C : Boosting Model

Figure 9. Learning structure design of Model_C.

3.2.4. Model_D: Learning Model by Voting Model_A, Model_B, and Model_C

Model_D is an ensemble of Model_A, Model_B, and Model_C in the form of voting (Figure 10). Algorithm 2 presents the pseudocode for the structural design of Model_D.

Algorithm 2 Model_D structure

```

1:  model_A = Mped_Rnn_Model()
2:  model_B = C3D_model()
3:  model_C = AdaBoostClassifier(base_estimators = [(C3DWrapper()),
4:  (MPED_RNN_Wrapper())])
5:  model_D = VotingClassifier(estimators = [('MPED-RNN', model_A),
6:  ('C3D', model_B), ('AdaBoostClassifier', model_C)], voting = 'hard')
7:  model_D.fit(X_train, y_train)
8:  pred = model_D.predict(X_test)
9:  print('VotingClassifier Accuracy:', round(accuracy_score(y_test, pred),4))

```

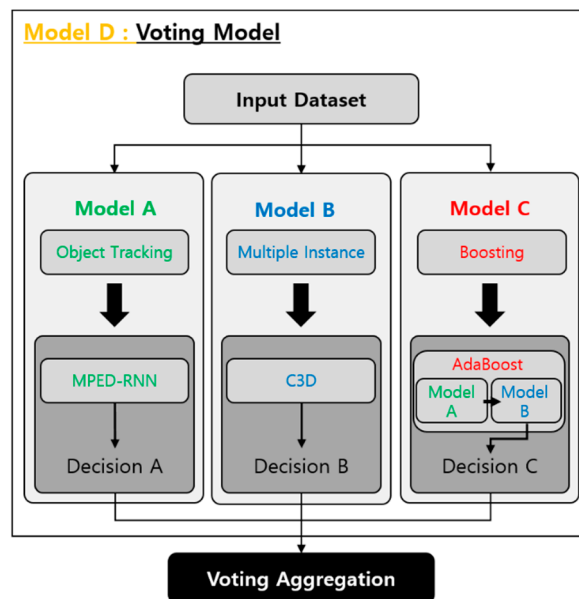


Figure 10. Learning structure design of Model_D.

The accuracy of each model was measured, and the model with the highest accuracy was selected. The F1-score of the selected model was then measured. The selected model

discriminated whether a learner was learning. The model was simulated to give the learner permission to perform the learning activity and continue to the next lesson.

4. Implementation

4.1. Experimental Setup and Performance Comparison of Learning Models

This research team used the MPED-RNN, C3D, AdaBoost Classifier, and Voting Classifier to extract the features of the video dataset. Keras, matplotlib, numpy, opencv-python, Pillow, tensorflow, imutils, imageio, scikit-learn, pandas, joblib, datetime, pickle, and time APIs were also used. The dataset was 240×320 pixels in size at 30 fps.

The learning themes were categorized into nature exploration, emergency response, and historical experiences. Thirteen behaviors were categorized into two types: participation and non-participation in education. Planting, fruit picking, harvesting, and milking were classified as educational participation activities related to nature exploration. Help, extinguish, check temp, CPR, and escape were classified as educational participation activities related to emergency response. Hurray, draw, and shoot were classified as educational participation activities related to historical experience. A total of 251 data points were used for learning.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Negative} + \text{False Positive}} \quad (1)$$

The attributes of the utilized learning data were categorized based on data type, included content, and labeling items, and are presented in Tables 1 and 2.

Table 1. Data type and included content.

	Educational Activity Behavior Video	Educational Activity Behavior XML File
Content	Video containing 12 types of educational activity behaviors	Labeling for the 12 types of educational activity behaviors
Quantity	259	259
Format	MP4	XML

Table 2. Labeling Items.

Item		Description
folder		Parent Folder Name (Organized based on each event)
filename		Original Video File Name (Matched with XML file)
hader	duration	Video Duration
	fps	Frames per Second (FPS)
	frames	Total Frame Count
	location	Location within the Video
	time	Time Period
	population	Number of Participants Engaging in the Educational Activity
event	eventname	Educational Activity Name
	starttime	Educational Activity Start Time
	duration	Educational Activity Duration
	objectname	Object Name
	position	Object's Positional Information
	frame	Frame in which the Object Appears
object	keypoin	Specific Location of the Object
	actionname	Action Name
	frame	Frames where the Action Begins and Ends

For the learning of Model_A, Model_B, Model_C, and Model_D, the dataset was divided in the ratio of 8:1:1 for training, validation, and test, that is, 189 pieces of data for training, 31 for validation, and 31 for testing. As shown in Table 3, there was no significant difference in terms of accuracy between Model_C and Model_D. However, Model_C was selected, as it had a lower learning complexity than Model_D.

Table 3. Comparison of the test results of each model.

Model	Test Video	Correct	Accuracy
A	31	25	0.80645
B	31	26	0.83870
C	31	27	0.87096
D	31	27	0.87096

The F1-score of the selected Model_C was measured by dividing it into precision and recall and then deriving the harmonic average value and specificity (see Figure 11).

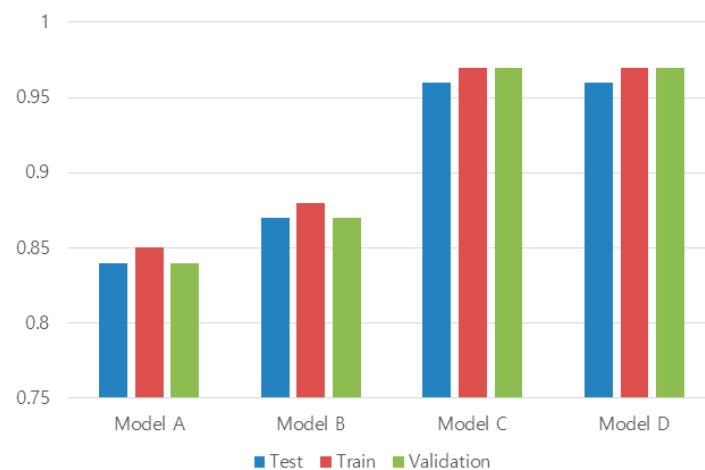


Figure 11. Accuracy of the training results for each model.

4.2. Simulation of Educational Content Applying the F1-Score and AI of the Selected Learning Model

Learning was categorized into 3 themes, and 12 learning behaviors were classified. Nature exploration was categorized into the learning behaviors of planting (N01), fruit picking (N02), harvesting (N03), and milking (N04). Emergency response was categorized into the learning behaviors of help (E01), extinguish (E02), check temperature (E03), CPR (E04), and escape (E05). Historical experience was categorized into the hurray (H01), draw (H02), and shoot (H03) learning behaviors. After coding the learning behaviors as above, the F1-score was obtained as shown in Table 4. After completing the testing of the learning model, the accuracy of the model was evaluated based on the F1-score evaluation index. The precision and recall were examined, and the harmonic mean and specificity were derived.

$$\text{Precision} = \frac{\text{Ture Positive}}{\text{True Positive} + \text{False Positive}} \quad (2)$$

$$\text{Recall} = \frac{\text{Ture Positive}}{\text{True Positive} + \text{False Negative}} \quad (3)$$

$$\text{F1 Score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

$$\text{Specificty} = \frac{\text{Ture Negative}}{\text{True Negative} + \text{False Positive}} \quad (5)$$

Table 4. F1-scores of the selected Model_C.

		Actual												
		N01	N02	N03	N04	E01	E02	E03	E04	E05	H01	H02	H03	
Predicted	N01	15	0	1	1	0	0	1	0	0	0	1	0	19
	N02	0	31	0	0	0	0	0	0	0	0	0	0	31
	N03	1	0	19	1	0	0	0	1	0	0	0	0	22
	N04	1	0	1	14	0	0	0	1	0	0	0	0	17
	E01	0	1	0	0	22	0	0	0	0	3	0	0	26
	E02	0	0	0	0	0	11	0	0	0	0	0	2	13
	E03	0	1	0	0	0	0	13	0	0	0	0	0	14
	E04	1	0	1	0	0	0	0	21	0	0	1	0	24
	E05	0	0	0	0	0	0	0	0	19	0	0	0	19
	H01	0	1	0	0	3	0	0	0	0	22	0	0	26
	H02	1	0	1	0	0	0	0	2	0	0	17	0	21
	H03	0	0	0	0	0	3	0	0	0	0	0	16	19
Total Count		19	34	23	16	25	14	14	25	19	25	19	18	251

Table 4 lists the F1-score results for the selected model. Table 5 lists the precision and recall values based on Table 4. Equations (2)–(5) were used to obtain the precision, recall rate, harmonic mean, and outliers, respectively. In this study, precision was measured at 0.875, recall at 0.872, harmonic mean at 0.873771, and singularity at 0.779428.

Table 5. Precision and recall results of the selected Model_C.

Precision	N01	0.789473684	Recall	N01	0.789473684
	N02	0.911764706		N02	1
	N03	0.826086957		N03	0.863636364
	N04	0.875		N04	0.823529412
	E01	0.88		E01	0.846153846
	E02	0.785714286		E02	0.846153846
	E03	0.928571429		E03	0.928571429
	E04	0.84		E04	0.875
	E05	1		E05	1
	H01	0.88		H01	0.846153846
	H02	0.894736842		H02	0.80952381
	H03	0.888888889		H03	0.842105263
	Average Precision	0.875019733		Average Recall	0.872525125
F1-Score		0.873771			

Figure 12 shows a scene in which five predicted behaviors most similar to a specific behavior were selected in real time using Model_C. Among the predicted behaviors that fluctuated in real time, the behavior with the highest probability was selected as the predicted behavior.



Figure 12. Top 5 behavioral predictions.

Figure 13 shows the screen that recognized that the educational activity had been completed according to the behavior prediction in Figure 12 and gave the learner permission to proceed to the next step. The user could proceed to the next training step by selecting the “Next Scene” option.

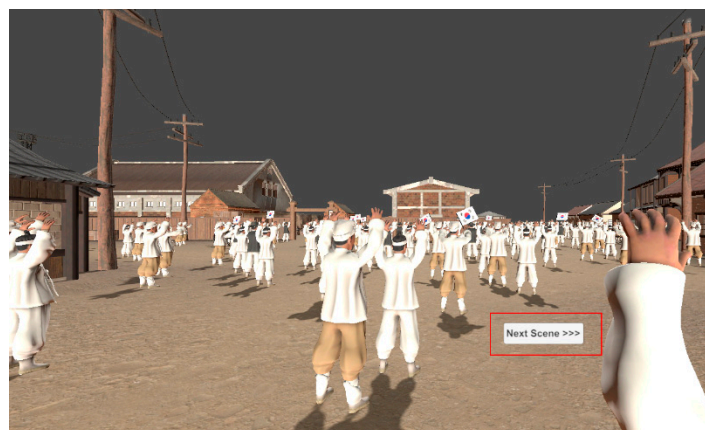


Figure 13. Giving learners the permission to carry on to the next lesson based on AI judgment.

5. Conclusions

With the increasing popularity of non-face-to-face education, educators and learners have become familiar with this learning process. However, certain disadvantages of non-face-to-face education have been pointed out, such as reduced learners’ concentration and participation in the learning. Metaverse-based education is active in solving the problem of reduced concentration, but improving the immersion for a learner with low control skills remains a challenge.

In this paper, an artificial intelligence system is proposed to address the issue of learners’ inability to proceed with education due to their lack of manipulation skills and the resulting decline in their immersion. The proposed system is a sustainable virtual reality educational content that can progress in education regardless of the level of manipulation proficiency by determining the learner’s engagement in educational activities. To do this, we first created VR educational content for learners’ learning activities. We then compared and selected four learning models to implement highly accurate AI. We simulated the chosen learning model to evaluate whether the learner was actively engaged in the educational process and had been granted the authorization to transition to the subsequent lesson phase. Consequently, with the aid of AI, we established an environment where even users with limited proficiency could effortlessly navigate a VR educational content. We anticipate that VR education, conducted in an environment devoid of elements that could diminish learners’ immersion, will culminate in enhanced learning outcomes.

This study has limitations regarding its generalization, as it was conducted within a predefined scenario with a limited set of educational activities. Moreover, to discern the diverse and numerous actions of learners, artificial intelligence requires a vast amount of training data with significant diversity. We plan to secure the data by extracting videos of various user behaviors and animations from virtual reality content engine programs from multiple perspectives. Additionally, we intend to expand our study to a continuous experimental environment where new models are designed, compared, and analyzed, aiming to enhance the system's reliability.

Author Contributions: Conceptualization, J.L.; Investigation, J.L.; Methodology, J.L.; Software, J.L.; Writing—original draft, J.L.; Writing—review and editing, Y.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lee, H.-Y. *Current Status and Implications of the Edutech Market*; Korea International Trade Association, IITTRADE FOCUS: Seoul, Republic of Korea, 2020; pp. 1–26.
2. Jinhaksa, 7 out of 10 High School Seniors Taking Online Classes, Decreased Focus and Quality of Education. Available online: <https://www.jinhak.com/IpsiStrategy/NewsDetail.aspx?ContentID=822785> (accessed on 30 June 2023).
3. Gye, B.-K.; Kim, H.-S.; Lee, Y.-S.; Son, J.-E.; Kim, S.-W.; Baek, S.-I. Analysis of experiences and perceptions of remote education in elementary and middle schools due to COVID-19. *KERIS* **2020**, *10*, 1–46.
4. Kang, Y. A study on problems of remote learning General English Class and improve plans in the age of COVID-19, Focusing on the core competence of D University. *J. Humanit. Soc. Sci.* **2021**, *12*, 1013–1022.
5. Noh, Y.; Lee, K.-K. A Study on Factors Affecting Learner Satisfaction in Non-face-to-face Online Education. *Acad. Cust. Satisf. Manag.* **2020**, *22*, 107–126. [\[CrossRef\]](#)
6. Lee, L.-H.; Tristan, B.; Zhou, P.; Wang, L.; Xu, D.; Lin, Z.; Kumar, A.; Bermejo, C.; Hui, P. All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda. *arXiv* **2021**, *14*, 1–66.
7. Hong, S.-C.; Kang, S.-H.; Ahn, J.-M.; Lim, S.-H. Gathertown-based metaverse campus and analysis of its usability. *J. Digit. Contents Soc.* **2022**, *23*, 2413–2423. [\[CrossRef\]](#)
8. Jeong, Y.; Lim, T.; Ryu, J. The effects of spatial mobility on metaverse based online class on learning presence and interest development in higher education. *Korea Educ. Rev.* **2021**, *27*, 1167–1188.
9. Lee, J.K.; Kim, Y.C. A study on the immersion evaluation system to increase the educational effect of educational smart contents. *e Bus. Stud.* **2020**, *21*, 85–98. [\[CrossRef\]](#)
10. Lee, H.S.; Park, M.S.; Han, D.S. A Study on VR Advertising Effects According to the Personal Characteristic of Consumer: Focus on the Mediating Effects of Presence. *J. Media Econ. Cult.* **2022**, *20*, 83–125. [\[CrossRef\]](#)
11. Sidik, F. Actualizing Jean Piaget's Theory of Cognitive Development in Learning. *Pendidik. Dan Pengajaran* **2020**, *4*, 1106–1111.
12. Kilag, O.K.T.; Ignacio, R.; Lumando, E.B.; Alvez, G.U.; Abendan, C.F.K.; Quiñanola, N.M.P.; Sasan, J.M. ICT Integration in Primary School Classrooms in the Time of Pandemic in the Light of Jean Piaget's Cognitive Development Theory. *Int. J. Emerg. Issues Early Child. Educ.* **2022**, *4*, 42–54. [\[CrossRef\]](#)
13. Lee, D.J.; Kim, M. University students' perceptions on the practices of online learning in the COVID-19 situation and future directions. *Multimed. Assist. Lang. Learn.* **2020**, *23*, 359–377.
14. Kyoung, J.H.; Cheol, K.D. The effects of physical environment on student's satisfaction and class concentration in college education services. *Korean Bus. Educ. Rev.* **2017**, *32*, 355–377.
15. Shen, B.; Tan, W.; Guo, J.; Zhao, L.; Qin, P. How to promote user purchase in metaverse? A systematic literature review on consumer behavior research and virtual commerce application design. *Appl. Sci.* **2021**, *11*, 11087. [\[CrossRef\]](#)
16. Han, S.-Y.; Kim, H.-G. User experience: VR remote collaboration. *SPRI Issue Report* **2021**, *IS-112*, 1–17.
17. Zhao, G.; Fan, M.; Yuan, Y.; Zhao, F.; Huang, H. The comparison of teaching efficiency between virtual reality and traditional education in medical education: A systematic review and meta-analysis. *Ann. Transl. Med.* **2021**, *9*, 1–8. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Balsam, P.; Borodzicz, S.; Malesa, K.; Puchta, D.; Tymńska, A.; Ozierański, K.; Kołtowski, Ł.; Peller, M.; Grabowski, M.; Filipiak, K.J.; et al. OCULUS study: Virtual reality-based education in daily clinical practice. *Cardiol. J.* **2019**, *26*, 260–264. [\[CrossRef\]](#)
19. Craddock, I.M. Immersive Virtual Reality, Google Expeditions, and English Language Learning. *Access. Technol. Librariansh. Lib. Technol. Rep.* **2018**, *54*, 7–9.

20. Della Corte, M.; Clemente, E.; Checcucci, E.; Amparore, D.; Cerchia, E.; Tulelli, B.; Fiori, C.; Porpiglia, F.; Gerocarni Nappo, S. Pediatric Urology Metaverse. *Surgeries* **2023**, *4*, 325–334. [\[CrossRef\]](#)
21. Ryu, I.-Y.; Ahn, E.-Y.; Kim, J.-W. Implementation of historic educational contents using virtual reality. *J. Korea Contents Assoc.* **2009**, *9*, 32–40. [\[CrossRef\]](#)
22. Ruan, B. VR-Assisted Environmental Education for Undergraduates. *Adv. Multimed.* **2022**, *2022*, 1–8. [\[CrossRef\]](#)
23. Tlili, A.; Huang, R.; Shehata, B.; Liu, D.; Zhao, J.; Metwally, A.H.; Wang, H.; Denden, M.; Bozkurt, A.; Lee, L.-H.; et al. Is Metaverse in Education a Blessing or a Curse: A Combined Content and Bibliometric Analysis. *Smart Learn. Env.* **2022**, *9*, 1–31. [\[CrossRef\]](#)
24. Lee, H.; Woo, D.; Yu, S. Virtual Reality Metaverse System Supplementing Remote Education Methods: Based on Aircraft Maintenance Simulation. *Appl. Sci.* **2022**, *12*, 2667. [\[CrossRef\]](#)
25. Yoon, J.; Yee, D. A Study of General Philosophy Education in the Metaverse -Focusing on the 'VRChat College' Case. *Korean J. Gen. Educ.* **2022**, *16*, 275–288. [\[CrossRef\]](#)
26. Baig, M.; Shuib, L.; Yadegaridehkordi, E. Big data in education: A state of the art, limitations, and future research directions. *Int. J. Educ. Technol. High. Educ.* **2020**, *17*, 1–23. [\[CrossRef\]](#)
27. Zhai, X.; Neumann, K.; Krajcik, J. Editorial: AI for tackling STEM education challenges. *Front. Educ.* **2023**, *8*, 1–3. [\[CrossRef\]](#)
28. Dimitriadou, E.; Lanitis, A. A critical evaluation, challenges, and future perspectives of using artificial intelligence and emerging technologies in smart classrooms. *Smart Learn. Environ.* **2023**, *10*, 1–26. [\[CrossRef\]](#)
29. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23 August 2017; IEEE: New York, NY, USA, 2017; pp. 1–6.
30. Song, W.Y.; Lee, K.W.; Park, S.C. Deep Learning based Pedestrian Tracking Framework using Body Posture Information. *Korean J. Comput. Des. Eng.* **2020**, *25*, 256–266. [\[CrossRef\]](#)
31. Sultani, W.; Chen, C.; Shah, M. Real-world anomaly detection in surveillance videos. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 6479–6488.
32. Mosavi, A.; Hosseini, F.S.; Choubin, B.; Goodarzi, M.; Dineva, A.A.; Sardooi, E.R. Ensemble boosting and bagging based machine learning models for groundwater potential prediction. *Water Resour. Manag.* **2021**, *35*, 23–37. [\[CrossRef\]](#)
33. Kumari, S.; Kumar, D.; Mittal, M. An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. *Int. J. Cogn. Comput. Eng.* **2021**, *2*, 40–46. [\[CrossRef\]](#)
34. Insafutdinov, E.; Andriluka, M.; Pishchulin, L.; Tang, S.; Levinkov, E.; Andres, B.; Schiele, B. ArtTrack: Articulated Multi-Person Tracking in the Wild. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1293–1301.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.