Research Article

Design and Implementation of Virtual Reality Interactive Product Software Based on Artificial Intelligence Deep Learning Algorithm

Yi Wang [iD] **and Junhai Sun** [iD]

School of Information and Engineering, Shanghai Zhongqiao Vocational and Technical University, Shanghai 201514, China

Correspondence should be addressed to Junhai Sun; 19440224@smail.cczu.edu.cn

The aim of this study is to improve the interactive needs of artificial intelligence in the virtual reality environment. Based on the in-depth study of the interactive needs of virtual reality, a virtual reality interactive glove based on nine-axis inertial sensor and realized by artificial intelligence deep learning algorithm is designed. The AI deep learning algorithms employed include the KNN, SVM, Fuzzy, PNN, and DTW algorithms. Static gesture recognition is relatively simple, dynamic gesture recognition needs to a dynamic real-time gesture sequence data starting point and end point planning, by building the directed graph structure, quickly retrieving the global optimal solution, and determining gesture starting point, with dynamic planning to solve the minimum distance between two points, avoid the graph, and improve efficiency. The results showed that by 50 gestures such as select object, attract object, zoom object, rotate object, shoot small box, exhale menu, and close the menu, the recognition rate is 100%, 94%, 96%, 100%, 92%, 100%, and 100%. The motion data of finger and palm are captured by nine axis sensor, and the gesture recognition is carried out by using artificial intelligence deep learning algorithm. It is proved that the artificial intelligence deep learning algorithm can effectively realize the design of virtual reality interactive product software.

1. Introduction

The artificial intelligence we often talk about is actually a process of simulating people's thinking, consciousness, and other contents. In other words, it is like having a machine that has a human brain that can think for itself and create for itself, and unlike the human brain, it has an advantage in speed and memory. Artificial intelligence is a very complex subject, just like the human body, with a complex structure and a wide range of contents, mathematics, computer science, computer programming, information theory, cybernetics, biology, mechanical automation, even psychology, and philosophy; it can be said that there are many different kinds of artificial intelligence, which is one of the reasons why there are so many researchers in artificial intelligence. However, there is no denying that China's artificial intelligence technology is still in the development stage, and there

are still many technologies to be broken through. The emergence and development of artificial intelligence have brought great changes to our life, work, and study; we can even see a lot of content about artificial intelligence around us, such as robots and mechanical arms in factories, all belonging to the category of artificial intelligence. Artificial intelligence is a new technology science to research and develop theories, methods, technologies, and application systems used to simulate, extend, and expand human intelligence. As an interdisciplinary discipline, philosophy and cognitive science, mathematics, computer science, biology, psychology, information theory, and so on are all the research categories of artificial intelligence.

Therefore, domestic and foreign research studies continue. Combined with the characteristics of virtual reality interactive classroom, Wilson et al. put forward a design scheme of virtual reality interactive classroom based on the

deep learning algorithm [1]. Li et al. discussed the elements of the virtual reality-based deep learning model, basic model construction method, and deep learning path design, providing model reference for institutions and personnel engaged in the field of deep learning [2]. Kwon et al. first introduced the workflow of DBN, a deep learning algorithm, and summarized the computational characteristics of this algorithm. Assembly language is used based on instruction set to convert classification functions into assembler and evaluate program performance. Secondly, based on the Hadoop ecosystem, the application of the BDMISS system in the sharing of big data medical information resources is analyzed [3]. With the development of the virtual reality industry, the major smart device manufacturers have launched different virtual reality interaction handles in the virtual reality interaction. Although virtual reality interactive gamepads can provide good user experience at present, the problems existing in virtual reality gamepads, such as unnatural touch and insufficient simulated gestures, cannot meet the needs of virtual reality for high immersion [4]. With the development of virtual reality technology, human-computer interaction technology of virtual reality requires high precision capture and high simulation of hand movements.

Based on the current research, this paper proposes the design and implementation of virtual reality interactive product software based on the artificial intelligence deep learning algorithm; with the development of virtual reality technology, the traditional computer mouse and keyboard are not suitable for the interactive needs of virtual reality environment; the need for virtual reality interaction is becoming more and more prominent. Based on the in-depth study of the interactive needs of virtual reality, a virtual reality interactive glove based on nine-axis inertial sensor and realized by the artificial intelligence deep learning algorithm is designed. The finger and palm movement data are captured by nine-axis sensors, using the artificial intelligence deep learning algorithm for gesture gesture recognition, with the help of the supporting software development kit (SDK) software, to realize the synchronous restoration of the complete motion process of the hand part in the virtual reality environment, as well as gesture interaction intention recognition, to achieve the purpose of interaction through natural hand movements in the virtual reality environment. It is proved that the artificial intelligence deep learning algorithm can effectively realize the design of virtual reality interactive product software.

2. Materials and Methods

2.1. Deep Learning Algorithm Model.

Typical deep learning models include Convolutional Neural Network (CNN), Deep Belief Networks (DBNs), and Stacked Auto Encoder (SAE) models, which are described as follows:

(1) The ownership values of the convolutional neural network are initialized with different small random numbers before training, and supervised training is performed. The training of a convolutional neural network has two stages:

(a) *Forward Propagation Stage*. Extract a sample (X, Y_P) from the sample set, input X to the network, W is the weight of the network, F is the mapping function, the information is transmitted from the input layer to the output layer through a first-level transformation, and the corresponding actual output is calculated as shown in the following formula:

$$O_P = F_n\left(\cdots \left(F_2\left(F_1\left(XW_1\right)W_2\right)\cdots\right)W_n\right). \quad (1)$$

(b) In the backward propagation stage, the difference between the actual output O_P and the ideal output Y_P is calculated, as shown in the following equation:

And the weight matrix is adjusted in a way to minimize the error. In order to avoid the display and extraction of features, the feature detection layer of the convolutional neural network learns according to the training data, so that the features are implicitly extracted, and compared with other neural networks, since the weights of neurons are the same, in a feature map, parallel learning of the network is performed on the surface. Typically, the wavelet transform uses a discrete form. Considering that we are mainly interested in a continuous signal with a resolution of 2, it can be defined by $a = 2^m$ and $n2^m$ as a discrete wavelet transform, where m and n are integers. Then, the decomposition formula of the wavelet is shown in the following equation:

$$f(t) = \sum_{m,n} c_{m,n}\psi_{m,n}(t). \quad (2)$$

Among them, $\psi_{m,n}(t) = 2^{-m/2}\psi(2^{-m}t - n)$, ψ has a very special form, so that $\psi_{m,n}(t)$ constitutes a standard orthonormal basis, as shown in the following equation:

$$c_{m,n} = \langle f, \psi_{m,n}\rangle = \int \psi_{m,n}(t)f(t)\mathrm{d}t. \quad (3)$$

(2) As a probabilistic generative model, DBNs can treat a directed graph of multiple random variables as a superposition of multiple RBMs. In DBNs, the RBM is used at the layer far from the display layer, while the Bayesian belief network is used at the part close to the display layer. The adjustment uses the training result of each layer of RBM as the input of the next layer until all RBMs are trained. After the training is completed, the BP algorithm is used to fine-tune backwards according to the sample label value. The specific process can be divided into two steps:

(a) The first step is pretraining, initializing the parameters of the deep belief network at the beginning, then training the RBM of each layer

separately, using the output of each layer as the input of the next layer, and retaining each layer to get the weight W_{ij}.

(b) The second step is the fine-tuning stage. After the first step is completed, the entire network is adjusted using the gradient downward algorithm and the sample label value. At this time, the method adopted by DBNs is similar to the BP algorithm. The weights of each layer are only trained in advance, not randomly initialized, so only a small number of iterations can get better results.

Both in terms of algorithm efficiency and time, DBN algorithms have very good results, especially in big data. At present, the DBN algorithm has been widely used in speech, image, and document, and the research on its theory and model has also increased.

(3) The stacked self-encoding network is composed of stacks of some structural units, which is similar to DBN, except that the stacked self-encoding network uses an auto encoder (AE) as a structural unit. In the AE model, the encoder $c(\cdot)$ encodes the input x into the representation $c(x)$ and then uses the decoder $g(\cdot)$ to decode from the $c(x)$ representation to obtain the reconstructed input equation as follows:

$$r(x) = g(c(x)). \tag{4}$$

The AE model is trained by minimizing the reconstruction error $L(r(x), x)$ when the reconstruction error is as shown in the following equation:

$$L(r(x), x) = \|r(x) - x\|^2. \tag{5}$$

And when the hidden layer is linear, the $c(x)$ training network projects the input into the data principal component space. At this time, the role of the AE model is equivalent to the principal component analysis method; when the hidden layer is nonlinear, the obtained representation can be stacked into multiple layers and the AE model can obtain the multimodal input distribution. Given $c(x)$, the mean square error criterion is extended to the case where the reconstructed negative log-likelihood function is minimized, as shown in the following equation:

$$RE = -\log p(x|c(x)). \tag{6}$$

Since the sparse term in the energy function has a fixed representation and the generated features have stronger geometric transformation invariance, when the input x_i of each layer is binary, the resulting loss function is shown in the following equation:

$$-\log p(x|c(x)) = -\sum_i \log g_i(c(x)) + (1 - x_i)\log(1 - g_i(c(x))). \tag{7}$$

2.2. Embedded Hardware System Design

The virtual reality interactive glove developed in this project uses STM32CortexM4 high-speed processor, the operating frequency of the system up to 168 MHz meets the processing requirements of the high-speed motion capture algorithm, which is used to collect the data of the nine-axis inertial sensor on the finger and the back of the hand, and the nine-axis inertial sensor adopts Bosch high-performance motion sensor and high-speed motion capture frequency of 400 kHz/s; Bluetooth communication module adopts BLUEMOD SR high-speed Bluetooth module, which can provide the communication rate of 1.3 Mbps. The whole hardware system can meet the real-time and effectiveness of virtual reality hand interactive data communication. Due to the need to collect the data of six nine-axis sensors, the hardware system uses its own independent hardware bus module for data acquisition of each nine-axis sensor, the system uses three interintegrated circuits (IICs) and three serial peripheral interfaces (SPIs) of CortexM4 CPU to connect six nine-axis sensors, respectively. The whole hardware platform system includes CortexM4 CPU, Bluetooth communication module, power module, each finger nine-axis sensor, and the nine-axis sensor on the back of the hand (see Figure 1) [5].

2.2.1. CortexM4 CPU

The CPU of CortexM4 mainly completes the initialization of each module of the system, the data acquisition of each finger nine-axis sensor, and the nine-axis sensor on the back of the hand, as well as motion posture algorithm processing, Bluetooth wireless communication, watchdog system protection, system error detection and reporting. The initialization of each module also includes the processor itself hardware IIC module initialization, SPI module initialization, universal asynchronous receiver/transmitter, UART module initialization, timer initialization, GPIO initialization, AD module initialization, and NVROM module initialization; it also includes the initialization of each nine-axis sensor and Bluetooth module [6].

2.2.2. Bluetooth Communication Module

The Bluetooth communication module is connected with the CPU of CortexM4 through the UART interface to realize the wireless data communication between the Virtual reality interactive glove and the upper-layer application software. The Bluetooth communication module receives the communication commands of the upper-layer application software and sends them to the CPU and receives the motion attitude data packaged by the CPU and sends them to the upper-layer application software.

2.2.3. Power Module

The power module is designed with switching power supply. The 3.7 V lithium battery is used to convert the 3.7 V power supply into a stable 2.8 V power supply required by the system. The PSU supports a wide input voltage from 2.9 V to 7.8 V to a stable 2.8 V voltage.
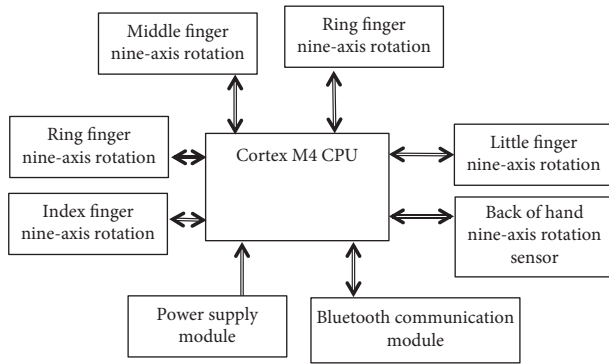
FIGURE 1: System framework of hardware platform of virtual reality interactive glove.

2.2.4. Nine-Axis Finger Sensor. The virtual reality interactive gloves developed in this project include five finger nine-axis sensors and nine-axis sensors on the back of the hand; the nine-axis sensors adopt Bosch high-performance motion sensors; the other three finger nine-axis sensors are connected to the CPU through the IIC interface; and the other two finger and hand back nine-axis sensors are connected to the CPU through the SPI interface. A three-axis acceleration sensor, a three-axis gyroscope, and a three-axis magnetometer sensor are integrated in a single nine-axis sensor chip to capture the movement process data of each finger and hand.

2.3. Embedded Firmware Design. The design of virtual reality interactive glove firmware adopts multithread management and collaborative processing mechanism; data acquisition, data processing, and data communication are independently processed; semaphore and event events are used between multiple threads to ensure effective synchronous data transmission. Among them, the main thread completes the initialization of the hardware system and obtains the data processed by the sensor data acquisition subthread; the motion attitude algorithm was processed to generate the overall motion model data of the hand and packet the data [7].

2.4. Motion Attitude Detection Principle of Nine-Axis Sensor. Attitude refers to the pitch, roll, and heading of a moving object in the process of motion. On Earth, attitude generally refers to the pitch, roll, and heading of the vehicle in the Earth coordinate system. Generally speaking, in the motion control of aircraft, only knowing its real-time attitude can control its next action according to the need. Attitude is used to describe the angular position relationship between the fixed frame and the reference frame of a rigid body, generally using Euler angles, quaternions, matrices, or axis angles as mathematical representations; using different representations in different fields, each representation has its own advantages. The representation methods used in this project are Euler angle and quaternion [8].

The Euler angle is used to describe the orientation of a rigid body in a three-dimensional Euclidean space. For a

reference frame in 3D, the orientation of any coordinate system can be represented by three Euler angles. The reference system, also known as the laboratory reference system, is stationary; the coordinate system is fixed to the rigid body and rotates as the rigid body rotates.

As shown in Figure 2, the fixed coordinate system *oxyz* is made from the point *o* and the coordinate system *ox′y′z′* is fixed to the rigid body. With the axes *oz* and *oz′* as the basic axes and the vertical planes *oxy* and *ox′y′* as the basic planes, the angle θ from the axis *oz* to *oz′* is called the nutation angle. The vertical line *oN* of the plane *zoz′* is called the nodule line, which is also the intersection of the basic planes *ox′y′* and *oxy*. In the right-handed coordinate system, angle β should be measured counterclockwise when viewed from the positive end of *oN*. The angle α measured from the fixed axis ox to the pitch line *oN* is called the precession angle, and the angle δ measured from the pitch line *oN* to the moving axis *ox′* is called the rotation angle. From the positive end of *oz* axis, angle α is measured counterclockwise. From the positive end of *oz′* axis, angle δ is also measured counterclockwise [9].

Quaternions are widely used for computer drawing (and related image analysis) to indicate the rotation and orientation of 3D objects. Quaternions are also used in cybernetics, signal processing, attitude control, physics, and orbital mechanics, to represent rotation and orientation.

The automatic heading reference system for nine-axis sensor motion attitude detection is "automatic heading reference system." The AHRS algorithm uses quaternions for pose solution and is widely used in quarticopters.

The control of motion is not involved in this project, but the pitch, roll, and heading of the nine-axis sensor are utilized to obtain the attitude of the nine-axis sensor in real-time. The position of the finger during the motion is simulated by restoring the finger position of the nine-axis sensor, and a number of nine-axis sensors are used to jointly complete the simulated reduction of the complete hand motion posture.

2.5. Finger Movement Detection Scheme. This project adopts a wearable design scheme, which sets a nine-axis sensor on the back of each finger and the back of each hand. Among them, the nine-axis sensor of the thumb back is arranged in the phalanx of the thumb tip, and the nine-axis sensor of the other four fingers is arranged in the middle phalanx, and the nine-axis sensor on the back of the hand is located in the middle of the back of the hand. Since the finger has multiple phalanges, and in this project, the motion posture of the whole finger is captured by a nine-axis sensor; therefore, it is also necessary to study the characteristics of flexible movement of fingers and establish a multidegree of freedom mathematical model for each finger, including the differences in the motion characteristics of several phalanxes of each finger during the movement process, and the correlation characteristics of the phalanx where the nine-axis sensor is located with other phalanx movements during the movement of the finger [10].
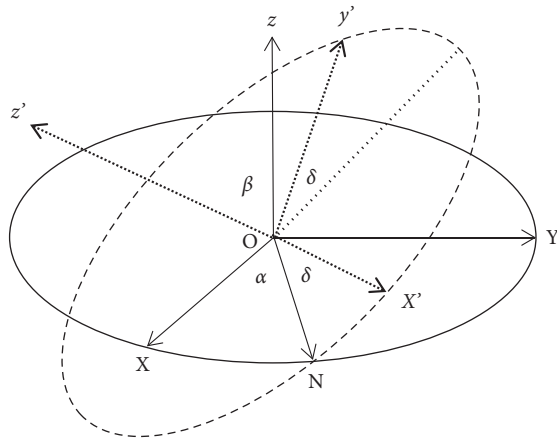
FIGURE 2: Euler angle rotation coordinate system.

FIGURE 3: Dynamic gesture recognition process of the DTW algorithm.

The three phalanxes of the index finger were numbered as bone 0, bone 1, and bone 2, and the three nine-axis sensors were worn on the back of the fingers of bone 0, bone 1, and bone 2, respectively; a large number of data of the entire bending process of the same finger were collected to obtain the movement characteristic curve model of each bone. Taking bone 1 as a reference, the motion characteristics of bone 0 and bone 2 can be used to establish the mathematical model of motion. Let the motion of bone 1 be x, and the mathematical model of bone 0 and bone 2 is established; the motion characteristics of bone 0 conform to the following mathematical model: $y = A - \mathrm{Tan}h(B - K^*x)/C$ The motion characteristics of bone 2 conform to the following mathematical model: $y = (e(a + b^*x) - K)/(1 - K)$. The same method as above can be used to build the corresponding mathematical model of other fingers.

3. Results and Discussion

AHRS attitude solution makes full use of the magnetic field and gravity field to correct the gyroscope drift, but AHRS also has the problem of relying heavily on the relationship between the magnetic field and gravity field; AHRS will not work properly without gravity and magnetic fields. In general, the more orthogonal the magnetic field and gravity field, the better the attitude measurement effect. And if the magnetic field and the gravitational field are parallel, like the magnetic field and the weight field are in the same direction at the north and south of the geomagnetic field, at this point, the course angle cannot be measured, and the course angle error will become larger and larger at high latitudes, which is the defect of AHRS [11].

In the process of AHRS attitude calculation, geomagnetic data are ignored when the local magnetic field is 0 in the three-axial directions. When gravity is 0 in all three axes, the AHRS attitude solution ignores gravity data. This project refers to the AHRS algorithm proposed by Sebastian Madgwick and uses the gradient descent method to solve the attitude of quaternion. First, the acceleration and magnetic intensity are normalized to simplify the subsequent floating-point calculation, and the characteristics of the Earth's magnetic field pointing to the true north are used, the gravity
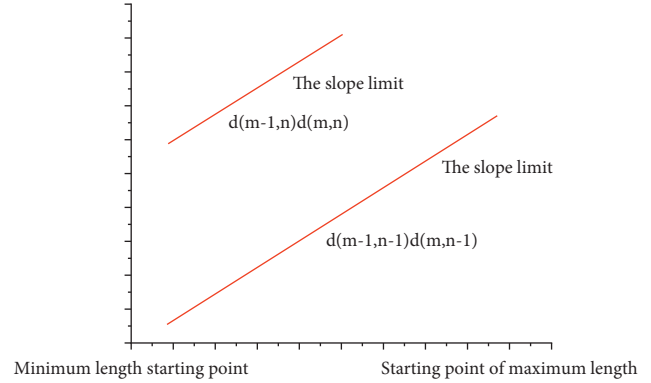
and magnetic flux errors were estimated, and the PROPORTION integration differentiation (PID) error compensation was made for the gyroscope through the gravity and magnetic intensity data; finally, the attitude fusion algorithm of the quaternion gradient descent method is realized.

The artificial intelligence deep learning algorithm is used for gesture recognition application, which is divided into two steps: training learning and recognition. Step 1: virtual reality interactive gloves are used to collect user gesture data, the artificial intelligence deep learning algorithm was used to train and learn the collected gesture data, and the gesture recognition mathematical model library was generated. Step 2: the gesture recognition mathematical model library generated in step 1 is put into gesture recognition application software, the real-time gesture data of users is collected through virtual reality interactive gloves, the real-time gesture data are sent to the gesture recognition mathematical model library interface, and real-time gesture recognition results can be obtained. In the application implementation of gesture recognition using the artificial intelligence deep learning algorithm, it is divided into static gesture recognition and dynamic gesture recognition; the artificial intelligence deep learning algorithms adopted include KNN algorithm, SVM algorithm, Fuzzy algorithm, PNN algorithm, and DTW algorithm. Static gesture recognition is relatively simple, while dynamic gesture recognition needs to determine the starting point and the end point of a dynamic real-time gesture sequence data. By constructing a directed graph structure, the global optimal solution is quickly retrieved and the starting point of gesture is determined. The minimum distance between two points of the graph is solved by dynamic programming to avoid traversal of the graph and improve efficiency. The process of using the DTW algorithm to recognize dynamic gestures is shown in Figure 3.

In order to verify the implementation effect of this project, a supporting SDK software is developed based on C++, which is used to provide third-party developers with the development of virtual reality application based on the project's virtual reality interactive gloves; 3D gesture restoration software and gesture recognition software were also developed based on QT and Unity 3D to observe the gesture

TABLE 1: Test results of virtual reality gesture recognition.

Hand gestures	Number of repeated tests	Number of successful identification	Recognition (%)	Description of test results
Ray selected object	50	50	100	The rays are straight, and the sensory needs to adjust the rays in the development of virtual reality content effect
Draw the object	50	47	94	
Object scale	50	48	96	You need both hands to be standard to trigger
Rotating	50	50	100	
Shoot out cubes	50	46	92	You need to be good with your hand gestures, and you need to do a standard trigger gesture
Breathe out menu	50	50	100	
Close the menu	50	50	100	The thumb bend is critical. When different people test, the thumb bend is too low, and there is a time when it is triggered by mistake

restoration effect and gesture recognition effect in real-time (see Table 1).

4. Conclusions

This paper proposes the design and implementation of virtual reality interactive product software based on the artificial intelligence deep learning algorithm. Through the hardware system design of virtual reality interactive glove, embedded system design, C++ SDK interface design, and artificial intelligence deep learning algorithm research and design. The experimental results show that the virtual reality interactive product based on the nine-axis sensor can realize gesture motion capture and gesture recognition, and use gestures 50 times to indicate the following: select an object, attract an object, zoom the object, rotate the object, shoot a small square, call out the menu, and close the menu. The recognition rate is 100%, 94%, 96%, 100%, 92%, 100%, and 100%. The finger and palm movement data are captured by nine-axis sensors, using the artificial intelligence deep learning algorithm for gesture recognition, with the help of the supporting software development kit (SDK) software, in the virtual reality environment; the complete motion process of the hand is restored synchronously, and the intention of gesture interaction is recognized, so as to achieve the purpose of natural hand interaction in the virtual reality environment. It is proved that the artificial intelligence deep learning algorithm can effectively realize the design of virtual reality interactive product software. In gesture motion capture, its lack of spatial positioning function can only move around a reference point; if you rely on the product's own inertial sensor for spatial positioning and movement calculation, there will be a very big error, so in the future project, products in virtual reality application need a third-party positioning equipment for auxiliary spatial positioning for further research.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare there are no conflicts of interest.

References

[1] J. C. Wilson, S. Nair, S. Scielzo, and E. C. Larson, "Objective measures of Cognitive Load Using Deep Multi-modal Learning," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 1, pp. 1–35, 2021.

[2] D. Li, Y. Zhao, M. Luo, and R. Cui, "Design and implementation of a course answering system based on error correction and trie tree language model," *Journal of Physics: Conference Series*, vol. 1621, no. 1, Article ID 012057, 2020.

[3] Y. C. Kwon, J. W. Jang, H. Lim, and O. Choi, "Feasibility analysis of deep learning-based reality assessment of human back-view images," *Electronics*, vol. 9, no. 4, p. 656, 2020.

[4] H. Zhang, "The literature review of action recognition in traffic context," *Journal of Visual Communication and Image Representation*, vol. 58, pp. 63–66, 2019.

[5] G. Yang, D. Liu, and C. Ma, "Design and implementation of roadside intelligent information interaction system based on edge computing," *Journal of Physics: Conference Series*, vol. 1486, no. 2, Article ID 022022, 2020.

[6] F. M. Asch, T. Abraham, M. Jankowski et al., "Accuracy and reproducibility of a novel artificial intelligence deep learning-based algorithm for automated calculation of ejection fraction in echocardiography," *Journal of the American College of Cardiology*, vol. 73, no. 9, p. 1447, 2019.

[7] C. Bi, X. Fang, Z. Geng, and L. Dong, "Design and implementation of cancer patient survival prediction model based on ensemble learning method," *E3S Web of Conferences*, vol. 271, no. 3, Article ID 04030, 2021.

[8] Y. Fan, "Criminal psychology trend prediction based on deep learning algorithm and three-dimensional convolutional neural network," *Journal of Psychology in Africa*, vol. 31, no. 3, pp. 292–297, 2021.

[9] H. Yu, H. Gong, and Y. Wang, "Design and implementation of fault injection based on abstract syntax tree of c program," *IOP Conference Series: Materials Science and Engineering*, vol. 715, no. 1, Article ID 012034, 2020.

[10] S. Ahmad, S. A. Salman, and A. A. Abu Ebayyeh, "Design and implementation of education and training graphical user interface (gui) based on ni labview for the festo mps pa compact workstation," *International Review of Automatic Control (IREACO)*, vol. 12, no. 2, p. 67, 2019.

[11] T.-Y. Heo, K. M. Kim, H. K. Min, S. M. Gu, J. H. Kim, and J. Yun, "Development of a deep-learning-based artificial intelligence tool for differential diagnosis between dry and neovascular age-related macular degeneration," *Diagnostics*, vol. 10, no. 5, p. 261, 2020.