

Different Types of Neural Networks and Applications: Evidence from Feedforward, Convolutional and Recurrent Neural Networks

Yumin Pan *

Sino-Canada School No.1 Kangli Road Lili Town, Wujiang Suzhou 215212, China

* Corresponding Author Email: 2018010956@student.cup.edu.cn

Abstract. Neural networks have achieved great process in the 90 years since they were officially introduced in 1943. Because of its wide application and huge research and development potential, this technology attracts more and more scientific and technological workers to the research of neural networks. Neural network technology is an essential component of AI development, and it is a significant indicator of a country's overall strength. In this paper, this study will demonstrate Feedforward Neural Network, Convolution Neural Network and Recurrent Neural networks and evaluate them through datasets from kaggle.com. and CSDN (China IT community). Through this paper, readers can have a better outlook and understanding of the operating principles of each type of neural network as well as their specific jobs (what kind of jobs they specialized in) and each application of these neural networks. So that this paper can promote readers' thoughts and help them start learning neural networks or be a supplement or reference for future scholars. In the end, this paper will present the outcome, which is the evaluation of the accuracy, loss curve, and accuracy curve of neural networks.

Keywords: Computer science, Artificial Intelligence, Machine learning, Neural Network.

1. Introduction

A neural network is a collection of interconnected nodes, units, and simple processing elements whose function is roughly based on an animal's neurons [1]. In other words, is a machine-learning algorithm that mimics the way biological neurons transmit signals to each other.

The period from 1943 to 1949 was the period of initial attempts at neural networks. The establishment of this field was marked by the development of the first neural network models in 1943. The year between 1969 and 1974 is a period that the development of neural networks suffered most of the difficulties and setbacks when exploiting this new field. It is not until the year 1980 is the time for innovation of neural networks. Nowadays, is a transformational era for neural networks and this field has made significant progress. For example, the research team Schmidhuber invented deep feedforward neural networks and recurrent neural networks. [2].

In the fields of classification, clustering, model recognition, and prediction, neural networks have been widely utilized, which makes neural networks provide effective support in dealing with complex and non-complex problems in many fields, which greatly helps the work efficiency of industries and facilitates people's life as well [3].

The paper's purpose is to introduce neural networks and the principles behind three different types of them. that let readers deeply understand what is neural networks. Through the real cases of neural networks, this will stimulate readers to think about the importance of neural networks to all aspects of human development. In this paper, the literature review section will review the important research and development in this field in the past and different types of neural networks are being utilized in different fields. In methodology, it will demonstrate each of the three types of operating principles and the neural networks application. The experiment will provide the information on experiment in detail such as the evaluation tool, the dataset, the process of the experiment, and some outcome demonstration.

2. Literature review

Initially, the concept of neural networks arose independently. But with the pages of time, the emergence of new backgrounds and new disciplines has made the natural combination of neural network technology and other technologies form a set of computational processes with solid theories that are highly efficient in solving practical problems in various fields. Today neural networks are no longer considered a separate field of study but are part of new contexts and new disciplines [4].

The perceptron, proposed by Rosenblatt in 1958, was the basis for neural networks and support vector machines [2]. David Rumelhart, Geoffrey Hinton, and Ronald Williams proposed the backpropagation algorithm in 1986, which was a method that was effective in training multi-layer neural networks and enabled them to solve more complex non-linear problems. In 1997, Sepp Hochreiter and Jurgen Schmidhuber proposed a short-duration memory network (LSTM), an improved recurrent neural network (RNN) that solved the problem of disappearing gradients and exploding gradients. In natural language processing, speech recognition, and other fields, LSTM is widely utilized. Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton proposed AlexNet in 2012, which was a model that had a multi-layer convolutional neural network architecture was the winner of the ImageNet Large-scale Visual Recognition Challenge in 2012. AlexNet's success has sparked a revolution in deep learning. These are several groundbreaking inventions in the field of neural networks, which have laid the foundation for the development of neural networks and continue to promote technological innovation in such an area.

In the domain of mental health, Psychologists can use neural networks to help predict and evaluate the health of prisoners in prison and identify the factors that affect the mental health of prisoners. In this way, prisons can build a healthy intervention environment based on the results obtained, so that prisoners can be more effective in rehabilitating to increase the social workforce [5]. In business, a neural network can predict the likelihood of a company going bankrupt based on a series of financial ratios that determine the 'health' of the company. This involves using neural networks to discover the weaknesses of the company being tested and the factors that could cause the company to fail [6]. In agriculture, a wide variety of independent variables based on neural networks make it possible to forecast the results of agricultural production, disease and pest verification, smart weed management, and crop quality grading. [7].

3. Methodology

The aim of this main focus of this study is through the demonstration of the working principles and application cases of each three types of neural networks to let readers have a deeper understanding of what are they and the importance of neural network applications in real life as well. The data which support the paper are mainly from references from three educational websites: Google Scholar, Baidu Scholar, and cnki. This paper is also included some information and data provided by experts and professors in this area as a supplement. In this paper, the method to classify the types of neural networks is based on the network structure. As a result, this study has taken three of them as illustrations. They are FNN, CNN, and RNN.

3.1. Feedforward Neural Networks (FNN)

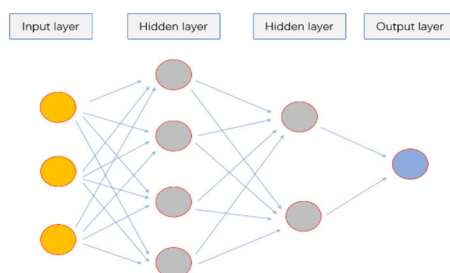


Fig. 1. Feedforward Neural Networks' layout (Photo/Picture credit: Original)

As shown in Fig.1, One of the most basic neural network models is a Feedforward Neural Network (FNN), also known as Multiple-Layer Perceptron (MLP). It is trained with backpropagation learning algorithms and is of the most popular neural network available today. It consists of multiple neurons, each of which receives the output from the neurons of the previous layer, and is weighted and processed by certain weights and biases to finally obtain the output of the neurons of the current layer, which in turn serves as the input of the neurons of the next layer. The network information stream is unidirectional and can only move from the input layer to the output layer, so it is called a feedforward neural network.

The feedforward neural network is usually composed of an input layer, several hidden layers and an output layer, as shown in Fig.1, when the input layer receives the original data input, the hidden layer transforms it and extracts the feature to a certain extent., and the hidden layer's results are used to predict the model's value in the output layer [8]. Each layer consists of more than one neuron, and the connections between the neurons are weighted, which can be learned to optimize through backpropagation algorithms. Feedforward neural networks are powerful models that are capable of handling nonlinear classification and regression tasks.

FNN is an artificial neural network in which the nodes do not form loops. FNN that has hidden layers can be trained to perform and solve any classification and regression problems [9]. Application cases include image classification and credit scoring.

Image classification: The input layer's node is determined by the image pixel value in the image classification task, then activation functions are used to pass the signal to the hidden layer (such as ReLU, Sigmoid, etc.) and connection weights. There can be multiple hidden layers, and each hidden layer can learn different feature representations. Finally, the output layer uses classifiers, such as Softmax, to map features onto predefined category labels, enabling image classification.

Credit scoring: In the credit scoring task, FNN can make risk predictions based on various attributes of the customer, such as income, credit history, etc. These attributes are used as input layer nodes, feature extraction, and representation learning are carried out through multiple hidden layers, and finally, the predicted default probability is given by the output layer. According to this probability, the financial institution can give the customer a credit rating and decide whether to approve the loan.

3.2. Convolutional Neural Networks (CNN)

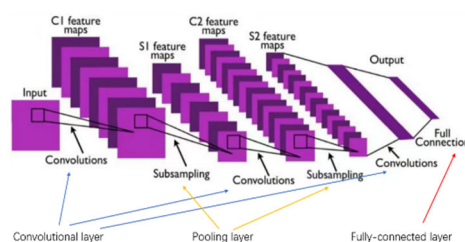


Fig. 2. Convolutional Neural Networks' layout (Photo/Picture credit: Original)

As shown in Fig.2, feedforward neural networks that have a deep structure and include convolutional computation are known as convolutional neural networks (CNNs). Networks of convolutional neurons have been proposed by the mechanism of the Receiving Domain in biology. Convolutional neural networks are designed to process data that has a grid-like structure. For instance, a one-dimensional grid formed by regular sampling on the timeline is the form of time series data and image data is viewed as a grid of pixels in two dimensions.

In Fig.2, The convolutive layer will dictate the output of the neuron and connect it to the local by calculating the scalar product between its weight and the areas that are connected to the input volume [10].

In Fig.2, the pooling layer is primarily responsible for reducing the size of the feature map, which in turn reduces the amount of calculations and parameters, and enhancing the effectiveness of model training. The functions of the pooling layer include: 1. Dimension reduction: Pooling reduces the

height and width of the feature graph and reduces the computational complexity. This helps to alleviate the overfitting phenomenon and improve model generalization ability. 2. Translation invariance: Pooling operations can provide translation invariance to some extent. When the object in the image is shifted slightly, the pooled feature map can still retain the key information. 3. Feature selection: The pooling operation can filter features to retain significant features. For example, maximum pooling preserves the maximum value within each local area to highlight strong features; Average pooling calculates the average within a local area, helping to preserve the smoothing feature.

In Fig.2, in the latter layers of CNN, the fully connected layer is often used to integrate the characteristics extracted by the convolution and pooling layers and carry out the final classification or regression task. The functions of the fully connected layer include: 1. Feature integration: The fully connected layer incorporates the local characteristics of the convolution and pooling layers into a global feature vector to capture the higher-order relationships between features. 2. Classification or regression: The fully connected layer can be used to implement the final classification or regression task. The classification task commonly uses the Softmax activation function in the output layer to map the feature vector to the probability distribution of each class. In a regression task, a linear activation function or other appropriate activation function can be used to predict continuous values. 3. Nonlinear mapping: Activation functions in the fully connected layer (such as ReLU, Sigmoid, etc.) introduce nonlinear mapping into the network to improve the expressiveness of the model.

CNN is a unique feedforward neural network that has a local connection, weight sharing, and pooling, which is suitable for processing data with grid structure (such as images and speech signals). Application cases include Computer vision and analysis of medical imagery.

Computer vision: In computer vision tasks, CNN extracts image features and classifies them through convolution layers, pooling layers, and fully connected layers. The layer of convolution operates a convolution kernel to slide over the input image to capture local features. The pooling layer is used to decrease the size of the characteristics map, decrease the calculation, and enhance the model's robustness. The fully connected layer maps convolution and pooled features to class labels. Through this layered approach, CNNs can automatically learn advanced features in images and achieve accurate image recognition.

Medical image analysis: CNN can be used for the automatic diagnosis of medical images. First, medical images (such as X-rays, CT, MRI, etc.) are fed into the CNN. Then, image features are extracted and classified through multiple convolutions, pooling, and fully connected layers. Finally, the output layer gives diagnostic results, such as the presence of the tumor and the location of the lesion [11].

3.3. Recurrent Neural Networks (RNN)

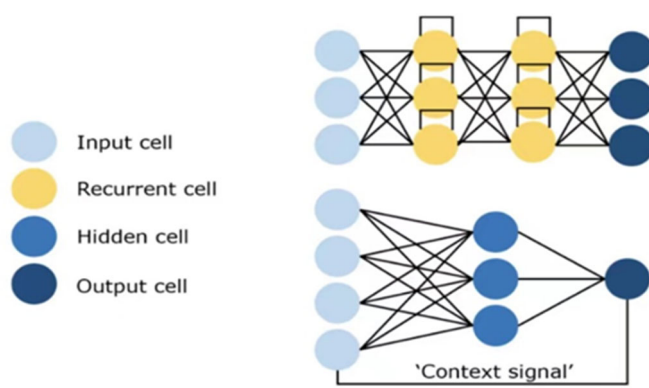


Fig. 3. Recurrent Neural Networks' layout (Photo/Picture credit: Original)

A network that sends feedback signals to each other is called a recurrent neural network (RNN) [12]. Processing time series data requires a specific type of neural network. The main feature of RNN is the existence of cyclic connections in the network, which allows the network to remember previous information. Here are the main components of RNN and what they do:

Input cell: In Fig.3, the input cell is responsible for receiving timing data. When processing sequential data (such as text, time series, etc.), the network receives data from the input cell for each step. The input cell can be a simple vector or a more complex structure, such as the Embedding Layer, which maps discrete data (such as words) to a continuous vector space.

Hidden cell: In Fig.3, the hidden cell is the core part of the RNN and contains Recurrent Units. The masked layer is informed of the input data of the current temporal step and the masked state of the previous temporal step at each temporal step. The hidden state contains information about previous time steps, giving the RNN memory power. The cyclic units in the hidden cell can adopt different structures, such as simple RNN units, Long Short-Term Memory units (LSTM), or Gated Recurrent units (GRU). The main functions of the hidden cell are: 1. Learning and preserving long-term and short-term dependencies in time series data. 2. The information from the previous time step is integrated with the input of the current time step to generate a new hidden state.

Recurrent cell: In Fig. 3, the recurrent cell is the core components of the network that processes timing data and retains historical information. The main functions of the recurrent cell are: 1. Processing time series data: The loop unit enables RNNs to process data with time sequence, such as text, time series data, etc. At each time step, the cyclic unit receives the input data of the current time step and, combined with the hidden state of the previous time step, generates a new hidden state. This structure allows the network to share parameters between different time steps to accommodate sequences of different lengths. 2. Retention of historical information: The loop unit implements the memory function by storing information about previous time steps in a hidden state. At each time step, the loop unit updates the hidden state, integrating the information from the previous time step with the input from the current time step. This enables RNNs to capture long - and short-term dependencies in time series data. 3. Learning temporal features: Cyclic units can learn complex features and patterns in temporal data. Depending on the task requirements, the cyclic Unit can adopt different structures, such as a simple RNN unit, a Long Short-Term Memory unit (LSTM), or a Gated Recurrent Unit (GRU). These different cyclic unit structures allow for more efficient learning and retention of features and dependencies in time series data.

Output cell: In Fig.3, the hidden cell's hidden state is used by the output cell for generation of the final output. The output cell can be a simple linear transformation, an activation function, or a more complex structure such as the Softmax classifier. The main functions of the output cell include: 1. Map the timing features learned by the RNN to the target space (such as classification labels, regression values, etc.). 2. In sequence generation tasks (such as machine translation, text summarization, etc.), the output cell may feed the generated results back to the input cell, forming a closed loop.

An RNN is a neural network with an internal state that is capable of processing time series data. In RNNs, certain results are fed back into the network, giving it memory capability. Application cases include:

Natural language processing: In machine translation tasks, RNN can be used to process text sequences of variable length. A typical RNN structure is a coder and a decoder. Cyclic units (like LSTM or GRU) are used by the encoder to encode the source language text sequence into a fixed-length vector representation; The decoder then decodes the representation into a sequence of text in the target language. Because RNN has temporal memory capabilities, it can capture long-distance dependencies in text, enabling accurate machine translation.

Speech recognition: RNNs can be used to convert speech signals into text. In the speech recognition task, the speech signal is first split into frames, and features (like the Mel frequency cepstrum coefficient) are obtained. These features are then fed into an RNN (typically using a bidirectional LSTM or GRU) to capture timing information through the loop unit. Finally, the output layer is decoded using the Connectionist Temporal Classification (CTC) loss function to map successive frame labels onto the final text sequence.

4. Experiment

4.1. Evaluation tool

To assess neuronal network performance, a few key indicators are available:

1. Accuracy: Accuracy is the relationship between the number of appropriately planned samples and the total number of samples, which is the most common evaluation indicator used in classification assignments.
2. Precision: The ratio of true cases (TP) to the number of samples predicted to be positive is known as precision. The proportion of samples predicted to be positive cases that actually turn out to be positive is measured by this method.
3. Recall: The recall rate is the ratio of actual cases (PT) to the quantity of samples that are truly positive. It measures the proportion of the sample that is actually positive that is predicted to be positive.

There are also some tools that can evaluate the performance:

1. Confusion matrix: It is a table that demonstrates the accuracy of a model in predicting different categories, in which contains true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).
2. K-Fold Cross-Validation: Cross-validating the K-fold is a technology that is used to do the model performance evaluation. The data set is divided into K subsets, and each time, a single subset is used as a test set. The other subsets are utilized as a training set, resulting in a total of K training and testing. Lastly, the model performance evaluation is based on the average value of K test results.
3. TensorBoard: TensorBoard is a tool that allows for visualizing the process of preparing neural networks, which can display the change of loss function, accuracy and other indicators, as well as the calculation diagram of the model, parameter distribution and other information.

4.2. Database

The experiment of Feedforward, Convolutional and Recurrent Neural Networks, this experiment uses MNIST as a database, which is a classic machine-learning dataset containing grayscale images of handwritten numbers. It is widely used in academic research and algorithm testing for image classification and pattern recognition tasks. Each of the 60,000 training samples and 10,000 test samples in the dataset is a 28x28 pixel image. The task is to correctly classify these images into numbers from 0 to 9. This experiment of RNN uses 10 classifications. Although iris is more suitable for training Feedforward Neural Networks and CIFAR-10 is more suitable for training Convolutional Neural Network, the comparison would lack meaning if using different dataset. This study chooses MNIST for several reasons: 1. Diversity: The MNIST dataset contains 70,000 handwritten digital images with a high diversity. This makes it possible to provide enough samples to learn and generalize when training different types of neural networks. 2. Applicable to different networks: While the MNIST dataset is commonly used to train convolutional neural networks (CNNs), it can also be used to train feedforward neural networks (FNN) and recurrent neural networks (RNN). For FNN, the image can be flattened into a one-dimensional vector as input; For RNNs, you can think of an image as a sequence, with each row of pixels as a time step. 3. Challenging: The MNIST dataset is challenging, but moderately difficult, and suitable for evaluating the performance of different types of neural networks on image classification tasks. 4. Widely used: MNIST datasets are widely used in the field of machine learning and deep learning, so the training results can be compared with other methods to assess the effectiveness of the model.

4.3. Model hyperparameter Settings

In the hyperparameter setting of Feedforward Neural Networks, the model has a number of input neurons that are defined as 4, the number of neurons in the hidden layer is set at 6, and the number of neurons in the outgoing layer is set at 3. The activation function used in this experiment is Logistic,

the loss function is cross-entropy loss, and the optimizer is gradient descent. Softmax classification is used in the output layer.

In the hyperparameter setting of Convolutional Neural Networks, the exponential decay learning rate was used in this experiment. The learning rate's initial value was set at 0.001, and the decay step was set to the number of lots in the training package. ($500/128=390$), and the decay rate was adjusted to 0.9. Adam was adopted as the optimization algorithm. The cross entropy loss is the loss function that is employed. Softmax is used in the output process.

In the hyperparameter setting of Recurrent Neural Networks, this experiment will train the training data n times, to save time, it just trains 1 epoch. The number of batches trained is 64. This experiment considers 28-time points and a line of information includes 28-pixel points. The input size is 28, which means one line of information is read in each of the 28 steps. The learning rate equals 0.01. The optimizer used in this experiment is Adam, and the loss function is cross-entropy loss. The RNN has one layer and 64 hidden units.

4.4. Outcome demonstration

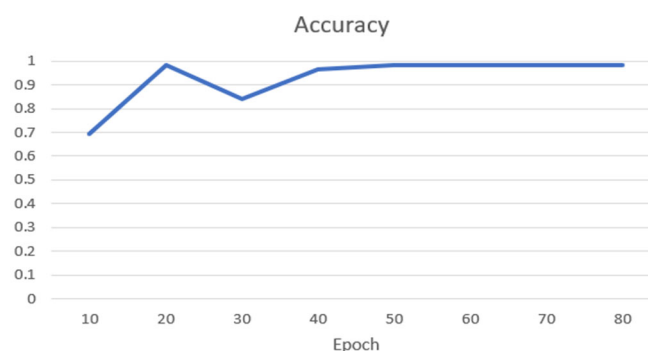


Fig. 4. The accuracy curve of Feedforward (Photo/Picture credit :Original)

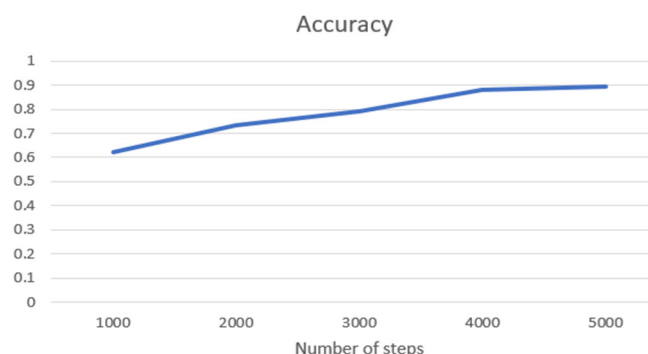


Fig. 5. The accuracy curve of Convolutional (Photo/Picture credit :Original)

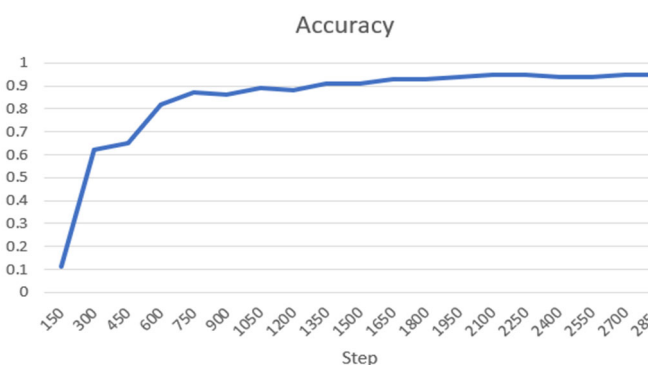


Fig. 6. The accuracy curve of Recurrent (Photo/Picture credit :Original)

Summarize: With the ever-increasing times of training, the outcome of the accuracy in Fig 4, Fig 5 and Fig 6 will have an upper trend despite some fluctuations whereas the fluctuation range will become smaller. Finally, the figure as shown on these graphs will infinitely approach a certain value.

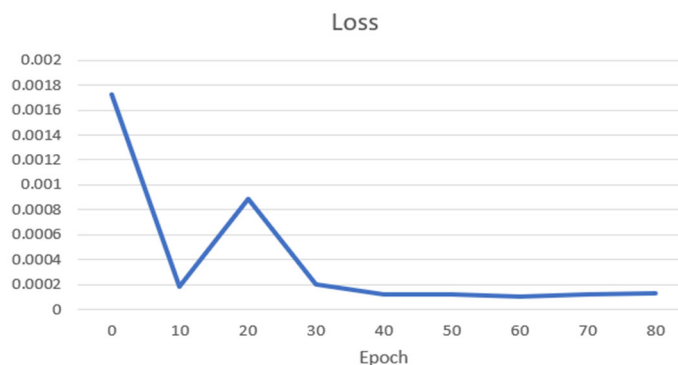


Fig. 7. The loss curve of Feedforward (Photo/Picture credit :Original)

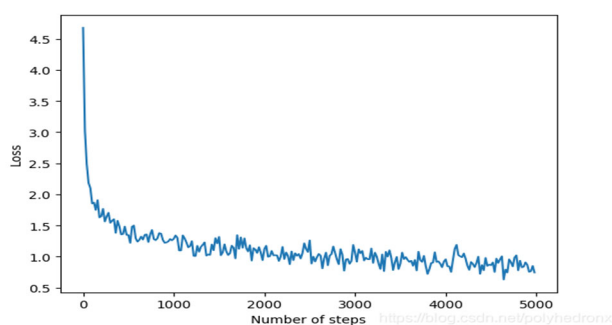


Fig. 8. The loss curve of Convolutional (Photo/Picture credit :Original)

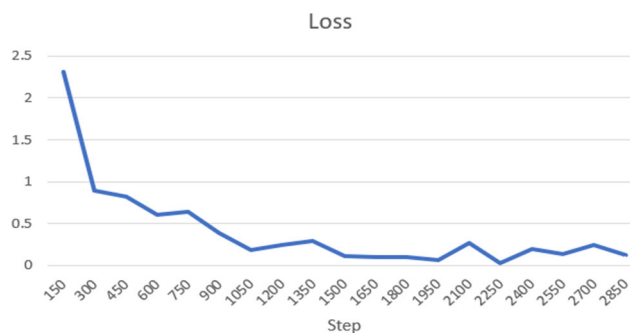


Fig. 9. The loss curve of Recurrent (Photo/Picture credit :Original)

Summarize: With the ever-increasing number of steps that have taken place, the amount of loss as shown in Fig 7, Fig 8 and Fig 9 will drop significantly at first then the figure will fluctuate in a certain range and become stable at a low pace.

5. Conclusion

The neural network is a technology that has a variety of applications. Through the analysis of this paper, this research finds that different types of neural networks can be competent for different jobs and tasks to realize their value and contribute to society. It can facilitate our life and contribute to the improvement of the quality of people's lives and the level of science and technology in the country through assistance from the industry. However, the discipline of neural networks has also fallen into a low ebb and has been attacked and criticized by some people. In the future, there are also some challenges such as Data Privacy and Security, ethics and human values etc. However, with the

increasingly fierce high-tech competition and the pains-taking exploration of neural network scholars, it finally broke through the fog with a pioneering attitude to break through the thorns and became an avant-garde discipline in the 21st century.

References

- [1] Gurney, K. An introduction to neural networks. CRC press (2018).
- [2] Macukow, B. Neural networks—state of art, brief history, basic models and architecture. In Computer Information Systems and Industrial Management: 15th IFIP TC8 International Conference, CISIM 2016, Vilnius, Lithuania, September 14-16, 2016, Proceedings 15 (pp. 3-14). Springer International Publishing (2016).
- [3] Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4(11) (2018).
- [4] Prieto, A., Prieto, B., Ortigosa, E. M., Ros, E., Pelayo, F., Ortega, J., & Rojas, I. Neural networks: An overview of early research, current frameworks and new challenges. *Neurocomputing*, 214, 242-268 (2016).
- [5] Allahyari, E., & Moshtagh, M. Predicting mental health of prisoners by artificial neural network. *BioMedicine*, 11(1), 26 (2021).
- [6] Vellido, A., Lisboa, P. J., & Vaughan, J. Neural networks in business: a survey of applications (1992–1998). *Expert Systems with applications*, 17(1), 51-70 (1999).
- [7] Kujawa, S., & Niedbała, G. Artificial neural networks in agriculture. *Agriculture*, 11(6), 497 (2021).
- [8] Svozil, D., Kvasnicka, V., & Pospichal, J. Introduction to multi-layer feed-forward neural networks. *Chemometrics and intelligent laboratory systems*, 39(1), 43-62 (1997).
- [9] Egmont-Petersen, M., de Ridder, D., & Handels, H. Image processing with neural networks—a review. *Pattern recognition*, 35(10), 2279-2301 (2002).
- [10] O'Shea, K., & Nash, R. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458* (2015).
- [11] Kusumoto, D., & Yuasa, S. The application of convolutional neural network to stem cell biology. *Inflammation and regeneration*, 39(1), 1-7 (2019).
- [12] Grossberg, S. Recurrent neural networks. *Scholarpedia*, 8(2), 1888 (2013).