Research on Machine Learning with Algorithms and Development

Liqiang Yu [1],* Xinyu Zhao [2], Jiaxin Huang[3], Hao Hu [4],Bo Liu [5]

[1] Computational Social Sciences, The University of Chicago,Irvine CA, USA
[2] Information Studies,Trine University,Phoenix, USA
[3] Information Studies,Trine University,Phoenix USA
[4] Software Engineering,Zhejiang University,Hangzhou ,China
[5] Software Engineering,Zhejiang University,Hangzhou,China
*Correspondence Author

Abstract: *Machine Learning, as one of the key technologies in the field of artificial intelligence, has made significant advancements in recent years. This study provides a relatively systematic introduction to machine learning. Firstly, it gives an overview of the historical development of machine learning, and then focuses on the analysis of classical algorithms in machine learning. Subsequently, it elucidates the latest research advancements in machine learning, aiming to comprehensively explore the applications of machine learning in various domains and discuss potential future directions.*

Keywords: Machine Learning; Algorithm; Development.

1. INTRODUCTION

With the support of modern information technology, computer technology has provided a solid foundation for the development of intelligent artificial technology. Intelligent computing technology is supported by computers and involves interdisciplinary knowledge such as statistics, approximation theory, algorithmic complexity theory, and convex part theory. Through computer technology, the system can continuously improve its performance by utilizing its own learning experience. Based on computer regular information data, intelligent computing technology seeks information patterns, obtains knowledge and experience, and realizes the intelligent development of computer technology. This enables computers to learn autonomously, adapt to the environment, and take important steps towards artificial intelligence. The intelligent progress of computer technology has not only accelerated the speed of problem solving, but also provided more efficient solutions for various industries. With the continuous evolution of computer technology, intelligent computing will continue to lead technological innovation and lay a solid foundation for future artificial intelligence applications. This development trend indicates that computer technology will demonstrate strong potential for intelligent applications in a wider range of fields.

2. THE EVOLUTION OF MACHINE LEARNING

Although artificial intelligence has not only emerged in recent years, it has always appeared in the public eye as a science fiction element. Since AlphaGo defeated Li Shishi, artificial intelligence has suddenly become a topic of discussion, as if humans have created machines that surpass human intelligence. The core technology of artificial intelligence, machine learning and its subfields of deep learning, have become the apple of people's eyes for a while.

Herb proposed a learning approach based on neuropsychology in 1949, which is known as the Herb learning theory. The general description is: assuming that the persistence or repeatability of reflex activity leads to sustained changes in cells and increases their stability, when neuron A can continuously or repeatedly stimulate neuron B, the growth or metabolic processes of one or two neurons will change. From the perspective of artificial neurons or artificial neural networks, this learning theory simply explains the correlation relationship between nodes in recurrent neural networks (RNNs), that is, when two nodes change simultaneously, there is a strong positive correlation between nodes; If the two changes are opposite, it indicates a negative weight correlation.

In 1950, Alan Turing created the Turing test to determine whether a computer is intelligent. The Turing test suggests that if a machine is able to engage in dialogue with humans, but cannot be identified as its machine identity, then the machine is considered intelligent. This simplification enables Turing to convincingly

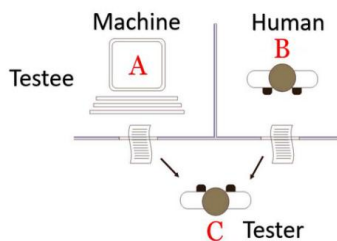demonstrate that "thinking machines" are possible.

Figure 1: Turing Test

In 1952, IBM scientist Arthur Samuel developed a checkers program. This program can provide better guidance for subsequent actions by observing the current position and learning an implicit model. Samuel discovered that as the running time of the game program increased, it could achieve better and better follow-up guidance. Through this program, Samuel refuted Providence's notion that machines cannot surpass humans, writing code and learning like humans. He coined the term "machine learning" and defined it as a research field that can provide computer capabilities without explicit programming.

In 1957, Rosen Blatter proposed the second model based on the background of neurosensory science, which is very similar to today's machine learning model. This was a very exciting discovery at the time, which was more applicable than Herb's ideas. Based on this model, Rosen Bratt designed the first computer neural network - the perceptron, which simulates the operation of the human brain.

In 1967, the nearest neighbor algorithm emerged, allowing computers to perform simple pattern recognition. The core idea of the KNN algorithm is that if most of the k nearest samples in the feature space belong to a certain category, then the sample also belongs to that category and has the characteristics of the sample in that category. This is the so-called "minority obeys majority" principle.
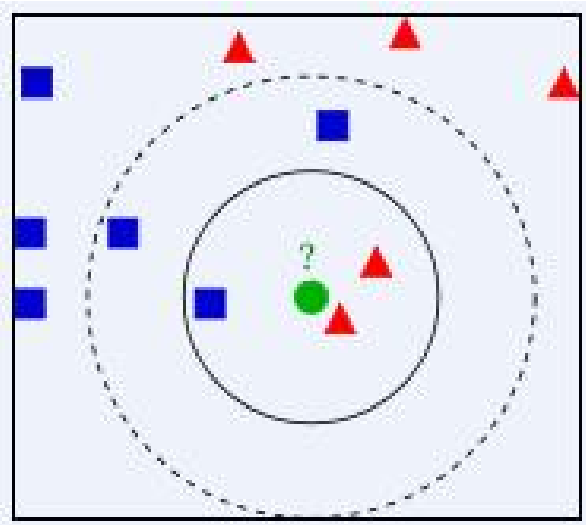

Figure 2: KNN

From the mid-1960s to the late 1970s, the development of machine learning was almost at a standstill. Both theoretical research and computer hardware limitations have encountered significant bottlenecks in the development of the entire field of artificial intelligence.

Although Winston's structural learning system and Hayes Roth's logic based inductive learning system made significant progress during this period, they could only learn a single concept and were not put into practical application. However, neural network learning machines have entered a downturn due to theoretical defects and have not achieved the expected results.
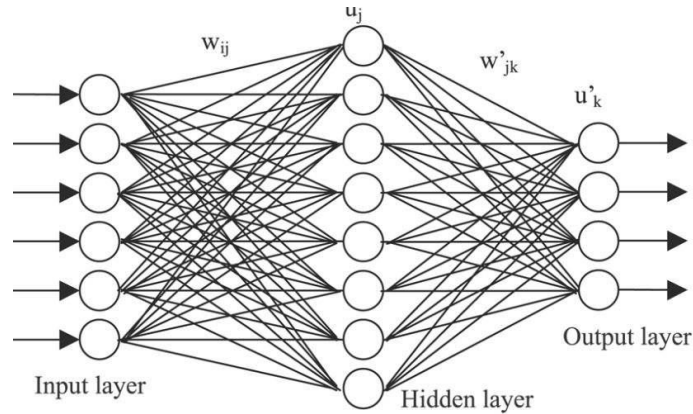
Figure 3: Multi-Layer Perceptron Model

Paul J. Werbos specifically proposed a multi-layer perceptron model (Figure 3) in the neural network backpropagation (BP) algorithm in 1981. Although the BP algorithm was proposed under the name "reverse mode of automatic differentiation" as early as 1970, it was not until then that it truly took effect, and to this day, the BP algorithm remains a key factor in neural network architecture. With these new ideas, research on neural networks has accelerated again.

In another lineage, Quinlan proposed a very famous machine learning algorithm in 1986, which we call "decision tree", more specifically the ID3 algorithm. This is another breakthrough point for mainstream machine learning algorithms. In addition, the ID3 algorithm has also been released as a software that can find more real-life cases with simple planning and clear inference, which is exactly the opposite of the neural network black box model.

In 1990, Schapire first constructed a polynomial level algorithm, which was the original Boosting algorithm. One year later, Freund proposed a more efficient Boosting algorithm. However, these two algorithms share a common practical flaw, which is that they both require prior knowledge of the lower bound for weak learning algorithms to learn correctly.

In 1995, Freund and Schapire improved the Boosting algorithm and proposed the AdaBoost (Adaptive Boosting) algorithm, which has almost the same efficiency as Freund's Boosting algorithm proposed in 1991, but does not require any prior knowledge about weak learners, making it easier to apply to practical problems.
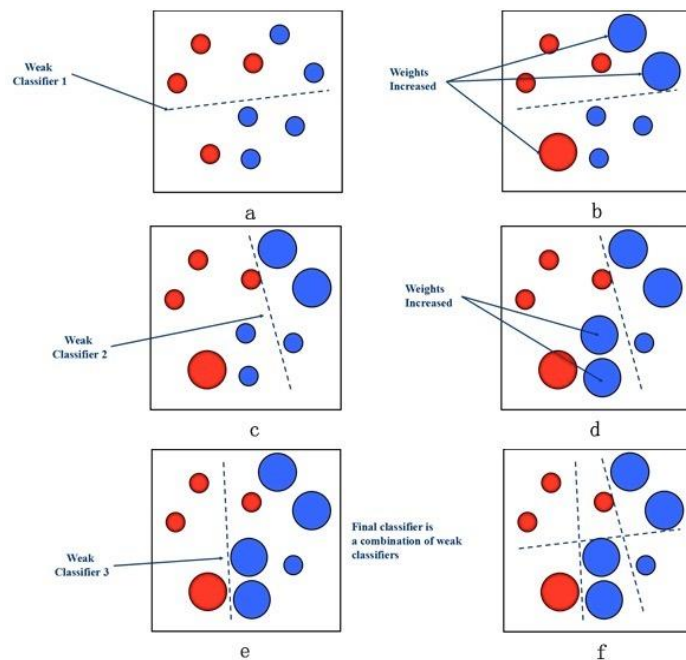

Figure 4: Boosting Algorithm

The emergence of support vector machines is another important breakthrough in the field of machine learning, and this algorithm has a very strong theoretical position and empirical results. During that period, machine learning research was also divided into two schools: NN and SVM.

However, after the proposal of support vector machines with kernel functions around 2000, SVM achieved better performance in many tasks previously occupied by NN. In addition, SVM can also utilize all the profound knowledge about convex optimization, generalized marginal theory, and kernel functions relative to NN. Therefore, SVM can vigorously promote the improvement of theory and practice from different disciplines.

Hinton, a leader in the field of neural network research, proposed the Deep Learning algorithm for neural networks in 2006, which greatly improved the capabilities of neural networks and posed a challenge to support vector machines. In 2006, Hinton, the leader of machine learning, and his student Salakhutdinov published an article in the top academic journal "Scince", ushering in a wave of deep learning in both academia and industry.

The success of deep learning does not stem from advances in neuroscience or cognitive science, but from the driving force of big data and the significant improvement of computing power. It can be said that machine learning is created by the joint efforts of academia, industry, entrepreneurship, and other fields. The academic community is the engine, the industrial community is the driving force, and the entrepreneurial community is the vitality and future. The academic and industrial communities should have their own responsibilities and division of labor. The responsibility of the academic community is to establish and develop the discipline of machine learning, and cultivate specialized talents in the field of machine learning; Large projects and engineering should be driven by the market and implemented and completed by the industry.

3.   IN DEPTH EXPLORATION OF MACHINE LEARNING ALGORITHMS

3.1 Supervised Learning Algorithms

Supervised learning is the most common and widely used algorithm in machine learning, which includes decision trees, support vector machines, and neural networks.

Decision tree is a model based on a tree structure that recursively partitions a dataset to achieve prediction of target variables. It's simple and intuitive characteristics make it widely used in data mining and classification problems. Decision trees have strong interpretability and are easy to understand and interpret, but they are also prone to overfitting and require optimization measures such as pruning. Che C et al.[1] are committed to solving the challenge of cancer literature classification in biomedical texts. They have planned a unique dataset containing over 6 pages of extensive documents, using the random forest tree method to tackle classification tasks, and combining multiple decision trees to improve accuracy and robustness. They have demonstrated outstanding performance in dealing with complex classification tasks and have been widely applied in the fields of machine learning and data science.

Support Vector Machine (SVM) is a supervised learning algorithm that performs well in classification and regression. The basic idea is to find a hyperplane that can effectively separate different categories and maximize the classification interval. The application of SVM mapping and kernel techniques in high-dimensional space makes it suitable for complex nonlinear problems. However, for large-scale datasets, SVM has a high computational complexity and requires careful selection of kernel functions and parameters. There is currently significant controversy over the detection of earthquake precursors, making remote sensing technology a warning tool. Saed A et al.[2] used machine learning support vector machines to estimate the total ionospheric electron content time series of GPS and evaluate earthquake precursors. By perturbing TEC data, SVM can identify stress accumulation signals deep in the crust, providing a new approach for earthquake prediction.

Neural network is an algorithm that simulates the structure of human brain neurons, learning complex patterns and relationships through multi-level neural networks. Neural networks in deep learning have achieved great success in fields such as image recognition and speech processing. However, the training of neural networks requires a large amount of data and computing resources, and the selection and adjustment of network structure is also a complex problem. Hu H et al. [3] addressed the challenges in casting defect detection by using the convolutional neural network model (Figure 5) Xception to accurately analyze product images, capture defects that are difficult to detect by the human eye, and improve the dataset and model generalization ability through data augmentation techniques, significantly improving defect recognition efficiency.
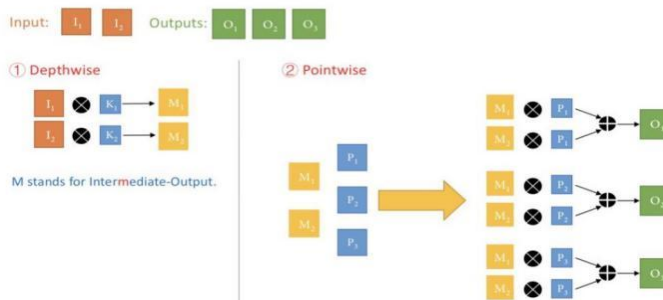
Figure 5: Depthwise Separable Convolution

Table 1: Model Results

Model	Precision	Recall	F1-Score
ResNet50	0.8456	0.7818	0.7850
CNN	0.9802	0.9790	0.9791
InceptionV3	0.9848	0.9846	0.9846
Xception	0.9852	0.9846	0.9847
Xception Aug	0.9916	0.9916	0.9916

Alzheimer's disease (AD) is a neurodegenerative disease in the elderly with no cure. Magnetic resonance imaging (MRI) is used to evaluate AD patients and provide information on brain atrophy. Research has shown that MRI features can predict the development of Alzheimer's disease. Lin Q et al. [4] assisted artificial neural network technology to accurately predict the progression from cognitive impairment to dementia, and suggested developing reliable models to assist clinical doctors in predicting early Alzheimer's disease.

3.2 Unsupervised Learning Algorithms

Unsupervised learning algorithms search for hidden patterns and structures in unlabeled data, including clustering algorithms, principal component analysis, and association rule learning.

The clustering algorithm divides data into similar clusters, resulting in higher similarity between data within the same cluster and lower similarity between different clusters. K-means clustering, hierarchical clustering, and other clustering algorithms are common and widely used in fields such as market analysis and image segmentation.

Principal Component Analysis (PCA) is a dimensionality reduction technique that maps high-dimensional data to low dimensional space through linear transformation, preserving the most important information in the dataset. PCA plays an important role in data visualization and feature extraction, but there is also an issue of information loss, which requires balancing accuracy and dimension selection in the dimensionality reduction process.

Association rule learning is used to discover the association relationships between items in a dataset, such as product combinations in shopping basket analysis. By mining association rules, businesses can be assisted in making precise recommendations and optimizing sales strategies. However, the computational cost of association rule learning on large-scale datasets is high and requires optimization processing.

3.3 Reinforcement Learning Algorithms

Reinforcement learning is an algorithm that formulates decision strategies through interactive learning between agents and the environment, including Q-Learning and deep reinforcement learning.

Learning is a value function based reinforcement learning method that formulates optimal strategies by learning the value functions of states and actions. Q-Learning performs well in fields such as transportation, gaming, and robot control, but is easily limited by dimensional disasters and sample efficiency in complex environments. Alnazir A M A et al. [5] used optimization techniques based on available space of discharge routes to solve the control problem of intersection capacity in urban transportation networks. They managed signalized intersections through deep Q-Learning agents and used density and velocity measurements to guide control strategies. In the

simulation micro model of real urban transportation networks, the discharge based controller has significantly alleviated operational problems and achieved superior results through comparative testing with other methods.
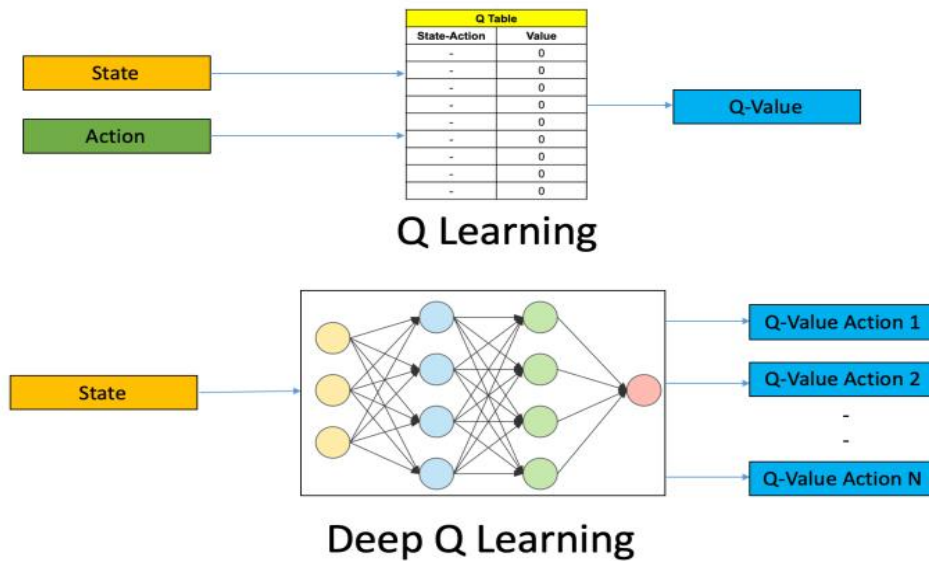

Figure 6: Q-Learning & Deep Q-Learning

Deep reinforcement learning combines deep neural networks with reinforcement learning to learn complex strategies through neural networks. The success of AlphaGo is a typical case of deep reinforcement learning, however, the training difficulty of deep reinforcement learning is high, and it needs to overcome problems such as stability and convergence. Che C et al. [6] focus on the interdisciplinary integration of mechanical engineering and computer science. Through deep learning, especially 2D convolutional neural networks, the original sensor data is transformed into accurate robot position prediction.

Through the training of asymmetric Gaussian loss functions, the mean square error is effectively reduced. With the continuous development of deep learning technology, more and more algorithms are proposed to solve various problems. However, a single algorithm is often difficult to achieve optimal results, so it is necessary to fuse and integrate multiple algorithms to achieve higher performance. The fusion and ensemble optimization of machine learning models based on swarm intelligence algorithms is a solution. Tianbo S et al. [7] combined adaptive gain control and clustering SWARM algorithm is used to implement a communication free system with adversarial agents. Adopting a biologically inspired Flocking algorithm, population observation is enhanced through adaptive gain control and partial Kalman filters. Finally, through the computer vision target recognition of SWARM drones, the simulation system was successfully transformed into practical and reliable applications.

4. THE DEVELOPMENT OF MACHINE LEARNING

The Go battle between AlphaGo and Li Shishi marked a huge advancement in artificial intelligence, with a 4:1 victory triggering profound attention to the development of artificial intelligence. This event highlights the powerful capabilities of machine learning and lays the foundation for showcasing the promising future of deep level machine learning. With the support of brain like computer cognitive technology, machine learning is bound to usher in greater development. It is crucial to conduct in-depth research on the performance, structure, learning, and functional models of machine learning to replace weak artificial intelligence and enhance intelligence in order to meet the needs of advanced development.

In the future, machine learning is expected to combine human cognition, learning, thinking, reasoning, and other aspects to enhance their abilities. Continuously upgrading, optimizing, and improving artificial intelligence is the key to ensuring the sustainable development of advanced science and technology. With the support of cloud computing, the Internet of Things and big data, machine learning will promote the development of digital technology, pay attention to the comprehensive role, promote to the height of interpersonal interaction, and realize the practical application of autonomous vehicle and other fields.

The widespread application of machine learning will bring great convenience to people's daily life and production,

promoting the development of personalized and intelligent fields such as education, finance, and healthcare. In the field of education, machine learning can customize personalized learning paths and educational resources to enhance learning effectiveness. In the financial field, machine learning can optimize risk management and investment strategies, and improve the efficiency of financial services. In the medical field, machine learning can assist doctors in diagnosis and treatment decisions, improving medical standards.

Overall, the continuous development of machine learning will bring revolutionary changes to society. Through continuous innovation and optimization, artificial intelligence will become a leader in future technological development, creating more convenient, intelligent, and personalized life and work experiences for people. In this process of continuous progress, people need to maintain a cautious focus on ethics and social impact, ensure that the application of machine learning meets ethical standards, and bring positive changes to society.

5. CONCLUSION

This study provides a detailed exploration of machine learning algorithms and a comprehensive analysis of their development trends, leading to the following conclusions. Machine learning, as a key technology in the field of artificial intelligence, has made significant achievements in supervised learning, unsupervised learning, and reinforcement learning. Supervised learning algorithms such as decision trees, support vector machines, and neural networks perform excellently in tasks such as classification and regression, providing powerful solutions for many fields. Unsupervised learning algorithms explore potential patterns in data through clustering, principal component analysis, and association rule learning, providing effective means for data analysis and feature extraction. Reinforcement learning algorithms have achieved remarkable results in fields such as automatic control and gaming due to their ability to learn optimal strategies in the interaction between agents and the environment.

However, machine learning still faces a series of challenges, including model interpretability, data privacy protection, and computational resource requirements. In the future, with the continuous evolution of technology, the continuous development of deep learning, the rise of adaptive learning systems, and research on interpretability and fairness will become key directions. In addition, federated learning, as an emerging learning framework, is expected to address issues of data privacy and security, promoting the application of machine learning in a wider range of fields.

The continuous progress of machine learning has opened up new prospects for the development of artificial intelligence, but it also requires continuous research and exploration to better adapt to the needs of society. Through continuous efforts, machine learning is expected to play a more important role in fields such as healthcare, finance, and transportation, and contribute more possibilities to building an intelligent society.

Acknowledgements

Comment

This work utilizes Random Forest to effectively analyze cancer case files, showcasing a very high medical applicability. Unlike previous research that often deals with shorter abstracts and concise summaries, it curated a unique dataset comprising documents with more extensive content, each exceeding 6 pages in length.

REFERENCES

[1] Che C, Hu H, Zhao X, et al. Advancing Cancer Document Classification with R andom Forest[J]. Academic Journal of Science and Technology, 2023, 8(1): 278-280.DOI: https://doi.org/10.54097/ajst.v8i1.14333.

[2] Saed A ,LeeAd G ,Nimrod I , et al.Using Support Vector Machine (SVM) with GPS Ionospheric TEC Estimations to Potentially Predict Earthquake Events[J].Remote Sensing,2022,14(12):2822-2822.DOI: https://doi.org/10.3390/rs14122822.

[3] Hu H, Li S, Huang J, et al. Casting product image data for quality inspection with xception and data augmentation[J]. Journal of Theory and Practice of Engineering Science, 2023, 3(10): 42-46.DOI: https://doi.org/10.53469/jtpes.2023.03(10).06.

[4] Lin Q, Che C, Hu H, et al. A Comprehensive Study on Early Alzheimer's Disease Detection through Advanced Machine Learning Techniques on MRI Data[J]. Academic Journal of Science and Technology, 2023, 8(1): 281-285.DOI: https://doi.org/10.54097/ajst.v8i1.14334.

[5] Alnazir A M A ,Ling H K ,Ee O N .Discharge control policy based on density and speed for deep Q-learning adaptive traffic signal[J].Transportmetrica B: Transport Dynamics,2023,11(1):1707-1726.DOI: https://doi.org/10.1080/21680566.2023.2243388.

[6] Che C, Liu B, Li S, et al. Deep learning for precise robot position prediction in logistics[J]. Journal of Theory and Practice of Engineering Science, 2023, 3(10): 36-41.DOI: https://doi.org/10.53469/jtpes.2023.03(10).05.

[7] Tianbo S, Weijun H, Jiangfeng C, et al. Bio-inspired Swarm Intelligence: a Flocking Project With Group Object Recognition[C]//2023 3rd International Conference on Consumer Electronics and Computer Engineering (ICCECE). IEEE, 2023: 834-837.DOI: 10.1109/ICCECE58074.2023.10135464.

[8] T. K E ,M. A H ,Abu A H , et al.Prediction of water distribution uniformity of sprinkler irrigation system based on machine learning algorithms[J].Scientific Reports,2023,13(1):20885-20885.

[9] Yuchen S ,Yanwen Q ,Ze Z , et al.Risk Factor Analysis and Multiple Predictive Machine Learning Models for Mortality in COVID-19: A Multicenter and Multi-Ethnic Cohort Study[J].Journal of Emergency Medicine,2023,65(6):e584-e593.

[10] Chenlong F ,Dongxing Z ,Li Y , et al.A multi-parameter control method for maize threshing based on machine learning algorithm optimisation[J].Biosystems Engineering,2023,236212-223.

[11] Shaikh F, Bou-Harb E, Crichigno J, et al. A machine learning model for classifying unsolicited IoT devices by observing network telescopes[A]2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)[C]. IEEE, 2018: 938-943.

[12] Liu L, Han P. Application of improved cuckoo algorithm to optimize generalized regression neural network in software quality prediction[C]//International Conference on Neural Networks, Information, and Communication Engineering (NNICE). SPIE, 2022, 12258: 94-100.

[13] You X , Yan G , Al-Masoudy M M ,et al.Application of novel hybrid machine learning approach for estimation of ultimate bond strength between ultra-high performance concrete and reinforced bar[J].Advances in engineering software, 2023.

[14] Song B , Shang K , He J ,et al.Assessing the Influence Level of Food Safety Public Opinion with Unbalanced Samples Using Ensemble Machine Learning[J].Scientific Programming, 2022.DOI:10.1155/2022/8971882.

[15] Daniël Reijsbergen, Aung Maw, Sarad Venugopalan, Dianshi Yang, Tien Tuan Anh Dinh, and Jianying Zhou. "Protecting the Integrity of IoT Sensor Data and Firmware With A Feather-Light Blockchain Infrastructure." 2022 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), IEEE, 2022, pp. 1-9.

[16] Dianshi Yang, Daisuke Mashima, Wei Lin, and Jianying Zhou. "DecIED: Scalable k-anonymous deception for iec61850-compliant smart grid systems." Proceedings of the 6th ACM on Cyber-Physical System Security Workshop, 2020, pp. 54-65.

[17] Zheng Yang, Tien Tuan Anh Dinh, Chao Yin, Yingying Yao, Dianshi Yang, Xiaolin Chang, and Jianying Zhou. "LARP: A Lightweight Auto-Refreshing Pseudonym Protocol for V2X." Proceedings of the 27th ACM on Symposium on Access Control Models and Technologies, 2022, pp. 49-60.

[18] Dianshi Yang, Abhinav Kumar, Stuart Ray, Wei Wang, and Reza Tourani. "IoT Sentinel: Correlation-based Attack Detection, Localization, and Authentication in IoT Networks." 2023 32nd International Conference on Computer Communications and Networks (ICCCN), IEEE, 2023, pp. 1-10.

[19] Chang Che, Qunwei Lin, Xinyu Zhao, Jiaxin Huang, and Liqiang Yu. 2023. Enhancing Multimodal Understanding with CLIP-Based Image-to-Text Transformation. In Proceedings of the 2023 6th International Conference on Big Data Technologies (ICBDT '23). Association for Computing Machinery, New York, NY, USA, 414–418. https://doi.org/10.1145/3627377.3627442.

[20] Li L, Xu C, Wu W, et al. Zero-resource knowledge-grounded dialogue generation[J]. Advances in Neural Information Processing Systems, 2020, 33: 8475-8485.

[21] Yao, J., Zou, Y., Du, S., Wu, H., & Yuan, B. (2023). Progress in the Application of Artificial Intelligence in Ultrasound Diagnosis of Breast Cancer. Frontiers in Computing and Intelligent Systems, 6(1), 56-59.

[22] Zhou, H., Lou, Y., Xiong, J., Wang, Y., & Liu, Y. (2023). Improvement of Deep Learning Model for Gastrointestinal Tract Segmentation Surgery. Frontiers in Computing and Intelligent Systems, 6(1), 103-106.

[23] Liu, B., Yu, L., Che, C., Lin, Q., Hu, H., & Zhao, X. (2023). Integration and Performance Analysis of Artificial Intelligence and Computer Vision Based on Deep Learning Algorithms. arXiv preprint arXiv:2312.12872.