

Enhancing Collaborative Machine Learning for Security and Privacy in Federated Learning

Mingwei Zhu^{1,*}, Jiaqiang Yuan², Guanghui Wang³, Zheng Xu⁴, Kuo Wei⁵

¹Computer Information System, Colorado state university, Fort Collins, CO, USA

²Information Studies, Trine University, Phoenix, AZ, USA

³Computer Science, Individual Contributor, Shanghai, CN

⁴Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA

⁵Computer Science, Individual Contributor, Shenzhen, China

*Corresponding author E-mail :zhumingwei666@gmail.com

Abstract: *In the age of the Internet, machine learning has revolutionized our lives, offering convenience and innovation. However, it also poses significant security and privacy risks that cannot be overlooked. With the vast amount of personal data we upload online, including search histories and location data, there's a growing concern about how this information is collected and potentially exploited by hackers or malicious actors. Moreover, in the realm of machine learning, there's a risk of attackers stealing training data and model results, which can disrupt algorithms and lead to substantial economic losses or even threats to human safety. Thus, safeguarding the security and privacy of machine learning processes has become paramount. This paper delves into the myriad security challenges and privacy risks associated with machine learning algorithms. It explores methods and technologies for securing federated learning, offering technical solutions to protect privacy while maintaining the efficiency and effectiveness of machine learning systems.*

Keywords: Machine learning; Security and privacy; Federal learning

1. INTRODUCTION

With the continuous expansion of machine learning applications, vulnerability attacks have also begun to shift from traditional software and hardware attacks to the field of machine learning. Vulnerability attacks in machine learning fall into two main categories. Attack training data: An attacker can modify the model by tampering with the training data. For example, for a capTCHA recognition model, an attacker could add certain images to the training set to trick the model into recognizing certain incorrect captCHA as correct. Attacking the model: An attacker can achieve the same effect as training data by directly attacking the trained model. For example, an attacker can intentionally input some malicious data, make the model misjudge and cause the system to crash or leak sensitive information.

Since machine learning uses a large amount of data for training, it is easy for attackers to obtain, and attackers can monitor the running state of the model, thus making the machine learning model the target of attack. Compared with traditional software and hardware attacks, machine learning vulnerability attacks have the following characteristics: the attacker does not need to know the internal structure of the model, but only needs to submit some malicious data sets or attack the training data set. Attackers can use powerful computing power and algorithms to generate large amounts of malicious data in a relatively short period of time. Attackers can attack different models, and the probability of success is high.

With the continuous expansion of machine learning applications, vulnerability attacks have gradually shifted to the field of machine learning, which is mainly divided into two categories: attacks on training data and attacks on models. An attacker can tamper with training data or attack the model directly, causing miscalculation or information leakage. Because machine learning is trained with large amounts of data, it is vulnerable to attacks. Compared with traditional attacks, machine learning vulnerability attacks do not need to understand the model structure, can use powerful computing power to quickly generate malicious data, and the attack success rate is high.

So in this case, the importance of federated learning algorithms is highlighted. Federated learning allows for model training while protecting data privacy, without the need to transfer data to a centralized server. This distributed learning approach can effectively reduce the risk of data leakage because the raw data is always kept local and only the model parameters are transmitted. In addition, Federated learning provides a secure aggregation mechanism that ensures user privacy during model updates. Therefore, federated learning algorithm plays a key role in

machine learning and provides an important technical means to protect privacy security.

2. RELATED WORK

The advantage of combining machine learning with federated algorithms to achieve privacy protection is that through model training on local devices, centralized transmission of sensitive data is avoided, thus reducing the risk of data leakage. At the same time, distributed computing and cross-border cooperation are used to improve the efficiency and speed of model training, and promote cooperation and innovation on a global scale. In addition, federated learning also allows personalized model training to meet the personalized needs of users while protecting user privacy, thus achieving a balanced development of data security and personalized services while protecting data privacy.

2.1 Machine Learning(ML)

Machine Learning (ML) is a branch of artificial intelligence that focuses on how to improve the performance of algorithms from experiential learning. Machine learning can automatically "learn" from sample data (i.e. training data) to obtain a mathematical model, and use this mathematical model to make predictions about unknown data. In recent years, machine learning theory and technology have made breakthrough progress, providing strong data and algorithm support for the vigorous development of machine learning in many fields such as computer vision, natural language processing and speech recognition, and promoting the scale and industrialization of machine learning. Realize the landing application of machine learning in multiple scenarios such as automatic driving, face recognition, intelligent medical treatment and intelligent risk control. Machine learning technology has been widely used in various fields of social life, and has excellent performance in practical applications. But machine learning itself still faces serious security concerns. For one thing, most machine learning models are not designed with the presence of attackers in mind. Although the model has excellent performance in predicting normal samples, in the real scenario, due to the possible existence of a large number of malicious users or even attackers, the machine learning model may face different degrees of security risks at all stages of the life cycle, resulting in the model failing to provide normal services or leaking the privacy information of the model. On the other hand, in addition to traditional cyber attacks, machine learning is also vulnerable to new types of attacks that exploit potential information processing algorithms and workflows, and security and privacy threats at the data, model, and application layers are diverse, hidden, and dynamically evolving. For example, some researchers have found that changing just 1 pixel can fool deep learning algorithms. Researchers have found that by interfering with road traffic signs, self-driving cars can misclassify traffic signs.

Generally speaking, attackers destroy the confidentiality, availability and integrity of the model by maliciously tampering with the training data and input samples of the model or stealing model parameters, which is the security and privacy problems faced by machine learning model.

2.2 Common security and privacy threats in machine learning

A complete machine learning phase includes data collection, cleaning, etc. before entering the training phase, and the testing phase that occurs between the training phase and the prediction phase. Among them, the training and prediction stage is the most important work, and common security and privacy threats also appear in the training and prediction stage. Common security and privacy threats in the training stage include privacy disclosure of training data and poisoning attacks; Common security and privacy threats during the prediction phase include adversarial attacks, privacy attacks, and privacy breaches of predicted data.

1. Security and privacy threats during the training phase

Privacy leakage of training data refers to the problem of data leakage that may occur during model training. When training data, Centralized Learning, Distributed Learning, or Federated Learning is often used. However, some scholars have pointed out that according to these three learning methods, data privacy disclosure will inevitably occur in the model training stage from the perspective of data collection or training methods. In the process of machine learning training, the model provider records some training data, and these training data often involve the user's personal privacy and other information. In the training stage, machine learning develops the model training

based on the training data set and obtains the decision hypothesis function based on the inherent characteristics of the learned data. The main security threat in the data training stage is Poisoning Attack.

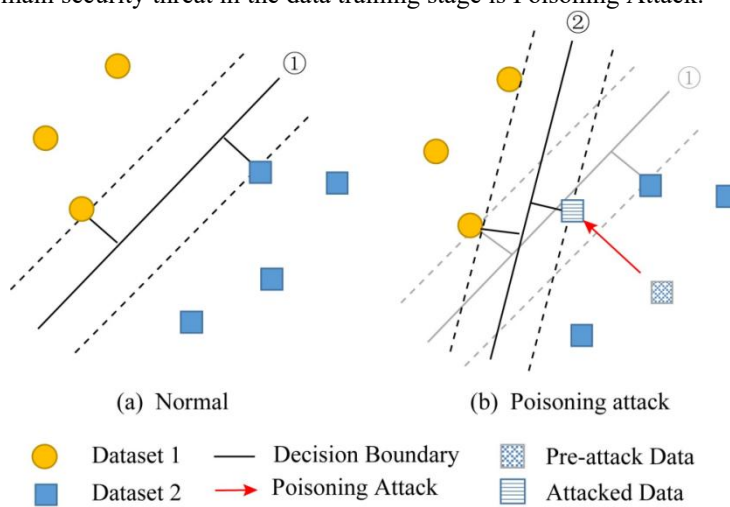


Figure 1: Data training phase of the poisoning attack architecture

Poisoning attacks are attacks in which the attacker changes the original data set by modifying, deleting or injecting bad data, so that the results of model training are biased [13]. Poisoning attacks in the data training phase refer to manipulating the predictions of machine learning models by attacking training data sets or algorithms during training or retraining. The methods of attacking training data include pollution source data, adding malicious samples to training data, modifying some labels in training data, deleting some original samples in training data, etc. On the one hand, since the performance of machine learning largely depends on the quality of the training data, high-quality data should generally be comprehensive, unbiased, and representative, while mislabeled or biased data added to the training data through an attacker's poison attack reduces the overall quality of the training data set. On the other hand, since the poisoning attack occurs before the training phase, the resulting contamination is difficult to resolve by adjusting the relevant parameters or adopting alternative models.

2.3 Machine learning security protection threats

Despite significant advances in machine learning technologies, most require centralized storage of graph data on a single machine. However, with the emphasis on data security and user privacy, centralized storage of data becomes insecure and unfeasible. Graph data is often distributed across multiple data sources (data silos), making it impractical to collect the required graph data from different places for privacy and security reasons. For example, a third-party company wants to train graph machine learning models for financial institutions to help them detect potential financial crime and fraud customers. Every financial institution has private customer data, such as demographics and transaction records. The customers of each financial institution form a customer graph, where the sides represent transaction records. Due to strict privacy policies and commercial competition, private customer data of each organization cannot be shared directly with third party companies or other organizations. At the same time, there may also be associations between institutions, which can be seen as structural information between institutions. The main challenge is to train graph machine learning models for financial crime detection based on private customer graphs and interagency structure information without direct access to each institution's private customer data.

2.4 Federated machine learning

Federated Learning (FL) is a distributed machine learning scheme that solves data silos through collaborative training. It enables participants (i.e. customers) to jointly train machine learning models without sharing their private data. Therefore, combining FL with graph machine learning is a promising solution to the above problems. In this paper, researchers from the University of Virginia propose Federated Graph Machine Learning (FGML). In general, FGML can be divided into two Settings depending on the level of structural information:

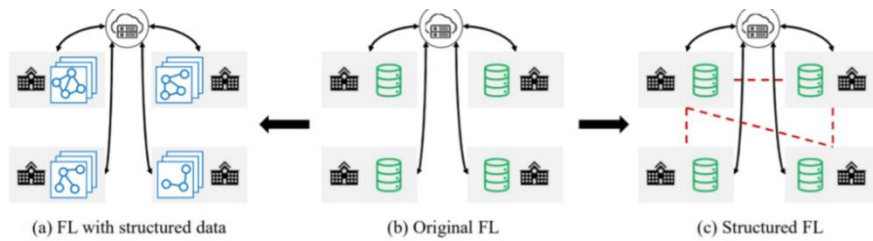


Figure 2: Three architectural models of Federation Learning (FL)

1. The first is FL with structured data, in FL with structured data, customers collaboratively train graph machine learning models based on their graph data, while keeping the graph data locally.
2. The second is structured FL, in structured FL, there is structural information between clients to form a client diagram. Client diagrams can be used to design more efficient joint optimization methods.

2.5 Research progress in federal learning safety

1. Differential privacy-based approach

This technique introduces noise to the sensitive properties of the client before the federated learning server shares a single update, so that the privacy of each user is protected. Kang Wei et al proposed a new framework based on differential privacy, which adds artificial noise to the parameters of the client before aggregation, that is, the noise aggregation before model aggregation [8]. The proposed scheme satisfies the differential privacy requirements of global data under a certain noise disturbance level of Gaussian noise by adjusting the variance appropriately, and gives the convergence limit of the trained model loss function. The experiment finds that better convergence performance will lead to lower protection ability, but under the condition of fixed privacy protection level, the model will be protected by the algorithm. Increasing the number of clients participating in the learning can improve the convergence, but there is also an optimal maximum number of aggregation. On this basis, they also propose a K-client random adjustment strategy, in which K clients are randomly selected to participate in each aggregation, so that there is an optimal value of K, and the best convergence performance is achieved at a fixed level of privacy protection.

2. Methods based on robust aggregation and homomorphic encryption

Due to the centrality of federated learning framework and the unreliability of clients, federated learning is vulnerable to attacks by malicious clients and servers. Yinbin Miao et al. designed a blockchain-based Byzantine robust Federated learning (PBFL) scheme for privacy protection [9-11]. They used cosine similarity to judge the malicious gradient uploaded by malicious clients. Provide a secure global model to resist poisoning attacks, and then adopt full-homomorphic encryption technology to provide a privacy protection training mechanism to achieve security aggregation, which can effectively prevent attackers from snooping on the local data of the client, and finally use blockchain technology, the server performs off-chain calculations and uploads the results to the blockchain. Xiaoyuan Liu et al. also adopted homomorphic encryption as the underlying technology and proposed a privacy enhanced FL(PEEL) framework to remove the malicious gradient through the logarithmic function [12]. PEEL can not only prevent the server from violating the user's privacy, but also ensure that malicious users cannot infer the membership identity by uploading the malicious gradient.

In addition, the model may be inaccurate due to the low data quality provided by some users (here called irregular users). Based on this problem, Guowen Xun et al proposed PPFDL, a federal learning framework for privacy protection with irregular users [13-15]. Highly integrated additive homomorphism and Yao's garble circuit technology ensure the confidentiality of all user information.

3. Methods based on secure multi-party computation and verification of the network

Aiming at the local gradient in the training process and the integrity of the aggregate result returned from the server, Guowen Xu et al proposed VerifyNet, the first privacy-protecting and verifiable federated learning framework [16]. They first proposed a double masking protocol to ensure the confidentiality of the user's local gradient in the

federated learning process. Allow a certain number of users to exit, but the privacy of these users is still protected by the cloud server, and then require each user to provide proof of the correctness of their aggregated results. They use a homomorphic hash function combined with pseudo-random technology as the underlying structure of VerifyNet. Allows the user to verify the correctness of the results returned from the server at an acceptable overhead.

3. METHODOLOGY

3.1 Centralized differential privacy

Centralized differential privacy is based on two adjacent data sets D and D' , which means that only one piece of data is different between D and D' . Differential privacy technology makes it impossible for users to distinguish whether data comes from data set D or data set D' from the obtained output data, so as to achieve the purpose of protecting data privacy.

The main difference between DP-SGD algorithm based on differential privacy and traditional random gradient descent algorithm (SGD) is that DP-SGD algorithm carries out gradient clipping and adds Gaussian noise during each iteration. The deep learning training algorithm based on differential privacy is as follows:

input: Training sample data:

$$\left/ (x_1, y_1), (x_2, y_2), \dots, (x_N, y_N) \right/ \tag{1}$$

Loss function:

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N L(\theta; x_i, y_i) \tag{2}$$

Learning rate: η ;

Gradient clipping boundary value: C ;

Gaussian noise standard deviation: σ , training sample size of each round: B ;

output: Model parameter: θ^T

Randomly initialize the model parameter θ_0

for each iteration $t=1,2,\dots,T$, update do with the following parameters

Set B_t with size B is randomly selected from the sample set

for each sample $i \in B_t$ do

Find the gradient value:

$$g_t(x_i, y_i) \leftarrow \nabla_{\theta_t} L(\theta_t; x_i, y_i) \tag{3}$$

Compared with the centralized differential privacy, the introduction of differential privacy technology in the federated learning scenario needs to consider not only the privacy security at the data level, but also the security at the user level. It not only requires the local data privacy security of each client, but also requires the information security between clients, that is, the user receives the local model of the client at the server side, and can neither infer which client uploaded it, nor infer whether a client participated in the current training.

3.2 Differential privacy protection applications

In the process of differential privacy protection for security data, in order to prove that the data set satisfies differential privacy, we must prove that the algorithm that produces it satisfies differential privacy.

Definition: Functions that satisfy differential privacy are often called mechanisms. We say that if for all adjacent data sets and all possible outputs, S .

$$\frac{\Pr[F(x)=S]}{\Pr[F(x')=S]} \leq e^\epsilon \quad (4)$$

If two datasets differ in the presence of only a single individual, they are considered adjacent datasets. Where F is usually a random function, so the probability distribution describing its output is not just a point distribution.

The important implication of this definition is that the output will be almost the same regardless of the data of any particular individual. In other words, the randomness built into F should be "sufficient" so that the output from the observation does not reveal which of the two is the input.

If the opponent cannot determine which of or is the input, then the opponent cannot determine whether my data exists in the input, let alone the content of the data.

The ϵ parameter in the definition is called the privacy parameter or privacy budget. A knob is provided to adjust the "amount of privacy" provided by the definition. When the value of ϵ is small, it requires a very similar output given a similar input, thus providing a higher level of privacy; When the value of ϵ is large, the similarity in the output is allowed to decrease, thus providing less privacy.

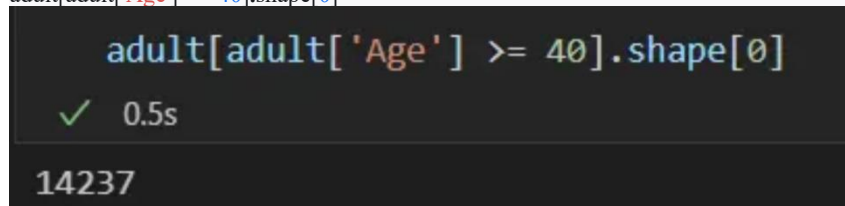
3.3. Practical application

Data Data:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
plt.style.use('seaborn-whitegrid')
adult = pd.read_csv("adult_with_pii.csv")
```

The output:

```
adult[adult['Age'] >= 40].shape[0]
```



```
adult[adult['Age'] >= 40].shape[0]
```

✓ 0.5s

14237

To make this process easier, some basic mechanisms have been developed in the field of differential privacy that describe exactly what kind of noise to use and how much to use. For an $f(x)$ function that returns a number, the following definition satisfies the ϵ difference privacy of $F(x)$:

$$F(x) = f(x) + \text{Lap}\left(\frac{s}{\epsilon}\right) \quad (5)$$

If the query counts the number of rows in the dataset with a particular property, and then we modify only one row of the dataset, the output of the query can be changed by up to 1.

Therefore, we can achieve differential privacy for the example query ϵ by using the Laplacian mechanism with a sensitivity of 1 and our choice.

Now, let's choose. We can use $\epsilon=0.1$ to sample from the Laplace distribution.

```
sensitivity = 1
epsilon = 0.1
adult[adult['Age'] >= 40].shape[0] + np.random.laplace(loc=0, scale=sensitivity/epsilon)
```



```

sensitivity = 1
epsilon = 0.1

adult[adult['Age'] >= 40].shape[0] + np.random.laplace(loc=0, scale=sensitivity/epsilon)
✓ 0.3s
14251.069541567282
    
```

Where f represents sensitivity and privacy budget. As you can see, the smaller the privacy budget, the greater the noise, and the resulting less availability and better privacy protection. Just like this picture. Privacy budget is directly proportional to availability.

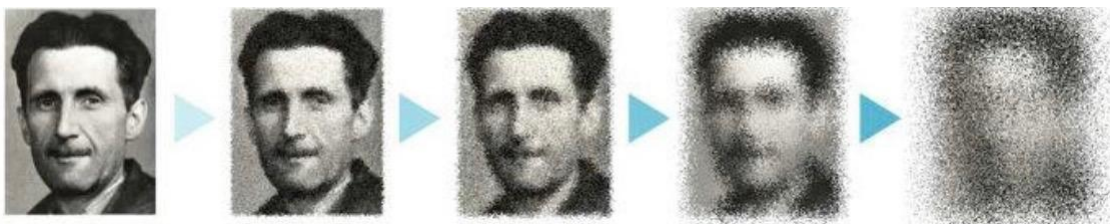


Figure 2: Privacy budget and availability proportional sequence

The proof for the above conclusion is also very simple. Firstly, it is mentioned in the previous two chapters that to satisfy $(\epsilon, 0)$ -dp, as long as the constraint $\text{MaaDivergence} < \epsilon$

$$\begin{aligned}
 \frac{p_x(z)}{p_y(z)} &= \prod_{i=1}^k \left(\frac{\exp(-\frac{\epsilon|f(x)_i - z_i|}{\Delta f})}{\exp(-\frac{\epsilon|f(y)_i - z_i|}{\Delta f})} \right) \\
 &= \prod_{i=1}^k \exp\left(\frac{\epsilon(|f(y)_i - z_i| - |f(x)_i - z_i|)}{\Delta f}\right) \\
 &\leq \prod_{i=1}^k \exp\left(\frac{\epsilon|f(x)_i - f(y)_i|}{\Delta f}\right) \\
 &= \exp\left(\frac{\epsilon \cdot \|f(x) - f(y)\|_1}{\Delta f}\right) \\
 &\leq \exp(\epsilon),
 \end{aligned} \tag{6}$$

Therefore, the Laplace mechanism is used for differential privacy protection, which protects individual privacy by adding random noise to the query results. The basic idea of differential privacy is to confuse the query results by adding some noise when statistics or query data, so as to prevent the private information of individuals from being inferred. The effectiveness of this approach depends on the concept of sensitivity.

Its Sensitivity refers to the maximum impact that a change in any single record in the input data set can have on the query results. The Laplace mechanism uses sensitivity to determine the amount of added noise to balance privacy protection and data availability. Specifically, for a query with a sensitivity of Δ , the Laplace mechanism will add random noise with a Laplacian distribution to the query result, where the size of the noise is proportional to the sensitivity, i.e., the standard deviation of the noise is Δ/ϵ , where ϵ is a privacy parameter, which controls the level of privacy protection.

Finally, by using Laplace mechanism, it is difficult for an attacker to infer sensitive information of an individual even after the query result is published, because the introduction of noise makes the query result no longer completely accurate. This approach is very useful in implementing smart security because it provides a way to allow statistical analysis and querying of data while protecting individual privacy.

4. CONCLUSION

With the continuous expansion of machine learning applications, vulnerabilities attacking machine learning have gradually shifted to the field of machine learning, which is mainly divided into two categories: attacks on training data and attacks on models. Attackers can achieve these attacks by tampering with training data or directly attacking trained models, resulting in misjudgments or information leaks. Because machine learning uses a large amount of data for training, it is easy for attackers to obtain and monitor the operating state of the model, making

the machine learning model a target for attack. Compared with traditional attacks, machine learning vulnerability attacks do not need to understand the model structure, and can use powerful computing power to quickly generate malicious data, and the attack success rate is high. Therefore, in the current situation, the importance of federated learning algorithms is highlighted.

Federated learning allows for model training while protecting data privacy, without the need to transfer data to a centralized server. This distributed learning approach can effectively reduce the risk of data leakage because the raw data is always kept locally and only the model parameters are transmitted. In addition, Federated learning provides a secure aggregation mechanism to ensure that user privacy is protected during model updates. Therefore, federated learning algorithms play a key role in machine learning and provide important technical means for protecting privacy security.

In summary, it can be concluded that federated learning, as an effective method to protect data privacy, provides important technical support for solving security and privacy problems in machine learning. With the continuous development of machine learning applications, federated learning algorithms will achieve greater breakthroughs and applications in the future. Through federated learning, we can achieve effective training and improvement of models while protecting personal privacy, promoting the safe development of machine learning technologies.

REFERENCES

- [1] Song Tianbo, Hu Weijun, Cai Jiangfeng, Liu Weijia, Yuan Quan, and He Kun. Bio-inspired swarm intelligence: a flocking project with group object recognition. In 2023 3rd International Conference on Consumer Electronics and Computer Engineering (ICCECE), pages 834–837. IEEE, 2023.
- [2] An Overview of the Development of Stereotactic Body Radiation Therapy. (2024). *Frontiers in Computing and Intelligent Systems*, 6(3), 56-60. <https://doi.org/10.54097/09nly12x>.
- [3] Yang, Le & Tian, Miao & Xin, Duan & Cheng, Qishuo & Zheng, Jiajian. (2024). AI-Driven Anonymization: Protecting Personal Data Privacy While Leveraging Machine Learning.
- [4] Cheng, Qishuo & Yang, Le & Zheng, Jiajian & Tian, Miao & Xin, Duan. (2024). Optimizing Portfolio Management and Risk Assessment in Digital Assets Using Deep Learning for Predictive Analysis.
- [5] Yao, Jerry, et al. "Progress in the Application of Artificial Intelligence in Ultrasound Diagnosis of Breast Cancer". *Frontiers in Computing and Intelligent Systems*, vol. 6, no. 1, Nov. 2023, pp. 56-59, <https://doi.org/10.54097/fcis.v6i1.11>.
- [6] Pan, Yiming, et al. "Application of Three-Dimensional Coding Network in Screening and Diagnosis of Cervical Precancerous Lesions". *Frontiers in Computing and Intelligent Systems*, vol. 6, no. 3, Jan. 2024, pp. 61-64, <https://doi.org/10.54097/mi3VM0yB>.
- [7] He, Yuhang, et al. "Intelligent Fault Analysis With AIOps Technology". *Journal of Theory and Practice of Engineering Science*, vol. 4, no. 01, Feb. 2024, pp. 94-100, doi:10.53469/jtpes.2024.04(01).13.
- [8] Cai, J., Ou, Y., Li, X., Wang, H. (2021). ST-NAS: Efficient Optimization of Joint Neural Architecture and Hyperparameter. In: Mantoro, T., Lee, M., Ayu, M.A., Wong, K.W., Hidayanto, A.N. (eds) *Neural Information Processing. ICONIP 2021. Communications in Computer and Information Science*, vol 1516. Springer, Cham. https://doi.org/10.1007/978-3-030-92307-5_32.
- [9] Pan, Linying & Xu, Jingyu & Wan, Weixiang & Zeng, Qiang. (2024). Combine deep learning and artificial intelligence to optimize the application path of digital image processing technology.
- [10] Wan, Weixiang & Sun, Wenjian & Zeng, Qiang & Pan, Linying & Xu, Jingyu. (2024). Progress in artificial intelligence applications based on the combination of self-driven sensors and deep learning.
- [11] Sun, Wenjian & Xu, Jingyu & Pan, Linying & Wan, Weixiang & Wang, Yong. (2024). Automatic driving lane change safety prediction model based on LSTM.
- [12] Du, S., Li, L., Wang, Y., Liu, Y., & Pan, Y. (2023). Application of HPV-16 in Liquid-Based thin Layer Cytology of Host Genetic Lesions Based on AI Diagnostic Technology Presentation of Liquid. *Journal of Theory and Practice of Engineering Science*, 3(12), 1-6.
- [13] H. Zhu and B. Wang, "Negative Siamese Network for Classifying Semantically Similar Sentences," 2021 International Conference on Asian Language Processing (IALP), Singapore, Singapore, 2021, pp. 170-173, doi: 10.1109/IALP54817.2021.9675278.
- [14] "Exploring New Frontiers of Deep Learning in Legal Practice: A Case Study of Large Language Models". *International Journal of Computer Science and Information Technology*, vol. 1, no. 1, Dec. 2023, pp. 131-8, <https://doi.org/10.62051/ijcsit.v1n1.18>.

- [15] Duan, Shiheng, et al. "Prediction of Atmospheric Carbon Dioxide Radiative Transfer Model Based on Machine Learning". *Frontiers in Computing and Intelligent Systems*, vol. 6, no. 3, Jan. 2024, pp. 132-6, <https://doi.org/10.54097/ObMPjw5n>.
- [16] "Unveiling the Future Navigating Next-Generation AI Frontiers and Innovations in Application". *International Journal of Computer Science and Information Technology*, vol. 1, no. 1, Dec. 2023, pp. 147-56, <https://doi.org/10.62051/ijcsit.v1n1.20>.
- [17] K.Tan and W. Li, "Imaging and Parameter Estimating for Fast Moving Targets in Airborne SAR," in *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 126-140, March 2017, doi: 10.1109/TCI.2016.2634421.
- [18] K. Tan and W. Li, "A novel moving parameter estimation approach offast moving targets based on phase extraction," 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 2015, pp. 2075-2079, doi: 10.1109/ICIP.2015.7351166.
- [19] He, Zheng & Shen, Xinyu & Zhou, Yanlin & Wang, Yong. (2024). Application of K-means clustering based on artificial intelligence in gene statistics of biological information engineering. 10.13140/RG.2.2.11207.47527.
- [20] Wang, Yong & Ji, Huan & Zhou, Yanlin & He, Zheng & Shen, Xinyu. (2024). Construction and application of artificial intelligence crowdsourcing map based on multi-track GPS data. 10.13140/RG.2.2.24419.53288.
- [21] Zheng, Jiajian & Xin, Duan & Cheng, Qishuo & Tian, Miao & Yang, Le. (2024). The Random Forest Model for Analyzing and Forecasting the US Stock Market in the Context of Smart Finance.