

Utilizing Machine Learning for Precise Audience Targeting in Data Science and Targeted Advertising

Huan Ji^{1,*}, Xiaonan Xu², Guangze Su³, Jiufan Wang⁴, Yufu Wang⁵

¹Engineering Management, Trine University, phoenix, AZ, USA

²Independent Researcher, Northern Arizona University, Flagstaff, USA

³Information Science, Trine University, Arizona, USA

⁴Independent Researcher, William & Mary, Williamsburg, VA, USA

⁵Computer Science & Engineering, Santa Clara University, Santa Clara, CA, USA

*Corresponding author E-mail: 95133hji23@my.trine.edu

Abstract: In the current era of technology, big data has become increasingly accurate in understanding individuals' interests and needs. Personalized advertising recommendations have become a reality through the use of deep learning and intelligent algorithms. Whether it is a shopping app suggesting your favorite style of clothes or receiving relevant news feeds after discussing hot pot with friends, these are not coincidences but the result of machine learning. Machine learning has revolutionized advertising by providing outstanding performance. It has replaced the traditional approach that relied on the experience and intuition of practitioners. Powerful data processing and analysis capabilities are used to extract potential associations and patterns from massive amounts of data, providing new possibilities for advertisers and marketers. This includes applications such as idea generation, recommendation optimization, and bid strategy optimization, which have injected new vitality into the industry. The use of machine learning in data science and targeted advertising has yielded impressive results, offering users and advertisers more precise targeting and engaging advertising experiences. This field holds great promise and potential, and will undoubtedly continue to shape the future of the advertising industry.

Keywords: Personalized advertising; Machine learning; Data science; Targeted audience.

1. Introduction

In the evolving landscape of a slowing global economy, advertisers are adjusting their strategies to align with the changing market dynamics. Whether it's brand advertising or performance advertising, advertisers are placing a greater emphasis on the tangible impact of their ad campaigns, rather than just aiming for broad reach. ROI (Return on Investment) has emerged as the pivotal metric in maximizing the value of each advertising budget. For brand advertising, especially in economic downturns, advertisers are shifting their focus towards the long-term value of their campaigns. Instead of solely increasing brand awareness, they are concentrating on cultivating deeper brand relationships and enhancing brand value. Through ingeniously crafted advertising creativity and more precise audience targeting, brand advertising can generate substantial brand benefits within the confines of an effective advertising budget, facilitating more sustainable growth.

In the realm of performance advertising, advertisers are now prioritizing actual conversions and business value in the face of fierce competition[1]. Click-through rates and exposure are no longer the sole concerns. Practical results and business value take center stage as advertisers aim to ensure that their advertising efforts yield genuine value for their businesses. In a competitive market environment with budget constraints, advertisers recognize that the adoption of personalized and highly targeted advertising content is crucial to reach their intended audience. By adjusting their strategies, both brand advertising and performance advertising are becoming more accurate and measurable. This approach ensures that each ad delivery is highly targeted and closely aligned with the actual business objectives[2-4]. This strategy

not only enhances ad targeting but also enables advertisers to achieve higher ROI even with limited budgets.

Personalization has always been a cornerstone of marketing, and in the current era of technological advancement, machine learning is taking personalized advertising to new heights. With machine learning, advertisers can create hyper-personalized experiences that not only enhance the effectiveness of their ads but also adapt to market fluctuations. Technologies like real-time data analysis, intelligent placement decisions, predictive analytics, and personalized recommendation systems empower advertisers with more flexible, market-relevant strategies and higher conversion rates.

As machine learning continues to advance, it will play a pivotal role in refining and optimizing advertising strategies, making ad campaigns more precise and responsive to the ever-changing market conditions.

2. Related Work

The relationship between[5] AI, ML and DL: Artificial intelligence is the broadest concept, machine learning is a method of implementing artificial intelligence, and deep learning is a specific technology in machine learning.

Artificial Intelligence (AI) : This is the broadest, topmost concept. The goal of artificial intelligence is to enable computers to perform complex tasks that require human intelligence.

Machine Learning (ML) : As a subfield of artificial intelligence, machine learning is a method for achieving artificial intelligence goals. It studies how algorithms can enable computers to learn from data and make predictions or decisions without explicit programming.

Deep Learning (DL) : Deep learning is a specific technique

in the field of machine learning that is inspired by the structure of the human brain and uses artificial neural networks to simulate how human neural networks work. Deep learning processes and analyzes complex data by building multi-layer neural networks that can automatically extract high-level features from the data.

The quest for precise robot positioning within logistics automation has been a focal point in recent research endeavors. A multitude of techniques and methodologies have been explored to address this critical challenge[6].

2.1. Machine Learning (ML)

Machine learning is a scientific discipline that studies algorithms from data. Specialized in the study of how computers simulate or implement human learning behavior to acquire new knowledge or skills, reorganize the existing knowledge structure to continuously improve its own performance. As mentioned above, to put it bluntly, it is to select algorithms based on existing data, and build models based on algorithms and data, and finally make predictions about the future.

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \quad (1)$$

Practical applications of machine learning:

① Personalized recommendation: Personalized refers to changing the user experience and the content presented to the user based on various factors, which may include the user's behavioral data and external factors; A recommendation usually refers to a system that presents the user with a list of items that may be of interest to the user. For example, brushing Tiktok[7-9], we will find that how certain types of content (like food) are particularly large? In fact, according to your operational data (likes, comments, retweets, completion rate, etc.), I think you may be interested in food and other types of content, so I will recommend more relevant content.

② Precision marketing: Find out the specific marketing objects from the user masses.

③ Customer segmentation: attempts to divide the user group into different groups, according to the given user characteristics of the customer group. For example, if you plan to go to the bank for a loan, the bank will judge whether you have the ability to repay according to your personal information and credit card use, and decide whether to loan

you money.

Predictive modeling and analysis: Model based on existing data, and use the resulting model to predict the future.

2.2. Machine learning and advertising recommendations

Personalisation has always been a crucial element of marketing, and machine learning is taking personalised advertising to new heights in the current context of technological advancement. Through the use of machine learning, advertisers can create hyper-personalised experiences that improve the effectiveness of their ads. This is achieved by adapting to market changes through real-time data analysis[10-11], intelligent placement decisions, predictive analytics, and personalised recommendation systems. Additionally, machine learning allows for more flexible and market-appropriate strategies, resulting in higher conversion rates.

1. Real-time data analysis and optimization: Through in-depth analysis of real-time data, machine learning enables advertisers to understand audience feedback and effects in real time during advertising campaigns. This real-time data analysis capability provides advertisers with more timely and accurate optimization solutions to ensure the flexibility and efficiency of advertising campaigns.

2. Predictive analysis optimization: Machine learning has the ability to predict advertising effects through deep learning of historical advertising data. Based on these predictive analytics, advertisers can adjust advertising strategies more specifically to improve the actual effectiveness of advertising and user engagement rates[12].

3.intelligent delivery decision: In the field of effective advertising, machine learning is not only a data analysis tool, but also the key to intelligent delivery decision. By learning and analyzing large amounts of data, machine learning systems can automatically adjust advertising strategies to maximize advertising effectiveness and reduce costs.

4.personalized recommendation system[13]: With the help of machine learning algorithms, advertisers can establish a more efficient personalized recommendation system. By digging deep into user behavior data, the system can accurately predict user interests, so that users can present more in line with their expectations of advertising content, improve user interaction and conversion.

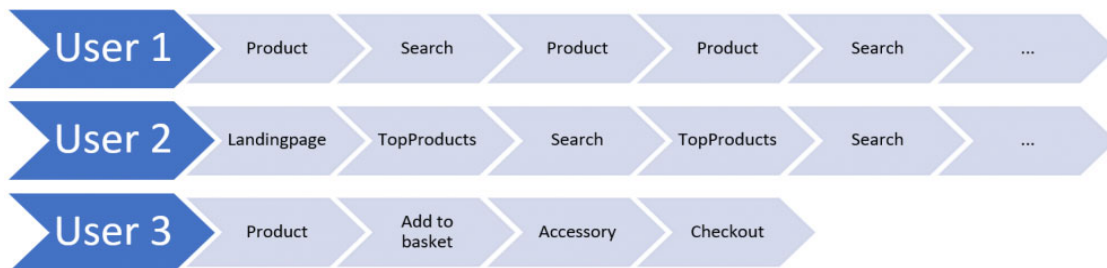


Figure 1. Three different user event streams

On a website, a user's intentions can often be inferred from their past interactions. These interactions leave a trail of events that include search queries, article page views, and email receipts. This data is used as the basis for the following techniques. The initial step is to collect or extract this data. The raw data is frequently stored in a web server or

database and may require optimization to make it usable.

3. Methodology

In the realm of e-commerce, websites frequently adopt a 'one size fits all' approach to their customers. However, this approach may not always cater to the diverse needs of all customers.

Customers may have varying intentions and goals when they visit a website, and understanding these intentions can significantly enhance their online experience. The objective of this study is to employ machine learning models to predict customer intentions using pre-existing customer behaviour data[14-18]. Anticipating our customers' intentions enables us to meet their needs and provide personalised recommendations and services, enhancing their overall experience on the site. The following section details the methodology and results of the experiment, demonstrating how a data-driven approach can improve the customer experience.

3.1. Training data

The initial step is to assign an event ID to each event. It may also be helpful to group similar events under a single event ID. This can be accomplished manually or by using LabelEncoder in sci-kit learning[19-23]. It is recommended to begin with 1 as the first ID since 0 is reserved for padding.

Step 2: Each event is represented by a number, and our data is simply a sequence of numbers. To ensure that each sequence [24-26]of events is of equal length, the input to the classifier must have a fixed size. We achieve this by populating the data to a predefined length. By default, pad_sequences fills in missing events with zeros before the sequence starts. If the sequence is longer than the required length, the beginning of the sequence is truncated. The resulting X is then obtained.

```
Code: import numpy as np
import tensorflow as tf
num_events = 8
example_seq_len = 10
example_events = [ [1, 2, 1, 2, 1], # user1
[3, 4, 2, 4, 1], # user2
[1, 5, 6, 7]] # user3x =
```

For this example, we will utilise information from another system to determine that the customer has made a purchase. Typically, this is also a top-down event. In this case, User 3 has made a purchase, resulting in a target label of 1.

```
Code: y = [0, 0, 1]------(x, y)
Output:(array([[0, 0, 0, 0, 0, 1, 2, 1, 2, 1],
[0, 0, 0, 0, 0, 3, 4, 2, 4, 1],
[0, 0, 0, 0, 0, 1, 5, 6, 7]]), [0, 0, 1])
```

In machine learning personalised recommendation applications, the number of unique events can reach thousands, while the length of the event stream is usually in the hundreds[27]. The number of events in the clickstream can vary for each user and session. However, it is necessary to limit the number of events at some point. The exact number depends on the data, but the 90th percentile is a good starting point.

3.2. Prediction sequence

To predict user intent, it is crucial to divide the data into time windows. One window (X) represents the click data before t0, while the second window represents the data after t0, where the target event Y is expected to occur. The click data is not a continuous numerical value, but the window method involves moving the window through the data, creating multiple sequences for a single user. For instance, a 6-hour time window can be utilised to forecast whether a customer will make a purchase within the next 2 hours. By sliding across the entire daily dataset, multiple sequences (X and Y) can be obtained.

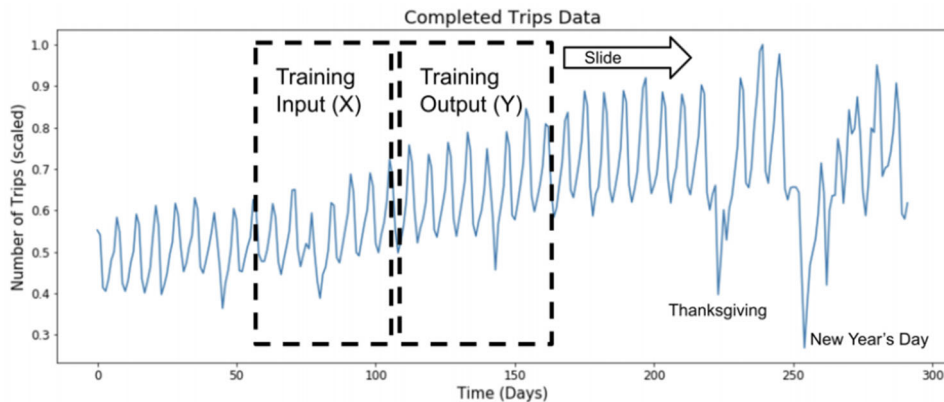


Figure 2. Prediction model architecture diagram

The model prediction results indicate that the model is not very useful when events that leak labels are included, such as clicking the 'pay' button to make a purchase prediction. There are no general rules about which events to exclude, but if the classifier performs exceptionally well on such tasks, it may have a leakage feature. It is important to note that intent prediction will never yield a high-precision model because the event data is usually not clean and specific enough.

3.3. Random sequence data

The subsequent stage involves inputting the data into the artificial neural network. Each event is assigned an internal representation in the neural network, known as an embed. This representation is learned by the network during training. The learning process of the network constructs an embedded space where each event is positioned based on its similarity to other events[28]. The use of such representations enables comparison of events (refer to word embedding). Furthermore, it allows for the handling of numerous distinct events without the need to

manage high-dimensional vectors. The quantity of events corresponds to the vocabulary in NLP.

Once the transformation from event ID to embedded representation (embedding layer) is complete, the sequence must be reduced to a single vector. LSTM is the standard method for this task. The Masking layer in the following example eliminates zeros from the sequence.

```
Code: import random
num_events = 1000
seq_len = 100
y = np.random.choice(2, 1000, replace=True)
x = np.random.randint(num_events, size=(1000, seq_len))
net_in = tf.keras.layers.Input(shape=(seq_len,))
emb = tf.keras.layers.Embedding(num_events, 8, input_length=seq_len,
mask_zero=True, input_shape=(num_events,))
net_in_mask = tf.keras.layers.Masking(mask_value=0)(emb)
lstm = tf.keras.layers.LSTM(64)(net_in_mask)
dense = tf.keras.layers.Dense(1, activation='sigmoid')(lstm)
model = tf.keras.Model(net_in, dense)
model.compile('adam', 'binary_crossentropy',
metrics=['acc'])
model.summary()
history = model.fit(x, y, epochs = 50, validation_split=0.2)
```

3.4. Data result

The dataset comprises approximately 13 million event streams exported using sliding window methods, with a class imbalance of about 1:100. Based on this data, the Conv1D architecture outperforms the LSTM architecture in several ways. Firstly, convolution is faster to calculate and therefore quicker

to train. Secondly, LSTM is more sensitive to hyperparameters, so the model becomes more robust with convolution, and even slightly more accurate[29]. The classifier below exhibits a significant difference in the Precision/Recall ratio and a slight difference in the ROC curve. Therefore, it is more suitable for predicting personalized targeted advertising messages to target customers using a convolutional network.

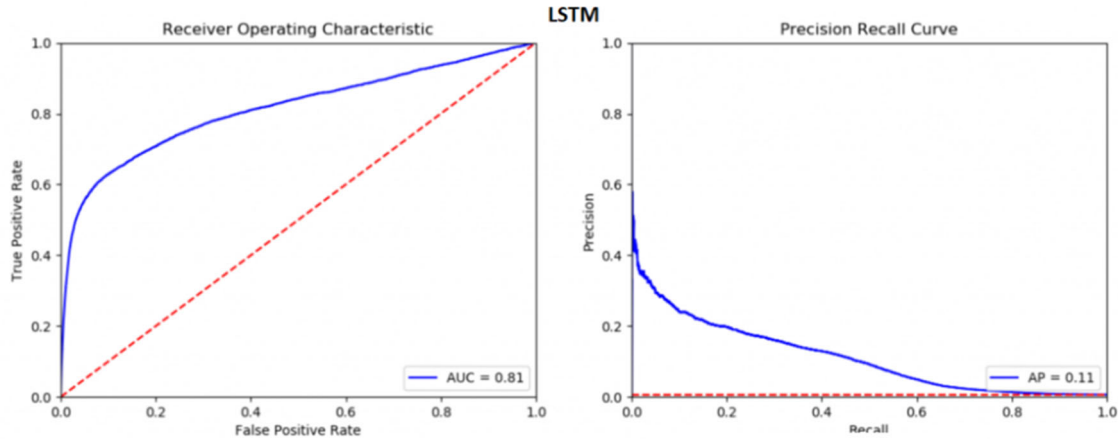


Figure 3. LSTM personalized predictive ROC curve

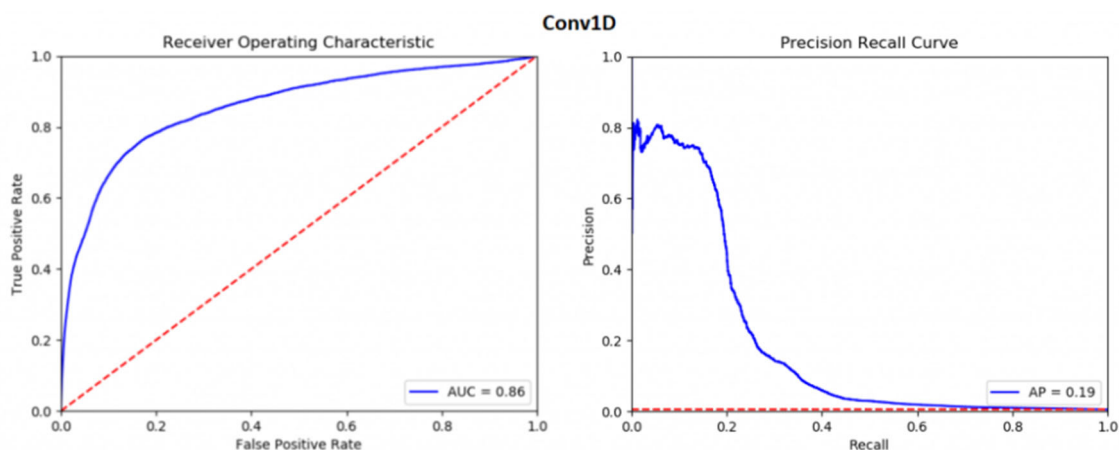


Figure 4. Precision-Recall curves for the experiment

In reality, data is rarely perfect. It often contains duplicate events, crawlers, bots, or other sources of noise. Before classifying sequences, ensure that your data is corrected by filtering out outliers, merging duplicate events, and splitting sessions from users based on a meaningful inactivity time. You may also consider incorporating time into the model, such as including 'time since last event' on top of the embedding vector. Additionally[30], it may be beneficial to use both a maximum and minimum sequence length. If the target event is rare, undersampling of negative sequences can be an option.

4. Conclusion

This article examines how big data can accurately understand the interests and needs of individuals in the current era of technology. It also discusses how deep learning and intelligent algorithms can be used to achieve personalized advertising recommendations. Examples of this include shopping apps suggesting preferred clothing styles and receiving relevant news feeds after discussing hotpot with friends. These outcomes are not accidental, but rather the result of machine learning. Machine learning has replaced

traditional advertising methods that relied on practitioner experience and intuition[31]. It leverages powerful data processing and analytics to extract potential associations and patterns from large amounts of data, providing advertisers and marketers with new possibilities for idea generation, recommendation optimization, and bid strategy optimization. The use of machine learning in data science and targeted advertising has produced impressive outcomes, offering users and advertisers more accurate targeting and engaging advertising experiences. This field has enormous potential and will undoubtedly continue to shape the future of the advertising industry.

Machine learning has significant advantages in the field of personalized AD recommendations. Through machine learning, advertisers can create hyper-personalized experiences that not only improve the effectiveness of their ads, but also adapt to changes in the market[32-34]. Through technologies such as real-time data analytics, intelligent placement decisions, predictive analytics, and personalized recommendation systems, advertisers have the flexibility to adapt their strategies and deliver market-relevant strategies that increase conversion rates. In addition, machine learning enables advertisers to better understand users, enabling more

precise targeting and personalized advertising content. As machine learning continues to evolve, it will play a key role in refining and optimizing advertising strategies, making campaigns more precise and responsive to the ever-changing market.

Acknowledgment

We would like to extend our sincere thanks to Professor Guanghui Wang and his outstanding contributions and professional insights in the field of artificial intelligence. In particular, we would like to thank him for his paper "The Application of Artificial Intelligence to The Bayesian Model Algorithm for Combining Genome Data," This paper provides an important reference and theoretical basis for this study. This paper discusses the application of artificial intelligence to genomic data in Bayesian model algorithm, which provides valuable inspiration and support for our research.

We would also like to thank Professor Guanghui Wang for his guidance and support in academic exchange and research cooperation. His research results have provided us with in-depth ideas and methods that have advanced our understanding and application of machine learning in data science and targeted advertising. With his help, we were able to delve into machine learning techniques for precise audience targeting and put forward innovative ideas and methods in this paper.

Finally, we would like to once again express our sincerest thanks and high respect to Professor Guanghui Wang. Thank you for his selfless sharing and professional support, and look forward to more academic exchanges and cooperation opportunities in the future to jointly promote the development of machine learning in the field of data science and targeted advertising.

References

- [1] 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>.
- [2] "Enhancing Computer Digital Signal Processing through the Utilization of RNN Sequence Algorithms". *International Journal of Computer Science and Information Technology*, vol. 1, no. 1, Dec. 2023, pp. 60-68, <https://doi.org/10.62051/ijcsit.v1n1.09>
- [3] Cai, Guoqing et al. "A deep learning-based algorithm for crop Disease identification positioning using computer vision." *International Journal of Computer Science and Information Technology* (2023): n. pag.
- [4] Chen, Wangmei, et al. "Applying Machine Learning Algorithm to Optimize Personalized Education Recommendation System". *Journal of Theory and Practice of Engineering Science*, vol. 4, no. 01, Feb. 2024, pp. 101-8, doi:10.53469/jtpes.2024.04(01).14.
- [5] 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.
- [6] "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>. (10)
- [7] "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>.
- [8] Du, Shuqian, et al. "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*, vol. 3, no. 12, Dec. 2023, pp. 1-6, doi:10.53469/jtpes.2023.03(12).01.
- [9] Dong, Xinqi, et al. "The Prediction Trend of Enterprise Financial Risk Based on Machine Learning ARIMA Model". *Journal of Theory and Practice of Engineering Science*, vol. 4, no. 01, Jan. 2024, pp. 65-71, doi:10.53469/jtpes.2024.04(01).09.
- [10] 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.
- [11] K. Jin, Z. Z. Zhong and E. Y. Zhao, "Sustainable Digital Marketing Under Big Data: An AI Random Forest Model Approach," in *IEEE Transactions on Engineering Management*, vol. 71, pp. 3566-3579, 2024, doi: 10.1109/TEM.2023.3348991.
- [12] Yu, L., Liu, B., Lin, Q., Zhao, X., & Che, C. (2024). Semantic Similarity Matching for Patent Documents Using Ensemble BERT-related Model and Novel Text Processing Method. *arXiv preprint arXiv:2401.06782*.
- [13] Liu, B., Zhao, X., Hu, H., Lin, Q., & Huang, J. (2023). Detection of Esophageal Cancer Lesions Based on CBAM Faster R-CNN. *Journal of Theory and Practice of Engineering Science*, 3(12), 36-42. [https://doi.org/10.53469/jtpes.2023.03\(12\).06](https://doi.org/10.53469/jtpes.2023.03(12).06)
- [14] Liu, Bo, et al. "Integration and Performance Analysis of Artificial Intelligence and Computer Vision Based on Deep Learning Algorithms." *arXiv preprint arXiv:2312.12872* (2023).
- [15] Liu, Bo, et al. "Integration and Performance Analysis of Artificial Intelligence and Computer Vision Based on Deep Learning Algorithms." *arXiv preprint arXiv:2312.12872* (2023).
- [16] 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>.
- [17] Tan, Kai, et al. "Integrating Advanced Computer Vision and AI Algorithms for Autonomous Driving Systems". *Journal of Theory and Practice of Engineering Science*, vol. 4, no. 01, Jan. 2024, pp. 41-48, doi:10.53469/jtpes.2024.04(01).06.
- [18] Wang, Sihao, et al. "Diabetes Risk Analysis Based on Machine Learning LASSO Regression Model". *Journal of Theory and Practice of Engineering Science*, vol. 4, no. 01, Jan. 2024, pp. 58-64, doi:10.53469/jtpes.2024.04(01).08.
- [19] Wei, Kuo, et al. "Strategic Application of AI Intelligent Algorithm in Network Threat Detection and Defense". *Journal of Theory and Practice of Engineering Science*, vol. 4, no. 01, Jan. 2024, pp. 49-57, doi:10.53469/jtpes.2024.04(01).07.
- [20] Zong, Yanqi, et al. "Improvements and Challenges in StarCraft II Macro-Management A Study on the MSC Dataset". *Journal of Theory and Practice of Engineering Science*, vol. 3, no. 12, Dec. 2023, pp. 29-35, doi:10.53469/jtpes.2023.03(12).05.
- [21] 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>.

- [22] Huang, J., Zhao, X., Che, C., Lin, Q., & Liu, B. (2024). Enhancing Essay Scoring with Adversarial Weights Perturbation and Metric-specific Attention Pooling. arXiv preprint arXiv:2401.05433.
- [23] Bao, W., Che, H., & Zhang, J. (2020, December). Will_Go at SemEval-2020 Task 3: An Accurate Model for Predicting the (Graded) Effect of Context in Word Similarity Based on BERT. In A. Herbelot, X. Zhu, A. Palmer, N. Schneider, J. May, & E. Shutova (Eds.), *Proceedings of the Fourteenth Workshop on Semantic Evaluation* (pp. 301–306). doi:10.18653/v1/2020.semeval-1.
- [24] Niu H, Li H, Wang J, et al. Enhancing Computer Digital Signal Processing through the Utilization of RNN Sequence Algorithms[J]. *International Journal of Computer Science and Information Technology*, 2023, 1(1): 60-68.
- [25] Su, J., Nair, S., & Popokh, L. (2023, February). EdgeGym: A Reinforcement Learning Environment for Constraint-Aware NFV Resource Allocation. 2023 IEEE 2nd International Conference on AI in Cybersecurity (ICAIC), 1–7. doi:10.1109/ICAIC57335.2023.10044182
- [26] 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.
- [27] Popokh, L., Su, J., Nair, S., & Olinick, E. (2021, September). IllumiCore: Optimization Modeling and Implementation for Efficient VNF Placement. 2021 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 1–7. doi:10.23919/SoftCOM52868.2021.9559076.
- [28] Zhuoran Lu, Zhuoyan Li, Chun-Wei Chiang, Ming Yin , strategic Adversarial Attacks in AI-assisted Decision Making to Reduce Human Trust and Reliance Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence Main Track. Pages 3020-3028. <https://doi.org/10.24963/ijcai.2023/337>