International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 1, Number 1, Year 2023
DOI: https://doi.org/10.62051/ijcsit.v1n1.19
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Machine Learning Model Training and Practice: A Study on Constructing a Novel Drug Detection System

Beichang Liu[1, *], Guoqing Cai[2], Jili Qian[3], Tianbo Song[4], Quan Zhang[5]

[1] Information Studies, Trine University, Salt Lake City UT, USA
[2] Information Studies, Trine University, Phoenix AZ, USA
[3] Information Studies, Trine University, Phoenix AZ, USA
[4] Computer Engineering, Arizona State University, Tempe, AZ, USA
[5] Information Studies, Trine University, Phoenix AZ, USA
*Corresponding Author: Beichang Liu (Email: alexfeynman2023@gmail.com)

ABSTRACT

Drugs, AIDS and terrorism are the three major public hazards in the world. Drug abuse seriously endangers social security and human life and health. The World Drug Report 2023 shows that the continued record supply of illicit drugs and increasingly flexible trafficking networks are exacerbating the global crisis and posing challenges for health services and law enforcement responses. The number of people injecting drugs worldwide in 2021 is estimated to be 13.2 million, 18% higher than previous estimates. At present, the world drug control situation is very serious, and it is necessary to increase the means to solve the drug problem and curb the spread of drugs, and efficient and accurate drug detection technology plays a very important role in drug control work. NPS, also known as "planning drugs" or "laboratory drugs", is a drug analogue obtained by the chemical structure modification of controlled drugs by criminals in order to evade the crackdown. It has similar or stronger excitatory, hallucinogenic, narcotic and other effects with controlled drugs, and has become a third-generation drug popular in the world after traditional drugs and synthetic drugs. Therefore, through the drug detection technology based on artificial intelligence machine learning technology, the current development of AI as a "drug detector" and the future space are analyzed.

KEYWORDS

Drug detection; Machine learning; NSP; Risk prediction.

1. INTRODUCTION

A large number of new psychotropic drugs appear on the illegal market each year. These substances can cause psychoactive effects similar to those of known illegal drugs; But because the way they are synthesized makes them chemically different, these drugs circumvent existing drug laws and are even harder to detect. Forensic laboratories use mass spectrometry to identify known man-made drugs in seized pills or powders. The rapid emergence of a large number of NPS in the world poses a major risk to public security, and poses a challenge to social governance and rule of law construction. Often, little is known about the adverse health effects and social harms of NPS, which poses considerable challenges for prevention and treatment. At the same time, the analysis of potentially large numbers of NPS is demanding, and monitoring, information sharing, early warning, and risk awareness are critical to dealing with this situation. But figuring out the structure of a new synthetic drug often requires weeks or months of work by chemists and other experimental techniques.

As a result, they are mostly outside the scope of existing drug laws, making them difficult to detect. Testing for synthetic drugs is usually done by the relevant forensic laboratory, which usually samples seized pills or powders and identifies them using mass spectrometry. Figuring out the structure of a new synthetic drug is not an easy task, and chemists often have to work for weeks or even months, in addition to other types of experimental techniques. Michael Skinnider at the University of British Columbia in Canada and colleagues used confidential data crowdsourced from forensic LABS around the world to train a machine-learning model to produce molecules with structures and properties similar to recent man-made drugs. The model then produced a database containing the structures of a billion potential new psychotropic drugs. Therefore, this paper tests the model with the newly collected data after the model training, and finds that this method can identify unknown man-made drugs using mass spectrometry alone. In cases where the exact structure is difficult to determine precisely, the model suggests a structure very similar to that of unknown man-made drugs.

2. RAPID DRUG DETECTION TECHNOLOGY

According to the Regulations on Drug Testing Procedures issued by the Ministry of Public Security, drug testing includes on-site testing, laboratory testing and laboratory retesting. Because of the problems of long detection cycle, expensive equipment and high detection cost, laboratory detection can not meet the requirements of on-site detection in anti-drug work. At present, the main technologies that can meet the requirements of field rapid detection are enzyme-linked immunoassay (ELISA), colloidal gold method, time-resolved fluorescence immunochromatography, quantum dot fluorescence immunochromatography, colloidal carbon method, etc.

2.1. Enzyme-linked immunoassay

Enzyme-linked immunosorbent assay (ELISA) is a heterogeneous immunoassay technique using generic enzymes as markers. The detection principle is to bind the antigen or antibody to the surface of the solid phase carrier, and use enzymes to label the antigen or antibody, and maintain the catalytic activity of the enzyme and the immune activity of the antigen or antibody. In assays, samples and enzyme-labeled antigens or antibodies react in different steps with antibodies or antigens on the surface of a solid phase carrier. By washing, other substances can be separated to retain the antigen and antibody complexes formed on the solid carrier. Finally, the amount of the enzyme bound to the solid carrier is proportional to the amount of the object detected in the sample. After the addition of an enzyme substrate, the enzyme catalyzes it to become a colored substance, through which the amount of this colored substance and the depth of the color change can be qualitatively or quantitatively the composition in the sample to be tested. Due to the high catalytic efficiency of the enzyme, the method can achieve relatively high detection sensitivity and can be applied to the detection of drugs and psychoactive substances.

2.2. Two-color fluorescence detection

The detection principle of the fluorescent-enhanced two-color visual drug detection probe is as follows: when methamphetamine and ketamine are not detected, the probe has no fluorescence emission and is in a fluorescence quenching state under 365nm excitation light due to the fluorescence resonance energy transfer mechanism between polydopamine nanoballs and copper nanoclusters. In the presence of methamphetamine and/or ketamine, the specific binding between the surface modified antibody of copper nanoclusters and the target detection substance is stronger than the π-π weak interaction between copper nanoclusters and polydopamine nanospheres. Therefore, under the excitation light of 365nm, the copper nanoclusters are separated from the polydopamine nanospheres. The detection probe showed blue and/or orange fluorescence, which enhanced the detection probe fluorescence signal. The enhancement of the fluorescence signal is proportional to the content of the object to be detected, so that the visual synchronous detection of multiple drugs based on fluorescence enhancement can be realized.

2.3. Quantum dot immunofluorescence technique

Quantum dots (also known as "semiconductor nanoparticles") are composed of group I~VI or group V elements whose radius is less than or close to the laser Bohr radius, and are a class of semiconductor nanoparticles that can accept excited light to produce fluorescence. Because of its wide excitation spectrum, narrow emission peak, polychromism and photobleaching resistance, it has become an important fluorescent marker in modern homogeneous immunoassay techniques.

Quantum dot immunochromatography uses quantum dot fluorescence probe as a marker, which not only fully embodies the advantages of simple, rapid and strong specificity of immunochromatography, but also demonstrates the high sensitivity and fluorescence characteristics of quantum dots, showing great development potential and broad application prospects. However, there are still some problems with the application of quantum dots. First, the transition of quantum dots between different states leads to the appearance of fluorescent scintillation, which affects the detection results of individual quantum dots. Secondly, with the gradual excitation of quantum dots, their fluorescence intensity will gradually become stronger, affecting quantitative detection. Finally, in the process of quantitative detection, the luminous intensity is required to be stable, and the change of intensity will affect the accuracy of the detection results. The effective solution of these problems will greatly promote the application of quantum dot fluorescence immune technology.

3. MACHINE LEARNING AND DRUG DETECTION

In recent years, with the emergence of new drugs, the traditional drug detection technology is facing great challenges. In this context, the development of machine learning technology has injected new vitality into the field of drug detection. At present, many research institutions and enterprises have begun to explore how machine learning can be applied to drug detection. This trend is not only reflected in the laboratory environment, but also includes portable inspection equipment and real-time monitoring systems. By combining advanced machine learning algorithms, researchers are able to more accurately identify and classify various novel drugs, improving the efficiency and accuracy of drug detection.

3.1. Role of Machine Learning in Drug Detection

Machine learning plays a key role in drug detection, mainly in the following aspects. First, by using supervised and unsupervised learning methods, machine learning models can be trained on large amounts of drug data to be able to identify new drugs. Second, machine learning can enable the learning of complex data patterns, enabling detection systems to better cope with the variability and diversity of new drugs. In addition, machine learning can also provide deeper information for drug control by analyzing large-scale data sets to find potential correlations and patterns.

3.2. Advantages

Compared with traditional drug detection methods, machine learning technology has obvious advantages. First, the machine learning model can continuously improve its performance in the process of continuous learning and adapt to changes in new drugs. Second, machine learning is excellent at processing large-scale data, enabling a more comprehensive analysis of the dynamics of the drug market. In addition, the automated nature of machine learning makes the drug detection process more efficient, reducing the need for human intervention. These advantages together provide a broad development prospect for the application of machine learning in drug detection, and provide a strong support for the society to combat drug crimes.

4. METHODOLOGY

"With the new technology of 'poison detection AI+', for an emerging NPS, it can quickly predict what kind of drug analogues it is, what toxicological activity it has, and whether it poses a harm." At present, there are more than 800 known and confirmed "harmful" NPS worldwide, but there are about 3,000 to 5,000 substances that have been reported but not confirmed toxicity, in addition to the unreported discovery of potential NPS community is much larger. Therefore, the application technology research of artificial intelligence in drug analysis and toxicity prediction focuses on substances that are harmful to human body (mainly drugs), and establishes a model that can identify the toxicity and amount of substances through specific technologies such as deep learning, so as to digitize harmful substances and evaluate substances by grading. Through toxicology studies, the AI is trained to score accurately. "The establishment of a safe range for substances that deviate from this range can quickly provide more dimensional scientific data on whether it is a 'drug.'".

4.1. Data collection and processing

The study focused on a class of drugs called NPS (novel psychoactive substances), or new psychoactive drugs. These new psychotropic drugs, often created by "street chemists," have the same hallucinogenic effects as drugs such as marijuana and heroin. In order to evade the law, the chemical structure of new psychotropic drugs is often not known. Right now, the pain point for law enforcement and medical authorities is how to detect them. For example, law enforcement intercepts a shipment of powder at an airport and needs to know what it is, or the medical department has a patient who overdosed today and needs to know exactly what the patient took.

First, we need to collect sample data including NSP drugs and non-drugs. The data set includes known NSP components and their characteristics, as well as data on other harmless or legal substances. Therefore, it is necessary to make a convolutional data network suitable for NSP drug characteristics, as shown in the figure below:

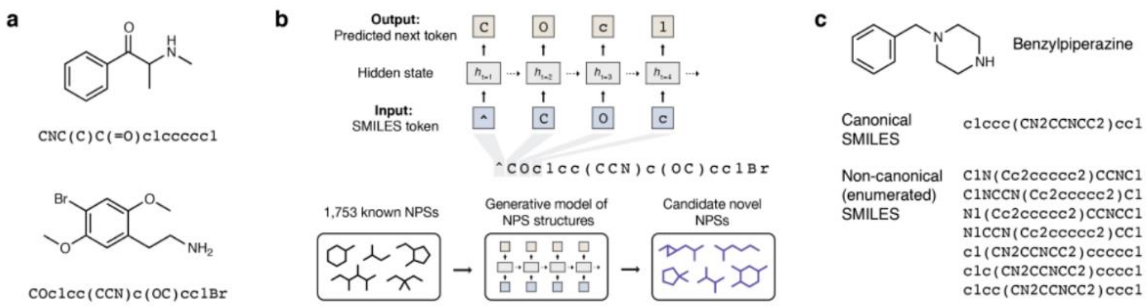

Figure 1. Establishment of data convolutional networks

At present, there are two common laboratory methods to obtain the structure: one is by nuclear magnetic resonance (NMR); Another is by mass spectrometry (MS). That is, when you get a sample, you first get its nuclear magnetic resonance or mass spectrometry, and then you go to a database to compare it.

4.2. Feature engineering and model training

In the experiment, the team trained chemical language model models (DarkNPS) on the structures of more than 1,700 new hallucinogens. This model uses SMILES (multiple simplified Molecular-input line-entry system) text to represent the molecular structure. Conceptually, the model is very similar to OpenAI's GPT-3, except that the input to GPT-3 is human language text, while the input to the model is a molecular text representation. This model can generate a large number of molecular expression texts. By modifying the model they got about a billion different outputs. Because molecular SMILES can be repetitive. That is, the same molecular structure can be expressed in different texts, and after removing unqualified expressions, the molecular structure of 8.9 million potential new psychotropic drugs is finally obtained. Next, the team used an existing mass spectrometry prediction model (CFM-ID) to calculate MS/MS mass spectra for each molecular structure. In the test, the system achieved a 68% Top-3 detection accuracy.

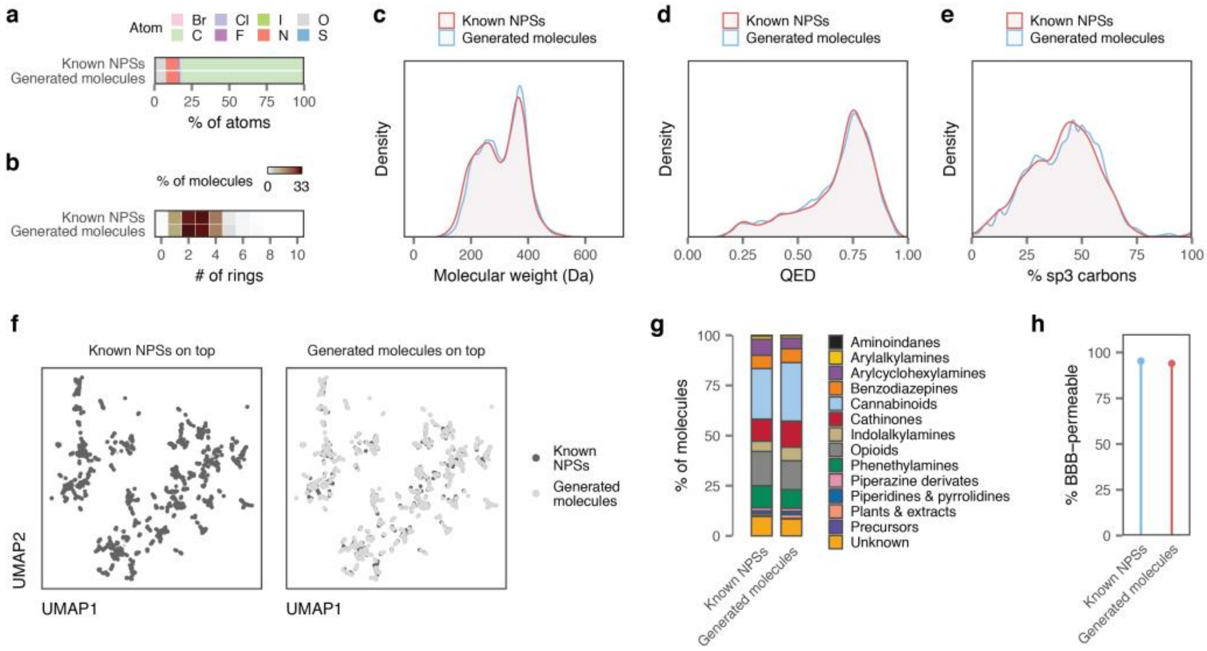

Figure 2. Detection model formed by machine learning algorithm

Through machine learning and other means,AI can accurately predict the molecular structure, targeted binding, toxicological activity and other aspects, and through the corresponding algorithm, you can know what impact this substance has on the human body. "This not only saves a lot of time, but also can more accurately 'portrait' for new harmful substances, avoiding mistakes and omissions."

4.3. Deployment models and improvements

The study focused on a class of drugs called NPS (novel psychoactive substances), or new psychoactive drugs. These new psychotropic drugs, often created by "street chemists," have the same hallucinogenic effects as drugs such as marijuana and heroin. In order to evade the law, the chemical structure of new psychotropic drugs is often not known.

In addition, the method of combining the molecular generation structure model of artificial intelligence and the model generated by mass spectrometry will provide a new idea for small molecule recognition, especially for biological detection samples. Other interesting applications could include the detection of stimulants, and the same method could also be used in some medical-related tests. As for the generative model itself, it can be used in drug development, as well as the detection of environmental pollutants.

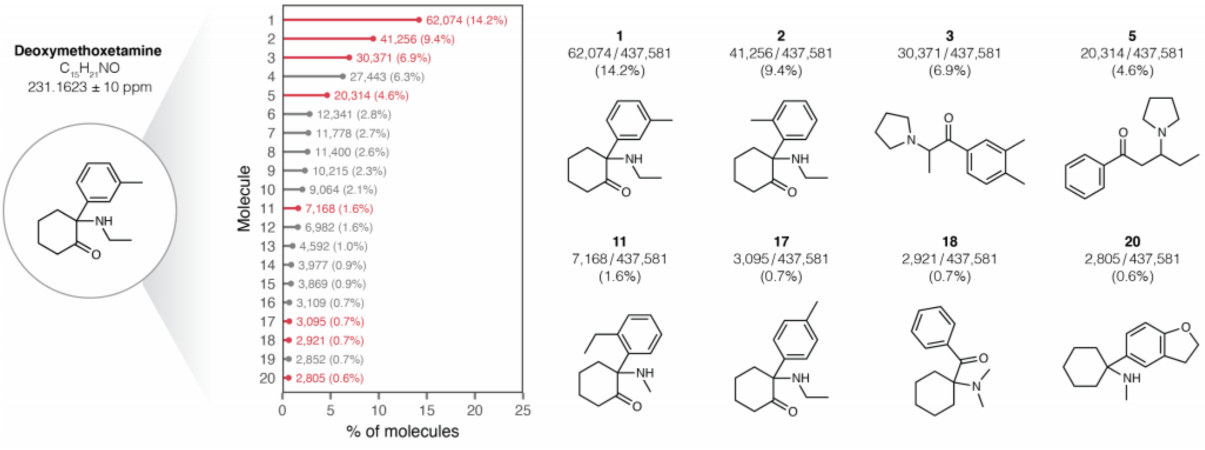

Figure 3. Principle of novel DMXE drug model

5. CONCLUSION

This study explores the intersection of artificial intelligence, machine learning, and drug detection to address the pressing issues in drug control.The application of machine learning algorithms, such as the DarkNPS model, demonstrates a promising avenue for predicting the molecular structures and toxicological activities of NPS.By leveraging large datasets and advanced algorithms, machine learning enhances the accuracy and efficiency of drug detection processes, surpassing the limitations of traditional methods. Furthermore, the integration of mass spectrometry prediction models and chemical language models illustrates a comprehensive approach to identifying potential new psychotropic drugs.The study's findings showcase a significant leap in the capability to rapidly detect and categorize drugs, offering a valuable tool for law enforcement and healthcare professionals.The deployment of the 'poison detection AI+' technology presents a practical solution for identifying harmful substances, contributing to a safer society.

In essence, the marriage of machine learning and drug detection technology not only addresses the immediate challenges posed by NPS but also establishes a foundation for continuous adaptation to evolving threats.The automated and data-driven nature of machine learning provides a powerful ally in the ongoing battle against drug-related crimes, offering a glimpse into a future where technology plays a pivotal role in preserving public health and safety.

ACKNOWLEDGEMENT

In the process of completing this article, I would like to express my sincere thanks to the journal Frontiers in Computing and Intelligent Systems. Thank you to the journal for providing me with a platform to present my research results and enable me to share my research on the application of deep learning in drug detection technology with the academic community. At the same time, I would like to express my sincere gratitude to authors such as Zhou Hong for their work. Their research results in the article "Improvement of Deep Learning Model for Gastrointestinal Tract Segmentation Surgery" provide important reference and inspiration for my article. In my research on drug detection technologies, the experience of improving deep learning models from their work has been a positive boost to my research.

This article provides strong support for my research and provides important guidance for my exploration in the field of deep learning and drug detection. Thanks again to the journal Frontiers in Computing and Intelligent Systems and Zhou Hong and other authors for their outstanding contributions to the academic community.

REFERENCES

[1] Li, Shulin, et al. “Application Analysis of AI Technology Combined With Spiral CT Scanning in Early Lung Cancer Screening”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 52-55, https://doi.org/10.54097/LAwfJzEA.

[2] Liu, Bo, et al. “Detection of Esophageal Cancer Lesions Based on CBAM Faster R-CNN”. Journal of Theory and Practice of Engineering Science, vol. 3, no. 12, Dec. 2023, pp. 36-42, doi:10.53469/jtpes.2023.03(12).06.

[3] “Based on Intelligent Advertising Recommendation and Abnormal Advertising Monitoring System in the Field of Machine Learning”. International Journal of Computer Science and Information Technology, vol. 1, no. 1, Dec. 2023, pp. 17-23, https://doi.org/10.62051/ijcsit.v1n1.03.

[4] Yu, Liqiang, et al. “Research on Machine Learning With Algorithms and Development”. Journal of Theory and Practice of Engineering Science, vol. 3, no. 12, Dec. 2023, pp. 7-14, doi:10.53469/jtpes.2023.03(12).02.

[5] 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).

[6] 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.

[7] Huang, J., Zhao, X., Che, C., Lin, Q., & Liu, B. (2024). Enhancing Essay Scoring with Adversarial Weights Perturbation and Metric-specific AttentionPooling. arXiv preprint arXiv:2401.05433.

[8] Tianbo, Song, 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), pp. 834-837. IEEE, 2023.DOI: 10.1109/mce.2022.3206678

[9] 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

[10] 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).

[11] “Implementation of Computer Vision Technology Based on Artificial Intelligence for Medical Image Analysis”. International Journal of Computer Science and Information Technology, vol. 1, no. 1, Dec. 2023, pp. 69-76, https://doi.org/10.62051/ijcsit.v1n1.10.

[12] “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.

[13] 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.

[14] 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.

[15] “A Deep Learning-Based Algorithm for Crop Disease Identification Positioning Using Computer Vision”. International Journal of Computer Science and Information Technology, vol. 1, no. 1, Dec. 2023, pp. 85-92, https://doi.org/10.62051/ijcsit.v1n1.12.

[16] 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.

[17] 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.

[18] S. Tianbo, H. Weijun, C. Jiangfeng, L. Weijia, Y. Quan and H. Kun, "Bio-inspired Swarm Intelligence: a Flocking Project With Group Object Recognition," 2023 3rd International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 2023, pp. 834-837, doi: 10.1109/ICCECE58074.2023.10135464.

[19] Zhang, Quan, et al. “Deep Learning Model Aids Breast Cancer Detection”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 1, Dec. 2023, pp. 99-102, https://doi.org/10.54097/fcis.v6i1.18.

[20] Jingyu Xu, Yifeng Jiang, Bin Yuan, Shulin Li, Tianbo Song，Automated Scoring of Clinical Patient Notes using Advanced NLP and Pseudo Labeling，arXiv preprint arXiv:2401.12994, 2024

[21] Xiaonan Xu, Bin Yuan, Yongyao Mo, Tianbo Song, Shulin Li, Curriculum Recommendations Using Transformer Base Model with InfoNCE Loss And Language Switching Method, arXiv preprint arXiv:2401.09699

[22] “A Deep Learning-Based Algorithm for Crop Disease Identification Positioning Using Computer Vision”. International Journal of Computer Science and Information Technology, vol. 1, no. 1, Dec. 2023, pp. 85-92, https://doi.org/10.62051/ijcsit.v1n1.12.

[23] Zhang, Yufeng, et al. "Manipulator Control System Based on Machine Vision." International Conference on Applications and Techniques in Cyber Intelligence ATCI 2019: Applications and Techniques in Cyber Intelligence 7. Springer International Publishing, 2020.

[24] Gao, Longsen, et al. "Autonomous Multi-Robot Servicing for Spacecraft Operation Extension." 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2023.