Application of AI technology in pulsar candidate identification (Postprint)

Authors: Wanqiong Wang, Jie Wang, Xinchen Ye, Yazhou Zhang, Jia Li, Xu Du, Wenna Cai, Han Wu, Ting Zhang, Yuyue Jiao, Jie Wang

Date: 2025-03-14T00:00:00+00:00

Abstract

As artificial intelligence (AI) technology has continued to develop, its efficient data processing and pattern recognition capabilities have significantly improved the precision and speed of decision-making processes, and it has been widely applied across various fields. In the field of astronomy, AI techniques have demonstrated unique advantages, particularly in the identification of pulsars and pulsar candidates. AI is capable of accurately and efficiently addressing the challenges of pulsar identification and classification. This paper systematically surveys commonly used AI models for pulsar candidate identification, analyzing and discussing the typical applications of machine learning, artificial neural networks, convolutional neural networks, and generative adversarial networks in candidate identification. Furthermore, it explores how the introduction of AI techniques not only enhances the efficiency and accuracy of pulsar identification but also provides new perspectives and tools for pulsar survey data processing, thus playing a significant role in advancing pulsar research and the field of astronomy.

Full Text

Preamble

Astronomical Techniques and Instruments, Vol. 2, January 2025, 27–43

Article Open Access

Application of AI Technology in Pulsar Candidate Identification

Wanqiong Wang[1], Jie Wang[1,2,3*], Xinchen Ye[1,3], Yazhou Zhang[1], Jia Li[1], Xu Du[1,2], Wenna Cai[1,2], Han Wu[1,2], Ting Zhang[1,2], Yuyue Jiao[1,2]

[1]Xinjiang Astronomical Observatory, Chinese Academy of Sciences, Urumqi 830011, China

[2]University of Chinese Academy of Sciences, Beijing 100049, China
[3]National Astronomical Data Center, Beijing 100101, China

*Correspondence: wangjie@xao.ac.cn

Received: September 10, 2024; Accepted: October 26, 2024; Published Online: November 16, 2024
https://doi.org/10.61977/ati2024050; https://cstr.cn/32083.14.ati2024050

Citation: Wang, W. Q., Wang, J., Ye, X. C., et al. 2025. Application of AI technology in pulsar candidate identification. *Astronomical Techniques and Instruments*, 2(1): 27−43. https://doi.org/10.61977/ati2024050.

Abstract: As artificial intelligence (AI) technology has continued to develop, its efficient data processing and pattern recognition capabilities have significantly improved the precision and speed of decision-making processes, leading to widespread application across various fields. In astronomy, AI techniques have demonstrated unique advantages, particularly in the identification of pulsars and their candidates. AI is able to address the challenges of pulsar celestial body identification and classification because of its accuracy and efficiency. This paper systematically surveys commonly used AI models for pulsar candidate identification, analyzing and discussing the typical applications of machine learning, artificial neural networks, convolutional neural networks, and generative adversarial networks in candidate identification. Furthermore, it explores how the introduction of AI techniques not only enhances the efficiency and accuracy of pulsar identification but also provides new perspectives and tools for pulsar survey data processing, thus playing a significant role in advancing pulsar research and the field of astronomy.

Keywords: AI technology; Candidate identification; Machine learning; Neural networks

---

1. INTRODUCTION

With the rapid development of computer science and artificial intelligence (AI) technology, their use in the field of astronomy has gradually grown, demonstrating unique advantages in the identification of pulsars and their candidates. Deep learning methods can accurately and efficiently process and analyze the massive astronomical observation data collected by radio telescopes for celestial object identification and classification [1]. This paper discusses several commonly used AI models and their typical applications in pulsar candidate identification. Focusing on the challenges of pulsar celestial identification and classification, we emphasize the use of artificial neural networks (ANNs) [2], convolutional neural networks (CNNs) [3], and generative adversarial networks (GANs) [4] in image recognition and data augmentation. AI not only expedites the identification of pulsar candidates but also enhances the depth and breadth of data analysis, leading to new perspectives and methods for astronomical research. This technology demonstrates immense potential and value in the identification of pulsar candidates and broader astronomical studies, indicating that it will play an increasingly important role in future scientific exploration.

2. COMMON AI MODELS FOR PULSAR CANDIDATE IDENTIFICATION

A crucial branch of computer science, AI [5] has become an important driving force for technological innovation and industrial transformation. Its core lies in simulating and augmenting human intelligence through research on theories, methods, technologies, and application systems, forming a new type of technological science. To address the challenges of searching for and identifying pulsar candidates, AI techniques provide various effective solutions and technical means. Specifically, machine learning (ML) algorithms [6] demonstrate robust learning and prediction capabilities when processing large-scale astronomical data; ANNs [7] efficiently process complex data by mimicking the operational mechanisms of biological neural networks; CNNs [1] excel in image recognition and analysis, accurately extracting pulsar signal features from massive datasets obtained by radio telescopes; and GANs [8] use adversarial training between generative and discriminative models to further enhance data generation and classification capabilities.

The application of these AI models not only improves the accuracy and efficiency of pulsar candidate identification but also provides new methods and tools for astronomical research, promoting a leap forward in astronomical data processing and analysis.

2.1. ML

ML [9] is a class of computer techniques that are trained on a large volume of data samples to obtain certain model parameters, enabling the analysis and prediction of new samples. Based on whether the training data contains features and labels, ML can be classified into three main types: supervised learning, unsupervised learning, and semi-supervised learning.

Supervised learning [10] refers to input data comprising input feature values and target values, where the output function value can either be a continuous value (as in regression problems) or a finite set of discrete values (as in classification problems). Given a training dataset (including features and labels), the model learns the mapping relationship between input features and target values, allowing it to predict new data. Common algorithms in supervised learning include linear regression [11], logistic regression [12], support vector machines (SVMs) [13], decision trees [14], random forests [15], and neural networks.

Unsupervised learning [16], by contrast, refers to scenarios in which input data consists solely of input feature values without target values, making it impossible to directly judge the characteristics of samples within the training data. This method is primarily used for clustering problems, analyzing the underlying structure of datasets to group data and uncover potential patterns or structures. Common unsupervised learning algorithms include K-means clustering [17], hierarchical clustering [18], principal component analysis, independent component analysis, and self-organizing maps.

Semi-supervised learning [19] merges the advantages of both supervised and unsupervised learning. In traditional supervised learning, each training data sample includes a target value (label), whereas in unsupervised learning, samples lack labeled target values. By using both labeled and unlabeled data for training, semi-supervised learning enhances model recognition performance. This method is particularly suitable for scenarios where labeling data is costly or difficult to obtain, and it has been widely applied in fields such as image classification and text classification.

Using various ML methods and techniques, data can be effectively processed and analyzed, providing corresponding solutions for different types of problems. This not only accelerates the rapid development of AI technology but also introduces new opportunities and challenges for research and applications in the field of astronomy.

2.2. Deep Learning

Deep learning [20] is a learning method that uses the multilayer perceptron structure to process nonlinear information. Deep learning models can automatically extract features and perform tasks such as classification, regression, and clustering. By training on large-scale datasets, deep learning not only improves the accuracy and generalization ability of models but also effectively avoids the laborious process of manual feature extraction and model training in traditional ML [21]. According to the types of data being processed, different network structures for learning tasks, and the connectivity and hierarchical relationships between neurons, deep learning can be classified into various types, including ANNs, CNNs, recurrent neural networks, long short-term memory networks, GANs, and autoencoders.

ANNs [22] are computational models inspired by the structure and function of the human brain' s neural networks to process information. They consist of numerous processing units, or neurons, interconnected through a series of connections [23] (as shown in Fig. 1 [Figure 1: see original paper]). Neurons receive input signals from other neurons, process these signals through an internal nonlinear transformation, and then output the results to other neurons. ANNs can perform complex parallel processing tasks using this structure and play a significant role in various fields, including classification, prediction, and pattern recognition. The basic structure of an ANN includes an input layer, hidden layers, and an output layer. The input layer receives external data, the hidden layers are responsible for data processing and feature extraction, and the output layer produces the final output of the network based on the processed information. The number of hidden layers and the number of neurons in each layer are critical parameters in network design, directly affecting the network' s performance and complexity. In practical applications, the number and configuration of neurons in the hidden layers are typically determined based on the specific problem and characteristics of the dataset. ANNs provide powerful technical support for handling complex data analysis and pattern recognition problems by simulating the information processing mechanisms of the human brain' s neural system. With advancements in computational power and continuous optimization of algorithms, ANNs are expected to play an increasingly important role in scientific research and industrial applications.

The ANN model is based on a learning process that uses the backpropagation algorithm, which adjusts the synaptic weights in the network to minimize the difference between the network' s output and actual results [24]. Through extensive iterative calculations, an optimal set of weight values is found, enabling the network to accurately map input data to the desired output. Nonlinear processing capability is another crucial characteristic of ANN models, allowing them to handle problems that traditional algorithms struggle with by learning and simulating complex functional relationships. Deep learning techniques, based on deep neural networks (DNNs), have become a popular research topic in the AI field and have demonstrated outstanding performance in various areas, such as image recognition and natural language processing [25].

CNNs [26] are designed to mimic the hierarchical feature extraction mechanism of the human brain' s visual cortex, making them particularly well-suited for tasks such as image recognition, classification, and localization. The core components of a CNN [27] include the input layer, convolutional layers, pooling layers, fully connected layers, and output layer, which work collaboratively (as illustrated in Fig. 2 [Figure 2: see original paper]) to enable the model to automatically learn and extract features from images, thereby facilitating the efficient processing of complex image data.

In 1998, Lecun et al. [28] proposed the LeNet-5 model (as shown in Fig. 3 [Figure 3: see original paper]), which marked the birth of CNNs. This model was initially designed for handwritten digit recognition and successfully demonstrated the potential of CNNs for specific image recognition tasks. LeNet-5 effectively extracts key features from input images through a series of convolutional and pooling operations, followed by classification decisions made in the fully connected layers. This process highlights the powerful ability of CNNs to automatically learn features. Subsequent research by Lou et al. further expanded the application of CNNs in the field of image recognition. By introducing convolutional operations to extract feature values from images, their research used these values for image recognition and analysis. Not only did this validate the effectiveness of CNNs in image recognition tasks, it also propelled the further development of image recognition techniques. By optimizing the size, number, and depth of the convolutional kernels, as well as adjusting the pooling strategies, the recognition accuracy and generalization ability of the model were enhanced, leading to significant results across a broader range of image recognition tasks [29]. As a powerful deep learning model, CNNs flexibly apply their unique structure and efficient feature extraction capabilities in fields such as image processing [30] and pattern recognition, making them an essential tool in modern AI research.

GANs [31] are a cutting-edge class of generative models that have garnered widespread attention in both academia and industry in recent years. A GAN consists of two main components: a generator and a discriminator (as shown in Fig. 4 [Figure 4: see original paper]). These two networks engage in a process of adversarial training, competing against each other to achieve a common optimization goal. The generator aims to learn the distribution of real data and generate new, seemingly real samples to deceive the discriminator, while the discriminator' s task is to differentiate whether the input samples come from real data or have been fabricated by the generator. This adversarial training mechanism endows GANs with powerful generative capabilities, enabling them to capture the intrinsic structures and distribution characteristics of complex data. GANs are widely applied in various fields, including sample data generation, image generation, image restoration, image-to-image translation, and natural language generation. In image generation, GANs can produce high-quality and diverse images, significantly advancing the fields of computer vision and graphics. In image restoration, GANs can effectively fill in missing parts of images, restoring their completeness and aesthetic appeal. In image translation, GANs facilitate transformations between different styles, resolutions, or modalities, providing robust tools for image editing and enhancement. In natural language generation, GANs also demonstrate substantial potential for generating fluent and coherent textual content. As a generative model with robust capabilities, GANs exhibit broad application prospects and profound research value across multiple domains, and they are poised to exert a significant impact on related fields, driving ongoing innovation and progress in related technologies.

3. APPLICATIONS OF AI IN PULSAR IDENTIFICATION

The application of AI in radio astronomy, particularly in the field of pulsar identification and timing, has not only enhanced data processing efficiency but also significantly broadened our understanding of the universe. By incorporating advanced algorithms such as deep learning and residual networks, scientists are now able to analyze complex astronomical data more accurately and swiftly, thus providing robust support for cutting-edge scientific areas such as pulsar physics and gravitational wave detection. Zhang et al. [32] noted that the introduction of AI techniques has notably improved the efficiency of pulsar searches and candidate identification, enabling a more comprehensive analysis of the spatial distribution and temporal evolution of pulsars. They also summarized the current AI models applied to pulsar candidate identification. Zhang et al. [33] introduced relevant ML theories for pulsar candidate identification and reviewed the research status of methods based on SVMs and neural networks in recent years. With continuous technological advancements, the prospects of AI applications in astronomy will become even broader, potentially leading humanity to explore more mysteries of the universe.

In modern astronomy, radio telescopes such as the Five-hundred-meter Aperture Spherical Telescope (FAST) [34,35], located in Guizhou, China, serve as major national scientific infrastructure projects. FAST [36] is the world’ s largest single-dish radio telescope and plays a crucial role in exploring cosmic mysteries. Particularly in pulsar survey projects, the introduction of AI methods has greatly enhanced the efficiency and accuracy of data processing and analysis in the face of massive observational data.

In the pulsar survey project, which generates millions of pulsar candidates, Xu et al. [37] employed AI methods to automatically screen and analyze the characteristics of FAST’ s vast data. They summarized existing pulsar candidate screening methods regarding test data, accuracy, and construction methods, emphasizing the importance of AI techniques in astronomical data processing. Notably, researchers often face challenges because of the large volume and complexity of data in the selection process of pulsar candidates, whereas AI can effectively recognize and classify potential pulsar candidates through efficient pattern recognition and data mining techniques.

3.1. Application of ML in Identifying Pulsar Candidates

In the field of pulsar research, the introduction of ML has significantly enhanced the process of identifying and filtering pulsar candidates. Devine et al. [42] first applied ML to single-pulse search candidate identification in 2016, proposing a two-stage method for recognizing and classifying dispersed pulse groups. By training and testing binary and multiclass versions of six ML algorithms (as shown in Table 2), they empirically evaluated 48 classifiers, achieving automated filtering with a performance efficiency rate of 41%, which notably improved the efficiency of candidate identification.

Punia et al. [43] conducted experiments using actual data from the HTRU survey, comparing the performance of learning algorithms such as SVMs, random forests, and DNNs (as shown in Table 3). They assessed the accuracy and efficiency of these algorithms in identifying pulsar signals. The study found that DNNs exhibited superior performance in pulsar detection, as demonstrated by their higher accuracy and faster detection speeds.

To address the time-consuming issue of labeling massive datasets, Balakrishnan et al. [44] proposed a semi-supervised generative adversarial network (SGAN) model that greatly expanded the training set using a small amount of labeled data, thereby improving the efficiency and accuracy of candidate identification. The model achieved an accuracy of 94.9% and an average F-score of 94.9% when trained on 100 labeled and 5,000 unlabeled candidates. When trained on a larger labeled dataset, it achieved an accuracy of 99.2% and average F-score of 99.2% with a recall rate of 99.7%.

To tackle the automatic classification problem of fast transient astronomical phenomena, Agarwal et al. [45] introduced a deep learning-based classification model called Fast Extragalactic Transient Candidate Hunter (FETCH) that uses 11 different deep-learning architectures. Each model achieved accuracy and recall rates of over 99.5% on a test dataset comprising real radio frequency interference and pulsar candidate data, demonstrating outstanding classification performance. The FETCH classifier enables its models to be deployed alongside any symbiotic search pipeline, facilitating rapid identification and classification of newly observed fast transient events.

Zhang et al. [46] explored single-pulse candidate identification methods based on FAST observational data by comparing seven different ML classifiers, with results indicating that the SPEGID Model performed best, yielding an F1-score of 95.1%, recall rate of 95.4%, and false positive rate of 4.7%, further enriching the theory and practice of pulsar identification.

For the problem of extracting large-scale and rapidly sampled scalar data, Liu [47] proposed a clustering-based pulsar candidate screening scheme. Comparative experimental results on the High Time Resolution Universe Survey II (HTRU2) pulsar dataset and the actual Fast Transient Bursts (FAST) observational dataset AOD-FAST indicated that this algorithm achieved an accuracy of 0.946 and recall rate of 0.905 on HTRU2, while on AOD-FAST, the F1-score was 0.846 and recall rate was 0.994, effectively meeting the needs of unsupervised and semi-supervised learning scenarios.

Wang et al. [48] designed a residual network model consisting of 15 layers (as shown in Fig. 6 [Figure 6: see original paper]) to replace CNNs in the pulsar image-based classification system (PICS). Their experimental results demonstrated that this model could classify 96% of real pulsars in the top 1% of all candidate pulsars. To evaluate the practical applicability of the model, they conducted tests using dual GPUs and a 24-core computer, showing that the model could classify over 1.6 million candidate pulsars daily.

With the development of radio telescope technology, there has been a significant increase in the quantities and types of pulsar candidate data. This poses a challenge in accurately identifying pulsar candidates. To address this issue, Liu et al. [49] proposed a deep learning model called the multimodal fusion-based pulsar identification model (MFPIM) (as shown in Fig. 7 [Figure 7: see original paper]), which uses a multimodal fusion technique to improve the efficiency and accuracy of pulsar candidate identification. In MFPIM, each diagnostic image of a pulsar candidate is considered as a modality, and multiple CNNs are employed to extract effective features from these diagnostic images. Experiments conducted on the FAST dataset demonstrated that MFPIM can effectively identify pulsars within the dataset with an identification accuracy exceeding 98%. Furthermore, to evaluate the robustness of the model, transfer learning was used to apply MFPIM on a high time-resolution cosmic dataset. The test accuracy and F1-score achieved in this scenario exceeded 99%. These results illustrate the effectiveness and reliability of MFPIM in pulsar identification tasks.

To address the adverse effects of multi-view heterogeneous data and class imbalance on traditional single-modal supervised classification methods for real pulsar and non-pulsar candidate objects, You et al. [50] proposed a multimodal semi-supervised learning method based on a pulsar candidate screening algorithm. The method uses a hybrid ensemble clustering scheme based on density and partitioning (as shown in Fig. 8 [Figure 8: see original paper]). It also incorporates feature-level fusion strategies and data partitioning strategies for parallelization. Experiments conducted on the HTRU2 and actual FAST observation data demonstrate the effectiveness of this algorithm in pulsar identification. On HTRU2, the parallel mode achieved precision and recall rates of 0.981 and 0.988, respectively. On the FAST data, the values of its parallel mode are 0.891 and 0.961. Moreover, as the number of parallel nodes increased within a certain range, the runtime significantly decreased.

In response to the extreme imbalance data issue faced in pulsar candidate selection, Zeng et al. [51] pre-trained a model using the ImageNet dataset (as shown in Fig. 9 [Figure 9: see original paper]), analyzed the performance of the model on the pulsar dataset, and evaluated the transferability of the model using the LogME method. They then proposed a pre-training method. The LogME results were reported for 14 models, with experimental evidence indicating that the F1-score reached 85.56 for the swin_{v2}s model, which achieved a precision of 0.9143 and a recall rate of 0.8040. This approach not only can be applied to pulsar screening tasks but also can enhance the model' s effectiveness, simplify the process, and improve timeliness.

3.2. Application of ANNs in Pulsar Recognition

Morello et al. [52] constructed a basic network model based on artificial neurons (as shown in Fig. 10 [Figure 10: see original paper]) and proposed a direct method for identifying pulsars using a neural network called SPINN. Using the SPINN algorithm to analyze 2,400 candidate stars generated during post-processing, they successfully discovered four new pulsars (as shown in Fig. 11 [Figure 11: see original paper]), accounting for 0.06% of the survey output. Among them, three are millisecond pulsars, while one has a signal-to-noise ratio below 10, providing new clues and evidence for further exploration of the mysteries of the universe.

Eatough et al. [53] introduced an innovative technique in the field of pulsar surveys, using the ANN algorithm to automatically identify credible pulsar candidates. By applying this method to reanalyze candidates in the Parkes Multibeam Pulsar Survey, 34 previously unrecognized pulsars were successfully discovered, significantly increasing the discovery rate of pulsars. Their research shows that approximately 92% of pulsars in a test sample of 2.5 million candidate pulsars can be recovered by their neural network.

Bates et al. [54] further explored the development and application of ANN in data processing pipelines, especially the potential in classifying pulsar candidates. Using basic statistical data obtained from pulsar candidate plots as inputs for ANN, pulsar candidates were effectively classified, leading to the discovery of 75 pulsars in the HTRU survey data processing. The results indicate that this model can reject over 99% of non-target candidates in the data processing stage while achieving blind detection capability for 85% of pulsars. This study not only demonstrated the strong capabilities of ANNs in astronomical data processing but also provided a more efficient and accurate identification method for future pulsar surveys, holding significant importance in the pulsar research field.

3.3. Application of CNNs in Pulsar Identification

Zhu et al. [55] proposed an innovative classification system, PICSAI, which features a two-layer hierarchical structure (as shown in Fig. 12 [Figure 12: see original paper]). This system is based on pulsar images and aims to identify candidate pulsars. By analyzing four key features–average pulse profile, time-phase diagram, frequency-phase diagram, and dispersion measure curve–the candidates were observed using pulsar PSR J1914+08 as an example (as shown in Fig. 13 [Figure 13: see original paper]). The incorporation of features with algorithms enabled the successful recovery of over 90% of legitimate candidate pulsars, with a corresponding false positive rate below 0.5%. This method demonstrated significant effectiveness in early pulsar candidate identification, providing a powerful tool for pulsar research.

Wang et al. [56] further advanced this field by designing a deep CNN (DCNN) to analyze the features of pulsar candidates in greater detail. The DCNN consists of eight convolutional layers (as shown in Fig. 14 [Figure 14: see original paper]), which are used to extract complex features from the sub-integrated and sub-band images of the candidates. The DCNN can automatically learn the intricate patterns of pulsar candidates, thus exhibiting excellent performance in classification tasks. A pooling layer is introduced to reduce the dimensionality of feature maps while retaining key information, followed by two fully connected layers for precise classification of the deeply analyzed 11-layer pulsar candidates. Experiments on the HTRU1 dataset showed that the model achieved a recall rate of 0.962 and an accuracy of 0.963. This not only improved the accuracy of pulsar identification but also provided a more robust tool for subsequent pulsar studies, advancing the field of astrophysics.

Yin et al. [57] proposed a deep CNN model (called AR_{Net}) that incorporates attention mechanisms and residual connections (as shown in Fig. 15 [Figure 15: see original paper]), aiming to enhance the accuracy of pulsar identification. The constructed AR_{Net} model strengthens the network's learning ability regarding pulsar features, enabling the model to exhibit better generalization and recognition efficiency when processing complex data. Experimental results indicate that this method achieved an F1-score of 0.9981, a recall rate of 0.9988, and an accuracy of 0.9975. Compared with the DCGAN+SVM method, all metrics —F1-score, recall rate, and accuracy—showed significant improvement. Comprehensive evaluations in multiple experiments demonstrated that the AR_{Net} model exhibits high recognition precision and reliability in distinguishing between positive and negative samples of pulsars.

To address the issue of detecting transients in survey images, Cabrera et al. [58] proposed a rotation-invariant CNN model (as shown in Fig. 16 [Figure 16: see original paper]) designed to classify transient candidate images as artifacts or real sources, particularly in high-transient survey (HiTS) contexts. Practical implementation of this classification model has shown a significant improvement in the overall accuracy of the HiTS pipeline, increasing from 98.96%$±0.03±$0.03%. Using Deep-HiTS, approximately 2,000 transients can be processed each night, successfully reducing the number of missed transients to about one-fifth.

3.4. Application of GANs in Pulsar Identification

Radford et al. [59] proposed a deep convolutional GAN (DCGAN) model (shown in Fig. 17 [Figure 17: see original paper]), which integrates a GAN with a CNN and uses convolutional operations to extract spatial features, significantly improving the clarity and quality of the generated images. This combination allows DCGAN to excel in image generation tasks, producing images that closely resemble real ones.

To address the issue of imbalanced datasets, Douzas et al. [60] introduced an imbalance data oversampling method based on conditional GANs (CGAN). They compared the performance of CGAN with various standard oversampling algorithms and found that CGAN could more effectively enhance model performance when dealing with imbalanced datasets, especially on binary classification tasks. CGAN notably improved the model' s performance on minority class samples, thereby enhancing the overall model' s generalization capability.

Wang [61] proposed a pulsar candidate recognition method based on a semi-supervised attention GAN (SA-SGAN; shown in Fig. 18 [Figure 18: see original paper]). The research results indicated that, compared with traditional semi-supervised GAN models, this method achieved a 4.41% increase in recall rate and a 3.39% increase in F1-score when handling small sample datasets.

Zhou [62] proposed a semi-supervised learning model called SSL-ATJD (shown in Fig. 19 [Figure 19: see original paper]), which was applied to pulsar candidate datasets to address recognition issues in imbalanced datasets. Experiments showed that, for labeled samples with (200, 3,800) time markers, the SSL-ATJD model achieved a recognition effect of 87.88%, slightly outperforming the CNN model, which had a recognition effect of 87.25% across all training samples (as shown in Table 4). This method not only improved the model' s accuracy in pulsar identification tasks but also effectively used limited labeled data. By leveraging the characteristics of GAN, it enhanced the model' s ability to recognize pulsar features, thus achieving better recognition results on imbalanced datasets.

Currently, significant progress has been made in AI models used for pulsar candidate identification. These models can efficiently and rapidly extract useful features from complex signals and classify them accordingly (as listed in Table 5).

4. DISCUSSION

This study conducted an in-depth research and analysis on the application of AI technology in identifying pulsar candidates. The improvement in observational capabilities has led to a sharp increase in data volume. Traditional data processing and analysis methods are no longer sufficient to handle such massive scales of data. Therefore, training AI models based on massive-scale big data will become key to enhancing data recognition capabilities. This approach will lead to new algorithms and model optimization strategies for pulsar candidate identification, effectively addressing key issues such as data processing, feature learning, and model transfer, providing strong technical support for the discovery and study of pulsars.

While AI addresses the issue of data volume, enhancing real-time data processing technology to automate and make systems more intelligent will be a future development direction in pulsar candidate identification. The aim of this approach will be to enhance processing efficiency and accuracy by reducing manual intervention. Furthermore, the integration and fusion of technologies are crucial for future development. By combining real-time data processing techniques with advanced AI methods such as ML, neural networks, and federated learning, a more powerful and comprehensive processing system can be constructed. To cope with the constantly changing scale of data and processing requirements in the future, the design of real-time data processing systems will emphasize scalability and flexibility to ensure that systems can adapt to the evolving astronomical observation environment. Introducing more powerful attention mechanisms, graph convolutional networks, or GANs and other AI methods will enable more complex data patterns to be addressed and improve model generalization ability. These new architectures are expected to play a greater role in pulsar candidate identification, enhancing accuracy and efficiency.

AI methods have made significant progress in pulsar recognition and have attracted widespread attention in the scientific community. However, there are still limitations in current AI techniques for pulsar recognition, which require further research and improvement. In pulsar recognition, different data sources and sensors are commonly used to obtain different types of information, such as radio waveband, X-ray waveband, and gamma-ray waveband data. It is challenging to effectively integrate multi-modal data to improve the recognition performance of pulsars. The cost of manually labeling a large amount of data is high, and the labeled samples may be limited or only have partial label information. How to efficiently identify pulsar candidates using limited labeled data and unlabeled data is an urgent problem that must be solved. The development of weakly supervised learning provides more possibilities for pulsar candidate recognition and helps improve recognition performance in situations where data annotation is difficult. Reinforcement learning is not commonly applied in pulsar recognition, but it could optimize the decision-making strategy of a system by interacting with the environment to improve the performance of pulsar recognition. Further research is needed on how to design appropriate state representations, reward functions, and suitable reinforcement learning algorithms for pulsar recognition.

Although AI has achieved important achievements in pulsar recognition, there are still limitations. In the future, AI techniques such as multi-modal learning, weakly supervised learning and reinforcement learning could be applied to the field of pulsar candidate recognition for further exploration and research to improve the accuracy and efficiency of pulsar recognition. In the study of pulsar identification, computer science provides powerful data processing capabilities, physics offers a profound understanding of pulsar characteristics, and astronomy provides observation data and research scenarios. Only by combining the expert knowledge and technologies from these fields can we comprehensively and deeply understand pulsars and unleash the full potential of AI in pulsar identification. Therefore, interdisciplinary cooperation not only promotes knowledge sharing among disciplines but also advances the boundaries of scientific research. By collaborating with experts in computer science, physics, astronomy, and other fields, we can collectively explore the application of AI in pulsar candidate identification, providing a more comprehensive and in-depth perspective for astronomical research. Interdisciplinary cooperation is also an important direction for future research. AI not only improves the efficiency and accuracy of pulsar identification but also provides new perspectives and tools for pulsar survey data processing, making significant contributions to pulsar research and the development of astronomy.

5. CONCLUSIONS

This study systematically investigated commonly used AI techniques for identifying pulsar candidates, analyzing and discussing the typical applications of ML, ANNs, CNNs, and GANs. It showed that the introduction of AI technology not only enhances the efficiency and accuracy of pulsar identification but also provides new perspectives on pulsar survey data processing. With the continuous development of technology and the enhancement of data processing capabilities, the prospects for AI in astronomy remain very broad, exerting profound influence on pulsar research and other astronomical research fields. Further research into innovative neural network architectures, multimodal data fusion, deep reinforcement learning, and collaboration with domain experts will further drive the application of AI in pulsar candidate identification, providing more accurate and comprehensive results for astronomical research, and this research is expected to achieve more breakthroughs in pulsar studies. In the future, it will be possible to focus on optimizing the quality and scale of pulsar datasets, improving the signal-to-noise ratio of the data; reducing noise interference; and exploring new denoising, feature extraction, and normalization preprocessing methods to optimize the quality of input data. Selecting appropriate features is crucial for the accurate identification of pulsar candidates. It would also be possible to study feature extraction methods in the time domain, frequency domain, and pulse profile; reduce feature dimensions and complexity; and improve classification performance through research on automatic feature selection algorithms. In the process of pulsar candidate identification, multimodal data fusion provides more comprehensive and accurate information. In the future, it would be possible to explore multimodal data fusion methods that integrate radio waveband data with optical, X-ray, or gamma-ray data to improve the accuracy and robustness of classification. By conducting research in data optimization and preprocessing, feature extraction and selection, model optimization and integration, uncertainty modeling and interpretability, and multimodal data fusion, the accuracy and efficiency of pulsar candidate identification can be improved.

ACKNOWLEDGEMENTS

This work is supported by the National Key R&D Program of China (2021YFC2203502, 2022YFF0711502); the National Natural Science Foundation of China (NSFC) (12173077); the Tianshan Talent Project of Xinjiang Uygur Autonomous Region (2022TSYCCX0095 and 2023TSYCCX0112); the Scientific Instrument Developing Project of the Chinese Academy of Sciences (PTYQ2022YZZD01); China National Astronomical Data Center (NADC); the Operation, Maintenance and Upgrading Fund for Astronomical Telescopes and Facility Instruments, budgeted from the Ministry of Finance of China (MOF) and administered by the Chinese Academy of Sciences (CAS); Natural Science Foundation of Xinjiang Uygur Autonomous Region (2022D01A360).

AUTHOR CONTRIBUTIONS

Wanqiong Wang conceived the idea, provided investigation support, and wrote the original draft. Jie Wang reviewed the manuscript, and performed the project administration and supervision roles. Xinchen Ye and Yazhou Zhang reviewed the manuscript and performed supervision. Jia Li, Xu Du, Wenna Cai, Han Wu, Ting Zhang, Yuyue Jiao provided investigation support. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Zhang, H. Y., Zhao, Z., An, T., et al. 2019. Pulsar candidate recognition with deep learning. *Computers & Electrical Engineering*, 73: 1−8.

Haykin, S. 2009. *Neural networks: A comprehensive foundation* (3rd Edition). London: Pearson.

Lou, G. X., Shi, H. Z. 2020. Face image recognition based on convolutional neural network. *China Communications*, 17(2): 117−124.

Zeng, Q. 2024. Research on image recognition method based on generated adversarial networks. Master thesis, Changsha University of Science and Technology.

Lin, Y. R., Ma, S. P. 1989. *An introduction to artificial intelligence*. Beijing: Tsinghua University Press. (in Chinese)

Haindavi, P., Kumar, S., Ganesh, G., et al. 2024. Discovery of astronomical objects in galaxies by means of deep learning. *MATEC Web of Conferences*, 392: 01123.

Singh, K. K., Dhar, V. K., Meintjes, P. J. 2022. Artificial neural networks for cosmic gamma-ray propagation in the universe. *New Astronomy*, 91: 101701.

García-Jara, G., Protopapas, P., Estévez, P. A. 2022. Improving astronomical time-series classification via data augmentation with generative adversarial networks. *The Astrophysical Journal*, 935(1): 23.

Zhou, Z. H. 2016. *Machine learning*. Beijing: Tsinghua University Press.

Cunningham, P., Cord, M., Delany, S. J. 2008. Supervised learning. In *Machine learning techniques for multimedia*. Springer, 21−49.

Neter, J., Wassermann, W., Kutner, M. 1974. *Applied linear statistical models: Regression, analysis of variance, and experimental designs*. Boca Raton: CRC Press.

Kleinbaum, D. G., Klein, M. 2002. *Logistic regression (A self-learning text)*. Berlin: Springer.

Cristianint, N., Shawe-Taylor, J. 2000. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge: Cambridge University Press.

Hastie, T., Tibshirani, R., Friedman, J. 2008. *The elements of statistical learning: data mining, inference, and prediction*. Berlin: Springer.

Breiman, L. 2001. Random forests. *Machine Learning*, 45: 5−32.

Olaode, A., Naghdy, G., Todd, C. 2014. Unsupervised classification of images: A review. *International Journal of Image Processing*, 8(5): 325−342.

Jain, A. K. 2010. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8): 651−666.

Kaufman, L., Rousseeuw, P. J. 2005. *Finding groups in data: An introduction to cluster analysis*. New York: Wiley.

Hady, M. F. A., Schwenker, F. 2013. Semi-supervised learning. In *Handbook on Neural Information Processing*, Springer Nature Link, 215−239.

LeCun, Y., Bengio, Y., Hinton, G. 2015. Deep learning. *Nature*, 521: 436−444.

Janiesch, C., Zschech, P., Heinrich, K. 2021. Machine learning and deep learning. *Electronic Markets*, 31(3): 685−695.

Kanwisher, N., Khosla, M., Dobs, K. 2023. Using artificial neural networks to ask 'why' questions of minds and brains. *Trends in Neurosciences*, 46(3): 240−254.

Smith, M. J., Geach, J. E. 2023. *Astronomia ex machina: a history, primer and outlook on neural networks in astronomy. Royal Society Open Science*, 10(5): 221454.

Wang, Z. H., Ye, Y., Liu, H. Y., et al. 2024. Spatio-temporal back-propagation algorithm for deep pulse neural network based on pulse sequence identity. *Journal of Electronics & Information Technology*, 46(6): 2596−2604. (in Chinese)

Chen Yuncai. 2024. Research on natural language processing technology based on artificial neural network. *Engineering and Technological Research*, 9(8): 93−95. (in Chinese)

Zhou, F. Y., Jin, L. P., Dong, J. 2017. Review of convolutional neural network. *Chinese Journal of Computer Science*, 40(6): 1229−1251. (in Chinese)

Chang, L., Demg. X. M., Zhou, M. Q., et al. 2016. Convolutional neural network in image understanding. *Acta Automatica Sinica*, 42(9): 1300−1312. (in Chinese)

Lecun, Y., Bottou, L. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11): 2278−2324.

Wang, W., Zhang, H. M. 2024. High-similarity image recognition method based on a convolutional neural network. *Journal of Heilongjiang University of Technology (Comprehensive edition)*, 24(3): 80−84. (in Chinese)

Luo, F. G., Song, Q., Qin, Y. C., et al. 2024. Research on the application of convolutional neural network in image recognition. *Computer and Information Technology*, 32(3): 51−54. (in Chinese)

Wang, J. Y., Yang, H. T., Li, G. Y., et al. 2021. Progress in the application of generating adversarial networks and their image processing. *Computer Engineering and Application*, 57(8): 10. (in Chinese)

Zhang, H. L., Wang, J., Zhang, Y. Z., et al. 2024. Review of artificial intelligence applications in astronomical processing. *Astronomical Techniques and Instruments*, 1(1): 1−15.

Zhang, C. J., Shang, Z. H., Chen, W. M., et al. 2020. A review of research on pulsar candidate recognition based on machine learning. *Procedia Computer Science*, 166: 534−538.

Nan, R. D. 2006. Five hundred meter aperture spherical radio telescope (FAST). *Science China (Physics, Mechanics & Astronomy)*, 49(2): 129−148.

Nan, R. D., Li, D. 2013. The five-hundred-meter aperture spherical radio telescope (FAST) project. In *Proceedings of IOP Conference Series: Materials Science and Engineering*.

Nan, R. D., Wang, Q. M., Zhu, L. C., et al. 2006. Pulsar observations with radio telescope FAST. *Chinese Journal of Astronomy and Astrophysics*, 6(Suppl 2): 304.

Xu, Y., Li, D., Liu, Z. J., etc. 2017. Application of artificial intelligence in the screening of pulsar candidates. *Progress in Astronomy*, 35(3): 304−315. (in Chinese)

Wu, X. J. 2017. Parkes radio telescope with pulsar survey discovery −Memorial pulsar discovery 50 years. *Science*, 69(6): 43−49. (in Chinese)

Liu, X. F., Lao, B. Q., An, T., et al. 2021. Research on pulsar candidate identification method based on deep residual neural network. *Acta Astronomica Sinica*, 62(2): 98−111. (in Chinese)

Sarkissian, J. 2010. The Parkes Pulsar Timing Array (PPTA). In *Proceedings of the Conference dedicated to Viktor Ambartsumian's 100th anniversary, Evolution of Cosmic Objects Through Their Physical Activity*.

Manchester, R. N. 2013. The international pulsar timing array. *Classical and Quantum Gravity*, 30: 55−61.

Devine, T. R., Goseva-Popstojanova, K., McLaughlin, M. 2016. Detection of dispersed radio pulses: A machine learning approach to candidate classification. *Monthly Notices of the Royal Astronomical Society*, 459(2): 1519−1532.

Punia, A., Sardana, A., Subashini, M. 2017. Evaluating advanced machine learning techniques for pulsar detection from HTRU survey. In *Proceedings of International Conference on Intelligent Sustainable Systems*.

Balakrishnan, V., Champion, D., Barr, E., et al. 2021. Pulsar candidate identification using semi-supervised generative adversarial networks. *Monthly Notices of the Royal Astronomical Society*, 505(1): 1180−1194.

Agarwal, D., Aggarwal, K., Burke-Spolaor, S., et al. 2020. A deep-learning based classifier for fast transient classification. *Monthly Notices of the Royal Astronomical Society*, 497(2): 1661−1674.

Zhang, B., You, S. P., Xie, X. Y., et al. 2023. Application of single-pulse search candidate identification based on machine learning to FAST observation CRAFTS data. *Progress in Astronomy*, 41(3): 415−428. (in Chinese)

Liu, Y. 2022. Unsupervised clustering analysis study of large-scale pulsar candidate body signals. Master thesis, Guizhou Normal University.

Wang, H. F., Zhu, W. W., Guo, P., et al. 2019. Pulsar candidate selection using ensemble networks for FAST drift-scan survey. *Science China Physics, Mechanics & Astronomy*, 62: 959507.

Liu, Y., Jin, J., Zhao, H. Y., et al. 2023. MFPIM: A deep learning model based on multimodal Fusion technology for pulsar identification. *The Astrophysical Journal*, 954(1): 86.

You, Z. Y., Pan, Y. R., Ma, Z., et al. 2024. Applying hybrid clustering in pulsar candidate sifting with multi-modality for FAST survey. *Research in Astronomy and Astrophysics*, 24(3): 035022.

Zeng, P., Yu, X. H., Liu, Z. J., et al. 2024. Application of the pre-trained model in the screening of pulsar candidates. *Automation Application*, 65(8): 227−231. (in Chinese)

Morello, V., Barr, E. D., Bailes, M., et al. 2014. SPINN: a straightforward machine learning solution to the pulsar candidate selection problem. *Monthly Notices of the Royal Astronomical Society*, 443(2): 1651−1662.

Eatough, R. P., Molkenthin, N., Kramer, M., et al. 2010. Selection of radio pulsar candidates using artificial neural networks. *Monthly Notices of the Royal Astronomical Society*, 407(4): 2443−2450.

Bates, S. D., Bailes, M., Barsdell, B. R., et al. 2012. The high time resolution universe survey VI: An artificial neural network and timing of 75 pulsars. *Monthly notices of the Royal Astronomical Society*, 427(2): 1052−1065.

Lyon, R. J., Stappers, B. W., Cooper, S., et al. 2016. Fifty years of pulsar candidate selection: from simple filters to a new principled real-time classification approach. *Monthly Notices of the Royal Astronomical Society*, 459(1): 1104−1123.

Jiang, Y., Jin, J., Yu, Y. L., et al. 2019. Denoising method of pulsar photon signal based on recurrent neural network. In *Proceedings of 2019 IEEE International Conference on Unmanned Systems*, 661−665.

Zhu, K. R., Kang, S. J., Zheng, Y. G. 2021. Searching for AGN and pulsar candidates in 4FGL unassociated sources using machine learning. *Research in Astronomy and Astrophysics*, 21: 015.

Cai, N. N., Han, J. L., Jing, W. C., et al. 2023. Pulsar candidate classification using a computer vision method from a combination of convolution and attention. *Research in Astronomy and Astrophysics*, 23: 10.

Guo, P., Duan, F. Q., Wang, P., et al. 2019. Pulsar candidate classification using generative adversary networks. *Monthly Notices of the Royal Astronomical Society*, 490(4): 5424−5439.

Liu, T. 2020. *Research on pulsar candidate selection method based on artificial intelligence*. Master thesis, Hunan University. (in Chinese)

Yin, Q., Li, Y. F., Li, J. J., et al. 2022. Pulsar-candidate selection using a generative adversarial network and ResNeXt. *The Astrophysical Journal Supplement Series*, 264: 2.

Zhu, W. W., Berndsen, A., Madsen, E. C., et al. 2014. Searching for pulsars using image pattern recognition. *The Astrophysical Journal*, 781: 117.

Wang, Y. C., Li, M. T., Pan, Z. C., et al. 2019. Pulsar candidate classification with deep convolutional neural networks. *Research in Astronomy and Astrophysics*, 19: 133.

Yin, Q., Wang, Y., Zheng, X., et al. 2022. Pulsar candidate recognition using deep neural network model. *Electronics*, 11(14): 2216.

Cabrera-Vives, G., Reyes, I., Förster, F., et al. 2017. Deep-HiTS: rotation invariant convolutional neural network for transient detection. *The Astrophysical Journal*, 836: 97.

Radford, A., Metz, L., Chintala, S. 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv: 1511.06434.

Douzas, G., Bacao, F. 2018. Effective data generation for imbalanced learning using conditional generative adversarial networks. *Expert Systems with Application*, 91: 464−471.

Wang, Z. L. 2023. *Pulsar candidate discrimination method based on semi-supervised learning*. Master thesis, Southwest University of Science and Technology. (in Chinese)

Zhou, L. Y. 2024. *Research of image classification based on generative adversarial nets and its application for pulsar candidate identification*. Doctoral thesis, Guizhou Normal University. (in Chinese)

Tan, C. M., Lyon, R. J., Stappers, B. W., et al. 2018. Ensemble candidate classification for the LOTAAS pulsar survey. *Monthly Notices of the Royal Astronomical Society*, 474(4): 4571−4583.

Note: Figure translations are in progress. See original paper for figures.