



International Journal of Data Science and Big Data Analytics

Publisher's Home Page: <https://www.svedbergopen.com/>



Review Article

Open Access

The Role of AI in Analyzing Astronomical Data: A Literature Review

Abid Hossain Rion^{1*} and Mashrufa Meghla Any²

¹Department of Artificial Intelligence & Astrophysics, MindBridge Innovative Network, Bangladesh. E-mail: abid@minresearch.org

²Department of Artificial Intelligence & Astrophysics, MindBridge Innovative Network, Bangladesh. E-mail: mashrufa@minresearch.org

Article Info

Volume 5, Issue 2, November 2025

Received : 13 June 2025

Accepted : 31 October 2025

Published : 25 November 2025

doi: [10.51483/IJDSBDA.5.2.2025.111-116](https://doi.org/10.51483/IJDSBDA.5.2.2025.111-116)

Abstract

The exponential growth of astronomical data, especially with increasing amounts of processed data from digital surveys and time-domain observations, has thrust Artificial Intelligence (AI) and Machine Learning (ML) methods into the forefront of astronomy. We review the literature to summarize recent research and outline how AI techniques have transformed, automated, and computerized the processing, classification, and interpretation of astronomical data. In the application of algorithms like artificial neural networks, support vector machines, random forest, deep learning, etc. AI has developed a core role in helping to complete astronomy tasks like star-galaxy classification, variable star identification, and transient event discovery. With sizes and scales of data from terabytes to petabytes, previous methods of supervised and unsupervised ML have become necessary, and in some cases, essential to provide to give useful meaning to the data. AI has had advantages over traditional data-processing methods, both in automating routine data processing tasks and in monitoring data sets to find rare astronomical events and help with decision-making in real-time, with applications for time-sensitive processes like detecting gravitational waves and transient events. AI methods will likely be increasingly sophisticated, reliable, and improve in utility, augmenting traditional statistical sampling methods, culminating in major astronomical discoveries. It could be concluded that AI methods are fast evolving and will increasingly shape future discoveries.

Keywords: Artificial intelligence, Machine learning, Astronomy, Astronomical data, Classification, Time-domain astronomy, Neural networks, Data mining, Sky surveys, Transient event detection

© 2025 Abid Hossain Rion and Mashrufa Meghla Any. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

1. The Role of AI in Analyzing Astronomical Data

The field of astronomy has rapidly transformed thanks to the rapid advancements in digital technology and

* Corresponding author: Abid Hossain Rion, Department of Artificial Intelligence & Astrophysics, MindBridge Innovative Network, Bangladesh. E-mail: abid@minresearch.org

2710-2599/© 2025. Abid Hossain Rion and Mashrufa Meghla Any. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

observational methods utilizing data-intensive techniques. To summarize, the 1990s heralded the establishment of high-energy, digital sky surveys, which eventually led to the big data era of astronomy, and with this pioneering transformation in observational methods and subsequent data analysis, massive terabyte-scale datasets, they soon grew to Petascale data streams, as observational information has increased in sensitivity and frequency (Djorgovski *et al.*, 2023). Manual approaches for data processing and categorization are insufficient to manage volumes of information; subsequently, the astronomical community has embraced Artificial Intelligence (AI) and Machine Learning (ML) as part of contemporary data analysis. The application of Artificial Intelligence (AI) in astronomy goes well beyond automation. Machine learning algorithms, particularly supervised learning approaches such as Artificial Neural Networks (ANN), support vector machines, decision trees, and random forests, have proved to be extremely successful tools in solving complex classification and regression problems (Zhooldideh Haghighi, 2023). These algorithms are now routinely applied to astronomical problems like star-galaxy separation, spectral classification, photometric redshift estimation, and the identification of rare or transient objects. With AI's rapid and accurate processing of high-dimensional data, it has become an essential part of both static sky surveys and dynamic time-domain studies. AI models being implemented into real-time decision-making for time-domain astronomy have been among the biggest recent advances. Observational data are continuously generated as synoptic surveys like the Zwicky Transient Facility (ZTF) and Vera C. Rubin Observatory's upcoming Legacy Survey of Space and Time (LSST) produce streams of data. AI techniques allow for the detection and classification of transient events (including supernovae, gravitational wave counterparts, and other astrophysical phenomena) within seconds of their occurrence (Djorgovski *et al.*, 2023; Fluke and Jacobs, 2019). This is significant due to the ephemeral nature of many high-interest astronomical phenomena. Not only can AI create improvements towards efficiency and scalability, but it can also facilitate entirely new modes of scientific discovery. Astronomers can apply unsupervised learning and anomaly detection techniques to identify previously unknown classes of objects or even rare ones of interest, without any labels beforehand. We can progress our understanding of the universe (Fluke and Jacobs, 2019). This indicates a shift from a model-driven analysis to a mixed method of data-driven research with a theoretical anchor. The purpose of this literature review is to illuminate the changing role of AI in the identification of astronomical data, primarily in terms of classification and discovery techniques. This review will comment on the current methods of AI techniques found in the literature while documenting the capabilities and limitations of said techniques. Finally, the review will offer a snapshot of rapidly evolving capabilities, which are ultimately shaping an ever-developing field that may fundamentally change how we study astronomy going forward.

2. Literature Review

The entry of AI and ML into astronomy has progressed from a new pursuit to an established component of the astronomical data analysis structure. The evolution of AI in astronomy stemmed from the expanding volume and variability of data being generated from digital sky surveys, shifting the astronomical community's reliance on AI from a luxury to an essential approach. Early attempts at automation had modest aspirations in the 1990s, as basic as automation on project tasks, like star-galaxy classification; first with decision trees, and then with neural networks. The primary attraction of these techniques was efficiency and data consistency. With catalogs containing millions, if not billions of sources, human classification of sources had become untenable (Djorgovski *et al.*, 2023). Since those early years, the field has burgeoned in methodological complexity and application. Fluke and Jacobs (2019) provide a comprehensive summary of various ML and AI applications across seven categories: classification, regression, clustering, forecasting, generation, discovery, and scientific insight. Classification has received the most attention, and many authors have investigated galaxy morphology, optical identification of objects, and photometric redshift. Classification-based methods, such as random forests, support vector machines, and convolutional neural networks, classified with high accuracy based on complex sets of multidimensional features from image, spectral, or photometric datasets and often matched or exceeded human classification accuracy – benefiting from large data sets across the visible, infrared, and radio regimes. Zhooldideh Haghighi (2023) reinforces this trend by examining classification algorithms applied to Sloan Digital Sky Survey (SDSS) data, applied here to the issue of classifying RR Lyrae variable stars as opposed to non-variable stars. The study showcased the comparative abilities of popular algorithms such as logistic regression, decision trees, support vector machines, and multilayer perceptrons,

and the evaluation metrics (such as F1-score and accuracy) were referenced to provide quantitative validations. The study concluded that ML methods not only improve accuracy but drastically decrease the time needed to classify large amounts of astronomical data.

As observational datasets continue to grow exponentially, the time savings alone could be the most important benefit of ML methods. Beyond classification, time-domain astronomy represents an especially fruitful direction for AI applications because transient events, such as supernovae, gamma-ray bursts, and counterparts to gravitational waves, are dynamic and must be recognized and acted upon in or near real-time. Synoptic surveys such as the ZTF generate massive amounts of data: their multidimensional time series can consist of petabytes of information (Djorgovski *et al.*, 2023). As a result, the idealized models must be quick to learn and adaptable. DNNs, XGBoost, and Gaussian Process Regression (GPR) were used to classify the transient events also but can also be used to detect anomalies, which could be signs of new astrophysical phenomena. Transformative developments have occurred in astronomy in the realm of unsupervised learning, particularly clustering and dimensionality reduction methods, such as k-means, Principal Component Analysis (PCA), and t-distributed Stochastic Neighbor Embedding (t-SNE), which have allowed astronomers to discover outliers and unassigned classes of objects.

We are entering an era of discovery-based science, whereby AI is a collaborator in stimulus generation, instead of an instrument solely for stimulus testing (Fluke and Jacobs, 2019). As multi-messenger astronomy expands (detecting and correlating data from gravitational wave, neutrino, and high-energy cosmic ray observatories), we expect AI systems to be increasingly helpful for correlating and interpreting complex, cross-modal signals. Despite these advances, there are still significant barriers to overcome. In particular, the scaling of algorithms as a function of data dimensionality is problematic. The curse of dimensionality can impact classification performance and lead to noise during classification, particularly in multiclass scenarios, as highlighted by Djorgovski *et al.* (2023). In addition, ML pipelines must include error bars, uncertainties, and incomplete datasets where possible, which is non-trivial and complex. The strategies on error, uncertainty, and incomplete datasets include probabilistic modeling, ensemble methods, and addressing reusable standardized datasets, and are still in the early stages of development.

Overall, the literature agrees that AI is a game-changer for astronomy, not simply a supporting role. AI systems can generally be incorporated into a range of processes within an astronomical pipeline, whether this is supervised classification, unsupervised clustering, or automated anomaly detection in real time. The ML implementations in astronomy are increasingly visible, from producing a feature list through noisy light curves, forecasting solar flares, completing missing spectral build-up, and so forth, which are as diverse as the sources of data in astronomy.

The integration of astronomy and AI not only reflects a change in methodologies but a grander transformation in discovery within the era of data science.

3. Methodology

This literature review takes a qualitative approach to review the current application of Artificial Intelligence (AI) in the field of astronomical data analysis based purely on peer-reviewed scholarly articles. The purpose of the review was to discover themes, methods, and applications of Machine Learning (ML) and AI in different areas of astronomical research. The current status of three foundational papers was considered valuable based on factors of relevance and comprehensiveness of AI techniques used in astronomy.

The first paper, Fluke and Jacobs (2019), provides a broad overview of AI and Machine Learning adoption in astronomy across many application areas. The source is structured in a comprehensive classification structure that outlines seven types of AI techniques: classification, regression, clustering, forecasting, generation, discovery, and insight. The paper includes a hierarchical maturity model that shows levels of AI integration at different levels of subfield, which is useful for comparative purposes.

Djorgovski *et al.* (2023), the second source, also maintains a focused discussion on the many possibilities of AI in the era of Big Data and time-domain astronomy, but it was highlighted because the authors address the challenges of real-time classification and anomaly detection in light of synoptic sky surveys and multi-

messenger astrophysics. It contributes insights to the use of deep learning architectures, probabilistic models, and anomaly detection in order to analyze petascale time-series data.

The third source, by Zhooldideh Haghighi (2023), is a practical comparison of supervised classification algorithms using data from the Sloan Digital Sky Survey (SDSS), and was chosen because of its emphasis on empirical measures, and it has benchmarks (accuracy values) of models, such as decision trees, multilayer perceptrons, and support vector machines. This is a robust study, as well, with regard to empirical assessments of the practical AI/SI methods/classes of variable stars, and the use of F1-scores and accuracy values for the purpose of modelling, which is also grounded for variable stars, in assessing the performance of AI methods.

Data for the review associated with this project were taken from those three works, focusing on their emphasis on classification methods, the types of astronomical data that were used (i.e., images, spectra, light curves), and the use of AI in large-scale survey pipelines.

Citations and examples were cross-referenced to identify trends in algorithm usage, implementation obstacles, and constraints that were faced, all specifically in the context of astronomy. Evidence that highlighted not only technical performance but that demonstrated scientific importance and implications of research using AI in the derivation of astrophysical knowledge was preferred.

By limiting the review to these selected, high-quality sources, the methodology promotes depth instead of breadth. Instead of attempting to catalogue the entire suite of AI-focused astronomy papers, this rigorousness allows for the synthesis of methodological tendencies to better understand the ability to deploy and use AI in transformatively new ways in support of astronomical discovery.

4. Results and Discussion

The utilization of Artificial Intelligence (AI), including Machine Learning (ML), has led to both an improvement in accuracy and efficiency in astronomical data analysis. All sources reviewed reveal a common theme: AI has become an important tool in handling the vast and multidimensional data produced by contemporary sky surveys. The three studies reviewed included AI techniques that showed strong accuracy in object classification, transient detection, photometric redshift estimation, and anomaly detection.

One of the interesting findings across all sources was that supervised learning algorithms performed best in classification. For example, Zhooldideh Haghighi (2023) conducted a study that compared classifiers on SDSS data, specifically for RR Lyrae variable stars. Support Vector Machines (SVM), decision trees, and multilayer perceptrons performed well, all receiving strong F1-scores and accuracy values. These results indicate that traditional supervised learning models can be relied on when enough labeled data are available. In particular, SVMs and neural networks performed well with the multiple photometric data sets with varying quality, because they can learn complex boundaries in high-dimensional feature space.

In contrast, Fluke and Jacobs (2019) define a more comprehensive approach for AI applications as seven functions: classification, regression, clustering, forecasting, generation, discovery, and insight. In their study, classification is the most advanced AI application because it is widely applied. However, they outlined the growing importance of unsupervised learning and anomaly detection. The discovery applications, like finding gravitationally lensed quasars or transients, not only foreshadow how AI will automate certain tasks we've done, but also how AI may help us discover new forms of science. One prime example is the discovery of extrasolar planets and the classification of galaxy morphologies with convolutional neural networks at a level of human expert astronomers!

Djorgovski *et al.* (2023) highlighted the role of AI in time-domain astronomy, where events can rapidly change, and the ability to organize, determine, and classify events as they happen is critical. Surveys like the Zwicky Transient Facility (ZTF) also necessitate real-time processing pipelines capable of processing a large volume of data, identifying acquired data as an important event, amidst a sea of noisy, typically incomplete data. Also, for real-time event filtering, prioritization, and classification to occur in today's surveys, ensemble models and deep neural networks have effectively been used. These systems not only reduce the burden on human researchers looking for phenomena but can also improve how allocation of telescope time occurs (for

example, identifying high-interest phenomena like supernova or gravitational wave counterparts to provide feedback and immediate follow-up).

A theme that was common across the studies was the issues associated with high-dimensional, noisy, and incomplete data. Both Fluke and Jacobs (2019) and Djorgovski *et al.* (2023) highlighted difficulties, including overfitting, missing features, and the curse of dimensionality as impediments to optimal AI performance. Dimensionality reduction techniques, such as PCA, t-SNE, and feature selection via XGBoosts, were identified as useful approaches to mitigating these issues. Additionally, probabilistic models and uncertainty quantification techniques are being identified as new areas of research to assist astronomers in better capturing observational errors and measurement noise.

Another significant theme of discussion was the need for interpretability and transparency in AI models. While deep learning has achieved notable performance progress, it is often at the detriment of interpretability. Transitioning astronomy towards human-and-AI collaborative and co-discovery frameworks situates the understanding of the model's rationale in the same importance as the model's output. Researchers have begun exploring tools like model cards, data sheets, and explainable AI techniques to help create a communicative trail so that domain experts can validate and trust AI-led conclusions.

In conclusion, as the literature reveals, AI is not only enhancing workflows in astronomy but also reshaping the possibilities in this field. From data reduction automation to discovering previously unobservable new phenomena, AI approaches have evolved into a vital and central role in astronomy. However, the success of AI in astronomy relies on improving algorithm performance, data workflows, and ongoing collaboration between astronomers and AI experts.

5. Conclusion

Artificial Intelligence (AI) technology has already made a huge impact on the astronomy field as it relates to how researchers store, process, classify, and interpret large astronomical data sets. As demonstrated in the literature we reviewed, the usage of Machine Learning (ML) methods has passed from a phase of novelty into operational need, driven by the volume of observations from modern surveys and observatories growing exponentially. In supervised learning strategies, supervised learning approaches like support vector machines, decision trees, and neural networks excel at accurately classifying data. Additionally, deep learning methods have enabled near real-time detection and analysis of transient events in time-domain astronomy.

In our reviewed literature, there was a growing consensus that AI methods can handle scale and provide scientific insight that is impossible to achieve via manual analysis. The advent of unsupervised learning and anomaly detection with ML methods opens more possibilities of discovery, broadening our ability to identify rarer and previously unknown astronomical objects and behaviors. Moving beyond automation to discovery shows a turn toward more "involvement" of AI in the scientific process.

Nonetheless, obstacles remain in terms of data quality, model explainability, and algorithm complexity. Future progress on the above four subtitles will ultimately depend on increasing cooperation between astronomers and data scientists, but perhaps more importantly, the evolution of more sophisticated and transparent AI systems that are more flexible. AI will be important to ensure that the next (and potentially better spectroscopic) era of astronomy can provide the next generation of discoveries, providing tools that will not only help us keep pace with the data but also improve the science.

References

- Djorgovski, S.G., Mahabal, A.A., Graham, M.J., Polsterer, K. and Krone-Martins, A. (2023). [Applications of AI in Astronomy](https://doi.org/10.48550/arXiv.2212.01493). In A. Choudhary, G. Fox and T. Hey (Eds.), *Artificial Intelligence for Science (In Press)*, World Scientific. <https://doi.org/10.48550/arXiv.2212.01493>
- Fluke, C.J. and Jacobs, C. (2019). [Surveying the Reach and Maturity of Machine Learning and Artificial Intelligence in Astronomy](https://doi.org/10.1002/widm.1349). *WIREs Data Mining and Knowledge Discovery*, 10(6), e1349. <https://doi.org/10.1002/widm.1349>

Zhoolideh Haghghi, M.H. (2023). [Analyzing Astronomical Data with Machine Learning Techniques \[Conference Paper\]](#). Proceedings of the ICRANet-ISFAHAN Astronomy Meeting (to be Published in *Astronomical and Astrophysical Transactions*). <https://doi.org/10.48550/arXiv.2302.11573>

Cite this article as: Abid Hossain Rion and Mashrufa Meghla Any (2025). [The Role of AI in Analyzing Astronomical Data: A Literature Review](#). *International Journal of Data Science and Big Data Analytics*, 5(2), 111-116. doi: 10.51483/IJDSBDA.5.2.2025.111-116.