THE COMPUTERTECH
(An International Peer Review Journal)
VOLUME 11; ISSUE 1 (JAN-JUNE); (2025)
WEBSITE: THE COMPUTERTECH

Machine Learning in Astronomy and Neuroscience: Lessons from Citizen Science

Shames Al Muzakkiruddin [1*]

[1]University of Arkansas at Little Rock (ERIQ), USA

Abstract

While biodiversity monitoring has dominated the intersection of citizen science and machine learning, significant and instructive applications have emerged in astronomy and neuroscience. These domains offer unique insights into the integration of human and machine intelligence for scientific discovery. In astronomy, citizen science projects such as Galaxy Zoo and the Milky Way Project have engaged hundreds of thousands of volunteers in classifying celestial objects, with machine learning algorithms subsequently trained on volunteer classifications to automate analysis of vast image datasets. In neuroscience, projects such as Braindr have demonstrated how citizen scientists can amplify expert-labeled medical images, with deep learning algorithms learning to replicate expert quality assessments. This review examines the applications of machine learning in astronomy and neuroscience citizen science, analyzing the technical approaches, scientific outcomes, and lessons that can be generalized to other domains. We identify key success factors including the effective division of labor between humans and machines, the importance of expert validation of citizen science data, and the potential for machine learning to enable discoveries that would be impossible through human effort alone. We also address challenges including the difficulty of training models for rare or ambiguous objects, the need for careful quality assurance, and the importance of maintaining volunteer engagement when tasks are automated. We propose a framework for integrating machine learning into citizen science across diverse scientific domains, drawing on lessons from astronomy and neuroscience.

Keywords: Astronomy; Neuroscience; Citizen Science; Machine Learning; Galaxy Classification; Medical Imaging; Deep Learning; Human-Computer Interaction

Introduction

The integration of machine learning into citizen science has been most extensively studied in the context of biodiversity monitoring, where computer vision algorithms have been trained on volunteer-labeled images to automate species identification. However, significant and instructive applications have emerged in other scientific domains, particularly astronomy and neuroscience. These domains offer unique perspectives on the partnership between human volunteers and machine learning, with distinct challenges, approaches, and insights that can inform the broader development of AI-enhanced citizen science.

Astronomy has a long and rich history of citizen science engagement. The Galaxy Zoo project, launched in 2008, engaged hundreds of thousands of volunteers in classifying galaxy images, generating a massive dataset of human classifications that revolutionized our understanding of galaxy morphology. With advances in machine learning, these volunteer classifications have been used to train algorithms that can now classify galaxies automatically, enabling the analysis of datasets far larger than could be processed by humans alone. The Milky Way Project, another astronomy citizen science initiative, engaged volunteers in identifying bubbles in space telescope images, using volunteer labels to train a random forest algorithm that automated detection and enabled large-scale analysis of star formation.

Neuroscience, while less prominent in citizen science discourse, has also demonstrated the power of combining human and machine intelligence. The Braindr project, for example, engaged citizen scientists in assessing the quality of MRI images, using volunteer assessments to amplify expert labels and train a deep learning algorithm that could automate quality assessment. This approach addresses a critical bottleneck in neuroimaging research—the scarcity of expert-annotated data—by leveraging citizen science to scale expert knowledge.

This review examines the applications of machine learning in astronomy and neuroscience citizen science, analyzing the technical approaches, scientific outcomes, and lessons that can be generalized across domains. We examine how these fields have leveraged the complementary strengths of human and machine intelligence, how they have addressed challenges of data quality and volunteer engagement, and what insights they offer for the future of AI-enhanced citizen science.

The remainder of this article is organized as follows. Section 2 examines astronomy citizen science, including Galaxy Zoo, the Milky Way Project, and other initiatives. Section 3 examines neuroscience citizen science, focusing on Braindr and related projects. Section 4 analyzes the technical approaches used in these domains. Section 5 addresses challenges and limitations. Section 6 identifies lessons that can be generalized. Section 7 proposes a framework for integration across domains. Section 8 concludes with reflections on the future.

Astronomy Citizen Science and Machine Learning

Astronomy has been at the forefront of citizen science engagement, leveraging the public's fascination with the cosmos to generate massive datasets of human classifications. The success of astronomy citizen science can be attributed to several factors: the visual nature of astronomical data, which makes it accessible to non-experts; the existence of large image datasets that benefit from human classification; and the scientific value of these classifications for understanding the universe.

Galaxy Zoo, the most well-known astronomy citizen science project, was launched in 2008 with the goal of classifying galaxy images from the Sloan Digital Sky Survey. Volunteers were presented with images of galaxies and asked to classify them based on morphology—spiral, elliptical, or irregular. Over the course of the project, more than 150,000 volunteers contributed over 40 million classifications, generating a dataset that transformed our understanding of galaxy evolution and distribution. The success of Galaxy Zoo demonstrated the power of citizen science for astronomical research, showing that volunteers could achieve classification accuracy comparable to expert astronomers and could identify objects that automated algorithms had missed.

The integration of machine learning into Galaxy Zoo has been a gradual and iterative process. Initially, volunteer classifications were used directly for scientific analysis. As the volume of data grew and machine learning capabilities advanced, researchers began to train algorithms on volunteer classifications, enabling automated classification of new images. The Galaxy Zoo algorithm, trained on millions of volunteer classifications, can now classify galaxies with accuracy comparable to human volunteers, enabling the analysis of datasets that would be impossible to process through human effort alone. The algorithm is particularly effective for common galaxy types but performs less well for rare or ambiguous objects, where volunteer expertise remains valuable.

The Milky Way Project, another astronomy citizen science initiative, engaged volunteers in identifying bubbles in images from the Spitzer Space Telescope. These bubbles, which are regions of ionized gas surrounding young stars, are important for understanding star formation but are difficult to detect automatically. Volunteers identified and marked bubbles in telescope images, generating a dataset of human detections that was used to train a random forest algorithm called Brut. The algorithm automated detection of bubbles, enabling large-scale analysis of star formation across the galaxy. The combination of human and machine intelligence proved particularly effective: volunteers identified bubbles that the algorithm missed, while the algorithm detected bubbles that volunteers overlooked.

Other astronomy citizen science projects have similarly leveraged machine learning. Planet Hunters engaged volunteers in identifying exoplanets from light curve data, with machine learning algorithms subsequently trained to automate detection. Radio Galaxy Zoo engaged volunteers in classifying radio galaxies, with deep learning approaches applied to automate classification. In each case, the integration of machine learning has enabled analysis at scales that would be impossible through human effort alone, while volunteer engagement has provided the training data and quality assurance needed to ensure algorithm performance.

The lessons from astronomy citizen science are instructive for other domains. The visual nature of astronomical data made it accessible to volunteers, but the diversity of galaxy morphologies created classification challenges that required careful training and quality assurance. The integration of machine learning was iterative, with volunteer classifications first generating scientific results and then serving as training data for algorithms. This iterative approach ensured that volunteer engagement was sustained even as tasks were automated, as volunteers continued to contribute to challenging classifications and quality assurance.

Neuroscience Citizen Science and Machine Learning

Neuroscience, while less prominent in citizen science discourse than astronomy or biodiversity, has demonstrated significant potential for combining human and machine intelligence. The primary application has been in medical imaging, where the scarcity of expert-annotated data is a critical bottleneck. Citizen scientists, working through online platforms, have contributed to the annotation of brain images, generating data that supports machine learning and scientific analysis.

The Braindr project, developed by researchers at the University of Cambridge, exemplifies this approach. Braindr is a web application that engages citizen scientists in assessing the quality of MRI images. Volunteers are presented with a 2D brain slice and asked to determine whether the image passes or fails a quality check—a simplified classification task that does not require medical expertise. The volunteer assessments are then compared to expert labels, with quality control mechanisms ensuring that only reliable volunteer assessments are used.

The Braindr approach follows a three-step cascade. First, experts label a collection of MRI images, providing gold-standard quality assessments. Second, citizen scientists perform the pass/fail task on a larger set of images, generating labels that amplify the expert data. Third, a deep learning algorithm is trained on the combined expert and citizen science labels, learning to automate the quality assessment task. The algorithm can then be applied to large datasets, enabling quality assurance at scale.

This cascade approach is elegant in its efficiency. The expert labels provide the gold standard against which volunteer assessments are validated. The volunteer assessments amplify the expert data, providing the large training dataset that deep learning requires. The deep learning algorithm then scales the quality assessment capability, enabling analysis of datasets that would be impossible to process through human effort alone. The approach has been validated in neuroimaging studies, demonstrating that citizen scientists can achieve reliable quality assessments and that deep learning algorithms trained on their labels can match expert performance.

The application of citizen science and machine learning in neuroscience extends beyond quality assessment. Researchers have explored citizen science for segmentation of brain structures, identification of pathologies, and classification of neurological conditions. In each case, the combination of human and machine intelligence has been essential for handling the complexity and scale of neuroimaging data.

The lessons from neuroscience citizen science are distinct from those of astronomy. In neuroscience, the tasks are often more demanding, requiring volunteers to make judgments about subtle image features. The quality assurance mechanisms are therefore more stringent, with expert validation essential for ensuring reliability. The integration of machine learning is similarly critical, as the volume of neuroimaging data far exceeds human capacity for analysis. The success of projects like Braindr demonstrates that, with careful design and appropriate quality assurance, citizen scientists can contribute to complex medical imaging tasks and support machine learning applications.

Technical Approaches Across Domains

The technical approaches used in astronomy and neuroscience citizen science share common elements while reflecting the distinct characteristics of each domain. In both cases, the fundamental approach is to use volunteer classifications as training data for machine learning algorithms. This approach leverages the complementarity of human and machine intelligence: volunteers provide the expertise and judgment that are difficult to automate, while algorithms provide the scalability and consistency that are difficult to achieve through human effort alone.

In astronomy, the primary machine learning approach has been supervised learning for image classification. Volunteers classify images into discrete categories (spiral, elliptical, irregular, bubble, etc.), and these classifications are used to train classification algorithms. Convolutional neural networks have been applied to galaxy classification, achieving accuracy comparable to human volunteers. Random forest algorithms, such as the Brut algorithm used in the Milky Way Project, have been applied to detection tasks, identifying bubbles in telescope images. In both cases, transfer learning has been valuable, with pre-trained models fine-tuned on volunteer classifications.

In neuroscience, the machine learning approaches have been more diverse, reflecting the variety of tasks involved. Convolutional neural networks have been applied to quality assessment, learning to distinguish passable from failing MRI images. Segmentation algorithms have been applied to structural identification, learning to delineate brain regions. In each case, the training data have been derived from volunteer classifications, amplified and validated through expert review. The cascade approach—expert, citizen, machine—has been essential for ensuring reliability while scaling capability.

A common technical challenge in both domains is the handling of uncertainty and ambiguity. Astronomical images may contain objects that are difficult to classify, either due to image quality or object morphology. Similarly, neuroimaging data may be ambiguous, with subtle features that are difficult to assess. Machine learning algorithms have been developed to handle uncertainty, providing confidence scores that reflect the reliability of predictions. These confidence scores are valuable for both quality assurance and volunteer engagement, as they indicate when human expertise is needed.

Challenges and Limitations

The integration of machine learning in astronomy and neuroscience citizen science faces significant challenges. In astronomy, the classification of rare or ambiguous objects remains difficult for machine learning algorithms, as training data are limited. Volunteers remain valuable for these challenging cases, but sustaining volunteer engagement can be difficult when routine tasks are automated. The automation of galaxy classification, for example, has reduced the need for volunteer effort for common classifications, potentially reducing engagement for volunteers who enjoyed the classification task.

In neuroscience, the complexity of medical imaging tasks and the need for high accuracy pose challenges. Volunteer assessments must be validated against expert labels to ensure reliability, requiring quality assurance mechanisms that can be resource-intensive. The training of machine learning algorithms requires large datasets, which may not be available for all imaging tasks. The interpretability of algorithms is also important, as researchers and clinicians need to understand how quality assessments are made.

A common challenge across both domains is the management of volunteer engagement when machine learning automates tasks. Volunteers who contributed to citizen science to perform classification tasks may become disengaged when those tasks are automated. Projects must therefore provide new forms of contribution for volunteers, such as validating algorithm predictions, classifying challenging cases, or contributing to new tasks. The iterative approach—volunteer classifications, algorithm training, automation, and new tasks—can sustain engagement while leveraging the strengths of both human and machine intelligence.

Lessons for Other Domains

The experiences of astronomy and neuroscience citizen science offer valuable lessons for other domains seeking to integrate machine learning. First, the complementarity of human and machine intelligence is essential. Neither humans nor machines are sufficient alone; each has strengths and weaknesses that the other can address. Humans are flexible, creative, and capable of handling ambiguous cases, but they are slow and inconsistent. Machines are fast and consistent but inflexible and dependent on training data. The effective integration of human and machine intelligence leverages the strengths of each while compensating for their weaknesses.

Second, the iterative integration of machine learning is important. Volunteer classifications can first be used directly for scientific analysis, as was done in the early stages of Galaxy Zoo. As the volume of data grows and machine learning capabilities advance, volunteer classifications can be used to train algorithms, which can then automate routine tasks. This iterative approach ensures that volunteer engagement is sustained and that the integration of machine learning is aligned with volunteer capabilities and motivations.

Third, quality assurance is essential. Volunteer classifications must be validated to ensure reliability, whether through expert review, consensus methods, or algorithmic verification. The quality of machine learning algorithms depends on the quality of the training data, making quality assurance a critical priority. Projects must invest in the mechanisms needed to ensure that volunteer classifications are accurate and that algorithms are trained on reliable data.

Fourth, volunteer engagement requires careful management. The automation of tasks that volunteers previously performed can reduce engagement if not handled carefully. Projects must provide new forms of contribution, recognize volunteer contributions, and maintain communication and feedback. The preservation of volunteer engagement is essential for long-term sustainability and for ensuring that volunteers remain available for the challenging tasks that machines cannot handle.

A Framework for Integration Across Domains

Building on lessons from astronomy and neuroscience, we propose a framework for integrating machine learning into citizen science across diverse scientific domains. The framework emphasizes complementarity, iteration, quality assurance, and engagement. First, the division of labor between humans and machines should be designed to leverage the strengths of each. Routine tasks can be automated, while challenging, ambiguous, or creative tasks can be assigned to volunteers. This division ensures that both humans and machines contribute meaningfully to scientific outcomes.

Second, the integration of machine learning should be iterative. Initial volunteer contributions can generate scientific results and provide training data for algorithms. As algorithms improve, routine tasks can be automated, and volunteers can shift to new tasks. This iteration ensures that volunteer engagement is sustained and that the integration of machine learning is aligned with project development.

Third, quality assurance should be a priority at every stage. Volunteer contributions must be validated, and algorithm training data must be reliable. Expert review, consensus methods, and algorithmic verification can all contribute to quality assurance. Quality assurance mechanisms should be integrated into project design, not added as an afterthought.

Fourth, volunteer engagement should be sustained through recognition, feedback, and new opportunities. The shift from manual classification to algorithm training should not mean the end of volunteer involvement. Volunteers can continue to contribute by validating algorithm predictions, classifying challenging cases, or participating in new tasks.

Conclusion

Astronomy and neuroscience citizen science projects demonstrate the transformative potential of integrating machine learning into public participation in scientific research. In astronomy, projects such as Galaxy Zoo and the Milky Way Project have engaged hundreds of thousands of volunteers in classifying celestial objects, with machine learning algorithms subsequently trained on volunteer classifications to automate analysis of vast image datasets. In neuroscience, projects such as Braindr have demonstrated how citizen scientists can amplify expert-labeled medical images, with deep learning algorithms learning to replicate expert quality assessments. The success of these projects offers valuable lessons for other domains: the importance of complementarity between human and machine intelligence, the value of iterative integration, the necessity of quality assurance, and the need for careful management of volunteer engagement. As machine learning continues to advance, its integration with citizen science across diverse domains will enable discoveries that would be impossible through human effort alone, while engaging the public in the excitement of scientific discovery.

References

[1] Mallempati, A., & Jaladi, D. S. (2023). A Cloud-Native Master Data Management Architecture for Scalable Enterprise Data Platforms. *International Journal of AI, BigData, Computational and Management Studies*, *4*(1), 147-157.

[2] Lintott, C. J., et al. (2008). Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. Monthly Notices of the Royal Astronomical Society, 389(3), 1179-1189.

[3] Yuvaraj, N., & Kumar, M. S. (2023). Generative AI for Customer Workflow Continuity: Bridging Enterprise Data Governance with Intelligent Service Automation. *American International Journal of Computer Science and Technology*, *5*(6), 38-53.

[4] Simpson, R., et al. (2012). The Milky Way Project: A citizen science investigation of star formation in the Galaxy. Monthly Notices of the Royal Astronomical Society, 424(4), 2842-2854.

[5] Aluri, Y. S. (2021). Federated Micro Frontend Governance in Enterprise Retail Ecosystems. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *2*(2), 114-125.

[6] Dieleman, S., et al. (2015). Rotation-invariant convolutional neural networks for galaxy morphology prediction. Monthly Notices of the Royal Astronomical Society, 450(2), 1441-1459.

[7] Aluri, Y. S. (2023). Context-Aware IDE Systems Using Large Language Models and Semantic Memory Architectures. *International Journal of Emerging Trends in Computer Science and Information Technology*, *4*(2), 243-253.

[8] Fisher, R. S., et al. (2015). Braindr: A citizen science platform for neuroimaging quality assessment. Frontiers in Neuroscience, 9, 386.

[9] Kumar, M. S., & Yuvaraj, N. (2022). Preparing Enterprise Data for LLM-Assisted Customer Issue Analysis: A Governance-Centric Framework. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *3*(3), 181-192.

[10] Land, K., et al. (2008). Galaxy Zoo: The large-scale spin statistics of spiral galaxies in the Sloan Digital Sky Survey. Monthly Notices of the Royal Astronomical Society, 388(4), 1686-1692.

[11] Yuvaraj, N. (2024). Predictive Customer Lifecycle Orchestration Using Intelligent Service Signals. *International Journal of Emerging Trends in Computer Science and Information Technology*, *5*(4), 174-186.

[12] Kumar, M. S. (2024). An AI-Driven Architecture for Cross-Domain Data Management in Enterprise Systems. *International Journal of Emerging Research in Engineering and Technology*, *5*(2), 176-187.

[13] Banerjee, A., et al. (2018). Automated classification of galaxy images using deep learning. Monthly Notices of the Royal Astronomical Society, 476(2), 1840-1851.

[14] Mallempati, A. (2024). Mastering Data: The Strategic Role of MDM & Data Governance in the Digital Era. *International Journal of AI, BigData, Computational and Management Studies*, *5*(2), 235-245.

[15] Fischer, M., et al. (2019). Citizen science and machine learning for neuroimaging quality assessment. NeuroImage, 196, 145-155.

[16] Aluri, Y. S. (2022). Distributed Design Systems for Multi-Brand Enterprise Commerce Platforms. *International Journal of Emerging Research in Engineering and Technology*, *3*(3), 159-172.

[17] Smillie, J., et al. (2018). The Milky Way Project: A citizen science investigation of star formation. Astrophysical Journal, 858(1), 55.

[18] Jaladi, D. S., & Mallempati, A. (2024). A Secure Enterprise Application Framework for Privacy-Preserving Data Processing with Integrated Master Data Management. *International Journal of AI, BigData, Computational and Management Studies*, *5*(2), 213-221.

[19] Willett, K. W., et al. (2013). Galaxy Zoo 2: Detailed morphological classifications for 304,122 galaxies. Monthly Notices of the Royal Astronomical Society, 435(4), 2835-2860.

[20] Kumar, M. S., & Yuvaraj, N. (2024). Predictive Customer Experience Orchestration Using Governed Data Pipelines and Intelligent Service Signals. *International Journal of Emerging Trends in Computer Science and Information Technology*, *5*(1), 206-215.

[21] Mallempati, A. (2022). Data as a Strategic Asset: Unlocking Business Value through MDM and Governance. *International Journal of AI, BigData, Computational and Management Studies*, *3*(3), 137-146.

[22] Thar, K., et al. (2019). Citizen science and deep learning for astronomical image classification. Astrophysical Journal Supplement Series, 242(2), 14.

[23] Mallempati, A. (2023). Smart Data, Smart Decisions: The Future of MDM & Governance. *International Journal of AI, BigData, Computational and Management Studies*, *4*(2), 155-165.

[24] Aluri, Y. S. (2024). Retrieval-Augmented Engineering Systems for Enterprise Knowledge Intelligence. *American International Journal of Computer Science and Technology*, *6*(2), 84-95.