A Convolutional Neural Network (CNN) Approach to Galaxy Merger Classification

Sohan Subudhi,[a] Alex Xu,[b] and Darshan Desai[c]

The University of Texas at Austin, Austin, Texas 78712, USA

[a]Corresponding author: sohan.subudhi@utexas.edu
[b]alex.xu@utexas.edu
[c]dashkad2004@utexas.edu

Abstract. Galaxy mergers provide critical insights into cosmic evolution, yet identification in survey data remains challenging. We introduce a rigorously verified dataset of 228 hand-classified mergers and 292 nonmergers, curated from Galaxy Zoo DR2 and DR5, VV objects, and Toomre merger candidates. Using this data, we trained a convolutional neural network with four convolutional blocks, utilizing L2 regularization, dropout, and augmentations like random rotations and flips. Our model achieved 74% test accuracy, demonstrating the dataset's utility. The findings indicate that improvements through synthetic data or domain adaptation would enhance machine learning performance. We present this dataset as a resource to advance automated merger classification.

INTRODUCTION

Past Work

Classifying galaxy mergers would shed light on the theorized hierarchical formation of the universe by enabling the reconstruction of galaxy assembly histories over cosmic time. Furthermore, identifying and categorizing merger events would help constrain the distribution and evolution of dark matter halos, improve estimates of merger rates that produce gravitational wave sources, and clarify the role of mergers in triggering active galactic nuclei activity [1].

However, large-scale manual classification of mergers proves difficult, given the time scale of galaxy merger events, the poor visibility of certain features of galaxies such as tails and bars, and uncertainty as to whether or not a merger has occurred. Modern hydrodynamic simulations effectively model clusters and galaxies en masse [2], spurring the development of numerous machine learning (ML) models for galaxy merger classification. Observationally based approaches in particular have applied convolutional neural networks (CNNs) to real survey imagery, using transfer learning to detect mergers [3] and to assemble large observational merger catalogs [4].

Past work by Pearson et al. [5] and Margalef-Bentabol et al. [6] provides context and support for our research. The first study addresses the challenges of temporal classification in mergers, using the IllustrisTNG simulation to train ML models capable of predicting whether a galaxy is in a premerger, ongoing, or postmerger stage. The second study evaluates the performance of different machine learning techniques for merger detection, emphasizing the need for models that generalize effectively across diverse datasets. It is important to note that our study differs by addressing a broad range of limitations identified in both of the above works. While the first study highlights the power of simulations to improve temporal accuracy, it underscores the need for more examples and better integration of intermediate stages, which our work looks to address. Similarly, the second study reveals significant drops in classification performance when models are applied to data from different domains (sources of merger data), highlighting the limitations of current methods in transferring knowledge across datasets.

Research Question

There is a gap between simulation and real-data performance among classifiers today [6] that can be addressed by gathering more real merger data to build stronger models. With new observations from the James Webb Space Telescope (JWST) and the Simonyi Survey Telescope capturing high-redshift images of potential galaxy mergers, we aim to support classification by constructing a verifiable dataset and subsequently developing an effective classifier trained upon it.

From a data collection standpoint, we define a galaxy merger as the result of multiple galaxies passing through each other, disturbing the structure of the involved galaxies and resulting in long-lasting changes to the shape and appearance of a galaxy, such as observable tidal effects and distortion in the shape of a galaxy. A nonmerger for this study is defined as a galaxy without obvious perturbation. It is important to note the lack of any tidal tails, bridges, or other types of visual distortion that are typically seen within a merger. The nonmergers typically exhibit a well-defined structure, such as the arms of a spiral galaxy or smooth elliptical morphology. We also chose to exclude any galaxies that appear irregular to avoid a potential confounding variable.

DATA SOURCES

The data distribution and our sources are depicted in Table 1. These sources were selected based on their demonstrated reliability in previous studies and their established prominence in the field [7]. Our sources comprise Galaxy Zoo DR5 [8], Galaxy Zoo Merger Project [9], VV objects [10], and Toomre [11] merger candidates. A merger candidate is defined as galaxies marked as interacting based on their respective catalogs. We randomly sampled these from each catalog to do manual classification.

TABLE 1. Merger image data sources.

Merger Galaxy Distribution			**Nonmerger Galaxy Distribution**		
Catalog (Survey)	**Candidates**	**Validated**	**Catalog (Survey)**	**Morphology**	**Number**
Galaxy Zoo Merger Project (SDSS)	54	54	Galaxy Zoo DR2 (SDSS)	Spirals	50
				Ellipticals	50
Galaxy Zoo DR5 (SDSS)	248	93	Galaxy Zoo DR5 (SDSS)	Round	46
				In-Between	75
VV-Objects (SDSS/NED)	219	86		Cigar-shaped	12
Toomre Mergers (PAN-STAARS)	9	9		Edge-on	8
				Spirals	51
Total mergers	530	242	**Total nonmergers**		**292**
After deduplication	**228**				

We classified these candidates by visual inspection in a double-blind setting, including images in our merger set if the majority of three authors determined it was a merger. When it appeared that a smaller galaxy was either far away from or cannibalized by a larger galaxy, we evaluated whether there were enough structural disturbances and distortions to indicate a merger interaction. Unclear cases where a majority of authors did not believe the candidate was a merger were rejected from both the merger and nonmerger sets to prevent potential misclassifications. Note that the nonmergers were obtained separately from Galaxy Zoo, where the galaxy morphology labels were more reliably crowd-sourced.

Since our final dataset concatenated objects from multiple catalogs, we performed a deduplication step to catch overlapping images both within and across surveys. Using the catalog centroids, we implemented a coordinate filter ($|\Delta$right ascension (RA)$| + |\Delta$declination (DEC)$| < 0.1°$). Although this overestimates true angular distance at higher declinations, it successfully isolated a subset of potential duplicates that we then manually verified by eye. The redshift data obtained through Galaxy Zoo indicated mergers were all low-z and less than 0.15. Redshift was excluded from this matching criteria since it was not universally available across our catalogs.

This process does not exclude a merger until manually verified by comparing duplicates and, in the worst case, leads to mergers that were not duplicated being excluded. After this process, the number of mergers in our dataset decreased from 242 to 228. In the case of duplicates, we kept the more centered copy and prioritized images from less-represented surveys to improve image-source balance. Images from the VV and Toomre catalogs were not centered on galaxies due to outdated RA and DEC values.

ANALYSIS

For our model, we utilized a CNN [12] with four convolutional blocks of 32, 64, 96, and 128 filters with a 3×3 kernel, followed by global average pooling and a 256-unit dense classifier. The source images were 512×512; to select the input resolution, we trained and evaluated three otherwise-identical models at 128×128, 256×256, and 512×512. The 256×256 model was chosen since it gave the best test accuracy, balancing spatial detail with computation complexity. Input sizes were kept to powers of two to down-sample cleanly through the four pooling stages. The remaining choices (block depth, filter progression, and regularization) follow standard design principles for a dataset of this size. We applied L2 regularization with $\lambda = 0.01$, batch normalization, and dropout rates between

30% and 50%. We used Adam as our optimizer with a learning rate of $\eta = 0.0001$ and binary cross-entropy as our loss function, which are conventional parameter settings. A fuller hyperparameter search is left for future work. We trained a model on the set of 228 mergers and 292 nonmergers detailed above. In training the model, we used a split of 0.65 train and 0.20 validation and 0.15 test. Additionally, the following data augmentations were applied while training: normalizing pixel values to the $[0,1]$ range, random rotations up to 20°, horizontal and vertical shifts of $\pm 15\%$ of the image width, and horizontal and vertical image flipping.

RESULTS

TABLE 2: Model performance metrics.

Class	Precision	Recall	F1 Score
Merger	0.65	0.92	0.76
Nonmerger	0.90	0.60	0.72
Balanced accuracy: 0.7583		*AUC: 0.8852*	

TABLE 3: Test-set confusion matrix.

		Predicted	
		Merger	Nonmerger
True	Merger	33	3
	Nonmerger	18	27

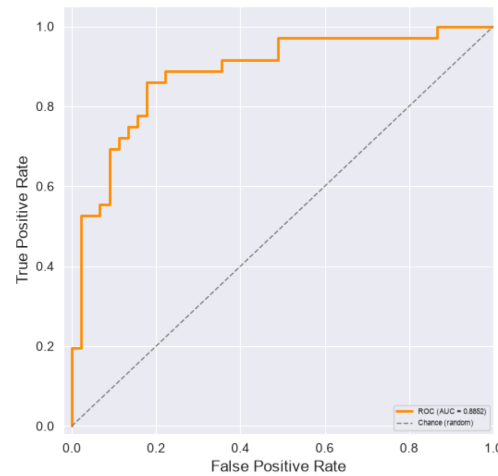

FIGURE 1: Test-set ROC (positive = merger).

The model achieved a test accuracy of $\sim$74%, a weighted recall of 74%, and a weighted precision of 79% (Table 2). For comparison, a baseline classifier that always predicts the majority class (in this case, the nonmerger class) has an accuracy of $\frac{45}{36+45} \times 100 = 55.6\%$ on the test set, so the model represents a substantial improvement over chance. This separability is reflected in the receiver operating characteristic curve (Fig. 1), which yields an area under the curve of 0.8852. As shown in the confusion matrix (Table 3), the model recovers 92% of true mergers but misclassifies 18 of the 45 nonmergers as mergers, indicating a mild bias toward predicting the *merger* class.

As per the misclassifications of the model (Fig. 2), images commonly thought to be of a nonmerger appear to be primarily black (empty) with a very small subject at the center of the image. The image composition is similar to that of higher redshift images belonging to nonmergers. Such images could potentially be better handled through a variable scale at which images are extracted depending on redshift. Images that were misclassified as mergers were typically spiral galaxies that have hazy and more separated spiral arms or somewhat elongated galaxies, indicating that those might be features the CNN responds strongly to; sharpening the model's ability to separate genuine interaction from lookalike structures is a promising direction for reducing them. Correct classifications associate mergers with clear interaction signatures such as tidal bridges, tails, or closely paired nuclei linking the merging components. Nonmergers are associated with isolated single galaxies, often compact and set against a sparse field, with no companion or interaction signature, indicating the model keys on the presence or absence of the morphological cues it associates with interaction.

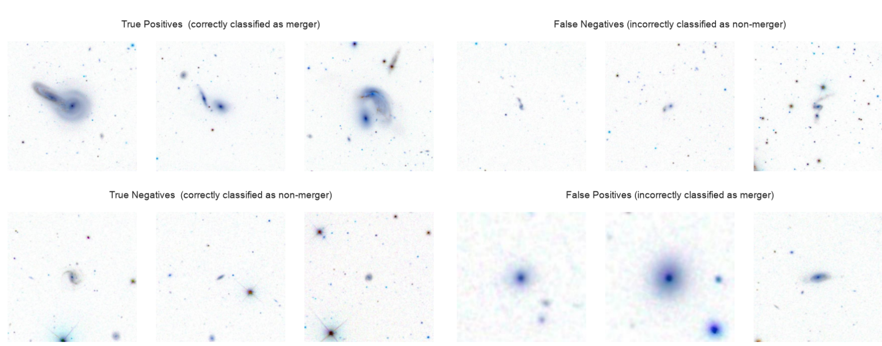

FIGURE 2: Select test set classifications. The inverse images are from SDSS and use SDSS-g, SDSS-r, and SDSS-i bandpasses.

CONCLUSION

The accuracy achieved by Margalef-Bentabol et al. [6] when using a CNN trained on Hyper Suprime-Cam survey images was lower than our final results, however, we believe that the low quantity of data may have impacted this. We support further testing with additional data; data gathered from different sources may also lead to different results. Our study advances galaxy merger classification by producing data for the implementation of domain adaptation to mitigate these biases, ensuring machine learning models are not constrained by patterns specific to any single survey or simulation. Within our experiment, we examined recall separately by source. Across the full dataset the model recovered mergers at 100% recall for VV objects, 96.3% for the Galaxy Merger Project sample, 93.5% for Galaxy Zoo DR5, and 88.9% for the Toomre candidates, with the mean predicted merger probability following the same ordering. The variation is modest, suggesting that catalog-level selection and labeling, rather than the imaging itself, drives the spread. We note, however, that our imaging is predominantly SDSS, so this dataset cannot by itself characterize the larger domain shift expected for surveys of markedly different depth and resolution, such as the Hyper Suprime-Cam imaging used by [6].

Our model demonstrates that mitigating class bias and taking certain data preprocessing steps, such as variable scaling based on red shift, are important factors to consider in a merger classifier. Future studies looking to build upon these results will have a well-defined list of mergers to draw from to assist in advancing this field beyond our work. In the future, we would like to augment our results by incorporating (manually verified) synthetic data obtained from IllustrisTNG. We also propose the use of domain adaptation [3] when training a better model on synthetic data. We would also like to customize the scale at which images are captured using redshift or another parameter, to standardize the size of the galaxy in the image. An interesting extension would be to predict the stage of merger evolution on a continuous scale from 0 (relaxed) to 1 (disturbed and strongly interacting). This may also allow for the classification of mergers as minor or major interactions based on interaction strength. We would also like to investigate distinguishing between wet and dry mergers through the incorporation of multi-band photometry, which may provide additional information regarding star formation, gas content, dust, and stellar mass.

ACKNOWLEDGMENTS

The authors would like to express sincere gratitude to Dr. Shyamal Mitra, Dr. Karl Gebhardt, the peer mentors, and other members of the Geometry of Space research stream for their invaluable guidance and mentorship throughout this research project.

REFERENCES

1. J. E. Barnes and L. Hernquist, Annu. Rev. Astron. Astrophys. **30**, 705–742 (1992).
2. D. Nelson, V. Springel, A. Pillepich, V. Rodriguez-Gomez, P. Torrey, S. Genel, M. Vogelsberger, R. Pakmor, F. Marinacci, R. Weinberger, et al., Comput. Astrophys. Cosmol. **6**, 2 (2019).
3. S. Ackermann, K. Schawinski, C. Zhang, A. K. Weigel, and M. D. Turp, Mon. Not. R. Astron. Soc. **479**, 415–425 (2018).
4. W. J. Pearson, L. E. Suelves, S. C.-C. Ho, N. Oi, S. Brough, B. W. Holwerda, A. M. Hopkins, T.-C. Huang, H. S. Hwang, L. S. Kelvin, et al., Astron. Astrophys. **661**, A52 (2022).
5. W. J. Pearson, V. Rodriguez-Gomez, S. Kruk, and B. Margalef-Bentabol, Astron. Astrophys. **687**, A45 (2024).
6. B. Margalef-Bentabol, L. Wang, A. La Marca, C. Blanco-Prieto, D. Chudy, H. Domínguez-Sánchez, A. D. Goulding, A. Guzmán-Ortega, M. Huertas-Company, G. Martin, et al., Astron. Astrophys. **687**, A24 (2024).
7. M. Walmsley, A. M. M. Scaife, C. Lintott, M. Lochner, V. Etsebeth, T. Géron, H. Dickinson, L. Fortson, S. Kruk, K. L. Masters, et al., Mon. Not. R. Astron. Soc. **513**, 1581–1599 (2022).
8. M. Walmsley, C. Lintott, T. Géron, S. Kruk, C. Krawczyk, K. W. Willett, S. Bamford, L. S. Kelvin, L. Fortson, Y. Gal, et al., Mon. Not. R. Astron. Soc. **509**, 3966–3988 (2022).
9. A. J. Holincheck, J. F. Wallin, K. Borne, L. Fortson, C. Lintott, A. M. Smith, S. Bamford, W. C.. Keel, and M. Parrish, Mon. Not. R. Astron. Soc. **459**, 720–745 (2016).
10. B. A. Vorontsov-Velyaminov, *Catalogue of Interacting Galaxies*. Available at http://www.sai.msu.su/sn/vv/.
11. A. Toomre, "Mergers and Some Consequences," in *The Evolution of Galaxies and Stellar Populations*, edited by B. M. Tinsley and R. B. Larson (Yale University Observatory, New Haven, CT, 1977), p. 401.
12. Y. LeCun, in *Connectionism in Perspective*, edited by R. Pfeifer, Z. Schreter, F. Fogelman, and L. Steels (Elsevier, Zurich, 1989), pp. 143–150.