A Novel Sector-Based Algorithm for an Optimized Star-Galaxy Classification

Anuamnchi Agastya Sai Ram Likhit, Divyansh Tripathi, Akshay Agarwal
Department of Physics, Department of Data Science and Engineering
Indian Institute of Science Education and Research, Bhopal
{anumanchi20, divyansh20, akagarwal}@iiserb.ac.in

Abstract

This paper introduces a novel sector-based methodology for star-galaxy classification, leveraging the latest Sloan Digital Sky Survey data (SDSS-DR18). By strategically segmenting the sky into sectors aligned with SDSS observational patterns and employing a dedicated convolutional neural network (CNN), we achieve state-of-the-art performance for star galaxy classification. Our preliminary results demonstrate a promising pathway for efficient and precise astronomical analysis, especially in real-time observational settings.

1 Introduction

Today is the age of Data-Driven astronomy, with sky surveys generating large amounts of data, and many new ones are lining up, such as the large synoptic survey telescope (LSST). One of the key motives of such surveys is to classify objects as stars or galaxies. However, manual classification can not be done for petabytes of data and large intra-class variation, which raises the need for an automated and robust classification model. Recently, several research works have been developed to help astronomers by automatically classifying the galaxies (Soumagnac et al., 2015; Ba Alawi & Al-Roainy, 2021; Chaini et al., 2022; Kim & Brunner, 2016; Garg et al., 2022). However, these models perform well but are complex. In contrast to the existing work, due to the complexity of our star-galaxy system, in this research, we have proposed the development of a classification approach utilizing a sector-based division of the sky. The prime motivation for such division can be seen in Figure 1, reflecting the variation present in different sectors and difficulties in classification. By utilizing these differences, we have developed a star-galaxy classification system that surpasses existing algorithms and yields a low computational cost.

2 Proposed Methodology

To address the star-galaxy classification challenge, we introduce a sector-based approach closely aligned with the Sloan Digital Sky Survey (SDSS)Almeida et al. (2023) observation patterns. For that sky is divided into thirty-six distinct sectors (Appendix A & Figure 2) by segmenting Right Ascension (RA) and Declination (Dec) intervals. *Right Ascension (RA) is akin to longitudinal lines on Earth and ranges from 0 hours to* 24 *hours, equivalent to* 0° *to* 360° *in celestial coordinates*. RA is divided into six equal intervals of 60° each, corresponding to 4-hour segments. Declination (Dec) which spans from the North celestial pole at $+90^o$ to the South celestial pole at -90° is segmented into six intervals of 30° each. Combining the divisions in RA and Dec, we obtain a total of 36 sectors which are defined by their specific RA and Dec range[1].

---

[1]The other necessary details are also provided in the Appendix.

Stars   Galaxies   Stars   Galaxies
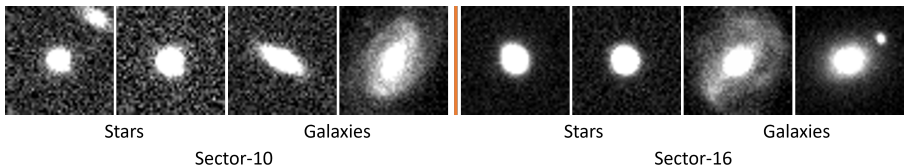Sector-10   Sector-16

Figure 1: A sample image reflecting the challenges in identifying star-galaxies in different sectors.

Table 1: Star-Galaxy Classification performance of the proposed and existing Algorithms in terms of accuracy (Acc), precision (P), recall (R), and F-1 score (F1).

Algorithm	Sector 10				Sector 16				Combined			
	Acc	P	R	F1	Acc	P	R	F1	Acc	P	R	F1
Proposed	**0.96**	**0.97**	**0.97**	**0.97**	**0.95**	**0.95**	**0.95**	**0.95**	**0.95**	**0.95**	**0.95**	**0.95**
CovNet	0.91	0.91	0.91	0.91	0.94	0.94	0.94	0.94	0.88	0.89	0.89	0.89
MargNet	0.94	0.94	0.94	0.94	0.93	0.94	0.94	0.94	0.92	0.93	0.92	0.92

Table 2: Confusion matrix reflecting the effectiveness in handling individual sectors and classes.

Algorithm	True ↓	Sector-10		Sector-16		Combined	
	Predicted →	Star	Galaxy	Star	Galaxy	Star	Galaxy
Proposed	Star	**935**	53	**963**	25	**1858**	123
	Galaxy	44	**968**	44	**968**	67	**1952**
CovNet	Star	**894**	89	**918**	70	**1742**	238
	Galaxy	85	**927**	42	**970**	217	**1802**
MargNet	Star	**952**	36	**907**	81	**1933**	48
	Galaxy	76	**936**	44	**968**	268	**1751**

Once the divided sector images are obtained, we have provided them as input to the proposed custom convolutional neural network model (Appendix B). As mentioned above, we asserted that the division of the sky into sectors can better help in star-galaxy classification even with the use of a simple model; therefore, a shallow model of 3 convolutional layers have been developed. The aim is not only to achieve higher accuracy but at a lower computational cost. Each convolutional layer is followed by max-pooling and ReLU activation functions. Two dense layers each containing 64 neurons are attached at the end to extract the features. The dropout is also added to reduce the impact of overfitting. The configuration of the proposed CNN is given in Table 3 (Appendix B).

3 EXPERIMENTS AND RESULTS

The primary focus of the evaluation is on sector-10 and sector-16 [2]. Each sector contributed $10,000$ images and the proposed classification model is trained separately on each sector to determine their effectiveness in handling individual sectors. Moreover, to evaluate the scalability of the proposed algorithm is also evaluated on the large-scale dataset achieved after combining the images of both sectors. Further, a comparison with existing SOTA algorithms: CovNet Kim & Brunner (2016) and MargNet Chaini et al. (2022) has also been performed to demonstrate the efficacy of the proposed approach. For a fair comparison, the existing models are trained on the same training-testing setting on which the proposed algorithm is trained.

The comparative results reported in Table 1 demonstrate that the proposed algorithm surpasses each existing algorithm by a significant margin. Further, the effectiveness of the proposed algorithm is not only for a single sector but also for each and in combination. For example, the proposed algorithm on a large-scale dataset comprising sector-10 and sector-16 yields an accuracy of $\mathbf{95.25\%}$ in comparison to 88.62% and 92.10% of CovNet and MargNet, respectively. Further, Table 2 shows the confusion matrix of the proposed algorithm. It shows that the proposed algorithm is not biased to any particular class and can effectively identify stars and galaxies with higher accuracy.

As shown in Table 4 of the Appendix B, the proposed algorithm took $\mathbf{25}s$/epoch on the data of the combined sectors as compared to $180sec$ and $1610sec$ taken by CovNet and MargNet, respectively.

4 CONCLUSION

We have proposed a novel and cost-effective algorithm for star-galaxy classification by handling sector-specific data. The efficacy of the proposed algorithm surpasses the existing algorithm back our idea of segregating the sky into sectors for better performance. In the future, we aim to develop an advanced architecture to tackle other sectors and improve the classification performance of the proposed approach by incorporating sector-specific auxiliary information. We believe the proposed research can advance the astronomical research by precisely identifying the celestial objects.

---

[2]More details on the selection of these sectors are provided in the appendix D

URM STATEMENT

The authors acknowledge that the key author of this work meets the URM criteria of the ICLR 2024 Tiny Papers Track.

REFERENCES

Andrés Almeida, Scott F. Anderson, Maria Argudo-Fernández, Carles Badenes, Kat Barger, et al. The eighteenth data release of the sloan digital sky surveys: Targeting and first spectra from sdss-v. *The Astrophysical Journal Supplement Series*, TBD(TBD):TBD, 2023. doi: 10.3847/1538-4365/acda98. URL https://doi.org/10.3847/1538-4365/acda98. arXiv:2301.07688 [astro-ph.GA].

Abdulfattah Ba Alawi and Ali Al-Roainy. Deep residual networks model for star-galaxy classification. pp. 1–4, 07 2021. doi: 10.1109/ICOTEN52080.2021.9493433.

Siddharth Chaini, Atharva Bagul, Anish Deshpande, Rishi Gondkar, Kaushal Sharma, M Vivek, and Ajit Kembhavi. Photometric identification of compact galaxies, stars, and quasars using multiple neural networks. *Monthly Notices of the Royal Astronomical Society*, 518(2):3123–3136, nov 2022. doi: 10.1093/mnras/stac3336. URL https://doi.org/10.1093%2Fmnras%2Fstac3336.

Prapti Garg, Tarini Chandra, Ritika Ahlawat, Neha Mittal, Ratneshwar Kumar Ratnesh, and Subodh Kumar Tripathi. Star galaxy image classification via convolutional neural networks. In *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)*, pp. 1156–1161, 2022. doi: 10.1109/ICOSEC54921.2022.9952065.

Edward J. Kim and Robert J. Brunner. Star–galaxy classification using deep convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 464(4):4463–4475, October 2016. ISSN 1365-2966. doi: 10.1093/mnras/stw2672. URL http://dx.doi.org/10.1093/mnras/stw2672.

M. T. Soumagnac, F. B. Abdalla, O. Lahav, D. Kirk, I. Sevilla, E. Bertin, B. T. P. Rowe, J. Annis, M. T. Busha, L. N. Da Costa, J. A. Frieman, E. Gaztanaga, M. Jarvis, H. Lin, W. J. Percival, B. X. Santiago, C. G. Sabiu, R. H. Wechsler, L. Wolz, and B. Yanny. Star/galaxy separation at faint magnitudes: application to a simulated Dark Energy Survey. *Monthly Notices of the Royal Astronomical Society*, 450(1):666–680, 04 2015. ISSN 0035-8711. doi: 10.1093/mnras/stu1410. URL https://doi.org/10.1093/mnras/stu1410.

The Sloan Digital Sky Survey. Sdss dr7 coverage. https://classic.sdss.org/dr7/coverage/404.php, 2008.

A SKY DIVISION

We want to highlight that this division aligns with the inherent structure of the SDSS, which scans the sky in stripes spaced 2.5^o apart in survey latitude. Using a 30° sector in declination, each sector encompasses 12 SDSS stripes. This alignment ensures data homogeneity within each sector, containing a complete and consistent set of SDSS stripes.

The 60° x 30° sector division strikes a balance between granularity and expansive sky coverage. It provides a detailed sky view while maintaining enough breadth for comprehensive sector-specific analysis. This division ensures that the data for each sector is substantial yet not overly dense, allowing computational algorithms to be applied effectively on a per-sector basis without requiring excessive resources.

Moreover, this sector division is not only tailored for the SDSS data structure but also remains universally applicable. The equatorial coordinate-based division offers a flexible foundation to integrate additional survey data in the future. Additionally, the polar regions, often challenging in sky segmentation due to projection distortion, are effectively managed in this division. The sectors that extend from +60° to +90° and -60° to -90° handle the polar regions without significant distortion.

Circular Representation of 36 Sectors

(a) Circular Representation

2D Representation of 36 Sectors
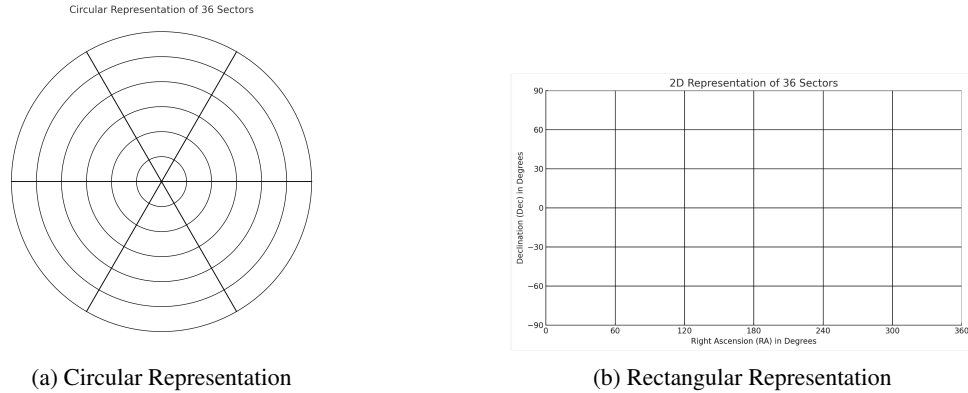
(b) Rectangular Representation

Figure 2: 2D Sky Sector Map

We utilized imaging data from SDSS-DR18, for which we crafted a tailored SQL query to extract metadata aligned with our sector-based methodology. An automated Python script processed this metadata and constructed URLs to download compressed FITS files. After downloading, we centered the celestial objects within 45x45 pixel frames based on their Right Ascension (RA) and Declination (Dec) and then converted the images into PNG format. To prepare the data for CNN analysis, we stacked images from all filters to create a five-channel .npy file and normalized pixel values for uniformity. We applied data augmentation techniques to increase the dataset's diversity and enhance the model's robustness.
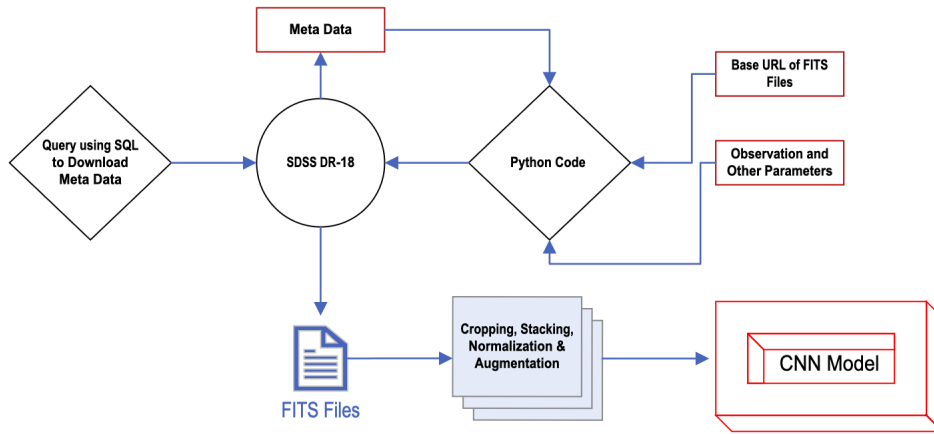

Figure 3: Data Workflow Diagram

B PROPOSED CNN ARCHITECTURE

Layer 1	Conv2D(32, 3×3 ×5), ReLU , MaxPool
Layer 2	Conv2D(64, 3×3 ×5), ReLU , MaxPool
Layer 3	Conv2D(128, 3×3 ×5), ReLU , MaxPool
Layer 4	Flatten
Layer 5	Dense (64) , Dropout(0.5)
Layer 6	Dense (1)

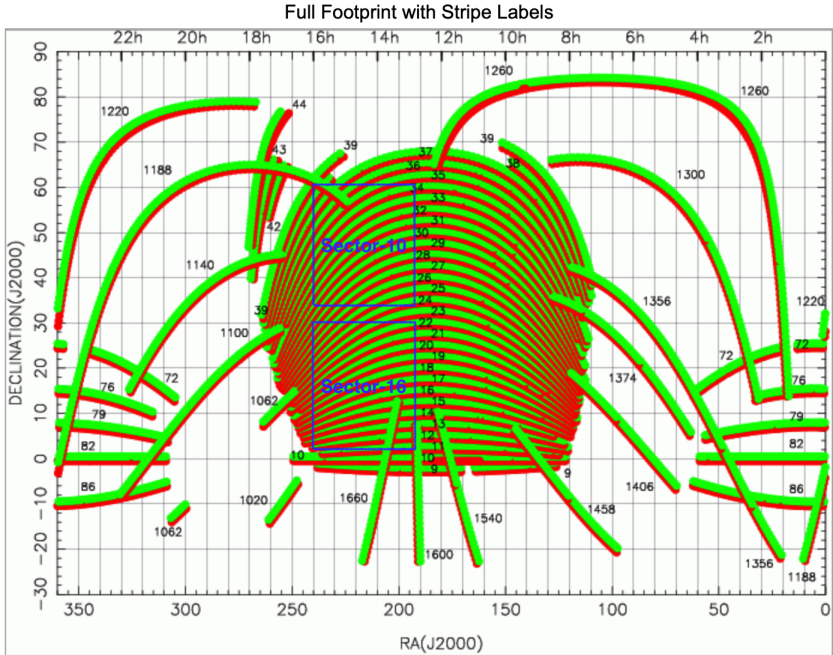Table 3: Configuration of the proposed CNN for star-galaxy classification.

Figure 4: SDSS Sky Coverage (The Sloan Digital Sky Survey, 2008)

All the networks are trained with a batch size of 32 and an "Adam" optimizer with a default learning rate of 0.001. The loss function used for the work is binary cross-entropy keeping in mind the binary nature of the problem.

Computational Cost Comparison			
Models	Sector-10	Sector-16	Combined
Proposed	**15s/epoch**	**13s/epoch**	**25s/epoch**
CovNet	80s/epoch	80s/epo	180s/epoch
MargNet	1000s/epoch	570s/epoch	1610s/epoch

Table 4: Comparison of Running Time per Epoch of Proposed and Existing Models

C   DATA SPLITTING FOR MODEL TRAINING AND TESTING

Our dataset consists of 20,000 augmented images, equally distributed across Sector-10 and Sector-16, with each sector containing 10,000 images (5,000 stars and 5,000 galaxies). We applied a train-test split of 0.2.

- **Individual Sector Analysis:** In each sector, 8,000 images are used for training, and 2,000 images are used for testing.
- **Combined Sector Analysis:** When sectors are combined, the total dataset comprises 20,000 images. Here, 16,000 images are used for training, and 4,000 images are used for testing.

D   CHOOSING SECTORS: AN OVERVIEW

We strategically chose sectors 10 and 16 for our analysis due to their high star and galaxy counts as seen in below Table 5. These sectors are not only dense but also centrally located within the SDSS sky coverage, as seen in Figure 4, making them ideal for developing predictive models. By starting with these sectors, we aim to establish a strong framework that can be implemented in other sectors as well.

Table 5: Available SDSS Data Across Sectors

Sector	RA_Range	Dec_Range	No_of_STARS	No_of_GALAXIES
1	0° to 60°	60° to 90°	273	159
2	60° to 120°	60° to 90°	5454	262
3	120° to 180°	60° to 90°	13133	31054
4	180° to 240°	60° to 90°	8239	33973
5	240° to 300°	60° to 90°	6628	2980
6	300° to 360°	60° to 90°	4771	381
7	0° to 60°	30° to 60°	21537	31868
8	60° to 120°	30° to 60°	26896	45414
9	120° to 180°	30° to 60°	125641	418122
10	180° to 240°	30° to 60°	113805	37085
11	240° to 300°	30° to 60°	49199	102023
12	300° to 360°	30° to 60°	20660	20119
13	0° to 60°	0° to 30°	102823	306967
14	60° to 120°	0° to 30°	41578	30375
15	120° to 180°	0° to 30°	128893	498274
16	180° to 240°	0° to 30°	104402	11901
17	240° to 300°	0° to 30°	39779	77124
18	300° to 360°	0° to 30°	94264	247358
19	0° to 60°	-30° to 0°	58354	185367
20	60° to 120°	-30° to 0°	7828	1920
21	120° to 180°	-30° to 0°	18351	28538
22	180° to 240°	-30° to 0°	16538	41186
23	240° to 300°	-30° to 0°	896	1144
24	300° to 360°	-30° to 0°	31461	79353
25	0° to 60°	-60° to -30°	0	0
26	60° to 120°	-60° to -30°	0	0
27	120° to 180°	-60° to -30°	0	0
28	180° to 240°	-60° to -30°	0	0
29	240° to 300°	-60° to -30°	0	0
30	300° to 360°	-60° to -30°	0	0
31	0° to 60°	-90° to -60°	0	0
32	60° to 120°	-90° to -60°	0	0
33	120° to 180°	-90° to -60°	0	0
34	180° to 240°	-90° to -60°	0	0
35	240° to 300°	-90° to -60°	0	0
36	300° to 360°	-90° to -60°	0	0

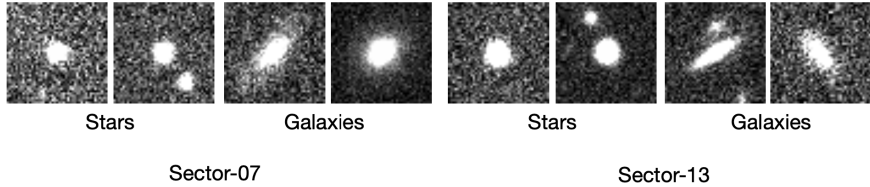

Figure 5: A sample image reflecting the challenges in identifying star-galaxies in sectors 7 and 13.

E OTHER SECTORS AND ZERO-SHOT RESILIENCY

We have also performed experiments in the following manners to evaluate the effectiveness and resiliency of the proposed approach.

Evaluation of Sectors 7 and 13: In this setting, we have downloaded the stars and galaxies images of two new sectors (7 and 13). Similar to setting on sectors 10 and 16, the images of these sectors are divided into training: used to train the model, and testing: used for evaluation. The proposed
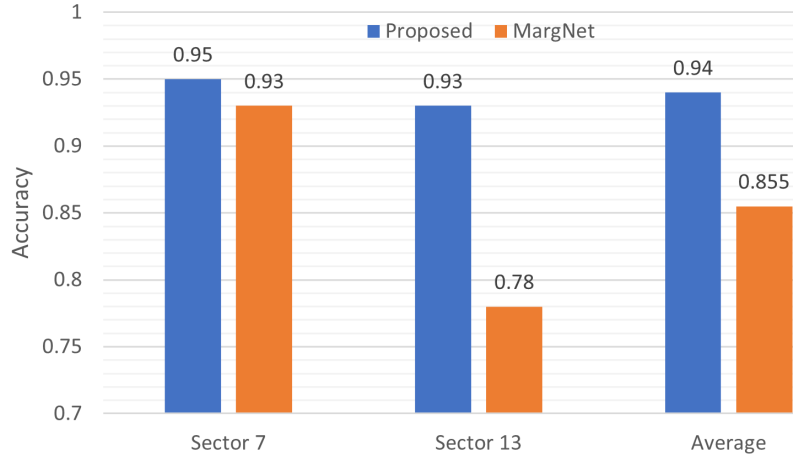
Figure 6: Star-Galaxy Classification performance of the proposed and the MargNet in terms of accuracy (Acc) on Sectors 7 and 13.

algorithm yields an accuracy rate of 0.95 and 0.93 in comparison to best best-performing existing algorithm namely MargNet, which yields an accuracy of 0.93 and 0.78 on sector 7 and 13 images, respectively. The samples of stars and galaxies concerning sectors 7 and 13 are shown in Figure 5 and the classification results are reported in Figure 6.

Zero-shot Sector Resiliency: In this setting, we have performed a 4-way zero setting experimental evaluation of the proposed algorithm and the MargNet. In other words, we have performed 4 fold cross-validation experiments, where in every fold one unseen sector is used for evaluation, and the remaining three sectors are used for training the classifier. It is observed that the proposed algorithm yields an average accuracy of 0.89 in comparison to an accuracy of 0.78 achieved by the MargNet. Further, the standard deviation of the proposed algorithm and the MargNet are 0.04 and 0.17, respectively. We believe the reliability against unseen sectors of the proposed algorithm further establishes the claim of its effectiveness in performing stars and galaxies classification.