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 due to 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 take care of the variations present in the different sector of the sky and hence lacks generalizability. 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 of such division can be seen from Figure 1 reflecting the variation present in different sectors. 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 (A) 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^o *to* 360^o *in celestial coordinates*. RA is divided into six equal intervals of 60^o 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^o is segmented into six intervals of

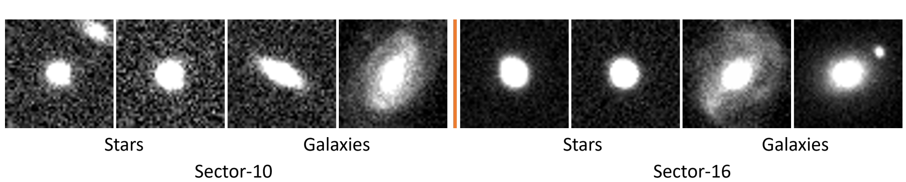

Figure 1: A sample image reflecting the challenges in identifying star-galaxies in different sectors.

Table 1: Star-galaxy classification (%) of the proposed and existing state-of-the-art algorithms.

Sector	**Proposed**	**CovNet**	**MargNet**
Sector-10	**96.55**	91.30	94.40
Sector-16	**95.15**	94.40	93.75
Combined	**95.25**	88.62	92.10

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**

30^o 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].

Once the divided sector images are obtained, we have provided them as input to the proposed custom convolutional neural network model (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 (Table 3); therefore, a shallow model of 3 convolutional layers each followed by max-pooling and ReLU activation has been developed. 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.

3 EXPERIMENTAL RESULTS AND ANALYSIS

In our study, we conducted several experiments to evaluate the effectiveness of the proposed CNN architecture. The primary focus of these experiments is on sector-10 (RA range from 180^o to 240^o and Dec range from 30^o to 60^o) and sector-16 (RA range from 180^o to 240^o and Dec range from 0^o to 30^o). 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.

On the computational front, the proposed algorithm took $\mathbf{25}$*s/epoch on the data of the combined sectors as compared to* 180*s/epoch and* 1610*s/epoch 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.

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

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`.

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.

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

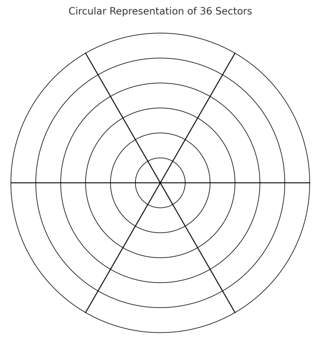

Figure 2: Sky Sector Map

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.

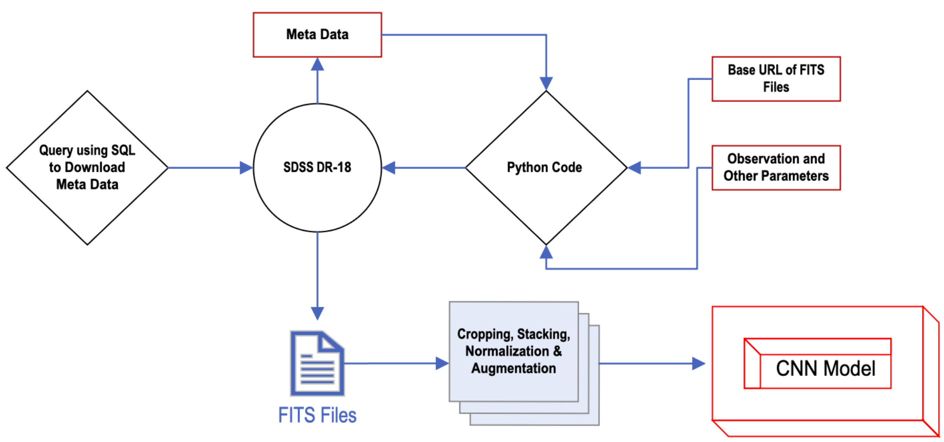

Figure 3: Data Workflow Diagram

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.

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

Table 5: Classification Report of Proposed and Existing Algorithms

Algorithm	Sector 10			Sector 16			Combined		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Proposed	97	97	97	95	95	95	95	95	95
CovNet	91	91	91	94	94	94	89	89	89
MargNet	94	94	94	94	94	94	93	92	92

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.