
Deep Learning for Classification of Low Surface Brightness Galaxies in Dark Energy Survey

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 Low Surface Brightness Galaxies (LSBGs) are faint, diffuse systems that are
2 difficult to identify and are often confused with imaging artifacts in wide-field
3 surveys. In this work, we apply convolutional neural networks (CNNs) to classify
4 LSBGs in the full six-year Dark Energy Survey (DES Y6) dataset, extending earlier
5 analyses on the Year 3 data. Training on the \sim 40,000 labeled Y6 objects used in
6 prior work, our CNN achieves 90.7% accuracy, surpassing traditional feature-based
7 methods. Applied to the full Y6 sample of \sim 70,000 labeled-objects, the network
8 reliably identifies artifacts, but struggles with ambiguous human-labeled LSBG
9 cases. Targeted augmentation reduces the fraction of LSBGs classified as artifacts
10 from 27.4% to 18.5%, bringing CNN predictions into closer alignment with human
11 labels. We also compare two classifiers that highlight a trade-off: one favors
12 completeness by retaining more LSBG candidates (at the risk of inaccuracy), while
13 the other favors purity by excluding ambiguous cases. Overall, CNNs classify
14 LSBGs with high efficiency and accuracy while also uncovering potential human
15 mislabels. With improved training data and stronger architectures, CNN-based
16 approaches will be indispensable for understanding the low-surface-brightness
17 universe in future large-scale surveys.

18 **1 Introduction**

19 **1.1 Low surface brightness universe and Identification of LSBGs**

20 Low Surface Brightness Galaxies (LSBGs) are diffuse stellar systems whose average luminosity
21 per unit area falls below that of the night sky, often by more than a magnitude. Their intrinsic
22 faintness makes them difficult to detect in traditional surveys, and only in the past few decades
23 have they been systematically studied [Impey and Bothun, 1997]. LSBGs are also among the most
24 dark-matter-dominated systems known, making them valuable laboratories for testing models of
25 dark matter and alternative theories of gravity. Studying these galaxies expands the census of galaxy
26 populations while providing constraints on galaxy formation, feedback processes, and the distribution
27 of dark matter in the universe [de Blok and McGaugh, 1997]. With modern surveys and deep-imaging
28 techniques, the number of known LSBGs continues to grow, offering new opportunities for refining
29 cosmological models.

30 The detection of LSBGs presents a major challenge in observational astronomy. Owing to their
31 diffuse profiles, LSBGs can be confused with a variety of imaging artifacts, including scattered light
32 from bright stars, cosmic rays, Galactic cirrus, and residual background fluctuations [Koda et al.,
33 2015]. Historically, the separation of real LSBGs from artifacts has relied heavily on visual inspection.
34 While effective in limited cases, this approach is subjective and ultimately impractical given the data
35 volumes produced by current surveys. These limitations have motivated the development of automated

36 classification methods. In particular, deep learning approaches based on convolutional neural networks
37 (CNNs) have achieved notable success in astronomical image analysis. For instance, CNNs applied
38 to the CFHTLS-Wide Survey detected faint tidal features in galaxies, outperforming conventional
39 methods when combined with regularization and data augmentation techniques [Walmsley et al.,
40 2019].

41 1.2 CNN-based classification of DES LSBGs

42 Tanoglidis et al. introduced *DeepShadows*, a CNN trained on Dark Energy Survey (DES) Year 3
43 (Y3) data to separate LSBGs from artifacts [Tanoglidis et al., 2020]. Using a large sample of visually
44 inspected galaxies and artifacts, the network achieved 92% accuracy, outperforming traditional ML
45 methods. Applied to Hyper Suprime-Cam (HSC) data, it reached 82.1% without retraining and 87.6%
46 with fine-tuning, demonstrating both strong performance within DES and adaptability across surveys.
47 Building on this, the full DES Year 6 (Y6) dataset offers new opportunities. Compared to Y3, Y6
48 is deeper, with improved photometric calibration that increases sensitivity to faint galaxies but also
49 introduces more diverse noise and artifacts.

50 2 Data and Methods

51 2.1 The Dark Energy Survey

52 The Dark Energy Survey (DES) is an optical and near-infrared imaging survey aimed at probing dark
53 energy through large-scale structure, galaxy clusters, and weak lensing. Using the 570-megapixel
54 Dark Energy Camera (DECam) on the 4-m Blanco Telescope in Chile, DES imaged $\sim 5,000 \text{ deg}^2$ of
55 the southern sky in five bands ($grizY$) between 2013 and 2019 [Abbott and the DES Collaboration,
56 2018]. The Year 6 (Y6) Gold catalog is the deepest and most uniform DES dataset to date, containing
57 ~ 669 million objects over nearly the full footprint. It reaches a depth of $i_{AB} \approx 23.4$ mag for
58 extended sources at $S/N \approx 10$ and delivers photometric uniformity better than 2 mmag [Bechtol
59 and Collaboration, 2025]. Compared to Year 3 (Y3), Y6 increases sensitivity to LSBGs while also
60 introducing more complex noise properties and artifact diversity.

61 2.2 LSBG catalog

62 The LSBG catalog was constructed in four stages:

- 63 1. Apply selection cuts using SOURCEEXTRACTOR parameters [Bertin and Arnouts, 1996].
- 64 2. Use a random forest classifier on these features, identifying $\sim 80,000$ candidates.
- 65 3. Perform visual inspection to reject false positives, leaving $\sim 50,000$ LSBGs and $\sim 30,000$
66 artifacts.
- 67 4. Fit single Sérsic profiles with GALFIT [Peng et al., 2002], yielding $\sim 40,000$ galaxies with
68 reliable light profiles.

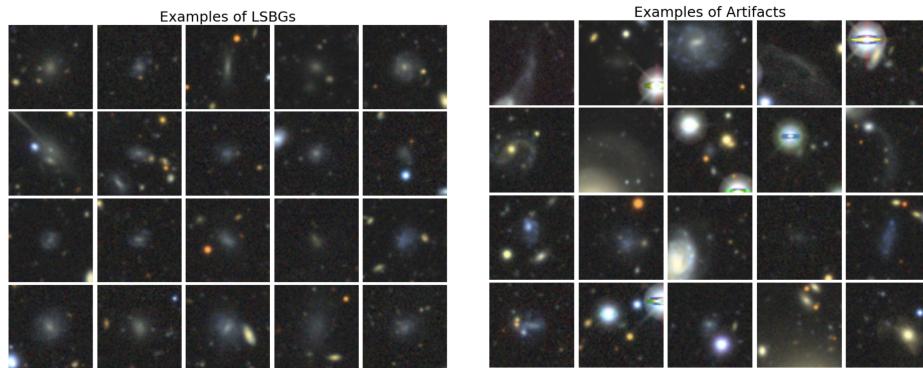


Figure 1: Representative cutouts from DES Y6. Left: low surface brightness galaxies (LSBGs); Right: artifacts.

69 The Y6 sample thus contains \sim 40,000 modeled LSBGs and \sim 30,000 artifacts. For comparison, the
70 Y3 analysis used \sim 40,000 visually labeled objects (20,000 LSBGs and 20,000 artifacts) [Tanoglidis
71 et al., 2020]. We matched those Y3 objects to their Y6 counterparts; nearly all were recovered in the
72 Y6 sample (Fig. 1).

73 2.3 CNN Architecture

74 We adopt a convolutional neural network (CNN) for binary classification of 64×64 RGB cutouts
75 (LSBG vs. artifact), following Tanoglidis et al. [2020]. The model has three convolutional blocks with
76 increasing filters (16, 32, 64), each with ReLU activation, 3×3 kernels, max pooling, dropout, and
77 L_2 regularization. A dense layer (1024 units) precedes the final sigmoid output. Batch normalization
78 is used throughout, and accuracy is the primary evaluation metric. Hyperparameters match those in
79 Table 1.

Table 1: CNN architecture used in this work.

Stage	Layers	Output
Input	$64 \times 64 \times 3$ RGB	$64 \times 64 \times 3$
Block 1	Conv2D (16, 3×3), ReLU, $L_2=0.13$, MaxPool 2×2 , Dropout 0.4	$32 \times 32 \times 16$
Block 2	Conv2D (32, 3×3), ReLU, $L_2=0.13$, MaxPool 2×2 , Dropout 0.4	$16 \times 16 \times 32$
Block 3	Conv2D (64, 3×3), ReLU, $L_2=0.13$, MaxPool 2×2 , Dropout 0.4	$8 \times 8 \times 64$
Flatten	—	4096
Dense	Dense(1024), ReLU, $L_2=0.12$	1024
Output	Dense(1), Sigmoid	1

80 3 Classification results

81 3.1 Main findings

82 We first test whether the Year 3 results can be reproduced on the same objects in Year 6. RGB cutouts
83 of \sim 40K “Y3 objects in Y6” were split into training, validation, and test sets, and used to train the
84 CNN described earlier. As shown in Fig. 2a, the training history indicates stable convergence without
85 overfitting. The model achieved 90.7% accuracy on the test set. For comparison, SVM and Random
86 Forest models trained on SOURCEEXTRACTOR features reached 82.9% and 80.6% accuracy.

87 We next evaluate the CNN on the full DES Y6 dataset of \sim 50K galaxies and 30K artifacts. After
88 removing \sim 10K galaxies with poor GALFIT fits or visual flags, the sample includes \sim 40K LSBGs
89 and \sim 30K artifacts. The classification results are shown in Fig. 2b. The network reliably identifies
90 artifacts, but 27% of human-labeled LSBGs are classified as artifacts. Re-inspection of low-confidence

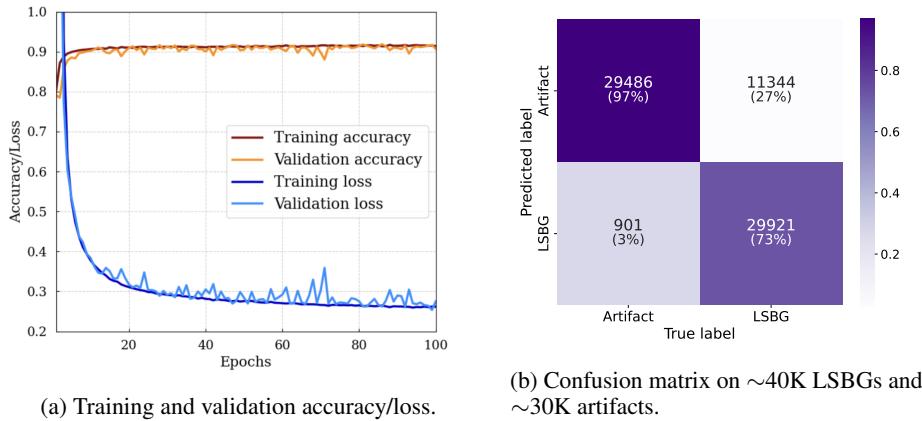


Figure 2: Training history (left) and CNN confusion matrix (right).

91 cases shows that many of these are in fact genuine LSBGs. This indicates that the network mainly
 92 struggles with ambiguous LSBGs, often classifying them as artifacts.

93 To address this, we explore augmenting the training dataset. A set of 380 cases that the network
 94 classified as artifacts but were confirmed as genuine LSBGs were expanded through flips (Fig. 3),
 95 yielding $\sim 1.5K$ augmented images. Retraining with this data reduced the fraction of LSBGs classified
 96 as artifacts from 27.4% to 18.5%. These results demonstrate that targeted augmentation improves
 97 agreement between CNN predictions and human labels.

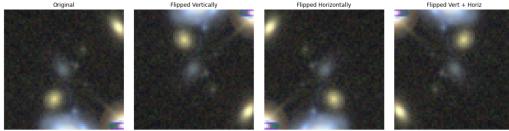


Figure 3: Example of data augmentation: original, vertical flip, horizontal flip, combined flip.

98 Then we compared two classifiers: **DeepSkyNet-Y6**, trained on $\sim 70K$ labeled Y6 objects, and
 99 **DeepSkyNet-Y3**, trained on 40K “Y3 objects in Y6” with augmentation. Both were evaluated on
 100 the same $\sim 20K$ test objects for comparison. As shown in Fig. 4, both classifiers performed well
 101 in identifying human-labeled artifacts. For human-labeled LSBGs, DeepSkyNet-Y6 retained more
 102 candidates, classifying 93% as LSBGs, while DeepSkyNet-Y3 retained only 76%. We also observed
 103 particularly interesting cases where both classifiers disagreed with the human labels—for example,
 104 objects labeled as artifacts but identified by both classifiers as LSBGs. Such cases likely point to
 105 mislabels in the original dataset.

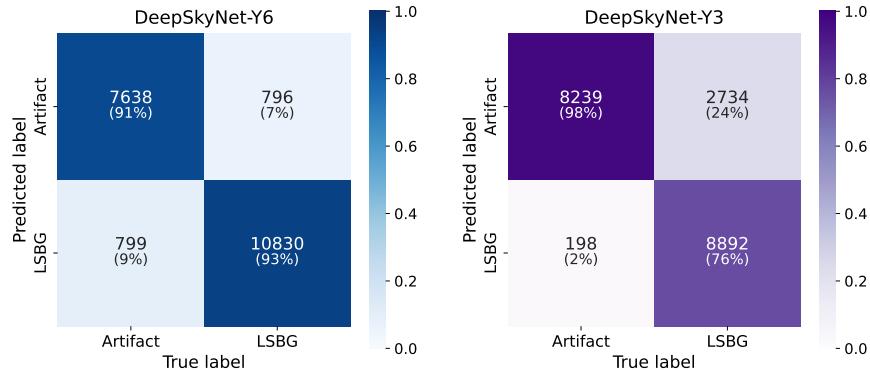


Figure 4: Confusion matrices summarizing the classification of $\sim 20K$ test objects for DeepSkyNet-Y6 (left, Blues) and DeepSkyNet-Y3 (right, Purples).

106 4 Discussion and Conclusions

107 In this work, we applied deep learning to classify low-surface-brightness galaxies (LSBGs) and
 108 artifacts in the full six-year DES dataset. Our results confirm and extend earlier Year 3 findings: CNNs
 109 achieve high accuracy, outperform traditional machine-learning pipelines, and expose inconsistencies
 110 in human labeling. Performance depends critically on the quality of the training data, but targeted
 111 augmentation and retraining already yield predictions that align more closely with human judgments.
 112 This demonstrates that the network’s behavior can be shaped through careful training, allowing it
 113 to classify in ways that better mirror human judgement, as seen in the comparison between the two
 114 classifiers. Looking ahead, the most immediate step is refining the DES Year 6 labels by correcting
 115 clear mistakes and flagging ambiguous cases. A finalized catalog will enable these networks to be
 116 applied across surveys to test generalizability and reduce reliance on manual inspection. At the
 117 same time, exploring alternative CNN architectures and systematic hyperparameter optimization,
 118 ideally with parallelized training, offers avenues for further gains. ML/AI methods have the immense
 119 potential to drive discovery in large-scale scientific surveys, and with better data and models, CNN-
 120 based approaches will undoubtedly advance the study of low-surface-brightness galaxies.

121 **References**

- 122 T. Abbott and the DES Collaboration. The dark energy survey: more than dark energy—an overview.
123 *Monthly Notices of the Royal Astronomical Society*, 460:1270–1299, 2018. (Commonly cited DES
124 overview; use the version appropriate to your context).
- 125 K. Bechtol and DES Collaboration. Dark energy survey year 6 results: Photometric data set for
126 cosmology (y6 gold). *arXiv preprint*, 2025.
- 127 E. Bertin and S. Arnouts. SExtractor: Software for source extraction. *Astronomy & Astrophysics
128 Supplement Series*, 117:393–404, 1996. doi: 10.1051/aas:1996164.
- 129 W. J. G. de Blok and S. S. McGaugh. The dark and baryonic matter content of low surface brightness
130 disk galaxies. *Monthly Notices of the Royal Astronomical Society*, 290:533–552, 1997. doi:
131 10.1093/mnras/290.3.533.
- 132 C. D. Impey and G. D. Bothun. Low surface brightness galaxies. *Annual Review of Astronomy and
133 Astrophysics*, 35:267–307, 1997. doi: 10.1146/annurev.astro.35.1.267.
- 134 J. Koda et al. Approximately a thousand ultra-diffuse galaxies in the coma cluster. *Astrophysical
135 Journal Letters*, 807:L2, 2015. doi: 10.1088/2041-8205/807/1/L2.
- 136 C. Y. Peng, L. C. Ho, C. D. Impey, and H.-W. Rix. Detailed structural decomposition of galaxy
137 images. *Astronomical Journal*, 124:266–293, 2002. doi: 10.1086/340952.
- 138 D. Tanoglidis, A. Ćiprijanović, and A. Drlica-Wagner. Deepshadows: Separating low surface
139 brightness galaxies from artifacts using deep learning. *Astronomy and Computing*, 35:100469,
140 2020. doi: 10.1016/j.ascom.2021.100469.
- 141 M. Walmsley et al. Identification of low surface brightness tidal features in galaxies using convo-
142 lutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 491:1554–1574,
143 2019.