

## Article

# Using Machine Learning to Profile Asymmetry between Spiral Galaxies with Opposite Spin Directions

Lior Shamir 

Department of Computer Science, Kansas State University, Manhattan, KS 66506, USA; lshamir@mtu.edu;  
Tel.: +1-785-532-4809

**Abstract:** Spiral galaxies can spin clockwise or counterclockwise, and the spin direction of a spiral galaxy is a clear visual characteristic. Since in a sufficiently large universe the Universe is expected to be symmetric, the spin direction of a galaxy is merely the perception of the observer, and therefore, galaxies that spin clockwise are expected to have the same characteristics of galaxies spinning counterclockwise. Here, machine learning is applied to study the possible morphological differences between galaxies that spin in opposite directions. The dataset used in this study is a dataset of 77,840 spiral galaxies classified by their spin direction, as well as a smaller dataset of galaxies classified manually. A machine learning algorithm was applied to classify between images of clockwise galaxies and counterclockwise galaxies. The results show that the classifier was able to predict the spin direction of the galaxy by its image in accuracy higher than mere chance, even when the images in one of the classes were mirrored to create a dataset with consistent spin directions. That suggests that galaxies that seem to spin clockwise to an Earth-based observer are not necessarily fully symmetric to galaxies that spin counterclockwise; while further research is required, these results are aligned with previous observations of differences between galaxies based on their spin directions.

**Keywords:** machine learning; galaxies; anisotropy



**Citation:** Shamir, L. Using Machine Learning to Profile Asymmetry between Spiral Galaxies with Opposite Spin Directions. *Symmetry* **2022**, *14*, 934. <https://doi.org/10.3390/sym14050934>

Academic Editors: Stefano Profumo and Chin-Ling Chen

Received: 5 April 2022

Accepted: 30 April 2022

Published: 4 May 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The morphology of a galaxy can provide useful information about its characteristics, and galaxy morphology has been studied extensively in the past century [1]. Basic classification of galaxy morphology includes the broad morphological types of elliptical, spiral and lenticular galaxies, and the classification can be extended to irregular [2] and peculiar [3] galaxies. Established and commonly used examples of more detailed systems for galaxy morphology classification are the Hubble sequence and the de Vaucouleurs system [4].

One of the important tools used for studying the morphology of galaxies are catalogs of galaxies annotated by their morphology. Such catalogs can be prepared by astronomers who analyze large sets of galaxy images manually [5–7]. In the era of Big Data and autonomous digital sky surveys, the vast astronomical pipelines reinforces the use of automation to annotate many millions of galaxies. Since manual analysis is limited by the amount of galaxies that can be analyzed, catalogs of galaxy morphology have been compiled by using the power of a large number of non-expert volunteers who inspect galaxy images manually, and collectively are able to provide morphological analysis of a large number of galaxies [8–10]. More recently, computer vision algorithms have been used to annotate galaxies automatically and generate catalogs [11–14].

A visually dominant morphological feature of a spiral galaxy is its spin direction. Spiral galaxies may have clockwise or counterclockwise oriented patterns, separating spiral galaxies into two classes. Galaxy spin direction may be considered to be merely a feature of the location of the observer, as a galaxy that may seem to be spinning clockwise to an observer on Earth might seem to be spinning counterclockwise to an observer located at some distant galaxy. Therefore, galaxies that spin clockwise are expected to be symmetric

to galaxies that spin counterclockwise in all of their morphological features, except from their spin direction.

Previous studies have shown clear evidence of photometric differences between clockwise and counterclockwise galaxies [15,16]. These experiments were based on photometric data taken from Sloan Digital Sky Survey [15], the Panoramic Survey Telescope and Rapid Response System [16], and Hubble Space Telescope [16]. For instance, a supervised machine learning experiment was performed such that the label of each sample was the galaxy spin direction (clockwise or counterclockwise), and the features of each sample were the 454 photometric measurements provided by the SDSS pipeline. The results showed that for ~64% of the galaxies, the machine learning algorithm was able to predict whether the galaxy is clockwise or counterclockwise based on the photometric variables [15]. That accuracy is significantly stronger than the random chance classification accuracy of 50% ( $p < 10^{-5}$ ) [15]. The experiment was also repeated using a galaxy dataset that was classified to galaxies that spin clockwise and galaxies that spin counterclockwise. That was done through a fully automatic process and without any human intervention, and provided similar results of ~65% classification accuracy [15].

Other studies showed evidence of asymmetry between the number of galaxies with opposite spin directions as observed from Earth [17–21]. This consistent evidence using different instruments and various data analyses suggests that certain differences can exist between galaxies with opposite spin directions. Here, a machine learning method is applied to test whether galaxies that spin clockwise are morphologically the same as galaxies that spin counterclockwise as observed from Earth; while each galaxy is different, comparison of a large number of galaxies enables a statistical analysis that can allow to identify possible morphological differences between galaxies that spin in opposite directions.

## 2. Data

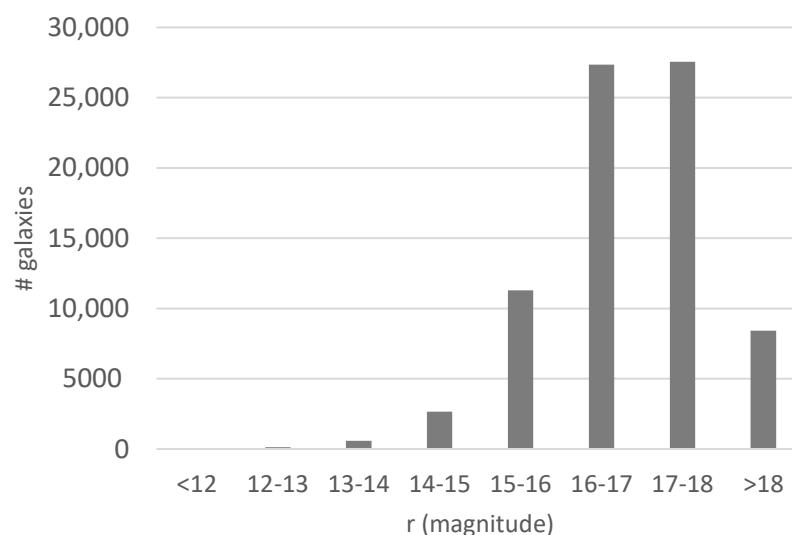
The main dataset used in the experiment contains 77,840 spiral galaxies that were classified automatically into clockwise and counterclockwise galaxies. The dataset is described in [21]. As described in [13], the galaxies were initially selected from SDSS DR8 such that the Petrosian radius of the galaxies was at least  $5.5''$ , and  $i$  magnitude of less than 18 to filter galaxies that are too small or too dim to identify their morphology.

Then, the galaxies were separated by their spin directions. That was done by applying the *Ganalyzer* algorithm [22], as was done in [19–21]. *Ganalyzer* first applies a transformation of the galaxy image into its radial intensity plots. The radial intensity plot is an image of  $360 \times 35$  pixels, such that the  $X$  axis is the polar angle of the pixel in the original image compared to the galaxy center, and the  $Y$  axis is the radial distance from the galaxy center in percents of the galaxy radius. Then, a peak detection algorithm is applied to identify groups of peaks in the horizontal lines of the radial intensity plot [22]. Since pixels on the arm of the galaxy are expected to be brighter than pixels that are not on the arm of the galaxy, the group of peaks identify the galaxy arms. Each vertical line made by the peaks detected in the neighboring horizontal lines of the radial intensity plot is a galaxy arm, and the slope of the line corresponds to the direction of the galaxy arm. A linear regression is then applied to each vertical line, and the sign of the regression coefficient reflects the direction of the curve. The direction of the curve indicates whether the arm is leading or trailing, and therefore, can deduce the spin direction of the galaxy. The *Ganalyzer* algorithm is fully described with examples and performance analysis in [19–22]. The use of the algorithm to generate the specific dataset of galaxy images used here is described in [21].

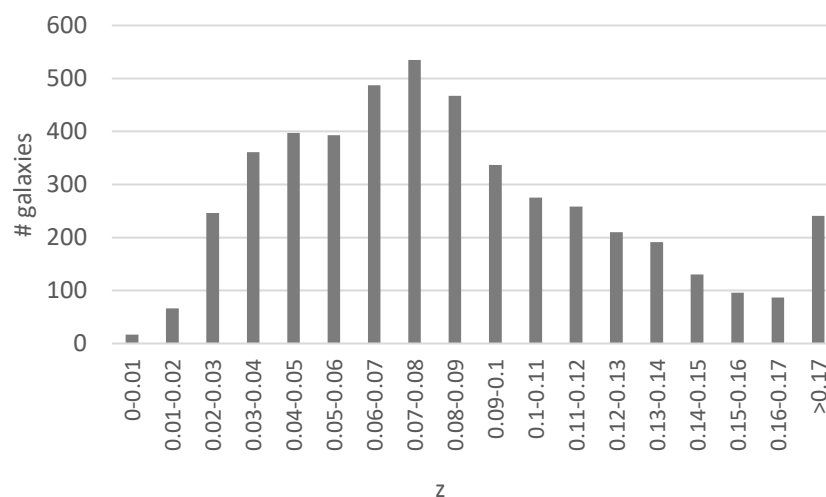
Because many galaxies, such as elliptical galaxies, do not have a visually clear spin direction, not all galaxies can be assigned with a spin direction by their visual appearance alone. To avoid galaxies with unclear spin direction, only galaxies that had at least 30 peaks in their radial intensity plots were used. Galaxies that did not have at least 30 peaks were not used in the analysis. Separating the galaxies to clockwise and counterclockwise galaxies

provided a dataset of 39,187 galaxies that spin clockwise and 38,653 galaxies that spin counterclockwise. The entire process of the galaxy annotation is described in [21].

Figures 1 and 2 show the distribution of the  $r$  magnitude and the redshift of the galaxies in the dataset, respectively. The vast majority of the galaxies do not have spectra, and therefore, just the subset of 10,281 galaxies that had spectroscopic information could be used for deducing the redshift distribution.



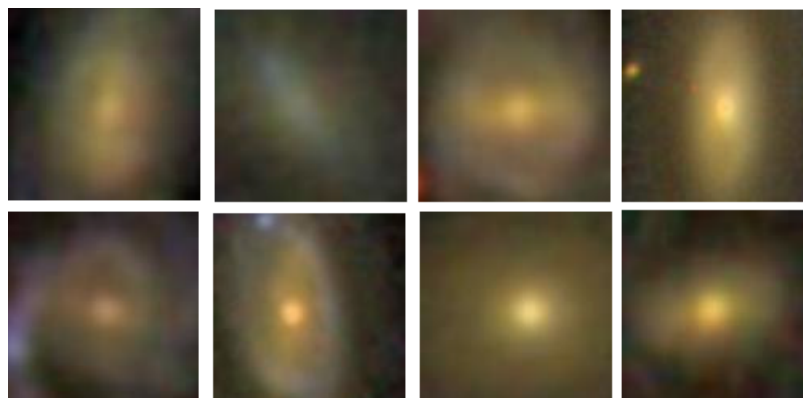
**Figure 1.** The  $r$  magnitude distribution of the galaxies in the dataset.



**Figure 2.** Redshift distribution of the galaxies in the dataset.

In addition to the dataset of automatically classified galaxies, another dataset that was used in the experiment was a dataset of 13,440 galaxies used in [15]. These galaxies were annotated and inspected manually by the author in a long labor-intensive process. Because the human perception is sensitive to the spin direction of the galaxy [23], the galaxies were mirrored randomly before they were annotated. Then, the galaxies were mirrored again to verify that the annotation was correct. The two annotations of each galaxy were compared, and in case of disagreement, the galaxy was inspected carefully for the third time to determine whether its spin direction can be identified clearly. Just 12 galaxies had 2 conflicting annotations. Because the third manual inspection did not lead to a clear identification of the spin directions of these galaxies, all of these galaxies were excluded

from the analysis. Figure 3 shows examples of galaxies that are not necessarily elliptical, but their spin direction could not be determined in an obvious and reliable manner.



**Figure 3.** Examples of galaxy images that are not necessarily elliptical, but their spin direction could not be identified clearly.

That process required substantial labor of about 150 h of work, but provided a clean dataset of 6941 clockwise galaxies and 6499 counterclockwise galaxies. For the rest of the galaxies, the spin direction could not be determined (e.g., edge-on galaxies). Galaxy pairs were also excluded from the experiment, and therefore, the weak spin magnitude correlation in galaxy pairs [24] did not affect the results. A final inspection step included random selection of 500 galaxies to verify that all galaxies are classified correctly, and found no errors in the classifications.

### 3. Machine Learning Algorithm

Galaxy morphology has been analyzed by using deep neural networks (DNNs), and specifically deep convolutional neural networks [12,14,25–30]; while DNNs have demonstrated excellent performance in the automatic classification of image data, there are several downsides to using DNNs for purposes related to analysis of subtle asymmetries. DNNs rely on a large number of data-driven non-intuitive rules that are determined automatically during the training process, and are very difficult to understand. The nature of deep neural networks, therefore, makes it more difficult to turn the empirical results such as classification accuracy into useful insights about the specific features that identify differences between the galaxy morphological types. Moreover, DNNs are also sensitive to background information that can lead to substantial biases [31]. These biases were also identified to be present among galaxy images, but are difficult to quantify due to the complex non-intuitive nature of the algorithm [32]. In the absence of specific measurements, and given the assumption that the DNNs can be biased, DNNs might not be a sound approach to identifying subtle asymmetries in the morphology of galaxies.

To provide a more informative machine learning analysis of a possible asymmetry between galaxies with opposite spin directions, the Wndchrm method was used [33]. Wndchrm is a non-parametric method that is not based on deep neural networks, and has been widely used to classify and analyze galaxy morphology [13,34]. Wndchrm implements a comprehensive and large set of numerical content descriptors that reflect multiple aspects of the image content, including textures [35,36], edges [37], fractals [38], statistics of the pixel intensities [39], polynomial decomposition [40,41], Radon features [42], and Gabor filters [43]. These features are filtered automatically by their Fisher discriminant scores [44], and classified by the Weighted Nearest Neighbor [33] or other pattern recognition algorithms as will be described in this section.

Wndchrm is non-parametric in the sense that all features are computed for all images. The user is not required to make any preliminary assumption about the data, and the most informative features are selected automatically by the pattern recognition algorithm, and

without involving decisions made by the user. The advantage of the approach compared to DNNs is that it uses defined features, and therefore, the analysis can identify specific morphological features that correlate with the asymmetry of spin directions of spiral galaxies. That is different from DNNs, which can provide high classification accuracy but does not excel in its ability to identify explainable differences between the classes; while the Wndchrm approach can identify specific defined image descriptors that can differentiate between the morphology of the galaxies, it is not certain that such features will be identified with statistical significance. However, the analysis attempts to identify such features. That analysis is not possible with DNNs, which often work as a “black box”, and does not allow to identify explainable attributes that differentiate between the different classes. An analysis with a DNN is described in Section 3.3.

### 3.1. Machine Learning Analysis of the Manually Annotated Galaxies

The manually annotated galaxies taken from [15] allowed a dataset of 12,000 galaxy images such that 5000 galaxies spinning clockwise and 5000 galaxies spinning counterclockwise galaxies were used for training, and 1000 from each class were used for testing. That allowed to train and test the Wndchrm image classifier described in Section 3. The classifier was able to differentiate between the two classes in accuracy of  $\sim 54.3\%$ . The classification accuracy is not high, but it is higher than 50% mere chance accuracy, when the prediction of the galaxy is done by guessing. Using cumulative binomial probability such that the number of trials is 2000 and the chance of success is 0.5, the probability to have 1085 or more successes by mere chance is  $\sim 0.00008$ .

Wndchrm was initially designed as a machine learning tool that can analyze images of cells [33], and therefore, its image content descriptors are rotationally invariant, so that mirroring an image is not expected to lead to a difference in the Wndchrm analysis. Therefore, while clockwise and counterclockwise galaxies are visually different from each other, Wndchrm is not expected to differentiate between these galaxies, and a difference detected by Wndchrm might reflect other differences between the galaxy images that are not directly related to the spin direction. However, it might still be possible that some of the features are sensitive to the spin direction, leading to the ability of the classifier to identify between clockwise and counterclockwise spiral galaxies.

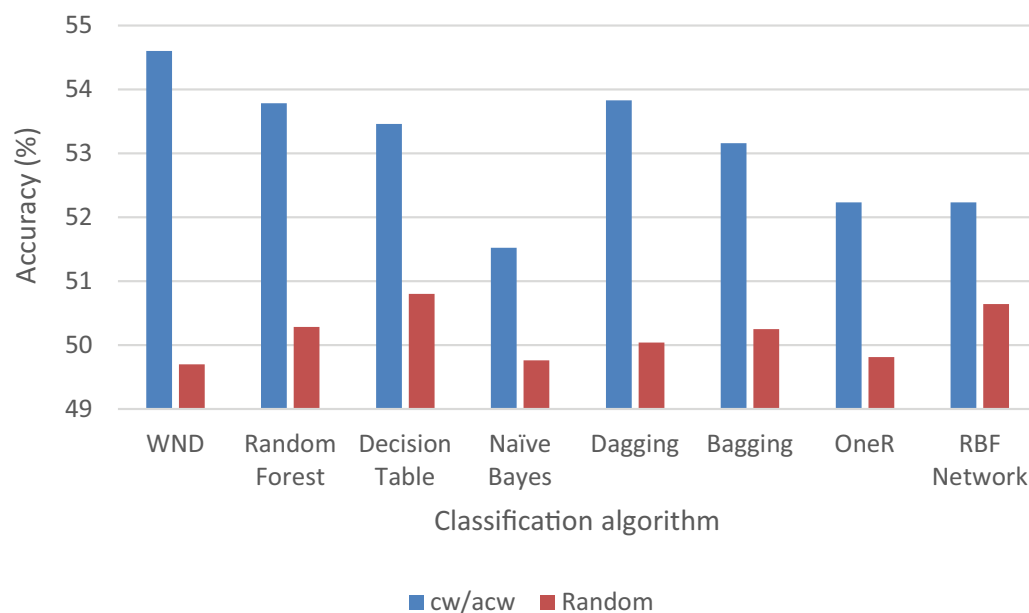
To completely eliminate the effect of the spin direction, two datasets were created such that each dataset had two classes, and all galaxies in both classes have the same spin direction. That was done by mirroring all images of one of the classes in each dataset. The first dataset contained one class of the original clockwise galaxy images and another class of the mirrored counterclockwise galaxy images. The second dataset contained the original counterclockwise galaxy images, while the clockwise galaxies were mirrored. That provided two datasets such that each dataset contained two classes, and all galaxies in the two classes had the same spin direction. The uncompressed TIFF file format was used, so that no compression can have an impact on the mirrored images.

While the TIFF format is not a common file format in astronomy, it is much more frequent in the field of machine vision. Normally, the TIFF format does not allow to deduce accurate photometric information that is available when using other formats such as FITS. However, in the case of this study, the important information is not the photometry, but the morphology of the galaxies, and therefore, the ability to deduce accurate photometry, is not a primary expectation from the image format. Because the TIFF images contain several color channels in a single image file, they provide more useful information to analyze the shape of the galaxy compared to FITS images, which normally provide a single color band. The TIFF files are not compressed, to avoid possible effect of the compression algorithm.

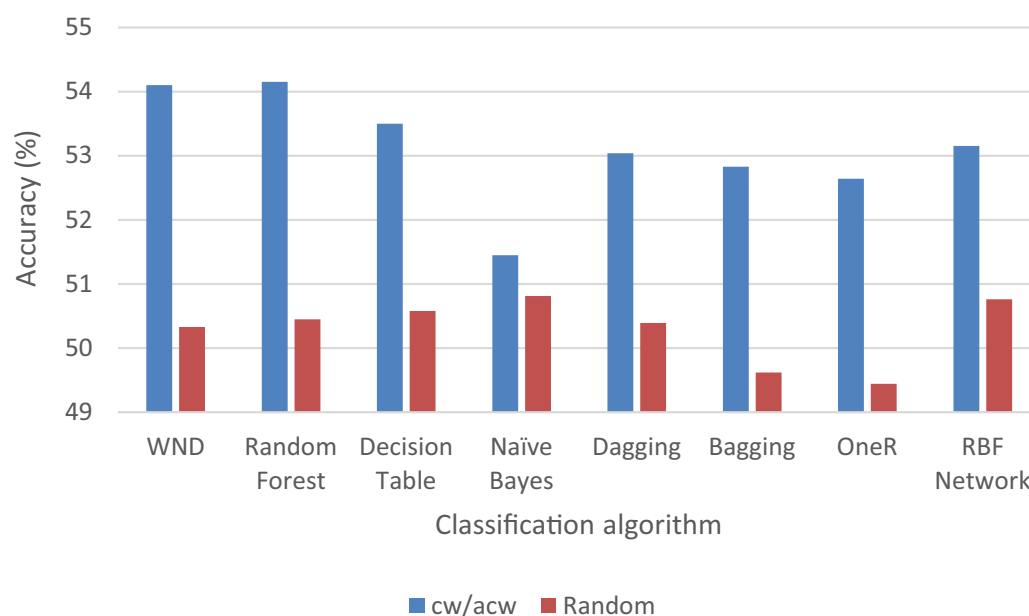
The machine learning experiment was then repeated using each of these datasets. The numerical image content descriptors were classified using several different pattern recognition algorithms. These supervised machine learning algorithms are Weighted Nearest Distance (WND) [33], as well as Random Forest [45], Decision Table [46], Naive

Bayesian classifier [47], Dagging [48], Bagging [49], OneR [50], and radial basic function (RBF) Networks [51], available as part of the Weka machine learning software [52,53].

The classification accuracy was also compared to the classification accuracy observed when the galaxies were separated randomly into clockwise and counterclockwise galaxies. Figures 4 and 5 display the classification accuracy of each of the classifiers when the clockwise galaxies are mirrored and when the counterclockwise galaxies are mirrored, respectively.



**Figure 4.** Classification accuracy of predicting the galaxy spin direction from the galaxy image when all clockwise galaxies are mirrored. The classification accuracy was also tested when the galaxies were separated randomly to two classes.



**Figure 5.** Classification accuracy of predicting the galaxy spin direction by its galaxy image when all counterclockwise galaxies are mirrored.

As the figures show, the classification accuracy of the dataset where the clockwise galaxies were mirrored was  $\sim 54.6\%$  ( $p \simeq 0.00002$ ), and the accuracy of the dataset in which the counterclockwise galaxies were mirrored was  $\sim 54.1\%$  ( $p \simeq 0.0001$ ). The observation



that clockwise and counterclockwise galaxies can be identified by a machine learning algorithm with higher accuracy than mere chance shows that even when the spin direction is the same, the classifier can still differentiate between clockwise and counterclockwise galaxies, indicating that there could be differences between these galaxies other than the spin directions. That means that when observing a large number of spiral galaxies, galaxies that spin in one direction can be morphologically different from galaxies spin in the opposite direction.

When assigning the galaxies with random spin directions, the classification accuracy using WND was  $\sim 49.7\%$ . The probability to have that classification accuracy by chance is  $\sim 0.37$ . All other machine learning algorithms provided similar results, i.e., around mere chance accuracy.

### 3.2. Machine Learning Analysis Using Computer-Annotated Data

To compare the results with a dataset that was annotated in a fully automatic manner, the same analysis was applied using the automatically classified dataset described in Section 2. For the experiment, the galaxies with spectra were used, providing a dataset comparable in size to the manually classified dataset, with 5142 counterclockwise galaxies and 5139 galaxies that spin clockwise.

As with the manually classified dataset, the clockwise galaxies were classified against the mirrored counterclockwise galaxies, and the counterclockwise galaxies were classified against the mirrored clockwise galaxies. That led to two different datasets, each with two classes, and in each dataset, the galaxy images in both classes had the same spin direction. The classification accuracy of the clockwise and counterclockwise galaxies using different pattern recognition algorithms are displayed in Figure 6.

As the graphs show, the classification accuracy is comparable to the classification accuracy of the dataset of manually classified galaxies described in Section 3.1. The random forest algorithm outperformed the WND classifier, and the Decision Table algorithm provided the highest classification accuracy. The naive Bayes and the OneR classifiers provide a classification accuracy very close to mere chance, but the other classifiers all provide a classification accuracy higher than random.

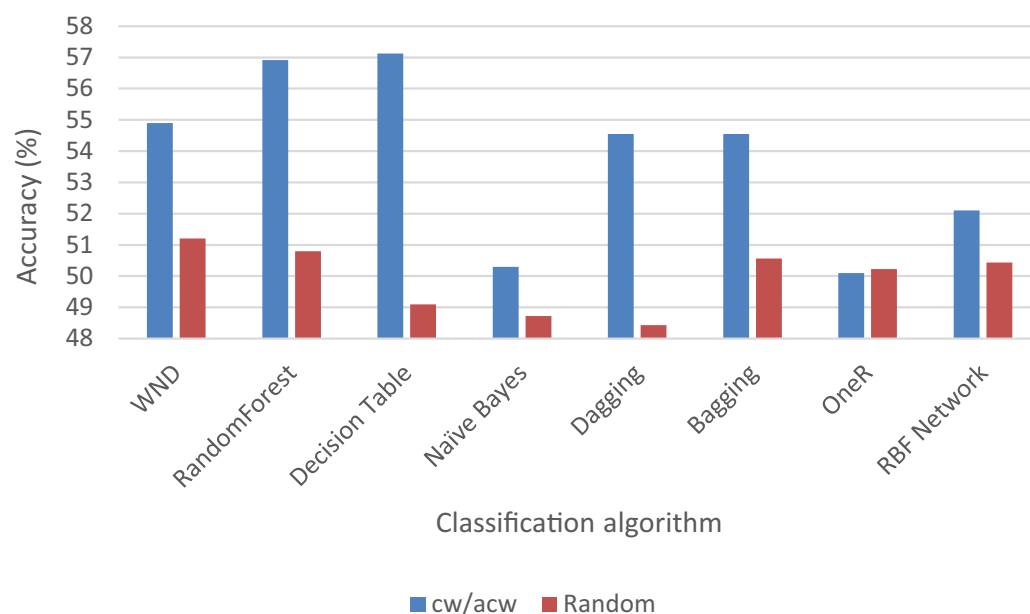
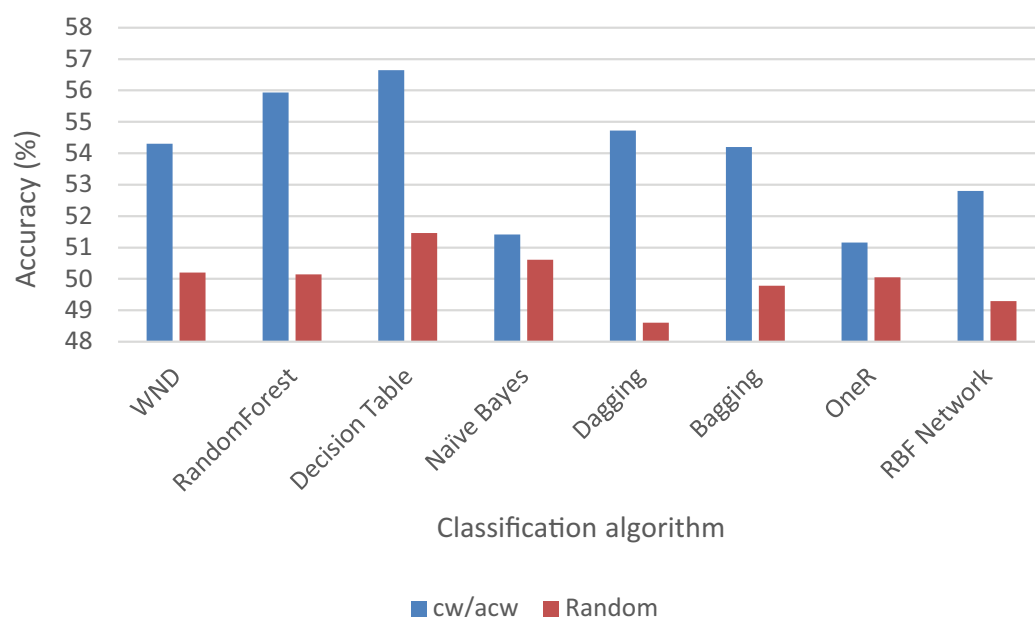


Figure 6. Cont.



**Figure 6.** Classification accuracy of the different algorithms when using automatically classified galaxies. The classification accuracy was measured when the clockwise galaxies are mirrored (**top**) and when the counterclockwise galaxies are mirrored (**bottom**).

### 3.3. Analysis Using a Deep Convolutional Neural Network

Another experiment used the same dataset used in Section 3.2, but the images classifier that was used was a deep convolutional neural network. The neural network that was used was of simple architecture as described in [32], and based on the LeNet-5 architecture [54]. The full description of the network is available in [32].

As before, 1000 images from each class were used for testing, and the rest for training and validation. Table 1 summarizes the classification accuracy observed with three different experiments. In the first experiments, the galaxies were not mirrored. That experiment was followed by two other experiments in which the clockwise or counterclockwise galaxies were mirrored to normalize the spin direction of the entire dataset. A fourth experiment was performed by assigning random spin directions to the galaxies. In all cases, the neural networks were trained from initial random weights, and without using any pre-defined weights in the form of transfer learning that might have an unexpected impact on the analysis.

**Table 1.** Classification accuracy between images of galaxies with opposite spin directions when using a deep convolutional neural network. In one experiment, the original images were used, while in the other experiments, the images in one of the classes were mirrored.

Dataset	Accuracy (%)
Original images	78.5
Clockwise mirrored	55.3
Counterclockwise mirrored	54.9
Random labels	50.8

As the table shows, the original images were classified in accuracy far higher than mere chance. That can be explained by the fact that CNNs are not rotationally invariant, and therefore, the CNN can differentiate between galaxies with opposite spin directions. When normalizing the images by mirroring one of the classes, the results become comparable and slightly higher than the results observed with the feature-based machine learning algorithms. When assigning the galaxies with random labels, the classification accuracy drops to random accuracy level.



### 3.4. Numerical Image Content Descriptors

Wndchrm uses a comprehensive set of 2885 numerical content descriptors of the visual data [33], weighted by their Fisher discriminant scores. More informative descriptors have a higher Fisher discriminant score, and therefore, a stronger impact on the classification decision. These image content descriptors are extracted from the raw pixels, as well as from different transforms of the image [33]. The groups of numerical image content descriptors with the highest cumulative Fisher discriminant scores are displayed in Figure 7. As the figure shows, numerous different numerical image content descriptors differentiate between galaxies that spin clockwise and galaxies that spin counterclockwise.

None of the numerical image content descriptors shown in Figure 7 show a statistically significant difference in the means measured in clockwise and counterclockwise galaxies. The features with the highest Fisher discriminant scores are the Zernike features [41] extracted from the Wavelet transforms of the galaxy images. The Zernike moment of degree  $m$  and angular dependence  $n$  is defined by  $A_{mn} = \frac{M+n}{\pi} \int \int f(x, y) [V_{mn}(x, y)]^* dx dy$ , where  $x^2 + y^2 \leq 1$  is the complex conjugate and  $V_{mn}(x, y)$  is the polar coordinate expression of the Zernike polynomial  $V_{mn}(r, \theta) = R_{mn}(r) \exp(jn\theta)$ .

The Zernike features used by Wndchrm are the magnitude of the complex numbers, leading to 72 descriptors based on the different degrees and angular dependences, up to a degree of 15 [33]. The means of these descriptors measured from clockwise and counterclockwise galaxies are displayed in Figure 8.

None of these features show a statistically significant difference, but it can be noticed that the feature values are somewhat higher for counterclockwise galaxies in the lower degrees, especially when the angular dependence is 0. Although the differences are not statistically significant, they can imply that counterclockwise galaxies have better fitness when the number of consistent changes in pixel intensity around the center is lower. That can happen if clockwise galaxies are flatter and more dense than counterclockwise galaxies, and have a less dominant bulge. It is important to note that these differences are not statistically significant, and the variety of shapes of spiral galaxies in each of the group makes it difficult to make clear characterization of all galaxies within each class.

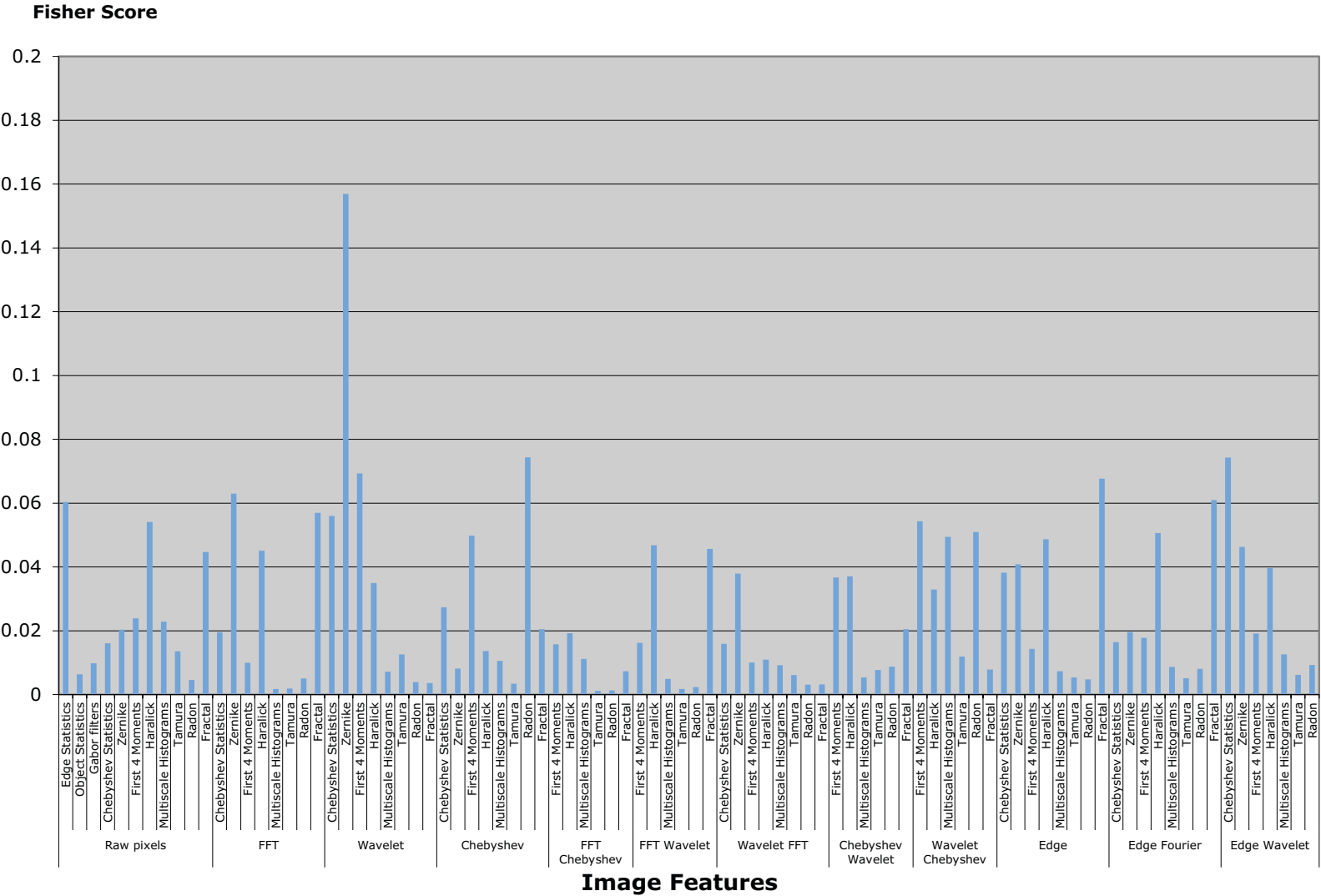
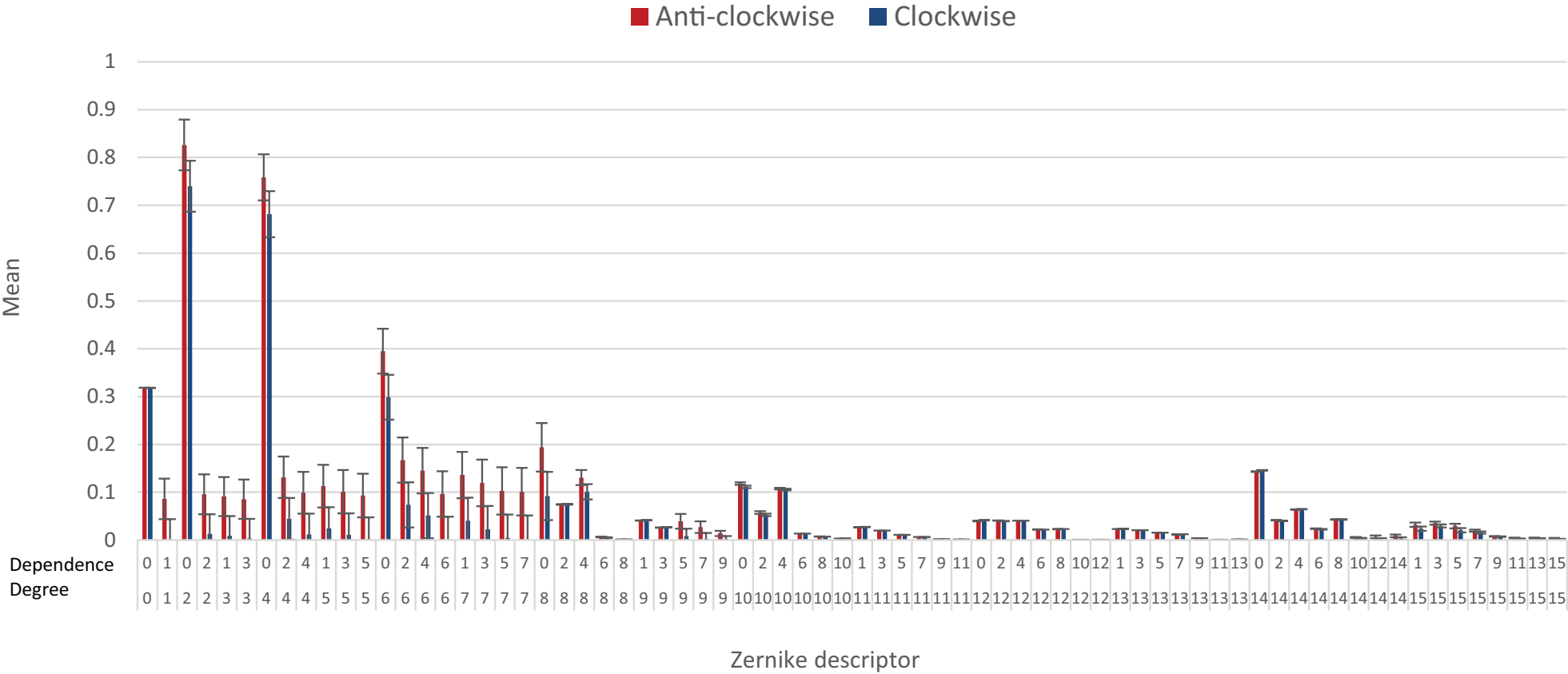


Figure 7. Fisher discriminant scores of the image content descriptors that differentiate between the clockwise and counterclockwise galaxies.

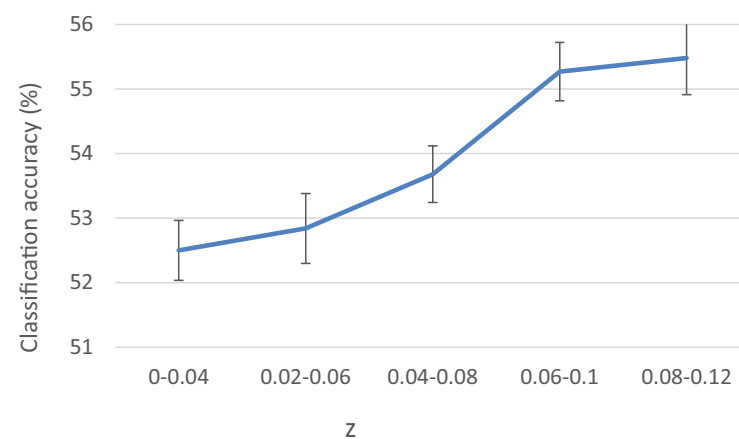


**Figure 8.** Means and standard errors of the Zernike features extracted from the Wavelet transform measured from clockwise and counterclockwise galaxies.

### 3.5. Redshift Effect on the Classification of Clockwise and Counterclockwise Galaxies

Another experiment that was performed aimed at testing the change in the accuracy of the automatic classification when the galaxies are limited to certain redshift ranges. For that purpose, the galaxies were divided into five redshift ranges of 0–0.04, 0.02–0.06, 0.04–0.08, 0.06–0.1, and 0.08–0.12. When  $z > 0.12$ , the number of galaxies drops sharply, and does not allow sufficient data to train and test the machine learning algorithm.

The ability of the algorithm to separate automatically between clockwise and mirrored counterclockwise galaxies (and vice versa) was done as described in Section 3, such that 1500 clockwise and counterclockwise galaxies in each redshift range were used for training, and 100 galaxies from each class for testing. Because the number of test galaxies was low, the experiment was done by using cross-validation, such that the classification of the galaxies in each redshift range was repeated 15 times, and in each run, 100 different galaxies per class were used for testing. Then, the accuracy of the 15 runs was averaged to deduce the classification accuracy. Figure 9 shows the average classification accuracy between images of clockwise galaxies and mirrored images of counterclockwise galaxies, and between counterclockwise galaxies and mirrored clockwise galaxies for each redshift range.



**Figure 9.** Classification accuracy using WND when the galaxies are limited to certain  $z$  ranges.

As the figure shows, the classification accuracy tends to increase when the redshift gets higher; while it is expected that galaxies with higher redshift would seem fainter and smaller, the galaxies in the dataset were all bright and had large surface size. In any case, the effect of the redshift is expected to impact both clockwise and counterclockwise galaxies equally. That means that if clockwise galaxies become smaller and fainter due to the higher redshift, galaxies spinning counterclockwise in the same redshift range are also expected to become smaller and fainter. Because both clockwise and counterclockwise galaxies are taken from the same field and the same redshift range, the impact on the redshift should be equal on both clockwise and counterclockwise galaxies, and therefore, is not expected to increase the classification accuracy when assuming that clockwise and counterclockwise galaxies are not different in their morphology.

## 4. Discussion

The nature of galaxy rotation is one of the greatest mysteries in space science, and it is not clear why and how galaxies rotate. Early astronomers assumed that the rotation of stars around their galaxy center is driven by gravity, as is the case of planets orbiting their star. That turned to be not nearly the case, as it has been shown and proven that unlike planets, the velocity of stars does not change significantly as their distance from the galaxy center increases [55]. That unexpected observation led to the assumption that the vast majority of the mass of a galaxy is made of dark matter that does not interact with light or any other radiation [55]. Other theories proposed that the rotation of galaxies is driven by different physics that does not agree with the known Newtonian physics [56–58], and that

modifications to Newtonian dynamics (MOND) can explain the anomaly of the galaxy rotation, as well as other observations such as the Hubble constant tension [59]. Despite five decades of research, no proven answer has been found, and research efforts are still being continued. In fact, more recent observations showed that the common assumption that galaxy rotation was initiated by gravitational interactions might not be correct, as rotating galaxies have been observed before they could interact with other galaxies and spin according to the current models [60]. Such observations agree with theories of primordial spin [61], and in such a case, it can be expected that the spin directions of galaxies at higher redshift are aligned [20].

This paper applies machine learning to study the symmetry of galaxies with opposite spin directions. Substantial previous work, starting in the 20th century, proposed that the number of galaxies with opposite spin directions is not necessarily equal within statistical fluctuations [17–21]. More recent work also proposed that the brightness and color of galaxies that spin in opposite directions is, on average, different [15,16]. This study aims at addressing a new type of asymmetry between galaxies that rotate in opposite ways, which is the morphology of the galaxies. The morphology of a galaxy is definitely linked to its color and distance from Earth, and therefore, differences in color and brightness of the galaxies can also be linked with their morphology, making the observation shown in this paper expected.

By analyzing the morphology using machine learning, this study shows that the morphology of galaxies that spin clockwise can be different from the morphology of galaxies that spin counterclockwise. Naturally, a single galaxy cannot be used to show such difference, and therefore, the study is done by analyzing a large number of galaxies, and classifying them by their images. The results show that a machine learning classifier can identify the galaxy spin direction based on its shape in accuracy higher than mere chance. That shows that spiral galaxies that spin in opposite directions are not necessarily symmetric on a large scale.

One of the observation is the link between the morphological differences and the redshift. That link might be considered unexpected, as the level of details of the galaxy images is expected to decline as the redshift gets higher, and therefore, an image classifier is expected to become less informative when classifying between these images of galaxies with higher redshift. On the other hand, it has been shown that the asymmetry between the number of galaxies that spin in opposite directions increases with the redshift [20]. Redshift is known to correlate with the morphology of the galaxies [62,63]. Therefore, a higher population of galaxies that spin in a certain direction at the higher redshift ranges is expected to lead to certain average differences in the morphology of these groups of galaxies. That is, if the morphology of the galaxies is linked to its redshift, and the distribution of clockwise and counterclockwise galaxies changes in different redshifts ranges, the morphology of the galaxies can become different.

Clearly, further research will be required to verify the observations and fully characterize and profile its nature. Future work will include the analysis of larger datasets, covering a larger footprint of the sky. Such analysis can allow to better profile the differences in morphology at different redshift ranges and different parts of the sky. Sky surveys such as the Dark Energy Survey (DES) and the Dark Energy Spectroscopic Instrument (DESI) Legacy Survey can provide such large datasets of galaxy images, and the Dark Energy Spectroscopic Instrument can provide the spectra of a high number of galaxies. Such data can allow to identify possible large-scale patterns exhibited by the possible asymmetry of the morphology of galaxies spinning in opposite directions.

**Funding:** The research was supported in part by NSF grants AST-1903823 and IIS-1546079.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The annotated SDSS galaxies are available at <https://people.cs.ksu.edu/~lshamir/data/assym> (accessed on 2 April 2022).

**Acknowledgments:** I would like to thank the two anonymous knowledgeable reviewers for the insightful comments and their help in improving the paper. The study was supported in part by NSF grant AST-1903823 and IIS-1546079.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Hubble, E.P. Extragalactic nebulae. *Astrophys. J.* **1926**, *64*, 321–369. [\[CrossRef\]](#)
2. Spitzer, L., Jr.; Baade, W. Stellar Populations and Collisions of Galaxies. *Astrophys. J.* **1951**, *113*, 413. [\[CrossRef\]](#)
3. Nairn, A.; Lahav, O. What is a peculiar galaxy? *Mon. Not. R. Astron. Soc.* **1997**, *286*, 969–978. [\[CrossRef\]](#)
4. De Vaucouleurs, G. Classification and morphology of external galaxies. In *Astrophysik IV: Sternsysteme/Astrophysics IV: Stellar Systems*; Springer: New York, NY, USA, 1959; pp. 275–310.
5. Baillard, A.; Bertin, E.; De Lapparent, V.; Fouqué, P.; Arnouts, S.; Mellier, Y.; Pelló, R.; Leborgne, J.F.; Prugniel, P.; Makarov, D.; et al. The EFIGI catalogue of 4458 nearby galaxies with detailed morphology. *Astron. Astrophys.* **2011**, *532*, A74. [\[CrossRef\]](#)
6. Malin, D.; Carter, D. A catalog of elliptical galaxies with shells. *Astrophys. J.* **1983**, *274*, 534–540. [\[CrossRef\]](#)
7. Arp, H.C.; Madore, B. *A Catalogue of Southern Peculiar Galaxies and Associations: Volume 1, Positions and Descriptions*; Cambridge University Press: Cambridge, UK, 1987; Volume 1.
8. Lintott, C.; Schawinski, K.; Bamford, S.; Slosar, A.; Land, K.; Thomas, D.; Edmondson, E.; Masters, K.; Nichol, R.C.; Raddick, M.J.; et al. Galaxy Zoo 1: Data release of morphological classifications for nearly 900,000 galaxies. *Mon. Not. R. Astron. Soc.* **2011**, *410*, 166–178. [\[CrossRef\]](#)
9. Willett, K.W.; Lintott, C.J.; Bamford, S.P.; Masters, K.L.; Simmons, B.D.; Casteels, K.R.; Edmondson, E.M.; Fortson, L.F.; Kaviraj, S.; Keel, W.C.; et al. Galaxy Zoo 2: Detailed morphological classifications for 304,122 galaxies from the Sloan Digital Sky Survey. *Mon. Not. R. Astron. Soc.* **2013**, *435*, 2835–2860. [\[CrossRef\]](#)
10. Holincheck, A.J.; Wallin, J.F.; Borne, K.; Fortson, L.; Lintott, C.; Smith, A.M.; Bamford, S.; Keel, W.C.; Parrish, M. Galaxy Zoo: Mergers-Dynamical Models of Interacting Galaxies. *Mon. Not. R. Astron. Soc.* **2016**, *459*, 720–745. [\[CrossRef\]](#)
11. Pović, M.; Aguerri, J.; Márquez, I.; Masegosa, J.; Husillos, C.; Molino, A.; Cristóbal-Hornillos, D.; Perea, J.; Benítez, N.; del Olmo, A.; et al. The ALHAMBRA survey: Reliable morphological catalogue of 22 051 early-and late-type galaxies. *Mon. Not. R. Astron. Soc.* **2013**, *435*, 3444–3461. [\[CrossRef\]](#)
12. Huertas-Company, M.; Gravet, R.; Cabrera-Vives, G.; Pérez-González, P.G.; Kartaltepe, J.; Barro, G.; Bernardi, M.; Mei, S.; Shankar, F.; Dimauro, P.; et al. A Catalog of Visual-like Morphologies in the 5 CANDELS Fields Using Deep Learning. *Astrophys. J. Suppl. Ser.* **2015**, *221*, 8. [\[CrossRef\]](#)
13. Kuminski, E.; Shamir, L. Computer-generated visual morphology catalog of ~3,000,000 SDSS galaxies. *Astrophys. J. Suppl. Ser.* **2016**, *223*, 20. [\[CrossRef\]](#)
14. Cheng, T.Y.; Conselice, C.J.; Aragón-Salamanca, A.; Agüena, M.; Allam, S.; Andrade Oliveira, F.; Annis, J.; Bluck, A.; Brooks, D.; Burke, D.; et al. Galaxy morphological classification catalogue of the Dark Energy Survey Year 3 data with convolutional neural networks. *Mon. Not. R. Astron. Soc.* **2021**, *507*, 4425–4444. [\[CrossRef\]](#)
15. Shamir, L. Asymmetry Between Galaxies with Clockwise Handedness and Counterclockwise Handedness. *Astrophys. J.* **2016**, *823*, 32. [\[CrossRef\]](#)
16. Shamir, L. Asymmetry between galaxies with different spin patterns: A comparison between COSMOS, SDSS, and Pan-STARRS. *Open Astron.* **2020**, *29*, 15–27. [\[CrossRef\]](#)
17. MacGillivray, H.; Dodd, R. The anisotropy of the spatial orientations of galaxies in the local supercluster. *Astron. Astrophys.* **1985**, *145*, 269–274.
18. Longo, M.J. Detection of a Dipole in the Handedness of Spiral Galaxies with Redshifts  $z \sim 0.04$ . *Phys. Lett.* **2011**, *699*, 224–229. [\[CrossRef\]](#)
19. Shamir, L. Handedness asymmetry of spiral galaxies with  $z < 0.3$  shows cosmic parity violation and a dipole axis. *Phys. Lett.* **2012**, *715*, 25–29.
20. Shamir, L. Patterns of galaxy spin directions in SDSS and Pan-STARRS show parity violation and multipoles. *Astrophys. Space Sci.* **2020**, *365*, 136. [\[CrossRef\]](#)
21. Shamir, L. Analysis of the alignment of non-random patterns of spin directions in populations of spiral galaxies. *Particles* **2021**, *4*, 2. [\[CrossRef\]](#)
22. Shamir, L. Ganalyzer: A tool for automatic galaxy image analysis. *Astrophys. J.* **2011**, *736*, 141. [\[CrossRef\]](#)
23. Land, K.; Slosar, A.; Lintott, C.; Andreescu, D.; Bamford, S.; Murray, P.; Nichol, R.; Raddick, M.J.; Schawinski, K.; Szalay, A.; et al. Galaxy Zoo: The large-scale spin statistics of spiral galaxies in the Sloan Digital Sky Survey. *Mon. Not. R. Astron. Soc.* **2008**, *388*, 1686–1692. [\[CrossRef\]](#)
24. Cervantes-Sodi, B.; Hernandez, X.; Park, C. Clues on the origin of galactic angular momentum from looking at galaxy pairs. *Mon. Not. R. Astron. Soc.* **2010**, *402*, 1807–1815. [\[CrossRef\]](#)



25. Dieleman, S.; Willett, K.W.; Dambre, J. Rotation-invariant convolutional neural networks for galaxy morphology prediction. *Mon. Not. R. Astron. Soc.* **2015**, *450*, 1441–1459. [\[CrossRef\]](#)
26. Cheng, T.Y.; Conselice, C.J.; Aragón-Salamanca, A.; Li, N.; Bluck, A.F.; Hartley, W.G.; Annis, J.; Brooks, D.; Doel, P.; García-Bellido, J.; et al. Optimizing automatic morphological classification of galaxies with machine learning and deep learning using Dark Energy Survey imaging. *Mon. Not. R. Astron. Soc.* **2020**, *493*, 4209–4228. [\[CrossRef\]](#)
27. González, R.E.; Munoz, R.P.; Hernández, C.A. Galaxy detection and identification using deep learning and data augmentation. *Astron. Comput.* **2018**, *25*, 103–109. [\[CrossRef\]](#)
28. Barchi, P.; de Carvalho, R.; Rosa, R.; Sautter, R.; Soares-Santos, M.; Marques, B.; Clua, E.; Gonçalves, T.; de Sá-Freitas, C.; Moura, T. Machine and Deep Learning applied to galaxy morphology—A comparative study. *Astron. Comput.* **2020**, *30*, 100334. [\[CrossRef\]](#)
29. Domínguez Sánchez, H.; Huertas-Company, M.; Bernardi, M.; Tuccillo, D.; Fischer, J. Improving galaxy morphologies for SDSS with deep learning. *Mon. Not. R. Astron. Soc.* **2018**, *476*, 3661–3676. [\[CrossRef\]](#)
30. Khan, A.; Huerta, E.; Wang, S.; Gruendl, R.; Jennings, E.; Zheng, H. Deep learning at scale for the construction of galaxy catalogs in the Dark Energy Survey. *Phys. Lett.* **2019**, *795*, 248–258. [\[CrossRef\]](#)
31. Lapuschkin, S.; Wäldchen, S.; Binder, A.; Montavon, G.; Samek, W.; Müller, K.R. Unmasking clever hans predictors and assessing what machines really learn. *Nat. Commun.* **2019**, *10*, 1096. [\[CrossRef\]](#)
32. Dhar, S.; Shamir, L. Systematic biases when using deep neural networks for annotating large catalogs of astronomical images. *Astron. Comput.* **2022**, *38*, 100545. [\[CrossRef\]](#)
33. Shamir, L.; Orlov, N.; Eckley, D.M.; Macura, T.; Johnston, J.; Goldberg, I.G. Wndchrm—An open source utility for biological image analysis. *Source Code Biol. Med.* **2008**, *3*, 13. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Shamir, L. Automatic morphological classification of galaxy images. *Mon. Not. R. Astron. Soc.* **2009**, *399*, 1367–1372. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Tamura, H.; Mori, S.; Yamawaki, T. Textural features corresponding to visual perception. *IEEE Trans. Syst. Man Cybern.* **1978**, *8*, 460–473. [\[CrossRef\]](#)
36. Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *6*, 610–621. [\[CrossRef\]](#)
37. Prewitt, J.M. Object enhancement and extraction. *Pict. Process. Psychopictorics* **1970**, *10*, 15–19.
38. Wu, C.M.; Chen, Y.C.; Hsieh, K.S. Texture features for classification of ultrasonic liver images. *IEEE Trans. Med. Imaging* **1992**, *11*, 141–152.
39. Hadjidemetriou, E.; Grossberg, M.D.; Nayar, S.K. Spatial information in multiresolution histograms. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA, 8–14 December 2001; IEEE: Kauai, HI, USA, 2001; Volume 1, p. 7176925.
40. Gradshteyn, I.; Ryzhik, I. *Table of Integrals, Series, and Products*; Translated from the Fourth Russian Edition; Jeffrey, A., Ed.; Academic Press: New York, NY, USA, 1994.
41. Teague, M.R. Image analysis via the general theory of moments. *J. Opt. Soc. Am.* **1980**, *70*, 920–930. [\[CrossRef\]](#)
42. Lim, J.S. *Two-Dimensional Signal and Image Processing*; Prentice Hall: Englewood Cliffs, NJ, USA, 1990; Volume 1, 710p.
43. Gabor, D. Theory of communication. Part 1: The analysis of information. *Electr. Eng. Part III* **1946**, *93*, 429–441. [\[CrossRef\]](#)
44. Bishop, C. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2007.
45. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
46. Kohavi, R. The power of decision tables. In *Machine Learning: ECML-95*; Springer: New York, NY, USA, 1995; pp. 174–189.
47. Lewis, D.D. Naive (Bayes) at forty: The independence assumption in information retrieval. In *Machine Learning*; Springer: New York, NY, USA, 1998; pp. 4–15.
48. Ting, K.M.; Witten, I.H. Stacking bagged and dagged models. In Proceedings of the International Conference on Machine Learning, San Francisco, CA, USA, 8–12 July 1997; Citeseer: San Francisco, CA, USA, 1997; pp. 367–375.
49. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [\[CrossRef\]](#)
50. Holte, R.C. Very simple classification rules perform well on most commonly used datasets. *Mach. Learn.* **1993**, *11*, 63–90. [\[CrossRef\]](#)
51. Moody, J.; Darken, C.J. Fast learning in networks of locally-tuned processing units. *Neural Comput.* **1989**, *1*, 281–294. [\[CrossRef\]](#)
52. Witten, I.H.; Frank, E.; Trigg, L.; Hall, M.; Holmes, G.; Cunningham, S.J. *Weka: Practical Machine Learning Tools and Techniques with JAVA Implementations*; Department of Computer Science, University of Waikato: Hamilton, New Zealand, 1999.
53. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The WEKA data mining software: An update. *Acm Sigkdd Explor. Newsl.* **2009**, *11*, 10–18. [\[CrossRef\]](#)
54. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [\[CrossRef\]](#)
55. Rubin, V.C. The rotation of spiral galaxies. *Science* **1983**, *220*, 1339–1344. [\[CrossRef\]](#)
56. Milgrom, M. A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis. *Astrophys. J.* **1983**, *270*, 365–370. [\[CrossRef\]](#)
57. Sivaram, C.; Arun, K.; Rebecca, L. MOND, MONG, MORG as alternatives to dark matter and dark energy, and consequences for cosmic structures. *J. Astrophys. Astron.* **2020**, *41*, 1–6. [\[CrossRef\]](#)

- 
58. Sivaram, C.; Arun, K.; Prasad, A.; Rebecca, L. Non-detection of Dark Matter particles: A case for alternate theories of gravity. *J. High Energy Phys. Gravit. Cosmol.* **2021**, *7*, 680. [[CrossRef](#)]
  59. Sivaram, C.; Arun, K.; Rebecca, L. The Hubble tension: Change in dark energy or a case for modified gravity? *Indian J. Phys.* **2021**, 1–4. [[CrossRef](#)]
  60. Neeleman, M.; Prochaska, J.X.; Kanekar, N.; Rafelski, M. A cold, massive, rotating disk galaxy 1.5 billion years after the Big Bang. *Nature* **2020**, *581*, 269–272. [[CrossRef](#)]
  61. Sivaram, C.; Arun, K. Primordial rotation of the universe, hydrodynamics, vortices and angular momenta of celestial objects. *Open Astron.* **2012**, *5*, 7–11. [[CrossRef](#)]
  62. Calvi, R.; Poggianti, B.M.; Fasano, G.; Vulcani, B. The distribution of galaxy morphological types and the morphology–mass relation in different environments at low redshift. *Mon. Not. R. Astron. Soc. Lett.* **2012**, *419*, L14–L18. [[CrossRef](#)]
  63. Soo, J.Y.; Moraes, B.; Joachimi, B.; Hartley, W.; Lahav, O.; Charbonnier, A.; Makler, M.; Pereira, M.E.; Comparat, J.; Erben, T.; et al. Morpho-z: Improving photometric redshifts with galaxy morphology. *Mon. Not. R. Astron. Soc.* **2018**, *475*, 3613–3632. [[CrossRef](#)]