Using Group-Equivariant Deep Learning to Determine the Chirality of Spiral Galaxies

Neil Power
Department of Physics and Astronomy
University of Manchester
Student ID: 10623021
neil.power-2@student.manchester.ac.uk

This project was performed in collaboration with Ezzy Cross,
and supervised by Prof. Anna Scaife.

Abstract

The spin direction of clockwise and anticlockwise spiral galaxies results from the initial conditions during their formation, and can therefore provide information about the early Universe. Volunteers classified the morphology of over 900,000 galaxies through the Galaxy Zoo 1 project, and these labels were used in combination with images from the Dark Energy Spectroscopic Instrument Survey (DESI) to train convolutional neural networks (CNNs). CNNs are inherently equivariant to image translation, and group-equivariant convolutional neural networks (G-CNNs) can extend this equivariance to other transformations such as rotation, improving their ability to identify images in different orientations. Several CNNs and G-CNNS were used, and models based on the ResNet architecture performed well, with either rotation-equivariance or a customised 'chirality equivariant' prediction function improving performance over default ResNet models. 'G_ResNet18_c', a model including both rotation equivariance and chirality equivariance achieved the lowest test loss at 0.3766 $\pm$ 0.0045, and did not find any significant chirality violation at (-0.48 $\pm$ 0.38)σ when classifying 8.7 million galaxies from the DESI survey.

Contents

1 Introduction **3**

2 Theoretical Background **3**
2.1 Galaxy Chirality . . . 3
2.2 Convolutional Neural Networks . . . 5
2.3 Group Equivariant Neural Networks . . . 6

3 Data Sources and Processing **7**
3.1 Galaxy Zoo 1 & Sloan Digital Sky Survey . . . 7
3.2 Dark Energy Spectroscopic Instrument Survey . . . 8
3.3 Data Pre-Processing . . . 8

4 Models **9**
4.1 LeNet . . . 9
4.2 ResNet . . . 10
4.3 General Implementation of Models . . . 12
4.4 Measuring Performance . . . 13

5 Results & Discussion **13**
5.1 Training and Testing Results . . . 13
5.2 Reflection Equivariance . . . 15
5.3 Effects of Rotation . . . 15
5.4 Representation of Chirality Violation . . . 17
5.5 DESI Dataset Predictions . . . 18

6 Conclusion **19**
6.1 Future Work . . . 20

General Notes **20**

References **20**

Appendix **23**

1 Introduction

Spiral galaxies can be either have clockwise spin (leading to a 'Z' shape) or anticlockwise spin (leading to an 'S' shape). The direction of this spin is dependent on the position of an observer, as a clockwise galaxy will appear anticlockwise if viewed from the opposite side. Pairing the spin direction of a galaxy with its recession velocity makes spiral galaxies chiral objects, allowing an observer-independent distinction between S and Z galaxies. Many studies have investigated the balance between the numbers of S and Z galaxies, as their spin direction may correlate with the density field of the early Universe [1]. Under the widely-accepted cosmological principle [2], the distribution of matter and this density field is thought to be isotropic, but any significant asymmetry in spin directions may provide evidence against this.

Machine learning tools are increasingly used in astrophysics, especially for large datasets such as deep sky surveys that may consist of images of millions of objects, beyond human classification. Convolutional neural networks (CNNs) are models designed for processing images, identifying features automatically and using these to create predictions of what these images represent. CNNs are translation equivariant, where that the placement of a galaxy within an image does not need to be identical for each image. However, they are not typically equivariant to other image transformations, such as rotation or reflection. As a result, CNNs need to be trained on multiple orientations of images to ensure they can accurately classify galaxy images of any orientation. Group equivariant neural networks (G-CNNs) can be made inherently rotation equivariant, and may therefore perform better at classifying galaxies at different orientations.

The Galaxy Zoo 1 project [3] ran from 2007 to 2009 and consisted of volunteers classifying images of galaxies from the Sloan Digital Sky Survey (SDSS) [4] based on the galaxy's shape, with categories such as 'clockwise spiral' or 'elliptical'. Over 900,000 galaxies were classified, and provide an excellent labelled training set for machine learning. If a model is trained on these, it can then create predictions for much larger datasets such as the Dark Energy Spectroscopic Instrument (DESI) survey [5]. The DESI survey contains over 8 million galaxies, an excellent sample to investigate if there is any significant spiral galaxy chirality violation observed. The key objectives of the project in order to achieve this are shown in Figure 1.

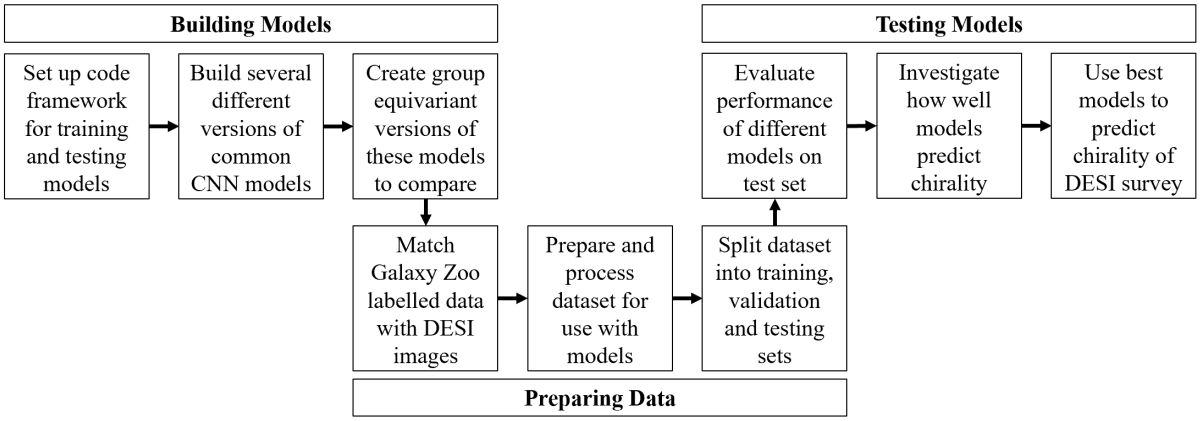


Figure 1: Objectives for this project.

2 Theoretical Background

2.1 Galaxy Chirality

Spiral galaxies are characterised by their spiral arms, extending outwards from their centre. The shape and number of these arms can vary significantly, from two large arms closely wrapped around the centre of the galaxy (a Hubble SBa type) to many thin loosely wrapped spiral arms

(a Hubble Sc type). The formation of spiral galaxies has been studied in detail by astronomers, and there are currently two leading theories for their formation. In the density wave theory, denser spiral arms rotate more slowly than the rest of the galaxy [6], creating a spiral structure. In the stochastic self-propagating star formation (SSPSF) model, objects such as supernovae produce shockwaves that induce star formation, organising into spirals as the galaxy rotates [7].

The vast majority of galaxies have trailing spiral arms, where the arms trail outwards, and so the rotation can be identified from the spiral direction of the arms. Whilst some galaxies may have leading spiral arms where the spiral is in the opposite direction to the galaxy's rotation, these are very rare [8]. The handedness of a galaxy's rotation (either clockwise or anticlockwise) is dependent on the location of the observer, but including the recession velocity from Earth allows astronomers to assign an objective chirality to galaxies. Figure 2 shows typical clockwise and anticlockwise galaxies, using images taken from the Dark Energy Spectroscopic Instrument (DESI) survey [5].

(a) Clockwise (Z) (b) Anticlockwise (S)

Figure 2: Examples from the DESI survey of a clockwise (OBJID 309464_4368) and anticlockwise (OBJID 305694_123) spiral galaxy.

The origin of a galaxy's initial rotation can be described by tidal torque theory. In this, the quadrupole moment of a protogalaxy (the initial collection of matter that forms a galaxy) interacts with the tidal fields of other nearby protogalaxies, transferring angular momentum [9]. Motloch et al. [1] used classifications from the Galaxy Zoo survey in order to find a tentative correlation (2.7σ) between the current spin direction of galaxies and reconstructions of the initial early Universe density field during their formation. 15,000 galaxies were used, with the study concluding that more labelled galaxies may improve the significance of the correlation and that this is a problem 'well suited' to machine learning.

Following the cosmological principle, the Universe is expected to be isotropic at a large enough scale, as the forces that govern the distribution of matter should themselves be both isotropic and homogeneous at large scales. This principle is central to many cosmological models, such as the Lambda-CDM model [2], one of the most commonly-used models describing dark energy and dark matter. Any evidence that the cosmological principle is incorrect, such as significant chirality violations at a cosmological scale, may suggest that a different model may be required to explain the early Universe. Chiral objects, such as spiral galaxies when their recession velocity is accounted for [10], can be used to investigate the isotropy of the Universe at extremely large scales and test for significant violations.

With modern large sky surveys creating increasingly larger datasets of astronomical objects, the chiral symmetry of spiral galaxies has been investigated many times [11]. Several studies found no overall asymmetry [12, 13, 14], whereas others have found asymmetry at varying degrees of significance [15, 16, 17]. Central to many of these works are both the Sloan Digital Sky Survey (SDSS) [4], providing one of the first large-scale sky surveys, and the Galaxy Zoo project [3], labelling galaxy morphology for nearly a million galaxies identified in SDSS images.

When investigating chirality violation, it is useful to calculate the significance of any violation relative to the size of the dataset. Equation 1 shows one way of calculating this. This equation results from assuming a distribution with no net chirality violation, calculating its variance and then using the central limit theorem to define significance [18]. Positive chirality violation significance (CVS) is defined as having more ACW galaxies than CW galaxies:

$$CVS = \frac{n_{ACW} - n_{CW}}{\sqrt{n_{ACW} + n_{CW}}}. \tag{1}$$

More recently, machine learning tools such as neural networks have been used to investigate chirality violations for spiral galaxies. These allow researchers to measure the number of galaxies in much larger datasets, such as the DESI [5] survey, which contains over 8 million objects, which would be impossibly time-intensive to classify manually. Galaxy Zoo 1 and subsequent data releases have been used as training data for several models [19, 18], as they provide one of the largest datasets of labelled galaxies.

One of the most recent studies involving building models for galaxy chirality classification was in 2023, where Jia, Zhu, and Pen [18] created a neural network based on the ResNet architecture. Their aim was to make a chirality equivariant model, such that even if it was trained on a biased dataset (with more of one type of galaxy correctly labelled than the other), that bias would not be trained into the model. When the model is used to make predictions on an unknown dataset, it should accurately return the difference in chirality for that dataset. This chirality equivariance was achieved by training the model with flipped and un-flipped images to ensure that the model selected clockwise and anticlockwise spirals using the same method.

This project aims to build on the work of Jia, Zhu, and Pen [18] by comparing the performance of a number of different machine learning models when predicting spiral galaxy spin, including their chirality equivariant model and models designed to be equivariant to galaxies at different rotations. All of the models used were convolutional neural networks.

2.2 Convolutional Neural Networks

A neural network (NN) is a type of machine learning model that passes data through a network consisting of layers of artificial 'neurons' or nodes, each of which uses non-linear mathematical functions to determine how to pass on information to the next neuron [20]. As the network is trained, the weights and biases of each node are updated repeatedly to fit the model to the data. Each piece of data passes through the network, reaching an 'output' node in the final layer of the network that represents the model's prediction for that data. At each training stage, known as an epoch, a loss function is used to calculate the difference between the true data and the model's predictions, and then optimisation algorithms are used to minimise this loss [21].

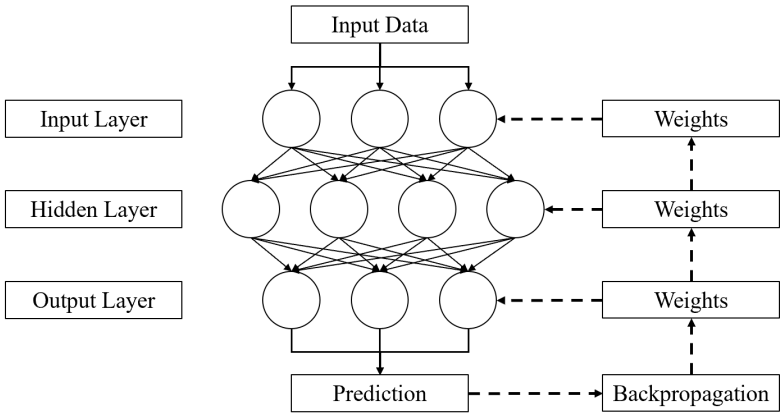


Figure 3: Schematic of a simple three layer neural network, showing neurons (circles) and the connections between them (arrows). The training loop is shown using dashed arrows.

Data used to train a network is typically split into three subsets: training, used in batches to train the network, validation, which the model uses after training on each batch to assess performance, and testing, used to test overall performance after the model has been fully trained. The progress made when training a model is measured in epochs, where one epoch is one complete sweep through all the batches in the training subset. Figure 3 shows the structure of a very simple three layer neural network, consisting of an initial input layer, a central (hidden) layer and an output layer. Neural network structures can vary significantly and can be tuned to a specific problem as very generally more layers allow a model to learn more complex relationships. The training loop (backpropagation) is also shown, where information about the model's performance is propagated back through the model and used to adjust the weights of components in the layers.

Activation functions are applied to the data at each node. These are typically non-linear, as introducing non-linearity allows the model to learn more complicated systems. The rectified linear unit (ReLU) function is a commonly used activation function [22], which returns its original input if the input is positive, and zero otherwise. The learning rate of a network determines how much these weights are varied in each training cycle - too high and the model will skip minima when fitting, but too low and the model will not learn at all. Weight decay is commonly used to prevent overfitting, which adds a penalty to the complexity of the weights. A common optimiser algorithm used for backpropagation is 'Adam' [23], an extension to stochastic gradient descent, which approximates the gradient needed to minimise the loss function. `Adam` improves on stochastic gradient descent by changing and updating the learning rate for each individual weight, rather than using one single learning rate for the entire model. An optimiser's learning rate can be set by a scheduler, which will adjust the rate over time. There are also a number of loss functions that can be used to measure how well a neural network performs, including cross-entropy loss [24].

Convolutional neural networks (CNNs) are a specific type of neural network, designed to process images. Unlike numerical datasets (such as a table or database with columns for each feature of the data), images do not have predefined features for a model to use. This means that the user must either select specific areas of the image for the module to use (time-consuming and difficult to generalise), or the model must be able to identify the features without user input. CNNs do this by convolving a kernel across the image, building up a map of the features of an image [25]. Kernels are transformation matrices of different parameters that identify features (such as object edges) when they are convolved with an image, also represented as a matrix or tensor. The kernel moves across the image according to a set 'stride', which determines how big a step the kernel will take each time. When the kernel passes over the map, it reduces the size of the map, and so padding can be added before convolution to preserve the original dimensions. As with a traditional neural network, CNNs contain an initial input layer, hidden layers (where at least one will be a convolutional layer), and a final output layer.

2.3 Group Equivariant Neural Networks

CNNs are translation-equivariant, such that identical objects in different shifted positions within an image should produce identical results. This is a result of the way that weights are distributed in CNNs - in order to minimise the number of parameters needed, weights are shared across different positions as the kernel is applied across the image [26]. However, they are not equivariant to other image transformations, such as reflection or rotation. For example, CNNs will consider an image rotated by 90° as a different image to the original. For datasets where subjects will be in a variety of orientations within an image, the networks must be trained with repeated images in different orientations. This 'learned equivariance' mean that CNNs may struggle to generalise compared to a network that is inherently equivariant [27].

Group-equivariant convolutional neural networks (G-CNNs) [28] are one solution to this generalisation problem. The groups used in G-CNNs are symmetry groups, consisting of a set

of distinct transformations acting on a vector space [29]. Symmetry groups acting on a space transform objects in that space, whilst preserving one or more properties of the object (for example, rotation preserves shape and size). G-CNNs can be made equivariant to a particular group or groups, and therefore a G-CNN equivariant to a group including rotation would be able to recognise rotated versions of objects, despite only being trained with objects in one orientation.

Group equivariance is achieved by using group-equivariant convolutions in the layers, where the map of features identified by the network will transform in the same way as the original image. The kernel in a group-equivariant convolution consists of a stack of conventional kernels representing elements of the transform group. When this kernel is convolved with an input, each kernel in the stack convolves, creating a multi-dimensional feature map. Increasing the number of transformations that a G-CNN is equivariant to can significantly increase computational cost, as at each convolution the output size depends on both the size of the input image or feature map, and the size of the kernel [29].

An advantage of G-CNNs is that similar network architectures to conventional CNNs can be used, meaning the layer types can often be switched from conventional to group-equivariant without needing to design a new structure. In a steerable G-CNN, the set of transformations that the kernel applies in convolution can be adjusted, enabling the equivariance of the network to be precisely controlled. One common choice is E(2)-equivariant CNNs, equivariant to the 2D Euclidian E(2) group of transformations, which encompasses translation, reflection and rotation.

Whilst there have been several uses of CNNs for galaxy chirality and morphology classification [19, 18, 30, 31], including models trained to be rotation-equivariant [32], steerable G-CNNs have not been widely used and may result in better performance.

3 Data Sources and Processing

3.1 Galaxy Zoo 1 & Sloan Digital Sky Survey

Galaxy Zoo is a decade-long citizen science project using the Zooniverse framework, where volunteers classify images of galaxies from various sky surveys. The Galaxy Zoo 1 (GZ1) project ran from 2007 to 2009 [3], where galaxies were sorted into six categories, including either clockwise spiral or anticlockwise spiral. Volunteers were presented with images from Data Release 7 of the Sloan Digital Sky Survey (SDSS) [4], taken using the Apache Point Observatory 2.5m telescope in New Mexico. Data Release 7 was part of the second phase of SDSS and mapped approximately a quarter of the sky over eight years, covering 350 million celestial objects.

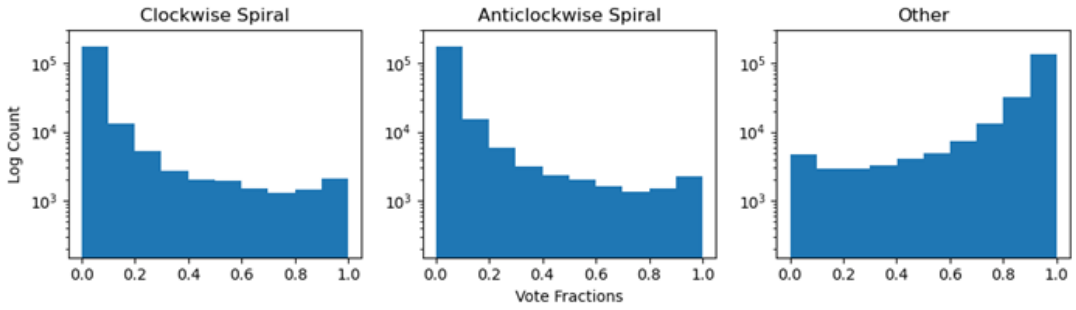


Figure 4: Histograms of the votes for each morphology class from Galaxy Zoo volunteers.

Over 900,000 galaxies were classified [33], with vote fraction scores assigned to six labels based on volunteers' confidence. These labels were elliptical, clockwise spiral, anticlockwise spiral, edge-on, merger and don't know. For this project, the anticlockwise spiral vote fractions

were labelled as 'ACW', the clockwise were labelled as 'CW', and the remaining categories combined into 'Other', signifying non-spiral galaxies or uncertain classifications. Figure 4 shows the histograms of vote fractions for each class. Each image typically has between 20 to 60 votes, with the majority of galaxies in the dataset labelled confidently as non-spirals. A galaxy is classified as CW or ACW if the vote fraction for either class was greater than 0.5, and Other otherwise. There is a slight bias towards anti-clockwise spiral galaxies in the GZ1 dataset, at 32189 (4.8%) compared to 30356 (4.5%) clockwise spiral galaxies at a threshold of 0.5. This has been attributed to human bias, as the 'selection function' used by humans favours spotting S shapes over Z shapes, and once this bias is accounted for and removed, there is not a significant violation of chirality in the GZ1 dataset [13, 14].

3.2 Dark Energy Spectroscopic Instrument Survey

The Dark Energy Spectroscopic Instrument (DESI) survey is a combination of several sky surveys [5]. Data Release 8 (DR8) includes images from several smaller surveys: Dark Energy Camera Legacy Survey (DECaLS), Beijing-Arizona Sky Survey (BASS) and the Mayall z-band Legacy Survey (MzLS), taken from 2015 to 2019. DESI DR8 spans a considerably larger area than SDSS DR7, covering 14,000 square degrees and containing more objects imaged at a higher resolution. 8.7 million galaxies were identified in the DESI survey [34]. The majority of the galaxies in the SDSS survey also appear in DESI, and so labels for galaxies classified in GZ1 can be matched with DESI images of the same galaxies.

3.3 Data Pre-Processing

Figure 5 shows a summary of the data pre-processing steps used, with the number of CW, ACW and other galaxies at each stage shown in Table 1.

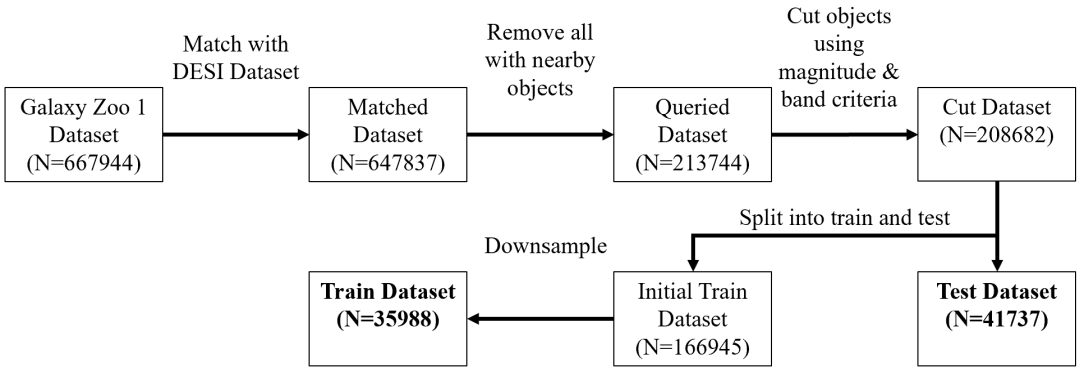


Figure 5: Steps taken to create testing and training datasets from the Galaxy Zoo 1 dataset, showing the number of galaxies selected at each stage.

Table 1: Datasets

Dataset	Total	ACW	CW	Other	CVS
Galaxy Zoo 1	667944	32189 (4.8%)	30356 (4.5%)	605399 (90.6%)	7.33
Matched	647837	31660 (4.9%)	29872 (4.6%)	586305 (90.5%)	7.21
Queried	213744	8711 (4.1%)	8160 (3.8%)	196873 (92.1%)	4.24
Cut	208682	8549 (4.1%)	8003 (3.8%)	192130 (92.1%)	4.24
Test	41737	1753 (4.2%)	1580 (3.8%)	38404 (92.0%)	3.00
Initial Train	166945	6796 (4.1%)	6423 (3.8%)	153726 (92.1%)	3.24
Train	35988	6423 (17.8%)	6796 (18.9%)	22769 (63.3%)	3.24

Firstly, galaxy classifications from the GZ1 catalogue were matched with images from the DESI catalogue in order to create a dataset of GZ1-classified galaxies with DESI images. Python's `astropy` [35] module was used to match galaxies from GZ1 to DESI based on their coordinates, using a match radius of 10 arcseconds. Of the 667,944 objects in the GZ1 catalogue, 647,837 were matched successfully.

The cuts used by Jia, Zhu, and Pen [18] were then performed on this matched dataset. Firstly, to remove images containing several clustered astronomical objects, where models may struggle to identify which is the galaxy to be classified, all galaxies with another object located within 1 arcsecond were removed. The `astroquery` package [36] was used to query the SDSS DR7 database, counting the number of objects within a 1 arcsecond radius. This reduced the number of galaxies in the dataset to 213,744. Then, several cuts were made to remove dim objects or objects with a high magnitude error. Galaxies that did not meet the following criteria were removed, leaving 208,682 objects.

- The error in magnitude of the R (red) band must be between 0 and 1.
- The galaxy half-light radius in R-band must be greater than 1 arcsecond.
- The relative error of the R-band galaxy half-light radius must be between 0 and 0.25.

After performing these cuts, the data was split into training and testing datasets, with the training dataset further split into training and validation sets. As the overall dataset is very imbalanced (approximately 4% ACW, 4% CW and 92% Other), a train-test split of 80%-20% was used to ensure there were at least 1500 galaxies of each spiral type in the testing dataset. The testing dataset contained 41,737 galaxies. Performing the train-test split before downsampling ensures that the test data has a balance similar to any datasets the model may be used to predict.

However, many CNN models may struggle to learn on a very unbalanced dataset. To account for this, the training/validation dataset was downsampled. Where P_{CW} is the fraction of voters who described the galaxy as clockwise, and P_{ACW} the proportion for anticlockwise, galaxies were downsampled as shown below:

- Keep 1 in 20 galaxies with $0 < \max(P_{CW}, P_{ACW}) \leq 0.1$
- Keep 1 in 5 galaxies with $0.1 < \max(P_{CW}, P_{ACW}) \leq 0.2$
- Keep 1 in 2 galaxies with $0.2 < \max(P_{CW}, P_{ACW}) \leq 0.3$
- Keep all other galaxies

The final training/validation dataset contains 35,988 galaxies. The distribution of CW, ACW and Other is 19%, 20% and 61% respectively. This is significantly more balanced and should avoid models learning to classify all galaxies as other to achieve high accuracies.

4 Models

4.1 LeNet

Developed during the 1990s, LeNet [37] was one of the first convolutional neural networks. The architecture of LeNet has seven layers: an input layer, two convolutional layers, three fully-connected layers and an output layer, as shown in Table 2.

An implementation of LeNet by Scaife and Porter [38] using the `PyTorch` library [39] was used. LeNet can be made into a steerable G-CNN by replacing the convolutional layers with group-equivariant layers. The `escnn` [40] library is an extension to `PyTorch` that contains group-equivariant versions of many common neural network layers. Similarly to the conventional LeNet, the implementation of a steerable 'G_LeNet' by Scaife and Porter [38] was used, with

some minor modifications to update the code to use the newer `escnn` [40] rather than the now-depreciated `e2cnn` library [41].

Layers used in LeNet and G_LeNet include:

- Input Layer - this reshapes the image from X pixels $\times$ Y pixels $\times$ N channels into a 1D array for processing.
- Convolution - this extracts features from the image data by sweeping a kernel across the data, building filters that identify features.
- G-Convolution - a convolutional layer where the kernel uses transformations that are equivariant to specific groups.
- ReLU (Rectified Linear Unit) function - this is a non-linear activation function used to transform a layer [22].
- Max-pool - a pooling layer that downsamples input data, selecting the maximum value of a small region instead of the entire region.
- Average pool - similar to max-pooling, except the average of a region is used.
- Global average pool - this is a more extreme version of average pooling, where the entire input data is pooled.
- Fully connected layer - a layer where every node is connected to every other node
- Dropout layer - this randomly removes the output of some nodes, preventing overfitting
- Softmax - this scales the outputs into probabilities for each class, summing to 1.

Table 2: LeNet and G_LeNet architectures

LeNet	**G_LeNet**	**Shape**	**Output Size**
Input Layer	Input Layer		160x160x3
Convolution	G-Convolution	5x5, padding 1	158x158x6
ReLU	ReLU		
Max-pool	Max-pool	2x2	79x79x6
Convolution	G-Convolution	5x5, padding 1	77x77x16
ReLU	ReLU		
Max-pool	Max-pool	2x2	38x38x16
	Global Average Pool		
Fully-connected	Fully-connected		120x1x1
ReLU	ReLU		
Fully-connected	Fully-connected		84x1x1
ReLU	ReLU		
Dropout (p = 0.5)	Dropout (p = 0.5)		84x1x1
Fully-connected	Fully-connected		3
Softmax	Softmax		3

4.2 ResNet

Residual neural networks (ResNet) are a family of neural networks first proposed by He et al. [42] that use residual functions in each layer when training the network. This allows ResNet networks to contain more layers than other CNN types without significantly increasing complexity.

A 18 layer network and a 50 layer network were used, labelled ResNet18 and ResNet50 respectively. Both versions were implemented using untrained, prebuilt models from the `PyTorch` Python library [39]. The architectures of ResNet18 and ResNet50 are shown in Tables 3 and 4. Unlike LeNet, ResNet contains blocks of convolutional layers. ResNet18 contains 'BasicBlock' blocks, containing two sets of a batch normalisation (BN) layer, ReLU activation function and

convolutional layer (CL) with a 3×3 kernel as shown in Equation 2. Batch normalisation rescales the outputs of a layer to a mean of zero and standard deviation of 1. Resnet50 uses 'Bottleneck' blocks, containing three sets of BN, ReLU, and CL, as shown in Equation 3. These blocks are indicated by square brackets in Tables 3 and 4.

$$\text{BasicBlock: } \begin{bmatrix} \text{Batch Normalisation, ReLU, } 2 \times 2 \text{ Convolution} \\ \text{Batch Normalisation, ReLU, } 2 \times 2 \text{ Convolution} \end{bmatrix} \tag{2}$$

$$\text{Bottleneck: } \begin{bmatrix} \text{Batch Normalisation, ReLU, } 1 \times 1 \text{ Convolution} \\ \text{Batch Normalisation, ReLU, } 3 \times 3 \text{ Convolution} \\ \text{Batch Normalisation, ReLU, } 1 \times 1 \text{ Convolution} \end{bmatrix} \tag{3}$$

As well as these default ResNet models, a general group-equivariant ResNet (G_ResNet) model was created using the `escnn` library. This model followed the same central structure as non-equivariant ResNet, but used custom versions of the BasicBlock and Bottleneck blocks built using implementations of layers from the `escnn` library, with the model updated to use geometric tensors and field types. The model was created to be equivariant to `c8` transformations, which include translation and 8 discrete rotations (45° rotations about any point on an image), with the option to extend this to include reflection equivariance. Similar to the default ResNet, 18 and 50 layer version can then be initialised from this. The differences between ResNet18 and G_ResNet18 are outlined in Table 3. The standard convolutional layers were replaced by group-equivariant convolutional layers from the `escnn` library [40].

Table 3: 18-layer ResNet model architectures

Resnet18	**G_ResNet18**	**Shape**	**Output Size**
Input Layer	Input Layer		160x160x3
Convolution 1	G-Convolution 1	7x7,64	80x80x64
Max-pool	Max-pool	3	40x40x64
Convolution 2	G-Convolution 2	$2 \times \begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix}$	40x40x64
Convolution 3	G-Convolution 3	$2 \times \begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix}$	20x20x128
Convolution 4	G-Convolution 4	$2 \times \begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix}$	10x10x256
Convolution 5	G-Convolution 5	$2 \times \begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix}$	5x5x512
Average Pool	Average Pool		1x1x512
Fully connected	Fully connected	3	3
Softmax	Softmax		3

A 'chirality-equivariant' (CE) ResNet50 model following the structure set out in Jia, Zhu, and Pen [18] was also built, labelled CE_Resnet50. This model aims to reduce any issues that the biased GZ1 dataset may cause by training the model on each image twice, once in with the original image and once with the image mirrored. Figure 6 shows the method used by this model to make predictions. Instead of classifying images into 3 labels: CE, ACW and 'Other', the model classifies each image into CW or Other for the original, and ACW or Other for the flipped. The CW and ACW probabilities are then combined with the average of the Other probabilities and then softmaxed. As well as a modified predict function that implements this flipping, CE_Resnet50 also adds four more fully-connected layers, as shown in Table 4. contrasting ResNet50, G_ResNet50 and CE_Resnet50.

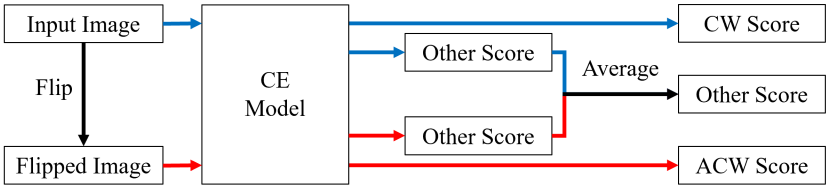


Figure 6: Chirality-equivariant prediction function

Table 4: 50-layer ResNet model architectures

Resnet50	**G_ResNet50**	**CE_Resnet50**	**Shape**	**Output Size**
Input Layer	Input Layer	Input Layer		160x160x3
Convolution	G-Convolution	Convolution	7x7,64	80x80x64
Max-pool	Max-pool	Max-pool	3x3	40x40x64
Convolution	G-Convolution	Convolution	$3 \times \begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix}$	40x40x256
Convolution	G-Convolution	Convolution	$4 \times \begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix}$	20x20x512
Convolution	G-Convolution	Convolution	$6 \times \begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix}$	10x10x1024
Convolution	G-Convolution	Convolution	$3 \times \begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix}$	5x5x2048
Average Pool	Average Pool	Average Pool	1x1	1x1x2048
Fully connected	Fully connected		3	3x1x1
		Fully Connected	$2 \times [512]$	512x1x1
		Fully Connected	$2 \times [64]$	64x1x1
		Fully Connected	2	2x1x1
Default Predictor	Default Predictor	Flipped Predictor		3
Softmax	Softmax	Softmax		3

4.3 General Implementation of Models

Models were built in Python using the `PyTorch` machine learning library [39]. Developed by Walmsley [43], the `galaxy-datasets` dataloader was used to import, rescale, rotate and crop the image files. When loading in an image, each pixel value was rescaled from 0-255 to 0-1. Images were then randomly rotated between 0 and 360° in order to increase the resilience of each model to image rotation. Finally, each image was resized from 424x424 to 160x160 to convert from a angled square into a square aligned with the x and y axes, ensuring no padding was added from the rotation.

Each CNN used the AdamW [23] optimiser to train the network weights. Cross-entropy loss was used as a loss function, providing a metric of how well the neural network is predicting data labels compared to their true value. A simple step scheduler, `StepLR` [39] was used to reduce the learning rate by a set value of 0.85 every set number of steps.

All machine learning models have hyperparameters, inherent properties of the model that can be varied to determine how the model learns. The same hyperparameters were used in all models, as listed in Table 5. Models used a random seeded train-test split to ensure that any differences in results were due to the models themselves rather than the selection of data used in training. For training, the data was randomly shuffled into batches.

Table 5: Model hyperparameters

Parameter	Description	Value
Batch size	Number of images the model uses per training batch	100
Learning rate	How much the model adjusts itself by after calculating loss	0.0001
Number of epochs	Number of epochs to train the model for	100
Step size	Number of steps to update the learning rate	5
Gamma	Rate to adjust the learning rate by per step size	0.85

4.4 Measuring Performance

There are a number of ways to measure the performance of a neural network. The loss (cross-entropy loss) is a common metric when training a model that predicts labels with a probability value, where a lower loss suggests the model fits better. Accuracy is also a useful metric, but as both the true and predicted labels are probabilities, a direct equality comparison is not possible. Instead, the class with the maximum probability from both true and predicted labels can be compared to calculate accuracy. Additionally, the chirality violation of the predictions was calculated using Equation 1, and then compared to the actual chirality violation of the testing and training datasets.

5 Results & Discussion

The models set out in Section 4 allow a number of structural properties to be changed, and then their effect on a model's overall performance investigated. A number of models were built encompassing different combinations of the properties below. Equivariant models are labelled with a 'G_' prefix, whereas a 'CE' prefix or '_c' suffix indicates the model used a 2-class CE predict function.

- Model Architecture: LeNet or ResNet
- ResNet version: 18 layers or 50 layers
- Training augmentation: randomly rotate images or keep original rotation
- Rotation Equivariance: G-CNN or CNN
- Classification: 2-class CE predict function or 3-class predict function

5.1 Training and Testing Results

To measure the average metrics and metric uncertainties for each model, five runs of each model were trained on the GZ1 training dataset. The validation accuracies during training are shown for the first runs of selected models in Figure 7. A clear distinction is seen in training between ResNet and LeNet-based models, with the latter failing to significantly decrease loss and increase accuracy whilst training. After no significant improvement, training for the LeNet models was stopped after 60 epochs, whereas ResNet models were trained for 120 epochs, with the model saved after the last epoch as losses stabilised. The performances of ResNet18, ResNet50 and CE-ResNet50 are broadly similar, plateauing at around 60 epochs. The group equivariant ResNets achieve the highest training accuracy and lowest loss.

The mean accuracies and losses during training (calculated on the validation dataset) are shown in Table 6 for all of the default models, as well as G_ResNet18_c and G_ResNet50_c trained with a chirality-equivariant predict function. G_ResNet50 achieves the lowest loss in training, whereas G_ResNet50_c reaches the highest accuracy. Neither LeNet model succesfully trained on the dataset, achieving 62-64% accuracy by guessing 'Other' for most images (63.3% of the training set was 'Other').

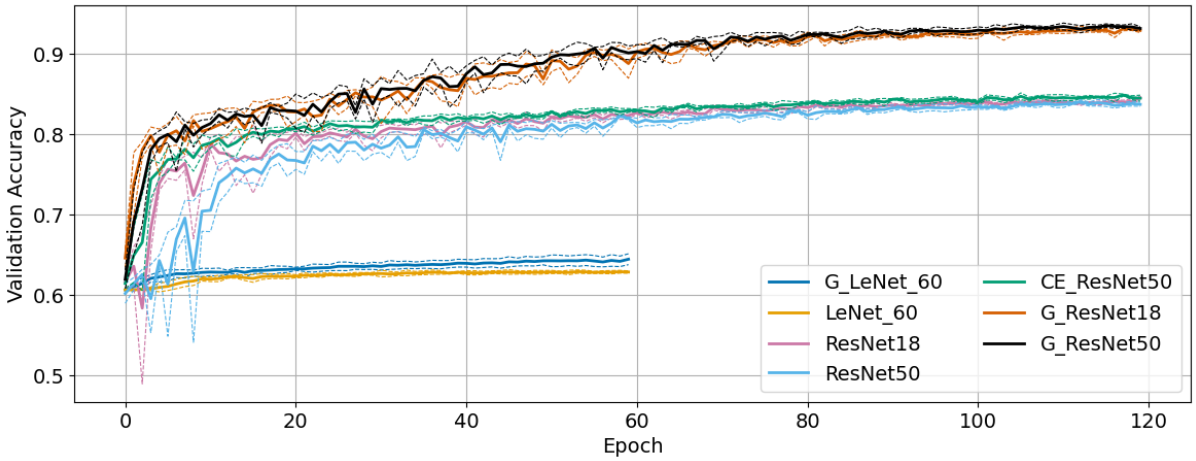


Figure 7: Training accuracy per epoch for selected models, averaged over five runs with the standard deviations shown with dashed lines.

Table 6: Training and testing metrics

Model	**Validation**		**Test**	
	Loss	**Accuracy (%)**	**Loss**	**Accuracy (%)**
LeNet	0.9195 ± 0.0021	62.79 ± 0.20	0.5280 ± 0.0052	91.25 ± 0.10
G_LeNet	0.9028 ± 0.0070	64.32 ± 0.78	0.4899 ± 0.0130	91.80 ± 0.15
ResNet18	0.6216 ± 0.0007	84.31 ± 0.17	0.3850 ± 0.0013	94.64 ± 0.09
ResNet50	0.6229 ± 0.0016	83.71 ± 0.40	0.3890 ± 0.0030	94.46 ± 0.22
CE_ResNet50	0.6152 ± 0.0009	84.57 ± 0.20	0.3777 ± 0.0022	**95.07 ± 0.13**
G_ResNet18	0.5676 ± 0.0012	92.86 ± 0.25	0.3775 ± 0.0019	94.51 ± 0.16
G_ResNet50	0.5636 ± 0.0015	93.36 ± 0.39	0.3820 ± 0.0025	94.43 ± 0.14
G_ResNet18_c	0.5692 ± 0.0012	92.76 ± 0.21	**0.3766 ± 0.0045**	94.40 ± 0.32
G_ResNet50_c	**0.5606 ± 0.0032**	**93.73 ± 0.50**	0.3805 ± 0.0042	94.42 ± 0.24

Each run of each model was then tested on the GZ1 testing dataset. The mean accuracies and losses are shown in Table 6. As shown in Table 1, the test dataset had a chirality violation of 3.00, and the models were trained on a dataset with a chirality violation of 3.24. Unlike the training dataset, the test dataset was not downsampled and is dominated by Other galaxies, making up 92% of the total. As a result, any accuracies of 92% may suggest that a model is simply selecting Other all of the time to minimise loss and maximise accuracy.

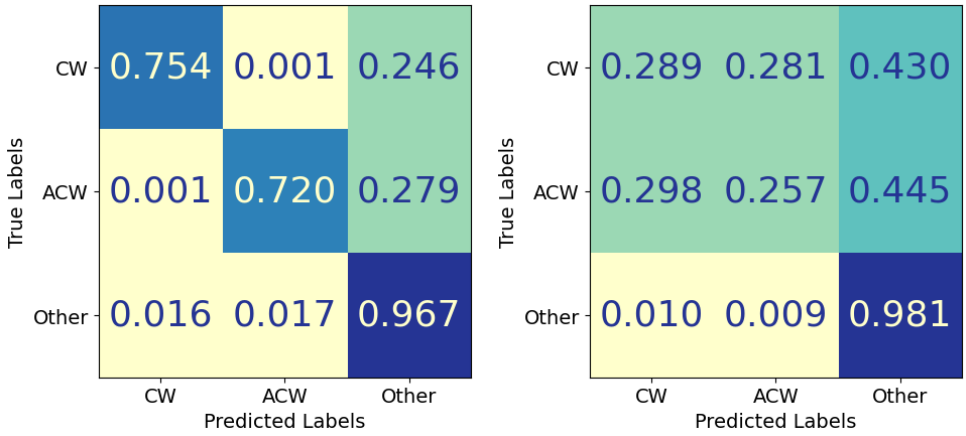
The lowest loss in testing was achieved by G_ResNet18_c, and the highest accuracy by CE_ResNet50. Once again, the LeNet models did not perform well, reaching the same accuracy as the percentage of 'Other' galaxies in the test dataset (91%). This suggests that classifiying galaxy chirality is too complex a problem for the simple 5-layer LeNet models. The significantly higher accuracy seen in testing by the G_ResNet models does not appear to extend to testing, possibly due to the large amount of uncertain galaxy images with spiral vote fractions of approximately 0.1-0.3.

There is a significant difference in the time taken to train the different models, and then use the models to generate predictions. Models were trained on a NVIDIA A100 64GB GPU, with the default ResNet models taking approximately 1 hour to train for 120 epochs, and the G_ResNet18 and G_ResNet50 models taking 2 and 3 hours respectively. Adding the chirality equivariant predict function does not significantly increase the performance of the default ResNet models, but increases the training time for G_ResNet18_c and G_ResNet50_c to six hours each. The custom predict function also doubles the memory needed when training or testing a model, as both the original images and flipped images for each batch are stored in

memory.

5.2 Reflection Equivariance

The group-equivariant G_ResNet18 and G_ResNet50 neural networks are steerable and the transformations that they are equivariant to can be adjusted. For some scenarios, reflection equivariance may be useful, although not in the case of galaxy chirality violation as reflecting a clockwise spiral turns it into an anticlockwise spiral. Figure 8 shows two confusion matrices (actual labels against predicted labels on test dataset, showing the proportion of matching labels) for G_ResNet18 with and without reflection equivariance. The reflection equivariant model (Figure 8b) cannot distinguish between CW and ACW galaxies and incorrectly misclasses most spiral galaxies as 'Other'.



(a) Rotation Equivariant (b) Rotation/Reflection Equivariant

Figure 8: Confusion matrices for G_ResNet18 with and without reflection equivariance.

5.3 Effects of Rotation

In order to investigate the importance of models being trained to recognise different rotations of images (either by inherent rotation equivariance or by randomly rotating the training dataset), several models were trained without the random image rotation during training. These models are labelled with '_nr' for no rotation in training.

Table 7: Test metrics without rotation in training

Model	Loss	Accuracy (%)
ResNet18_nr	0.4514 ± 0.0075	92.18 ± 0.29
ResNet50_nr	0.4579 ± 0.0077	92.32 ± 0.37
CE_ResNet50_nr	0.4075 ± 0.0023	94.27 ± 0.06
G_ResNet18_nr	0.4119 ± 0.0298	93.33 ± 1.10
G_ResNet50_nr	0.4215 ± 0.0136	93.39 ± 0.83

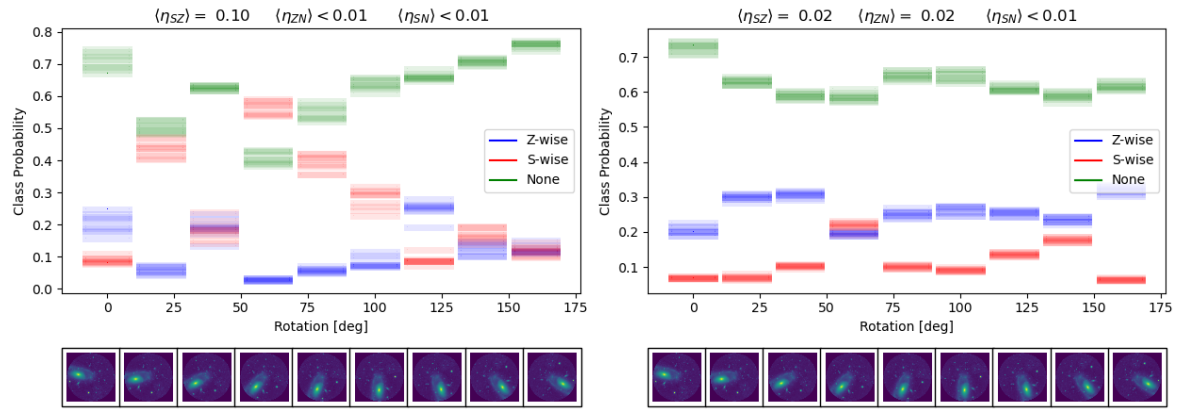
The test results of the models are shown in Table 7. Contrasting these with Table 6, most models perform several percent better with image rotation in training. Considering accuracy, only the equivariant ResNets and CE_ResNet50 perform well, as 92% of the test dataset is 'Other', so any model with accuracy close to this or below may simply be predicting 'Other' for every image.

These results show that adding some degree of rotation augmentation in training significantly improves the performance of ResNet models. Galaxies can be in a number of orientations, and measures such as adding random rotation in training can help models to generalise and

recognise more spirals. Consequently, models that are inherently rotation equivariant should perform particularly well.

Another way to determine the effect of rotation on a model is by investigating how model predictions on the same image vary when the image is rotated. In order to assess how the model's certainty varies with rotated images (as well as just the predicted probability), a dropout layer was introduced before the last convolutional layer. This dropout layer randomly selects nodes with a probability of 0.5 to 'drop' and pass an output of zero onto the next layer, preventing a small number of nodes from determining the model output. This also has the effect of providing an uncertainty in the model's output [44], as by running the model multiple times with dropout enabled, the standard deviation of the outputs can be used as an estimated uncertainty. Adding a dropout layer was not found to significantly affect model performance on the full test dataset, achieving an accuracy of $(94.67 \pm 0.63)\%$ for CE_ResNet50, within uncertainty of the accuracy achieved without a dropout layer.

Following the method used by Scaife and Porter [38], each sample in the test dataset was passed through the model 10 times with dropout enabled, each at 9 rotations from 0 to 180°. In order to quantify the variation in model predictions over rotations, a distribution-free overlap index η [45] was used. This index applies to two classes, so was calculated for each combination of the three predicted classes: CW, ACW and Other. The average of this index, $\langle\eta\rangle$, then provides a metric of model confidence over rotation, where a lower overlap index means the predictions for each class are more separated, and the prediction has a higher confidence. Figure 9a and 9b show how the range of predicted probabilities for each image rotation varies for each class for both CE_ResNet50 and G_ResNet50, with the lower overlap of 0.02 between ACW and CW ($\langle\eta_{SZ}\rangle$) for the group-equivariant model showing how the predictions vary less over rotations.



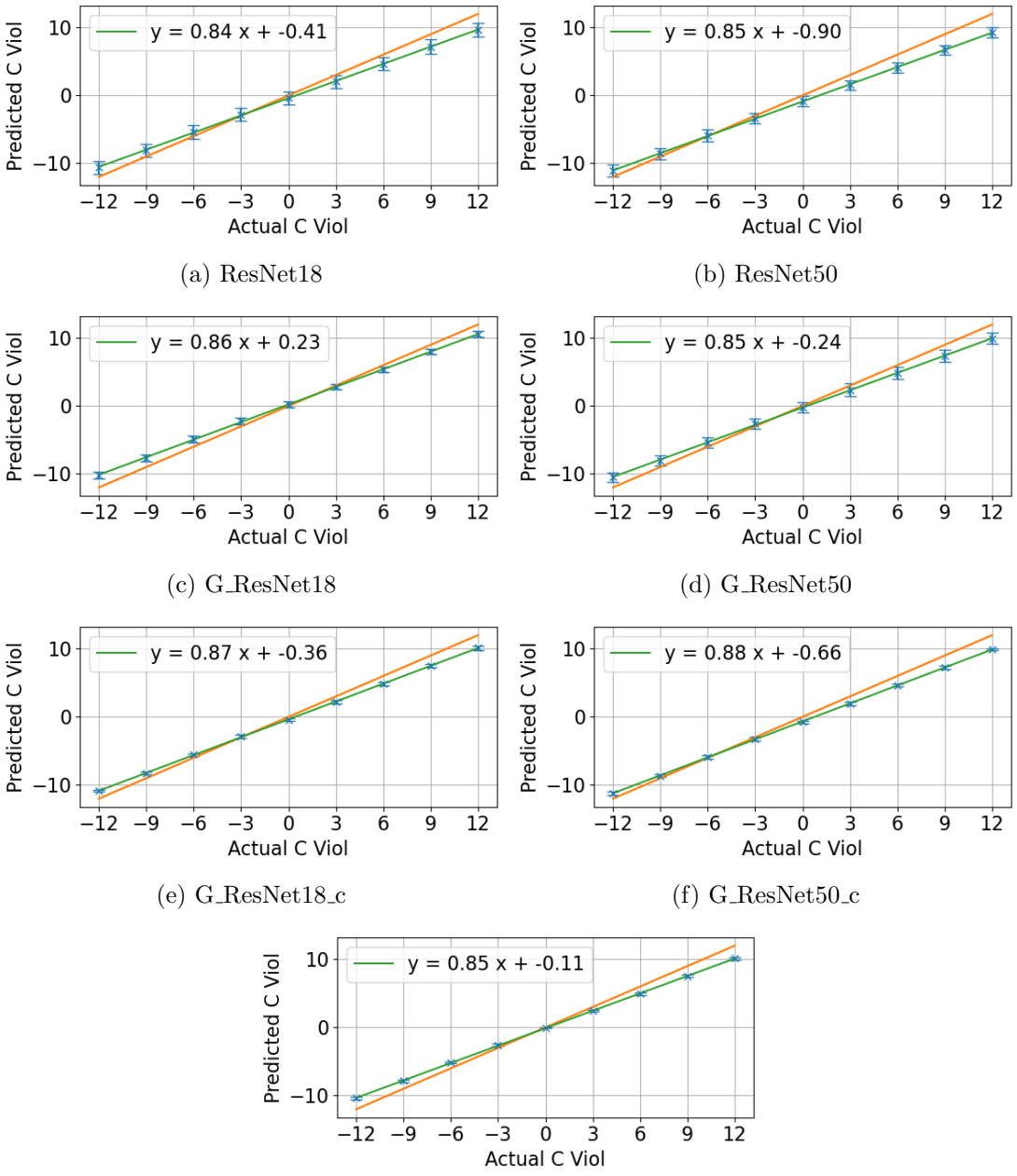
(a) CE_ResNet50 (b) G_ResNet50

Figure 9: Plots showing the variation in predicted probabilities for each class as an input image is rotated through 180°.

Overlap indexes for 1000 randomly selected test dataset images were calculated for the G_ResNet50 and CE_ResNet50 models to investigate the effect of rotation equivariance. The rotation-equivariant G_ResNet50 produced a lower $\langle\eta_{SZ}\rangle$ for 19.3% of the images, whereas CE_ResNet50 had a lower overlap for 11.2%. For $\langle\eta_{ZN}\rangle$, this was 4.2% and 7.3%, and for $\langle\eta_{SN}\rangle$ this was 4.5% and 5.3%. All other images produced overlap values within 0.01 of each other. From this small sample size, rotation equivariance appears to improve the stability over rotation between clockwise and anticlockwise galaxies, but does not significantly affect other class combinations. Using 1000 galaxies out of the approximately 40,000 in the test set is a reasonably small sample but was limited by time constraints as calculating the overlap required passing each image to both models 90 times.

5.4 Representation of Chirality Violation

As well as considering accuracy on a test set, it is important to investigate how well model predictions represent the actual chirality violation of datasets. In order to do this, a test dataset (labelled as the chirality test set) was created using only S or Z spiral galaxies from the main testing dataset described in Table 1, where a galaxy was considered S or Z if the vote fraction was greater than 0.5. This created a dataset of 2333 galaxies. The chirality of this dataset was then varied from -12 to 12 in steps of 3, keeping the overall number of galaxies constant. At each step, the 5 versions of each model predicted the number of S and Z spirals in the dataset, and these were then used to calculate the predicted chirality violation.



(a) ResNet18

(b) ResNet50

(c) G_ResNet18

(d) G_ResNet50

(e) G_ResNet18_c

(f) G_ResNet50_c

(g) CE_ResNet50

Figure 10: Plots of set chirality violation significance of a test dataset against the chirality violation significance of each model's predictions

Figure 10 shows the predicted and actual chirality violation significances for each model. Each plot also includes an 'ideal' CVS, with a gradient of one and zero intercept, expected if the

model exactly recreates the CVS of the test dataset. Polynomial fits to the predicted results show all of the models perform reasonably well, with no significant difference between the default, group-equivariant or chirality-equivariant models. The lower gradients (0.84-0.88) are likely due to the number of galaxies identified being less than the total. It is also difficult to assess whether the chirality equivariant models successfully account for the known chirality difference (CSV=3σ) of the training dataset. The fit intercepts decrease moving from G_ResNet18 and G_ResNet50 (shown in Figures 10c and 10d to their chirality equivariant versions (Figures 10e and 10f), but increases from ResNet50 (Figure 10b) to CE_ResNet50 (Figure 10g). More than 2333 galaxies may be needed in order to see a clearer effect.

As the significance of the chirality violation also depends on the total number of spirals identified (Equation 1), it is useful to investigate how many galaxies in the chirality test set were correctly identified as spirals. Table 8 shows the total number of spiral galaxies (ACW or CW with probability of 0.5 or greater). G_ResNet18_c correctly identified the most spirals, categorising 76% of the 2333 galaxies as either CW or ACW when averaged across all of the set chirality test sets, whereas the default ResNet18 and ResNet50 identified the least at 71.8%. Considering the equivariant models, the 18-layer versions identify slightly more galaxies correctly than the 50-layer versions on average, although the 50-layer predictions have a lower standard deviation, showing the total predicted number is more consistent across model runs.

Table 8: Number of spirals identified in chirality test dataset

Model	**Total Spirals (N=2333)**
ResNet18	1675 (71.8%) ± 24
ResNet50	1674 (71.8%) ± 39
CE_ResNet50	1692 (72.5%) ± 28
G_ResNet18	1726 (74.0%) ± 40
G_ResNet50	1686 (72.3%) ± 27
G_ResNet18_c	1772 (76.0%) ± 49
G_ResNet50_c	1748 (74.9%) ± 31

5.5 DESI Dataset Predictions

The ResNet models were then used to identify spirals in the full DESI dataset of 8.7 million galaxies. The 35988 galaxies used in training and the 41737 used in testing were removed. Table 9 shows the predictions of each type of galaxy in the DESI dataset.

Table 9: Model predictions on unlabelled DESI dataset

Model	**ACW**	**CW**	**Other**
ResNet18	119515 (1.4%) ± 6911	108011 (1.3%) ± 2279	8384134 (97.4%) ± 5022
ResNet50	130157 (1.5%) ± 5087	132413 (1.5%) ± 11446	8349090 (97.0%) ± 13118
CE_ResNet50	111952 (1.3%) ± 2765	112272 (1.3%) ± 2587	8387436 (97.4%) ± 5348
G_ResNet18	133378 (1.5%) ± 10019	117627 (1.4%) ± 2564	8360655 (97.1%) ± 8946
G_ResNet50	129642 (1.5%) ± 7201	118142 (1.4%) ± 2634	8363876 (97.1%) ± 7873
G_ResNet18_c	122902 (1.4%) ± 8296	123142 (1.4%) ± 8336	8365616 (97.1%) ± 16632
G_ResNet50_c	126247 (1.5%) ± 5806	126661 (1.5%) ± 5899	8358752 (97.1%) ± 11705

Table 10 shows these identified galaxies as a ratio of anticlockwise to clockwise, as well as the significance of any chirality violation calculated using Equation 1. Models with the CE predict function, designed to ensure they have identical selection functions for clockwise and anticlockwise galaxies, do not predict any significant chirality violation, with a ratio of 1.00 ± 0.00 and chirality violation of -0.48 to -0.82. CE models also have similar error ranges on the number of CW and ACW galaxies identified, compared to the asymmetry for models such

as ResNet50, with a 3.9% error for ACW and a 8.6% error. This is likely to be a result of the symmetrical selection functions of the CE models, whereas the non-CE models may have asymmetrical selection functions, possibly following the volunteer bias towards ACW galaxies in training classification.

There is no significant difference in the number of galaxies identified by 18-layer ResNets compared to 50-layer ResNets, suggesting that the available additional complexity of 50-layer models is not necessary. Whilst CE_ResNet50 identified fewer galaxies (1.3%) than the CE group equivariant ResNets (1.4-1.5%), it had a significantly smaller error on the number of galaxies identified. All of the models have a high uncertainty for chirality violation, particularly ResNet50 with a $> 400\%$ error. This results from the high variation in the number of both ACW and CW galaxies identified. The overall proportion of spiral galaxies to other galaxies (approximately 1.5/1.5/97%) is lower than the proportion identified in the SDSS by Galaxy Zoo 1 volunteers at 4/4/92%. The clearer images produced by the DESI survey may mean more other galaxies are identified, and as the DESI survey is not labelled it is difficult to assess how accurate these proportions are to the true values. However, they are likely to be an underestimate, as shown by Table 8, where models only confidently identified 72-76% of spirals in an all-spiral dataset.

Table 10: Chirality of model predictions

Model	**ACW/CW Ratio**	**CVS**
ResNet18	1.11 ± 0.08	23.92 ± 18.81
ResNet50	0.99 ± 0.10	-3.98 ± 23.79
CE_ResNet50	1.00 ± 0.00	-0.68 ± 0.55
G_ResNet18	1.14 ± 0.10	31.07 ± 23.17
G_ResNet50	1.10 ± 0.06	22.93 ± 14.73
G_ResNet18_c	1.00 ± 0.00	-0.48 ± 0.38
G_ResNet50_c	1.00 ± 0.00	-0.82 ± 0.26

6 Conclusion

A number of different convolutional neural networks were built, varying the model architecture (LeNet or ResNet), the number of layers (18 or 50), whether models were group equivariant, and whether a chirality-equivariant prediction function by Jia, Zhu, and Pen [18] was used. These models were tested on a set of DESI images, classified as clockwise spiral, anticlockwise spiral or other by volunteers as part of the Galaxy Zoo 1 project. The trained models were then used to create predictions of the difference between the number of clockwise and anticlockwise galaxies on the full DESI survey of 8.7 million galaxies.

Considering loss and accuracy on the test dataset, all ResNet-based models performed similarly, achieving approximately 94% accuracy. LeNet models were unable to train, suggesting spiral classification is too complex for a shallower network. G_ResNet18 and G_ResNet50 achieved higher validation accuracies in training by 8% compared to non-equivariant models, but this same accuracy increase was not seen in testing. When models were trained without random rotation augmentation in training, their accuracy decreased by 1-2%. All models preserved preserved the chirality violation significance of a known dataset well in testing, with no clear difference between default, group-equivariant or chirality-equivariant models.

There was not a significant difference between the performance of 18 and 50-layer versions of ResNet, suggesting that the additional complexity of ResNet50 was not needed. However, the group equivariant models did take significantly longer to train (approximately twice as long, or six times as long with the CE predictor), as a result of their increased complexity. Adding the chirality-equivariant prediction function adds a similar performance increase over conventional

ResNet without significantly increasing training time, although the CE predict function does require more memory as each image is passed to the model twice.

When the models were used to predict the number of spiral galaxies in the full DESI dataset, the three CE models did not find any significant violation in spin chirality. All models showed significant variation in the number of spiral galaxies identified over multiple runs with large errors, with the non-CE models showing an asymmetrical variation between the variation in the number of clockwise and anticlockwise galaxies identified.

6.1 Future Work

As with any machine learning project, more data would be advantageous for use in training, especially given the vast differences in galaxy shapes, sizes, orientations, spiral arms, brightness and many other factors. However, data for training is limited by the need for labelled data - Galaxy Zoo 1 took several years to classify just under a million galaxies, and this was only possible due to the work of thousands of volunteers. Instead, more use could be made of the already available data, including investigating to see the effects of different pre-processing cuts on model performance. For example, the cut to remove galaxies with other objects within a 1 arcsecond radius removed approximately 400,000 objects, and the effect of removing or reducing this cut may provide more training images. Additionally, there are approximately 250,000 additional galaxies classified by Galaxy Zoo 1 that are missing spectra and were not included in the training set, yet may still be useful for training if they can be matched with images from the DESI survey.

More investigation could also be performed into the effects of other data augmentation in training beyond rotation, including translation, cropping/scaling, colour. Whilst CNNs should be reasonably equivariant to translation due to their weight sharing, adding random translation during training may still improve performance, especially given adding random rotation during training still improved the performance of rotation-equivariant models. Throughout this project, the threshold for classifying a galaxy as a clockwise or anticlockwise spiral was a vote fraction or predicted probability of 0.5 or greater, and the effects of varying this threshold could also be investigated.

The group-equivariant ResNets used in this project were created by building a general ResNet module in the same format as the `PyTorch` ResNet, meaning any specific version of G_ResNet (including ResNet18, ResNet152, WideResNet etc.) can be easily created and applied to other problems.

General Notes

The code used in this project, including models and model weights, is available on GitHub at https://github.com/neil-power/mphys-galaxy. This project made extensive use of Galahad, the Radio Astronomy Research Cluster at the Jodrell Bank Centre for Astrophysics.

References

[1] Pavel Motloch et al. "Observational detection of correlation between galaxy spins and initial conditions". In: *Nature Astronomy* 5 (3 Mar. 2020), pp. 283–288. DOI: 10.1038/s41550-020-01262-3.

[2] Pavan Kumar Aluri et al. "Is the Observable Universe Consistent with the Cosmological Principle?" In: *Classical and Quantum Gravity* 40 (9 July 2022). DOI: 10.1088/1361-6382/acbefc.

[3] Chris J. Lintott et al. "Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey". In: *Monthly Notices of the Royal Astronomical Society* 389 (3 2008), pp. 1179–1189. ISSN: 13652966. DOI: 10.1111/j.1365-2966.2008.13689.x.

[4] Kevork N Abazajian et al. “The Seventh Data Release of the Sloan Digital Sky Survey”. In: *The Astrophysical Journal Supplement Series* 182 (2009), pp. 543–558. DOI: 10.1088/0067-0049/182/2/543.

[5] Arjun Dey et al. “Overview of the DESI Legacy Imaging Surveys”. In: *The Astronomical Journal* 157 (5 May 2019), p. 168. ISSN: 0004-6256. DOI: 10.3847/1538-3881/ab089d.

[6] C. C. Lin and Frank H. Shu. “On the Spiral Structure of Disk Galaxies.” In: *The Astrophysical Journal* 140 (Aug. 1964), p. 646. ISSN: 0004-637X. DOI: 10.1086/147955.

[7] M. W. Mueller and W. D. Arnett. “Propagating star formation and irregular structure in spiral galaxies”. In: *The Astrophysical Journal* 210 (Dec. 1976), p. 670. ISSN: 0004-637X. DOI: 10.1086/154873.

[8] I. I. Pasha and M. A. Smirnov. “On the direction of rotation of the spirals in galaxies”. In: *Astrophysics and Space Science* 86 (1 Aug. 1982), pp. 215–224. ISSN: 0004-640X. DOI: 10.1007/BF00651844.

[9] P. J. E. Peebles. “Origin of the Angular Momentum of Galaxies”. In: *The Astrophysical Journal* 155 (Feb. 1969), p. 393. ISSN: 0004-637X. DOI: 10.1086/149876.

[10] S. Capozziello and A. Lattanzi. “Spiral galaxies as chiral objects?” In: *Astrophysics and Space Science* 301 (1-4 Jan. 2006), pp. 189–193. ISSN: 0004640X. DOI: 10.1007/S10509-006-1984-6.

[11] Lior Shamir. “Large-Scale Asymmetry in the Distribution of Galaxy Spin Directions—Analysis and Reproduction”. In: *Symmetry 2023, Vol. 15, Page 1704* 15 (9 Sept. 2023), p. 1704. ISSN: 2073-8994. DOI: 10.3390/SYM15091704.

[12] Masanori Iye and Hajime Sugai. “A catalog of spin orientation of southern galaxies”. In: *The Astrophysical Journal* 374 (June 1991), p. 112. ISSN: 0004-637X. DOI: 10.1086/170101.

[13] Kate Land et al. “Galaxy Zoo: the large-scale spin statistics of spiral galaxies in the Sloan Digital Sky Survey”. In: *Monthly Notices of the Royal Astronomical Society* 388 (4 Aug. 2008), pp. 1686–1692. ISSN: 00358711. DOI: 10.1111/j.1365-2966.2008.13490.x.

[14] Wayne B. Hayes, Darren Davis, and Pedro Silva. “On the nature and correction of the spurious S-wise spiral galaxy winding bias in Galaxy Zoo 1”. In: *Monthly Notices of the Royal Astronomical Society* 466 (4 May 2017), pp. 3928–3936. ISSN: 0035-8711. DOI: 10.1093/MNRAS/STW3290.

[15] H. T. MacGillivray et al. “The anisotropy of the spatial orientations of galaxies in the Local Supercluster”. In: *Astronomy and Astrophysics* 145 (1 1985), pp. 269–274. ISSN: 0004-6361.

[16] Michael J Longo. “Detection of a dipole in the handedness of spiral galaxies with redshifts z 0.04”. In: *Physics Letters B* 699 (2011), pp. 224–229. DOI: 10.1016/j.physletb.2011.04.008.

[17] Lior Shamir. “Analysis of spin directions of galaxies in the DESI Legacy Survey”. In: *Monthly Notices of the Royal Astronomical Society* 516 (2 Aug. 2022), pp. 2281–2291. DOI: 10.1093/mnras/stac2372.

[18] He Jia, Hong-Ming Zhu, and Ue-Li Pen. “Galaxy Spin Classification. I. Z-wise versus S-wise Spirals with the Chirality Equivariant Residual Network”. In: *The Astrophysical Journal* 943.1 (2023). ISSN: 1538-4357. DOI: 10.3847/1538-4357/aca8aa.

[19] Mike Walmsley et al. *Zoobot: Deep learning galaxy morphology classifier*. 2022.

[20] David E Rumelhart, Bernard Widrow, and Michael A Lehr. “The basic ideas in neural networks”. In: *Communications of the ACM* 37 (3 Jan. 1994), pp. 87–92. ISSN: 0001-0782. DOI: 10.1145/175247.175256.

[21] Shun-ichi Amari. “Backpropagation and stochastic gradient descent method”. In: *Neurocomputing* 5 (4-5 Jan. 1993), pp. 185–196. ISSN: 09252312. DOI: 10.1016/0925-2312(93)90006-O.

[22] Andrea Apicella et al. “A survey on modern trainable activation functions”. In: *Neural Networks* 138 (Jan. 2021), pp. 14–32. ISSN: 0893-6080. DOI: 10.1016/J.NEUNET.2021.01.026.

[23] Diederik P Kingma and Jimmy Lei Ba. “Adam: A Method for Stochastic Optimization”. In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (Jan. 2014).

[24] Anqi Mao, Mehryar Mohri, and Yutao Zhong. “Cross-Entropy Loss Functions: Theoretical Analysis and Applications”. In: *Proceedings of Machine Learning Research* (2023).

[25] Waseem Rawat and Zenghui Wang. "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review". In: *Neural Computation* 29 (9 Sept. 2017), pp. 2352–2449. ISSN: 0899-7667. DOI: 10.1162/neco_a_00990.

[26] Osman Semih Kayhan and Jan C. van Gemert. "On Translation Invariance in CNNs: Convolutional Layers can Exploit Absolute Spatial Location". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Mar. 2020), pp. 14262–14273. ISSN: 10636919. DOI: 10.1109/CVPR42600.2020.01428.

[27] Karel Lenc and Andrea Vedaldi. "Understanding image representations by measuring their equivariance and equivalence". In: *International Journal of Computer Vision* (Nov. 2014).

[28] Taco S. Cohen and Max Welling. "Group Equivariant Convolutional Networks". In: *33rd International Conference on Machine Learning, ICML 2016* 6 (Feb. 2016), pp. 4375–4386.

[29] Taco S. Cohen and Max Welling. "Steerable CNNs". In: *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings* (Dec. 2016).

[30] Nour Eldeen M. Khalifa et al. "Deep Galaxy: Classification of Galaxies based on Deep Convolutional Neural Networks". In: *Computer Vision and Pattern Recognition* (Sept. 2017).

[31] Mitchell K. Cavanagh, Kenji Bekki, and Brent A. Groves. "Morphological classification of galaxies with deep learning: comparing 3-way and 4-way CNNs". In: *Monthly Notices of the Royal Astronomical Society* 506 (1 June 2021), pp. 659–676. DOI: 10.1093/mnras/stab1552.

[32] Sander Dieleman, Kyle W. Willett, and Joni Dambre. "Rotation-invariant convolutional neural networks for galaxy morphology prediction". In: *Monthly Notices of the Royal Astronomical Society* 450 (2 June 2015), pp. 1441–1459. ISSN: 0035-8711. DOI: 10.1093/MNRAS/STV632.

[33] Chris Lintott et al. "Galaxy Zoo 1: Data release of morphological classifications for nearly 900 000 galaxies". In: *Monthly Notices of the Royal Astronomical Society* 410 (1 2011), pp. 166–178. ISSN: 13652966. DOI: 10.1111/j.1365-2966.2010.17432.x.

[34] Mike Walmsley et al. "Galaxy Zoo DESI: Detailed morphology measurements for 8.7M galaxies in the DESI Legacy Imaging Surveys". In: *Monthly Notices of the Royal Astronomical Society* 526 (3 Oct. 2023), pp. 4768–4786. ISSN: 0035-8711. DOI: 10.1093/MNRAS/STAD2919.

[35] Adrian M Price-Whelan et al. "The Astropy Project: Building an open-science project and status of the v2. 0 core package". In: *The Astronomical Journal* 156 (3 2018), p. 123.

[36] Adam Ginsburg et al. "astroquery: An Astronomical Web-querying Package in Python". In: *The Astronomical Journal* 157 (3 Feb. 2019), p. 98. ISSN: 1538-3881. DOI: 10.3847/1538-3881/AAFC33.

[37] Yann LeCun et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86 (11 1998), pp. 2278–2323. ISSN: 00189219. DOI: 10.1109/5.726791.

[38] Anna M M Scaife and Fiona Porter. "Fanaroff-Riley classification of radio galaxies using group-equivariant convolutional neural networks". In: *MNRAS* 000 (2015), pp. 1–11. DOI: 10.5281/zenodo.4288837.

[39] Adam Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: Curran Associates, Inc., 2019, pp. 8024–8035.

[40] Gabriele Cesa, Leon Lang, and Maurice Weiler. "A Program to Build E(n)-Equivariant Steerable CNNs". In: *ICLR 2022* (2022).

[41] Maurice Weiler and Gabriele Cesa. "General $E(2)$-Equivariant Steerable CNNs". In: *Advances in Neural Information Processing Systems* 32 (Nov. 2019). ISSN: 10495258.

[42] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *Computer Vision and Pattern Recognition* (Dec. 2015).

[43] Mike Walmsley. galaxy-datasets. URL: https://github.com/mwalmsley/galaxy-datasets.

[44] Yarin Gal and Zoubin Ghahramani. "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning". In: *33rd International Conference on Machine Learning, ICML 2016* 3 (June 2015), pp. 1651–1660.

[45] Massimiliano Pastore and Antonio Calcagnì. "Measuring distribution similarities between samples: A distribution-free overlapping index". In: *Frontiers in Psychology* 10 (May 2019), p. 455421. ISSN: 16641078. DOI: 10.3389/FPSYG.2019.01089/BIBTEX.

Appendix

Feature Identification

The *shap* library can be used to display Shapely values for areas of the image. Shapely values are a concept originating from game theory, and show the mean contribution to a model's output for features. Figure 11 shows examples of ACW, CW, and other galaxies predicted by ResNet18, along with an additional image of a cat in an anticlockwise spiral position (trailing tail), which the model also correctly identifies.

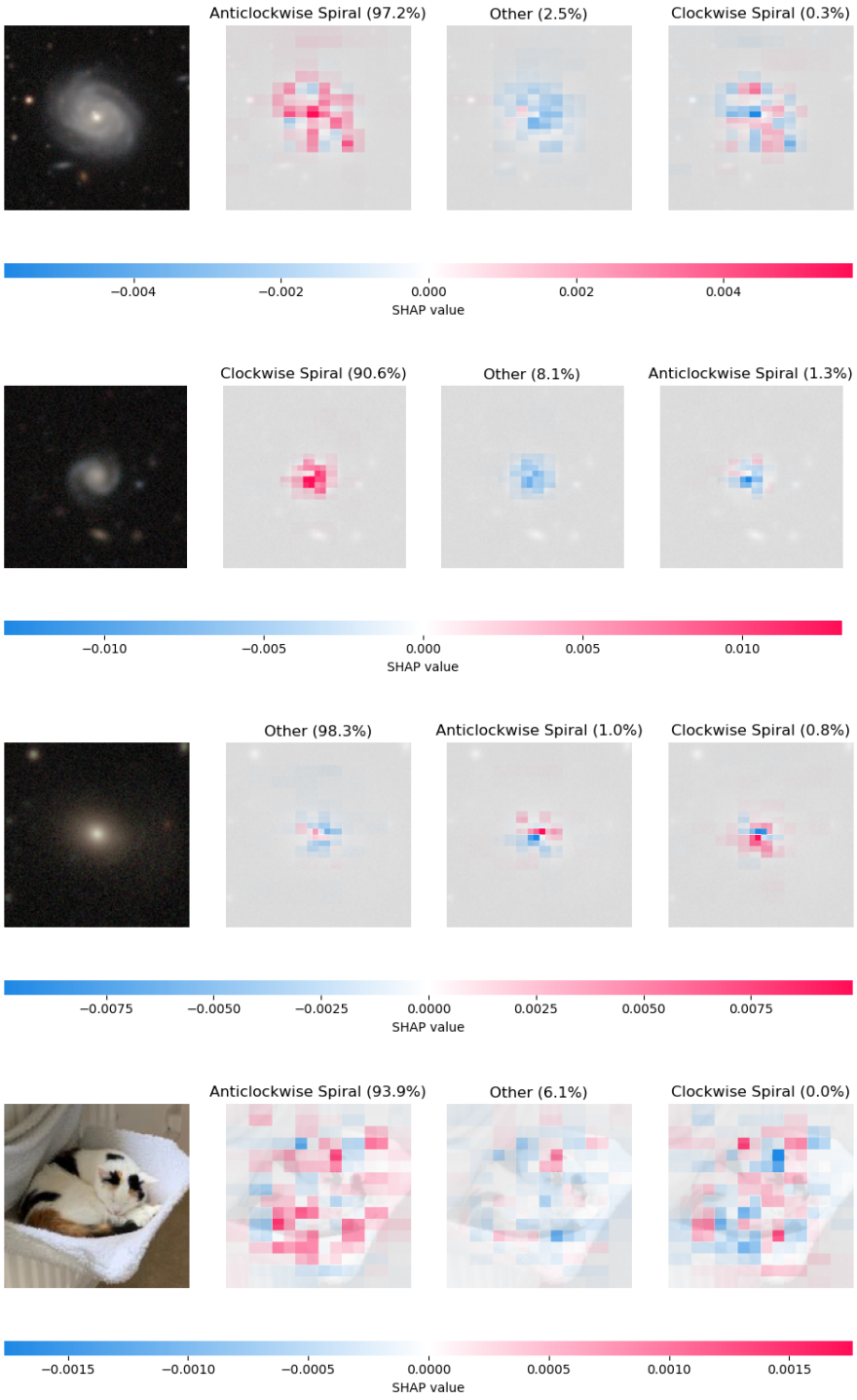


Figure 11: Feature importance visualisations for anticlockwise, clockwise and other galaxies. Cat photo: Natalie Lines