Rijksuniversiteit Groningen

Bachelor Thesis

Investigating the bar-AGN connection using deep learning

Abstract

This thesis aims to re-evaluate the connection between the presence of bars and AGN activity. We make use of machine learning to identify the galaxy morphology of HSC galaxies with redshift $0.1 \leq z \leq 0.55$. We made use of a model with an accuracy of 86.9% to identify our most confident bars and a model with an accuracy of 81.8% to identify our most confident non-bars. We make use of a mid-infrared WISE colour criterion and SDSS optical line emission to identify AGNs in our sample. We found an overall AGN fraction of 1% $\pm$ 0.1% and 1.2% $\pm$ 0.1% for bars and non bars, respectively for the WISE data and 13.2% $\pm$ 1.8% in bars and 14.0%$\pm$ 1.0% in non-bars for the SDSS data. We also evaluated f_{AGN} as functions of redshift and r-band magnitude. These plots did not show any evidence that bars are a prominent mechanism in triggering AGNs.

rijksuniversiteit
groningen

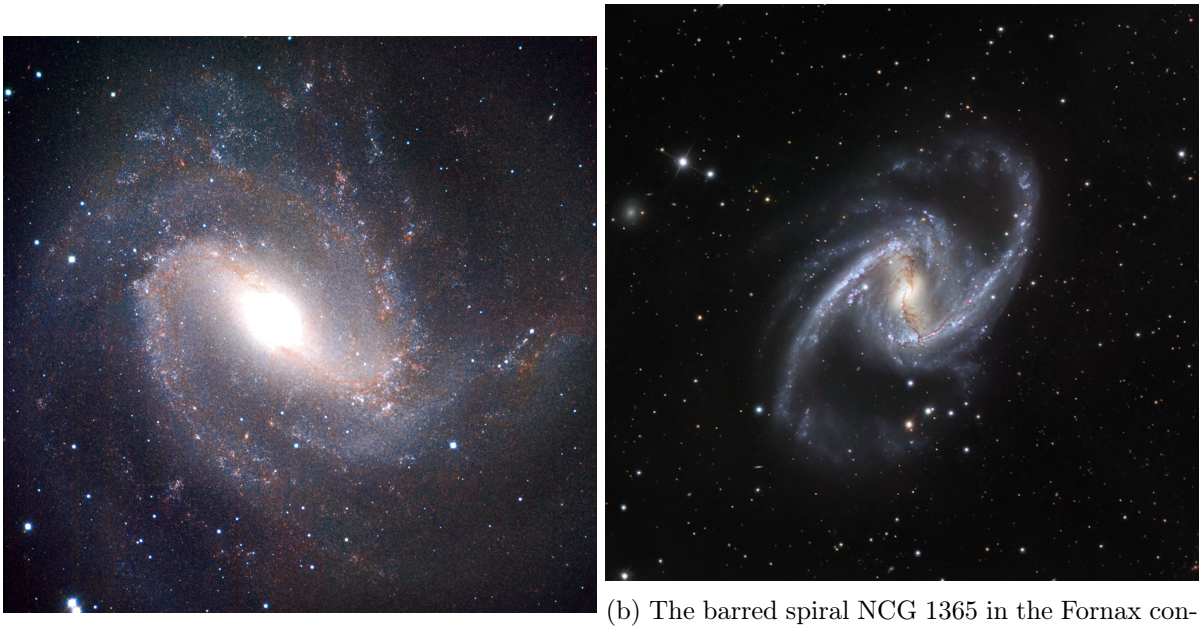
Author:
M. van Asselt
s3744531

Supervisors:
L. Wang
B. Margalef Bentabol

Contents

1 Introduction

Galactic bars are some of the most striking features in astronomy. They provide some beautiful images of galaxies, showing an interesting large-scale feature, as shown in Figure 1. Aside from looking pretty, bars also play an important role in galaxy evolution. For example, they can aid in the creation of spiral arms (Mo, Bosch, and White 2010). They have also been speculated to be one of the mechanisms that trigger Active Galactic Nuclei (AGNs) (Laurikainen, Salo, and Buta 2004; Goulding et al. 2017) However, this is still a very controversial topic. There is a lot of disagreement between studies that have previously looked at this bar-AGN connection. Some found that there was a connection between the two (Silva-Lima et al. 2022; S. Oh, K. Oh, and Yi 2011; Alonso, Coldwell, and Lambas 2013), while other did not find such a connection (Cisternas et al. 2015; Lee et al. 2012).



(a) Barred spiral Messier 83, or NGC 5236, in the hydra constellation. ESO

(b) The barred spiral NCG 1365 in the Fornax constellation. ESO/IDA/Danish 1.5 m/ R. Gendler, J-E. Ovaldsen, C. Thöne, and C. Feron.

Figure 1: Examples of bars in spiral galaxies. Both images are from ESO.

Most of these studies were done in the low-redshift ($z \leq 0.1$) and were done using the Sloan Digital Sky survey (SDSS). Due to the resolution of SDSS the morphological classification is limited to a lower redshift sample (Hart et al. 2016). However, a newer survey, the Hyper Suprime-Cam SSP Survey (HSC), has a better resolution and is deeper than SDSS (Aihara et al. 2017). This improved resolution allows for large-scale details, such as bars, to remain identifiable at higher redshifts. In this thesis we will thus attempt to identify bars at a higher redshift than previous studies have done in order to re-evaluate the bar-AGN connection.

In this thesis, we will be training a machine learning algorithm to identify barred galaxies from HSC images. This will be done using the morphological classification of galaxies done by Galaxy Zoo 2 (GZ2). The algorithm will then be used to classify the galaxies in an HSC sample, constructed with redshift limit $0.1 \leq z \leq 0.55$, as barred or non-barred galaxies.
We will then evaluate the AGN fraction (f_{AGN}) in both barred and non-barred galaxies using

a mid-infrared catalog (Stern et al. 2005) and optical line-emission catalog (Brinchmann et al. 2004). The AGN fraction will be evaluated as a function of redshift and as a function of magnitude.

In section 1, we will be explaining the theory behind AGNs, bars and their suspected connection in more detail. We will also be expanding on some of the previous studies done about this topic. Section 2 describes the different surveys that this thesis uses. The machine learning algorithm will be explained in section 3.1. The training of the model and its application will be discussed in section 3.2 and 3.3, respectively. In section 4 the results will be presented. These results will be discussed in section 5. Finally, we will give a brief summary and our final conclusions in section 6.

1.1 AGN activity

There are several types of galaxies where unusual activity is observed, such as Seyfert galaxies, quasars and radio galaxies. While it used to be believed that these activities originated somewhere else, it is now known that all this activity takes place in the nucleus of the galaxy or is fueled by activity in the nucleus (Kutner 2003). We say that these galaxies have Active Galactic Nuclei (AGNs). AGNs are dense central regions that emit light that is not produced by the stars in that region or by the heated gas. Active nuclei can emit radiation throughout the entire electromagnetic spectrum (Sparke and Gallagher 2007).

One example of AGN host-galaxies are Seyfert galaxies. These are spiral galaxies with a small, bright nucleus. Seyfert galaxies are characterized by radiation that does not seem to originate from the stars, but rather from the nucleus. The nucleus shows strong and broad absorption lines coming from high excitation (Mo, Bosch, and White 2010). We can divide these galaxies up into two categories, Seyfert 1 and Seyfert 2. Seyfert 1 galaxies have very broad (permitted) lines and Seyfert 2 galaxy have lines (permitted and forbidden) with much narrower velocity widths. Seyfert galaxies are very bright at the IR, UV, X-ray and visual wavelengths (Sparke and Gallagher 2007). Seyferts are thought to make up 2 to 5% of all spiral galaxies (Kutner 2003).

Quasars (also referred to as quasi-stellar objects, or QSOs) are another example of AGNs. They have a similar optical spectra to Seyfert 1 galaxies, as they also have broad emission lines. The difference is that quasars are the brighter objects, as they have nuclei much brighter than their host galaxies. An AGN is considered to be a quasar if $L_V \geq 10^{11} L_\odot$, and a Seyfert 1 otherwise (Sparke and Gallagher 2007).

As mentioned before, AGNs such as Seyferts emit brightly at the IR, UV, X-ray and visual wavelengths. There are thus many ways to identify AGNs. We will be focusing on the mid-infrared (MIR) and the visual wavelengths.

The first way to identify AGN activity is through optical line analysis. For instance, there is the classification system developed by Baldwin, Phillips, and Terlevich 1981, where they show that most emission-line spectra of extra-galactic objects can easily be classified using line intensity ratios. They classified the objects by dominating excitation process (normal HII regions, power-law photo-ionisation or shock-wave heating) or as a planetary nebula. This was done using plots of the relative intensities of the strongest lines [OII]λ3727, H$\beta\lambda$4861, [OIII]λ5007, [OI]λ6300, H$\alpha\lambda$6563 and [NII]λ6584, avoiding lines that could provide difficulties due to the blending together with other emission or absorption lines. They tried to use lines that were as close together in wavelength as possible to avoid unnecessary error in the redshift correction. They found that the classification could be done using the following plots:(λ6584)/(λ6563) vs (λ5007)/(λ4861) and the plot of (λ3727)/(λ5007) vs (λ5007)/(λ4861), (λ6584)/(λ6563) or (λ6300)/(λ6563). Figure 2 shows an example of how Brinchmann et al. 2004 used the BPT line ratio diagram to identify star-forming galaxies, AGN host galaxies and composite galaxies. Where composite galaxies are galaxies that show star formation and AGN activity.
The optical identification method yields more AGN candidates than other wavelengths but also comes with some shortcomings. For example, it strongly favours bright and unobscured sources and will therefore not be able to identify most of the obscured AGNs that we might find at other wavelengths.(Padovani et al. 2017)

If an object has a line-of-sight neutral hydrogen column density of $N_H \simeq 1.5 \cdot 10^{24}$ cm^{-2}, it

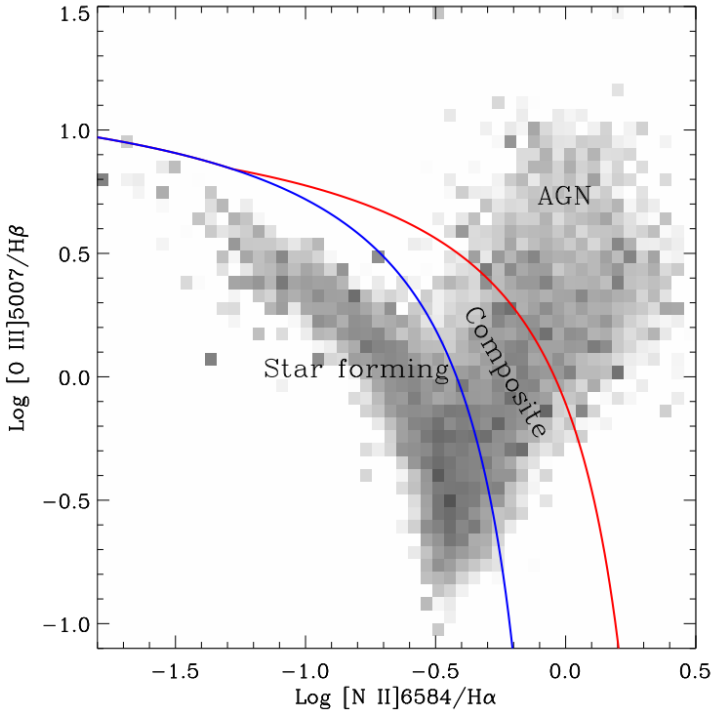


Figure 2: The BPT line ratio diagram can, for example, be used to identify star-forming, composite and AGN host galaxies (Brinchmann et al. 2004)

is considered to have Compton-thick obscuration. Compton-thick AGNs have been show to make up $20^{+9}_{-6}\%$ of the AGN population (Burlon et al. 2011). As discussed before, optical line analysis will not be able to identify most of these obscured sources. Even the deepest X-ray surveys seem to miss a lot of these object (Mateos et al. 2013). However, mid-infrared light (MIR) has been shown to still be visible even if it is from an obscured source (Lacy et al. 2004). This is because the light itself goes relatively unabsorbed (re-phrase this) but the MIR emission from dust also traces the absorption in other wavelengths (Donley et al. 2012). This makes MIR selection an excellent method for the identification of heavily obscured AGNs.

Stern et al. 2012 used a simple criterion using WISE data to reliably identify AGNs with MIR colour selection. They used the two bluest, most sensitive passbands, W1 [$3.4\mu m$] and W2 [$4.6\mu m$], to put a constraint on the resulting relative Vega magnitude. The Infrared Array Camera (IRAC) on the Spritzer Telescope also offered a colour selection for AGNs (*Insert citation*). However, in order to distinguish between AGNs and high-redshift galaxies ($z \geq 1.3$), they typically required all four colour bands (Stern et al. 2012). WISE is a much shallower survey so it does not suffer from this kind of contamination, hence why we can use a much simpler two band colour selection. Figure 3 depicts the completeness and reliability of different thresholds that can be put on the W1-W2 colour. These values were determined using the IRAC colour candidates of Stern et al. 2005 as the truth sample. Stern et al. 2012 has shown that the criterion $W1 - W2 \geq 0.8$ identifies 61.9 ± 5.4 AGN candidates per deg^2, identifies 78% of Spitzer-IRAC candidates and has a 95% reliability, which is determined by how many galaxies are correctly identified in the truth sample.

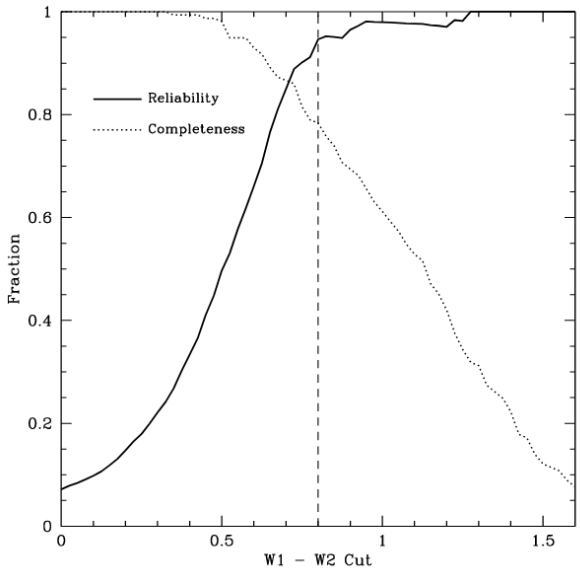


Figure 3: Reliability (solid line) and completeness (dotted line) of WISE AGN selection as a function of W1-W2 colour selection. The completeness is the percentage of Spitzer infrared candidates found by Stern et al. 2005. (Stern et al. 2012)

1.2 Bar activity

Different kinds of galaxies are usually classified according to Hubble's classification scheme. In Figure 4 we can see the scheme, with all the different classification: elliptical types (E0 to E7), lenticulars (S0) and spirals. Spirals are subdivided into "normal" (Sa, Sb, Sc) and barred-spirals (SBa, SBc, SBc). This split is why Hubble's Classification Scheme is also referred to as Hubble's Tuning Fork.

We often refer to elliptical and lenticular (S0) galaxies as "early-type" galaxies and to spirals as "late-type" galaxies. We also refer to late- and early-type spirals depending on whether they are on the right or left side of the scheme. However it is important to note that the Hubble diagram is not an evolutionary diagram. The naming convention does not have a correlation to the evolution stage of the galaxy.

While not present in all spiral galaxies, bar can help with the evolution of spiral arms. Simulations have show that a rigidly rotating bar in disk galaxy can create spiral arms without any other external triggers. In these models, the gas in the disk would eventually reach a steady-state with a trailing spiral structure (Mo, Bosch, and White 2010).
As for the formation of the bar itself, there are two main theories: formation due to gravitational instabilities in the disk or through tidal galaxy encounters which serve as an external trigger (Noguchi 1996).

Bars do not only serve as triggers for spiral arm formation, as they are also theorised to help trigger AGN activity. Super Massive Blackholes (SMBHs), the ones we find at the centre of most galaxies, are likely formed soon after bulge formation and are left-over from past quasar activity. In order to see any new nuclear activity from here there needs to be an inflow of new matter into the SMBH. We thus need a mechanism to remove the angular momentum from the gas in the ex-

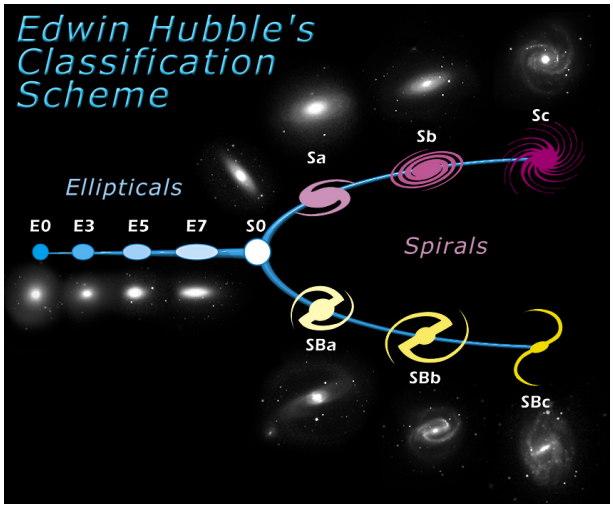


Figure 4: The Hubble sequence or Hubble tuning fork depicts a classification of galaxy morphology. We first have the elliptical galaxies and then the spiral galaxies split off into barred spirals and "normal" spirals.

ternal regions so that it can flow inwards. The presence of bars in a galaxy can cause an angular momentum exchange that can transport kpc-scale gas down to merely parsec scales, close to the galactic nucleus (Laurikainen, Salo, and Buta 2004; Goulding et al. 2017; Silva-Lima et al. 2022).

1.3 Bar and AGN connection

As mentioned before, the relation between bars and AGN is not well understood, with studies showing contradictory results. Some studies find that bars help trigger AGN activity, while other studies do not find a connection between the two.

For example, Lee et al. 2012 looked at a volume-limited sample of ∼9000 galaxies from the SDSS 7th data release at low redshift ($0.02 \leq z \leq 0.055$). At first glance they found that the bar fraction in AGN type galaxies was ~ 2.5 times higher than in non-AGN galaxies. Additionally they saw that the AGN fraction in galaxies with strong bars was roughly twice as high as the fraction in non-barred galaxies. However, they explained this trend through the fact that the fraction of strong bars increases with (u-r) colour and for more massive galaxies (related to a higher velocity dispersion σ) and that AGN host galaxies are on average more massive and redder than non-AGN galaxies. When these fractions were studied at a fixed (u-r) and σ, no correlation was found (see fig 5).

Cisternas et al. 2015 looked at a sample of 95 AGNs over a redshift range of $0.15 < z < 0.84$. They found that the bar fraction of AGN host galaxies decreases from $71 \pm 10\%$ at $z \sim 0.3$ to 35 ± 7 at $z \sim 0.8$, which is a similar evolution as in inactive galaxies. They also found that the strength of the AGN activity is not related to the presence of a bar. They finally conclude that AGN activity is independent of bars in both their occurrence and strength.
Both Lee et al. 2012 and Cisternas et al. 2015 acknowledge that bars may indeed play a role in funneling gas into the central regions but it is not the driving mechanism behind AGN activity.

Like Lee et al. 2012, Silva-Lima et al. 2022 also used SDSS data to evaluate the connection

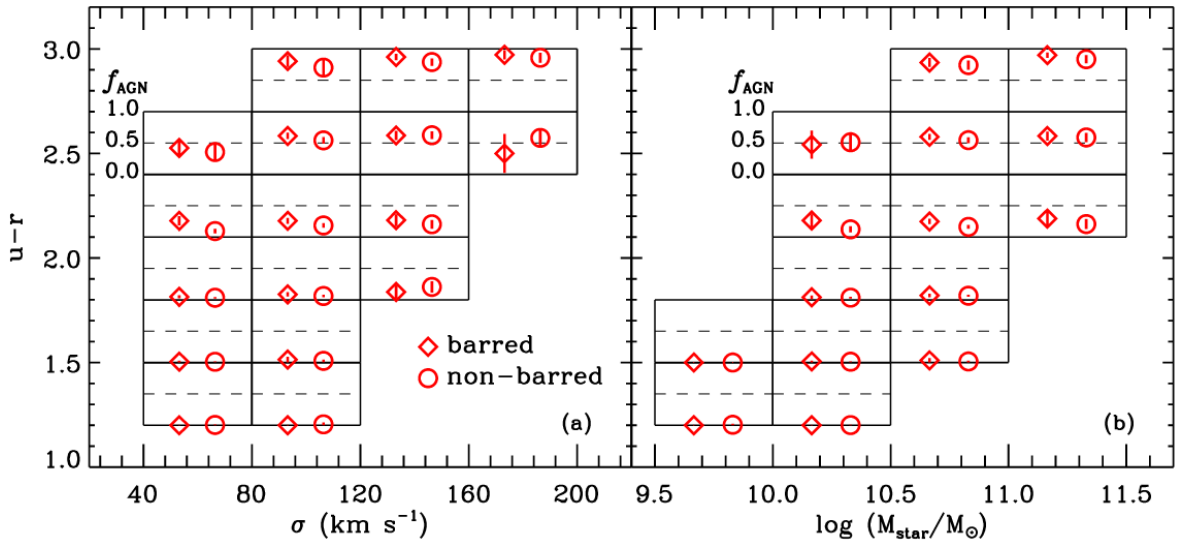


Figure 5: Dependence of fraction of AGNs on the bar fraction for fixed ranges of (u-r) colours and the velocity dispersion (σ) or the mass. (Lee et al. 2012)

between the presence of bars and AGN activity over a redshift range of $0.02 \leq z \leq 0.07$. They look at a sample with the same distribution of a number of parameters, amongst which the stellar mass of the galaxy M, the stellar mass of its bulge $M_{*,bulge}$ and (g-r) colour. In their work they found that the AGN fraction in barred galaxies is higher than that in non-barred galaxies. They also found that the accretion parameter tends to be higher in barred galaxies than in those without a bar, which is what we expect to see if bars are indeed a driving mechanism for gas transport to the central regions. However, no connection was found between bar-strength and AGN activity level. This could mean that bars are responsible for bringing gas to the center of the galaxies, building a gas reservoir but a different mechanism would be responsible for the feeding process on parsec scales.

S. Oh, K. Oh, and Yi 2011 evaluated the connection for nearby, bright galaxies ($0.01 \leq z \leq 0.05$) ($M_r < -19$). They inspected the effect of bars for fixed galaxy properties such as stellar mass, black hole mass, gas contents, etc. In addition to a clear connection between bars and AGN activity, they also found evidence that the bar strength also affects the formation of AGNs. They found a higher AGN fraction in galaxies with a longer bar. The effect of bars also seemed to increase in bluer and less massive galaxies.
Alonso, Coldwell, and Lambas 2013 got results that agree with the findings by S. Oh, K. Oh, and Yi 2011. They used a selection of AGNs from the SDSS Data Release 7 to inspect the bar fraction. They found a bar fraction of 28.5%. Like Silva-Lima et al. 2022, they also found that the effect of bars seems to increase in bluer, less massive galaxies. Finally, they also found that barred galaxies show a higher accretion rate onto the central black holes than there non-barred counterparts.

As mentioned before, some of these results are very contradicting. All of these studies except Cisternas et al. 2015 use galaxies from the Sloan Digital Sky Survey and look at a similar redshift range. Lee et al. 2012 found that once they fixed the colour and mass of the galaxies, the AGN fraction in barred and non-barred galaxies became very similar. However, Silva-Lima et al. 2022 and S. Oh, K. Oh, and Yi 2011 also fixed galaxy properties such as the stellar mass

and the colour but still found a connection between bars and non-bars.
Silva-Lima et al. 2022 and Cisternas et al. 2015 also mentioned that they did not find a correlation between bar-strength and the central black hole accretion rate, while Alonso, Coldwell, and Lambas 2013 did find that stronger bars show a higher accretion rate.

One thing that all of the previously mentioned studies agree on is that f_{agn} in bars seems to decrease for redder galaxies. Cisternas et al. 2015 is the only one of these studies to look at a redshift beyond $z = 0.1$. The other studies made use of SDSS, which is not a deep enough survey for identifying galaxy mythologies or AGN hosts at higher redshift. The Hyper Suprime-Cam SSP Survey is a much deeper survey that is able to provide higher resolution images of galaxies at higher redshifts. In this study we will thus be using HSC images to explore the AGN-bar connection at higher redshifts.

2 Databases

This thesis uses data from the Hyper Suprime-Cam SSP Survey. We use two separate data set, one for training the machine learning algorithm and one for evaluating the AGN-bar connection.

For the training sample we want to construct a data set of galaxies for which we can obtain their morphological classification. For the morphological classification we use a Galaxy Zoo 2 catalogue. The final training set comes from HSC sources that match with the Galaxy Zoo 2 sources.
The other HSC sample has to consist of galaxies for which we can obtain an AGN classification. For this we create two separate samples. The first sample consists of matches with an SDSS catalogue which has classifications for optical AGNs. The second sample is matched with the WISE all-sky survey for which we use mid-infrared colour selection to classify AGN host galaxies.

In the following sections we will expand upon the SDSS, HSC and WISE surveys, as well as the Galaxy Zoo 2 project.

2.1 SDSS and Galaxy Zoo 2

The Sloan Digital Sky Survey (SDSS) is a spectroscopic and imaging survey that covers a third of the celestial sky (Stoughton et al. 2002). This survey uses a 2.5 meter telescope with a large-format mosaic CCD camera that does imaging in five broad bands (*ugriz*) and has two digital spectrographs to obtain the spectra of a million galaxies and 100 000 quasars (York et al. 2000). The five filters have design effective wavelengths of 3550Å, 4770Å, 6230Å, 7620Åand 9130Åre-spectively. The detection limit for point sources in 1" seeing was estimated to be reached at magnitudes 22.3, 23.3, 23.1, 22.3, and 20.8 in the respective u',g',r',i',z' filters (York et al. 2000).

Brinchmann et al. 2004 uses the 8th data release of SDSS to study the physical properties of the low-redshift universe. This was done using the classification system developed by Baldwin, Phillips, and Terlevich 1981 (BPT) that was discussed in this paper earlier on in Section 2.1. This study uses a subset of Blanton et al. 2003 Sample 10, which consists of 149 660 galaxies with spectroscopic observations, $14.5 < r < 17.77$ and $0.005 < z < 0.22$. The redshift limit leads to a mass limit of $M_* > M_\odot$. All objects were classified according to 6 categories: Star-forming galaxies (SF), Composite galaxies (C), AGN's, Low Signal-to-Noise AGN's (Low S/N LINER), Low Signal-to-Noise star-forming galaxies (Low S/N SF) and Unclassifiable objects (UnClass). This last class mostly consists of galaxies with very weak or no emission lines. The amount of galaxies in each class is shown in Table 1. For this research we will consider the composite, AGN and the Low S/N AGN classes as AGNs. These make up a total of 23.8 % of the sample. The match of this sample with our HSC sample gave us a subset consisting of 2024 galaxies.

Galaxy Zoo 2 (GZ2) (Willett et al. 2013) was released as a follow up of the first data release of Galaxy Zoo. This project provides morphological classifications of galaxies from SDSS. GZ2 classified the morphology of over 300 000 of the brightest and the largest galaxies of SDSS Data Release 7 2009ApJS..182..543A. A known issue from GZ2 is that galaxies at a higher redshift (and thus further away) are scaled up to have a bigger angular size than they appear to have. This results in a lower resolution, which can make detailed features, such as bars, less distinguishable. This leads to biases in the classification of the galaxies (Hart et al. 2016).

The classification made by Galaxy Zoo is determined through a question tree to which vol-

Subsample	Number	Percent
All	146994	100
SF	39141	26.6
C	14372	9.8
AGN	8836	6.0
Low S/N LINER	11752	8.0
Low S/N SF	29115	19.8
UnClass	43778	29.8

Table 1

unteer citizen answer after visual inspection of the images. The question tree consists of 11 different questions about the morphology of the galaxies and has a total of 39 possible responses (Willett et al. 2013). For our purposes the most relevant question was T02: "Is there a sign of a bar feature through the centre of the galaxy". Figure 6 depicts what path of the decision tree leads to this question.

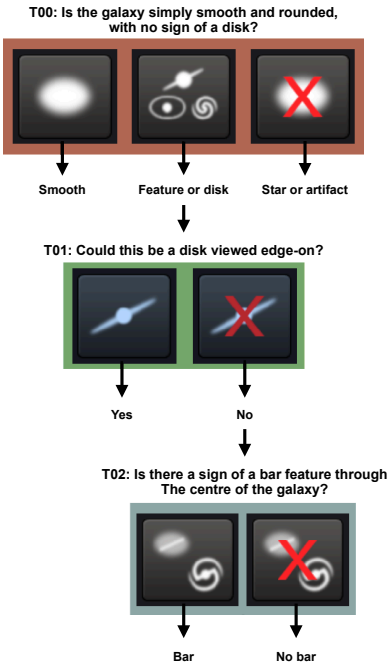


Figure 6: Question tree used in Galaxy Zoo 2. This figure only includes the questions that follow up to T02: "Is there a sign of a bar feature through the centre of the galaxy. The full question tree can be found in Dieleman, Willett, and Dambre 2015.

The catalog from Hart et al. 2016 provides a morphological classification of all the galaxies. However, it also provides a debiased fraction of the given answers for each galaxy, which aims to make the votes distribute themselves as consistently as possible. This catalog allows us to only focus on the questions that are relevant for a specific morphology query and to implement different thresholds. For this research we are only interested in whether a galaxies has a bar feature or not. As mentioned before, this corresponds to question T02. We used the responses to

this question to determine which of the galaxies were barred galaxies and which were non-barred galaxies.
The catalog provides a vote fraction for question T02, and therefore, we need to choose a threshold for the vote fraction, f, to classify them into bars or non-bars. Table 2 shows the number of galaxies classified in each class for different thresholds. Only galaxies that have a counterpart in the HSC survey are included. The threshold labeled as "standard" in this table, is considered as the morphological classification assigned by Hart et al. 2016.
We also checked the numbers for a threshold of f=0.5. If more than 50% of the votes answered "no" to the question, we classified it as a non-barred galaxy and if more than 50% answered "yes", it was classified as a barred galaxy. The unclassified objects for this threshold are likely a result of the debiasing.
A threshold of f=0.8 was also set. The galaxy was considered a non-barred galaxy if more than 80% of the votes answered "no" to the question and if more than 80% answered "yes", it was classified as a barred galaxy.
The threshold of 0.8 filters out the uncertain classifications much stricter than the 0.5 threshold

Classification	standard	0.5	0.8
Bars	1525	1413	637
Non-bars	5241	4387	3500
Unclassified	0	966	2629

Table 2: Number of galaxies in each category for different thresholds on the question: "Is there a sign of a bar feature through the centre of the galaxy?". This was done using the debiased votes from Hart et al. 2016.

but as a result it also loses a lot of galaxies. If a galaxy is not classified it can also not be used for the training data. However, in the case of machine learning it is important to have as little contamination as possible to make the model more accurate. We judged the data set with threshold 0.8 to be large enough to offer sufficient training data. We will thus use the Hart et al. 2016 debiased table with a threshold of f>0.8 on the question T02:"Is there a sign of a bar feature through the centre of the galaxy?" as our method of classification for our training set.

2.2 HSC

The Hyper Suprime-Cam SSP Survey uses a wide field imaging camera, the Hyper Suprime-Cam (HSC), mounted on the 8.2-meter Subaru telescope in Hawaii (Aihara et al. 2017). It uses a total of 9 filters. The range of each filter is shown in Figure 7. There are 5 broad-band filters (g, r, i, z, y) and an additional 4 narrow-band filters. The survey has three different survey depths, for our purposes we will only be looking at the HSC Wide layer. This layer covers 3 target fields with a total coverage of $\simeq$ 1400 deg^2 and has a depth of around $i \sim 26$ mag and 26.1 mag in the r-band (Aihara et al. 2017). In the second public data release the depth improved from 25.9 mag to 26.2 mag (Aihara et al. 2019).

Figure 8 shows a comparison in resolution between SDSS and HSC. The left column shows the SDSS images. This image is obtained as a combination of the the g, r and i filters of the SDSS survey. The right column shows the HSC images in the I-band. These objects were selected on the basis of the vote fraction on question T02 from GZ2:"Is there a sign of a bar feature through the centre of the galaxy?". These galaxies had vote fractions around the 0.5 and would thus not be included in our training set due to uncertainty. The HSC images clearly show a higher

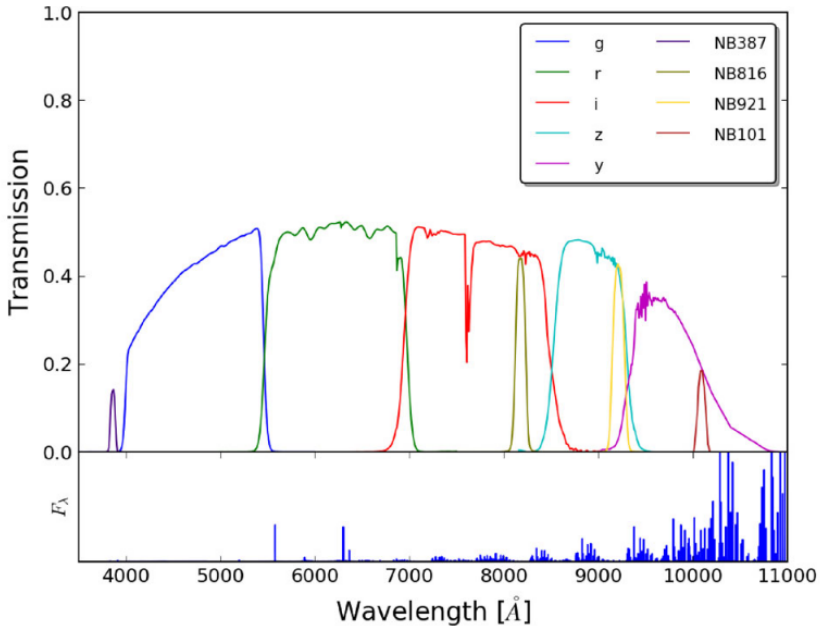


Figure 7: Aihara et al. 2017

resolution and make the galaxy features more identifiable compared to the SDSS images. This higher resolution also allows us to run our model on higher redshift galaxies because the galaxy features should still be distinguishable.

A catalog was constructed from DR2 of the HSC SSP survey consisting of 18 280 galaxies. For this work, we use the GAMA09 field, taken in the wide layer of HSC. The coverage of this field is shown in Figure 9. GAMA09 covers a sky area of ~ 62 degree^2. The catalog has a limiting spectroscopic redshift of $0.1 \leq z \leq 0.55$ and a r-band magnitude between $15 \leq r \leq 24.5$. Figure 10 shows the distribution of the full sample over the redshift and r-band magnitude.

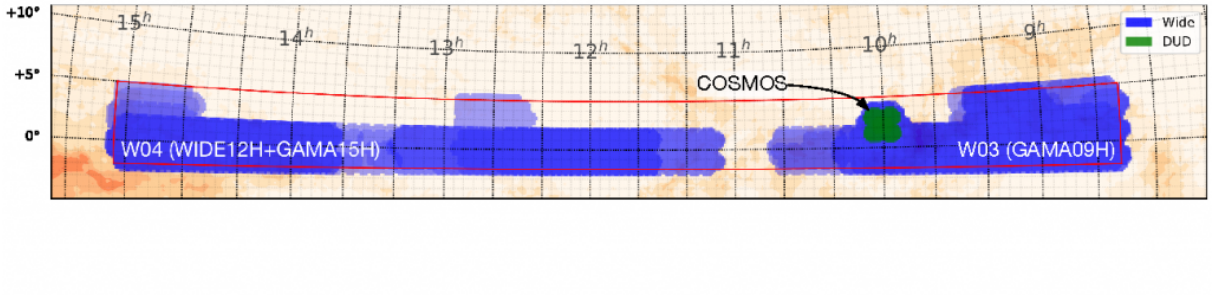


Figure 9: The catalog used for the results consists of galaxies in the GAMA09 field. This field covers an area of ~ 62 deg^2. (Aihara et al. 2019)

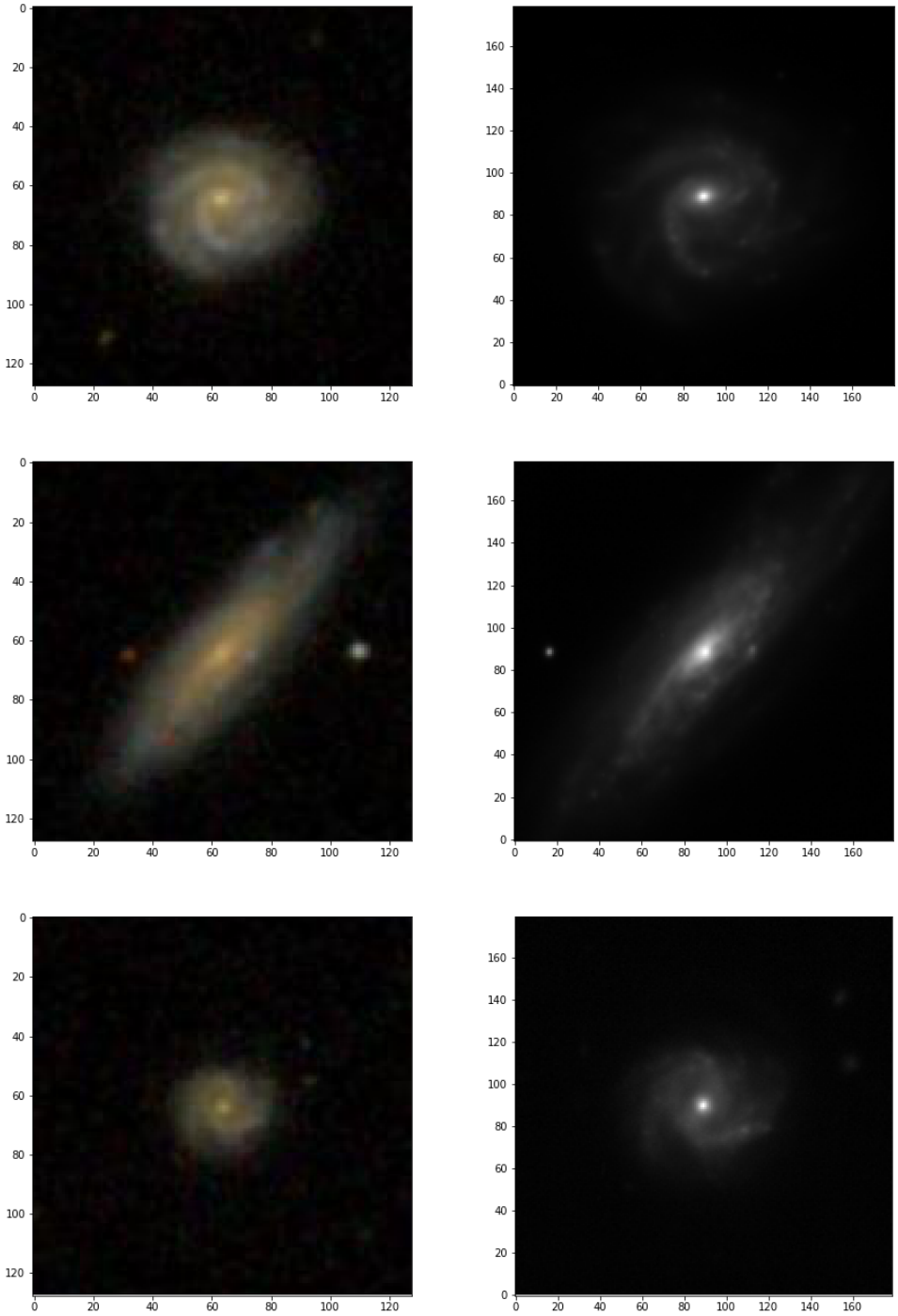


Figure 8: The same objects, shown as an SDSS image and HSC image. On the left shows the SDSS using the g,r and i and on the right the HSC image in the i band.

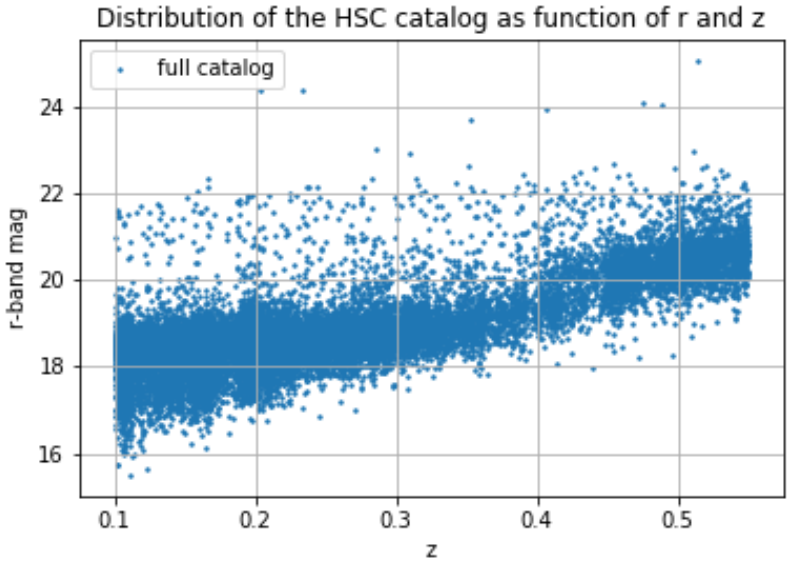


Figure 10: The r-band magnitude vs the spectroscopic redshift for the full catalog.

2.3 WISE

The Wide-field Infrared Survey (WISE) completed it's first All-Sky survey in July of 2010. It covered 95% of the sky with a minimum coverage of 5 exposures per positions over 4 passbands, W1 (3.5 μm), W2 (4.6 μm), W3 (12 μm) and W4 (22 μm) (Wright et al. 2010). Figure 11 shows a colour-colour diagram of some interesting WISE sources.

As mentioned before Stern et al. 2012 developed a colour criterion for WISE to identify mid-infrared AGNs. We created a HSC sample for MIR AGNs consisting of 17 080 galaxies that have a counterpart in the WISE All-Sky survey. The matching of these objects was done based on their positions (RA and Dec) using a 3" cone radius.

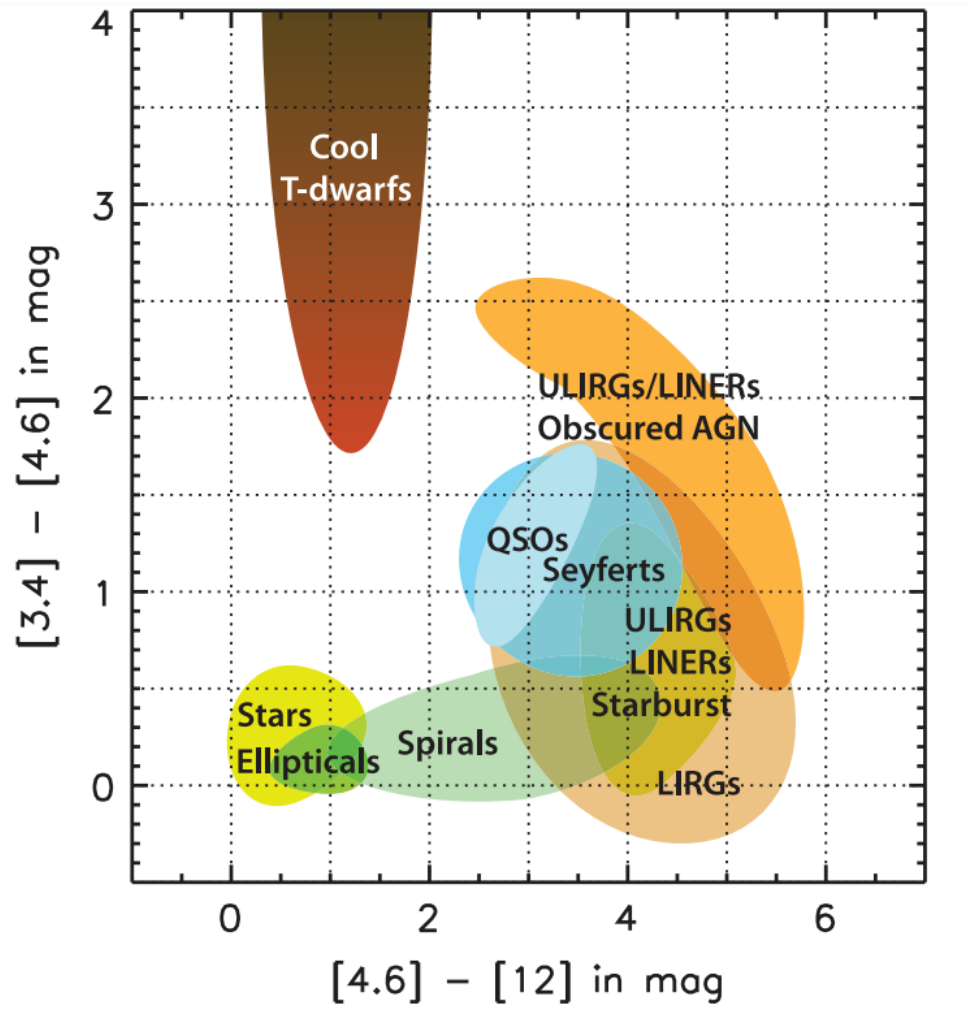


Figure 11: Some of the main sources detected by WISE shown in a colour-colour diagram (Wright et al. 2010).

3 Methodology

Machine learning is a branch of artificial intelligence and computer science to automate the analysing of data. It uses a set of training data to teach a model or algorithm to recognise patterns so that it can learn to make predictions for new data. When training a machine learning algorithm, in addition to a training set, a validation set and a test set is also used. The validation set is used to check the performance of the model and to determine the best set of hyper-parameters for the model. The test set is a set that should only get fed to the model once. This is done when we are satisfied with the training, to evaluate the true performance of the model. It is thus very important that the model has not seen the test set before because we want to know how well the model can predict new data.

There are 4 main types of machine learning: supervised, unsupervised, reinforcement and semi-supervised learning. Here we will be focusing on supervised learning.
Supervised machine learning, as the name suggests, requires some form of guidance. This comes in the form of a annotated data. This can be used for classification or regression problems.

In a regression problem, the algorithm learns to predict a continuous value, for example, the expected sales or for the expected population growth. In a classification problem,on the other hand, the algorithm learn to classify data into a discrete numbers of labels. For example, if the goal is to train a model to be able identify bananas from apples, if you provide the model with a training set of images of these two categories with the corresponding label attached, the model will learn to classify objects under one of these two classes.
This is also the machine learning method used in this research, as we want our images to be classified as a barred or a non-barred galaxy.

3.1 CNN's

Neural Networks are machine learning algorithms that make use of different nodes that are interconnected to predict the desired output from an input. The name comes from the fact that these nodes can be viewed as analogous the the neurons in the brain that interact with each other. The nodes can be arranged in different patterns and layers, this is what we call the architecture of the network.There are three different types of layers: the input layer (which received the input data), the output layer (the output from the network) and, in between those two layers, the hidden layers. These hidden layers consist of mathematical functions with weights attached to them that connect nodes with different layers (Figure 12).

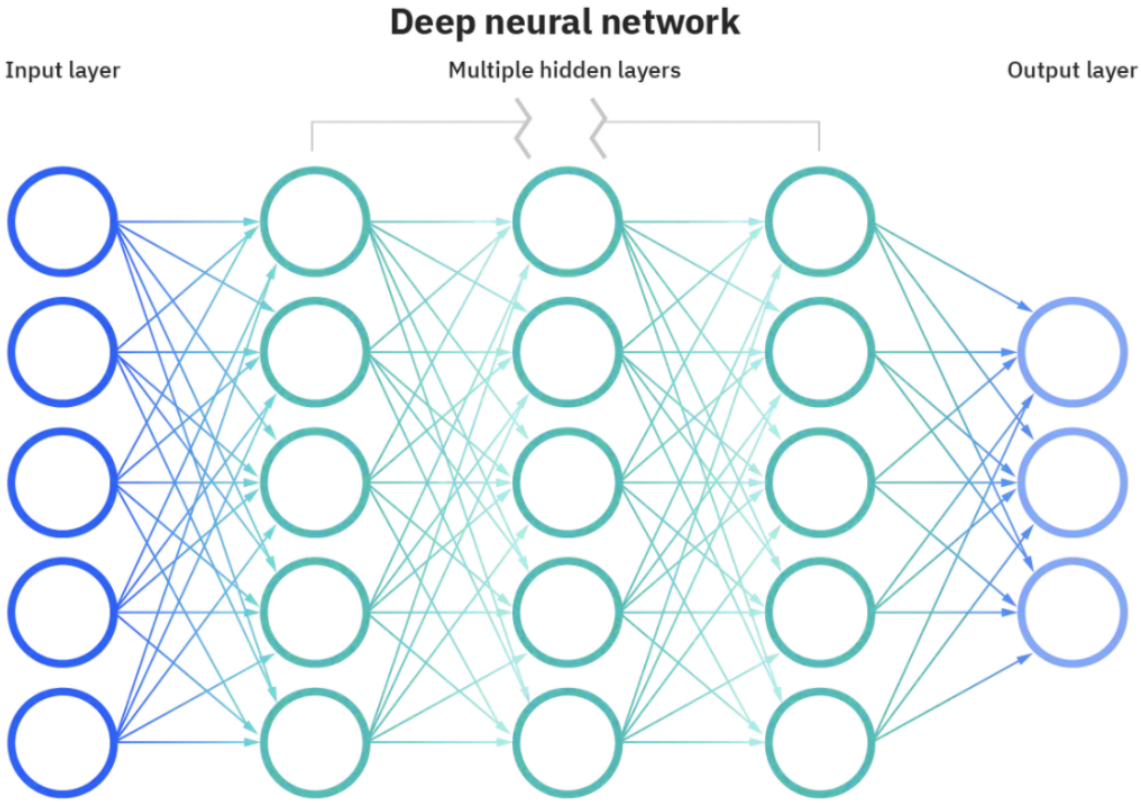


Figure 12: Neural networks have hidden layers connecting the input layer with the output layer. The nodes are interconnected with different weights attached to each connection. Image taken from: https://www.ibm.com/cloud/learn/neural-networks

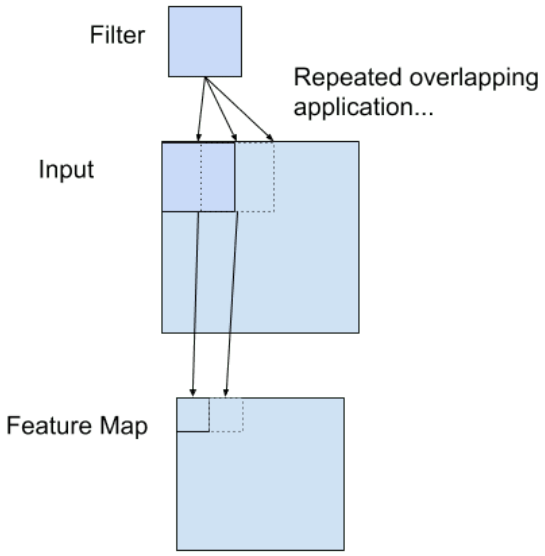


Figure 13: Filters scan sections of the input array to map onto a feature map. It works its way from left to right and top to bottom. Image from: https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/

These weights are learned during training, using an optimization algorithm, to minimize the error in the outputs. The error is determined by the loss function. The loss function is also a mathematical function that can take different forms depending on the problem. For example, the cross-entropy loss (used for classification problems) calculates the cross-entropy between the predicted and true labels, or the mean squared error and mean absolute error (used for regression problems), which compute the mean of squares of errors or the mean of the absolute errors, respectively, between labels and predictions.
In order to quantify the performance of the model, we need to define and use a metric. The metric function can be similar in form to the loss function, except that the metric function does not get used during the training, while the loss function does. One commonly used metric function is the accuracy, which calculates how often the predictions match the true labels.

If a network has many more hidden layers than a traditional neural network we consider it to be deep neural network. Convolutional Neural Networks (CNNs) are a type of deep learning architectures commonly used for image classification. It gets its name from the use of convolutions layers.
Convolutional layers are composed of several filter. These filters perform the convolutional operation over the input of the layer, and are able to detect certain patters (depending on the filter), that is why convolutional layer are said to be feature extractors. The filters are smaller than the input data so that it can "scan" sections of the input array multiple times. It works its way over the input data from left to right and from top to bottom (see Figure 13). The filter itself consists of different weights that will map the input data to the output data. The value of these weights is to be determined by the optimization algorithm.

A convolutional layer with a single filter will map a 2D array to another 2D array. However, it

is common that each convolutional layer has several filter, each of them detecting a particular feature in the input map. The outputs of the filters will then be stacked together to create a 3D array, with as many layers as there were filters.
A filter also does not have to move over the array 1 pixel at a time. Depending on the "stride" of the filter can take bigger steps. A bigger stride will result in a smaller output array.

In order to reduce our array to one single value (the final result), we can also make use of flattening and dense layers. Flattening will arrange all the data points as a 1D array. Dense layers change the dimensions of the array. This is done through matrix multiplication, if we multiply an array with dimensions m x n and a dense layer with dimensions n x k, the resulting array will have the dimensions m x k. Dense layers are the classification part of the CNN. We thus use these as final layers to obtain a prediction.

Training a CNN from scratch can be a very resource intensive process and requires a lot of training data. If a training set is too small or the different categories are not balanced, it is possible to make use of data augmentation to increase the amount of data. Data augmentation is the process of performing transformations on the existing data in order to increase the data set without needing new data. For images this can for example be image rotation, flipping vertically and horizontally.

CNNs have shown to be very useful in the field of astronomy. For example, Davies, Serjeant, and Bromley 2019 used CNNs to identify gravitational lenses in images with a completeness of 77%. Cavanagh, Bekki, and Groves 2021 used them to classify galaxy morphology of elliptical, lenticular, spiral and irregular galaxies. They constructed 3- and 4-way architectures to obtain 83% and 81% accuracy respectively. But most interestingly, they also used binary classification between the different galaxies to obtain a best result of 98% accuracy between ellipticals and spirals, and a lowest accuracy of 78% between irregulars and spirals.

3.2 Training

In this work we use a CNN architecture for a supervised classification problem. The goal is for the model to learn to classify galaxy images as "barred" or as "non-barred". For this kind of binary classification the model will return any value between 0 and 1, with 0 and 1 corresponding to the different classes. Because we have a supervised problem, we will need already labeled training data. We obtained these labels from the debiased table from Galaxy Zoo 2 created by Hart et al. 2016. This table was matched with HSC objects, resulting in 6766 matches.The matching was done using Topcat, where we matched the table on their RA and Dec coordinates with a maximum error of 1 arcsec.
In order to then clean up the data we used a threshold of 0.8 on T02 (see fig 6) to find our most confident barred galaxies (where $\geq$ 80% answered "Yes" to T02) and our most confident non-bars (where $\geq$ 80% answered "No" to T02). This threshold gives us 637 barred-galaxy samples, 3500 non-barred galaxy samples and thus 2629 galaxies that do not fit either of these classes, these will be labeled "other". This split of the labels can also be seen in Table 3.

In order to classify galaxies into three classes using a binary classification model, we use the approach of "one vs all", in which three separate models are trained. Each model learns to classify each class vs all the rest. For that reason, we create 3 training sets: bars vs all, non-bars vs all and other vs all. These 3 different models allow us to identify the most confident bars, the most

Classification	Number	Percent
All	6766	100
Bars	637	9.4
Non-Bars	3500	51.7
Other	2629	43.3

Table 3: Division of the full data set used. The second column shows the total amount of images found for each label and the third column shows the percentage with respect to the full set. The bars and non-bars where determined using the Hart et al. 2016 table with a threshold of 0.8 on the question "Is there a sign of a bar feature through the centre of the galaxy".

confident non-bars and the galaxies that are not certain. All three models will then return a value between 0 and 1, with 0 corresponding to "bar","non-bar" or "other" and 1 corresponding to the complementary sample ("all"). The final classification can then be obtained by taking the lowest of the three values, which gives us the most confident classification.

The training, validation and testing set are split as 81%, 10% and 9% respectively. This was done in python by taking a random set of indices using the build in python function "random.sample()".

To prepare our training data for the model, we first normalise the images so that all pixels have a value between 0 and 1. Any possible dead pixels are also replaced by the median value of the image. We also make sure that all the images have the same size. They are all cut to have the dimensions 176x176x1.
Lastly, we use data augmentation to even out the classes. Table 3 shows that the class for "bars" is much smaller than "non-bars" and "other". Especially in our "bars vs all" training set this would give a $\sim 1:10$ ratio. To increase the size of the bars set we use rotations of 90, 180 and 270 degrees and horizontal and vertical flips on our images until we have set that is roughly the same size as the "all" set. The same is done for the "non vs all" and "other vs all" data sets.

The architecture of the model used is depicted in figure 14. As mentioned before, we start with our input data formatted as a (176,176,1) array. The first convolutional layers consists of 32 filters with a filter size (11,11) and a stride of (2,2). This maps our input to a (88,88,32) array. The second layer has 64 filters of size (9,9). The third layers has 128 filters with size (5,5) and the fourth layer has 256 filter of size (3,3). All the layers have a stride of (2,2), use a "LeakyReLu" activation function and a dropout of 0.5 to prevent over-fitting to the training data. After these 4 convolutions the data gets flattened to a one dimensional array. We then have two dense layers to condense our data to first 64 and then 32 pixels. Both of these dense layers use the activation function "relu". Lastly we obtain the output by compressing the data to a single value using a "sigmoid" activation function.

As mentioned before, the weights of the different layers are determined using an optimization algorithm that minimizes the error in our outputs. In our model we use the optimizer known as Adam. This name comes from Adaptive Moment Estimation. It's a stochastic optimization method that only uses first-order gradients, which makes for a computationally efficient optimization that has little memory requirements (Kingma and Ba 2014).
The optimizer also needs a loss function to compute the performance of the model. The loss

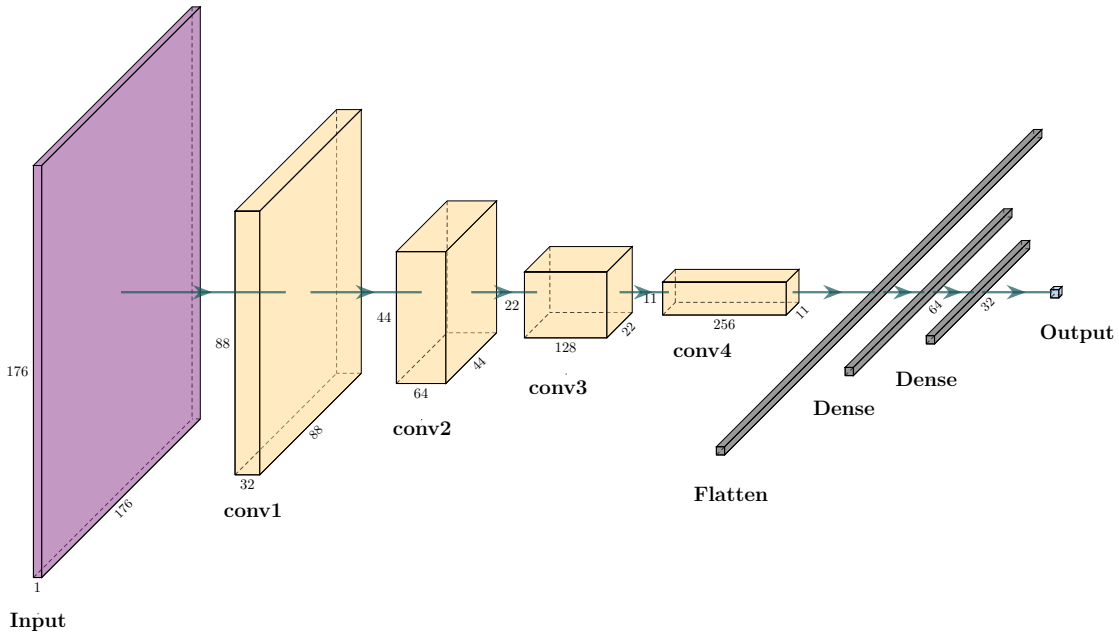


Figure 14: Architecture of the network. We use four different convolutional layers before flattening the array and then using two dense layers to obtain the output

function that is used in this model is the binary cross entropy. This loss function needs the true label (y_true) and the predicted label (y_pred) as inputs. Because we are using binary classification the true label will have a value of either 0 or 1 and the predicted label will take any value between 0 and 1. The cross entropy loss is then calculated using the following equation 1. Where N is the amount of images.

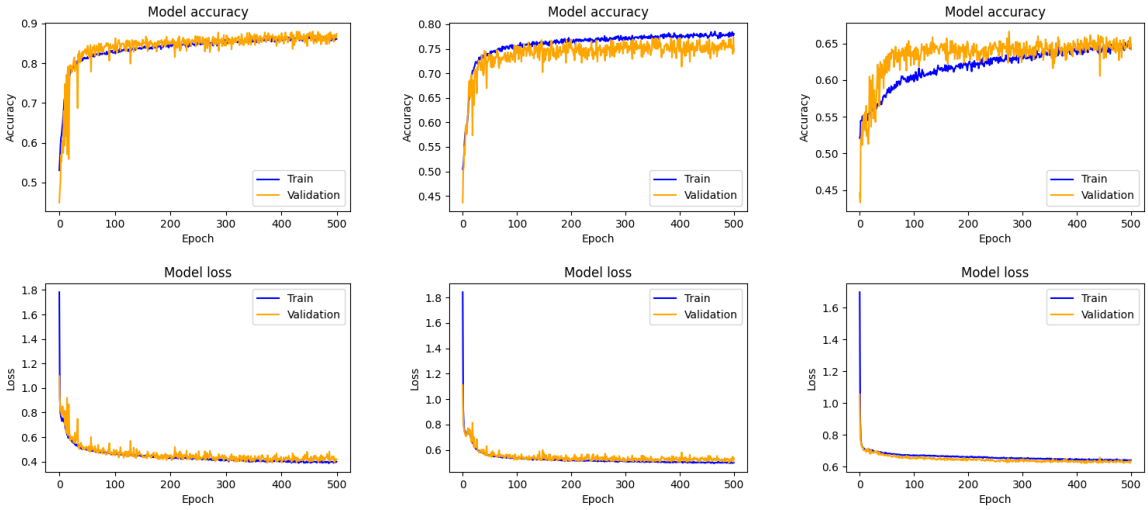
$$Loss = \frac{1}{N}\sum_{i=1}^{N} y\_true \cdot log(y\_pred) + (1 - y\_true) \cdot log(1 - y\_pred) \tag{1}$$

The training for the three models was done using the same architecture, shown in figure 14. The training was done over 500 epochs. The epoch with the best validation accuracy was kept. Figure 15 shows the training of each of the model over all the epochs. For the "bars vs all", the model reached a validation accuracy of 0.87990, the "non vs all" best epoch had an accuracy of 0.77214 and for "other vs all" the maximum accuracy was 0.66643.

Figure 16 shows the confusion matrices for all three models. The main diagonal shows how many images were identified correctly. The accuracy given for each model tells us the fraction of correctly identified images, whereas the precision tells us how many of the "positives" where actually correct. So in case of the "bars vs all" model we have:

$$\text{Accuracy} = \frac{\text{True bars}}{\text{True bars} + \text{False bars}}$$

Where "True bars" are bar images correctly classified by the model and "False bars" are bar images that the model classify as non-bars.
When the classes are not balanced, the precision tends to give a better idea of the performance



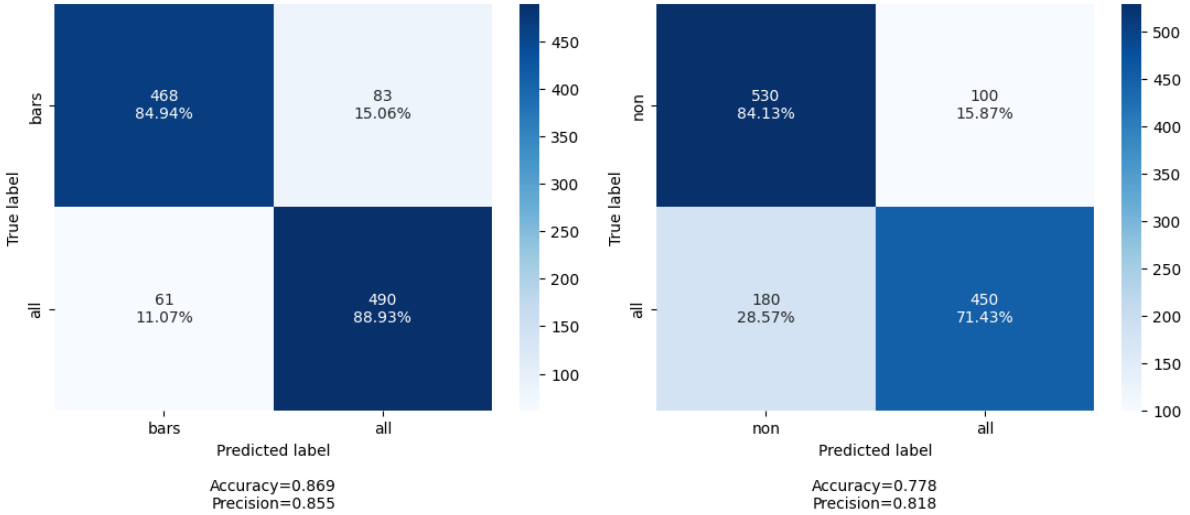
(a) Bars vs all (b) non-bars vs all (c) Other vs all

Figure 15: Training graph for each of the models. All models were trained for 500 epochs and the best epoch was chosen as final model.

than the accuracy. All these values were obtained on the testing set.
Table 4 summarises the accuracy and precision for each model.

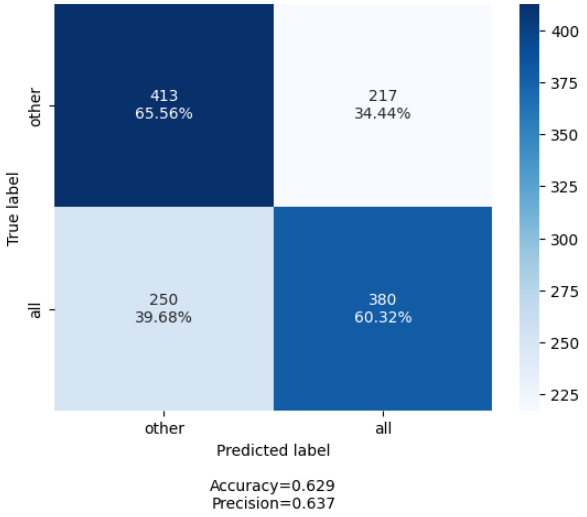
Model	Accuracy	Precision
Bars vs All	0.869	0.855
Non vs All	0.778	0.818
Other vs All	0.629	0.637

Table 4: Final accuracy and precision for the three models



(a) Bars vs all confusion matrix

(b) non-bars vs all confusion matrix



(c) Other vs all confusion matrix

Figure 16: Confusion matrices for the three different binary classifications

3.3 Application

The final models were applied to HSC data. The data set that was used is comprised of galaxies from the GAMA09 field, and are in the redshift range $0.1 \geq z \geq 0.55$. The images were prepared in the same manner as the training data set. First the images were normalised to have values between 0 and 1 and then they were cut to the dimensions (176,176,1).

The classification can be obtained after running all three models for each image, and then selecting the model that gives the closest value to 0, which is the most confident result. However, as shown in figure 16c, the "other vs all" model only has an accuracy of 0.629. For this reason we have decided to not use this model. The classification will thus only be done using the "bars vs all" and the "non vs all" models. The predictions returned by the model are values between 0 and 1. The 0 value denotes the prediction of barred or non-barred galaxy, depending on the model, and 1 is the "all" classification.

If the "bars vs all" returns a prediction <0.5, we classify the galaxy as a barred galaxies and if the "non vs all" returns a prediction <0.5, the galaxy gets classified as a non-barred galaxies. If both models return a prediction <0.5, we take the class that returns the lowest of the two values. After running the models and analysing the predictions, we find 4224 bars and 13106 non bars. We also matched the data points with the used SDSS and WISE catalogs. Table 5 shows how many barred and non-barred galaxies we find in each survey. Figure 17 shows a random set of galaxies that were identified as bars.

Classification	HSC	WISE	SDSS
Bars	4224	3945	454
Non-Bars	13106	12247	1463

Table 5: Amount of bars and non-bars the CNN model found in the HSC images. The WISE and SDSS columns show the amount of matches found with HSC for each of the classifications.

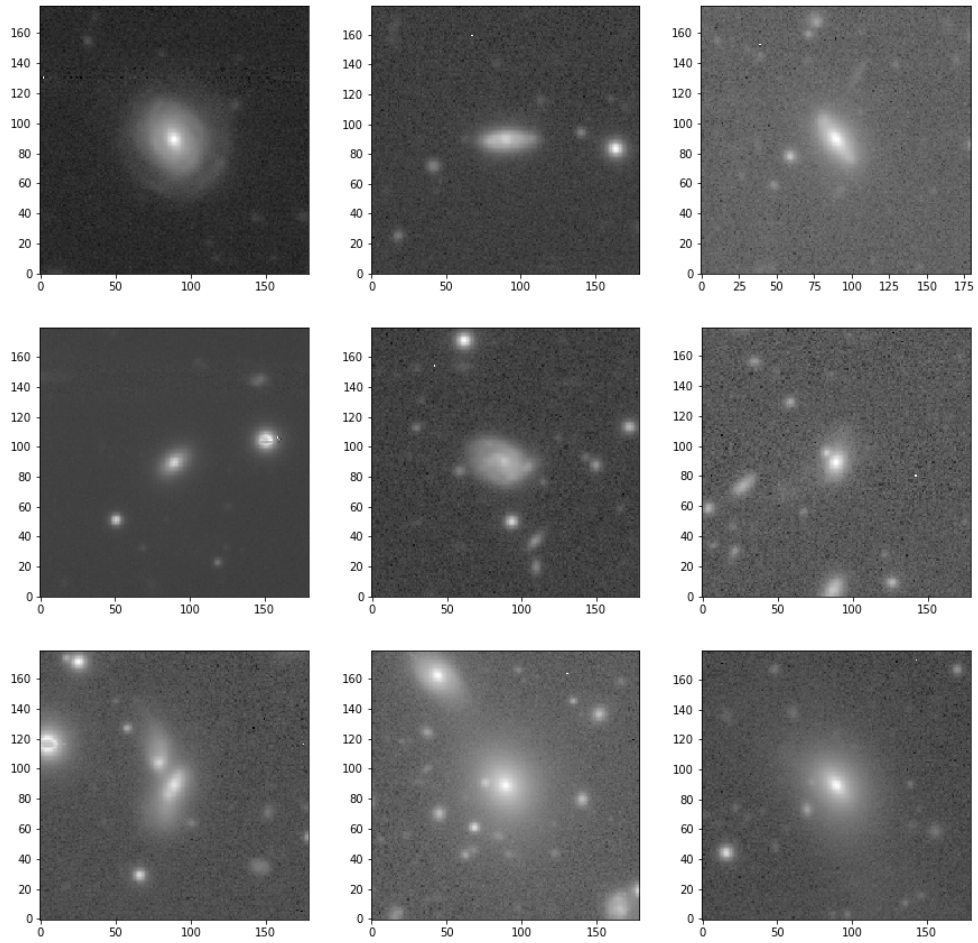


Figure 17: Examples of galaxies that were classified as bars

Figure 18 shows how the classified galaxies are distributed over redshift and r-band magnitude. Most of the galaxies are at the lower end of our redshift range and we see a slight "dip' around $z \sim 0.4$. The magnitude peaks around $r \sim 19$ and we see that there are hardly any data points in the tails of the r-band distribution. The training set data has a limiting magnitude of $r \leq 18$, so there are not many overlapping data points.

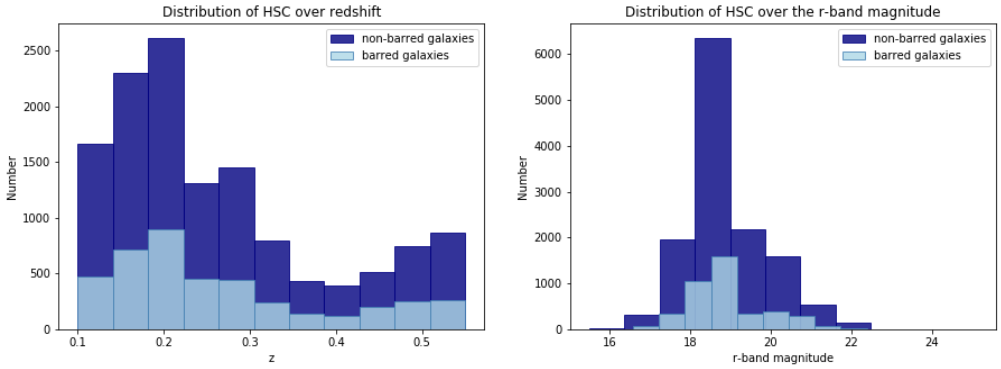


Figure 18: Distribution of the HSC barred and non-barred galaxies plotted over the redshift and the r-band magnitude.

The bar fraction found in the HSC data is shown in figure 19. This fraction was calculated using equation 2. To calculate this fraction we only use the number of most confident bars and most confident non-bats to avoid any contamination with unidentified barred galaxies.

$$f_{bar} = \frac{N_{bar}}{N_{bar} + N_{non}} \tag{2}$$

We notice a very clear downwards trend as we go to brighter magnitudes. Since the model was trained for lower magnitudes it is expected that it has trouble identifying bars in fainter galaxies. While less apparent, we can also identify a similar downwards trend for decreasing redshift. This could be due to the fact that the training set had a limiting redshift of $z \leq 0.25$, thus for higher redshift the models reliability goes down due to the difference in the redshift

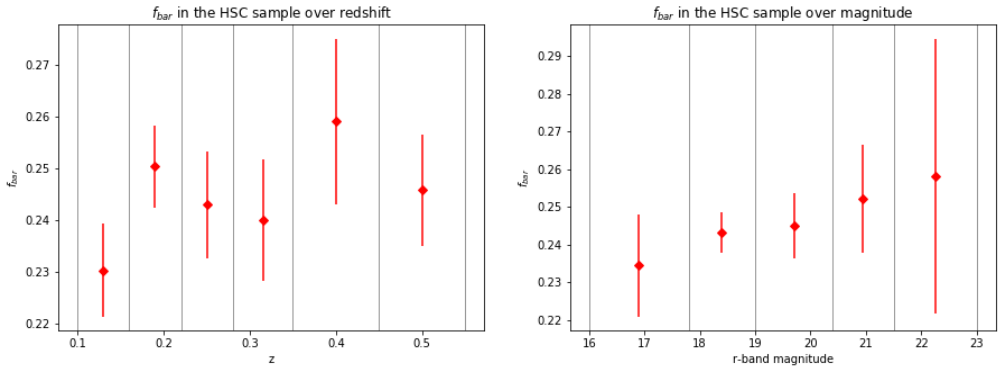


Figure 19: The fraction of bars plotted over the redshift and the r-band magnitude. The fraction of bars is defined by $\frac{N_{bar}}{N_{galaxies}}$.

4 Results

Here we will evaluate if we can identify any link between the presence of a bar in a galaxy and AGN activity. We will look at the AGN fraction in the most confident bars and the most confident non-bars that our model found. We only compare our most confident sub-classes rather than for the whole catalog to minimize any contamination.

AGN fractions are obtain using two methods for AGN selection. The first one (optical selection) selects AGN using the BPT diagram on SDSS data, and the second (mid-infrared selection), identifies AGNs from the colour criterion $W1 - W2 \geq 0.8$ applied to WISE data. The optical and mid-infrared AGNs will be evaluated separately.

4.1 AGN fraction

The AGN fraction over the full redshift and magnitude range is shown in Table 6. The errors in the fractions are obtained through basic error propagation and considering the error in the number of AGNs and the numbers of bars to be $\sqrt{N}$, where N is the corresponding galaxy count. At first glance these fractions are very close together. They do not indicate that bars influence

AGN classification	Barred galaxies	non-barred galaxies
WISE	$1\% \pm 0.1\%$	$1.2\% \pm 0.1\%$
SDSS	$13.2\% \pm 1.8\%$	$14.1\% \pm 1.0\%$

Table 6: AGN fraction found barred galaxies and non-barred galaxies. AGN classification was done using the WISE colour selection $[W1 - W2 \geq 0.8]$ and using the BPT diagram on SDSS emission spectra.

AGN activity, as we then would expected the barred galaxy to show a higher AGN fraction. However, as discussed in section 1.3, the bar-AGN connection is dependant on many different variables. We thus would like to evaluate $_{AGN}$ as a function of redshift and magnitude.

We first explore our results as a function of redshift. We show the distribution of our data set over redshift, depicted in Figure 20, to determine which redshift bins would give an even distribution of data points, and thus an accurate representation of f_{agn}. The resulting fractions are plotted in Figure 21.

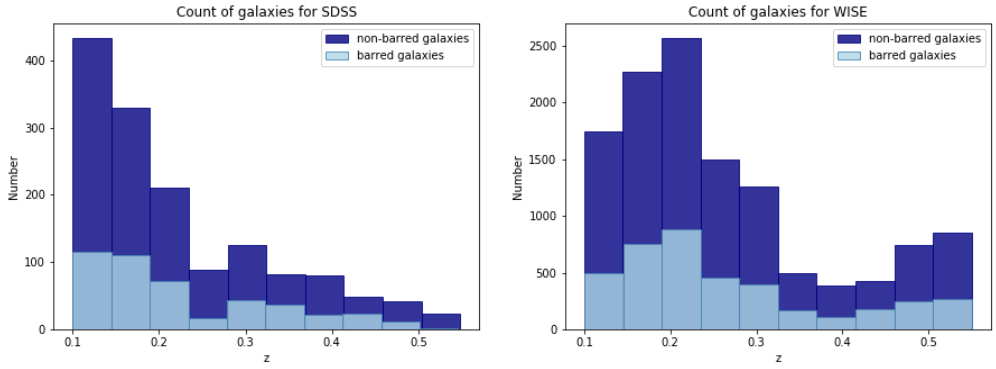


Figure 20: Distribution of data point for WISE and SDSS galaxies over redshift bins

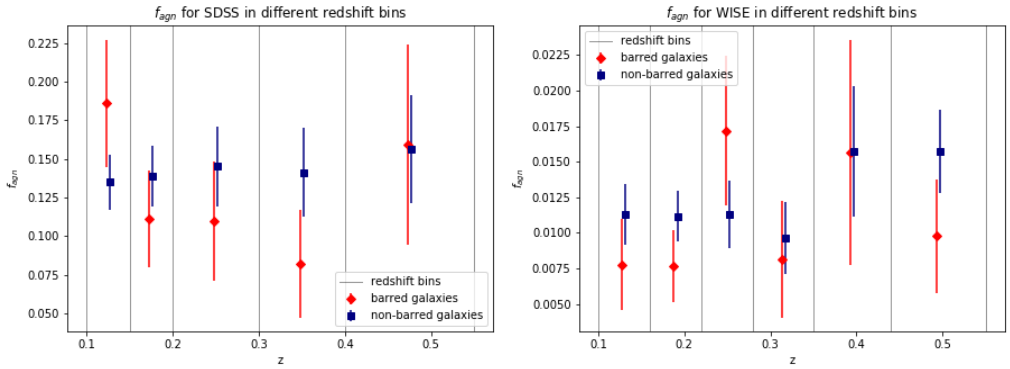


Figure 21: The fraction of AGN activity found in both barred and non-barred galaxies using the BPT diagram on SDSS data on the left and WISE colour selection on the right. The fraction is shown in different redshift bins of varying sizes, dependent on the amount of data points per bin

As we can see, f_{AGN} for barred and non-barred galaxies are very similar, we see this in both SDSS and WISE. However, while the WISE AGNs do not seem to follow a certain trend, for the SDSS AGNs we can see that the AGN fraction in barred galaxies seems to decrease over redshift, while f_{AGN} seems to stay fairly constant. This could be a result of the fact that we are increasing the redshift range compared to our training set but could also be due to different redshift effects. For $z \geq 0.4$ we see this fraction increase again but Figure 20 also shows that there are very few data points in this bin.

Figure 22 shows the distribution of galaxies over r-band magnitude. The WISE catalog does cover a larger magnitude range than the SDSS catalog so we have a few matches for $r \geq 22$, however these are too few data points to give us a realistic view of the AGN fraction at this magnitude range. We have thus elected to only take into account galaxies with $r \leq 22$. The resulting AGN fractions are plotted in Figure 23.

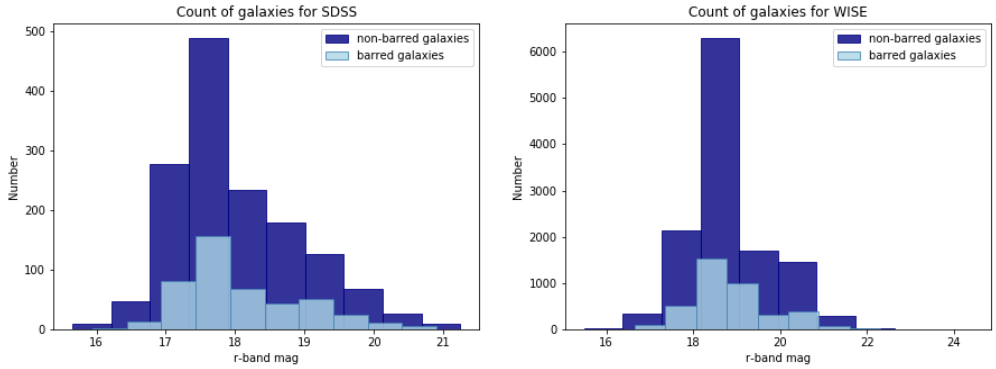


Figure 22: Distribution of data point for WISE and SDSS galaxies over magnitude bins

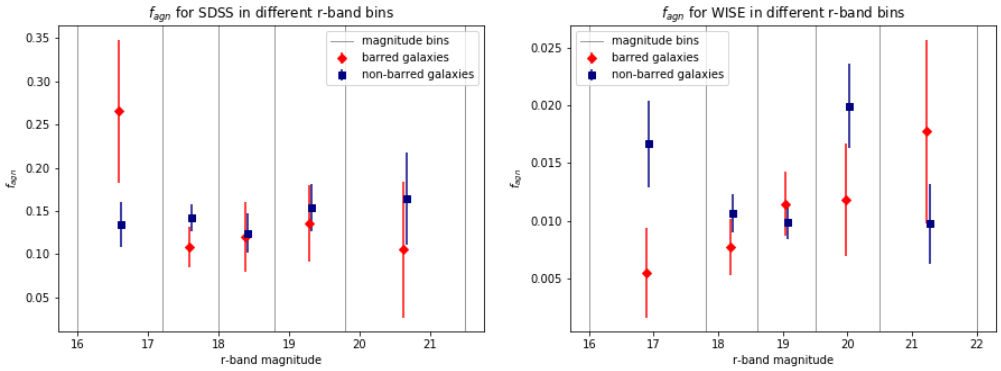


Figure 23: The fraction of AGN activity found in both barred and non-barred galaxies using the BPT diagram on SDSS data on the left and WISE colour selection on the right. The fraction is shown in different magnitude bins of varying sizes, dependent on the amount of data points per bin. The bin in WISE for $g \geq 22$ was excluded due to a lack of data points.

The SDSS data seems to show a similar trend as in Figure 21. All the fractions lay relatively close together but we notice that f_{agn} for the barred galaxies decreases for fainter magnitudes. The WISE selection seems to show an opposite correlation. In this plot the AGN fraction in non-barred galaxies does not follow any trend but the fraction for barred galaxies increases for higher magnitudes.

In Table 1 we see that the SDSS catalog has quite a few unclassifiable objects. This class mostly consists of galaxies with weak or no emission lines. Because we are looking at relatively high redshifts and magnitudes compared to the SDSS sample, it is very likely that this class dominates in our data set. Hence we also show the AGN fraction of SDSS excluding the unclassifiable objects in Figure 24. The AGN fraction in this case is calculated by taking the number of identified AGNs over the sum of the number of AGNs and star-forming galaxies: $f_{AGN} = \frac{N_{AGN}}{N_{AGN}+N_{SF}}$. The composite galaxies are only counted in the number of AGNs to avoid counting them twice.

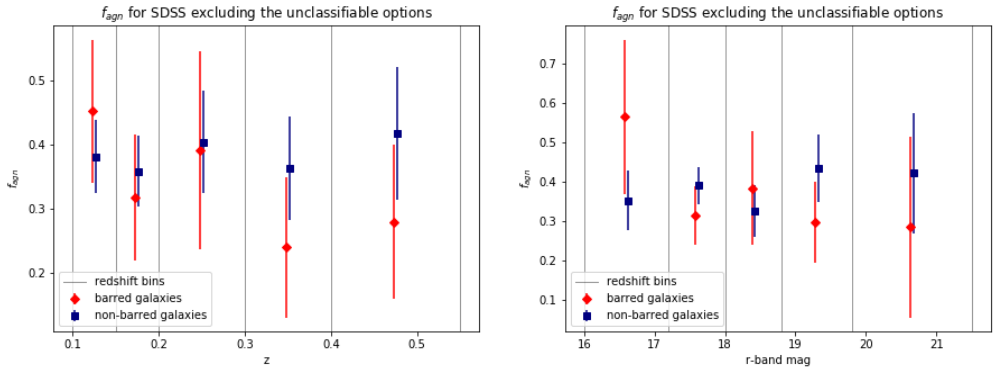


Figure 24: f_{AGN} for SDSS data where the unclassifiable options where excluded. The fraction is calculated as $f_{AGN} = \frac{N_{AGN}}{N_{AGN}+N_{SF}}$. The composite galaxies are only included in N_{AGN} here. On the left f_{bar} is plotted over the redshift and on the right it is plotted over the r-band magnitude

The fraction of AGNs as a function of redshift shown in Figure 24 have similar trends as Figure 21. The AGN fraction in the non-barred galaxies seems to stay fairly constant around $f_{AGN} \sim 0.38$, whereas the barred galaxies show a lot more variation over the different redshifts. We again notice a downwards trend with the bin $0.2 \leq z \leq 0.3$ being a slight outlier, suggesting that barred galaxies might be more susceptible to redshifts effects.
Figure 24 also shows that the trend shown by f_{AGN} as function of magnitude is fairly consistent with 23. Here too, we notice that f_{AGN} decreases for fainter magnitudes.

To summarise, we find that the overall fractions of AGN activity in barred galaxies and non-barred galaxies give no indication that bars trigger AGN activity. This was further confirmed by evaluating f_{AGN} as a function of redshift and as a function of magnitude. None of these plots showed that the fraction of AGN was higher for barred galaxies.
SDSS did seem to show that the AGN fraction in bars decreases for higher redshifts, indicating that barred galaxies might be subject to redshift effects.
The SDSS data also shows f_{AGN} decreasing for fainter magnitudes. However, the WISE data seems to shows the exact opposite effect. For this data set the AGN fraction increases for fainter magnitudes.

5 Discussion

After presenting our results, we will now take some time to discuss our findings. First we will discuss some of the caveats of this research. We'll discuss the limitations of the databases that were used and the caveats of our models. We will also discuss the implications of our finding and compare this with previous studies. Finally, we will talk about the future work to expand and improve this work.

5.1 Data and model limitations

One of the aims of this research was to improve the depth at which bar identification is done. While the HSC SSP survey indeed has a better resolution and is deeper than the Sloan Digital Sky Survey, the identification at higher redshifts has still proven to be difficult.
The model "bars vs all" has the highest accuracy of the three models. Yet it only has an accuracy of 86.9% and a precision of 85.5%. This means that out of the 4224 identified bars, ~ 612 are likely misidentified as bars. Because we compare the predictions with the model "non vs all", we hopefully will have caught some of these misidentifications. However, this model also only has a precision of 81.8%. In the extreme case where non of the misindentifications are caught by the other model, we would be dealing with 14.5% and 18.2% contamination in the bars and non-bars sample respectively.
Additionally, the model was trained using a sample with a limiting redshift $z \leq 0.25$ and magnitude $m_r \leq 17$. However, the HSC data has a limiting redshift $z \leq 0.55$ and limiting magnitude $16 \leq m_r \leq 25$. The model is thus run for higher redshifts and fainter magnitudes than it was trained with. In section 3.3 we discussed that the bar fraction is not consistent over an increasing redshift. We would expect to see f_{bar} decrease for higher redshifts (Sheth et al. 2008; Melvin et al. 2014). However, our model shows a slight increase in f_{bar} for higher magnitudes. We thus note that our model becomes less reliable for higher redshifts and fainter magnitudes.

Figure 17 and the appendix show examples of galaxies that have been identified as bars. These images show quite a few objects that would not be classified as a barred-galaxy through visual inspection. It would be possible to filter out some of these contaminations manually by doing visual inspection of the galaxies. However, for our sample size this would be extremely time consuming so due to time constraints we were not able to do this. This does mean that the results should be viewed with this error in mind.

For the MIR-AGN selection the WISE colour criterion $[W1 - W2 \geq 0.8]$ was used. However, the WISE colour selection mainly works well to a depth of $W2 \sim 15.0$ (Stern et al. 2012). Assef et al. 2013 shows that the completeness and reliability of this AGN selection reduces for fainter magnitudes.
Figure 25 shows the distribution of galaxies in the W2. The top figure shows the distribution of the whole WISE sample. This distribution peaks around $W2 \sim 15$ and extends only to a magnitude of ~ 17. The bottom figure shows the distribution of WISE selected AGNs. It shows the full sample but also the AGNs with barred host-galaxies and the ones with non-barred host-galaxies. The full AGN sample shows a similar distribution as the full WISE sample. The non-barred set seems to dominate the lower part of the magnitude range. It still has a significant amount of data points past the limit of $W2 \sim 15.0$, but the majority are below this limit. The barred set, on the other hand, shows very few galaxies under the limit of $W2 \sim 15.0$.
This leads us to conclude that the AGN fraction galaxies in the WISE date is too unreliable to draw any conclusions from. For non-barred galaxies f_{AGN} can still be an accurate representation

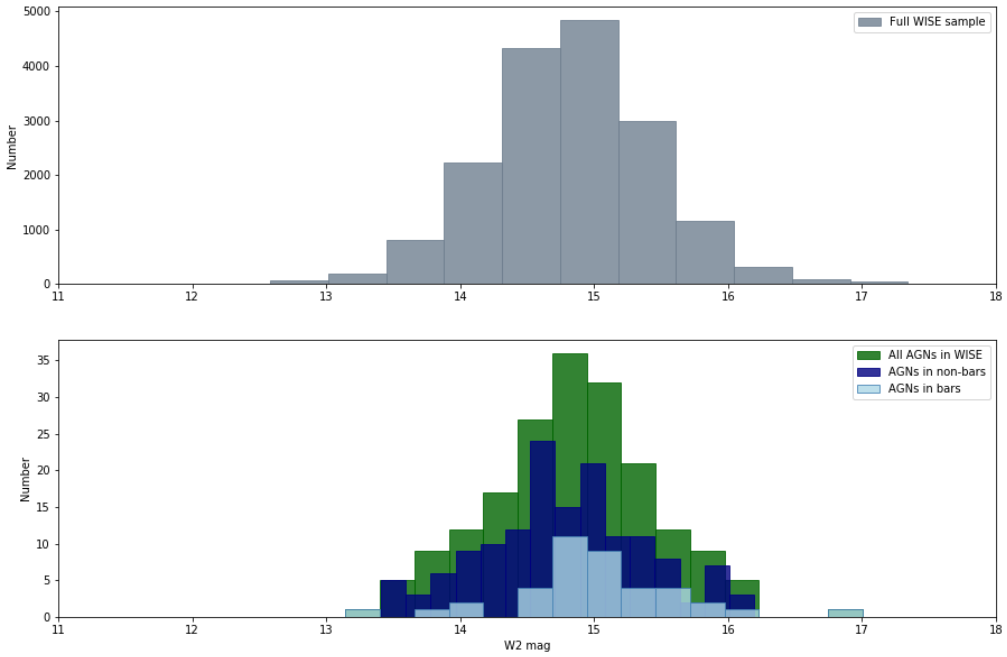


Figure 25: Distribution of galaxies in WISE over the W2 band magnitude. The top plot shows the distribution of the whole sample and the bottom plot shows the distribution of the AGNs that were obtained from the $[W1-W2 \geq 0.8]$ criterion. Here we also show the difference between AGNs in barred and non-barred galaxies.

for the magnitudes $W2 \leq 15.0$ but the barred-galaxies present too few data points under this limit to paint an accurate picture.

5.2 Implications

The fraction of AGNs that we find is significantly lower than what we would expect to find considering the used catalogs and previous research.
Using the same optical selection as we used in this work for the whole SDSS catalog, we find that the AGNs represent $\sim 24\%$ of the whole sample. However, in our results we find an AGN fraction of around 15%. This difference could arise from the fact that we are looking at a relatively high redshift and magnitude. In section 4.1 we explained that this could lead to a higher fraction of unclassified objects. We thus also plotted the results excluding this class. By excluding this class the expected AGN fraction would be $\sim 34\%$. This corresponds fairly well with the fractions found in Figure 24. The lower AGN fraction is thus indeed likely explained by the fact that we are looking at objects with a higher redshift.
We still do not find evidence that bars trigger AGN activity. Although this fraction is only the fraction of AGNs over a sample of AGN and star-forming galaxies rather than the fraction over a broader sample of galaxies.

We also noticed that f_{AGN} decreases for higher redshifts in the barred galaxies. Bar effects

playing a more prominent role for bluer redshifts is consistent with previous work (Lee et al. 2012; Cisternas et al. 2015; S. Oh, K. Oh, and Yi 2011; Alonso, Coldwell, and Lambas 2013). However, this redshift effect could be explained by an unreliability in our model. As mentioned in the previous section, the model becomes more unreliable for higher redshifts and fainter magnitudes. This could lead to more misidentified bars, which would result in a lower AGN fraction.

5.3 Future prospects

This thesis has shown that machine learning is a promising method to re-evaluate the bar-AGN connection at higher redshift. However, there are still many caveats that future studies could improve upon.

One of the main caveats of this thesis is limited obscured AGN classification. Deeper mid-infrared surveys could help improve this. There are also studies that improve upon the already existing criteria, such as Assef et al. 2013. They re-evaluated the WISE [W1-W2] colour criterion to work for magnitudes $W2 \leq 17$.

Future bar-catalogues could help improve the training of the model. Currently comprehensive classifications of galaxy morphology for large scale-structures such as bars is fairly limited to GZ2. Galaxy Zoo:Hubble (Willett et al. 2016) provides morphological classification out to $z \sim 4$. However, the Hubble survey considers deep field rather than wide field and thus only covers a limited amount of sky area. Future surveys or projects such as Galaxy Zoo could greatly improve the available bar catalogues and allow for models that are trained with deeper data.
Furthermore, our model only learned to identify barred and non-barred galaxies. A model that can identify the strength of the bar or can differentiate between elliptical and spiral galaxies would allow to explore these different properties further. The bar strength in particular could have a significant effect on the AGN activity (Silva-Lima et al. 2022,S. Oh, K. Oh, and Yi 2011).

Finally, this thesis only evaluated the AGN fraction as a function of redshift and magnitude. Works such as Lee et al. 2012 or Silva-Lima et al. 2022 also look at galaxy properties such as the stellar mass and colour. Additionally, Kim and Choi 2020 also found the effect of bars also depends on the galaxy environment. A stellar mass limited sample that take into consideration these different properties could tell us much more about the nature of AGNs and bars.

6 Conclusion

In this thesis we used a machine learning algorithm to classify barred and non-barred galaxies. We constructed an HSC sample with redshift limit $0.1 \leq z \leq 0.55$ and an r-band magnitude limit of $15.5 \leq m_r \leq 25$. We used the BPT classification system on an SDSS catalogue to identify optical AGNs. Additionally, we used the mid-infrared colour criterion $[W1-W2] \geq 0.8$ on WISE data to identify obscured AGNs.
We evaluated the AGN fraction in our most confident barred and non-barred galaxies. Here we found an overall AGN fraction of $1\% \pm 0.1\%$ and $1.2\% \pm 0.1\%$ for bars and non bars, respectively for the WISE data and $13.2\% \pm 1.8\%$ in bars and $14.0\% \pm 1.0\%$ in non-bars for the SDSS data. We also evaluated f_{AGN} as functions of redshift and r-band magnitude but none of the plots showed evidence that bars trigger AGN activity. The SDSS sample showed a decreasing AGN fraction in barred galaxies for both an increasing redshift and for fainter magnitudes. The WISE sample did not show any dependence on redshift but showed an increasing f_{AGN} for barred

galaxies for fainter magnitudes. However, this effect was deemed unreliable due to the WISE colour selection not being reliable for fainter magnitudes.
Finally, summarise our conclusions as:
1) We do not see any increase in f_{AGN} in barred galaxies compared to non-barred galaxies. Leading us to conclude that bars do not play a significant role the triggering of AGN activity.
2) We see a decreasing bar fraction for higher redshift and for fainter magnitude. However, this effect is likely caused by limitations in our model.

Acknowledgements

Firstly, I would like to thank everyone who helped with proof-reading this thesis. I really appreciate everyone's help and useful feedback!
Secondly, I'm very grateful to Antonio La Marca helping me out with all the different databases and data sets that I had to use. He helped me a lot with learning how to use the different catalogs.
Most of all, I would like to thank both of my supervisors. I would like to thank Lingyu Wang for being a wonderful supervisor. She helped at every step of the way and was always very helpful when I ran into issues. She helped me a lot with gaining a better understanding of the subject matter and I felt like I could always come to her with questions!
I would like to thank Berta Margalef Bentabol, for having an infinite amount of patience with helping me debug the code. I have spend many hours in her office with a code that was not working or questions about machine learning and she was always very willing to help.
I am very grateful for the help of both of my supervisors. This thesis definitely would not have been possible with out them!

This publication makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration.

The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from the Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University.

This paper makes use of software developed for the Large Synoptic Survey Telescope. We thank the LSST Project for making their code available as free software at http://dm.lsst.org

This paper is based [in part] on data collected at the Subaru Telescope and retrieved from the HSC data archive system, which is operated by the Subaru Telescope and Astronomy Data Center (ADC) at National Astronomical Observatory of Japan. Data analysis was in part carried out with the cooperation of Center for Computational Astrophysics (CfCA), National Astronomical Observatory of Japan. The Subaru Telescope is honored and grateful for the opportunity of observing the Universe from Maunakea, which has the cultural, historical and natural significance in Hawaii.

References

Aihara, Hiroaki et al. (Sept. 2017). "The Hyper Suprime-Cam SSP Survey: Overview and survey design". In: *Publications of the Astronomical Society of Japan* 70.SP1.

Aihara, Hiroaki et al. (Dec. 2019). "Second data release of the Hyper Suprime-Cam Subaru Strategic Program". In: 71.6, 114, p. 114.

Alonso, M. S., G. Coldwell, and D. G. Lambas (Jan. 2013). "Effect of bars in AGN host galaxies and black hole activity". In: *Astronomy &amp Astrophysics* 549, A141.

Assef, R. J. et al. (July 2013). "MID-INFRARED SELECTION OF ACTIVE GALACTIC NUCLEI WITH THEiWIDE-FIELD INFRARED SURVEY EXPLORER/i. II. PROPERTIES OFiWISE/i-SELECTED ACTIVE GALACTIC NUCLEI IN THE NDWFS BOÖTES FIELD". In: *The Astrophysical Journal* 772.1, p. 26.

Baldwin, J. A., M. M. Phillips, and R. Terlevich (Feb. 1981). "Classification parameters for the emission-line spectra of extragalactic objects." In: 93, pp. 5–19.

Blanton, Michael R. et al. (Aug. 2003). "The Galaxy Luminosity Function and Luminosity Density at Redshift z = 0.1". In: *ApJ* 592.2, pp. 819–838.

Brinchmann, J. et al. (July 2004). "The physical properties of star-forming galaxies in the low-redshift Universe". In: *MNRAS* 351.4, pp. 1151–1179.

Burlon, D. et al. (Feb. 2011). "Three-year Swift-BAT Survey of Active Galactic Nuclei: Reconciling Theory and Observations?" In: *ApJ* 728.1, 58, p. 58.

Cavanagh, Mitchell K, Kenji Bekki, and Brent A Groves (June 2021). "Morphological classification of galaxies with deep learning: comparing 3-way and 4-way CNNs". In: *Monthly Notices of the Royal Astronomical Society* 506.1, pp. 659–676.

Cisternas, Mauricio et al. (Apr. 2015). "THE ROLE OF BARS IN AGN FUELING IN DISK GALAXIES OVER THE LAST SEVEN BILLION YEARS". In: *The Astrophysical Journal* 802.2, p. 137.

Davies, Andrew, Stephen Serjeant, and Jane M Bromley (May 2019). "Using convolutional neural networks to identify gravitational lenses in astronomical images". In: *Monthly Notices of the Royal Astronomical Society* 487.4, pp. 5263–5271.

Dieleman, Sander, Kyle W. Willett, and Joni Dambre (Apr. 2015). "Rotation-invariant convolutional neural networks for galaxy morphology prediction". In: *Monthly Notices of the Royal Astronomical Society* 450.2, pp. 1441–1459.

Donley, J. L. et al. (Mar. 2012). "IDENTIFYING LUMINOUS ACTIVE GALACTIC NUCLEI IN DEEP SURVEYS: REVISED IRAC SELECTION CRITERIA". In: *The Astrophysical Journal* 748.2, p. 142.

Goulding, A. D. et al. (July 2017). "Galaxy-scale Bars in Late-type Sloan Digital Sky Survey Galaxies Do Not Influence the Average Accretion Rates of Supermassive Black Holes". In: *ApJ* 843.2, 135, p. 135.

Hart, Ross E. et al. (July 2016). "Galaxy Zoo: comparing the demographics of spiral arm number and a new method for correcting redshift bias". In: *Monthly Notices of the Royal Astronomical Society* 461.4, pp. 3663–3682. ISSN: 0035-8711.

Kim, Minbae and Yun-Young Choi (Oct. 2020). "The Relative Role of Bars and Galaxy Environments in AGN Triggering of SDSS Spirals". In: *The Astrophysical Journal Letters* 901.2, p. L38.

Kingma, Diederik P. and Jimmy Ba (2014). *Adam: A Method for Stochastic Optimization.*

Kutner, Marc L. (2003). *Astronomy: A Physical Perspective.* 2nd ed. Cambridge university press.

Lacy, M. et al. (Sept. 2004). "Obscured and Unobscured Active Galactic Nuclei in the Spitzer Space Telescope First Look Survey". In: *ApJS* 154.1, pp. 166–169.

Laurikainen, Eija, Heikki Salo, and Ronald Buta (May 2004). "Comparison of Bar Strengths and Fractions of Bars in Active and Nonactive Galaxies". In: *The Astrophysical Journal* 607.1, pp. 103–124.

Lee, Gwang-Ho et al. (Apr. 2012). "DO BARS TRIGGER ACTIVITY IN GALACTIC NUCLEI?" In: *The Astrophysical Journal* 750.2.

Mateos, S. et al. (July 2013). "Uncovering obscured luminous AGN with WISE". In: *Monthly Notices of the Royal Astronomical Society* 434.2, pp. 941–955.

Melvin, Thomas et al. (Jan. 2014). "Galaxy Zoo: an independent look at the evolution of the bar fraction over the last eight billion years from HST-COSMOS". In: *Monthly Notices of the Royal Astronomical Society* 438.4, pp. 2882–2897. ISSN: 0035-8711.

Mo, Houjun, Frank van den Bosch, and Simon White (2010). *Galaxy formation and evolution*. Cambridge university press.

Noguchi, M. (1996). "Barred Galaxies: Intrinsic or Extrinsic?" In: *International Astronomical Union Colloquium* 157, pp. 339–348.

Oh, Seulhee, Kyuseok Oh, and Sukyoung K. Yi (Dec. 2011). "BAR EFFECTS ON CENTRAL STAR FORMATION AND ACTIVE GALACTIC NUCLEUS ACTIVITY". In: *The Astrophysical Journal Supplement Series* 198.1, p. 4.

Padovani, P. et al. (Aug. 2017). "Active galactic nuclei: what's in a name?" In: *The Astronomy and Astrophysics Review* 25.1.

Sheth, Kartik et al. (Mar. 2008). "Evolution of the Bar Fraction in COSMOS: Quantifying the Assembly of the Hubble Sequence". In: *The Astrophysical Journal* 675.2, pp. 1141–1155.

Silva-Lima, Luiz A. et al. (May 2022). "Revisiting the role of bars in AGN fuelling with propensity score sample matching". In: *Astronomy &amp Astrophysics* 661, A105.

Sparke, L. S. and S. Gallagher (2007). *Galaxies in the Universe: An introduction*. 2nd ed. Cambridge university press.

Stern, Daniel et al. (Sept. 2005). "Mid-Infrared Selection of Active Galaxies". In: *The Astrophysical Journal* 631.1, pp. 163–168.

Stern, Daniel et al. (July 2012). "Mid-infrared Selection of Active Galactic Nuclei with the Wide-Field Infrared Survey Explorer. I. Characterizing WISE-selected Active Galactic Nuclei in COSMOS". In: *ApJ* 753.1, 30, p. 30.

Stoughton, Chris et al. (Jan. 2002). "Sloan Digital Sky Survey: Early Data Release". In: 123.1, pp. 485–548.

Willett, Kyle W. et al. (Sept. 2013). "Galaxy Zoo 2: detailed morphological classifications for 304 122 galaxies from the Sloan Digital Sky Survey". In: *Monthly Notices of the Royal Astronomical Society* 435.4, pp. 2835–2860.

Willett, Kyle W. et al. (Oct. 2016). "Galaxy Zoo: morphological classifications for 120 000 galaxies iniHST/ilegacy imaging". In: *Monthly Notices of the Royal Astronomical Society* 464.4, pp. 4176–4203.

Wright, Edward L. et al. (Dec. 2010). "The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance". In: 140.6, pp. 1868–1881.

York, Donald G. et al. (Sept. 2000). "The Sloan Digital Sky Survey: Technical Summary". In: 120.3, pp. 1579–1587.

A Bar examples

Below some plots are shown of HSC images that were identified as bars using the method described in section 3.3. Please note that these images were selected randomly so some galaxies might be shown twice.