Searching for Red Geyser Galaxies with Machine Learning

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Arun Ravi

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

April, 2023

Supervisor: Dr. Yogesh Wadadekar

Certificate

This is to certify that this dissertation entitled Searching for Red Geyser Galaxies with Machine Learning towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Arun Ravi at Indian Institute of Science Education and Research under the supervision of Dr. Yogesh Wadadekar, Associate Professor, NCRA - TIFR, Pune , during the academic year 2022-2023.

Dr. Yogesh Wadadekar

Committee:

Dr. Yogesh Wadadekar

Dr. Susmita Adhikari

This thesis is dedicated to my school teachers, who fostered my curiosity and helped to keep it intact through the rigmarole of school education

Declaration

I hereby declare that the matter embodied in the report entitled Searching for Red Geyser Galaxies with Machine Learning are the results of the work carried out by me at the NCRA - TIFR, Pune, Indian Institute of Science Education and Research, Pune, under the supervision of Dr. Yogesh Wadadekar and the same has not been submitted elsewhere for any other degree.

Arun Ravi

Acknowledgments

I would like to thank my thesis guide, Dr. Yogesh Wadadekar for his guidance throughout the duration of this thesis. His insights into the physics of galaxy evolution and machine learning helped me push the problem far enough to warrant a novel contribution at the intersection of machine learning and astronomy.

I would also like to thank Dr. Susmita Adhikari, my thesis expert, for her timely inputs and inspiring questions that made me think about my problem in aspects I would not have approached otherwise.

A special thanks to Dr. Namrata Roy for inspiring this problem, and also providing us with her personal catalog of red geysers for our work. I'd also like to thank Dr. Niharika Sravan for her insightful comments that helped me place my problem in the larger scale of current progress in the machine learning for astronomy community.

Many thanks to Ashwin Samudre and Kavin Kumar for discussions on ML and the physics of galaxies respectively.

A shoutout to my friends for the food, entertainment and company that kept me relaxed and composed to work on my thesis.

A special thanks to my school teachers - Mr. Manoj Kumar and Mrs. Sudha Brahmadattan, whose support inspired me to pursue physics for my undergraduate studies.

Last but definitely the most of them all, I would like to thank my parents for staying by me and supporting me throughout this journey. I couldn't have done it without them.

Abstract

Red geysers are an important class of low star forming galaxies that show telltale signs of AGN maintainence mode driven feedback mechanism that keeps them quenched. There are many questions left to be answered about the nature of these low luminosity AGNs in maintaining quenched state - such as which stage of maintainence is dominant in the local galactic population In order to answer these questions, one requires a statistically significant sample of red geysers to conduct a study of the distribution of their properties. From a sample of ~ 4700 in an earlier data release of the SDSS-MaNGA survey, 139 red geysers were identified by manually inspecting each example. The latest data release of ~ 10000 galaxies has not been scoured for red geysers yet. Our goal is to build an automated machine learning model to solve this problem. We present our results with different models and discuss a novel algorithm based on the few shot learning paradigm that can perform the task with $\sim 99\%$ accuracy.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Among gravitationally bound objects in the universe, galaxies are of intermediate size and have some of the richest variations in properties. The interplay between the underlying physics and how they manifest as emergent galaxy properties are difficult to model directly, and thus form very interesting objects to study. They are workshops for testing theories of cosmology and structure formation, and influence astrophysical activity on smaller scales such as star formation and evolution [1].

The variations in galaxy properties and evolutionary histories are influenced by local (group, cluster) and large scale (sheet, filament) environmental factors, and due to internal (AGN, chemical enrichment) processes. This complex interplay of factors manifests itself as a rich plethora of observational properties, which are essential to constrain our understanding of the physical processes that drive galaxy formation and evolution [2][3].

To observe their time evolution, we can study galaxies that are very far away. Since the speed of light is taken as a constant across cosmic timescales, we will observe these galaxies as they were in the past. By connecting how distributions of galaxy properties were at different times in cosmic history, we can understand how galaxies evolved with time[4].

At a given redshift, it is more complicated to study the time evolution of galaxies. The evolutionary timescales of galaxies are of the order of 10^8 years, and are thus impossible to observe directly in a human lifetime. To study them, we collect different galaxies at different stages of their evolutionary history and attempt to understand how their properties have changed. To accomplish

this, we need observed properties of large populations of galaxies that span the different stages of evolution of galaxy populations[4][3].

One of the critical properties of a galaxy is its ability to convert cold molecular gas into stars. This star formation process can be measured by the stellar mass content of a galaxy. Comparing it with the total mass tells us how "good" of a star forming galaxy it is; one typically uses this "specific star formation rate" as a measure of how actively a galaxy converts its gas to stars[4].

From redshift $z \sim 2$ onwards, it has been observed that the global star formation rate has been declining [5], and the fraction of quenched galaxies has been increasing [6]. This hints at the existence of mechanisms that suppress star formation over large timescales, as otherwise it would be difficult to reconcile with there being large reservoirs of molecular gas in the gravitational potential well of these galaxies that can potentially condense and form stars.

There must be some kind of "feedback mechanism" that ensure that the star formation in these galaxies continues to be suppressed. Some of the proposed drivers include gravitational effects due the presence of the bulge [7] and stellar winds that heat the ISM [8]. However the most popular candidate yet has been the winds, jets and outflows from the central AGN of the host galaxy [9]. AGN activity is of two types - firstly a violent "quasar mode" emission that has powerful outflows which heats the gas in and around the galaxy thereby preventing it from forming stars or by starving the galaxy of star forming gas by ejecting it during these outflows [10]. Secondly, there is the more peaceful "maintenance mode" of AGN activity which mostly deposits its energy as radiation to the medium of the galaxy, thereby suppressing star formation [9].

There are very few known candidates for which the maintenance mode of AGN activity suppresses star formation in the nearby universe; these are typically found in stellar clusters [11]. We are yet to extensively observe candidates from "typical" populations in the cosmic neighbourhood that stay quenched with this mechanism [11]; understanding their properties and behaviour is critical to solving the problem of why global star formation has been on the decline for the last 10 billion years [12]. Performing large survey projects and searching through its vast database is essential to identify these candidates.

Observational projects such as the Sloan Digital Sky Survey (SDSS) set out to observe and characterise large samples of galaxy populations. Earlier observations were restricted to global properties of galaxies and lacked spatially resolved information which would give better understanding of how certain properties were distributed within galaxies. To obtain these, one would

have to perform time intensive long slit spectroscopy, which among other science based drawbacks was limited to collecting only a few samples from certain targeted galaxies every observing run. The "**Ma**pping **N**earby **G**alaxies with **A**pache Point Observatory" (MaNGA) project used the more modern integral field unit (IFU) spectroscopy to obtain resolved spectra over 4 observing runs, covering $\sim$ 10000 galaxies from the SDSS. This comprehensive survey gives an opportunity to perform comprehensive studies of resolved galaxy properties and their statistical distribution in the near field.

Using the MaNGA survey, [13] discovered a new class of quenched galaxies, dubbed "red geysers". The behaviour of these galaxies as observed using spatially resolved information pointed at maintenance mode AGN activity being responsible for their quenched nature, and opened a new population of galaxies for further investigating the effect of AGN maintenance mode on quenching. The AGN signatures in red geysers were confirmed by [12] with radio observations. Investigation of their HI content using the HI-MaNGA results by [14] revealed no statistically significant difference in the neutral gas content of red geysers, which they infer as an intermittent phase of red geyser evolution where the AGN is about to switch off and the energy from the SMBH is deposited in the ISM.

Currently only 139 of these red geysers are known, which does not provide us a statistically significant population of galaxies for studying their properties. While the currently known sample was recovered manually by sorting through each galaxy in the MaNGA sample from earlier MaNGA releases (which had $\sim$ 4700 galaxies), we wish to automate the process for the latest release ($\sim$ 10000) which is more than twice the size of the previous sample. We employ deep learning techniques for this problem. Our efforts in this direction are motivated by the speed of inference and increasingly tractable nature of AI/ML techniques in the sciences, particularly in astronomy.

This thesis currently summarises the proof-of-concept study carried out to study the relative performances of different ML algorithms for this problem. The Theory chapter provides a basic overview on galaxy evolution and provides sufficient theoretical background and resources for the reader to further appreciate the depth of the problem. The Data chapter deals with the instrumentation used for observation, data collecting and cleaning procedure adopted to prepare the data for our machine learning models. The Methods chapter provides details on the ML models used. The Results and Discussion chapter details the results of our experiments and provides the scope for further work in this direction.

Chapter 2

Theory

The light of a galaxy is made of the emission from each one of its individual component stars, and the radiation from the gas and dust. In similar lines, the spectrum of a galaxy is a combination of all the individual spectra of its composite substrates. The key to understanding the evolution of galaxy properties lies in understanding the collective properties of the stellar populations.

2.1 Population Synthesis

Stellar population synthesis models attempt to explain the spectra of galaxies using the spectra of the stars in conjunction with the superposition principle as suggested above. Since the physics and evolution of stars is well known, we can use the distribution of stellar populations and how they evolve in time to understand how galaxies and their light changes with time. By using spectral energy distribution (SED) fitting techniques, one can recover parameters that are physically useful, such as stellar mass, star formation rate and history, metallicity and dust content, etc. For a detailed review on the topic, an interested reader can refer to [15]. Here we present a simple model and briefly describe the underlying physics followed by the procedure involved in extracting these physical parameters.

2.1.1 Model Assumptions

The hypothesis is that stars are formed from collapsing gas, possibly influenced by the surrounding medium to collapse into stars of different mass. The distribution of masses in this initial stage is called the initial mass function (IMF) $\phi(m)$ and is normalised as

$$\int_{m_L}^{m_U} m\,\phi(m)\,dm = 1\,M_\odot$$

Since the minimum mass required for a star to ignite is $0.08M_\odot$, the lower limit for integration is taken as this value. The upper limit is taken to be $m_U \sim 100M_\odot$, which is the observed upper limit for stars.

An ansatz for the form of the function $\phi(m)$ can be obtained by observations of young stellar populations, the most popularly used one is the Salpeter IMF [16] defined as :

$$\phi(m) \propto m^{-2.35}$$

This distribution function is not universal - it has been observed to vary with mass, but not with environment. It has been hypothesized that high redshift galaxies follow a different IMF (to explain their properties) as do starburst galaxies [17].

Despite the lower mass stars dominating in numbers, the bulk of the luminosity comes from the high mass stars, as for stars in the main sequence, $L \propto M^3$.

We define the star formation rate as the mass of gas converted to stars per unit time :

$$\psi(t) = -\frac{dM_{gas}}{dt}$$

The metallicity Z of the ISM influences the metallicity of the stars as it is this gas that collapses to form stars. The stellar properties themselves depend on the metallicity. Note that stellar processes result in increase of the metallicity in the star, which means the ISM has $\dot{Z}(t) \geq 0$.

Suppose we have a group of stars of age t' and metallicity Z. The energy emitted per wave-

length per time interval is then $S_{\lambda,Z}(t')$, normalized to $1M_{\odot}$. One can then write the total spectral luminosity of the galaxy as

$$F_{\lambda}(t) = \int_0^t dt' \psi(t-t') S_{\lambda,Z(t-t')}(t')$$

Note how $F_{\lambda}(t)$ depends on the star formation history and evolution of the metallicity of the ISM.

2.2 Spectral Evolution

When a stellar population is formed, the spectra is dominated by UV emission from luminous stars. As these massive stars burn out their fuel, the flux below 1000 Ådrops after $\sim 10^7$ yrs and diminishes after $\sim 10^8$ yrs. As these stars evolve into red supergiants, the NIR emission increases.

From 10^8 yr to 10^9 yr, the emission in NIR is high and the emission in low frequencies continues to be low. However, after $\sim 3 \times 10^9$ yrs, the UV flux increases as there is an increase in contribution from the blue stars (which were formerly in the asymptotic giant branch) and newly born white dwarfs. After 4×10^9 yrs, the spectrum evolves very little.

After $\sim 10^7$ yrs, there is a break in the spectrum at 4000 Ådue to the an increase in opacity owing to strong transitions of CaII and Balmer lines in the stellar atmospheres. This is a telltale sign of the age since star formation began.

2.3 Realistic Star Formation Models

The above discussion considered a toy model of star formation in many ways, but one important fundamental assumption was that all stars were formed at the same time. In reality, star formation is spread out over time and space, and we would like to describe it with a more realistic model. The standard model for star formation rate does just this, and is given as :

$$\psi(t) = \frac{H(t-t_f)}{\tau} exp\left[-(t-t_f)/\tau\right]$$

Here H is the Heaviside step function, t_f is the time at which star formation onset occurred and τ is the characteristic duration of star formation. The most significant achievement for this model is that it can reproduce the colour evolution of the galaxy.

2.4 Galactic Gas Contribution to the Spectrum

The HII regions of galaxies also contribute to the emission from galaxies besides the stars. However, this diminishes after $\sim 10^7$ yrs and does not significantly add to the broad-band emission from galaxies. Since these are the biggest contributors to emission from these galaxies, they are used as a diagnostic for star formation rate (SFR) and metallicity.

2.4.1 Star Formation and $H\alpha$ maps

Estimating SFR from SED fitting is a tedious and unreliable task :

- Reddening of galaxies due to dust, metallicity and age has a degeneracy which requires high quality data to be broken, which is expensive to collect
- The choice of dust and SFH model is not very strongly constrained

Thus, proxy measures involving monochromatic indicators are employed to estimate the SFR - the most commonly used ones being UV, Hα and total IR. The staple review on the topic is provided by [18]. Here we shall briefly discuss Hα as a star formation proxy.

The presence of young stars implies that there will be a lot of high frequency emission blueward of the visible spectrum. These energetic photons will be absorbed by the nebulous gas in the galaxy and re-emitted in higher wavelength regions like Hα and so on.

Calibrated measurements by [19] and [20] show that one can use the Hα luminosity $L_{H\alpha}$ to obtain the SFR using the relation

$$SFR[M_\odot yr^{-1}] = 7.9 \times 10^{-49} L_{H\alpha}[Js^{-1}]$$

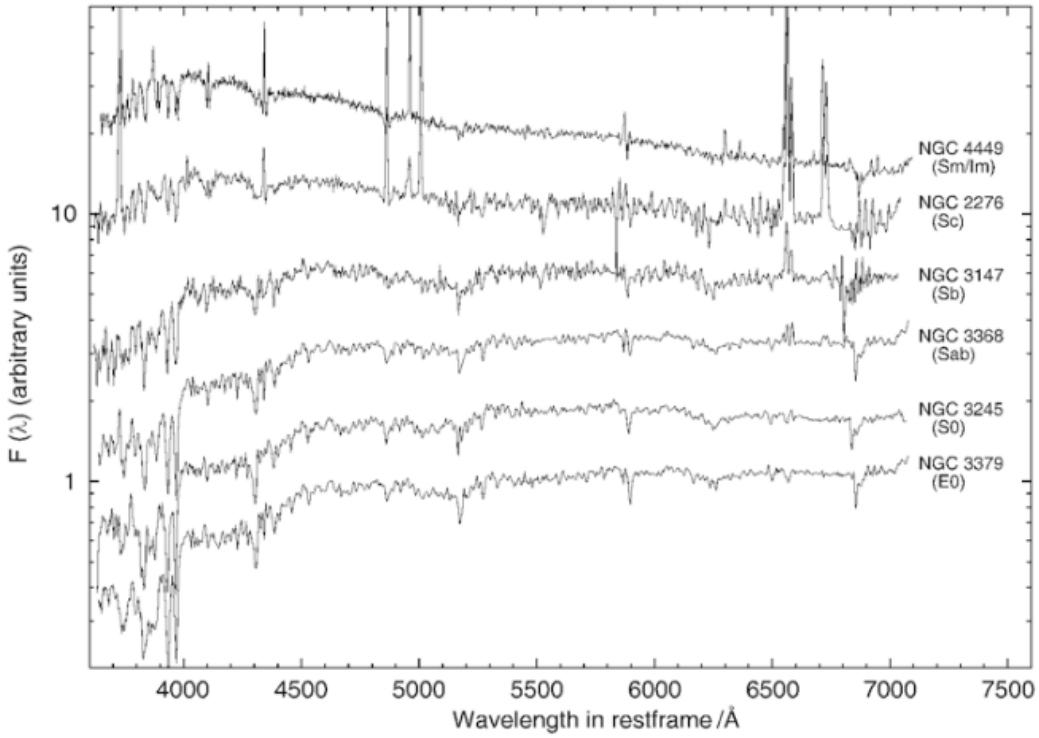


Figure 2.1: Spectra of different galaxy morphologies, from early Hubble types (E0) to late Hubble types (Sc) and dwarf/irregulars (Sm/Im). There is difference in absolute flux among the different morphologies. The late types have stronger emission features and lack a strong 4000 Åabsorption feature, both characteristics of ongoing star formation.

Similar relations exist for other spectral lines as well.

2.5 Spectra of Galaxies

Similar to how the photometric image of a galaxy is composed of its individual stars and gas, the spectra of a galaxy is composed of the emission and absorption from its constituent stars and gas. This is what leads to exciting features that can be used to study the physical composition of the galaxy, some of which display trends. In late Hubble types, we see :

- Bluer spectrum
- Stronger emission lines

- Weaker absorption lines
- Smaller 4000 Åbreak

All characteristics of recent/ongoing star formation events in the galaxy. This is as opposed to the early type galaxies that predominantly have

- Lower star formation, hence redder spectrum from red stars
- No HII regions which provide emission lines
- Pronounced 4000 Åbreak because of older stellar population

2.6 SED Fitting for SFR

To obtain the SFR for a galaxy from its spectra, one performs SED fitting. [15] and [21] provide excellent reviews on the topic. We shall briefly discuss the procedure here.

The procedure is an inverse problem where there is an attempt to recreate the spectrum (usually across all observed bands) using stellar population synthesis (SPS) models. Many ansatz star formation histories are attempted and SPS models are run to recreate the observed spectrum. This is then used to calculate the number of young stars and assign the SFR for the galaxy. Inverse modelling techniques such as Bayesian inference are popular, and in recent years SED fitting has also been done use ML techniques.

2.7 Galaxy Evolution in a Nutshell

The scope of this section is to provide a brief overview of galaxy evolution in near field cosmology i.e. at redshifts $z \lesssim 0.1$. We will briefly discuss galaxy formation followed by feedback processes that impact galaxy evolution in the near field. The interested reader can refer to a more comprehensive introduction from textbooks such as [4] or a detailed understanding through the canon of [1].

2.7.1 Formation

The first luminous structures formed in dark matter haloes $\sim 10^9$ yrs after the Big Bang. The collapse of baryonic matter (mostly HI) into disc like structures and their subsequent cooling and "condensing" to form the first stars happened at around this time. These were the earliest galaxies.

2.7.2 Supernova Feedback

Once the heaviest stars start going supernova, they eject their stellar matter (which has increased in metallicity) and heat the surrounding medium. The heating prevents gas from cooling and clumping to form new stars, and the violent explosion expels neutral gas from the galaxy's gravitational potential well, which also suppresses star formation.

2.7.3 Mergers

Mergers can take the form of minor mergers (galaxy mass ratio of $\lesssim 1:3$) or major mergers (mass ratio $\gtrsim 1:3$). Minor mergers involve the smaller galaxy getting stripped of its gas by ram pressure stripping when it enters the halo of the larger galaxy, thereby halting star formation in the minor galaxy. The stripped gas is available for the major galaxy to attract and form stars.

Major mergers are more disruptive to both galaxies; almost the entire galactic structure is lost. The resulting galactic material falls into the gravitational potential, and the gas getting shocked and forming overdensities condense to form stars. The time scale of infall is $\gtrsim 10^7$ yrs, which allows for the earliest stars to go supernova and provide feedback that depletes star forming gas as well as heat it. The first resulting galaxy is an elliptical galaxy, and the resulting material forms a disk around it forming a bulgy disk/spiral galaxy.

2.7.4 Black Hole and AGN Feedback

Feedback from AGNs and the SMBH at the center of galaxies is the most important internal mechanism affecting the evolution of galaxies. [22] provides an understanding of the current status of

our understanding of the topic, and [23] summarises the topic. Here, we will paint a brief picture of the same.

It is known that AGN feedback is necessary for replicating the observed mass and star forming rate distributions of galaxies from simulations. AGNs have a double effect on star formation in galaxies - they can enhance it (in rare occasions) or suppres it. When there is a smaller satellite orbiting the AGN host and it receives a shock from the outflow of matter from the AGN, there is an enhancement in SFR. The outflows and jets themselves expel a lot of radiation and heat that heats up the ISM, thereby preventing gas from cooling sufficiently to form stars. They also physically move gas from the galaxy to the outside, thereby depriving the galaxy of star forming gas. AGNs also suppress star formation by heating up the IGM surrouding the host galaxy - this process prevents cool gas from flowing into the host galaxy and enabling new star formation.

Different AGN types play different roles in quenching star formation. Among ”radio mode” AGN (these are the low luminosity counterparts of the more stronger quasar mode AGN), there are radiatively efficient and radiatively inefficient AGN types. Radiatively efficient AGN are powerful drivers of outflows and are responsible for the two-sided radio jets that has come to be popularly associated with AGN. These AGN types suppress star formation by driving out the gas physically during their violent emission episodes. Radiatively inefficient AGN are responsible for dissipating a large fraction of their energy into the ISM of the galaxy, thereby heating it and maintaining the quenched state. These AGNs also have radio jets associated with them.

2.8 Galaxy Classification by Star Formation

Galaxies can be segregated on the basis of their star formation rates. The distribution of galaxies on a stellar mass - star formation rate plot reveals to us three main populations of galaxies by way of their SFR. We attempt to understand each population in this section.

2.8.1 Star Forming Galaxies

These are colloquially termed ”blue” galaxies since they have a higher luminosity in the blue side of the spectrum due to ongoing star formation producing young stars that strongly emit UV light. The stereotypical star forming galaxy is a spiral galaxy [4], though there are blue ellipticals and

star forming S0s widely reported [24] [25]. These galaxies are aligned along what is roughly a straight line. This is named the *star forming main sequence* (SFMS), but it bears no similarity to the main sequence we are familiar with in a Hertzsprung-Russel Diagram in that the position on this sequence does not uniquely determine the properties of the galaxy. The low scatter of the main sequence indicates that galaxies grow their stars by secular processes and not by stochastic external processes. These star forming galaxies are mostly dominated by disk galaxies, and are mostly rotationally supported[4].

Starburst Galaxies

Starburst galaxies are those with anomalously large amounts of ongoing star formation. These are typically caused due to a recent merger, and are thus common in disk and irregular galaxies. They can also be driven by the formation of bars in galaxies that drive gas towards the center of the galaxy that enhances star formation[4].

2.8.2 Quenched Galaxies

As discussed earlier, quenched galaxies are those that have stopped or have reduced their star formation in recent history. This is typically characterized by the presence of old, redder stars which gives it a characteristic red colour. They are thus popularly referred to as ”red and dead” galaxies, and are mostly constituted by ellipticals. There are a few odd disky quenched galaxies as well, which are mostly due to loss of star forming gas by ram pressure stripping in high density environments like the vicinity of larger galaxies and in galaxy clusters [26]. The main reason elliptical galaxies are quenched is because of the lack of star forming gas in the galaxy and in their vicinity., with the major driver being AGN and SMBH feedback. These galaxies are mostly supported by dispersion, though towards the lower mass end there are also fast rotating galaxies. [27] provides an account of the kinematic properties of galaxies on their stellar mass-SFR diagram.

2.8.3 Green Valley Galaxies

There is a small population of galaxies in between the above two populations dubbed ”green valley galaxies” (see [28] for a review). These are low star forming galaxies that are typically in ”transit”

from the star forming sequence to the red and dead sequence due to various quenching mechanisms. The fact that the absolute number of green valley galaxies are low in number in the local universe compared to the star forming and quenched population hints that the process of quenching is fast, and these galaxies make it to the quenched population very fast.

2.9 The AGN Primer

In the local universe, quiescent galaxies dominate the galaxy population at stellar masses above $\sim 2 \times 10^8 M_\odot$ [29]. Once quenched, the galaxy still continues to attract star forming gas from the IGM. There has to be a feedback mechanism that continues to suppress star formation in quenched galaxies. Some of the possible candidates are :

- Physical loss of star forming gas by mergers, supernovae and AGN jets
- Heating of the ISM by strong radiative processes from the inside of the galaxy
- Gravitational effects induced by galaxy bulges

However, the most reliable candidate for keeping quenched galaxies quenched seems to be AGN activity. The center of galaxies host an "active galactic nucleus" - a region that has higher than normal emission in certain wavelengths that are not of stellar origin. The sources of most AGNs are the radiative processes from the accretion disks around black holes. The emission activity at centre also drives winds, outflows and jets. A review of AGNs can be found in [30].

There are two ways in which AGN activity could continue to quench a galaxy :

- **Quasar Mode Feedback :** In luminous AGNs and massive quasars, there is a large amount of energy released in radiative processes from the accretion of matter into the black hole which heat the surrounding gas, which prevents it from cooling sufficiently to form stars. These objects also drive powerful gas outflows, which drains the galaxy of its star forming gas
- **Radio Mode Emission (Maintenance Mode) :** There are also less luminous AGNs, which are black holes that accrete at a lower rate and thus emit radiation at a lower flux. This can only cause the surrounding gas to heat up and this suppress star formation

The number of galaxies observed exhibiting radio mode emission is quite low as their flux is pretty low. Thus, observations have been limited to only those in the vicinity of the Milky Way.

2.10 Red Geysers

Red geysers were first discovered by [13] and were further characterized in [12]. They are an exciting class of galaxies because their properties seem to suggest that their quiescence may be maintained by AGN radio mode feedback. They were first discovered using optical band observations in the MaNGA survey. Their quiescence was characterized by their colour i.e. their flux was such that $NUV - r > 5$. They also had large scale winds of ionized gas outflows. The salient feature of these outflows was that it aligned with a bisymmetric enhancement in the spatial distribution of strong emission lines like $H\alpha$ and $[OIII]$. This manifested itseelf in the emission width maps of these lines as a bisymmetric pattern. The ionized emission in these galaxies also extend throughout the galaxy.

Another characteristic feature is that of the gas kinematics of these galaxies, whose existence strongly hint at there being outflowing winds. The gradient of the gas velocity field is in alignment with the aforementioned bisyymetric feature observed in emission. However, it is misaligned with the major and minor axis of the stellar velocity field. Typical values of this gas velocity field can reach up to $300 km/s$, which is an important fact to consider in trying to understand the origin of these features.

These descriptions are necessary, but are they sufficient? In other words, does searching for galaxies with these properties provide only red geysers, or are there any "contaminant" galaxies as well?

One feature that has other classes of galaxies common with it is the kinematics. Early type galaxies with accreted disks will show very similar kinematic features, some of these accreted disks may even have the $H\alpha$ EW feature of bisymmetric patterns.

This begs the question - are red geysers really driven by AGNs, or are they simply accreted disks? The accreted disk formation is a very unique formation process - the accreted gas coming into the galaxy will be acted upon by the torques in its gravitational potential well, aligning it with the major or minor axis. Thus, the misalignment in the stellar and gas velocities will not be

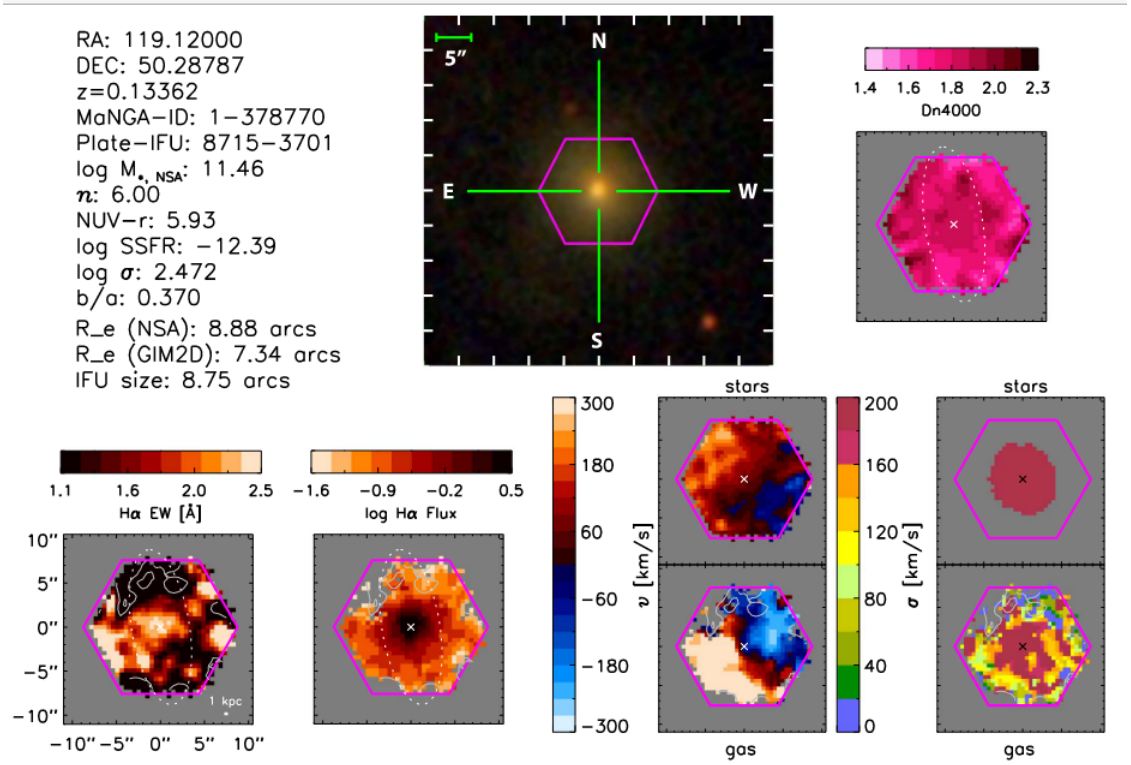


Figure 2.2: A red geyser candidate detected using MaNGA, courtesy [12]. Clockwise from top left : The sky position and other observable properties, along with instrumental details of the candidate galaxy; the target SDSS galaxy's MaNGA IFU coverage; the D_n4000 map of the galaxy; the stellar and gas velocity and velocity dispersions of the galaxy; the $H\alpha$ flux and EW maps of the candidate galaxy.

randomly oriented, as is the case in red geysers, but will be at 90^o (polar disk) angle or at $0^o/180^o$ (corotating disk /counterrotating disk) angle. Another simple argument puts this question to rest - the RMS gas velocities ($V_{rms} = \left(\sqrt{V^2+\sigma^2}\right)$) typically found in red geysers are different from that one would get from modelling the accreted disk gas velocity by $\sim 100 km/s$, which is a very strong discrepancy given the simulated model itself predicts gas velocities of $\sim 200 km/s$. For this they dervied escape velocities from the stellar velocity maps, and predicted that $15-20\%$ of the gas content would escape the galaxy because it would exceed the escape velocity. The accreted gas disk model also fails to prove why the bisymmetric pattern is to some degree randomly oriented with respect to the kinematic axis.

Since we have eliminated the closest competitor to explaining how red geysers work, can we say with absolute certainty that it is indeed the AGN which is driving these winds? One of the first steps in this direction is showing that red geysers have a greater possibility to host an AGN than other quenched galaxies in similar environments. Since the red geyser has very low Eddington ratio, one has to devise special methods to detect the faint radio signatures from them. [12] performed stacking of red geysers observed in radio using the VLA and compared the flux with a "control" sample of quiescent galaxies. They found the stacked red geysers had a larger radio flux than the control sample, hinting at the red geysers indeed having higher AGN activity than their other quenched counterparts.

2.10.1 $H\alpha$ disturbed galaxies

A visual inspection of the MaNGA galaxies reveal another similar class of galaxies. They have very similar gas content - median $H\alpha$ EW value is ~ 0.5 Å, compared to red geysers' ~ 0.8 Å. They also have high gas velocity ($\sim 250 km/s$) and gas velocity dispersion ($\sim 200 km/s$) compared to stellar velocity values of $\sim 60 km/s$. However, their $H\alpha$ EW maps do not show a bisymmetric pattern, but a twisted and disturbed pattern as seen in the Fig.

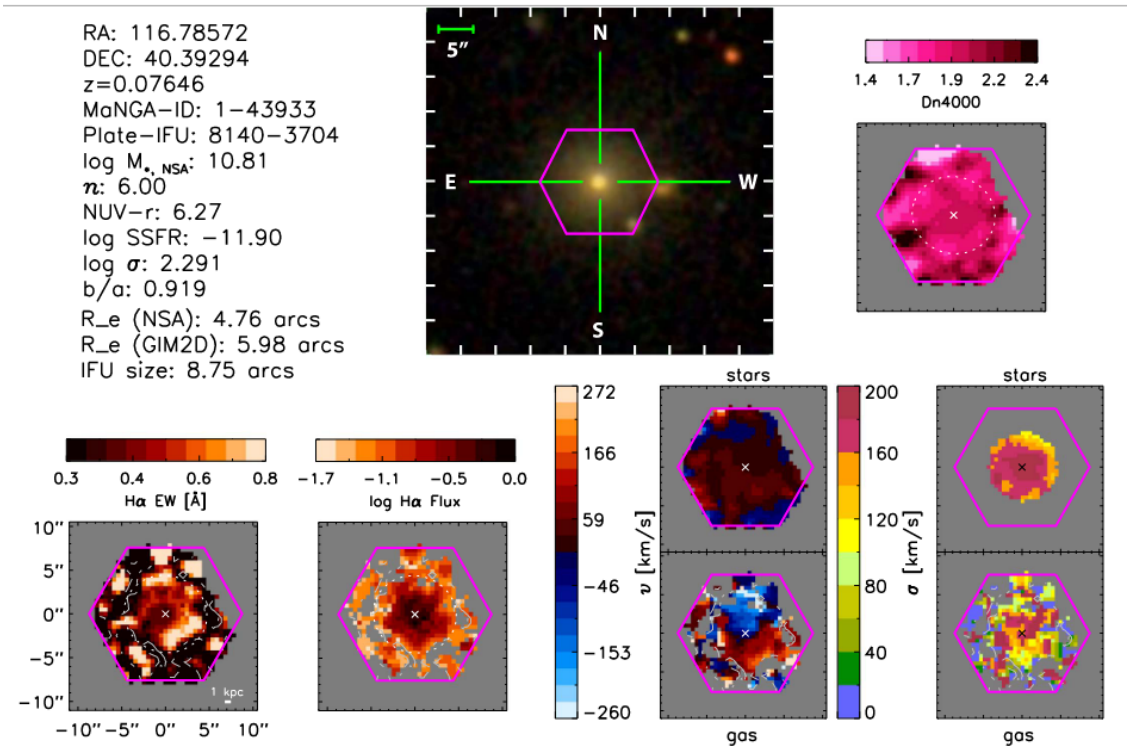


Figure 2.3: A Hα disturebed galaxy detected using MaNGA, courtesy [12]. Clockwise from top left : The sky position and other observable properties, along with instrumental details of the candidate galaxy; the target SDSS galaxy's MaNGA IFU coverage; the D_n4000 map of the galaxy; the stellar and gas velocity and velocity dispersions of the galaxy; the $H\alpha$ flux and EW maps of the candidate galaxy. Note how when compared to Fig 2.2, the EW map has more disordered distribution of $H\alpha$ enhancements.

Chapter 3

Data

3.1 The Sloan Digital Sky Survey

The Sloan Digital Sky Survey (SDSS) [31] is the first digitized optical survey. Operating out of the Apache Point Observatory in New Mexico, it collected spectroscopic redshifts as well as photometric data from a large part of the sky. This was possible because of two major pioneering breakthroughs in technology at the time it was commissioned :

- 30 highly efficient sky scanning CCD cameras that could capture images in five (u, g, r, i, z) photometric bands
- A spectrograph that consisted of optical fibres that could be inserted into punched holes towards a designated target that could collect 640 spectra simultaneously

The hundreds of gigabytes of data collected every night was efficiently post processed and stored by software pipelines, which ultimately produced detailed catalogues of stellar and galactic properties. The 9th Data Release was the last of the photometric data released, with having observed 10^9 objects and $\sim 35\%$ of the sky.

The redshift survey managed to go as deep as $z \sim 0.7$, whereas the deepest quasar was at $z \sim 5$. The imaging survey managed to capture one quasar at $z \sim 6$. Each one of these surveys are conducted as a part of projects with specific science goals - the Baryon Oscillation Spectroscopic

Survey (BOSS)[32] is one such example which sought to observe the baryon acoustic oscillations and had $\sim$ 800000 spectra collected in the SDSS-III run.

The project facility at the Apache Point Observatory has undergone modernization, with the manual plug plates for collecting spectra being replaced by automated robotic arms for the same purpose. Currently, the SDSS-V is in operation, procuring stellar spectra, mapping SMBHs in select galaxies and analyzing clouds of interstellar gas in the Local Volume [33]

3.2 Spatially Resolved Information

Traditionally, spectra were observed for the entirety of the source. Spectra were expensive to collect, and thus one had to work with low resolutions to capture the spectra of even extended objects. In many science cases, it is desired to have spectra for each individual pixel we observe on the object, as it can help us understand the physics of astronomical objects at a smaller resolution. Some possible use cases are in studying stellar populations, galactic structures, active galaxies, clusters, high redshift galaxies and gravitational lenses.

3.2.1 Long Slit Spectroscopy

Long slit spectroscopy (LSS) is a method to obtain spatial and spectral information of an extended source. One places a slit on the source (see Fig) and passes the slit of light through a diffraction grating. Along the axis of the slit, we observe the source's spatial features, and perpendicular to the slit axis, we observe its spectral features.

3.2.2 Integral Field Spectroscopy

Integral field spectroscopy (integral field unit based spectroscopy) seeks to overcome some of the defects of LSS like

- Reduce wavelength dependent slit losses due to differential atmospheric refraction
- Mitigates seeing effects by taking simultaneous shots across the spatial extent of the source

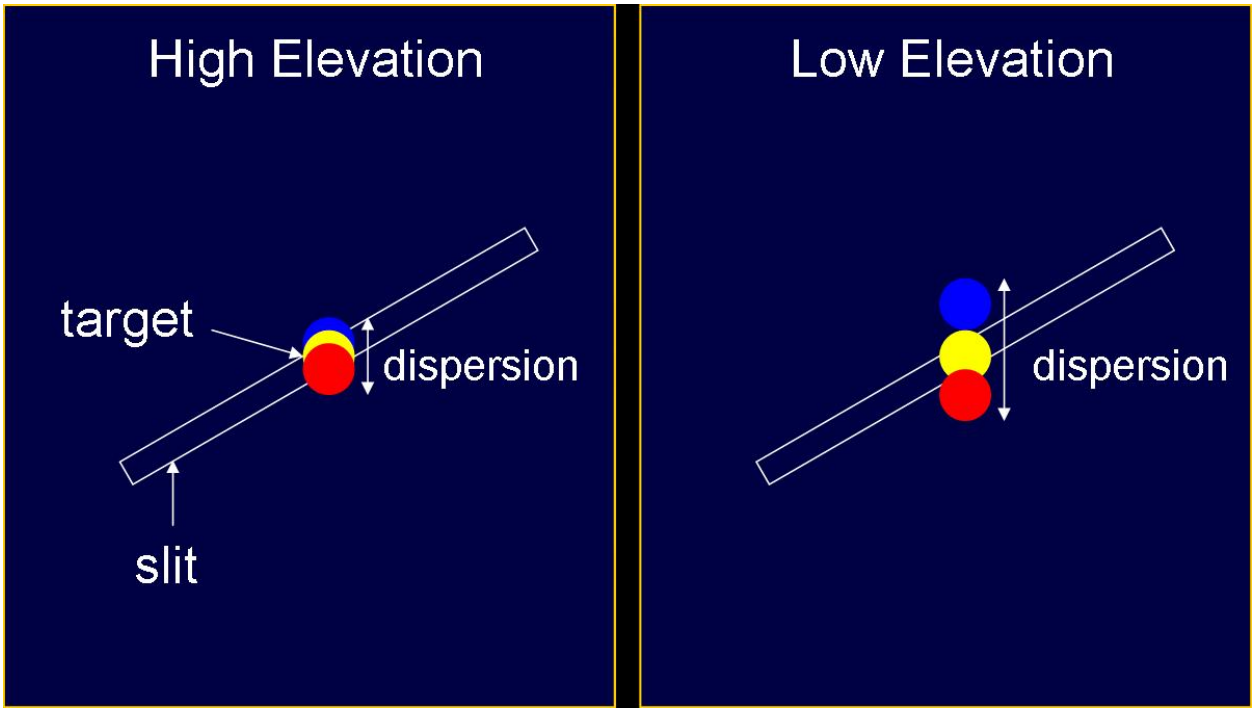


Figure 3.1: Effect of atmospheric dispersion on light from an extended source observed using long slit spectroscopy

- Reliable coverage of the extent of the object compared to unidimensional single slit exposure

The ultimate goal of IFU Spectroscopy is to provide a 3D datacube - 2 dimensions with the spatial extent (x,y coordinate, RA/Dec measurements) and a third dimension containing the values at different wavelengths. Each one of these data points is called a spatial pixel or a *spaxel*. A schematic visualization of an IFU datacube is given in the Fig.

There are different modes by which one performs IFU spectroscopy - the most common ones are :

- **Lenslet Array** : The input image is split using a microlens array, each focused image is passed through a spectrograph. One can tilt the microlens array to ensure the image coverage does not overlap, but the spectrum thus produced extends over a small range.
- **Fiber Array** : Some instruments have fibres attached to lenslets, some don't. Irrespective of the presence of a lenslet, both types are composed of a 2D array of fiber bundles that lead to the slit of a spectrograph

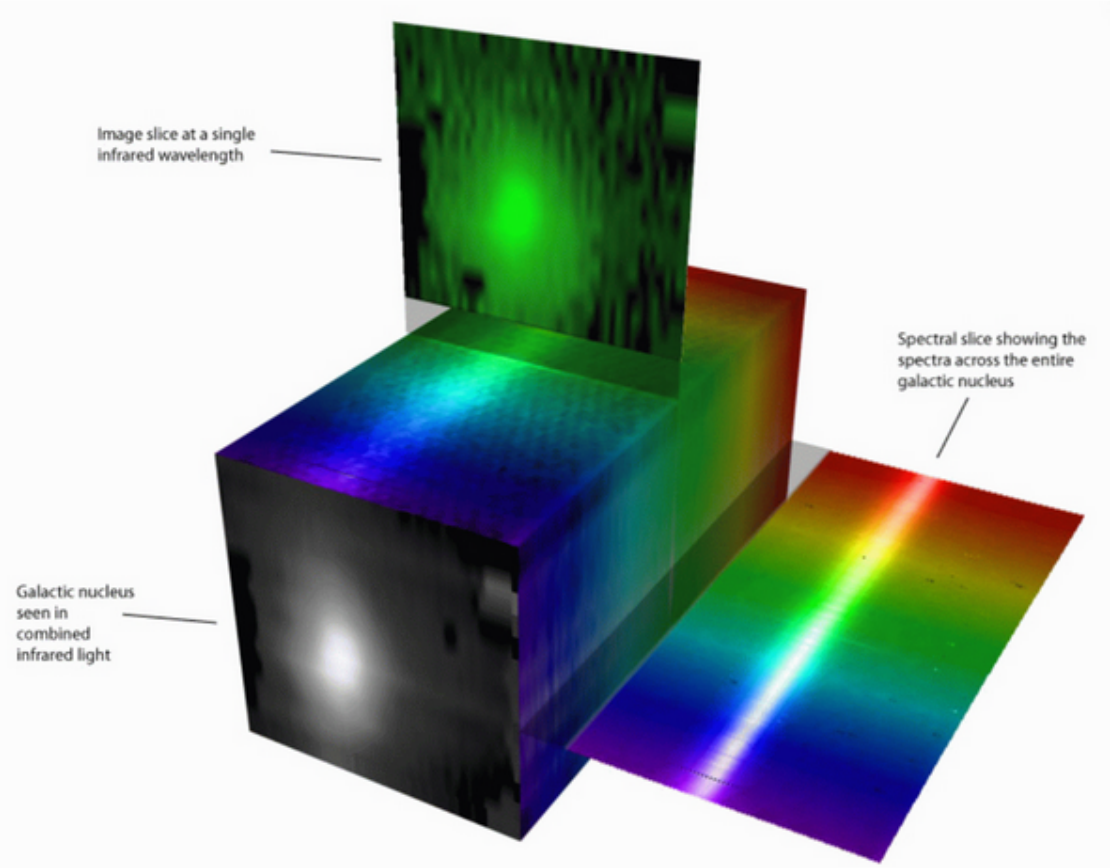


Figure 3.2: Cartoon representation of an IFU datacube

Of these, the fiber array based detector is the most popular one and is what is being used for state of the art large scale surveys. The advantage of using a fibre array is that there is more complete sky coverage, and the detector pixels are more efficiently used. The disadvantages are that the sky sampling is not contiguous and the image is subject to focal ratio degradation.

3.3 IFU Surveys

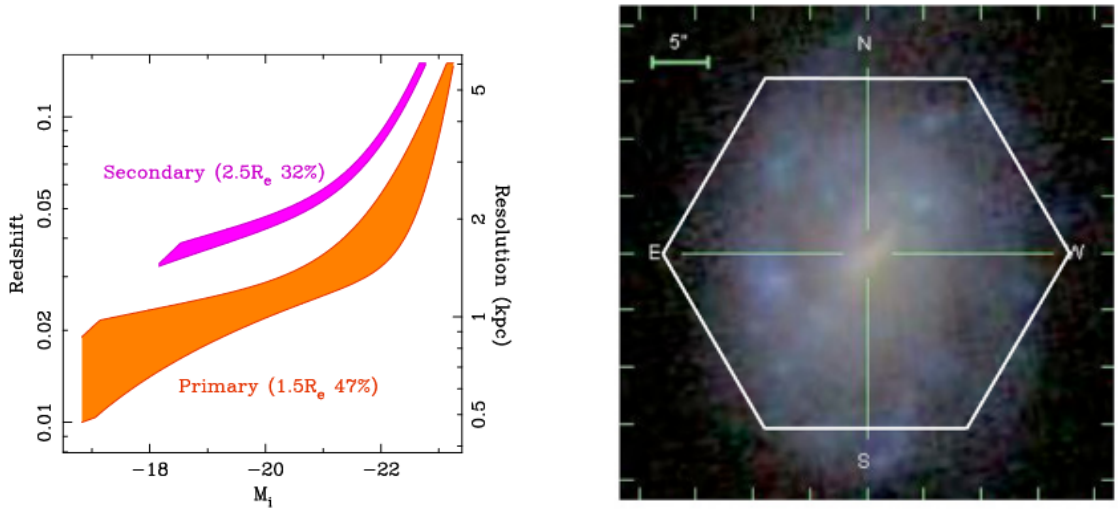
Apart from collecting spatially resolved information from single sources, it would be scientifically useful to collect them for a large sample of extended objects. IFU surveys do exactly that, and we list a few of the prominent ones here. It is also interesting to note that most IFU surveys are done with galaxy evolution as the main science goal, highlighting the importance of the technique for galaxy science.

- The SAURON project ([34]) used a lenslet array mechanism to collect resolved data from nearby early type galaxies. The focus was on studying the gas and stellar kinematic properties, as well as the line strength distribution of these galaxies.
- The CALIFA([35]) and SAMI([36]) survey used a fiber array to study kinematics, stellar populations, chemical properties and mass distributions among different components of a galaxy.
- The MaNGA survey ([37]) also uses a fiber array to study resolved stellar and gas kinematics, star formation and quenching processes, map stellar populations, and perform dynamic modelling of the galaxy components.

We will look at the MaNGA survey in particular, as it is relevant to our problem.

3.3.1 MaNGA Survey

The MaNGA survey (**Ma**pping **N**earby **G**alaxies with the **A**pache Point Observatory) shines above all other IFU surveys for one major reason - it is simply the largest and most comprehensive IFU survey done yet, with the latest data release having $\sim$ 10000 galaxies with spatially resolved



(a) MaNGA Redshift Coverage

(b) MaNGA IFU galaxy coverage

Figure 3.3: Left : Distribution of MaNGA galaxies observed across all observation runs; Right : A MaNGA target galaxy with the hexagonal IFU coverage

spectra. The statistically significant samples one can get from the datacubes of this survey help address the vast repertoire of science problems this survey hopes to address by means of collecting spectra.

The survey does not have any cuts on environment, morphology, inclination angle or size and is thus an excellent sample of the local universe. It however places a cut on the mass range of the galaxies, which still spans an impressive 3 orders of magnitude. It also has aimed for collecting galaxies such that their sample has a uniform mass distribution. This implies their sample is not volume limited, and requires one to perform a volume correction as described in [38] when performing statistical studies.

The data releases were both private and public - a certain amount of data was released for members only, but was made publicly available in the next release cycle which followed shortly.

3.4 Data for the problem

3.4.1 The Marvin API

To access MaNGA data, we use the Marvin API [39]. Its backend consists of a pipeline that takes the MaNGA datacubes (which are FITS files), flags bad data spaxels and present them in a user friendly manner. The tool has a GUI web interface as well as a Python interface, and has inbuilt functions for plotting and remotely accessing the data products. One can access the data for any galaxy of choice by providing the MaNGA ID (a unique identifier for each MaNGA galaxy) or plate-ifu ID (identifier for each plate and IFU used for observation). I used the MaNGA ID for all my work, and in cases where each MaNGA ID had multiple plate-ifus associated with it, I went for the one with the higher signal to noise ratio.

3.4.2 Selecting the red geyser sample

We obtained the currently known red geyser sample via private correspondence with Dr. Namrata Roy. Her criteria for obtaining red geysers from the MPL-5 release can be found in [12]. Here, we briefly mention the procedure followed by her in obtaining this sample.

The first cut is to select galaxies with $NUV - r > 5$. This UV-optical cut ensures quiescent galaxies are selected, and in the MPL-5 release, this accounts for 40% of all galaxies. To remove dust dominated galaxies, there is a cut on star formation rate made, with those with $SFR\left[M_{\odot}yr^{-1}\right] < -2$. To ensure there are no young stars in the galaxy, the EW map of D_n4000 absorption feature is also demanded to have a minimum value of 1.4 Å. If the misalignment in the stellar and gaseous velocity maps is $0^o, 90^o$ or 180^o, these are also excluded as they are galaxies with accreted disks. The spatially resolved gas velocities are expected to be very high ($\sim 300km/s$), and the gas velocity dispersion is also high ($\lessapprox 200km/s$).

To completely remove all contaminants that might resemble red geysers, we ensure all accreted disk galaxies are purged from the sample. This is done by removing all edge on galaxies with $b/a < 0.3$. This process might also remove some genuine red geysers, but that is something we will have to live with at the moment. The kinematic λ_{R_e} and the ellipticity measure ε reveal that red geysers are fast rotating early type galaxies. To further remove any accreted disk type galaxy,

a demand is made that the gas velocity dispersion be $> 60 km/s$ as this ensures regular rotators are excluded.

In summary, the five step procedure is as follows (taken from [12]) :

- Quiscent with rest frame $NUV - r > 5$
- Bisymmetric feature in the Hα EW map
- Bisymmetric feature roughly algined with ionized gas kinematic axis, misaligned with stellar kinematic axis
- Large spatially resolved gas velocity values ($\sim \pm 300 km/s$)
- Very low SFR (log SFR $[M_\odot\ yr^{-1}] < -2$)

3.5 Procedure

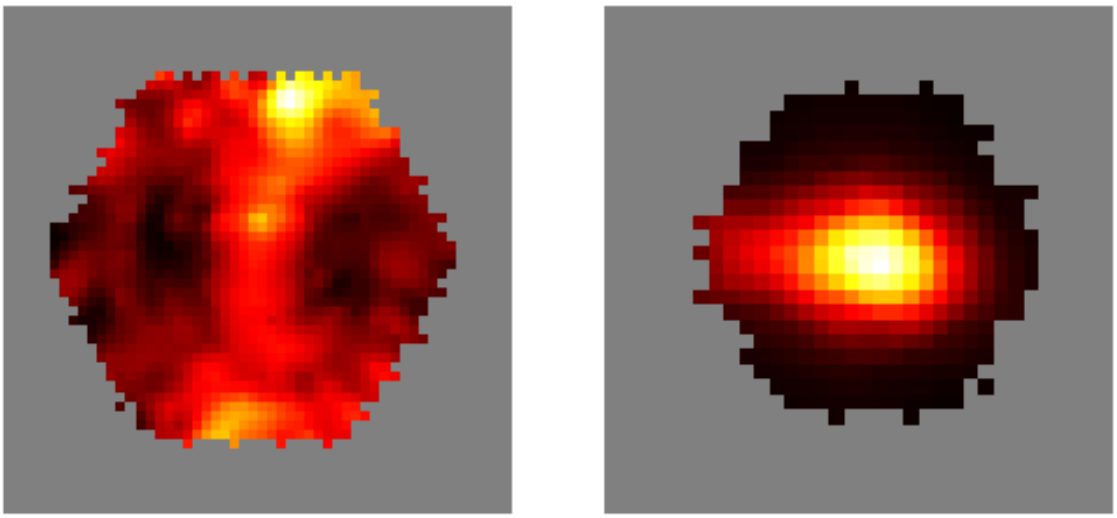
Here I outline the procedure adopted to obtain each of the MaNGA data products for my work. It is to be noted that the API calls to download and prepare each one of these images was extremely time consuming, and had to be run for a long time ($\sim$ days) to completely obtain the datasets.

As for creating the image itself for using in my model, since the Marvin API provided data as a matrix, I used `matplotlib.imshow` with `hot` colour scheme to colour the image. The pixel values outside the hexagon of observation were masked using masks available with each MaNGA galaxy observed, and coloured grey. The minimum and maximum value chosen was determined by the minimum and maximum value of the masked image, thereby providing excellent enhancements to the colour variation across the image. The data was stored in the PNG format in Google Drive.

3.5.1 Hα maps

For the domain adaptation step (to be discussed later in Methods), we require Hα maps of star forming, green valley and quenched galaxies. Using the recipe in [28], I distinguished the three as

- Star forming if $sSFR \geq -10.8$



(a) Red Geyser (b) Non Red Geyser

Figure 3.4: Representative examples of each class of galaxies used in the classification problem

- Quenched if $sSFR \leq -11.8$
- Green Valley if $-11.8 < sSFR < -10.8$

Using the entire final release of the MaNGA survey, I classified the sample as described above.

3.5.2 Red Geysers

In the penultimate release of the MaNGA survey, consisting of $\sim$ 4700 galaxies, 139 were identified as red geysers. With the MaNGA IDs of these in hand, I obtained the MaNGA IDs of the other galaxies from publicly available data. Since the most telltale sign of a red geyser is its bisymmetric Hα EW feature, I used the Marvin Python API to obtain the Hα EW maps of the entire sample. The 4700 odd prior data release examples were used for training, validation and testing purpose [discussed later in Methods].

Chapter 4

Methods

In this chapter, we shall motivate the need for data driven methods for solving the problem of identifying red geysers, and provide an overview of the various ML algorithms used in solving the problem.

4.1 Why Data Driven Methods?

One can approach the problem of identifying red geysers in many ways - manually sorting through the entire dataset and hosting a citizen science project are two of the simplest ways to do so.

However, these methods have their drawbacks. While one can work through $\sim$ 10000 galaxies in a fortnight, it is not practical use of human resources. In the current era of Big Data Astronomy where larger surveys with terabytes of data are collected every night and 10^4 observations every night are becoming the norm, this method does not scale well either. Citizen science projects are not very reliable sources of data as untrained non-experts without active participation in the science project is a quality and ethical issue.

With the rapid rise of AI/ML as a tool for performing scientific data analysis in the past decade, it is tempting to try and implement techniques from this field for our problem. ML algorithms are fast, easy to use and deploy and major strides have been made in making them interpretable and tractable. Given these pros, we decide to use them for our problem of identifying new red geysers.

4.2 Machine Learning

4.2.1 Background

Machine learning is a method used to identify relationships in data without explicitly programming the instructions for how to do so. In this sense, it is agnostic to domain of use.

Let us suppose we have a dataset **D** such that

$$\mathbf{D} \equiv \{(\vec{x}_i, \vec{y}_i)\}_{i=1}^{n} = \mathbf{X} \times \mathbf{Y}$$

of size n where $\vec{x} \in \mathbf{X}$ is the *feature vector* quantifying different details of each element of **D** and $\vec{y} \in \mathbf{Y}$ is an *output vector* having some desired information about the object that we would like to predict. The goal of machine learning is to identify a function f defined as :

$$f : \mathbf{X} \longrightarrow \mathbf{Y}$$

where

$$f(\vec{x}) = \vec{y}$$

The idea is that once we have identified a function f that can predict $\vec{y}$ with some reliability, we can use it on an instance of $\vec{x}$ that has unknown $\vec{y}$ to predict a reasonable expectation for what $\vec{y}$ should be. The central problem of machine learning is to identify methods that can obtain good approximations of f for different types of $\vec{x}$.

The form of $\vec{x}$ varies; in some instances, it can be as simple as an element of a table of data. In certain areas such as astronomy, a lot of data collected is in the raw image format and can take the form of a $p \times q$ image. The specific form of the function f is different in both cases.

Procedure for obtaining a useful function approximator

The basic strategy employed to find f is to input an instance of $\vec{x}$, identify what the output $f(\vec{x})$ is, and iteratively change f such that $f(\vec{x}) \to \vec{y}$. It is standard practice to perform this iteration towards

a "good" f (referred to as *training*) using a *training set* $\mathbf{D}_{train} \subset \mathbf{D}$, perform reality checks during training (referred to as *validation*) using a *validation set* $\mathbf{D}_{valid} \subset \mathbf{D}$ and *test* the performance of f on a *test set* $\mathbf{D}_{test} \subset \mathbf{D}$.

We impose that $\mathbf{D}_{train} \cap \mathbf{D}_{valid} \cap \mathbf{D}_{test} = \phi$ so as to ensure that the model does not show a performance boost in its evaluation metrics due to predicting data it has already learnt how to fit.

4.2.2 Supervised and Unsupervised learning

In supervised learning, one has access to the $\vec{y}$ during training, whereas in unsupervised learning, there is no $\vec{y}$ available during the training procedure.

Principal Component Analysis

Principal Component Analysis (PCA) is an unsupervised learning technique that is used for dimensionality reduction [40]. It does so by identifying directions in the high dimensional space that have maximum variance, and projecting the image along those directions with maximum variance. What this does is accentuates the discriminatory features of the data, thereby making it easier to perform downstream tasks like classification.

Images are an unusual case when it comes to dimensionality reduction using PCA, as they are not only high dimensional but have some spatial information embedded in the form of relative location of entries in their data instances. While one obtains a set of "eigen-images" when performing PCA on images, it is difficult to say whether these truly represent features of the images that are of any significance for classification, as the PCA algorithm treats the entire image as a 1D array thereby discarding any of the semantic features of the images.

K Nearest Neighbour Classification

The k-nearest neighbor (kNN) algorithm is a simple but effective non parametric, supervised classification algorithm which classifies data points based on the class of its k nearest neighbors in the training dataset. [41] [42]

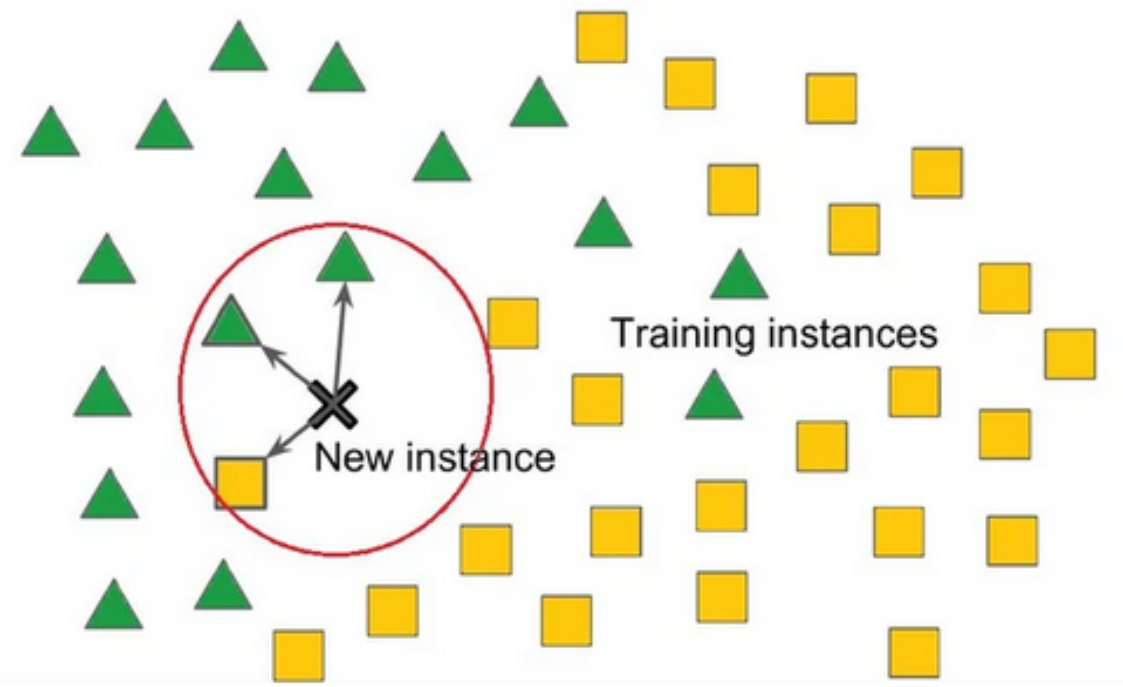


Figure 4.1: Graphical representation of the k-nearest neighbour classification algorithm

To explain with an example, suppose we have a classification problem, where we have to classify a ball with varying greyness as being black or white. We project the ball's features on the space where we wish to perform the kNN classification. This is typically chosen as a low dimensional space to minimise unrelated variations due to the "curse of dimensionality", and to reduce computation time. In our case, since we have only one discriminatory feature i.e. the greyness of the ball, we perform kNN classification in this 1D space. This space has already been populated by all the labelled examples at our disposal.

We then compute the distance (using a distance metric; usually the Euclidean distance is used) to the nearest neighbours of our test ball, and select the top k balls. We assign to the ball the class which has the maximum number of representatives in the list of k closest balls.

The choice of k is very important parameter here; it decides how our classifier behaves. A smaller value of k will make a more flexible classifier but be more sensitive to noise in the data whereas a larger k will give a smoother "classification boundary" in the classification space, at the expense of poor classification in regions that have complex data distributions.

Some advantages of using a kNN algorithm for classification are :

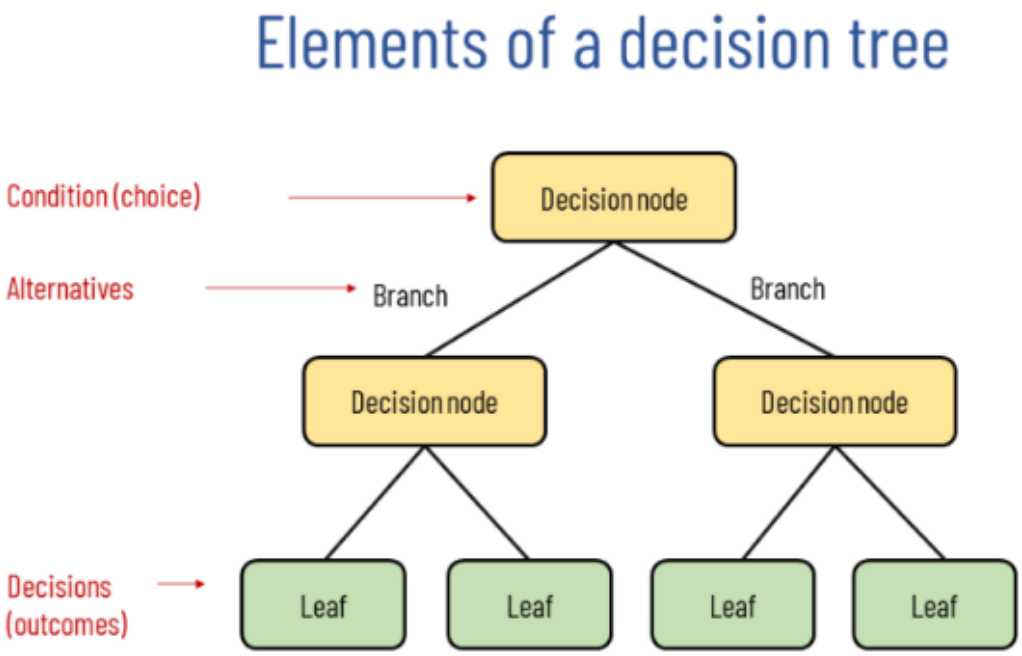


Figure 4.2: Graphical representation of the working principle of a decision tree for classification

- Easy to interpret
- Applicable to a wide variety of data
- Non parametric, so does not make any assumptions about the underlying distribution

However, there are some drawbacks as well :

- The distance computation is a fairly computationally expensive process, especially in high dimensional spaces
- It is sensitive to the choice of the distance metric
- Performs poorly in high dimensional feature spaces

Decision Trees

A decision tree is a non paramteric, supervised classification algorithm that produces a graphical representation of all possible decisions and their potential consequences, and is built by recursively

splitting the data based on the value of certain attributes, until each branch ends in a terminal node or a decision. [43]

The easy interpretability of decision trees makes them a very useful model to deal with for classification tasks. However, they are not good enough to perform classification using raw pixel values, and have to undergo a feature extraction step before classification.

Decision trees work best when predicting the class label based on attributes. It does so by selecting the attribute that maximizes information gain, which is measured as a reduction in entropy that results from splitting the data on that attribute. This process is repeated recursively until each ”leaf” node contains instances that have the same class label.

Of the many drawbacks of decision trees, the main one is overfitting to the test dataset. To avoid overfitting, various techniques such as pruning, setting a minimum number of instances per leaf, or using ensemble methods like random forests are used which mitigate the effects of overfitting in different ways.

4.2.3 XGBoost

The state of the art machine learning model used extensively in the industry, Kaggle competitions and research problems is XGBoost [44]. In order to understand XGBoost, it is important to understand certain key ideas that are briefly discussed here :

Ensemble Learning

In ensemble learning methods, one combines many ”weak learning” models (these are individually poorly performing models) to create a ”strong learning” model that has better predictive capacity than any of the individual models. They are particularly useful in improving accuracy and robustness[citation, citation] by working around overfitting, except in the case of noisy datasets using certain approaches.

Ways to implement ensemble learning include :

- **Bagging :** Also called bootstrap aggregation. In this technique, one bootstraps the data by

picking samples with repetition to form sub-datasets, followed by ”voting” on predictions made by multiple weak learners to identify the correct prediction.

- **Boosting :** From a classification perspective, one starts with a dataset D1 and a model M1, trains it and marks the misclassified examples. A new dataset D2 is made where the misclassified samples from D1 are given greater weightage and run through an updated model M2. This allows for poorly classified examples to be properly classified in the final model. This process is repeated iteratively till a certain preset threshold of misclassified samples is reached. Then, the weak learners are combined in a manner such that they are tuned to classify the misclassified examples they were specialized on, and the final classification is done by voting on the predictions of each of the weak learners.

Gradient Boosting

Gradient boosting is a form of boosting where the iterative procedure of training weak learners to perform classification on the given dataset is formalized as a gradient descent problem. It does so by providing targets for subsequent models to reach in terms of a target function.

XGBoost is a particular implementation of this boosting algorithm that is fast and parallelized. Instead of stacking trees one on the other, it parallelizes them to speed up the inference and training procedure.

4.3 Deep Learning

4.3.1 Model Architecture

ResNet

Residual Networks (ResNets) are a deep learning architecture that are very good at image classification problems [45]. It is capable of having large depths (many layers) as it has skip connections which address the vanishing gradient problem efficiently.

The skip connections address the vanishing gradient problem during backpropogation by pro-

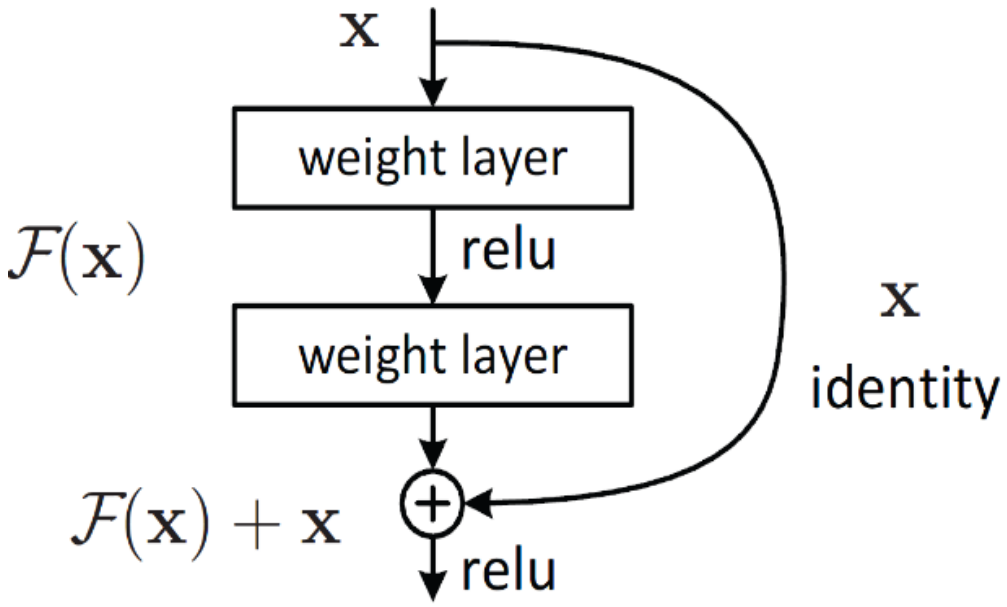


Figure 4.3: A simplified pictorial representation of a ResNet, displaying the skip layer protocol

viding gradient highways which allows the gradient to travel across convolutional layers, which can diminish the gradient.

This allows for constructing very deep layers, which can effectively learn good discriminative filters for image classification.

4.3.2 Data Augmentation

Since we have less data, can we do something to make sure that we can get *more* data from the less data? Data augmentation is a procedure by which one can achieve this. [46] [47]

Traditionally, in common machine learning settings, data augmentation is applied to reduce

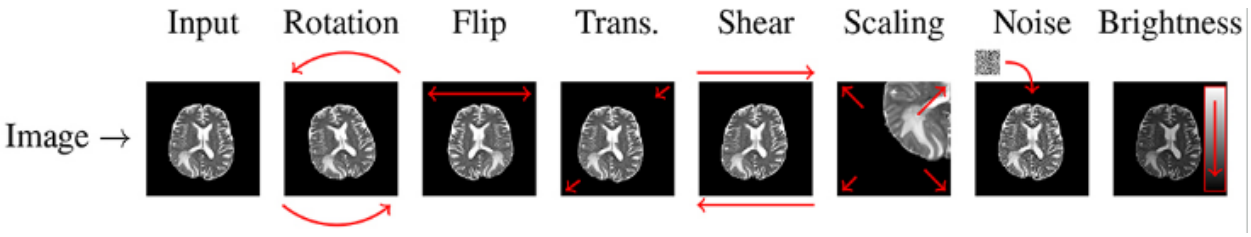


Figure 4.4: An assortment of augmenting transformations on an image

overfitting to the model data during training. For image classification, one applies various transforms that alter certain meta features of the image, leaving most of the image content similar. That's not to say there aren't any transformations that change the character of the image itself!

Some of the most commonly used transformations are show in the Fig.

Is it correct to "increase" the training data size by performing data augmentation? For the case of very small image datasets like ours, applying transforms to attempt this can infact lead to overfitting [48], and is also unphysical for the following reasons :

- Colour transformations remove essential information about the relative pixel values in the $H\alpha$ EW maps
- The hexagonal IFU structure gives us a six fold rotational symmetry and three reflective symmetries. These do not add any additional information about the structure of the red geyser, and with such a small sample will overfit the model
- Normalizing the image will create $H\alpha$ EW maps of galaxies that do not exist

4.3.3 Transfer Learning

One of the most common ways in which classification tasks with less training data is dealt with is by adopting a transfer learning approach. [49] [50]

Suppose that $\mathbf{D_S} \equiv \mathbf{X_S} \times \mathbf{Y_S}$ is the source dataset. We have a large number of labelled samples in this dataset, on which we train a model $f_S : \mathbf{X} \longrightarrow \mathbf{Y}$. Suppose that this model f_S is a very good (by means of some measure of how correct it is on the given data).

Suppose that we have a related dataset $\mathbf{D_T} \equiv \mathbf{X_T} \times \mathbf{Y_T}$. By related, we mean that there is some similarity between $\mathbf{D_S}$ and $\mathbf{D_T}$ in the sense that a human can perceive them to have some commonalities, but with the constraint that $\mathbf{D_S} \neq \mathbf{D_T}$. And also suppose that we do not have a lot of labelled examples in $\mathbf{D_T}$. Can we use the fact that $\mathbf{D_S}$ and $\mathbf{D_T}$ are related to carry over the model f_S to identify a model that can classify $\mathbf{X_T}$ efficiently, without extensive training?

When we train a model on a source dataset $\mathbf{D_S}$, it will learn the distribution of the data, called an *inductive bias*. When this model encounters an unseen example from the same distribution,

it will classify it with a decent degree of accuracy. In transfer learning, we push the limit of what it means for an example to belong to the same distribution, by choosing data from a related distribution instead.

How would one implement this in the case of a deep neural network? In the case of image classification, one of the most common ways of implementing transfer learning is by starting with a model with very good inductive biases, where good is defined as very generalizable source distribution. We then freeze all layers except the final few layers of the model and retrain these layers using the target dataset. These last few layers have weights corresponding to the highest level discriminative features for classification which are specific to the problem distribution at hand. In this process of freezing out the lower level layers, we can retain the lower level features which were learnt during training on the source dataset, which we can combine with the specific features learnt for the target dataset to achieve a better classification accuracy.

4.4 Few Shot Learning

Humans are extremely efficient at image classification - we are able to learn discriminative features of objects from a small sample of their images, and generalise these to classify the objects and their images successfully. Since image classification using CNNs is largely based on how biological neural networks, can we recreate this data efficient classification using artificial neural networks too?

Starting from a untrained model, it is a rule of thumb that standard machine learning techniques require $\sim$ 10 times more data than the number of model parameters to reliably fit a regression model. For deep neural architectures, it is recommended to have 1000 images per class for a classification model. For transfer learning and pre-trained models, the number of training examples required is lesser, by about 1.2-1.5 times.

Our case is unique in that we only have $\sim$ 140 galaxies in our sample which we can use for training. It is then imperative that we use data efficient classification methods for our problem. So far, we have addressed this issue using an imbalanced training dataset, but can we do better?

We try the ”few shot learning” approach of classification to solve our task [51] [52]. This paradigm follows the philosophy of **meta learning**, where the idea is to *learn how to learn*. One

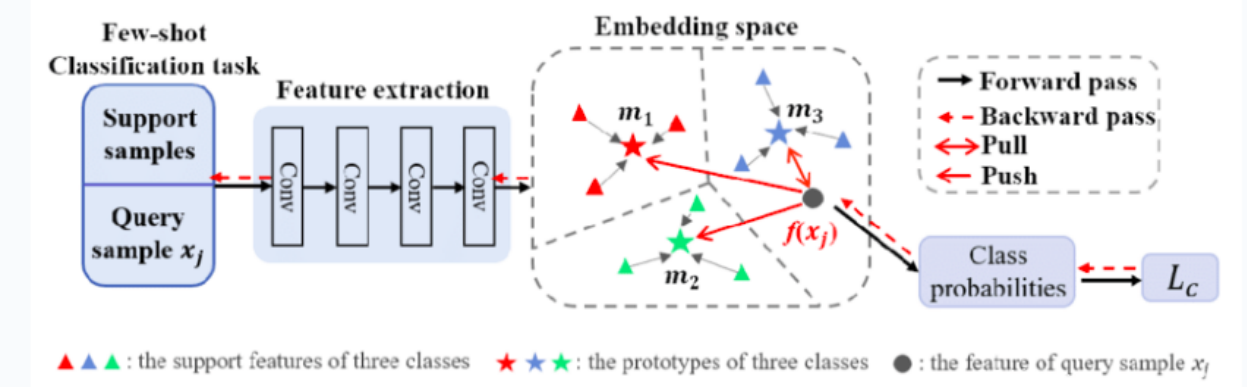


Figure 4.5: A schematic representation of a prototypical network implementation

trains the model on multiple classification tasks so that it may be tested and used to perform classification on hitherto unseen classes.

While there are plenty of techniques to approach few shot learning, we chose to work with prototypical networks.

4.4.1 Prototypical Networks

If we take a child to a zoo and ask it to identify animals it has never seen before, it is highly unlikely it will be able to do so satisfactorily. However, if we show it images of different animals and ask to identify the animal type based on the image representations, it is bound to correctly identify the correct animal with a high degree of certainty.

Prototypical networks were introduced as a method to perform few shot learning [53]. The philosophy is very similar to how a child (a metaphor for an untrained network) is able to identify hitherto unseen animals (classes) from representative images of the animals (image prototypes). The extensibility of prototypical networks is immense - it is not only adept at classification using few ”training” images but it also generalises to new classes very easily. We will not discuss the extensions of this method here but only that which interests us for our problem.

The approach is as following - we start by defining an *image embedding map* from the space of images to a lower dimensional *embedding space*

$$f_\phi : \mathbf{R}^{p\times p} \longrightarrow \mathbf{R}^m$$

Here f_ϕ is a function f with parameters ϕ.

This function is not a given when we start the problem. These functions are non trivial to define explicitly and we thus use the universal function approximation property of neural networks to represent this function. Our goal is to identify a good functional form (neural network architecture) and its parameters (weights and biases of the network) for this embedding function.

Suppose that our problem requires us to classify K classes of images. Let S_k be the set of images which belong to class k. In the embedding space $\mathbf{R}^m$, we define the prototype representation of the shot images as follows :

$$\mathbf{c_k} = \frac{1}{|S_k|} \sum_{(\mathbf{x_i}, y_i) \in S_k} f_\phi(\mathbf{x_i})$$

Let us define a metric d on this embedding space :

$$d : \mathbf{R}^m \times \mathbf{R}^m \longrightarrow [0, \infty)$$

Our prototypical network then can produce a distribution over classes for an input image $\mathbf{x}$ by means of a softmax on the distances between the embedding and prototypes in the embedding space :

$$p_\phi(y = k|\mathbf{x}) = \frac{exp(-d(f_\phi(\mathbf{x}), c_k))}{\sum_{k'} exp(-d(f_\phi(\mathbf{x}), c'_k))}$$

By increasing the probability for an image to lie close to its target class, we can hope to achieve the desired embedding function. We can achieve this by defining a loss function which takes smaller values when the function parameters ϕ take values that place the image embedding close to the prototype of the shot images :

$$J(\phi) = -\sum_{k'} log[p_\phi(y = k|\mathbf{x})]$$

The training proceeds by iteratively attempting to minimize this loss function using the algo-

rithm below :

Algorithm 1 Computing the loss for updating parameters. We have N (number of examples in the training set), K (number of classes in the training set, N_S (number of support examples per class, N_Q (number of query samples per class), RANDOM SAMPLE (S,N) (random sample of N elements from set S)

for k in $\{1,\ldots,K\}$ **do**
  $S_K \leftarrow$ RANDOM SAMPLE$(\mathbf{D}_K, N_S)$
  $Q_K \leftarrow$ RANDOM SAMPLE$(\mathbf{D}_K \setminus S_K, N_Q)$
  $\mathbf{c_k} \leftarrow \frac{1}{|S_K|} \sum_{(\mathbf{x_i}, y_i) \in S_k} f_\phi(\mathbf{x_i})$
end for
$J \leftarrow 0$
for k in $\{1,\ldots,K\}$ **do**
  for (x,y) in Q_K **do**
    $J \leftarrow J + \alpha \left[d(f_\phi(\mathbf{x}) + log \sum_{k'} exp(-d(f_\phi(\mathbf{x}), c'_k)) \right]$
  end for
end for

Chapter 5

Results and Discussion

The problem of identifying red geysers from the latest MaNGA release is an onerous task. The conventional method, as detailed in the Data section, requires one to sift the complete catalog after applying selection cuts followed by manually sorting through the residual dataset to identify the bisymmetric pattern.

In order to apply machine learning models to automate this task for us, we choose to classify whether or not an input Hα EW map has bisymmetric features or not. The rationale is that the $H\alpha$ bisymmetric feature is the most visually distinctive and discriminative characteristic of the red geyser population. It is also the most simplest for a human to recognise, and thus we expect the machine to easily identify this pattern. From a selection perspective, it is also the most tell-tale sign of the existence of a red geyser, closest to being the smoking gun, as there are other galaxy populations which share some of the other properties of red geysers.

In the first section, I will first present all the results obtained from our experiments. Then I will discuss their implications in the following section. Generalization and how it sets up future work will be discussed in the sections afterwards. This chapter will conclude with a summary of results and their implications in the broader scheme of solving the problem of identifying red geysers.

5.1 Image Classification Baseline Models

Since this is the first model of its kind, we need to implement a baseline model with respect to which we can test our more advanced models.

5.1.1 The Naive Approach - ZeroR Classifier

One of the baselines used to assess whether a classification model is performing well or not is the random classifier - this is a model that randomly assigns a class to the input object. In a binary classification task, one can assign a probability p with which the model assigns a sample as belonging to the positive class. In this case, during testing, if the test sample has n_+ positive samples and n_- negative samples, the accuracy would be given as :

$$Accuracy = \frac{n_+ p + n_-(1-p)}{n_+ + n_-}$$

if $p = 0.5$, we see that we should have a baseline accuracy of 0.5.

However, in the case of an imbalanced dataset, we can remodel the accuracy baseline. Note that in the limit of $n_+/n_- \ll 1$, the accuracy tends to $1-p$. If we set $p = 0$, i.e. the model always predicts the larger negative class, we would get an accuracy $\to 1$. This is called a **ZeroR Classifier**.

For our example with $n_+ = 139$ and $n_- = 4587$, the ZeroR Classifier has an accuracy of 96.97%.

5.1.2 kNN Classifier

Since I will be using a metric based deep learning model, for comparision I implemented a simple ML based kNN to compare the results.

Implementing a kNN classifier directly on images without performing any feature extraction is a poor idea as kNN performs poorly on classifying high dimensional data and there is disregard for the structural features of the image. This is because they have features such as the bisymmetric

structures which need to be semantically encoded, and a simple subtraction would not be able to capture the invariance of alignment of this jet feature.

We perform feature extraction using an unsupervised technique - PCA. Despite being a simple and reliable feature extractor, there are some drawbacks :

- Feature extraction step is not interpretable
- The results of feature extraction are stochastic (can be reproduced with seed)
- PCA tries to minimize reconstruction error and not maximize the difference in features between the many classes

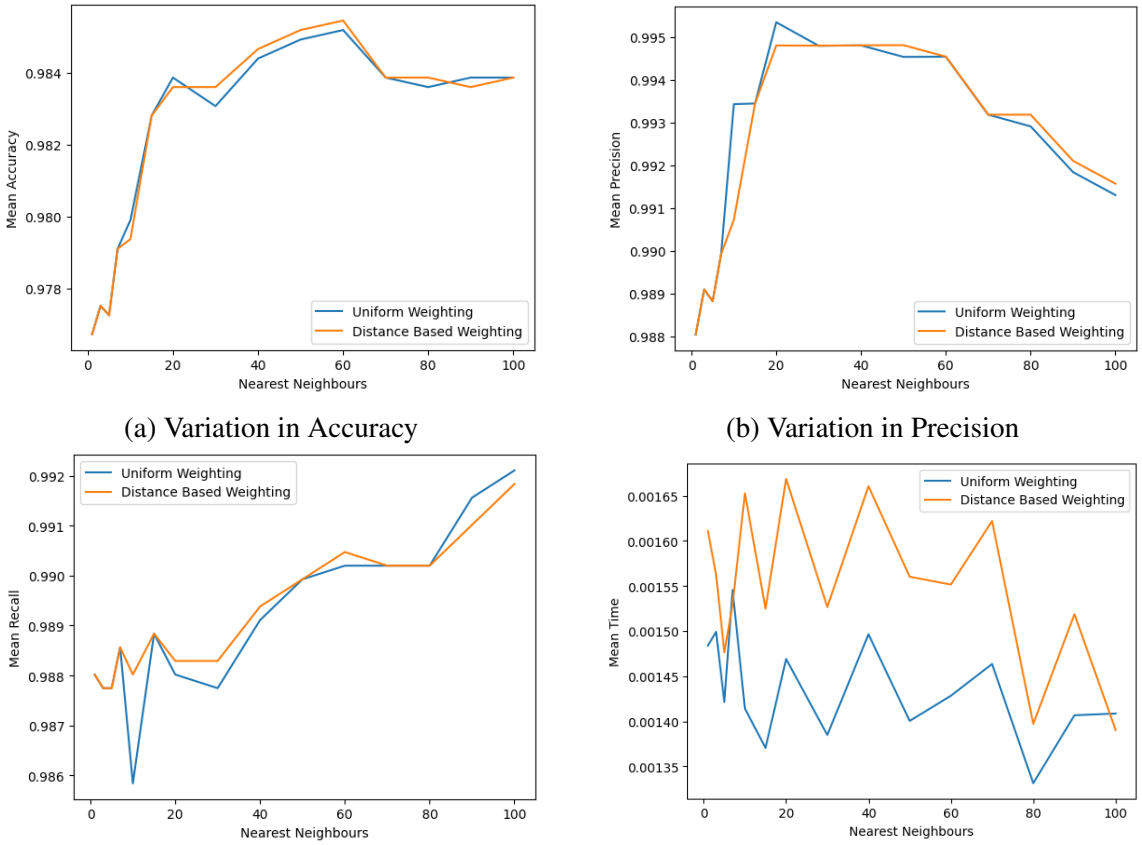
However, the PCA step does provide us with ”useful” features that we can use to implement standard ML techniques.

Before performing feature extraction, we reduce the dimensions of each image to (64,64) as it becomes too expensive to perform PCA otherwise. There were two important hyperparameters to optimize - the number of principal components to choose when performing PCA and the number of neighbours to consider when performing kNN i.e. the k in kNN. A common mistake is tuning the hyperparameters using the output from the test set - this is incorrect usage as the model is then optimized to perform well on the test set itself, and will very obviously yield good results. What is required is a seperate validation set to perform hyperparameter tuning on, which is set aside apart from the training and test sets at the beginning of the program.

For this purpose, we use the `GridSearchCV` option of the `sklearn.model_selection` package which performs a grid search over input hyperparameters to identify the best performing model using cross-validation.

Even though we would like to maximize recall, we do not choose hyperparameters that optimize on recall. We instead choose to optimize on accuracy. The reason being that optimizing on recall will push to model to mark all examples as red geyser, as evident from Fig 5.1c. We thus stick to accuracy.

Our hyperparameters are chosen as $k = 60$. We however choose `weights=uniform` as it is a faster method.



(a) Variation in Accuracy

(b) Variation in Precision

(c) Variation in Recall

(d) Time for Running

Figure 5.1: Hyperparameter tuning for the KNN model

We have a final accuracy of 0.9841, a precision of 0.7353 and a recall of 0.8064

Our experiments reveal that the PCA did a good job of extracting the desired features - the kNN algorithm on the PCA vectors performed better than the brute force ZeroR Classifier.

5.1.3 Decision Trees

Similar to kNN, we used PCA to extract features from the images and we used a decision tree algorithm provided by `scikit-learn` to perform the same binary classification task. We used a reduced image size of (64, 64) as in the previous problem, and set aside a validation set to perform hyperparameter tuning.

Apart from providing us with one more baseline test, the performance of decision trees serves

as a benchmark to compare the XGBoost performance, as we saw in the Methods section how XGBoost is an ensemble learning method with decision trees as its foundational stone.

By default, the program uses a `gini` criteria to quantify the difference for performing the split. We chose hyperparameter values for the trees as follows - maximum tree depth as 2, minimum sample split as 2 and minimum samples leaf as 1. Our choice was motivated by the following considerations - the space over which data is distributed is specifically covered when there is atmost one leaf per node, the generalizability best when the depth is low and most logical and interpretable (if required) when the sample split is 2.

The classification accuracy was 0.9831, the precision was 0.7419 and the recall was 0.7419.

5.2 XGBoost

5.2.1 Choice of Parameters

The preferred booster for standard XGBoost applications is a gradient boosted tree (`gbtree`). The other important hyperparameters such as `learning_rate` (η), `minimum_split_loss` (γ) and `maximum_tree_depth` are taken to be the standard default values of 0.3, 0 and 6 respectively.

5.2.2 Evaluation of Predictions

Since XGBoost is an excellent learner of the training data, we expect it to perform satisfactorily well on the test data. It exceeded expectations and had a classification accuracy of 0.9831, a precision of 0.9923 and a recall of 0.9902.

The one issue with XGBoost is that despite performing really well on this dataset, it has no guarantee of performing well on a different dataset as it is very poor to generalization (its weak learners are trees).

Batch Size	Epochs	Valid Tasks	Valid Frequency	Valid Accuracy	Test Accuracy
128	200	20	10	0.583	0.5375
128	20	20	10	0.55	0.5745
128	50	20	4	0.567	0.561
128	30	20	5	0.533	0.568
128	20	5	4	0.8	0.5970

Table 5.1: Training a ResNet on Hα maps for domain adaptation

5.3 Domain Adaptation Experiments

Since there are no pretrained models on the MaNGA dataset at the present, as discussed in the Methods and Data sections, we train a 10 layered ResNet to classify the galaxies based on their star forming rates into star forming galaxies, green valley galaxies or quenched galaxies. The results of the training and testing is provided in Table 5.1

First thing to note is that the test accuracy barely reaches 60%. Since this is a different dataset than our red geyser dataset (see Data), the baseline metrics for comparision are also different. The baseline here is simply a random classifier, which comes out to have an accuracy of 50%.

Our model thus performs only marginally better than a random classifier. There may be many reasons for this :

- The model was not trained for a sufficiently large number of epochs
- The encoder is not able to efficiently learn the features present in the images
- Since we are dealing with Hα maps and the principal discriminatory component is overall magnitude of the galaxy image, the 10 layered ResNet maybe too deep and may be missing out on extracting the correct features
- The presence of masked pixels in the training dataset could cause certain regions of the network to be improperly trained

One seemingly strong indication that our model is unable to pick up useful features is the training run with lower number of epochs having a better classification accuracy when compared

to the other ones. This means the pre-existing inductive biases are better at picking up differences if any than the ones learned by the model.

Possible directions in trying domain adaptation include :

- Working with a different dataset for creating a model for domain adaptation
- Training for more epochs on the dataset to learn the parameters more effectively

5.4 Prototypical Networks

5.4.1 Choosing the optimal model - Hyperparameter tuning

Shot Size

In few shot learning, the number of shots required to predict the class is an important choice of hyperparameter during training as well as testing. During training, the goal is to learn a generalizable representation by trying to minimize a suitable loss function, whereas in testing we typically aim to obtain a large enough accuracy.

The number of shot images thus chosen is of importance - in the limit of a large sample, a larger shot size will produce a better prototype, whereas a smaller shot size will be prone to anomalous shot examples dominating the contribution to the prototype. However, having a larger shot size will have an overhead in calculating the prototype, whereas smaller shot size will be faster during training and testing. It is imperative that we decide an optimal shot size for training and testing.

Another important question regarding shot size is whether they have to be the same for training and testing. This is very dependent on the size of these sets - sometimes we may be constrained by the number of examples during training or testing to use a smaller shot size.

We use a shot size of 5 as it represents the best among accuracy of prediction and speed of inference.

Encoder	Classifier	Accuracy	Precision	Recall
-	ZeroR	0.9697	0.0300	0.0000
PCA	kNN	0.9841	0.7353	0.8064
PCA	Decision Trees	0.9831	0.7419	0.7419
ResNet	**ProtoNet**	**0.9905**	0.5798	**0.9928**
PCA	XGBoost	0.9831	0.9923	0.9902

Table 5.2: Classification, Precision and Recall Scores for different ML techniques used

Depth of the model

As we saw earlier, the ResNet is a demonstrably good choice for image classification tasks. The choice of the number of ResNet layers in the model is an important hyperparameter - having a lot of layers may not provide a significant computational advantage over lesser number of layers. In fact, it is recommended not to have too many layers as the model will have too many parameters and thus be prone to overfitting. A larger model will also take longer to train, since there are more parameters whose gradients need evaluation.

By training and testing on the $H\alpha$ flux dataset, it was identified that the most optimal depth for the encoder was a 10 layer ResNet.

5.4.2 Prototypical Network Implementation

We implemented prototypical networks as described in the Methods section. We trained the model using "classic training", where we sample data from each class, compute its prototype and compare the distance to the feature space representation of the image. We perform validation on a regular basis to ensure our model is not overfitting on the training data.

In our case, we trained for 50 epochs with a batch size of . We performed 15 validation tasks every 5 epochs. We achieved a validation accuracy of 1.0, our training loss went down to 0.000459 and we achieved a test accuracy of 99.05%

5.4.3 Classification Metrics for ProtoNets

In a binary classification task with a class imbalance, it is improper to consider the classification accuracy with much merit, as the class imbalance will tip the classification accuracy towards higher values. In fact, this is clearly exemplified by the high classification accuracy of the most naive classifier aka the ZeroR Classifier.

Instead, we look at precision and accuracy, both measures of how ”clean” our classifier is in classifying the positive class. Precision tells us how contaminated our sample is, i.e. how many objects that are not positive get misclassified as positive; and recall tells us how complete our sample is, i.e. whether we've managed to catch all objects that are of the positive class when classifying.

We can express this in the form of a confusion matrix, as in table.

While ideally we would like to have a very high recall and precision, it is not always the case. For our particular example, since we would like to identify as many red geysers as possible, we demand that we have a high recall and are satisfied with a reasonably large precision.

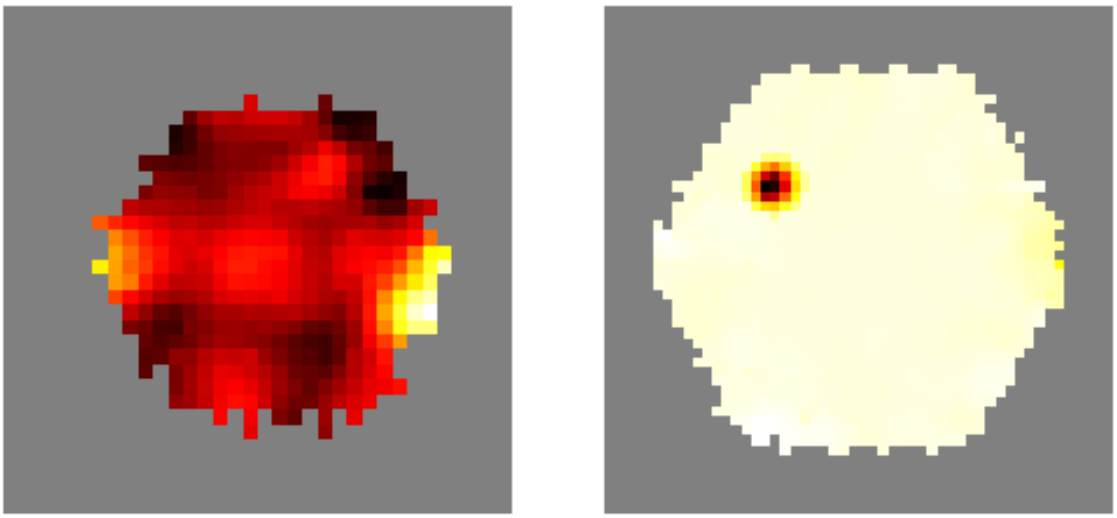
This is indeed the case; we have an exceptionally high recall > 0.99, which we hope can translate into identifying ne red geyser galaxies.

5.4.4 Misclassified galaxies

On closer inspection of the misclassified galaxies we find that the false negative example is actually a very poorly rendered image of the galaxy. This serves as a very good reality check for our model, as we know it is not spuriously performing well on our final test sample.

		Actual Classes	
		Red Geyser	Non Red Geyser
Predicted Classes	Red Geyser	138	100
	Non Red Geyser	1	4487

Table 5.3: Confusion matrix for the entire dataset classified using ProtoNets with a ResNet encoder



(a) False Positive (b) False Negative

Figure 5.2: Representative examples of each class of galaxies used in the classification problem

Many of the false positive galaxies have bisymmetric structures in them and require of a seperate inspection by an expert. Many others, like in Fig 5.2a, have a many centers of enhancements in their Hα EW maps. This is very characteristic of the Hα disturbed galaxies seen in the Data section. It implies our model is performing well to predict galaxies with these features, and they are all interesting enough to be worthy of a closer inspection before deciding whether or not to classify them as red geysers or not. Their numbers are also very manageable; they are of the order of the number of red geysers themselves so it would not be too taxing on the part of the expert to manually examine each one of them.

5.5 Generalizing to unseen examples

The task of generalizing to previously unseen training examples is the bane of ML models; the training procedure involving minimizing the loss function only penalizes the model if it deviates too far from predicting the manner in which the training data is distributed. One can add more constraints; techniques such as regularization (penalizing the model for having too many parameters that can lead to overfitting) and dropout (dropping certain fractions of neurons during training at every iteration) are some ways in which the overfitting issue can be solved. However, a faithful fit over the complete space of data is still not possible and this leads to very prominent drawbacks

when attempting to generalize over test and other similar unseen datasets.

5.5.1 XGBoost

The poor generalizing capabilities of decision trees is a well known phenomena; their inability to create partitions such that their leaves create a cover over the space of data is the chief culprit [cite Bengio 2010] apart from their tendency to fit the noise in the data. Algorithms like XGBoost provide an improved performance on the test set by creating an ensemble of tree based learners; however they still suffer from their tendency to fit the noise in the dataset and the mathematical constraint as described in [Bengio 2010] that prevents them from generalizing well.

In our case, despite performing with the best overall performance parameters among all other models, XGBoost fared pretty low when generalizing.

Few Shot Learning methods such as prototypical networks work very well for generalizing to unseen classes

5.6 Future Scope

So far, our discussion has been mainly on how the model has performed on the training and test datasets. The excellent performance on this smaller subsample means we have a working model that has passed the proof-of-concept test.

The immediate next step would be to test it on the full dataset. Preliminary excursions in this direction were not very fruitful; the models failed to generalize to the larger dataset. There can be many reasons for this :

- XGBoost is known to be a poorly generalizable algorithm; it may not be suitable to work on this problem
- PCA as a dimensionality reduction tool for images only manages to minimize the reconstruction error by design, this is not something we really need for our problem
- ResNet encoders of small size are also poor at generalizing to related datasets

Some remedies for these drawbacks include :

- Instead of using tree based gradient boosting, other options such as linear function based boosting could be implemented
- SimCLR [54] is a contrastive learning based dimensionality reduction techinque; its core tenet is to maximize the difference between different classes and is thus ideal for classification tasks
- We can try implementing deeper ResNets for our task

Our models after passing preliminary tests are very close to realizing its ultimate goal of identifying new red geysers from the latest MaNGA release.

We have also demonstrated that the few shot approach can also be used in an astronomy case. A natural extension would be to use it in other astronomy use cases where there is not a lot of training data available. Some prominent examples requiring an urgent intervention include the process of detecting gravitational waves and fast radio bursts in time domain astronomy.

Bibliography

[1] *Galaxy Formation and Evolution - NASA/ADS*. URL: https://ui.adsabs.harvard.edu/abs/2010gfe..book.....M?bbbRedirect=1 (visited on 04/01/2023) (cit. on pp. 5, 14).

[2] James Binney and Michael Merrifield. *Galactic astronomy*. Princeton series in astrophysics. Princeton, NJ: Princeton University Press, 1998. ISBN: 978-0-691-00402-0 978-0-691-02565-0 (cit. on p. 5).

[3] Linda S Sparke and John S Gallagher Iii. "Galaxies in the Universe: An Introduction, Second Edition". en. In: () (cit. on pp. 5, 6).

[4] Peter Schneider. *Extragalactic Astronomy and Cosmology: An Introduction*. en. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015. ISBN: 978-3-642-54082-0 978-3-642-54083-7. DOI: 10.1007/978-3-642-54083-7. URL: https://link.springer.com/10.1007/978-3-642-54083-7 (visited on 04/01/2023) (cit. on pp. 5, 6, 14, 16, 17).

[5] O. Ilbert et al. "Galaxy Stellar Mass Assembly between 0.2¡z¡2 from the S-COSMOS survey". In: *The Astrophysical Journal* 709.2 (Feb. 2010). arXiv:0903.0102 [astro-ph], pp. 644–663. ISSN: 0004-637X, 1538-4357. DOI: 10.1088/0004-637X/709/2/644. URL: http://arxiv.org/abs/0903.0102 (visited on 04/10/2023) (cit. on p. 6).

[6] John Moustakas et al. "PRIMUS: Constraints on Star Formation Quenching and Galaxy Merging, and the Evolution of the Stellar Mass Function from z = 0-1". In: *The Astrophysical Journal* 767 (Apr. 2013). ADS Bibcode: 2013ApJ...767...50M, p. 50. ISSN: 0004-637X. DOI: 10.1088/0004-637X/767/1/50. URL: https://ui.adsabs.harvard.edu/abs/2013ApJ...767...50M (visited on 04/10/2023) (cit. on p. 6).

[7] Marie Martig et al. "Morphological Quenching of Star Formation: Making Early-Type Galaxies Red". In: *The Astrophysical Journal* 707 (Dec. 2009). ADS Bibcode: 2009ApJ...707..250M, pp. 250–267. ISSN: 0004-637X. DOI: 10.1088/0004-637X/707/1/250. URL: https:

//ui.adsabs.harvard.edu/abs/2009ApJ...707..250M (visited on 04/10/2023) (cit. on p. 6).

[8] Charlie Conroy, Pieter G. van Dokkum, and Andrey Kravtsov. "Preventing Star Formation in Early-Type Galaxies with Late-Time Stellar Heating". In: *The Astrophysical Journal* 803 (Apr. 2015). ADS Bibcode: 2015ApJ...803...77C, p. 77. ISSN: 0004-637X. DOI: 10.1088/0004-637X/803/2/77. URL: https://ui.adsabs.harvard.edu/abs/2015ApJ...803...77C (visited on 04/10/2023) (cit. on p. 6).

[9] Feng Yuan and Ramesh Narayan. "Hot Accretion Flows Around Black Holes". In: *Annual Review of Astronomy and Astrophysics* 52.1 (Aug. 2014). arXiv:1401.0586 [astro-ph], pp. 529–588. ISSN: 0066-4146, 1545-4282. DOI: 10.1146/annurev-astro-082812-141003. URL: http://arxiv.org/abs/1401.0586 (visited on 04/10/2023) (cit. on p. 6).

[10] A. C. Fabian. "Observational Evidence of Active Galactic Nuclei Feedback". In: *Annual Review of Astronomy and Astrophysics* 50 (Sept. 2012). ADS Bibcode: 2012ARA&A..50..455F, pp. 455–489. ISSN: 0066-4146. DOI: 10.1146/annurev-astro-081811-125521. URL: https://ui.adsabs.harvard.edu/abs/2012ARA&A..50..455F (visited on 04/10/2023) (cit. on p. 6).

[11] A. Cattaneo et al. "The role of black holes in galaxy formation and evolution". In: *Nature* 460 (July 2009). ADS Bibcode: 2009Natur.460..213C, pp. 213–219. ISSN: 0028-0836. DOI: 10.1038/nature08135. URL: https://ui.adsabs.harvard.edu/abs/2009Natur.460..213C (visited on 04/10/2023) (cit. on p. 6).

[12] Namrata Roy et al. "Detecting Radio AGN Signatures in Red Geysers". In: *The Astrophysical Journal* 869 (Dec. 2018). ADS Bibcode: 2018ApJ...869..117R, p. 117. ISSN: 0004-637X. DOI: 10.3847/1538-4357/aaee72. URL: https://ui.adsabs.harvard.edu/abs/2018ApJ...869..117R (visited on 04/01/2023) (cit. on pp. 6, 7, 19–22, 29, 30).

[13] Edmond Cheung et al. "Suppressing star formation in quiescent galaxies with supermassive black hole winds". In: *Nature* 533 (May 2016). ADS Bibcode: 2016Natur.533..504C, pp. 504–508. ISSN: 0028-0836. DOI: 10.1038/nature18006. URL: https://ui.adsabs.harvard.edu/abs/2016Natur.533..504C (visited on 04/01/2023) (cit. on pp. 7, 19).

[14] Emily Frank et al. "The H I content of red geyser galaxies". In: *Monthly Notices of the Royal Astronomical Society* 519 (Mar. 2023). ADS Bibcode: 2023MNRAS.519.3312F, pp. 3312–3318. ISSN: 0035-8711. DOI: 10.1093/mnras/stac3784. URL: https://ui.adsabs.harvard.edu/abs/2023MNRAS.519.3312F (visited on 04/08/2023) (cit. on p. 7).

[15] Charlie Conroy. “Modeling the Panchromatic Spectral Energy Distributions of Galaxies”. en. In: *Annual Review of Astronomy and Astrophysics, vol. 51, issue 1, pp. 393-455* 51.1 (Aug. 2013), p. 393. ISSN: 0066-4146. DOI: 10.1146/annurev-astro-082812-141017. URL: https://ui.adsabs.harvard.edu/abs/2013ARA%26A..51..393C/abstract (visited on 03/30/2023) (cit. on pp. 9, 14).

[16] Edwin E. Salpeter. “The Luminosity Function and Stellar Evolution.” In: *The Astrophysical Journal* 121 (Jan. 1955). ADS Bibcode: 1955ApJ...121..161S, p. 161. ISSN: 0004-637X. DOI: 10.1086/145971. URL: https://ui.adsabs.harvard.edu/abs/1955ApJ...121..161S (visited on 03/30/2023) (cit. on p. 10).

[17] Albert Sneppen et al. “Implications of a Temperature-dependent Initial Mass Function. I. Photometric Template Fitting”. en. In: *The Astrophysical Journal* 931.1 (May 2022). Publisher: The American Astronomical Society, p. 57. ISSN: 0004-637X. DOI: 10.3847/1538-4357/ac695e. URL: https://dx.doi.org/10.3847/1538-4357/ac695e (visited on 03/30/2023) (cit. on p. 10).

[18] Robert C. Kennicutt. “Star Formation in Galaxies Along the Hubble Sequence”. en. In: *Annual Review of Astronomy and Astrophysics* 36 (1998), pp. 189–232. ISSN: 0066-4146. DOI: 10.1146/annurev.astro.36.1.189. URL: https://ui.adsabs.harvard.edu/abs/1998ARA&A..36..189K/abstract (visited on 03/30/2023) (cit. on p. 12).

[19] Robert C. Kennicutt Jr., Peter Tamblyn, and Charles E. Congdon. “Past and Future Star Formation in Disk Galaxies”. In: *The Astrophysical Journal* 435 (Nov. 1994). ADS Bibcode: 1994ApJ...435...22K, p. 22. ISSN: 0004-637X. DOI: 10.1086/174790. URL: https://ui.adsabs.harvard.edu/abs/1994ApJ...435...22K (visited on 03/30/2023) (cit. on p. 12).

[20] Piero Madau, Lucia Pozzetti, and Mark Dickinson. “The Star Formation History of Field Galaxies”. In: *The Astrophysical Journal* 498 (May 1998). ADS Bibcode: 1998ApJ...498..106M, pp. 106–116. ISSN: 0004-637X. DOI: 10.1086/305523. URL: https://ui.adsabs.harvard.edu/abs/1998ApJ...498..106M (visited on 03/30/2023) (cit. on p. 12).

[21] Robert C. Kennicutt and Neal J. Evans. “Star Formation in the Milky Way and Nearby Galaxies”. en. In: *Annual Review of Astronomy and Astrophysics, vol. 50, p.531-608* 50 (Sept. 2012), p. 531. ISSN: 0066-4146. DOI: 10.1146/annurev-astro-081811-125610. URL: https://ui.adsabs.harvard.edu/abs/2012ARA%26A..50..531K/abstract (visited on 03/30/2023) (cit. on p. 14).

[22] C. R. Mulcahey et al. “Star Formation and AGN Feedback in the Local Universe: Combining LOFAR and MaNGA”. In: *Astronomy & Astrophysics* 665 (Sept. 2022). arXiv:2206.01195 [astro-ph], A144. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/202142215. URL: http://arxiv.org/abs/2206.01195 (visited on 03/30/2023) (cit. on p. 15).

[23] Francoise Combes. “AGN Feedback and Its Quenching Efficiency”. In: *Frontiers in Astronomy and Space Sciences* 4 (2017). ISSN: 2296-987X. URL: https://www.frontiersin.org/articles/10.3389/fspas.2017.00010 (visited on 03/31/2023) (cit. on p. 16).

[24] Arif Babul and Martin J. Rees. “On dwarf elliptical galaxies and the faint blue counts.” In: *Monthly Notices of the Royal Astronomical Society* 255 (Mar. 1992). ADS Bibcode: 1992MNRAS.255..346B, pp. 346–350. ISSN: 0035-8711. DOI: 10.1093/mnras/255.2.346. URL: https://ui.adsabs.harvard.edu/abs/1992MNRAS.255..346B (visited on 04/10/2023) (cit. on p. 17).

[25] Himansh Rathore et al. “Star-forming S0 Galaxies in SDSS-MaNGA: fading spirals or rejuvenated S0s?” In: *Monthly Notices of the Royal Astronomical Society* 513 (June 2022). ADS Bibcode: 2022MNRAS.513..389R, pp. 389–404. ISSN: 0035-8711. DOI: 10.1093/mnras/stac871. URL: https://ui.adsabs.harvard.edu/abs/2022MNRAS.513..389R (visited on 04/10/2023) (cit. on p. 17).

[26] Ben Moore et al. “Galaxy harassment and the evolution of clusters of galaxies”. In: *Nature* 379 (Feb. 1996). ADS Bibcode: 1996Natur.379..613M, pp. 613–616. ISSN: 0028-0836. DOI: 10.1038/379613a0. URL: https://ui.adsabs.harvard.edu/abs/1996Natur.379..613M (visited on 04/10/2023) (cit. on p. 17).

[27] Bitao Wang et al. “SDSS-IV MaNGA: The kinematic-morphology of galaxies on the mass versus star-formation relation in different environments”. In: *Monthly Notices of the Royal Astronomical Society* 495 (June 2020). ADS Bibcode: 2020MNRAS.495.1958W, pp. 1958–1977. ISSN: 0035-8711. DOI: 10.1093/mnras/staa1325. URL: https://ui.adsabs.harvard.edu/abs/2020MNRAS.495.1958W (visited on 10/08/2022) (cit. on p. 17).

[28] Samir Salim. “Green Valley Galaxies”. In: *Serbian Astronomical Journal* 189 (2014). arXiv:1501.01963 [astro-ph], pp. 1–14. ISSN: 1450-698X, 1820-9289. DOI: 10.2298/SAJ1489001S. URL: http://arxiv.org/abs/1501.01963 (visited on 03/31/2023) (cit. on pp. 17, 30).

[29] Kevin Bundy et al. “The Mass Assembly History of Field Galaxies: Detection of an Evolving Mass Limit for Star-Forming Galaxies”. en. In: *The Astrophysical Journal* 651.1 (Nov. 2006). Publisher: IOP Publishing, p. 120. ISSN: 0004-637X. DOI: 10.1086/507456. URL:

https://iopscience.iop.org/article/10.1086/507456/meta (visited on 04/10/2023) (cit. on p. 18).

[30] Robert Antonucci. "Unified models for active galactic nuclei and quasars." In: *Annual Review of Astronomy and Astrophysics* 31 (Jan. 1993). ADS Bibcode: 1993ARA&A..31..473A, pp. 473–521. ISSN: 0066-4146. DOI: 10.1146/annurev.aa.31.090193.002353. URL: https://ui.adsabs.harvard.edu/abs/1993ARA&A..31..473A (visited on 04/10/2023) (cit. on p. 18).

[31] Donald G. York et al. "The Sloan Digital Sky Survey: Technical Summary". In: *The Astronomical Journal* 120 (Sept. 2000). ADS Bibcode: 2000AJ....120.1579Y, pp. 1579–1587. ISSN: 0004-6256. DOI: 10.1086/301513. URL: https://ui.adsabs.harvard.edu/abs/2000AJ....120.1579Y (visited on 04/01/2023) (cit. on p. 23).

[32] Shadab Alam et al. "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample". In: *Monthly Notices of the Royal Astronomical Society* 470 (Sept. 2017). ADS Bibcode: 2017MNRAS.470.2617A, pp. 2617–2652. ISSN: 0035-8711. DOI: 10.1093/mnras/stx721. URL: https://ui.adsabs.harvard.edu/abs/2017MNRAS.470.2617A (visited on 04/01/2023) (cit. on p. 24).

[33] Andrés Almeida et al. *The Eighteenth Data Release of the Sloan Digital Sky Surveys: Targeting and First Spectra from SDSS-V*. arXiv:2301.07688 [astro-ph]. Jan. 2023. DOI: 10.48550/arXiv.2301.07688. URL: http://arxiv.org/abs/2301.07688 (visited on 04/01/2023) (cit. on p. 24).

[34] R. Bacon et al. "The SAURON project - I. The panoramic integral-field spectrograph". In: *Monthly Notices of the Royal Astronomical Society* 326 (Sept. 2001). ADS Bibcode: 2001MNRAS.326...23B, pp. 23–35. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.2001.04612.x. URL: https://ui.adsabs.harvard.edu/abs/2001MNRAS.326...23B (visited on 03/31/2023) (cit. on p. 27).

[35] S. F. Sánchez et al. "CALIFA, the Calar Alto Legacy Integral Field Area survey. I. Survey presentation". In: *Astronomy and Astrophysics* 538 (Feb. 2012). ADS Bibcode: 2012A&A...538A...8S, A8. ISSN: 0004-6361. DOI: 10.1051/0004-6361/201117353. URL: https://ui.adsabs.harvard.edu/abs/2012A&A...538A...8S (visited on 03/31/2023) (cit. on p. 27).

[36] Scott M. Croom et al. “The Sydney-AAO Multi-object Integral field spectrograph”. In: *Monthly Notices of the Royal Astronomical Society* 421 (Mar. 2012). ADS Bibcode: 2012MNRAS.421..872C, pp. 872–893. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2011.20365.x. URL: https://ui.adsabs.harvard.edu/abs/2012MNRAS.421..872C (visited on 03/31/2023) (cit. on p. 27).

[37] Kevin Bundy et al. “Overview of the SDSS-IV MaNGA Survey: Mapping nearby Galaxies at Apache Point Observatory”. In: *The Astrophysical Journal* 798 (Jan. 2015). ADS Bibcode: 2015ApJ...798....7B, p. 7. ISSN: 0004-637X. DOI: 10.1088/0004-637X/798/1/7. URL: https://ui.adsabs.harvard.edu/abs/2015ApJ...798....7B (visited on 03/31/2023) (cit. on p. 27).

[38] David A. Wake et al. “The SDSS-IV MaNGA Sample: Design, Optimization, and Usage Considerations”. In: *The Astronomical Journal* 154 (Sept. 2017). ADS Bibcode: 2017AJ....154...86W, p. 86. ISSN: 0004-6256. DOI: 10.3847/1538-3881/aa7ecc. URL: https://ui.adsabs.harvard.edu/abs/2017AJ....154...86W (visited on 03/31/2023) (cit. on p. 28).

[39] Brian Cherinka et al. “Marvin: A Tool Kit for Streamlined Access and Visualization of the SDSS-IV MaNGA Data Set”. In: *The Astronomical Journal* 158 (Aug. 2019). ADS Bibcode: 2019AJ....158...74C, p. 74. ISSN: 0004-6256. DOI: 10.3847/1538-3881/ab2634. URL: https://ui.adsabs.harvard.edu/abs/2019AJ....158...74C (visited on 03/31/2023) (cit. on p. 29).

[40] Jonathon Shlens. *A Tutorial on Principal Component Analysis*. arXiv:1404.1100 [cs, stat] version: 1. Apr. 2014. DOI: 10.48550/arXiv.1404.1100. URL: http://arxiv.org/abs/1404.1100 (visited on 04/10/2023) (cit. on p. 35).

[41] Zhongheng Zhang. “Introduction to machine learning: k-nearest neighbors”. en. In: *Annals of Translational Medicine* 4.11 (June 2016). Number: 11 Publisher: AME Publishing Company, pp. 218–218. ISSN: 2305-5847, 2305-5839. DOI: 10.21037/atm.2016.03.37. URL: https://atm.amegroups.com/article/view/10170 (visited on 04/10/2023) (cit. on p. 35).

[42] Gongde Guo et al. “KNN Model-Based Approach in Classification”. en. In: *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*. Ed. by Robert Meersman, Zahir Tari, and Douglas C. Schmidt. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2003, pp. 986–996. ISBN: 978-3-540-39964-3. DOI: 10.1007/978-3-540-39964-3_62 (cit. on p. 35).

[43] J. R. Quinlan. “Induction of decision trees”. en. In: *Machine Learning* 1.1 (Mar. 1986), pp. 81–106. ISSN: 1573-0565. DOI: 10.1007/BF00116251. URL: https://doi.org/10.1007/BF00116251 (visited on 04/10/2023) (cit. on p. 38).

[44] Tianqi Chen and Carlos Guestrin. “XGBoost: A Scalable Tree Boosting System”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. arXiv:1603.02754 [cs]. Aug. 2016, pp. 785–794. DOI: 10.1145/2939672.2939785. URL: http://arxiv.org/abs/1603.02754 (visited on 04/10/2023) (cit. on p. 38).

[45] Kaiming He et al. *Deep Residual Learning for Image Recognition*. arXiv:1512.03385 [cs]. Dec. 2015. DOI: 10.48550/arXiv.1512.03385. URL: http://arxiv.org/abs/1512.03385 (visited on 04/10/2023) (cit. on p. 39).

[46] Connor Shorten and Taghi M. Khoshgoftaar. “A survey on Image Data Augmentation for Deep Learning”. In: *Journal of Big Data* 6.1 (July 2019), p. 60. ISSN: 2196-1115. DOI: 10.1186/s40537-019-0197-0. URL: https://doi.org/10.1186/s40537-019-0197-0 (visited on 04/10/2023) (cit. on p. 40).

[47] Luis Perez and Jason Wang. *The Effectiveness of Data Augmentation in Image Classification using Deep Learning*. arXiv:1712.04621 [cs]. Dec. 2017. DOI: 10.48550/arXiv.1712.04621. URL: http://arxiv.org/abs/1712.04621 (visited on 04/10/2023) (cit. on p. 40).

[48] Hoo-Chang Shin et al. *Medical Image Synthesis for Data Augmentation and Anonymization using Generative Adversarial Networks*. arXiv:1807.10225 [cs, stat]. Sept. 2018. DOI: 10.48550/arXiv.1807.10225. URL: http://arxiv.org/abs/1807.10225 (visited on 04/10/2023) (cit. on p. 41).

[49] Fuzhen Zhuang et al. *A Comprehensive Survey on Transfer Learning*. arXiv:1911.02685 [cs, stat]. June 2020. DOI: 10.48550/arXiv.1911.02685. URL: http://arxiv.org/abs/1911.02685 (visited on 04/10/2023) (cit. on p. 41).

[50] Karl Weiss, Taghi M. Khoshgoftaar, and DingDing Wang. “A survey of transfer learning”. In: *Journal of Big Data* 3.1 (May 2016), p. 9. ISSN: 2196-1115. DOI: 10.1186/s40537-016-0043-6. URL: https://doi.org/10.1186/s40537-016-0043-6 (visited on 04/10/2023) (cit. on p. 41).

[51] Yaqing Wang et al. *Generalizing from a Few Examples: A Survey on Few-Shot Learning*. arXiv:1904.05046 [cs]. Mar. 2020. DOI: 10.48550/arXiv.1904.05046. URL: http://arxiv.org/abs/1904.05046 (visited on 04/10/2023) (cit. on p. 42).

[52] Archit Parnami and Minwoo Lee. *Learning from Few Examples: A Summary of Approaches to Few-Shot Learning*. arXiv:2203.04291 [cs]. Mar. 2022. DOI: 10.48550/arXiv.2203.04291. URL: http://arxiv.org/abs/2203.04291 (visited on 04/10/2023) (cit. on p. 42).

[53] Jake Snell, Kevin Swersky, and Richard S. Zemel. *Prototypical Networks for Few-shot Learning*. arXiv:1703.05175 [cs, stat]. June 2017. DOI: 10.48550/arXiv.1703.05175. URL: http://arxiv.org/abs/1703.05175 (visited on 04/10/2023) (cit. on p. 43).

[54] Ting Chen et al. *A Simple Framework for Contrastive Learning of Visual Representations*. 2020. arXiv: 2002.05709 [cs.LG] (cit. on p. 58).