



Faculty of Sciences

Department of Computer Science and Information Systems

A MULTI LAYER PERCEPTRON CLASSIFIER FOR RADIO EMISSION SOURCES ASSOCIATED WITH GALAXY CLUSTERS

by

KELETSO BONTLE DICHABA

Student ID number: 12000587

BSc (Applied Sciences) (BIUST)

A Dissertation/Thesis Submitted to the Faculty of Sciences in Partial Fulfilment of the
Requirements for the Award of the Degree of Master of Science in Computer Science of BIUST

Supervisor(s): Dr., Hlomani Hlomani

Department of Computer Science and Information Systems

Faculty of Science, BIUST

hlomanihb@biust.ac.bw

Supervisor(s): Professor., James Okwe Chibueze

Centre for Space Research

North-West University (NWU)

James.Chibueze@nwu.ca.za

January, 2023

DECLARATION REGARDING THE WORK AND COPYRIGHT

Candidate (please write in caps or type) _____ KELETSO BONTLE DICHABA _____

Student ID: _____ 12000587 _____

Thesis Titled: _____ A MULTI LAYER PERCEPTRON CLASSIFIER FOR RADIO EMISSION SOURCES
ASSOCIATED WITH GALAXY CLUSTERS _____

I, the **Candidate**, certify that the Thesis is all my own original work and that I have not obtained a degree in this University or elsewhere on the basis of any of this work.

(If the thesis is based on a group project, then the student must indicate the extent of her / his contribution, with reference to any other theses submitted or published by each collaborator in the project, and a declaration to this effect must be included in the thesis)

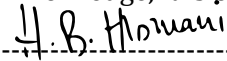
This dissertation/thesis is copyright material protected under the Berne Convention, the Copyright and Neighbouring Rights Act, Act. No. 8 of 2000 and other international and national enactments, in that behalf, on intellectual property. It must not be reproduced by any means, in full or in part, except for short extracts in fair dealing; for researcher private study, critical scholarly review or discourse with an acknowledgement, without the written permission of the office of the Postgraduate School, on behalf of both the author and the BIUST.



Signed: _____ Date: ----05/01/2023-----

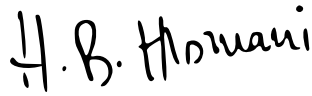
Primary Supervisor (please write in caps or type) _____ DR. HLOMANI HLOMANI _____

I, the Candidate's **Primary Supervisor**, hereby confirm that I have inspected the above titled thesis and, to the best of my knowledge, it is based on the original work of the candidate.

Signed: _____  _____ Date: -----10/01/2023-----

Certification

The undersigned certifies that he/she has read and hereby recommends for acceptance by the faculty of Science a thesis titled: **A MULTI LAYER PERCEPTRON CLASSIFIER FOR RADIO EMISSION SOURCES ASSOCIATED WITH RADIO GALAXY CLUSTERS** , in fulfilment of the requirements for the degree of Master of Science in Computer Science of the BIUST.



.....
Dr. Hlomani Hlomani

(Supervisor)

Date: 10/01/2023
.....



.....
Prof. James O. Chibueze

(Co-Supervisor)

Date: 10/01/2023
.....

Acknowledgements

Acknowledgement is given to my supervisors Dr. Hlomani and Prof. Chibueze. Thank your for the opportunity to explore a passion and your tireless patience. Your assistance has been phenomenal. I will forever be grateful for your guidance and support. Thank you.

Acknowledgement is also given to the Botswana International University of Science and Technology (BIUST), the North West University (NWU) and the South African MeerKAT radio telescope without which this research would not be possible. Many thanks.

A special thank you to my close friends, family, colleagues who all became mentors and motivators at one point or another and made this journey worth it. Thank you for the support.

Dedication

To little Kai and my family, a small cluster that is nature's biggest masterpiece and my early heaven.

Contents

List of Tables	4
List of Figures	5
1 Introduction	10
1.0.1 Introduction to the problem	10
1.0.2 Research background	15
1.0.3 Problem statement	20
1.0.4 Main objective	21
1.0.5 Specific objectives	21
1.0.6 Questions	21
1.0.7 Hypothesis	22
1.0.8 Significance of the study	22
1.0.9 Limitations of the study	23
1.0.10 Chapter summary	23
1.0.11 Thesis overview	24
2 Literature review	25
2.0.1 Introduction	25
2.0.2 Significance of galaxy clusters from the bigger picture	25
2.0.3 Types of radio emissions	27
2.0.4 Radio emission classification	29
2.0.5 Radio emission source classification approaches	30
2.0.6 Source finding software	30
2.0.7 Machine learning	33
2.0.8 Multi Layer Perceptron	37
2.0.9 Multi Layer Perceptron applications in astronomy	39

2.0.10	Back-propagation	41
2.0.11	Mathematical representation of back-propagation	42
2.0.12	General applications of back-propagation	42
2.0.13	Significance of using a Multi Layer Perceptron	43
2.0.14	Chapter summary	44
3	Methodology	45
3.0.1	Scikit-learn	45
3.0.2	Multi Layer Perceptron architecture	46
3.0.3	The experimental framework	47
3.0.4	MeerKAT observations	48
3.0.5	Data collection	49
3.0.6	Data preparation	55
3.0.7	PyBDSF	56
3.0.8	Experimental Setup	61
3.0.9	Performance metrics	62
3.0.10	Confusion matrix	63
3.0.11	Precision	65
3.0.12	Recall	65
3.0.13	Accuracy	65
3.0.14	F1_score	66
3.0.15	Chapter Summary	67
4	Results and discussion	68
4.0.1	Introduction	68
4.0.2	The Multi Layer Perceptron training results	68
4.0.3	The Multi Layer Perceptron testing results	71
4.0.4	Confusion matrix results 1	71
4.0.5	Accuracy results 1	77
4.0.6	Precision results 1	78
4.0.7	Recall results 1	79
4.0.8	F1_score results 1	80
4.0.9	Confusion matrix results 2	81

4.0.10	Accuracy results 2	83
4.0.11	Precision results 2	84
4.0.12	Recall results 2	84
4.0.13	F1_score results 2	85
4.0.14	Discussion	85
4.0.15	Confusion matrix	85
4.0.16	Model accuracy	86
4.0.17	Precision	87
4.0.18	Recall	89
4.0.19	F1_score	90
4.0.20	Number of sources in a field on the model performance	91
4.0.21	Chapter summary	92
5	Conclusion and recommendation	93
5.0.1	Introduction	93
5.0.2	Thesis overview	93
5.0.3	Objectives overview	94
5.0.4	Hypothesis overview	95
5.0.5	Questions overview	96
5.0.6	Conclusion	97
5.0.7	Recommendations	99
	Bibliography	100

List of Tables

3.1	The hyperparameters for the Multi Layer Perceptron model implemented.	46
3.2	Description of the cluster sample for this study.	52
3.3	A list of the data feature used for the training of the MLP.	54
3.4	The frequency sub-bands per image with corresponding source population for the three classification labels ‘S’, ‘C’ and ‘M’.	60
4.1	The accuracy results for the 21 galaxy clusters used for testing the Multi Layer Perceptron model and the total number of sources for each galaxy cluster.	78
4.2	The precision results for the 21 galaxy clusters used for testing the Multi Layer Perceptron model.	79
4.3	The recall results for the 21 galaxy clusters used for testing the Multi Layer Perceptron model.	80
4.4	The F1_score results for the 21 galaxy clusters used for testing the model.	81
4.5	Accuracy results of the Multi Layer Perceptron in classifying sources in varied frequency sub-bands and the number of sources used for testing.	83
4.6	Precision results for the Multi Layer Perceptron in classifying sources in varied frequency sub-bands and the number of sources used for testing.	84
4.7	Recall results for the Multi Layer Perceptron in classifying sources in the different frequency sub-bands and the number of sources used for testing.	84
4.8	F1_score results for the Multi Layer Perceptron in classifying sources in the different frequency sub-bands and the number of sources used for testing.	85

List of Figures

1.1	The electromagnetic spectrum characterised by wavelength. https://www.miniphysics.com/electromagnetic-spectrum_25.html	11
1.2	The W50 nebula shown in optical wavelength (a) vs in optical and radio (b). <i>Credit : NRAO/AUI/NSF, K. Golap, M. Goss; NASA's Wide Field Survey Explorer (WISE)</i>	12
2.1	A cosmic web simulation by Springel <i>et al</i> [66]	26
2.2	Two radio galaxy clusters with various extended emission sources located at different areas of the cluster.	30
2.3	An MLP with one hidden layer.	38
2.4	Back-propagation from a Multi Layer perceptron	41
3.1	The Multi Layer Perceptron architecture implemented.	47
3.2	The experimental framework	47
3.3	Multi-Frequency Synthesis (MFS) image of A3376	55
3.4	Flow chart for the source detection process using PyBDSF	57
3.5	Results of the interactive mode where outputs from each computation stage are produced.	58
3.6	Image hierarchy used by PyBDSF when processing an image. It shows how Gaussian grouping occurs and the class labels assigned to each confirmed radio emission source are deduced.	60
3.7	Source population from the eight data sets plotted against the three class labels denoted by 'S', 'C' and 'M'	61
3.8	Flow chart for the experimental framework used to evaluate a Multi Layer Perceptron for prediction radio emission sources.	62

3.9	A $N \times N$ confusion matrix developed for the three class labels namely 'S', 'C' and 'M' where the classification results can be calculated for each class.	64
4.1	The MLP loss curves for different number of hidden layers.	69
4.2	The MLP validation curves for different number of hidden layers. . . .	70
4.3	Confusion matrix for MACS_J0257.6-2209 and MACSJ_0025.4-1222B	72
4.4	Confusion matrix for RXCJ0225.1-22928 and PLCK_G200.9-28.2	72
4.5	Confusion matrix for MACS_J0257.6-2209B and J2340.1-8510	73
4.6	Confusion matrix for J0820.9-5704 and J1410.4-4246	73
4.7	Confusion matrix for J1240.2-4825 and J0543.4-4430	74
4.8	Confusion matrix for J0626.3-5341 and J0528.9-3927	74
4.9	Confusion matrix for J0449.9-4440 and J0108.5-4020	75
4.10	Confusion matrix for J0314.3-4525 and Abell_2895	75
4.11	Confusion matrix for J0216.3-4816 and Abell_4038	76
4.12	Confusion matrix for Abell_548 and Abell_13	76
4.13	Confusion matrix for galaxy cluster Abell_2744	77
4.14	Confusion matrix for Sub-band 1 and Sub-band 2.	81
4.15	Confusion matrix for Sub-band 3 and Sub-band 4.	82
4.16	Confusion matrix for Sub-band 5 and Sub-band 6.	82
4.17	Confusion matrix for Sub-band 7 and Sub-band 8.	83
4.18	A graph showing the effect of the number of sources in a field on the model performance using the four metrics	91

Abbreviations

- VLA - The Very Large Array
- Jy - Jansky, a unit of spectral flux density
- Arcsec - Arcsecond, an angular measurement
- SKA - The Square Kilometre Array
- ASKAP - The Australian SKA Pathfinder
- MeerKAT - The Meer Karoo Array Telescope
- MWA - Murchison Widefield Array
- A3376 - Abell 3376
- PyBDSF - Python Blob Detection and Source Finder
- ICM - Intra-cluster medium
- MFS - Multi-Frequency Synthesis
- Hz - Hertz, a unit of frequency

Measurements

- Parsec - Parallax of one arc-second which is a distance measurement used in astronomy.
- 1 Mpc - 3.26 million light years
- $1 M_{\odot} - 2 \times 10^{30} \text{ kg}$
- 1 MHz - $1 \times 10^6 \text{ Hz}$
- 1 GHz - $1 \times 10^9 \text{ Hz}$

Abstract

The Square Kilometre Array (SKA) is an emerging next-generation radio telescope project estimated to be completed a decade from now. Once completed it will make it easier and faster to survey space in greater detail. This will produce petabytes of data which will be higher by a factor of one hundred when compared to existing radio telescopes presenting an array of scientific and technical challenges. To address the challenges individual and combined solutions are required by utilizing it's precursors. Processing data in real time has since become an integral part of future data processing pipelines that will be designed and implemented with the aim of identifying and selecting important data for further analysis from the bulk of the incoming data while discarding what is not needed or considered unnecessary. It also makes the traditional manual handling of data an unfeasible approach. To this end this study explores the automatic classification of radio emission sources associated with galaxy clusters through the use of a Multi Layer Perceptron (MLP) as a way to eliminate the manual inspection that is often required to confirm individual sources and determine its feasibility as a preliminary eliminator of unnecessary data. This was achieved by implementing an MLP to train on over 400 000 radio emission sources and then predicting the class label on over 100 000 sources. The three class labels were 'S' which denoted a single Gaussian source in a given island, 'C' which denoted a single Gaussian source in an island with another source and 'M' which denoted a multi Gaussian radio source in a given island. The MLP achieved an average accuracy of 0.89, a precision of 0.82, a recall of 0.81 and a F1_score of 0.80. The results proved that an MLP serves as a good classifier for radio emission sources associated with radio galaxies.

Chapter 1

Introduction

1.0.1 Introduction to the problem

The electromagnetic spectrum describes the characteristics and distribution of electromagnetic radiation that either emanates or is absorbed by a given body [1] and form waves as they travel through a medium. The wavelength is described as the distance between two successive wave peaks and can be used to classify the different types of radiation into gamma rays, X-rays, ultraviolet, visible light, infrared, microwaves and radio waves.

As shown in Figure 1.1 gamma rays have the shortest wavelength in the electromagnetic spectrum followed by x-rays, ultraviolet, visible light, infrared, microwave, with radio waves having the longest wavelength. The wavelength is inversely proportional to the energy of a wave and so longer wavelengths have lower energy and the energy increases as you move towards shorter wavelengths in the electromagnetic spectrum as shown in Equation 1.1 below

$$E = hv = \frac{h}{\lambda} \quad (1.1)$$

Where E is the energy, h is the Planck's constant, v is the frequency and λ is the wavelength. The frequency however which is the number of waves passing through a fixed point at a given point in time is directly proportional to the wavelength, consequently shorter wavelengths have higher frequencies with gamma rays having a higher frequency as compared to other radiation emissions. In the field of astronomy different forms of radiation require specialised telescopes to study space at different

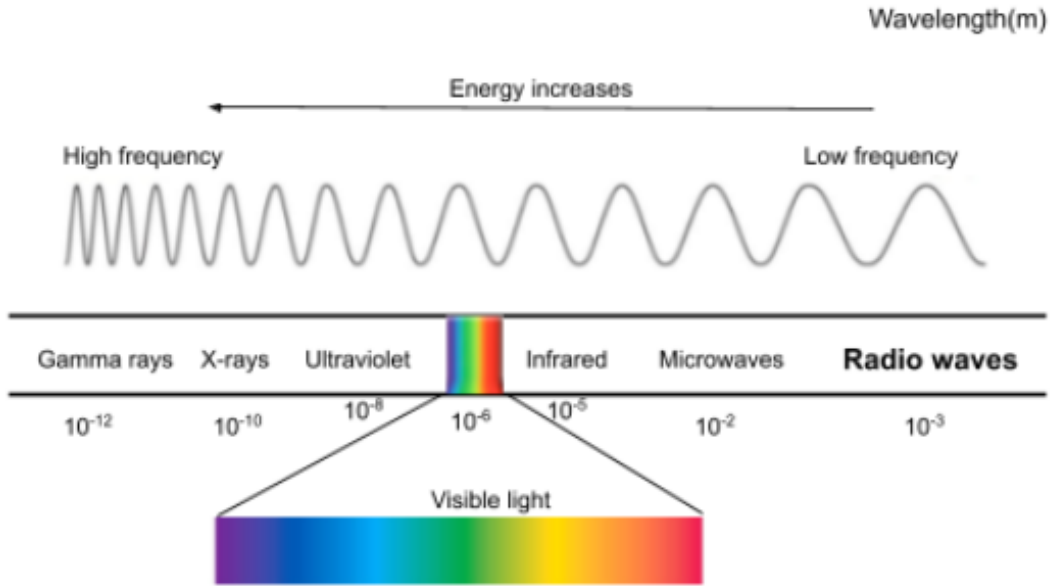
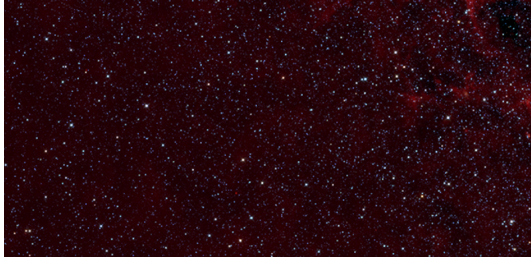


Figure 1.1: The electromagnetic spectrum characterised by wavelength.
https://www.miniphysics.com/electromagnetic-spectrum_25.html

wavelengths since various radiation emitting sources provide insight regarding the formation and evolution of our universe.

To demonstrate the significance of studying various radiation emitting sources at different wavelengths found in the universe, Figure 1.2 shows an image of space observed in the optical wavelength (a). The second image (b) is of the same spot in space observed in the optical and radio wavelengths in which the W50 "Manatee" nebula, which is the remains of a star explosion shown in green is clearly visible. This source would otherwise be invisible in the absence of the radio wavelength observations. The study of previous optical observations in the radio wavelength often reveals insights on the processes that occur within those celestial bodies and has led to the discovery of new objects such as pulsars and quasars.

Radio astronomy, as a sub-field of astronomy studies objects from our universe that emit significant amounts of radio waves as compared to other forms of radiation which range from as low as 13.36 - 13.41 MHz to as high as exceeding 1000 GHz [2]. These sources include planets, stars and galaxies. The radio emission from these sources can be either thermal emission which depends on the emitting source's temperature or



(a) The W50 nebula is imperceptible against an optical background



(b) The visible W50 nebula observed in the radio wavelength

Figure 1.2: The W50 nebula shown in optical wavelength (a) vs in optical and radio (b). *Credit : NRAO/AUI/NSF, K. Golap, M. Goss; NASA's Wide Field Survey Explorer (WISE).*

non-thermal emission which is independent of an object's temperature and include synchrotron radiation, Compton Scattering and maser emission. Radio telescopes offer radio astronomers the opportunity to study and record data about radio emitting objects in detail.

Around the 1980s, the Very Large Array was considered the state of the art radio telescope with a sensitivity reaching 100 mJy and a resolution up to tenths of arcsec [3] which surpassed any existing radio telescope at the time. However in the 1990s at the International Astronomical Union meeting that centered around radio interferometry [4], the idea of designing and implementing a more superior radio telescope began to materialise to what we now call the Square Kilometre Array (SKA) project that is to be completed by late 2030 [5].

A square kilometre equates to a million square metres of data collecting area for the SKA which is the largest thus far when compared to radio telescopes currently in operation. Once the Square Kilometre Array is up and running, which is approximated to be a decade from now it will produce petabytes of data which will be higher by a factor of one hundred when compared to existing radio telescopes [6]. Due to the unprecedented sensitivity and resolution from the SKA, it will make it easier and faster to survey space in greater detail, exceeding the Hubble Space Telescope's image resolution which is currently considered very powerful. The SKA has precursor telescopes which are prototypes that have been paving the way to study the scientific and technological requirements of the SKA and include the Australian SKA Pathfinder

(ASKAP) [7], The Meer Karoo Array Telescope (MeerKAT) [8] and the Murchison Widefield Array (MWA) [9]. This study focuses on the MeerKAT precursor and the data it produces.

One of the initial primary objectives of the SKA was to study and measure neutral hydrogen(HI) present in space and while the science goals have evolved over time measuring HI is still maintained as a key science goal together with understanding the occurrence of black holes, dark energy and the epoch of re-ionization. Other scientific goals for the SKA include the study of the formation of stars and their systems, planets, pulsars, galaxies and exploring the existence of extraterrestrial life. With time it was discovered that HI could be used as a tracer for large scale astronomical structures [4]. One of these being clusters of galaxies which may belong to a group, a cluster or a supercluster.

Besides galaxy clusters there exist galaxy groups. An example of a galaxy group would be the local group in which the Milky Way galaxy belongs to. Galaxy groups are relatively small compared to clusters with the local group consisting of 24 galaxies within a radius of 1 Mpc [10]. About three quarters of known galaxies belong to a cluster.

Galaxy clusters which are the subject for this study, are the largest gravitationally bound and most stable structures found in the cosmic web with majority of the gravitational force resulting from dark matter. They are formed from a hierarchical accretion of smaller galaxies to create gigantic clusters that grow in hubble time. Hubble time being the time it took for the universe to grow to it's current size [11] which is considered a very slow process. A cluster comprises of dark and luminous matter which include stars and can have a few hundred to tens of thousands of galaxies within their structure and usually reach 5 Mpc or more in size with temperatures reaching approximately 10^8 K.

Another type is called superclusters which are significantly larger compared to galaxy groups and galaxy clusters. They are formed from the merging of individual galaxy clusters creating a gigantic cosmic structure that spans more than 7 Mpc. An estimated 85% of a galaxy cluster mass is dark matter with galaxies making up a very small portion of the overall mass which is about (3%). The baryonic proportion of

the mass which makes up the rest of the cluster mass is made up of hot plasma, the intracluster medium (ICM) which is found between sub clusters and makes up 12% of the total mass [12].

Galaxy clusters are interesting for a number of reasons besides their magnitude. One of the most interesting scientific phenomena to study is cosmic magnetism. Until a couple of years ago cosmic magnetism was like any other scientific process until studies showed that it had a larger contribution to how the universe is structured and it has now become one of the scientific goals for the SKA [5]. Cosmic magnetism contributes significantly to star formation, stellar activity, pulsars, formation of cosmic rays and accretion disks. Studies such as [13], [14], [15],[16], [17] are now focused on understanding how cosmic magnetism has affected the early formation of the universe with the assertion that magnetic field lines during the earlier times support the big bang theory's nucleosynthesis limitations, the interstellar medium, halos and galaxy evolution. Galaxy clusters therefore provide the optimum environment for studying these astrophysical processes because of their large and dynamic atmosphere. Besides galaxy clusters, cosmic magnetism can be found in voids of the cosmic web and filaments and has been proven to be distributed according to the structure of the universe instead of the previous assumption that it was evenly dispersed.

There are several challenges concerning the handling of astronomical data from radio telescopes such as from the MeerKAT and the Evolutionary Map of the Universe (EMU). The high sensitivity and resolution of such radio telescopes imply a faster survey of the sky resulting in significantly large volumes of data being produced. It is estimated that EMU will detect and classify around 70 million galaxies [18] while the SKA will detect 62000 sources per square degree up to 1 mJy [19] and with the MeerKAT having a Science Data Processor (SDP) input rate of 4 terabits [20]. This makes the traditional manual handling of data an unfeasible approach as suggested by [21]. This also poses challenges in the fundamental data processing steps such as data preparation, calibration and it's imaging.

There exists a number of studies aimed at addressing the aforementioned challenges with regards to radio emitting sources. These include studies aimed at automated

source detection and classification. Among the existing solutions there is the application of artificial intelligence in the field of radio astronomy in particular the use of machine learning algorithms for the automatic detection and classification of radio sources found in galaxy clusters as host environments. This approach has been dominant when compared to others. Examples of these include [21], [22] and [23].

This study explores the automatic classification of radio emission sources from galaxy clusters through the use of machine learning as a way to eliminate the manual inspection that is often required to confirm individual sources. Particularly exploring the application of a Multi Layer Perceptron as an automatic radio emission source classifier. It's application as an eliminator of unnecessary data in real time and the paucity of studies in the field of radio astronomy as a classifier will be demonstrated subsequently which necessitated this study.

1.0.2 Research background

Artificial intelligence is a broad term that refers to research that focuses on developing ideas, theories, methods and technological systems that can be used to expand the intelligence of humans through the utilization of machines that comprehend problem solving, reasoning and language [24]. Machine learning exists as an artificial intelligence domain with research that enables machines to acquire knowledge by learning from experience without being given precise instructions on how to do so through programming [25], [26]. Since it's inception, machine learning has made advancements in a number of algorithms developed through complex statistics and mathematical computation among others.

Within machine learning sub-fields, there exists supervised machine learning with a number of algorithms and applications essential to establishing the fundamental relationship between labelled data features and target class for future predictions of unlabelled data. Supervised learning algorithms are trained on labelled data to predict the class outcome when given data the model has not seen before. It associates labels with the proper training sample provided and is often referred to as inductive models since they generate models based on the mapping learned between the input

and output and then being able to predict the output given new input data [27], [28]. Models well trained on labelled data ensure that predicted data belong to pre-defined classes that have been confirmed. This becomes an easy task to classify unlabelled data when assigning class labels. The next generation radio surveys such as the future SKA surveys will generate vast amounts of data with many unknown and undiscovered radio sources. Highly accurate predictor models give an almost guarantee that classification of known radio sources and a significant portion of previously undiscovered sources will be assigned a class label all done in a significantly shorter time frame. Given the assumption that even though a source is newly discovered it holds a level of correlation with previously discovered sources therefore that increases the chances of it being identified as a valid source and assigned a class label, an assumption this study aims to verify.

A Multi Layer Perceptron is an example of a feed-forward neural network which is a function approximator and typically contains several layers of neurons which are biologically inspired. Each neuron found in one layer of the network is connected to each neuron found in the previous layer. The value of each neuron connection varies as the connection weights are adjusted during training. The first layer is the input layer which comprises of the network inputs. This is proceeded by parallel hidden neurons that compute a weighted sum of inputs to reduce the variance between the output of the network and the true value of the output referred to as the neuron function that is transferred between layers with no feedback hence the feed forward neural network [29], [30]. Parameters for the network are established using back propagation [31], [32]. The concepts of feed forward networks and back propagation have significant contribution in classifications problems in the field of astronomy being that they can handle high dimensional and complex data. They also demonstrate the significant potential of using machine learning to automate manual processes in radio astronomy and other fields, eliminating the strenuous process of individual data inspection. Since they provide promising results, it is also important to study their mechanism when handling data to ensure that the correct models are assigned to the correct task/problem.

A Multi Layer Perceptron was successful in illustrating its efficiency as a class predictor by identifying sources as either stars or galaxies as part of the procedure developed by Andreon *et al* [33]. Other studies include Odewahn *et al* [34] which demonstrated that a Multi Layer Perceptron performed just as well as existing software packages used by astronomers for the classification of galaxies and stars. Back propagation ensures learning of complex non linearly separable classes as compared to linearly separable classes where convergence during training is guaranteed and a solution vector formed. Cases of complex non linearly separable class problems encompass [35], [36], [37]. Studies that explore the use of machine learning in particular feed forward neural networks in the field of astronomy include works by Andreon *et al* [38] that successfully employed a feed forward neural network for the detection and classification of resolved and unresolved sources against a noisy background in wide field imaging. The success of the study was succeeded by another study by the same main author Andreon *et al* [33] where a detection, deblending of overlapping sources and classification procedure of sources in wide field images was developed which demonstrated the application of three neural networks from detection to the classification of sources.

In view of the fact that the SKA is on a revolutionary path of its own, it also comes with its own set of scientific and technical challenges. These include how to handle massive amounts of incoming data transmitted in real time that will be produced once it is in full operation, processing techniques compatible with this next generation radio telescope, data transportation and storage, analysis and interpretation together with the computing power necessary to perform the previously mentioned tasks. For the aforementioned challenges, individual and combined solutions are required to address the big data obstacles presented by the SKA by utilizing its precursors. Processing data in real time has since become an integral part of future data processing pipelines that will be designed and implemented with the aim of identifying and selecting important data for further analysis from the bulk of the incoming data while discarding what is not needed or considered unnecessary.

A number of studies [21], [22], [39], [23], [40] have addressed automated source

detection and classification of galaxy clusters and their sources in order to completely remove the unfeasible aspect of requiring human input for identifying or confirming observed sources using machine learning algorithms. Studies such as Vafaei *et al* [41] have focused on the detection of faint sources which are often missed because of their reduced appearance due to correlated noise among other factors. Other studies including Feder *et al* [42] have addressed source detection in crowded fields by use of cataloguing techniques and Bonaldi *et al* [43] which summarises results from the SKA data challenge in handling detection and classification in crowded fields. Algorithms employed by previous studies include Convolutional Neural Networks(CNNs) and Faster Recurrent Convolutional Neural Networks(R-CNNs) with a focus on using radio images. Images are often very large, require tailored software for visualization and analysis and therefore require high computing power. This is a disadvantage because of the amount of resources being used for studying radio images. Additionally real time analysis becomes challenging for cases where images are used for radio source detection and classification. This study aims to demonstrate that the use of an alternative approach to detection and classification of galaxy cluster radio emission sources reduces the need for computational, storage and analysis resources by the use of numerical data.

Published studies have addressed the application of supervised and unsupervised models for detection and classification of radio emitting sources in particular those associated with galaxy clusters to handle the potential massive data from the next generation radio telescopes. Such studies include [22], [44], [45], [46] as mentioned earlier. The models have produced results comparable to existing source finding tools while demonstrating that the detection and classification of radio sources can be performed without human input. Supervised learning models have potential as classifiers of radio sources with several advantages. They find the relationship between the given input and expected output which once established can be used to predict the most likely output for any given unlabelled data [47]. Prediction models are advantageous since they can significantly reduce the computation expected from unsupervised models in determining what is likely a source and which category the source belongs to in huge data sets and in addition data labels. In the case of radio

astronomy they can serve as preliminary estimators of a potential source and the class they most likely belong to. With regards to real time analysis pipelines a supervised model can be an eliminating tool of what is considered unnecessary data based on the odds of it being a possible source. They are also considered the most simple predictive models when compared to time series, outliers, and clustering models. For this reason supervised machine learning algorithms warrant further exploration as possible radio emission classifiers.

Supervised classification tasks are also very popular in the medical field. It deals with enormous volumes of data which is often high dimensional. For instance the field of omics which involves prediction and pattern recognition related to genes and the Deoxyribonucleic acid (DNA). This field is of vital importance for diagnosis and treatment of diseases such as cancer diagnosis [48], [49], [50]. A Multi Layer Perceptron as a form of a deep supervised learning architecture has also been employed in the medical field to solve an array of detection and diagnostic problems. These include [51], [52], [53], [54], [55]. The studies focused on the diagnosis of thyroid disease, detection of brain tumors by using magnetic resonance imaging and the diagnosis of diabetes during it's onset stages. The latter indicating the capability of a Multi Layer Perceptron as a predictor model by surpassing the convolutional neural network (CNN) and recurrent neural network in accuracy when it came to early diagnosis of diabetes. The aforementioned studies support the versatility of supervised machine learning. In particular the application of Multi Layer Perceptron neural networks which is applied in various scientific fields all holding qualitatively different data. It thus necessitates it's performance on data produced by next generation telescopes as a predictor model. Something that has not been employed previously.

Catalogs in the field of astronomy summarizes information from observations and such information is used for studying astrophysical processes. A recent cataloguing technique is by Feder *et al* [42] which proposed a probabilistic cataloger using multi-band images in crowded fields. This approach outperformed the more conventional techniques in the recovery of point sources while maintaining a significantly low false detection rate. It was also much faster compared to Portillo *et al* [56] from which the

basis of the research was made. While this study is noteworthy it does not address the presence of extended sources which mainly comprise of galaxies. It is also not known how the probabilistic cataloguer performs in radio data. Other notable probabilistic cataloguing approaches include X-ray observations [57] and gamma ray [58] with the first use of probabilistic cataloguing of point sources in overpopulated fields being in 2013 by Brewer *et al* [59]. With catalogs, astronomers can study the populations and distribution of sources, their morphology, and the underlying astrophysical processes in order to understand our universe [60], [61], [62]. However catalogers still face challenges in terms of the identification and classification of extended radio sources and their variety. A lot of extended sources still require human input for verification since their populations are small, but what happens when large surveys such as the SKA are in operation and extended source population increase? Human input then becomes unfeasible and the application of automated classification techniques such as supervised machine learning algorithms become advantageous since computationally they are less complex, the model learns faster in comparison to unsupervised learning algorithms. Additionally there exist deep learning algorithms suitable for classification tasks that are easy to implement such as a Multi Layer Perceptron.

1.0.3 Problem statement

The Square Kilometre Array will produce petabytes of raw data once in full operation which presents a challenge since the storage demand from the SKA cannot be met with the current technology. To address this the bulk of the incoming data will have to be processed in real time in order to eliminate what is not required and keep the necessary data for further analysis. In addition the manual inspection of sources is unfeasible given the large amount of data expected. A Multi Layer perceptron presents an important opportunity to address the data storage demand, source detection and classification challenges mentioned as an automated source classifier and a preliminary data eliminator. A Multi Layer Perceptron neural network is great in pattern recognition and classification which can be applied in predicting classes for any large number of radio sources by training the model on radio emission sources to predict classes on unseen data. Since the supervised model maps the relationship between the input data and the output the mapping can then be used to identify true sources

in real time, discarding unnecessary data. The paucity in the application of a Multi Layer Perceptron as a radio emission source classifier also presents an opportunity to study how it performs in a classification task. This is because the majority of classifiers already employed are CNN based which are known to be computationally exhaustive.

1.0.4 Main objective

The main objective of this research is to design and implement a Multi Layer Perceptron classifier for galaxy cluster radio emission sources and establish factors that affect the performance of the network.

1.0.5 Specific objectives

Given the above main objective, the specific objectives for this research are to:

1. Collect data to be used in the Multi Layer Perceptron model.
2. Prepare the data for training and testing of the Multi Layer Perceptron.
3. Design and implement a Multi Layer Perceptron algorithm for the classification of radio emission sources associated with radio galaxy clusters.
4. Train and test the model using the data obtained.
5. Evaluate the model performance on classification while determining factors affecting its performance through the tuning process.

1.0.6 Questions

Below are questions formulated for this research:

1. How does a multi layer perceptron perform at a classification task of galaxy cluster radio emission sources?
2. What hyperparameters can be modified in the network which affect the performance of the model?
4. Does the frequency sub-band used in different data sets have any effect on the overall performance of the classifier?

5. Is a Multi Layer Perceptron suitable for galaxy cluster radio emission sources classification?
6. Can the Multi Layer Perceptron model be implemented as a predictor model for real life detection and classification of galaxy cluster sources?

1.0.7 Hypothesis

The hypothesis for this study was formulated based on the performance of previous studies that applied machine learning to the detection and classification of radio emitting sources where performance was compared to state of the art source finding tools. Based on the results of previous studies such as Wu *et al* [46], Gheller *et al* [63], accuracy above 80% was considered as a model that performed well in a detection and classification task. For this research there exists a hypothesis and a null hypothesis.

Hypothesis (H_1) - A multi layer perceptron neural network performs significantly well in the task of supervised classification of radio emission sources associated with galaxy and can be a good predictor model obtaining accuracy greater than 80%.

Null hypothesis (H_0) - Astronomical data is significantly different from other data types that have been largely used to assess the performance of a multi layer perceptron. Thus does not perform well in the task of classifying radio sources. Consequently it will not be a good predictor model obtaining accuracy less than 80%.

1.0.8 Significance of the study

This research in its unprecedented nature, evaluates the performance of a multi layer perceptron in a supervised classification task of galaxy cluster radio emission sources. It provides an opportunity to study how well an MLP learns in classifying radio astronomy sources. It also answers whether a multi layer perceptron has a future in the data processing pipeline of radio data. This study also contributes by identifying and implementing efficient big data classification models for challenges associated with large surveys such as the square kilometre array (SKA) and the evolutionary map of

the universe (EMU) by exploring ways in which big data classification tasks can be made efficient and tackling ways to safely remove unnecessary data bits while ensuring an almost full recovery of sources for large surveys. Since it uses data from the SKA precursor the MeerKAT, this approach also gives insight on how a multi layer perceptron performs on data from particularly high resolution and high sensitivity data which comes with it's own set of challenges.

1.0.9 Limitations of the study

The possible classification of labels used for this study falls within sources found in galaxy clusters. The labels will be generated from a source finding tool used in radio astronomy for data analysis and thus limits the number of labels that will be used to only three. There are several classification criteria that could be used in this instance but they fall outside the scope of this research. In addition the actual data elimination falls beyond the scope of this research since it focuses on training an MLP and testing it to assess its recovery of correct class labels given unseen data in real time.

1.0.10 Chapter summary

This chapter addressed the gap that necessitated the study of a Multi Layer Perceptron as a radio source classifier for galaxy clusters. It explored it as a potential predictor model to augment existing studies that focused on feed forward neural networks using back propagation. It demonstrated the significance of such a study in relation to next generation telescopes that will amass large data amounts. There exists a need to comb through the data for important information and analysis purposes. It also surveyed previous studies related to supervised machine learning algorithms for classification and Multi Layer Perceptron studies in both astronomy and other scientific fields such as medicine. Lastly there was a discussion on a Multi Layer Perceptron as an alternative classification approach to the cataloguing technique. In conclusion studying how a Multi Layer Perceptron performs in a classification task of radio emission sources found in galaxy clusters affords an intuition on how it works, behaves and how it's advantageous features could be implemented in future tasks such as serving as preliminary data eliminator.

1.0.11 Thesis overview

This section of the thesis will give an overview of the thesis in it's entirety.

Chapter 1: The current chapter provides insight on the problem to be addressed by this study, the background information and objectives for this study among many aspects.

Chapter 2: This chapter provides a survey of existing classification techniques and a literature review of a Multi Layer Perceptron.

Chapter 3: This chapter describes a detailed description of the methodology carried out for this research.

Chapter 4: This chapter presents the results obtained after carrying out the study, how performance was evaluated and interpretation of the results obtained.

Chapter 5: This chapter provides conclusions drawn from the results with respect to the objectives, thesis questions and the hypothesis and recommendations for further studies.

Chapter 2

Literature review

2.0.1 Introduction

This chapter will evaluate existing studies in the domains of machine learning and source finding tools in relation to radio source detection and classification while highlighting galaxy clusters and their significance. This is to analyze the gap this research aims to fill. It will also cover existing Multi Layer Perceptron studies while highlighting it's suitability and why it was chosen for this study.

2.0.2 Significance of galaxy clusters from the bigger picture

The cosmic web can be defined as a network that depicts how dark matter, filaments, voids and dense regions filled with galaxies are dispersed within the universe with all these components making up what we know as the web [64]. As a large scale structure it has four components which are filaments, sheets, knots and voids [65]. Figure 2.1 shows a simulation by Springel *et al* [66] of how the cosmic web looks like. Other simulations of the cosmic web include Rodríguez *et al* [67], Metuki *et al* [68], and Xu *et al* [65]. Galaxy clusters reside at the points where the filaments converge and an example of the complexity of the universe, an environment which is believed to have originated from a single yet simple point in time.

The study of galaxy clusters is from the use of advanced radio telescopes that survey the sky in search of these structures and often results in the interpretation of robust theoretical calculations and simulations. This often leads to the understanding of

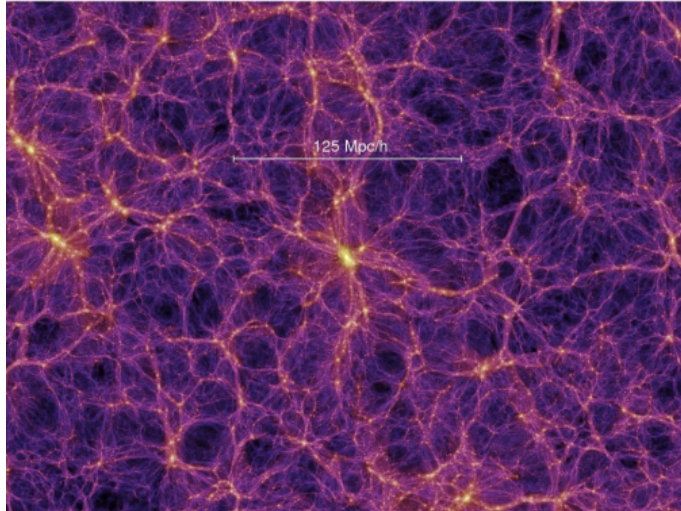


Figure 2.1: A cosmic web simulation by Springel *et al* [66]

the nature of cosmic structures which in turn limits their behaviours to certain parameters e.g constraining the nature of dark matter [66]. In addition, studies such as Larson & Tinsley [69] have shown that the mergers of galaxies triggers a burst of formation of stars making it an interesting phenomenon to study. The majority of the gravitational energy that results from galaxies merging turns into thermal energy through a process known as low-Mach number shocks which has produced substructures often classified as radio halos, relics and active galactic nuclei (AGN) [70], [71]. Observations of galaxy clusters include the study of their constituents, morphological and spectroscopic properties. These studies constrain how clusters were formed, how they evolved and provide meaningful insight on how the universe began and expanded. Clusters are ideal candidates for studying synchrotron radiation [72]. Understanding how galaxy clusters originated has proven to be challenging in terms of using statistical methods due to the large dynamic environment within them [73]. Thus detailed observations provide a better chance of resolving systemic processes of clusters, Chibueze *et al* [60] being a recent case in point.

The formation of galaxies is a focal point for the latest research being that galaxy formation environment provides an intriguing laboratory for studying processes on a larger magnitude. Studying the environment itself is a challenge and Metuki *et al* [68] has shown that the formation of galaxies and galaxy structures is a complicated process. Environments occupied by galaxy clusters can help explain the origin of

their formation and how they have evolved through time with early-type galaxies occupying more dense fields and the less dense fields of the cosmic web being occupied by late-type galaxies [74] [75]. Substructures being discovered in galaxy clusters are an indication that the merging and accumulation within galaxy clusters is an on-going process. This has been evidenced by the appearance of several peaks of bright regions as well as disrupted morphology in radio images. Thus a key point in understanding how the process of merging and accretion happens in radio clusters lies in the identification of known and unknown substructures in images and characterizing them which largely depends on reliable source finding techniques [76].

Studying galaxy cluster populations gives insight on the cosmic background radiation and often require studying them in detail in radio and it's counterpart wavelengths, it also helps explain mass distributions of these galaxies [70], [77]. Substructures found in galaxy clusters are a direct result of the different processes that occur and include radio relics and halos [78]. The variety of morphology found in galaxy clusters is influenced by the environment. When radio jets from a galaxy travel in the interstellar medium it bends the jets to some degree. This results in different morphological shapes of the radio emissions. The degree of morphing is dependent on the density of the ICM and the movement of the galaxy in relation to the ICM [79]. The morphology of sources in addition to the location determines the class assigned to the sources [80]. Galaxies also provide model constraints and parameters for cosmological sources and their processes through simulations and statistical calculations [65] [81].

2.0.3 Types of radio emissions

Radio emission sources refer to sources that emit significantly large amounts of radio waves. While almost all objects found in space emit some radio waves, there exists sources that emit radio waves more than others and include some nebula, quasars, pulsars, and galaxies. There exists two types of radio emission which is spectral line emission and continuum emission. Spectral line emission is emitted at one specific wavelength such as an optical spectral line. When observing line radiation, a telescope is set to a particular frequency. Spectral line emission is used to detect molecules, element or atom found in space since they emit at a

specific frequency such as neutral hydrogen molecular lines which emit at the 21 cm line. Molecules which also emit radiation at a given spectral line include water, ammonia, hydroxyl and carbon monoxide. There are also masers which refer to the magnifying microwave emission. In comparison, continuum emissions such as the one found in galaxy clusters spans over a large range of wavelengths. Such sources can be observed in an ideal wavelength from any selected frequency in the range. There are two types of continuum emission that exists which is thermal and non-thermal continuum emission. Thermal continuum emission include blackbody radiation where warm or heated objects release waves. The frequency of such objects increases and peaks with the increase of its temperature. It also includes free-free emission which refers to the interaction between free electrons that are deflected releasing energy. Non-thermal continuum emission includes synchrotron radiation which has no relation to the temperature of the body. It occurs when high speed electrons motion is coupled in a magnetic field releasing synchrotron radiation. For this study, the interest is in the continuum emission found in galaxy clusters in large-scales.

In a given galaxy cluster they can be as little as two thousand to as many as six thousand sources, with the number of sources expected to increase when the state of the art radio telescope, the SKA becomes fully operational. Precursors such as the MeerKAT and ASKAP as pathfinders for the SKA are important telescopes to study since they use significantly more advanced technology which surpasses existing radio telescopes. This makes the MeerKAT an ideal candidate for this study because of the type of data it produces from it's observations which will be similar to the future complete SKA telescope. There is an almost guarantee that new sources will be discovered as a result of the high sensitivity and resolution of the telescope. The complexity of the data produced does not make it easier to go through the large data or to identify new sources or to classify them by eye. Norris [82], for example states that it will be unlikely that human input would be sufficient to distinguish between source and instrumental artefacts from data such as the one obtained from the pathfinder ASKAP. Given this, Norris suggests studying how to identify likely sources from existing data which could then be used to identify new sources in future based on the experience obtained. Norris [6] also mentions that machine learning techniques would be more

advantageous in this regard as compared to the manual approaches being implemented currently.

Norris [6] further states several challenges facing continuum surveys such as the availability of automated source extracting software suitable for next generation telescopes, insufficient characterization of background and noise levels in images by source finding software used in the field of radio astronomy.

2.0.4 Radio emission classification

Radio emission sources can be grouped into simple or complex sources based on their morphologies. Simple radio emission sources are compact and normally appear as point like sources in images. Complex sources exist in a variety of unlimited shapes and are often extended in nature. Although extended sources substitute a small portion of the overall population of radio emission sources, more have been discovered recently with more most likely to be discovered by future surveys. There is no direct relation between extended emission sources and any individual galaxy within a galaxy cluster which suggests that the emissions are a result of the cluster activities which have indicated magnetic fields and cosmic rays in the ICM. They have been classified as cluster radio shocks, radio halo, and lastly revived active galactic nuclei (AGN) fossil plasma according to Van Weeren *et al* [83]. Van Weeren *et al* further states that radio halos can be further divided into giant and mini halos with the potential of intermediate sources with examples of extended emission sources shown in Figure 2.2. The compact sources appear as the more circular and point like sources in red, outnumbering the extended radio emission sources in the two images. Majority of existing radio emission source catalogs maintain this population fraction where there are more compact sources as compared to extended sources, which is how they naturally occur in nature.

Following source detection comes the classification of the identified sources on the basis of their morphology as an example and dividing them into simple (compact) sources or complex (extended) radio emission sources. The automatic extraction of extended continuum emission sources such as halos from images is well known to be a difficult task [82]. While the majority of radio emission sources existing in nature

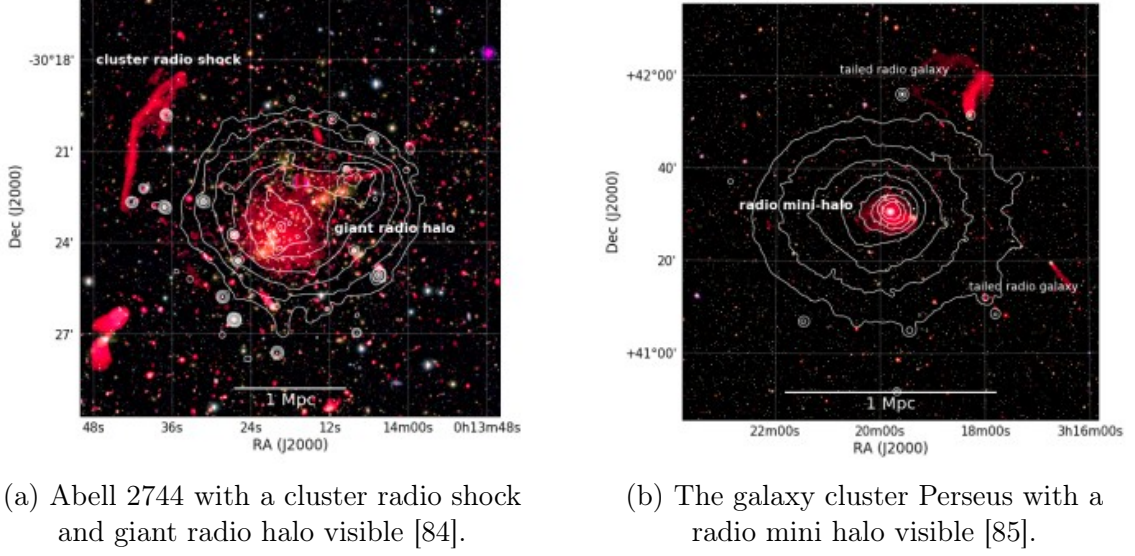


Figure 2.2: Two radio galaxy clusters with various extended emission sources located at different areas of the cluster.

are simple and single point sources, extended sources substitute a portion of the radio emitting source population from galaxy clusters and because of their significantly low numbers when compared to single point sources, it becomes harder to classify them using automated approaches.

2.0.5 Radio emission source classification approaches

There are two approaches discussed in this section with relation to how sources are extracted from images and then classified accordingly which is mostly based on the type of morphology they have. A third approach was discussed in the previous chapter which discussed the cataloguing approach. The two approaches discussed are the common go to software used by radio astronomers for their data analysis and the machine learning algorithms and techniques implemented to fill the gap caused by the lagging of technological advancement by existing source finding systems.

2.0.6 Source finding software

Traditional data analysis software in radio astronomy to some extent has advanced over time to meet the newer and more complex data being produced by next generation telescopes. An example of a more recent development is a consolidated source analysis platform called Collaborative and Integrated platform for Radio Astronomical Source

Analysis (CIRASA) developed by Riggia *et al* [86] aimed at an improved analysis, extraction and source cataloguing process. This was achieved by combining the best features from already existing source finding tools while incorporating new features for an all round hybrid tool. One of it's advantages is, it was tailored for the SKA. However it possesses some general functions that could be utilised by the broader astronomy community in domains such as astronomers observing space in gamma, x-ray or optical. It also utilises High Performance Computing (HPC) systems for faster processing of data given that the SKA precursor ASKAP will find approximately 70 million sources from it's EMU survey [87]. Given this, the platform is still on it's testing stages and unavailable for public use as of the time this study was carried out.

Some source finders have upgraded to support multi thread runs and include AEGEAN [88], PySE which is a source extractor for radio images [89] and SOFIA [90]. Other source tools with optimised processing capabilities include SELAVY [91], a source finder in interferometric images by Lucas *et al* [92], DUCHAMP [93] and CAESAR by Riggi *et al* [94] that focuses on optimising existing source finding tools. They however still struggle with the extraction of extended sources such as sources that exist with two or more components and fail to provide an estimate of the source size during it's extraction process [95].

Hopkins *et al* [96] published a study in which several source finding tools were compared to evaluate their strengths and weaknesses as a way of exploiting their strengths to create a better version of them. Each software was selected based on their merits. The tools chosen for the study included The Python Blob Detector and Source Finder (PyBDSF, formerly known as PyBDSM) [97], AEGEAN, SFind [98], Blobcat [99], SExtractor [100], Selavy [91], PySE and IMSAD [101]. Given that some of the software for the challenge were considered state of the art, the results of challenge were discouraging. Some of the areas of assessment included recovery of isolated sources which would be less of a challenge as compared to the extraction of sources in crowded planes, recovery of sources given a certain threshold, the measurements of source flux densities, the recovery rate of simple and more complex sources with the latter being an area the various software failed dismally [102]. In some cases some software created sources that did not exist in the given images.

The results obtained from the comparison of existing source finding tools by Hopkins *et al* [96] were less than satisfactory, the study failed to address extended radio emission sources. It had a bias towards point like sources with only a small variation of slightly extended sources. Meaning extended sources had a marginal focus and their challenges such as their complex morphology structure being addressed even less. In addition while simulated data is a good substitute for training and testing models, may fail to provide the wide range of complex structures that exist naturally in nature which may not be understood by trained models. A similar study was by Frean *et al* [103] which involved source detection of faint sources in noisy data.

Complex extended and low surface brightness source extraction has been addressed by Johnston & Johnston-Hollitt [95] by using Circle Hough Transform which created a sample of shapes that could be used to represent the various morphologies of extended sources and be used to extract them from images. Johnston & Johnston-Hollitt's model provided an in depth study of extended sources which is rarely performed, shedding light on the challenges faced by extended source extraction. The study also mentioned just how easily their algorithm would be confused in the presence of artifacts, an undesired object that is occasionally found in data such as from the MeerKAT and is often confused by source finders as a source while that is not the case. In general the approach had merits but it was undertaken without a consideration of technologically advanced telescopes such as the SKA precursors with ASKAP described as a complex machine that no individual could fully understand every single bit of by Norris [102].

The outcomes from the study by Hopkins *et al* [96] where several source finding tools were compared included results from determining which tool had the overall best performance. Better performing tools in general comprised of PyBDSF, AEGEAN, BLOBCAT and PySE just to mention a few. It is noted that while the aforementioned software tools generally performed better when compared to the others, no one tool performed well in all aspects. They each possessed notable strengths and weakness but still far from what radio astronomers consider an ideal source finding tool. Among these is PyBDSF which a significant portion of the radio astronomy community

utilises on a daily basis for their data analysis and imaging and which was chosen for this reason as well as its availability and user friendliness for this study. One aspect of PyBDSF as one of the high performing software tools was its high reliability to low signal to noise ratio, the trade off being that its completeness is low i.e it produces a high number of correctly identified sources in images yet does not completely find all sources as expected. It also performed well in estimating the flux density and produced a higher positional accuracy of sources in an image. It thus performed better than the other source finders as stated by the author.

There are other studies that were carried out to determine which software tools performed well given a set of criteria. They include works by Hancock *et al* [88] which compared four source finders considered to be widely used by the astronomy community at the time and the resulting source finder that was designed and implemented to supplement the existing source finders after the performance evaluation from the study was AEGEAN. Westerlund *et al* [104] designed and implemented an evaluator that could automatically be used to determine the accuracy of a source finder. DUCHAMP was tested by Westmeier *et al* [105] as the standard source finder for ASKAP. When it came to its source characterisation, DUCHAMP was affected by several system errors due to the presence of noise in the data produced by the ASKAP telescope. It did however perform as expected when it came to reliability in low signal to noise ratio data. The study of neutral Hydrogen as one of the science goals of the SKA prompted the analysis of source finding tools used in order to determine the optimal one to meet the requirements for its survey of Neutral Hydrogen. DUCHAMP was among the five source finders compared. It was stated by Popping *et al* [106] that all the five source finders were still in their developmental stages and there was room for improvement given that the source finders performed differently when it came to the wide range of source morphologies.

2.0.7 Machine learning

Machine learning can be defined as the study of statistics based models and algorithms that are used to perform a computational task by providing computers with the capability to learn from data without being programmed explicitly to do so [107],

[108]. They are automated models that find distinct patterns in data and make decisions based on those discovered patterns. They also have the ability to adapt when provided with new data. Machine learning techniques have a range of real life applications with automated processes such as prediction models, pattern and image recognition, classification, product recommendation systems and gaming e.g the famous checkers program by Samuel [108] and many other applications. Due to the demand of automated systems with the ability to handle large data volumes in the field of astronomy, it has become increasingly popular. For radio astronomy in particular, this application began quite recently but several studies have already been published with the results produced indicating that the performance of machine learning algorithms as source detecting and classification tools are comparable to already existing source finding software systems and in some instance perform better given that they require minimal human input.

There are several studies carried out that focused on automated source detection and classification. These approaches to applying machine learning in particular supervised machine learning in radio imaging are slowly gaining traction and have shown promising results thus far. Each approach has tackled certain issues and there is yet to be an approach that tackles all if not most challenges in one model. As it stands an-all-in one automatic approach is expected in the future but is yet to be developed. For now studies attempt to provide a solution for a particular problem. Proctor [109] studied the morphology of selected sources from the Faint Images of the Radio Sky at Twenty Centimeters (FIRST) survey. A source was designated a class based on it's shape, the number of components for each source and the type of bent tailed source it was as a morphology type. This method was a mix of manual inspection and the use of a decision tree thus it was only partially automated and not ideal for use in large surveys such as the SKA or it's precursors. It did however demonstrate the robustness of a decision tree and it provided insight on the various morphology types of radio galaxies given that the study produced several tables containing sources that contained either single, double or multiple components looking at their proximity, X-shaped, hybrid, or ring morphology just to mention a few of the classes assigned. Other studies by Proctor in which decision

trees were used for the class assignment of radio sources based on it's probability of belonging to one group or the other included comparing two pattern recognition algorithms to classify obscure sources. A decision tree was selected over an Artificial Neural Network (ANN) because the ANN performed slower [110]. Another was the assessment of a group of decision trees to evaluate the quality of data feature sets [111].

A study by Aniyan & Thorat [21] was aimed at an automated radio galaxy classification using a Convolutional Neural Network as an attempt to automate radio galaxy classification. It was a significant step towards automation in radio images since it is believed to be one of the first applications of a CNN to radio galaxy classification. The results were comparable to existing source finding tools. It was observed that a general classifier which distinguished between the three classes performed badly. When compared to a fusion classifier which distinguished between the three binary classes followed by a majority voting classifier for the final class designation. The fusion classifier had precision ranging from 75% to 95%.

The application of a deep CNN eliminated the manual crafting of features used by some approaches and the manual inspection of sources. This provided a viable option for classification in large data sets. The use of a fusion classifier gave better results and also allowed for the discovery of strange and new sources compared to the more generic model. The results were also compared with manual probing but the use of a DCNN completed the process in a much shorter time period given the number of images. This was a good benchmark for automatic detection and classification of radio sources. A DCNN also fared better than traditional machine learning algorithms with regards to working with images that have a significant number of features to learn from. When contrasted to the traditional machine learning methods, a DCNN had the added advantage of using the knowledge gained in the initial layers of the same model to train in different surveys using transfer learning [112]. The challenges faced included a small training sample with a severe pre-processing and data augmentation process. There was very little variety for each class for the model to train leading to over-fitting by the model due to the bootstrapping that was employed with at least 200 sources per each source category.

Several sources were excluded from the data sample based on the presence of artifacts, sources which were too big to fit in the image dimensions and generally sources that did not have much information on their structure such as uncertainty regarding their morphology. This means that issues regarding sources with the aforementioned characteristics were never addressed. This made the model slightly less effective in the general sense of the classification of known and unknown radio sources. In addition the class that obtained the highest precision also obtained the lowest recall. This meant that the fusion classifier did not recover all the sources from that class.

While Aniyan & Thorat [21] focused on classifying radio galaxies into FRI and FRII and bent-tailed sources, Vafaei *et al* [113] focused on recovery of point like sources where there was a low signal to noise ratio. By enhancing the signal to noise ratio in images followed by blob detection in simulated MeerKAT interferometric data, it produced a pure and complete source catalog. When compared to PyBDSF it outperformed it under the set of performance evaluation metrics with purity relatively the same for both methods. An added advantage of the model termed DeepSource was it's ability to choose purity over completeness and vice versa in any given case. While DeepSource outperformed PyBDSF in the fraction of sources recovered, the author stated that robust source finding tools like PyBDSF will most likely still have a place in the data analysis pipeline given the unique lack of data training stage by the source finder in it's source detection process. The author also stated that given the approach developed, there is most likely a better approach than DeepSource which would also incorporate extended radio sources. This is an aspect this study aims to address by training a Multi Layer Perceptron model to predict radio emission sources found in galaxy clusters belonging to either compact or extended classes. Lukic *et al* [22] used PyBDSF as part of the source detection and classification approach when it was used to create labels for the various source morphologies prior to the classification by the model. It served as an automatic label generator given that the training data had no prior labels that could be used for their Convolutional Neural Network.

Other noteworthy CNN studies include the use of a deep Convolutional Neural Network for an online radio source classification model based on source morphology

by Alhassan [23], classifying radio source using Faster Region-based CNN by Wu *et al* [39], the classification of Active Galactic nuclei (AGN) radio sources using a combination of supervised and unsupervised steps in a autoencoder based on a CNN by Ma *et al* [40], comparing a CNN to a Gaussian fitting method by Lukic *et al* [45], a performance analysis of a CNN versus capsule networks by Lukic *et al* [44], a comparison of five ConvNet architectures for various image classification problems in radio astronomy by Martinazzo *et al* [114] and a comparison of CNN architecture for radio galaxy classification by Becker *et al* [115] .

Previous research has demonstrated the importance of automating classification in radio astronomy. It has also shown that machine learning is a viable approach with exceptional results. Compared to existing software, it shows promise with improvement. The related studies made a good foundation for further work in order to improve performance for classification which afforded the opportunity for this study in particular.

2.0.8 Multi Layer Perceptron

A Multi Layer Perceptron is a supervised ANN which can estimate the function that maps a nonlinear function [116]. It can be made up of several hidden layers excluding the input and output layers as shown in figure 2.3. Earlier research on a Multi Layer Perceptron neural network are from McCulloch and Pitts [117] which investigated the mathematical representation of biological neuron activity. For the majority of deep learning neural networks, training is very time consuming. Mostly due to the several point multiplications performed by the model [118]. Training for the ground breaking study by Krizhevsky *et al* [119] took on average six days which was mostly due to the state of the hardware for data processing, the size of their neural network itself and the size of their training sample. However there exists efficient neural networks such as the MLP that can be a substitute for neural networks that are time laden. The perceptron was developed by Rosenblatt [120] which was a single perceptron that was limited to performing only linear classification. But the introduction of multiple layers established learning of non-linear functions with increased complexity. In particular it establishes a function:

$$f(\cdot) : R^m \rightarrow R^o \quad (2.1)$$

where M is the input dimensions. Figure 2.3 shows an MLP with an input layer, one hidden layer and an output layer. The input layer represents a set of neurons for $x_i | x_1, x_2, x_3 \cdots x_m$ and represents the input features. The values that are transferred from the previous neuron to the next are converted through a linear summation of the weights and the x values. The output layer y produces the predicted value by the model after training. An MLP can approximate the function that is non-linear depending on whether it is a regression or classification task.

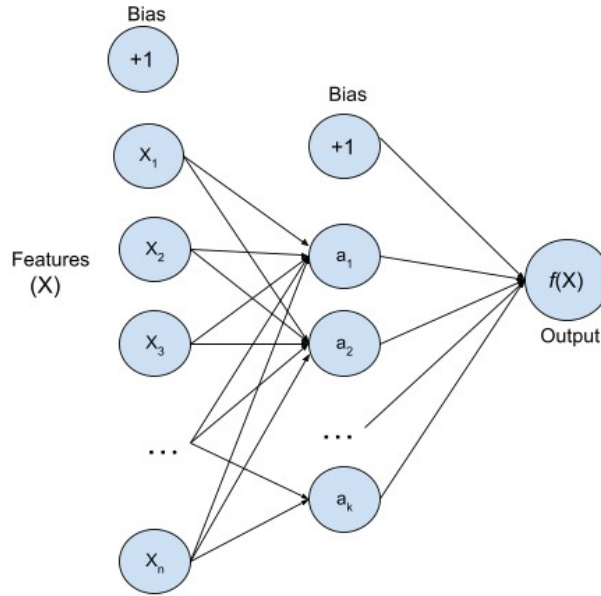


Figure 2.3: An MLP with one hidden layer.

For each input, the weights and bias are initialised and the network goes through a tuning process where the neural network adjusts and assigns weights and biases from the data used for training during each iteration. The learning process will end when the network has determined that the optimal parameters for classification have been determined. The learning rule modifies the weights to ensure that the predicted output approximates the correct output. It inputs a set of features and outputs a predicted vector that is an approximation of the true output value [121], with the output y computed by:

$$y = \varphi\left(\sum_{i=1}^n w_i x_i + b\right) = \varphi(\mathbf{W}^T \mathbf{X} + b) \quad (2.2)$$

where φ is the activation function that is non-linear, $\mathbf{W}$ the weights in vector form, $\mathbf{X}$ is the inputs in vector form and b is the bias. The predicted output is then compared to the expected output from which if they are not equal the weights can be adjusted. A Multi Layer Perceptron is capable of a binary and multi class categorization tasks. The activation function is given by:

$$g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (2.3)$$

as a hyperbolic tangent function. In binary classification $f(x)$ goes through a logistic function:

$$g(z) = \frac{1}{1 + e^{-z}} \quad (2.4)$$

resulting in values that fall between zero and one. For multiclass, $f(x)$ becomes a vector that goes through a softmax function instead of a logistic function given by:

$$softmax(z)_i = \frac{exp(z_i)}{\sum_{l=1}^k exp(z_l)} \quad (2.5)$$

where z_i is the i^{th} value in the softmax corresponding to the class i and k corresponds to the number of classes for the model. The class with the highest probability then becomes the output of the model. The Multi Layer Perceptron uses an efficient algorithm called back-propagation that successfully calculates the gradient through computing the partial derivative from one layer which is then used to compute the gradient of the previous layer.

2.0.9 Multi Layer Perceptron applications in astronomy

The history of the application of MLPs in the area of Multi-object Adaptive Optics (MOAO) is described succinctly by Gómez *et al* [122] from its use in establishing variations caused by atmospheric turbulence in the use of multiple-telescope arrays. These disturbance often impairs the quality of images which can be corrected through adaptive optics. Angel *et al* [123] demonstrated that the variations caused by the

disturbances could be derived using artificial neural networks, in particular an MLP where one hidden layer was used. Between the first study published [123] and the last study at the time of publication [124] there existed other studies which included [125], [126], [127], [128], [129] which employed some version of a Multi Layer Perceptron.

The Multi Layer Perceptron has also been popular in the branch of photometry such as establishing photometric redshifts in large and mixed surveys [130], the cataloguing of photometric redshifts [131] and the analysis of photometric redshift for multi-band surveys of quasars [132]. The use of an MLP for photometric redshift is detected from as early as 2002 with its popularity increasing over time from simple MLPs with a few hidden layers to hybrids that combine more than one MLP and combining MLPs with Quasi Newton Algorithms [132]. The MLPs from these previous studies have demonstrated their adaptability to being trained on homogeneous data which is a mix of simulated and real data. In addition to its rapid computations, good generalization and its implementation on hardware [133]. Needless to say the studies mentioned are no way near the vicinity of radio emission sources and their classification which is the main focus of this study.

In an analysis by Scaife & Porter [134] the weight computations by MLPs was compared to the weight sharing in CNNs that made CNNs perform significantly better. The study focused on implementing an equivariant CNN for the classification of Fanaroff–Riley radio galaxies. The reduction of efficiency by MLPs was attributed to the fact that CNNs generalised translated images better when compared to MLPs which mostly considered each image. While CNNs have been known to be highly successful in image data, MLPs are not commonly used for image classification on their own however through a proper selection of training algorithms that are efficient an MLP can thrive in image classification [135]. CNNs have implemented MLPs through the use of back-propagation by applying non-linear activation to each convolutional output layer [136]. Ralph et al [137] combined a convolutional autoencoder with Self-Organising Map (SOM) intended for source classification in astronomical images. This is a supervised method which has shown promise identifying outliers by using auto-encoded vectors rather than images while providing insight on the features related to

the source morphology. The auto-encoder utilised an MLP with the back-propagation algorithm for input data reduction. The use of vectors instead of the raw images for classification is similar to the approach intended by this study.

2.0.10 Back-propagation

Backpropagation is an algorithm that was first introduced during the 1970s with a significant study being published by Rumelhart *et al* [138] later on. It described back-propagation as an algorithm used for neural networks which regularly adjusts the value of the weights in order to reduce the difference between the predicted value by the network and the actual value of the output represented in Figure 2.4. The stages involved in back-propagation are initializing the weights, the feed forward pass, back-propagating of errors and the adjustment of the weights [139]. In that sense the neurons found in the hidden layers of the network would represent the more important feature from the data for the task at hand while making it easy to identify invariability from the neurons. It also demonstrated how learning took place faster in some networks while using back-propagation as compared to when gradient descent was used.

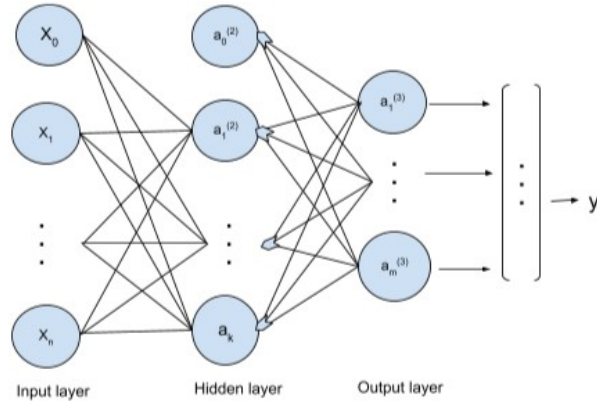


Figure 2.4: Back-propagation from a Multi Layer perceptron

2.0.11 Mathematical representation of back-propagation

The backbone of back-propagation is the partial derivative of the cost function with respect to a weight or bias denoted by:

$$\frac{\partial C}{\partial w} \quad (2.6)$$

Where expression 2.6 is the partial derivative of C with respect to w and

$$\frac{\partial C}{\partial b} \quad (2.7)$$

where expression 2.7 is the partial derivative of C with respect to b , C being the cost function and w being any weight and b being the bias. This shows how the cost changes as the biases or weights changes providing insight on the network behaviour through these changes. To implement back-propagation, there are two assumptions with regards to the cost function. The first one being for every training example x , the cost function can be computed as:

$$C = \frac{1}{n} \sum_x C_x \quad (2.8)$$

Where C_x is the cost function for each data example x . This assumption allows computing the average partial derivative of the cost function from the individual partial derivatives of each data point's cost function with respect to the weights and biases. The second assumption being for each output activation of the neural network, the cost can be computed as a function of it. Given that the y from equation 2.2 is a fixed value, the cost cannot be written in terms of it [140].

2.0.12 General applications of back-propagation

The loss function in a neural network is an excellent indicator of how good a neural network is at the classification task. It also produces the gradients during the training of the network [141]. Computing the errors between the predicted labels and the ground truth is done during training which results in what we call loss. Back-propagation then works to adjust these parameters. This is based on the computed gradient of the

loss given the values of the weights and biases. Consequently reducing the value of the loss function in particular the cross-entropy loss. It is considered an efficient algorithm because of the partial derivatives shown in 2.6 and 2.7 for each training example which can be computed simultaneously in one forward and backward pass through the network. Prior to the study by Rumelhart *et al* [138] given the limitations of the computations when training deep neural networks, the demonstration of how well back-propagation worked increased its popularity in neural networks thus reducing limitations of computation in deep neural networks. It is a simple and easy to implement algorithm when compared to some of the more complex deep learning algorithms.

Back-propagation has been applied across a variety of platforms that include medicine, physics, law enforcement and engineering. Wang and Bi [142] recently published a study where back-propagation was used to diagnose children who most likely had dyslexia. This model was an improvement of existing models and achieved the highest accuracy recorded to date which was 94%. A similar study was carried out to predict dyslexic students based on their grades by determining the relationship between college student performance and the medical condition. There existed a difficulty in determining the underlying relationship between the grades and having dyslexia. An MLP with a back-propagation algorithm became the most ideal neural network to implement because of its efficiency in approximating the likelihood that a student has dyslexia. This was done through learning from complex non-linear functions [143]. Further studies include [144] and [145] from the same main author. Another interesting study that used the back-propagation algorithm was the application of a CNN using back-propagation to identify number plates from stolen vehicles [146].

2.0.13 Significance of using a Multi Layer Perceptron

Previous studies have used MLPs in one form or the other, there has also been a comparison of different neural networks to identify which one works well to a given task which could either be source detection or classification. Deep learning in general has found its way into different branches of astronomy and have greatly assisted in automatic processes that were mainly manual inspections for years prior. For radio astronomy deep learning is still at an infant stage but has thus far proved the need

to employ more deep neural networks to help with handling vast amounts of data, working with complex and qualitatively different data such as data from the MeerKAT and ASKAP. The use of deep learning in radio astronomy is largely the application of CNNs to radio images. This leaves a gap in the evaluation of alternative deep learning algorithms such as employing an MLP to classify radio emission sources. Much research is required in order to understand the morphology of these sources especially extended radio emission sources which have proved that despite employing several deep learning algorithms there exists no one solution to satisfactorily classify them. Applying an MLP should provide better insight on how best to classify these sources and move one step further in developing an ideal classifier.

2.0.14 Chapter summary

This chapter discussed the importance of galaxy cluster in the grander scheme of things which is their role in the intricate system that is our universe, highlighting why it is necessary to have focused studies to understand their nature. It also covered what is required to achieve the science goals related to the study of galaxy clusters, the challenges associated with the data analysis and existing solutions to the said challenges. This was followed by the alternative solution proposed by this study and how it is suitable for addressing the data analysis challenges which is the classification of radio emission sources using a Multi Layer Perceptron. It should be noted that the use of an MLP in the radio astronomy domain is still at its infancy stage thus this study presents a novel opportunity to benchmark its performance as a classifier.

Chapter 3

Methodology

This section of the thesis describes the approach used in the design, implementation and application of a Multi Layer Perceptron as a classifier for radio galaxy cluster emission sources associated with galaxy clusters. The aim of the thesis was to develop a model to classify compact and extended radio emission sources in order to lessen the dependence on human input for the classification task by evaluating how well a Multi Layer Perceptron performs in classifying radio emission sources and its application as a preliminary data eliminator.

3.0.1 Scikit-learn

The Scikit-learn [147] library was used for the Multi Layer Perceptron implementation using its `MLPClassifier` class. The library's hyperparameters for the MLP classifier defined the structure of the network. The data was initially trained using the default hyperparameters to study the model's performance prior to tuning. After tuning, the hyperparameters shown in Table 3.1 were determined to be optimal after assessing the model performance. Some hyperparameters remained as default values while some were changed to optimize the model's performance as shown in Table 3.1. Attention was directed towards the number of hidden layers, the activation function, the solver and number of iterations.

HYPERPARAMETER	VALUE
Hidden layer sizes	(100)
Activation function	Identity
Solver	Stochastic gradient descent
Number of layers	12
Number of hidden layers	10
Alpha	0.0001
Batch size	Auto
Learning rate	Constant
Learning rate init	0.0005
Maximum iterations	200
Shuffle	True
Momentum	0.9
Number of iterations	371
Early stopping	True
Validation fraction	0.1
Number of iterations	38

Table 3.1: The hyperparameters for the Multi Layer Perceptron model implemented.

3.0.2 Multi Layer Perceptron architecture

The Multi Layer Perceptron implemented had 12 layers in total, which included the input and output layers. The total number of hidden layers for the model was 10. The number of input features for the model was 42 and the output layer had 3 possible output labels which corresponded to the three morphology types. The architecture is shown in Figure 3.1.

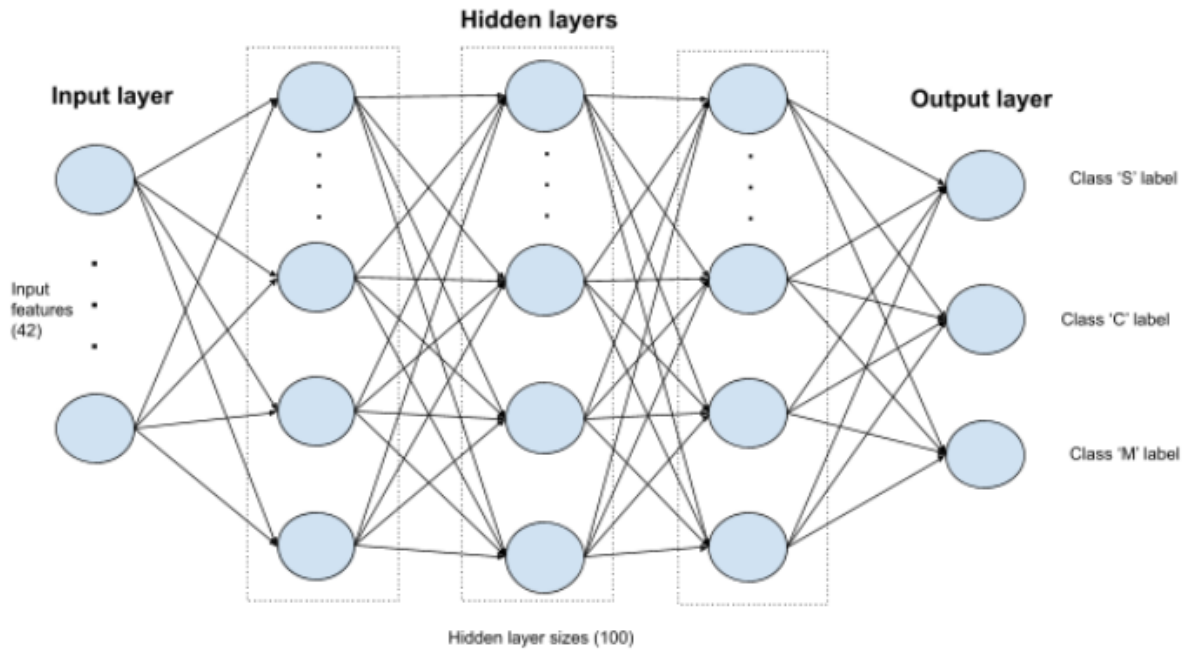


Figure 3.1: The Multi Layer Perceptron architecture implemented.

3.0.3 The experimental framework

To address the objectives previously stated (Chapter 1, section 1.5) which describes the specific objectives of this research the methodology was designed and implemented as shown in Figure 3.2:

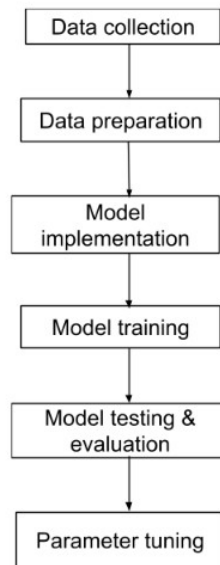


Figure 3.2: The experimental framework

Figure 3.2 depicts the linear workflow implemented for a Multi Layer Perceptron network to classify radio emission sources based on three classes. The three classes were ‘S’ for a single Gaussian fitted in a given island, ‘C’ for a single Gaussian fitted that lies in an island with another source and ‘M’ for a multi Gaussian fitted in a field. The workflow included collecting the data and preparing it to be fed into the model, implementing the model, training it and finally testing the model and evaluating its performance followed by parameter tuning to improve its performance. To address the objectives of the study, there were two training and testing data sets that followed the same workflow shown in Figure 3.2. It should be noted that this research methodology does not consider the classification of radio emission using alternative classification approaches such as labelling them as relics, halos or AGNs. It thus does not go deeper than that. There could also be a mapping of the predictions to the FITS images to view the morphology type of sources in order to get a better understanding of the way the model predicted the sources.

3.0.4 MeerKAT observations

The MeerKAT is a radio interferometer with 64 dishes in total and is located in the Northern Cape, South Africa. The 64-antenna array can be used to detect space from a declination lower than 45° with a combined collecting surface area of $9,000 \text{ m}^2$. As a precursor to the SKA, the MeerKAT is the most highly perceptive interferometer at the decimetre wavelength until the completion of the SKA. It is also optimised for the detection of point sources to accommodate the large-scale survey projects. The MeerKAT operates on a frequency range between 580MHz and 14.5GHz in the Ultra High Frequency (UHF) at 580 to 1015 MHz, the L and S frequency bands at 900 to 1670 MHz and 1.75 to 3.5 GHz respectively. Proposals submitted for observation time using the MeerKAT for large-scale surveys which would exceed 1000 hrs of observation resulted in 11 successful applications. Among the accepted proposals was LADUMA for the extensive study of atomic Hydrogen (HI) regarding its cosmic evolution over 9 billion years ago. It also includes the MIGHTEE, MHONGOOSE and ThunderKAT [148]. The MeerKAT Galaxy Cluster Legacy Survey (MGCLS) is also one such survey that tracks galaxy clusters and has resulted in 115 galaxy clusters observed at the L-band for approximately 1000 hours [149]. This data was made public

for use in scientific studies. The survey focused on investigating the polametric and continuum studies as well as investigating HI in the aforementioned galaxies. The survey has provided opportunities to do research on the source rich environments relating to source populations, morphologies, astrophysical processes and the surrounding environments that make them possible. An aspect which this paper investigates is the automated classification of the sources derived from the survey. As stated in the paper by Knowles *et al* [149] PyBDSF was used for the detection of sources from the observation images. This provided a list of compact and extended sources. They availed the collection of compact sources and state that the process of confirming extended radio emission sources is one exhaustive and time consuming process. This research provides an automated solution for both the compact and extended sources from MGCLS.

3.0.5 Data collection

The first data set comprised of a catalog of radio galaxy clusters in comma-separated values (CSV) files from MGCLS. The CSV files were shared by one of the authors and co-supervisor to this research [60]. There were 104 CSV files in total. Each CSV file listed all radio emission sources detected from each galaxy cluster. Table 3.2 lists the galaxy clusters used for the training and testing of the Multi Layer Perceptron. It provides the name of the cluster, its coordinates i.e J2000 Right Ascension (RA) and Declination (Dec), it's redshift (z) and the number of sources for each galaxy cluster.

No.	Galaxy cluster	RA_{J2000}	Dec_{J2000}	z	Source population
1.	Abell 13	3.3842	-19.5010	0.094	5561
2.	Abell 22	5.1608	-25.7220	0.142	6525
3.	Abell 33	6.7792	-19.5067	0.280	3579
4.	Abell 85	10.4529	-9.3180	0.055	4951
5.	Abell 548	86.7571	-25.6164	0.042	4789
6.	Abell 2485	342.1371	-16.1064	0.247	7734
7.	Abell 2597	351.3321	-12.1244	0.085	1780
8.	Abell 2645	355.3200	-9.0275	0.251	3687
9.	Abell 2667	357.9196	-26.0836	0.230	7699
10.	Abell 2744	3.5671	-30.3830	0.308	5478

11.	Abell 2751	4.0580	-31.3885	0.107	3831
12.	Abell 2811	10.5368	-28.5358	0.108	1375
13.	Abell 2813	10.8517	-20.6214	0.292	5120
14.	Abell 2895	19.5463	-26.9731	0.227	6253
15.	Abell 3365	87.0500	-21.9350	0.093	2290
16.	Abell 3376	90.4256	-39.9851	0.046	4627
17.	Abell 3562	202.7833	-31.6731	0.049	5768
18.	Abell 3667	303.1403	-56.8406	0.056	2248
19.	Abell 4038	356.8796	-28.2028	0.028	6674
20.	Abell S295	41.4000	-53.0380	0.300	7046
21.	Abell S1121	351.2844	-41.2118	0.190	3086
22.	MACS J0025.4-1222	6.3724	-12.3770	0.584	4051
23.	MACS J0257.6-2209	44.4179	-22.1628	0.322	4835
24.	MACS J0257.6-2209	44.4179	-22.1628	0.322	5094
25.	MACS J0417.5-1155	64.3942	-11.9089	0.440	6423
26.	PLCK G200.9-28.2	72.5871	-2.9493	0.220	4037
27.	RXC J0225.1-2928	36.3750	-29.5000	0.060	3576
28.	RXC J0510.7-0801	77.6846	-8.0200	0.220	4004
29.	RXC J0520.7-1328	80.1967	-13.5022	0.336	1804
30.	RXC J2351.0-1954	357.7704	-19.9133	0.248	6733
31.	J0014.3-6604	3.5767	-66.0775	0.155	8929
32.	J0027.3-5015	6.8388	-50.2511	0.145	6110
33.	J0051.1-4833	12.7967	-48.5597	0.187	7653
34.	J0108.5-4020	17.1383	-40.3500	0.143	7214
35.	J0117.8-5455	19.4604	-54.9239	0.251	8689
36.	J0145.0-5300	26.2596	-53.0139	0.118	7447
37.	J0145.2-6033	26.3196	-60.5650	0.184	6939
38.	J0212.8-4707	33.2246	-47.1328	0.115	6653
39.	J0216.3-4816	34.0796	-48.2731	0.163	6076
40.	J0217.2-5244	34.3025	-52.7469	0.343	6613
41.	J0225.9-4154	36.4775	-41.9097	0.220	7607
42.	J0232.2-4420	38.0700	-44.3475	0.284	8320

43.	J0303.7-7752	45.9433	-77.8692	0.274	5501
44.	J0314.3-4525	48.5825	-45.4242	0.073	8739
45.	J0317.9-4414	49.4938	-44.2389	0.075	6590
46.	J0328.6-5542	52.1563	-55.7128	0.086	6321
47.	J0336.3-4037	54.0779	40.6222	0.062	4037
48.	J0342.8-5338	55.7246	-53.6353	0.060	6544
49.	J0351.1-8212	57.7871	-82.2167	0.061	7231
50.	J0352.4-7401	58.1229	-74.0308	0.127	5952
51.	J0406.7-7116	61.6908	-71.2750	0.229	6448
52.	J0416.7-5525	64.1871	-55.4189	0.365	6664
53.	J0431.4-6126	67.8504	-61.4439	0.059	3415
54.	J0449.9-4440	72.4800	-44.6781	0.172	8228
55.	J0510.2-4519	77.5575	-45.3211	0.200	6639
56.	J0516.6-5430	79.1583	-54.5142	0.297	6541
57.	J0525.8-4715	81.4650	-47.2506	0.191	5194
58.	J0528.9-3927	82.2346	-39.4628	0.284	5698
59.	J0540.1-4050	85.0263	-40.8422	0.036	5703
60.	J0540.1-4322	85.0417	-43.3822	0.085	6286
61.	J0542.8-4100	85.7117	-41.0014	0.640	8317
62.	J0543.4-4430	85.8517	-44.5053	0.164	4908
63.	J0545.5-4756	86.3775	-47.9406	0.130	6882
64.	J0600.8-5835	90.2013	-58.5872	0.037	7922
65.	J0607.0-4928	91.7558	-49.4833	0.056	5536
66.	J0610.5-4848	92.6333	-48.8072	0.243	6271
67.	J0616.8-4748	94.2233	-47.8050	0.116	7178
68.	J0625.2-5521	96.3179	-55.3517	0.121	4447
69.	J0626.3-5341	96.5950	-53.6956	0.051	4192
70.	J0627.2-5428	96.8100	-54.4700	0.051	1801
71.	J0631.3-5610	97.8363	-56.1722	0.054	8837
72.	J0637.3-4828	99.3288	-48.4783	0.203	6993
73.	J0638.7-5358	99.6938	-53.9717	0.233	8016
74.	J0645.4-5413	101.3721	-54.2189	0.167	6240

75.	J0658.5-5556	104.6296	-55.9469	0.296	6190
76.	J0712.0-6030	108.0225	-60.5017	0.032	6472
77.	J0738.1-7506	114.5375	-75.1067	0.111	8300
78.	J0745.1-5404	116.2900	-54.0789	0.074	6884
79.	J0757.7-5315	119.4438	-53.2636	0.043	7069
80.	J0812.5-5714	123.1263	-57.2350	0.062	6182
81.	J0820.9-5704	125.2483	-57.0797	0.061	6224
82.	J0943.4-7619	145.8542	-76.3325	0.199	3890
83.	J0948.6-8327	147.1642	-83.4656	0.198	7233
84.	J1040.7-7047	160.1867	-70.7969	0.061	4498
85.	J1130.0-4213	172.5233	-42.2297	0.155	7121
86.	J1201.0-4623	180.2642	-46.3906	0.118	8042
87.	J1240.2-4825	190.0571	-48.4328	0.152	6935
88.	J1248.7-4118	192.1996	-41.3078	0.011	2663
89.	J1358.9-4750	209.7371	-47.8386	0.074	5483
90.	J1410.4-4246	212.6188	-42.7769	0.049	6267
91.	J1423.7-5412	215.9304	-54.2033	0.300	6595
92.	J1518.3-4632	229.5950	-46.5403	0.056	4586
93.	J1535.1-4658	233.7879	-46.9792	0.036	4543
94.	J1539.5-8335	234.8913	-83.5922	0.073	6703
95.	J1601.7-7544	240.4446	-75.7461	0.153	5714
96.	J1645.4-7334	251.3592	-73.5817	0.069	4037
97.	J1653.0-5943	253.2533	-59.7331	0.048	4214
98.	J1705.1-8210	256.2929	-82.1739	0.074	4650
99.	J1840.6-7709	280.1550	-77.1556	0.019	6314
100.	J2023.4-5535	305.8500	-55.5917	0.232	7525
101.	J2104.9-8243	316.2446	-82.7228	0.097	6841
102.	J2222.2-5235	335.5579	-52.5869	0.174	5437
103.	J2319.2-6750	349.8000	-67.8400	0.029	6071
104.	J2340.1-8510	355.0429	-85.1783	0.193	6390

Table 3.2: Description of the cluster sample for this study.

A galaxy cluster was represented with a single CSV file containing several radio emission sources in the order of thousands as shown in Table 3.2. In total there were 607 252 sources, 486 019 sources were used for the training of the Multi Layer Perceptron while 121 233 sources were used for the testing stage. Due to the split of the galaxy clusters files, there was a lower number of sources for the testing sample. Each CSV file had 42 features. The features are listed in Table 3.3. All 42 features were used in training of the Multi Layer Perceptron to classify radio emission sources.

The application of PyBDSF to the data was not required prior to training and testing since the data was already numerical and had been derived from the use of PyBDSF. The data was split 80/20 where 80% of the data was used for training the model and the remaining 20% was used for testing and evaluation of the model in the classification task. It is important to split the data such that the majority of the data is for training the model while the rest is assigned to the validation and testing stages. As an example Hemdan *et al* [150] used the 80/20 data set split also known as the Pareto principle for an ideal data split. Out of 104 files containing numerous radio emission sources, 83 files were used as the training set and 21 files were used as the testing data set. The files were chosen for the split instead of individual source in order to have an array of different galaxy clusters and sources. This also meant that the testing sources were small due to the varying source populations in the CSV files.

The second data set used for the training and testing of the MLP was obtained from the observations of a radio galaxy cluster by the MeerKAT. The data set was derived from radio images of the merging galaxy cluster Abell 3376 hereafter A3376. There were eight images each one an image of the galaxy cluster A3376 observed at different frequency sub-bands. For specific observations the target source is often observed at varying sub-bands in order to capture it in greater detail. What is absent in one image such as a radio source or an astrophysical phenomenon can often be found in another image of the same target thus observing a target in varying frequency sub-bands results in various images of the same target. The observations of the target source in different sub-bands also captured the fluctuation of source populations in each image. Among the radio images obtained was a Multi-Frequency Synthesis (MFS) image which was a composite image of all the seven sub-band images and made the

No.	Data feature
1.	Right ascension
2.	E_Right ascension
3.	Declination
4.	E_Declination
5.	Total_flux
6.	E_total_flux
7.	Peak_flux
8.	E_Peak_flux
9.	RA_max
10.	E_RA_max
11.	DEC_max
12.	E_DEC_max
13.	Maj
14.	E_Maj
15.	Min
16.	E_min
17.	PA
18.	E_PA
19.	Maj_img_plane
20.	E_Maj_img_plane
21.	Min_img_plane
22.	E_Min_img_plane
23.	PA_img_plane
24.	E_PA_img_plane
25.	DC_Maj
26.	E_DC_Maj
27.	DC_Min
28.	E_DC_Min
29.	DC_PA
30.	E_DC_PA
31.	DC_Maj_img_plane
32.	E_DC_Maj_img_plane
33.	DC_Min_img_plane
34.	E_DC_Min_img_plane
35.	DC_PA_img_plane
36.	E_DC_PA_img_plane
37.	Isl_Total_flux
38.	E_Isl_Total_flux
39.	Isl_rms
40.	Isl_mean
41.	Resid_Isl_rms
42.	Resid_Isl_mean

Table 3.3: A list of the data feature used for the training of the MLP.

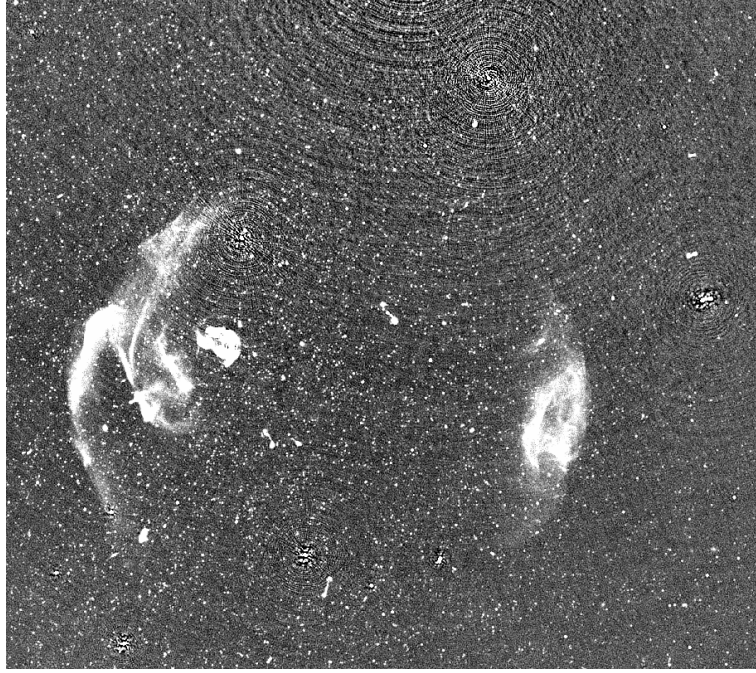


Figure 3.3: Multi-Frequency Synthesis (MFS) image of A3376

eighth image in the data set. Each image was a FITS file type and was approximately 400 MB in size. The data analysis and imaging of A3376 were kindly provided by Chibueze *et al* [60]. Figure 3.3 shows the Multi-Frequency Synthesis (MFS) image of A3376. The image contained thousands of radio emission sources with different morphology types from compact sources to extended sources.

3.0.6 Data preparation

A Flexible Image Transport System (FITS) is the standard file format in astronomy which was developed by Wells *et al* [151] to make it easier for different observatories to share images among themselves and with each other using a standard format. Previously each observatory designed and utilised their own software for processing images with a specific file format, which in turn was mostly incompatible with software from other observatories. The FITS format ensured uniform compatibility regardless of the software and equipment any observatory had. One of its main requirements as an image format was to be flexible enough to transfer n-dimensional structured data arrays regarding an image and its related text information, i.e parameters that give more details about the image and also be able to be used by other fields besides

astronomy.

The hierarchy of a FITS file consists of a primary header data unit(HDU) which is the initial header followed by it's primary array if it exists and provides the general information about the file [152]. The next record is a header with it's own data units arranged in a logical manner. A single FITS file can contain several headers and units with the headers encoded in the American Standard Code for Information Interchange(ASCII). Several FITS formats are available. This has since increased because of the surge in astronomy data processing. This includes random group format and prime data matrix [153]. The initial FITS was stored in magnetic tape and it's size was 23040 bits [154] but has since increased. The Flexible Image Transport System helps link metadata to the image as headers in the standard ASCII format. This is stored in a database that can be custom queried depending on the user intent [155].

Considering that an MLP takes a vector as input for the training and testing, the images of A3376 were firstly transformed into numerical data that could be used for training and testing by the model. In order to create the necessary vectors, PyBDSF was selected as a source finding tool for this purpose. A product of PyBDSF among several is a list of detected sources within the image together with their attributes.

3.0.7 PyBDSF

The Python Blob Detector and Source Finder (PyBDSF, formerly known as PyBDSM) [97] is a radio source detector software widely used in the field of radio astronomy and perceived as a state of the art source finder. Studies such as Vafaei *et al* [41] have compared their machine learning algorithms performance to PyBDSF. While others have included it, using PyBDSF in their machine learning process such as Lukic *et al* [22] as a source labelling tool. It's main purpose is to break down interferometry images such as the ones used in this study into radio sources that can be further used for research. Interferometry images are formed when signals from several telescopes are combined to produce higher resolution images of any given target source such as the MeerKAT. It breaks down an image's components into either wavelets, Gaussians or

shapelets in order to extract sources found in the image. Figure 3.4 shows the sequential flow of processes for PyBDSF source detection that was implemented in this study.

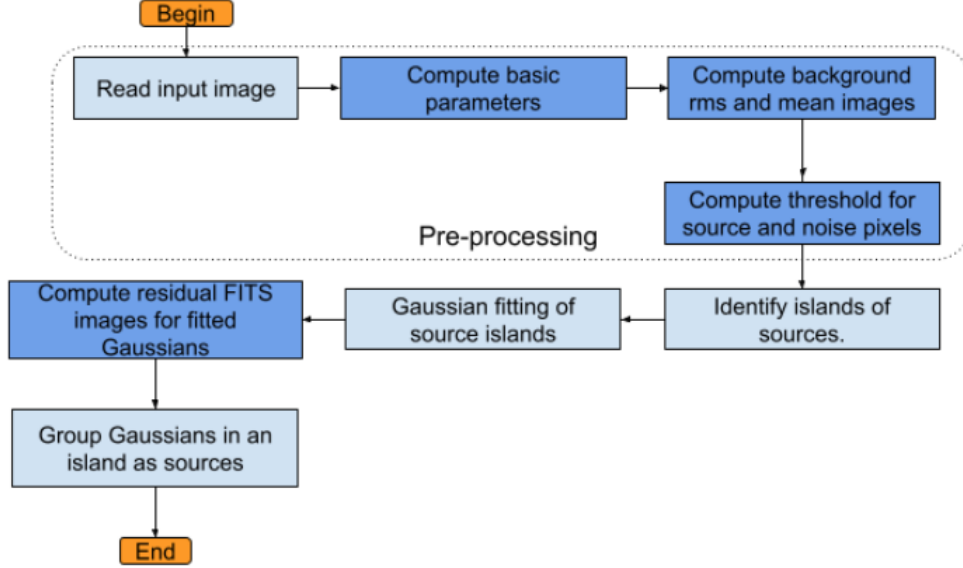


Figure 3.4: Flow chart for the source detection process using PyBDSF

The application of Python Blob Detector and Source Finder for source detection began by reading the input image which normally is a FITS file format [156]. It then produced a ‘ch0’ continuum image with a set of values describing the polarisation state of the image known as the Stokes l polarization parameters. A polarization parameter example is the total intensity of the optical beam. The ‘ch0’ image was the one used in the proceeding image computations. As part of the pre-processing, image statistics such as the dimensions of the image i.e whether it is a three dimensional or four dimensional image, the number of bits per data pixel, possible extensions in the FITS data set, image frequency, equinox and the telescope used for the imaging were computed. This was followed by the calculation of the root mean square and mean images of the background. This provided little information such as the dimensions of artifacts found in the image but helped optimise detecting possible islands in the image. A threshold that was used by PyBDSF for distinguishing noise and source pixel was specified which would be constant throughout the processing of the image. Thereafter islands with possible radio emission sources were identified and then fitted with a couple of Gaussians which were then flagged

to produce more legitimate Gaussians. One of the image products was the residual image in a FITS format produced from the difference between the image with noise and the image where noise had been eliminated. The fitted Gaussians were then shaped into individual sources depending on the number of Gaussians fitted per island.

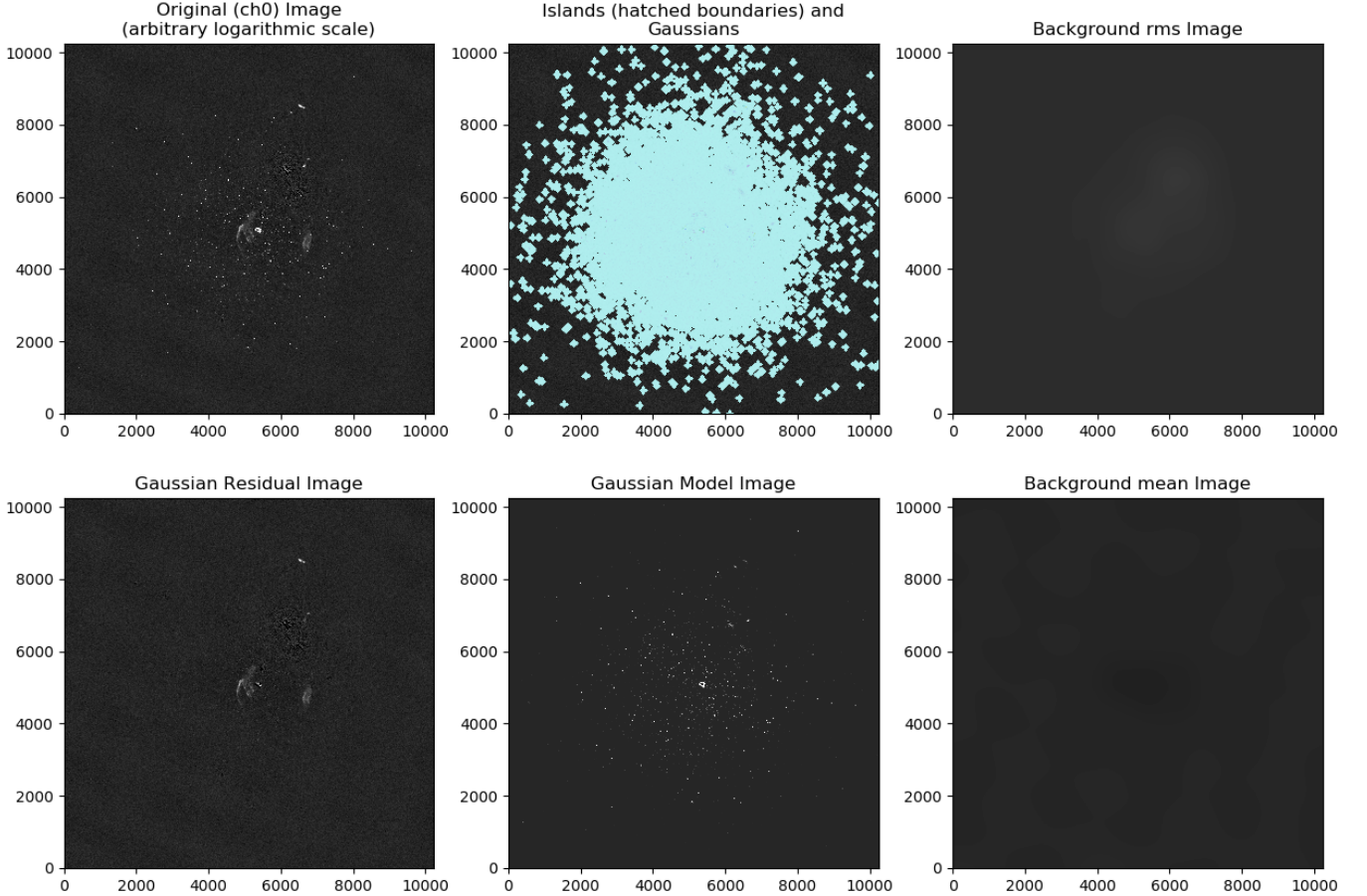


Figure 3.5: Results of the interactive mode where outputs from each computation stage are produced.

Figure 3.5 shows images produced after PyBDSF completed each stage of computations in order to detect radio sources. The top left image was the ‘ch0’ image used for all proceeding computations, followed by an image of all the possible islands that were identified by PyBDSF. The background rms and mean images were produced after they were calculated. The lower middle image was the image of the fitted Gaussians

followed by the Gaussian residual image. All the images were produced because the interactive mode prior to reading the input FITS file was set to ‘TRUE’ and the arrangement of the images in Figure 3.5 were not in chronological order in which PyBDSF processes an image. All eight images underwent the same process to produce eight CSV files. All values for the eight FITS images were set to default which means the threshold for source pixel was calculated and determined by PyBDSF which is usually 5-sigma and 3-sigma for pixel and the boundaries for islands respectively.

There were several exporting file formats for the source catalogs produced by PyBDSF which included CSV, FITS and ASCII to mention a few. For this study a CSV file format was selected as the output of the source catalog. The column names for each source catalog matched the ones from the first data set and in total the output columns for each data set was 42.

The last column in each CSV file was the S_Code which referred to the classification labels for each source designated by PyBDSF. The three class labels for a radio source were ‘S’ which denoted a single Gaussian in a given island, ‘C’ which denoted a single Gaussian but in an island that has another source and ‘M’ which denoted a multi Gaussian radio source. Once the islands are identified by PyBDSF where there are possible radio emission sources, the islands are broken down into Gaussians. These Gaussians are then assembled into sources. As shown by figure 3.6, Source 1 and Source 2 which resulted from a single Gaussian fit from Island 2 will be classified as single compact sources. On the other hand Source 3 where more than one Gaussian is fitted will be classified as an extended complex radio emission source. The Gaussian grouping was the last step in the PyBDSF flow chart which resulted in the generation of the source catalogs. The images with the different frequency sub-bands were not set up in any particular order for processing using PyBDSF.

Following the transformation from radio image data in order to detect all sources into CSV files, the next step was to analyse the data statistics contained in the CSV files and compare data populations. Table 3.4 shows the distribution of radio emission sources detected for each class where the data set column indicates the number of data points obtained from using PyBDSF. The class labels for the data sets and their

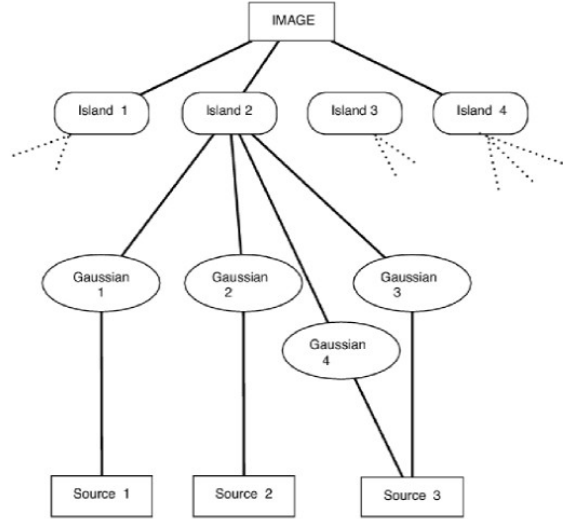


Figure 3.6: Image hierarchy used by PyBDSF when processing an image. It shows how Gaussian grouping occurs and the class labels assigned to each confirmed radio emission source are deduced.

population distributions are shown in the classification labels column.

In terms of radio emission source class distribution, the single Gaussian per island sources denoted by ‘S’ outnumbered the other two class labels significantly. In addition some data sets had an overall low number of radio sources for training and testing such as sub-band 6 and 7 and an even lower number of class ‘M’ radio sources as shown in Table 3.4. In total there 6359 sources used for data set 2. The radio emission sources from each sub-band was split 80/20 for training and testing.

Classification labels

DATA SET	S	C	M	TOTAL
Sub-band 1	2325	325	143	2793
Sub-band 2	4461	869	251	5581
Sub-band 3	3210	464	151	3825
Sub-band 4	1741	205	110	2056
Sub-band 5	3150	580	161	3891
Sub-band 6	1102	150	76	1328
Sub-band 7	1051	178	57	1286
Sub-band 8	7146	1231	344	8721
Total				6359

Table 3.4: The frequency sub-bands per image with corresponding source population for the three classification labels ‘S’, ‘C’ and ‘M’.

The skewed data distribution is further shown by Figure 3.7 where the data sets were plotted against the three class labels. This showed the margin by which the class label ‘S’ exceeded the other two class labels ‘C’ and ‘M’. Additionally it was apparent that the class label ‘M’ rarely had any spike increase across all the data sets which may have indicated that the availability of sources from the class label ‘M’ was not influenced by the frequency sub-bands variation but rather that the source population itself was quite rare in reality. Since this study explores the performance of a Multi Layer Perceptron in a multi-class classification task, the aspect of how it handles the imbalance will also be studied.

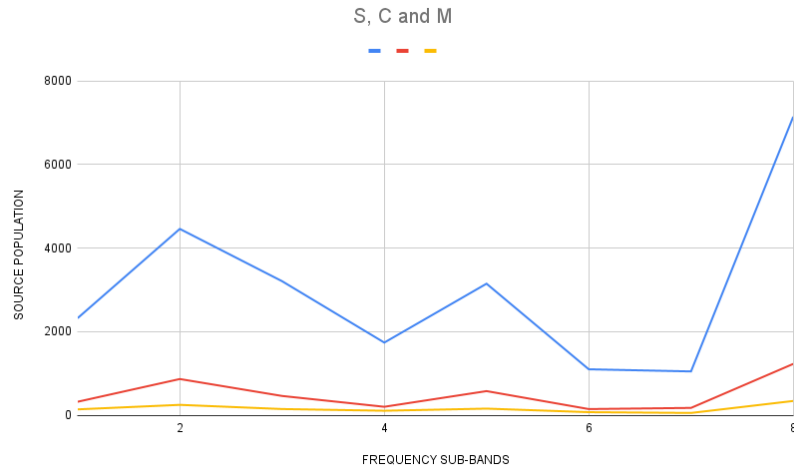


Figure 3.7: Source population from the eight data sets plotted against the three class labels denoted by ‘S’, ‘C’ and ‘M’

3.0.8 Experimental Setup

The experimental setup for this study was designed to train and test a Multi Layer Perceptron to predict class labels for radio emission sources. To do this each data set was loaded into the MLP network where it was read and the basic parameters of each data set computed such as the number of data entries, number of columns, class labels per data set and the data type for each data entry. Data cleaning involved removing any duplicates in the data sets, removing empty rows, rows with null values. Provided with this information the data sets were then normalised so that data entry values all ranged between zero and one to avoid the model being biased towards significantly large values. The ‘S_Code’ column of each data set was further encoded to three values

0, 1, & 2 which corresponded to class labels ‘C’, ‘M’ & ‘S’ respectively. The training and testing process involved splitting the data sets for training and testing. Thereafter, test the model on unseen data. Lastly the model performance was computed using the performance metrics discussed (Chapter 3, section 3.4). The model parameters were tuned given the model testing to determine the optimal parameters. Figure 3.8 depicts the summary of work flow for the experimental set up used. For the two data sets used, each was split into 80/20 where 80% of the data was used for training and 20% was used for testing the model.

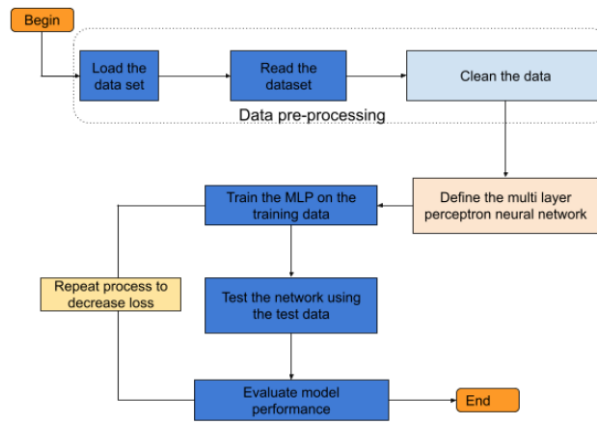


Figure 3.8: Flow chart for the experimental framework used to evaluate a Multi Layer Perceptron for prediction radio emission sources.

3.0.9 Performance metrics

Supervised machine learning problems can be split into three categories where an algorithm is developed to use patterns in data and inference to predict an output. These include regression, binary classification and multi-class classification problems. Regression problems involve predicting a continuous class or variable, while binary classification problems are based on only two variables and a multi-class classification involves more than two variables or classes. Due to the different nature of these problems, there are different performance evaluation metrics that can be applied to each category. Selecting one evaluation metric is not always enough to fully understand the problem being solved and how it was being solved by the model. Utilizing more than one metric can provide a bigger picture and give a better understanding. This section will cover evaluation metrics chosen for a multi-class classification problem

by a Multi Layer Perceptron that was used to predict radio emission sources based on three possible classes. The metrics used were accuracy, confusion matrix, recall and the F1 score with each metric providing insight on the performance of the model. Since this is a multi-class classification problem, the metrics used took into account the class cases for this study. Performance metrics applied in this study will be calculated individually for each class and then discussed later. Selecting several performance evaluation metrics for this study was essential since the data was highly unbalanced and the metrics guaranteed insight on how the model performed in order to understand how the MLP behaved and performed in predicting classes for radio emission sources.

Since there are various ways of viewing a model performance given it's aim, performance for this study was defined by how well the model correctly assigns a class label for each radio emission source. The aim was to ensure that all sources were correctly classified and assigned their right class label. This was important as the incorrect or false identification of emission sources equates to losses in the order of thousands. This is not ideal when dealing with real time elimination of undesired data for large radio surveys. Thus it is ideal to have a model that has a higher precision if there is a higher recovery of true positives.

3.0.10 Confusion matrix

The confusion matrix provides a comparison between predicted values and the actual values with the confusion matrix having $N \times N$ dimensions where N is the number of possible class outputs. For a three class classification problem, the confusion matrix will have a 3×3 dimension matrix. One dimension of the matrix contains the true values which are then mapped to the predicted values producing a tabular format of the model performance with regards to the following terms: True Positive(TP), True Negative(TN), False Positive(FP) and False Negative (FN). True Positive values implies the true value of the source and the model predicted value are the same and are normally found along the diagonal of the confusion matrix while a False positive implies that the model predicted a source as a positive when in fact the actual value was negative. Thus the model incorrectly predicted the class label for that particular

source. A True Negative implies the true label and the predicted label are the same which means the model correctly predicted a class as a negative where it's true value was actually negative.

All false predictions are normally found in the cells that are not part of the diagonal cells containing the True positives values. A False negative indicates that the true value was positive but the model incorrectly predicted it a negative. A confusion matrix allows for the computation of the model precision, specificity and recall. Since the confusion matrix also shows the model's lack of ability to correctly predict the class labels, it is a matrix of the model's confusion.

TRUE CLASS	C	TP_C	E _{CM}	E _{CS}
	M	E _{MC}	TP_M	E _{MS}
	S	E _{SC}	E _{SM}	TP_S
		C	M	S
		PREDICTED CLASS		

Figure 3.9: A N x N confusion matrix developed for the three class labels namely 'S', 'C' and 'M' where the classification results can be calculated for each class.

The first representation of a confusion matrix was by Pearson [157] where he tabled the relationship between two variables. This was seen as the first use of a confusion matrix. This was eventually followed by Townsend [158] where confusion matrix for the letters of the alphabets was made which resulted in a 26 x 26 matrix. It was later published by Kohavi and Provost [159] leading to it's popularity in the machine learning community. Several studies have used a confusion matrix for multi-class classification models performance evaluation in recent times such as Makhtar *et al* [160] in which predictive models were compared with regards to their confusion matrix. Others in-

clude [161], [162], [163], [164]. A confusion matrix for a multi classification can differ. Two examples are Krüger [165] and Tharwat [166]. Where the True Positives can be found in the diagonal cells for the matrix provided by Tharwat, the matrix provided by Krüger has only one True Positive cell at the center of the matrix.

3.0.11 Precision

Precision is used to evaluate how well the model correctly classifies the true positives out of all the positives computed by the confusion matrix. A high precision value shows that a very few negatives were incorrectly classified as positives. It is calculated as shown below:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Where TP is the number of True Positives and FP is the number of False Positives.

3.0.12 Recall

A high value for the recall indicates that a very small number of values were incorrectly classified as negatives. While precision evaluates how well the model correctly identifies true positives from all the positives computed, recall focuses on determining the proportion of true positives that were predicted correctly. Recall is defined by the following formula:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Where TP is the true positives and FN is the false negatives which are incorrectly classified positives as negatives.

3.0.13 Accuracy

There is a lot of emphasis on the accuracy of a machine learning model and while it is commonly applied as a performance evaluation metric, it does not provide the desired insight of the model performance. In this study, there is an imbalance of class. As such it does not become apparent through the use of accuracy how the model performs. The accuracy of the model is computed by determining the fraction of correctly predicted

values without any regards to whether it is a true positive or true negative. Accuracy can be calculated as shown below:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

Where TP is the number of True Positives for each class label, TN is the number of True Negatives, FP is the number of False Positives and FN is the number of False Negatives for each class label.

3.0.14 F1_score

F1_score is an accumulated performance score that is derived from the weighted average on the precision and recall of the model in measuring how successful it is. It is a median point between either using the number predicted positive values or the true positive values to measure the total number of correctly predicted positives. Also known as the harmonic mean. The F1_score is intuitive in relation to unbalanced data sets thus it is a suitable performance score metric of how successful the model is. For a multi-class classification task, the F1_score can be computed for each class label by :

$$F1_Score = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Where precision and recall is calculated for each class label as discussed in previous sections.

Generally the trade off between precision and recall is if the model is mainly biased towards the positives, then the recall will be high but the precision will be low. If the model is biased towards the negatives then the opposite would be true where the recall is low and the precision of the model is high. In theory the aim is to ensure that the recall, precision and accuracy values all together are closer to one.

3.0.15 Chapter Summary

The methodology chapter covered the design and implementation of an MLP in order to evaluate its performance in classifying radio emission sources associated with radio galaxy clusters. It described the data collection, data preparation, the design and implementation, its training and testing and model performance evaluation and tuning. It concluded by discussing the performance evaluation metrics that were adopted and how they assisted in describing the strengths and weakness of the model when it came to the testing phase.

Chapter 4

Results and discussion

4.0.1 Introduction

This chapter of the thesis will present and then discuss the results obtained from the design and implementation of a Multi Layer Perceptron in the classification of radio emission sources found in galaxy clusters. It will present results from the training of the model followed by the testing results. The results presented were derived from the use of the evaluation metrics discussed in chapter 3, section 3.4 in order to assess how well the model performed in predicting class labels for the testing data. This will be followed by the discussion on the factors affecting the model's training and testing performance.

4.0.2 The Multi Layer Perceptron training results

In order to assess how the Multi Layer Perceptron performs, there was an exploration of different hidden layers values and how they affected the training and testing of the model. Five different values were used to assess the effect of the number of hidden layers on the MLP's performance. The results are shown in graph 4.1 where the Multi Layer Perceptron with 1, 4, 6, 10 and 20 hidden layers was implemented. This is followed by the validation curves for the 5 different values hidden layers parameters in graph 4.2.

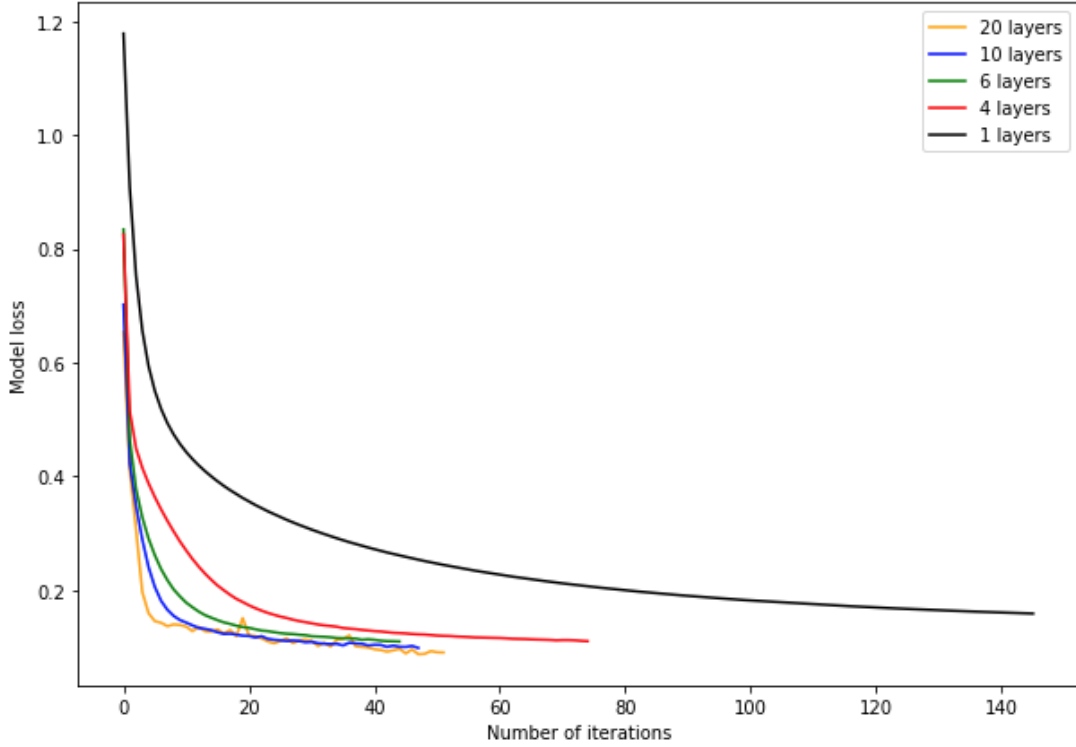


Figure 4.1: The MLP loss curves for different number of hidden layers.

Graph 4.1 shows that the Multi Layer Perceptron with one hidden layer took the longest time to converge with 146 iterations. This was followed by the Multi Layer Perceptron with 4 hidden layers which reached convergence after 75 iterations. The Multi Layer Perceptron with 6 hidden layers reached convergence after 45 iterations. The Multi Layer Perceptron with 10 hidden layers reached convergence after 48 iterations. Lastly the Multi Layer Perceptron with 20 hidden layers reached convergence after 52 iterations. The lowest loss obtained by the model during training was by the Multi Layer Perceptron with 20 hidden layers and the Multi Layer Perceptron with 10 hidden layers with a model loss of approximately 0.1.

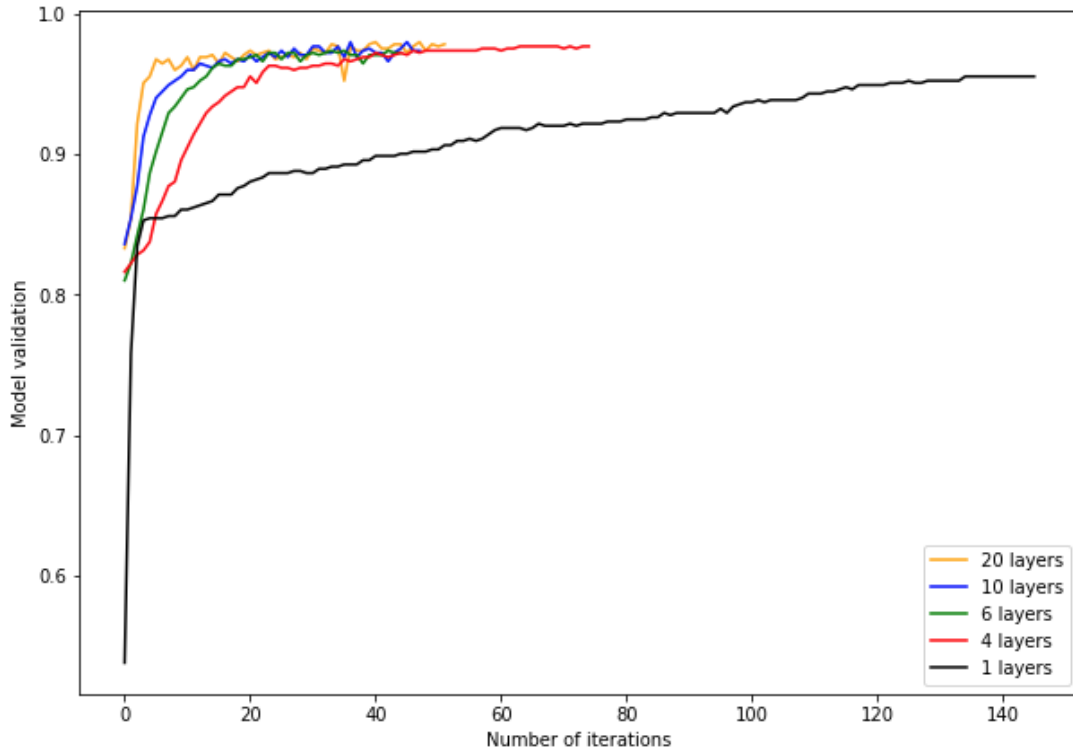


Figure 4.2: The MLP validation curves for different number of hidden layers.

The highest validation score was by the Multi Layer Perceptron with 20 hidden layers. This was followed by the Multi Layer Perceptron with 10 hidden layers. Then followed closely by the Multi Layer Perceptron with 6 hidden layers with the best validation score of approximately 0.97. The Multi Layer Perceptron with 1 hidden layer had a validation score of approximately 0.96.

4.0.3 The Multi Layer Perceptron testing results

The results presented are the outcomes of training and testing a Multi Layer Perceptron to classify radio emission sources that are associated with galaxy clusters. The three class labels were class 0, class 1 and class 2 which corresponded to whether a radio emission source was a double source, an extended radio emission source or a single compact source respectively. This section presents the confusion matrix accuracy results 1, recall results 1, F1_score results 1 for each of the 21 files used for testing. Additionally the results of the evaluation of a Multi Layer Perceptron model on data with different frequency sub-bands of the galaxy cluster A3376 are presented. This was to determine if applying a Multi Layer Perceptron model on data derived from observing a particular galaxy cluster at different frequency sub-bands had an impact on how the model performed in classifying radio emission sources from that galaxy cluster. A total of 8 data sets of the galaxy cluster A3376, with a 80/20 split for the training and testing were used. The results are presented as confusion matrix results 2, accuracy results 2, recall results 2 and F1_score results 2. In testing the performance of the MLP while evaluating it as a preliminary data eliminator, the testing mimicked real-time flow of data with the files each representing incoming data.

4.0.4 Confusion matrix results 1

Presented in this section is the confusion matrix for each data set used in the testing phase of the Multi Layer Perceptron model in classifying radio emission sources from unseen data. There were 21 confusion matrix corresponding to the 21 galaxy clusters the model was tested on.

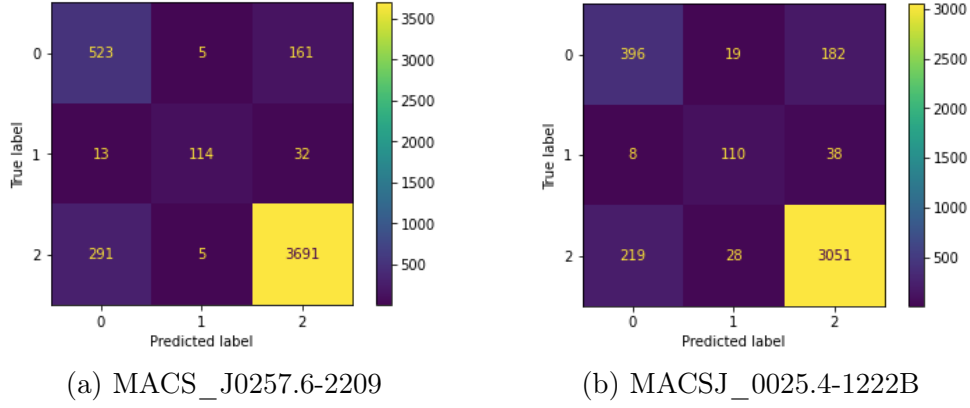


Figure 4.3: Confusion matrix for MACS_J0257.6-2209 and MACSJ_0025.4-1222B

Figure 4.3 shows the confusion matrix for two galaxy clusters namely MACS_J0257.6-2209 and MACSJ_0025.4-1222B. The class with the highest number of incorrectly classified sources was class ‘S’ where 291 and 219 sources were incorrectly classified as class ‘C’ where as they belonged to class ‘S’ for galaxy cluster MACS_J0257.6-2209 and MACSJ_0025.4-1222B respectively.

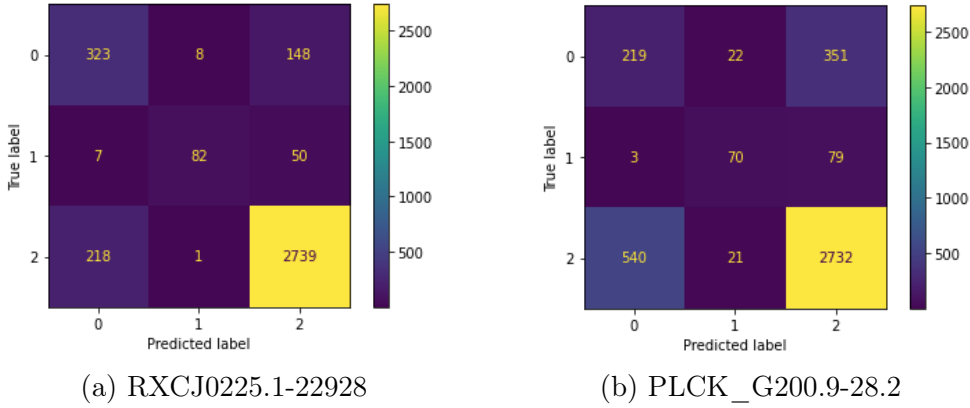


Figure 4.4: Confusion matrix for RXCJ0225.1-22928 and PLCK_G200.9-28.2

Figure 4.4 shows the confusion matrix for two galaxy clusters RXCJ0225.1-22928 and PLCK_G200.9-28.2. RXCJ0225.1-22928 has the lowest number of incorrectly classified sources overall which was 432 when compared to PLCK_G200.9-28.2 which had 1016 sources incorrectly classified overall.

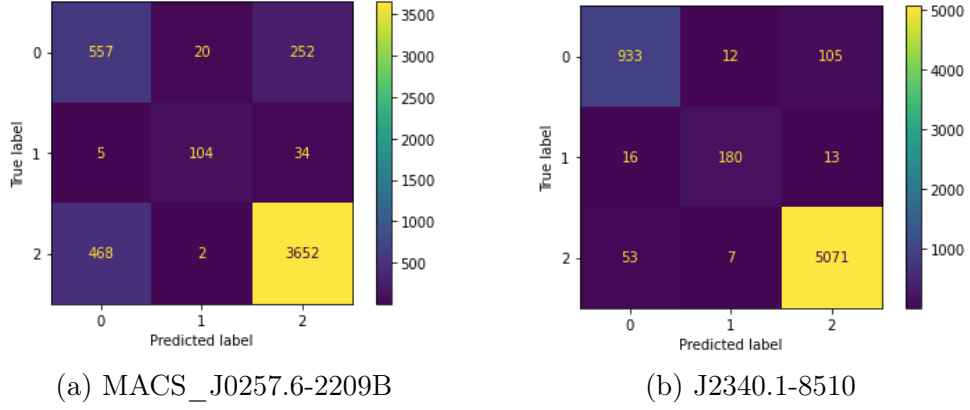


Figure 4.5: Confusion matrix for MACS_J0257.6-2209B and J2340.1-8510

Figure 4.5 presents the confusion matrix for two galaxy clusters MACS_J0257.6-2209B and J2340.1-8510. For the galaxy cluster MACS_J0257.6-2209B 557 sources were correctly identified as class ‘C’, 104 correctly classified as belonging to class ‘M’ and 3652 sources were correctly identified as class ‘S’. As for the galaxy cluster J2340.1-8510 933 sources were correctly assigned to class ‘C’, 180 sources correctly assigned to class ‘M’ and 5071 sources correctly identified as belonging to class ‘S’.

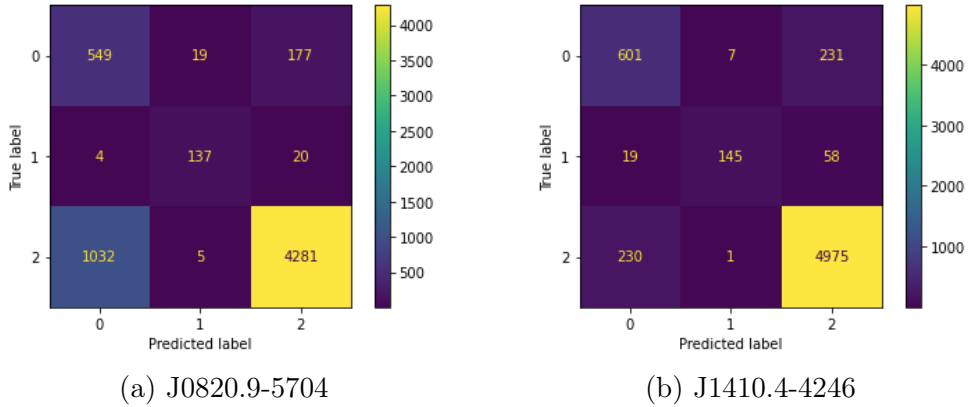


Figure 4.6: Confusion matrix for J0820.9-5704 and J1410.4-4246

Figure 4.6 presents the confusion matrix for two galaxy clusters J0820.9-5704 and J1410.4-4246. There was a high number of class ‘S’ misidentified sources from the galaxy cluster J0820.9-5704. In total 1032 sources were incorrectly identified as belonging to class ‘C’ were as the sources belonged to class ‘S’. When compared to the galaxy cluster J1410.4-4246 only 230 sources were incorrectly identified in that same category. This was followed by class ‘C’ sources incorrectly identified as class ‘S’

sources for both galaxy clusters J0820.9-5704 and J1410.4-4246 with 177 sources and 231 sources respectively.

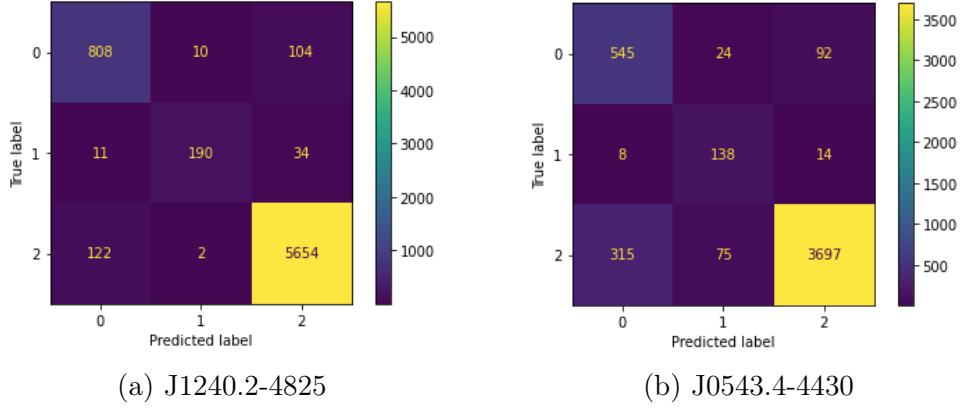


Figure 4.7: Confusion matrix for J1240.2-4825 and J0543.4-4430

Figure 4.7 shows the confusion matrix results for galaxy clusters J1240.2-4825 and J0543.4-4430. The galaxy cluster J1240.2-4825 had more sources when compared to the galaxy cluster J0543.4-4430. J1240.2-4825 had 6935 in total in which 6652 were correctly classified while 283 source were incorrectly classified. For the galaxy cluster J0543.4-4430 there were a total of 4908 sources, 4380 of which were correctly classified and 528 incorrectly classified by the model.

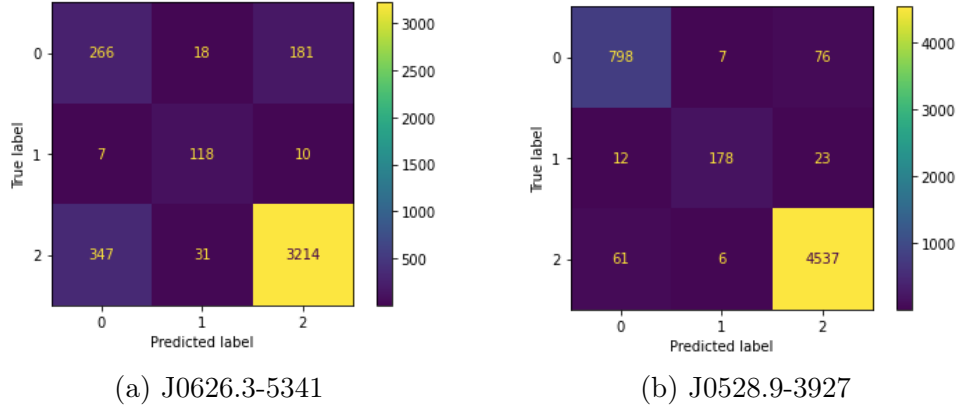


Figure 4.8: Confusion matrix for J0626.3-5341 and J0528.9-3927

Figure 4.8 presents the confusion matrix results for galaxy clusters J0626.3-5341 and J0528.9-3927. The lowest number of incorrectly classified sources for galaxy cluster J0626.3-5341 was 7 in which sources belonging to class ‘M’ were incorrectly identified as sources belonging to class ‘C’ by the model. The galaxy cluster J0528.9-3927’s lowest number of incorrectly identified sources per class was 6 where class ‘S’ sources

were incorrectly assigned the class ‘M’ label.

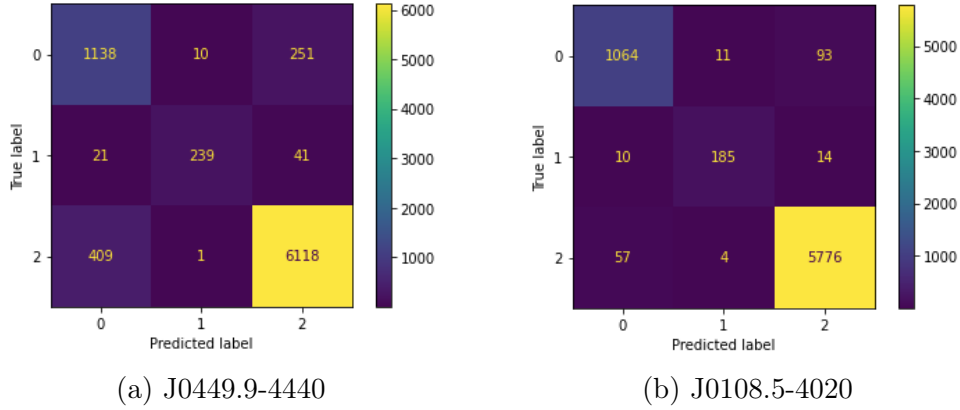


Figure 4.9: Confusion matrix for J0449.9-4440 and J0108.5-4020

Figure 4.9 shows the confusion matrix results for galaxy clusters J0449.9-4440 and J0108.5-4020. Out of a total of 1399 sources in class ‘C’ 1138 were correctly identified as belonging to the group, 239 out of 301 sources were correctly identified as belonging to class ‘M’ and out of 6528 sources in class ‘S’, 6118 were correctly classified as belonging to that group for the galaxy cluster J0449.9-4440. For the galaxy cluster J0108.5-4020 out of 1168 sources in class ‘C’, 1064 sources were correctly classified, 185 sources out of 209 sources were correctly assigned the class ‘M’ label and 5776 out of 5837 sources were correctly classified as class ‘S’ sources by the model.

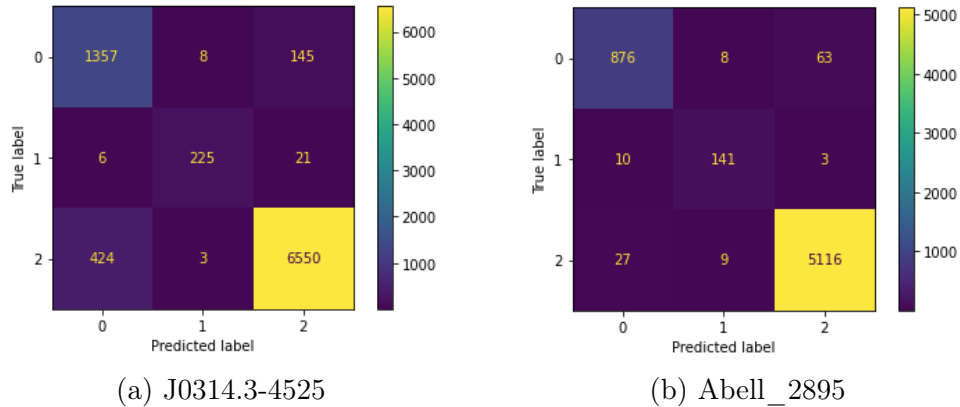


Figure 4.10: Confusion matrix for J0314.3-4525 and Abell_2895

Figure 4.10 presents the confusion matrix results for galaxy clusters J0314.3-4525 and Abell_2895. Out of 8739 sources, 8132 were correctly classified and 607 sources were

incorrectly classified for the galaxy cluster. For the galaxy cluster Abell_2895 out of a total of 6253 sources 6133 sources were correctly identified while 120 sources were incorrectly classified.

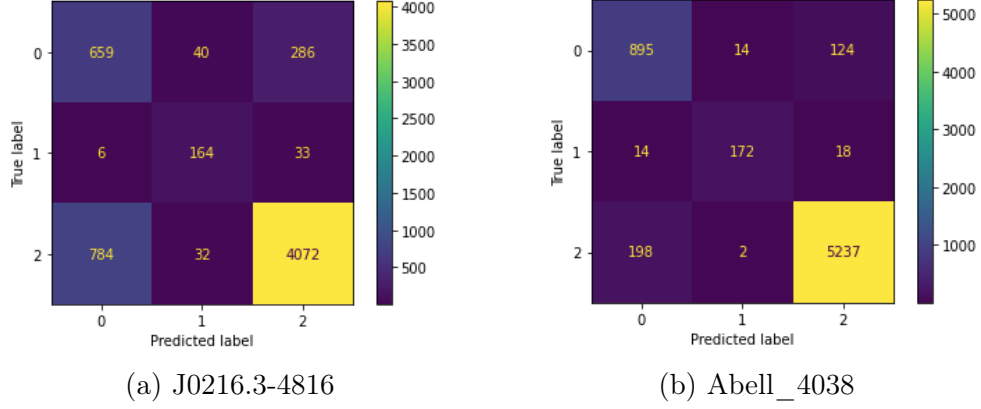


Figure 4.11: Confusion matrix for J0216.3-4816 and Abell_4038

Figure 4.11 shows the confusion matrix results for galaxy clusters J0216.3-4816 and Abell_4038. The lowest number of incorrectly classified sources were 6 sources from class ‘M’ which were incorrectly classified as class ‘C’ sources for the galaxy cluster J0216.3-4816 and the highest incorrectly classified was 784 class ‘S’ sources which were assigned the class ‘C’ label. For the galaxy cluster Abell_4038 the lowest number of incorrectly classified sources was 2 class ‘S’ sources that were assigned the class ‘M’ label and the highest number of incorrectly classified sources was 198 from the class ‘S’ sources misidentified as class ‘M’ sources.

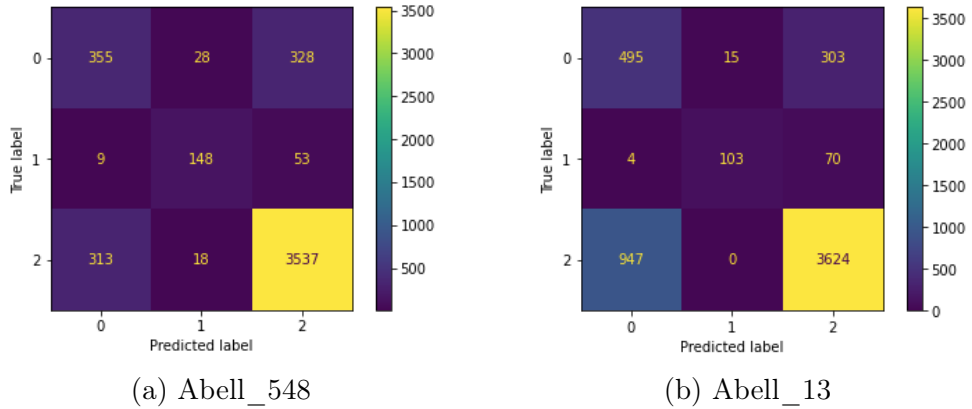


Figure 4.12: Confusion matrix for Abell_548 and Abell_13

Figure 4.12 shows the confusion matrix results for galaxy clusters Abell_548 and Abell_13. The galaxy cluster Abell_548 had the lowest number of incorrectly classified sources with 749 when compared to the 1339 misidentified sources from the galaxy cluster Abell_13.

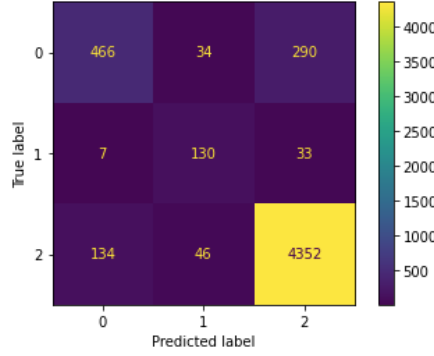


Figure 4.13: Confusion matrix for galaxy cluster Abell_2744

Figure 4.13 shows the confusion matrix results for the galaxy cluster Abell_2744. 466 sources were correctly classified as class ‘C’ sources, 130 were correctly classified as class ‘M’ sources and 4352 sources were correctly classified as class ‘S’ sources. The highest number of incorrectly classified sources was 290 class ‘C’ sources misidentified as class ‘S’ sources.

4.0.5 Accuracy results 1

This section presents the accuracy results with table 4.1 presenting the model accuracy for each of the galaxy clusters used for the testing of the model. The highest accuracy obtained by the model was 0.98 and the lowest accuracy obtained by the model was 0.75.

GALAXY CLUSTER	ACCURACY	TOTAL NO. OF SOURCES
RXCJ0225.1-22928	0.88	3576
MACSJ_0025.4-1222B	0.88	4051
MACS_J0257.6-2209	0.90	4835
PLCK_G200.9-28.2	0.75	4037
MACS_J0257.6-2209B	0.85	5094
J2340.1-8510	0.97	6390
J0820.9-5704	0.80	6224
J1410.4-4246	0.91	6267
J1240.2-4825	0.96	6935
J0543.4-4430	0.89	4908
J0626.3-5341	0.86	4192
J0528.9-3927	0.97	5698
J0449.9-4440	0.91	8228
J0108.5-4020	0.97	7214
J0314.3-4525	0.93	8739
Abell_2895	0.98	6253
J0216.3-4816	0.81	6076
Abell_4038	0.94	6674
Abell_548	0.84	4789
Abell_13	0.76	5561
Abell_2744	0.90	5492
Average accuracy	0.89	121 233

Table 4.1: The accuracy results for the 21 galaxy clusters used for testing the Multi Layer Perceptron model and the total number of sources for each galaxy cluster.

4.0.6 Precision results 1

This section presents the precision results shown in Table 4.2. The highest precision obtained by the model among the three classes was 0.99 for class ‘S’ sources in the galaxy cluster Abell_2895. The lowest precision being 0.29 for class ‘C’ in the galaxy cluster PLCK_G200.9-28.2. The highest precision obtained for the class ‘C’ was 0.94 and the highest precision obtained for the class ‘M’ sources was 0.96.

GALAXY CLUSTER	CLASS ‘C’	CLASS ‘M’	CLASS ‘S’
RXCJ0225.1-22928	0.59	0.90	0.93
MACSJ_0025.4-1222B	0.64	0.70	0.93
MACS_J0257.6-2209	0.63	0.92	0.95
PLCK_G200.9-28.2	0.29	0.62	0.86
MACS_J0257.6-2209B	0.54	0.83	0.93
J2340.1-8510	0.93	0.90	0.98
J0820.9-5704	0.35	0.85	0.96
J1410.4-4246	0.71	0.95	0.95
J1240.2-4825	0.85	0.94	0.98
J0543.4-4430	0.63	0.58	0.97
J0626.3-5341	0.43	0.71	0.94
J0528.9-3927	0.92	0.93	0.98
J0449.9-4440	0.73	0.96	0.95
J0108.5-4020	0.94	0.93	0.98
J0314.3-4525	0.76	0.95	0.98
Abell_2895	0.96	0.89	0.99
J0216.3-4816	0.45	0.69	0.93
Abell_4038	0.81	0.91	0.97
Abell_548	0.52	0.76	0.90
Abell_13	0.34	0.87	0.91
Abell_2744	0.77	0.62	0.93
Average precision	0.67	0.83	0.95

Table 4.2: The precision results for the 21 galaxy clusters used for testing the Multi Layer Perceptron model.

4.0.7 Recall results 1

This section presents the recall results shown in table 4.3 for each of the 21 galaxy clusters. The lowest recall for class ‘C’ was 0.37 and the highest recall obtained was 0.93. The lowest recall obtained for class ‘M’ was 0.46 and the highest recall was 0.92. For class ‘S’ sources the lowest recall was 0.63 and the highest recall obtained was 0.99.

GALAXY CLUSTER	CLASS ‘C’	CLASS ‘M’	CLASS ‘S’
RXCJ0225.1-22928	0.67	0.59	0.63
MACSJ_0025.4-1222B	0.66	0.71	0.93
MACS_J0257.6-2209	0.76	0.72	0.93
PLCK_G200.9-28.2	0.37	0.46	0.83
MACS_J0257.6-2209B	0.67	0.73	0.89
J2340.1-8510	0.89	0.86	0.99
J0820.9-5704	0.74	0.85	0.81
J1410.4-4246	0.72	0.65	0.96
J1240.2-4825	0.88	0.81	0.98
J0543.4-4430	0.82	0.86	0.90
J0626.3-5341	0.57	0.87	0.89
J0528.9-3927	0.91	0.84	0.99
J0449.9-4440	0.81	0.79	0.94
J0108.5-4020	0.91	0.89	0.99
J0314.3-4525	0.90	0.89	0.94
Abell_2895	0.93	0.92	0.99
J0216.3-4816	0.67	0.81	0.83
Abell_4038	0.87	0.84	0.96
Abell_548	0.50	0.70	0.91
Abell_13	0.61	0.58	0.70
Abell_2744	0.59	0.76	0.96
Average recall	0.74	0.77	0.91

Table 4.3: The recall results for the 21 galaxy clusters used for testing the Multi Layer Perceptron model.

4.0.8 F1_score results 1

This section presents the F1_score results of the Multi Layer Perceptron. Table 4.4 shows the F1_score obtained for each of the 21 galaxy clusters used for the testing of the model. The highest F1_score obtained by the model for class ‘C’ was 0.94 and lowest was 0.32. For class ‘M’ the highest F1_score obtained was 0.92 and the lowest F1_score obtained was 0.53. For class ‘S’ the highest and lowest F1_score obtained by the model was 0.99 and 0.85 respectively.

GALAXY CLUSTER	CLASS ‘C’	CLASS ‘M’	CLASS ‘S’
RXCJ0225.1-22928	0.63	0.71	0.93
MACSJ_0025.4-1222B	0.65	0.70	0.93
MACS_J0257.6-2209	0.69	0.81	0.94
PLCK_G200.9-28.2	0.32	0.53	0.85
MACS_J0257.6-2209B	0.60	0.77	0.91
J2340.1-8510	0.91	0.88	0.98
J0820.9-5704	0.47	0.85	0.87
J1410.4-4246	0.71	0.77	0.95
J1240.2-4825	0.87	0.87	0.98
J0543.4-4430	0.71	0.70	0.94
J0626.3-5341	0.49	0.78	0.92
J0528.9-3927	0.91	0.88	0.98
J0449.9-4440	0.77	0.87	0.95
J0108.5-4020	0.93	0.90	0.99
J0314.3-4525	0.82	0.92	0.96
Abell_2895	0.94	0.90	0.99
J0216.3-4816	0.54	0.75	0.88
Abell_4038	0.84	0.88	0.97
Abell_548	0.51	0.73	0.91
Abell_13	0.44	0.70	0.85
Abell_2744	0.67	0.68	0.95
Average F1_score	0.67	0.79	0.93

Table 4.4: The F1_score results for the 21 galaxy clusters used for testing the model.

4.0.9 Confusion matrix results 2

This section presents the confusion matrix computations of the Multi Layer Perceptron on classifying radio emissions found in the galaxy cluster A3376. Each data set was represented by a frequency sub-band in which the galaxy cluster was observed in. In total there were data sets 1-8 corresponding to frequency sub-bands 1-8.

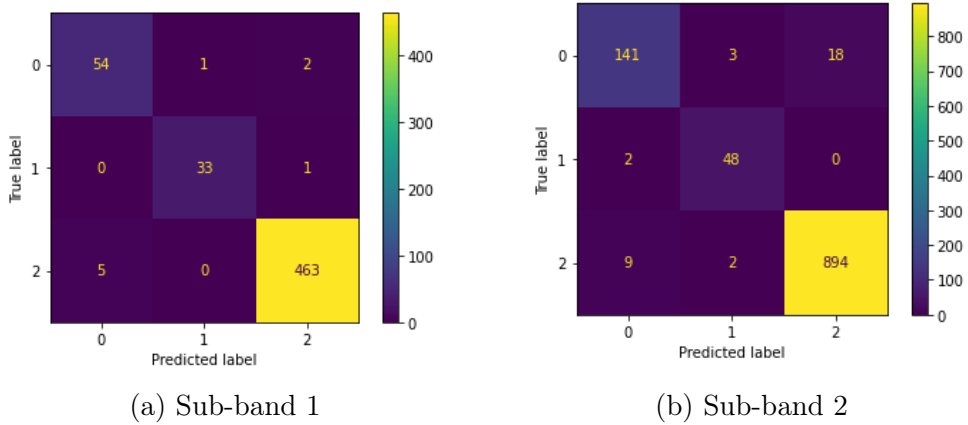


Figure 4.14: Confusion matrix for Sub-band 1 and Sub-band 2.

Figure 4.14 presents the confusion matrix results for Sub-band 1 and Sub-band 2. For Sub-band 1, 54 sources were correctly classified as belonging to class ‘C’, 33 sources were correctly classified as class ‘M’ sources and 463 sources correctly classified as belonging to class ‘S’. For Sub-band 2, 141 sources were correctly classified as belonging to class ‘C’, 48 sources were correctly classified as class ‘M’ and 894 source were correctly classified as belonging to class ‘S’.

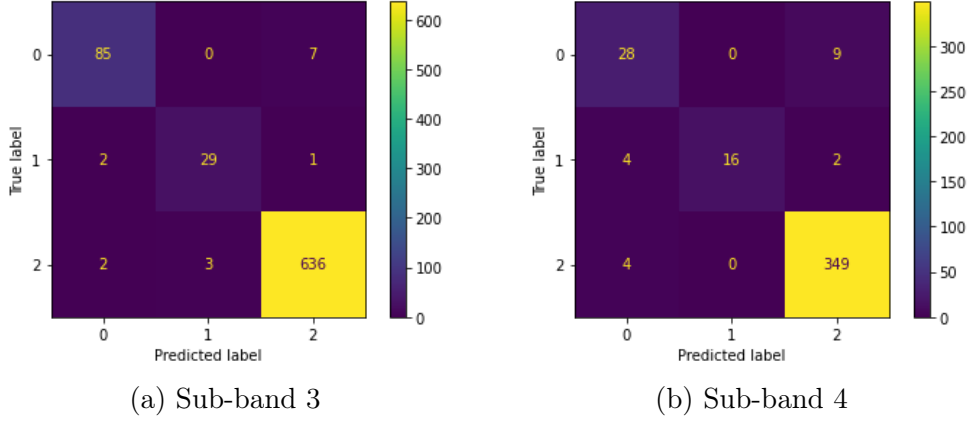


Figure 4.15: Confusion matrix for Sub-band 3 and Sub-band 4.

Figure 4.15 shows the confusion matrix results for Sub-band 3 and Sub-band 4. For Sub-band 3, 15 sources were incorrectly classified and for Sub-band 4, 19 sources were incorrectly classified. For correctly classified sources, 750 sources were correctly classified as either belonging to class ‘C’, class ‘M’ or ‘S’ and 393 sources were correctly classified for Sub-band 4.

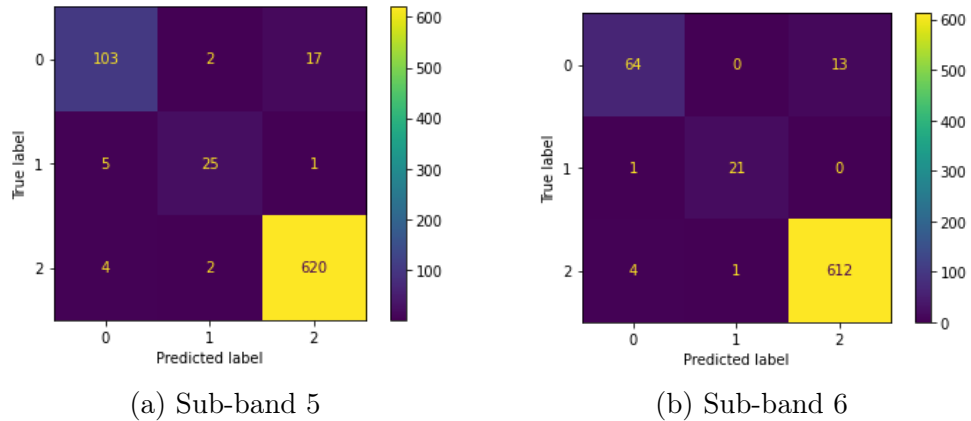


Figure 4.16: Confusion matrix for Sub-band 5 and Sub-band 6.

Figure 4.16 shows the confusion matrix for Sub-band 5 and Sub-band 6. Out of 779 sources, 748 sources were correctly classified and 31 sources were incorrectly classified

for Sub-band 5. Out of 716 sources, 697 sources were correctly classified for Sub-band 6.

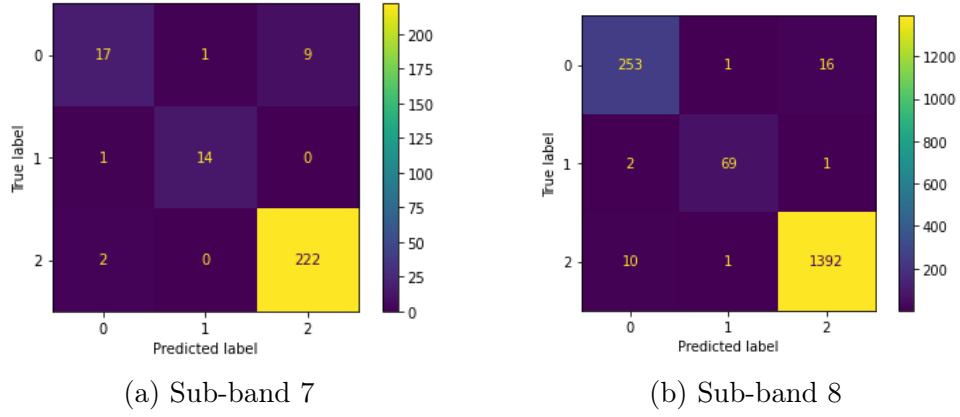


Figure 4.17: Confusion matrix for Sub-band 7 and Sub-band 8.

Figure 4.17 shows the confusion matrix for Sub-band 7 and Sub-band 8. For Sub-band 7, 263 sources were correctly classified and for Sub-band 8, 1714 sources were classified correctly.

4.0.10 Accuracy results 2

This section presents the accuracy results of the Multi Layer Perceptron in classifying sources in different frequency sub-bands.

FREQUENCY SUB-BAND	ACCURACY	TOTAL No. OF SOURCES
1	0.98	559
2	0.97	1117
3	0.98	765
4	0.95	412
5	0.96	779
6	0.97	716
7	0.95	266
8	0.98	1745
Average Accuracy	0.97	6359

Table 4.5: Accuracy results of the Multi Layer Perceptron in classifying sources in varied frequency sub-bands and the number of sources used for testing.

Table 4.5 shows the accuracy for each frequency sub-band and the number of sources used for testing the model. The highest accuracy obtained by the model was 0.98 and the lowest accuracy was 0.97.

4.0.11 Precision results 2

This section presents the precision results of the Multi Layer Perceptron in classifying sources in different frequency sub-bands. Table 4.6 shows the precision for each frequency sub-band used for testing the model. The highest precision obtained by the model for class ‘C’ was 0.96, for class ‘M’ was 1.00 and for class ‘S’ was 0.99.

FREQUENCY SUB-BAND	CLASS ‘C’	CLASS ‘M’	CLASS ‘S’
1	0.92	0.97	0.99
2	0.93	0.91	0.98
3	0.96	0.91	0.99
4	0.78	1.00	0.97
5	0.92	0.86	0.97
6	0.93	0.95	0.98
7	0.85	0.93	0.96
8	0.95	0.97	0.99
Average precision	0.91	0.94	0.98

Table 4.6: Precision results for the Multi Layer Perceptron in classifying sources in varied frequency sub-bands and the number of sources used for testing.

4.0.12 Recall results 2

This section presents the recall results of the Multi Layer Perceptron in classifying sources in the different frequency sub-bands. Table 4.7 shows the recall results for each frequency sub-band used for testing the model. The highest recall obtained by the model for class ‘C’ was 0.95, for class ‘M’ was 0.97 and for class ‘S’ was 0.99.

FREQUENCY SUB-BAND	CLASS ‘C’	CLASS ‘M’	CLASS ‘S’
1	0.95	0.97	0.99
2	0.87	0.96	0.99
3	0.92	0.91	0.99
4	0.76	0.73	0.99
5	0.84	0.81	0.99
6	0.83	0.95	0.99
7	0.63	0.93	0.99
8	0.94	0.96	0.99
Average recall	0.84	0.90	0.99

Table 4.7: Recall results for the Multi Layer Perceptron in classifying sources in the different frequency sub-bands and the number of sources used for testing.

4.0.13 F1_score results 2

This section presents the F1_score results of the Multi Layer Perceptron in classifying sources in the different frequency sub-bands. Table 4.8 shows the F1_score for each frequency sub-band used for testing the model. The highest F1_score obtained by the model for class ‘C’ was 0.95, for class ‘M’ was 0.97 and for class ‘S’ was 0.99.

FREQUENCY SUB-BAND	CLASS ‘C’	CLASS ‘M’	CLASS ‘M’
1	0.93	0.97	0.99
2	0.90	0.93	0.98
3	0.94	0.91	0.99
4	0.77	0.84	0.98
5	0.88	0.83	0.98
6	0.88	0.95	0.99
7	0.72	0.93	0.98
8	0.95	0.97	0.99
Average F1_score	0.87	0.92	0.99

Table 4.8: F1_score results for the Multi Layer Perceptron in classifying sources in the different frequency sub-bands and the number of sources used for testing.

4.0.14 Discussion

This section provides the discussion of the results on the performance of a Multi Layer Perceptron model in classifying radio emission sources associated with galaxy clusters. Firstly a discussion of its performance on the three classes used for classification followed by a discussion on the model performance on different frequency sub-bands of radio emission sources associated with the galaxy cluster A3376.

4.0.15 Confusion matrix

A confusion matrix gives a more detailed comparison report between the predicted values and the true values. The confusion matrix for each of the galaxy clusters used for testing the model indicate that the Multi Layer Perceptron had more correctly predicted sources as compared to incorrectly classified sources. The sum of values along the diagonal cells in the confusion matrix for most of the galaxy clusters was more than the sum of the remaining cells in the matrix which was all the incorrectly predicted values. An interesting trend shown by the confusion matrix from the 21 galaxy clusters

was how the model often confused class ‘C’ sources as class ‘S’ sources and vice versa. There was a significant number of misclassified sources among the two classes. Class ‘M’ had the lowest number of incorrectly classified sources in comparison. An example being from Figure 4.6 (a) which recorded 1032 incorrectly classified sources from class ‘S’ as class ‘C’ sources. Another example being Figure 4.4 (b) which shows 351 class ‘C’ sources incorrectly assigned to class ‘S’. This shows a strong correlation between the two classes that causes the model to incorrectly classify sources belonging to either of the two classes in such high numbers. The correlation may be attributed to the model failing to distinguish properly between class ‘S’ and class ‘C’ sources which could be that the two source classes have a significant similarity. In addition the confusion by the model may have been attributed by the Gaussian fitting method used by PyBDSF when breaking down interferometry images. PyBDSF defines class ‘C’ as sources with a Gaussian fitted in a field with another source which refers to double sources. This may be translated as two single sources belonging to class ‘S’. However it is unclear how the model determined it.

4.0.16 Model accuracy

The model accuracy was on average 89% correct in classifying the radio emission sources into either class ‘C’, class ‘M’ or class ‘S’. This accuracy value indicates that the model performed well in distinguishing between the three class sources and assigning them the proper class label in most of the instances. The accuracy was high in cases where there was a higher number of sources for the model testing as shown in Table 4.1 where the 97% accuracy was obtained for galaxy clusters with radio emission sources higher than 6000. Galaxy clusters with source populations between 3000 and 5000 obtained accuracy values in the region of 80%. There was an exception for the galaxy cluster PLCKG200.9-28.2 and Abell_13 with an accuracy of 75% and 76% respectively. While the higher the number of sources per galaxy cluster indicated that the model accuracy was higher this was not the case in other galaxy clusters. However it does indicate that there were more correctly classified radio emission sources than the incorrectly classified sources which is desirable in a classifier.

The model accuracy obtained for the classification of radio emission sources associated with the A3376 galaxy cluster was higher in comparison to the testing data set 1. In this case the model performance was evaluated to determine how it behaved when classifying radio emissions for a given galaxy cluster which was observed in different frequency sub-bands to find out if the model performance would improve or decline. The average accuracy for the model was 97% which was very high when compared to 89% accuracy obtained by the model previously. While a high accuracy may be desirable in a machine learning algorithm, overfitting may contribute to the high accuracy obtained by the model. Small data sets often contribute to the model overfitting. This was the case for some of the data sets used such as Sub-band 2 and Sub-band 4. There were even lower numbers for the class ‘C’ which had 76 and 57 sources for Sub-band 6 and Sub-band 7 respectively. This was for both the model training and testing. An increase in the class sample size would have reduced the model overfitting and provided a clearer picture.

Generally the accuracy shows a good model generalization even with a significant imbalance in terms of the class ‘M’ sources. This may have been contributed by the majority of sources belonging to class ‘C’ and class ‘S’ which generally scored high numbers in the model’s accuracy evaluation.

4.0.17 Precision

The model precision identified the ratio of true positives out of all positives which include both the true positives and false positives. The model was tested on approximately 120 000 radio emission sources from the 21 radio galaxy clusters. According to the evaluation of the Multi layer Perceptron model’s precision, it correctly identified 67% of the radio emission sources as belonging to class ‘C’ which is slightly above 50% and still a lower precision than desired. In some data sets the model obtained a precision higher than 67% such as 96% for the galaxy cluster Abell_2895, 94% for the galaxy cluster J0108.5-4020, 93% for the galaxy cluster J2340.1-8510 and 92% for the galaxy cluster J0528.9-3927. This shows the model performed well in some data sets and badly in some. Out of the 21 data sets used for training, the model obtained a precision higher than 50% with the remainder

having a precision of 45% and below. In comparison, the class ‘M’ precision was higher which indicated that on average 83% of the true positives were correctly identified. Class ‘S’ obtained the highest precision of 95% which is great since it is closer to 100% as the ideal precision value. The model performed very well in identifying the true positives for class ‘M’ and class ‘S’ but struggled in identifying the true class ‘C’ sources since it scored the lowest number of true positives on average.

In general, the model performed well since the precision for all three classes is above 50%. A precision of 50% is undesirable since it is similar to chance. This implies that regardless of the volume of data the model is used to classify sources, the number of true positives will always exceed the number of false positives except for cases in which sources belong to class ‘C’. The low precision for class ‘C’ indicates that the model would be less reliable in classifying the sources when given larger data sets. This defeats the purposes of eliminating doubt when classifying extended radio emission sources for any precision values lower than 50%. However identifying all the class ‘C’ sources which were incorrectly classified as belonging to either class ‘M’ or class ‘S’ helps identify the challenges the model encounters. This can be done by studying in detail the nature of those data points to explain why the model assigned them to the wrong class. This in turn gives insight on how the model could be improved when it comes to classifying class ‘C’ sources.

The precision score for class ‘M’ was 83% surpassing the precision of class ‘C’ sources. This indicates that the Multi Layer Perceptron achieved the objective of addressing the automated classification of extended radio emission sources which were rarely addressed by previous studies. In some cases an extended emission source is spread out with some parts residing a distance away. It becomes difficult to determine the true size of a source and it’s morphology. This might be caused by a source with several components. This can be easily mistook as single or double sources where as it is a part of a larger extended source. A multi component extended source could be identified or interpreted as multiple single component sources. In this case however the Multi Layer Perceptron scored high on the evaluation metric.

The precision of the model in classifying A3376 radio emission sources was lowest at 85% which was from the class ‘C’. For all the 8 Sub-band data sets, the precision score was higher. The model scored a precision of 100% in Sub-band 4 for class ‘M’ sources. This implies that the number of true class ‘M’ sources was high. The lowest precision the model scored for the class ‘M’ was 86%. For class ‘S’ the precision was above 90% showing that the majority of sources classified as class ‘S’ sources were class ‘S’ sources. Across all the three classes, the model precision demonstrated a high confidence level that the sources were assigned the correct class label. The precision however does not shed light on how many sources belonging to either of the three classes were incorrectly classified.

4.0.18 Recall

Recall determines how accurately the true positives were correctly classified by computing the portion of true positives out of all true positives and false negatives. If the recall of the model is high it indicates that less true positives were classified as negatives. The average recall obtained for class ‘C’ was 74% with the lowest recall being 37% from the galaxy cluster PLCKG200.9-28.2 which was also the only recall lower than 50%. For class ‘M’ and class ‘S’ the galaxy cluster PLCKG200.9-28.2 obtained the lowest recall. This was also the case for the accuracy, precision and the F1_score which will be discussed hereafter. It presents an interesting study to determine why this particular galaxy cluster scored very low values from evaluation metrics used. The confusion matrix for this particular galaxy cluster indicated that there was a significantly high number of class ‘S’ sources that were incorrectly classified as class ‘C’ sources. A total of 540 class ‘S’ sources which are compact, single Gaussian sources were assigned the class ‘C’ label which are extended multi Gaussian sources. In addition the model incorrectly classified 351 class ‘C’ sources as class ‘S’ sources which was the opposite of the previously mentioned 540 sources. While previously the model’s poor performance on class ‘C’ and class ‘S’ was caused partly by the data set imbalance, this does not adequately provide an explanation on how a significant number of sources were misclassified especially for two classes that are different with class ‘C’ being an intermediary class between the two. This misclassification also may have contributed to lower accuracy, precision and F1_score

obtained by the model.

The average recall for class ‘M’ sources was 77% which was slightly higher than the recall for class ‘C’ sources. The highest recall was 92% but the model recall could be improved to increase the number of correctly classified sources for the class label. For class ‘S’ the average recall was 91% with the highest recall obtained for a galaxy cluster being 99%. This is the ideal recall that indicates that almost all sources classified as class ‘S’ sources are indeed class ‘S’ sources. The recall for the classification of radio emissions associated with the galaxy cluster A3376 was 84%, 90% and 99% for class ‘C’, class ‘M’ and class ‘S’ respectively. Overall the Multi Layer Perceptron scored a good recall for all three class labels despite the recall for class ‘C’ being the lowest.

In view of the fact that the Multi Layer Perceptron model should be able to correctly classify radio emission sources by assigning the proper class labels, recall gives a clearer picture of how well the model does that. From the results of the recall it shows that the model did a good job of that. The higher the recall the more correctly classified emissions are. Given the performance of the model on class ‘C’ source classification the model correctly identified 74% of class ‘C’ sources from the 21 galaxy clusters. While 84% of class ‘C’ sources from the galaxy cluster A3376 were correctly classified.

4.0.19 F1_score

The model obtained an F1_score of 67% for the classification of class ‘C’ sources which was the lowest F1_score between the three classes. The F1_score for class ‘M’ sources was 79% and for class ‘S’ was 93%. The model performed well in predicting class ‘M’ and class ‘S’ sources. While the F1_score for class ‘C’ sources was lower it was well above 50%. The galaxy cluster PLCKG200.9-28.2 scored the lowest F1_score which was 32% indicating a significantly low performance by the model. Despite this the model obtained higher scores for some of the class ‘C’ sources in different galaxy clusters, some as high as 93% and 94% which is a good indicator that the model correctly predicts the sources belonging to class ‘C’. For class ‘M’ and class ‘S’, the F1_score values indicate that the model faces minimal challenges in predicting sources belonging

to the two class labels with the lowest score being 53% from class ‘M’ predictions. The F1_score for the classification of radio emission sources associated with the galaxy cluster A3376 was above 80% for all three class labels. This demonstrated that the model performed well in predicting the class labels the sources belonged to. There was a difference of 5% score between class ‘C’ and class ‘M’.

4.0.20 Number of sources in a field on the model performance

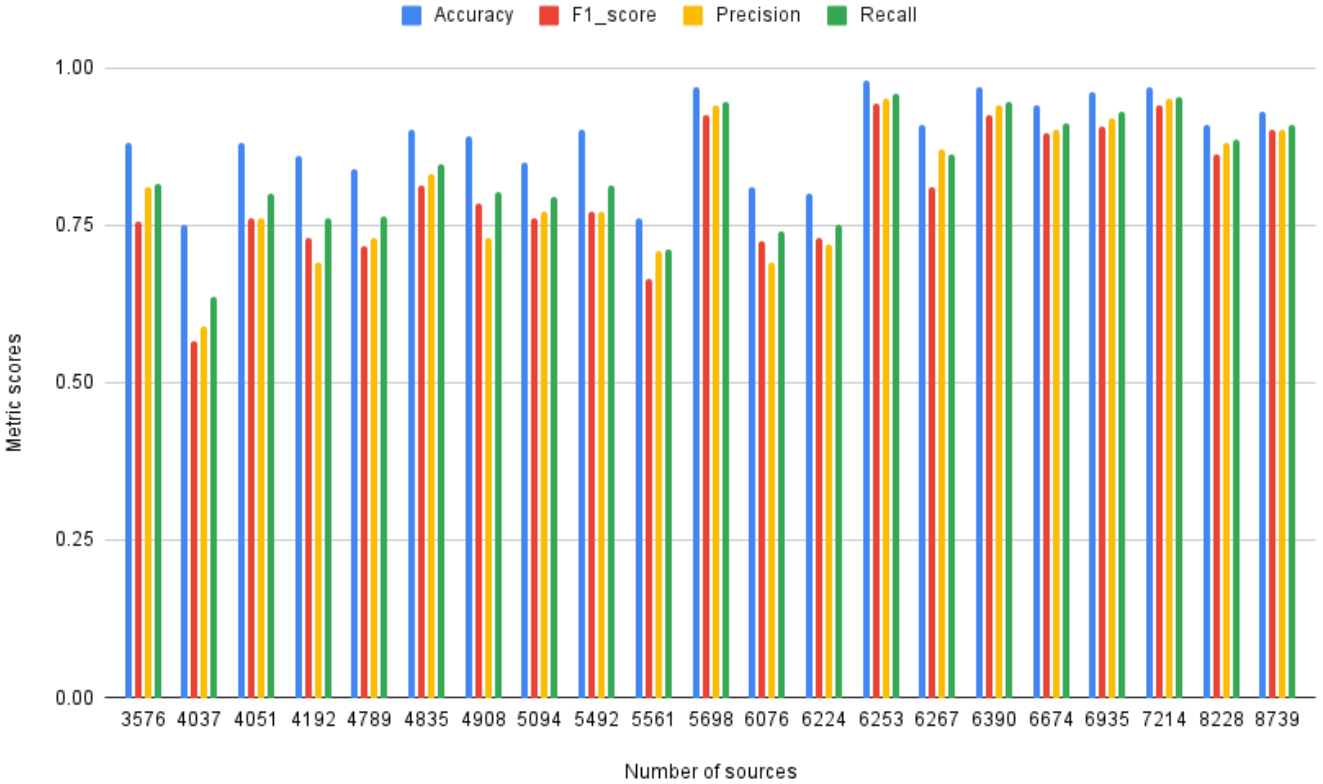


Figure 4.18: A graph showing the effect of the number of sources in a field on the model performance using the four metrics

Figure 4.18 depicts the model scores for each of the four evaluations metrics in relation to the number of sources per field. To explore the dependence of the accuracy, F1_score, precision and recall on the source population for each of the 21 galaxies used for training. The graph indicates that accuracy had the highest metric score with the model F1_score scoring the lowest. The metric scores increased with the increase in the number of sources per field indicating correlation between the number of sources and the score for each metric. The graph also indicated fluctuations in the score for

some fields where scores were low instead of the opposite such as fields with 6076 and 6224 sources obtained lower scores when the expectation was higher scores for the four metrics. Metric scores were generally higher for higher number of sources indicating that the performance of the model depended on the number of sources.

4.0.21 Chapter summary

This chapter presented and discussed the results of the performance evaluation of a Multi Layer Perceptron on a three class classification task. The results, shown by the accuracy, precision, recall and the F1_score indicated that the model scored high in classifying sources that belonged to either class ‘M’ or class ‘S’ indicative of the precision, recall and F1_score. The model however obtained the lowest values for class ‘C’ in the aforementioned metrics indicating that the model struggled with predicting sources belonging to class ‘C’. The results also revealed an interesting trend shown by the correlation between class ‘C’ and class ‘S’ sources that scored high values for incorrectly classified sources between the two classes. In addition there is evidence that the source finding tool contributed to the model performance just as the hyperparameters source population distribution.

Chapter 5

Conclusion and recommendation

5.0.1 Introduction

This chapter provides an overview of the thesis and what was achieved successfully, then presents the conclusion of the study and recommendations for further studies.

5.0.2 Thesis overview

The emergence of next generation radio telescopes has necessitated automated approaches to radio data analysis. One of which includes the detection of radio sources in survey data and the subsequent classification of the sources identified. There exists a number of studies aimed at resolving source identification and classification challenges for more automated processes. The majority of which are based on implementing a Convolutional Neural Network for image data which is an exhaustive approach computation wise. However previous studies have failed to address the classification of extended radio sources which substitute a portion of radio sources albeit at low numbers. As an alternative, a Multi Layer Perceptron can be used to classify these radio sources given its simplicity in implementation and its use of the back-propagation for weight and bias adjustments during training while results in a fast model. Moreover it addressed the storage challenge by large surveys such as the SKA by using the MLP as a preliminary data eliminator.

The main objective of this thesis was to implement a Multi Layer Perceptron model for the classification of compact and extended radio emission sources associated with

radio galaxy clusters. In addition, assessing its predictive performance in which there were several steps. This included the data collection and preparation for the training and testing of the model, the design and implementation of a Multi Layer Perceptron algorithm, the training and testing of the model and evaluating its performance using performance evaluation metrics that include accuracy precision, recall and F1_score. This was done to determine the suitability of a Multi Layer Perceptron for the classification of radio emission sources given already existing approaches. The properties of a Multi Layer Perceptron afforded an opportunity to study its performance when applied to a classification task using radio emission data which was also a novel approach to automating the classification process.

5.0.3 Objectives overview

This section of the chapter provides an overview of the specific objectives for this research, address if they were achieved and how they were achieved which all contributed to the completion of this research. The main objective of this research was to implement a multi layer perceptron predictor model for radio emission sources associated with radio galaxy clusters. Thereafter establish its likelihood as a classifier. This was done by evaluating its performance and determining factors that affect the performance of the network. The specific objectives included to:

1. Collect data to be used in the Multi Layer Perceptron model. This objective was successfully achieved. This was achieved through the acquisition of data from the MeerKAT. This was used for training and testing of the model. There were two data sets used. The first data set included 104 files of radio galaxy clusters each containing thousands of radio emission sources per file. This data set was to be used to train and test the model on a classification task of compact and extended sources. The second data set included 8 files of radio emissions sources associated with the radio galaxy A3376 that was to be used to assess whether data with different frequency sub-bands had any effect on the performance of the Multi Layer Perceptron model.
2. Prepare the data for training and testing of the Multi Layer Perceptron model. This objective was achieved successfully. The data preparation process involved

data cleaning to eliminate any errors in the data such as identifying and eliminating null values and redundant features, taking count of value counts of the class distributions. The data was also transformed to eliminate a bias towards larger values. This was done by normalizing the data so that it ranged from 0 to 1 in vector form. The output labels were encoded into values ranging from 0 to 2 to represent the three classification labels. The second data set included the use of PyBDSF to create a numerical catalogue of radio emission sources associated with A3376. This was then followed by the same data preparation from the first data set.

3. Design and implement a Multi Layer Perceptron algorithm for the classification of radio emission sources galaxy clusters associated with radio galaxy clusters. This objective was successfully achieved. The Scikit-learn library was used for designing and implementing the Multi Layer Perceptron network by defining the hyperparameters which defined the structure of the model prior to training.
4. Train and test the model using the data obtained. This objective was achieved successfully. The model was trained on 80% of the data and then tested on the remaining 20% for both data sets that were prepared previously. Convergence was reached during training and testing included using the model to predict class labels on unseen data.
5. Evaluate the model performance, factors affecting performance and determining any parameter tuning necessary. This objective was achieved successfully. The performance evaluation metrics used included the confusion matrix, accuracy, precision, recall and the F1_score. The metrics showed that the correlation between two class which affected the performance of the model. The MLP scored high in the performance metrics used for classifying compact and extended sources. It however struggled with the intermediary class. Besides the hyperparameters the source finder used affected the model's performance.

5.0.4 Hypothesis overview

For this research there existed a hypothesis and a null hypothesis.

Hypothesis (H_1) - A multi layer perceptron neural network performs significantly well in the task of supervised classification of radio emission sources associated with galaxy and can be a good predictor model obtaining accuracy greater than 80%..

Null hypothesis (H_0) - Astronomical data is significantly different from other data types that have been largely used to assess the performance of a multi layer perceptron thus does not perform well in the task of classifying radio sources nor will it be a good predictor model obtaining accuracy less than 80%.

The hypothesis (H_1) is accepted and the hypothesis (H_0) is rejected because the MLP scored an accuracy of 89% which exceeds the 80%.

5.0.5 Questions overview

There were several question formulated for this research. These included:

1. How does a multi layer perceptron perform at a classification task of galaxy cluster radio emission sources? The accuracy, precision, recall and the F1_score demonstrated that a Multi Layer Perceptron performs well in the classification task of radio emission sources. The model scored high metric values for the classification of compact and for extended sources. Taking this into account, the model achieved the goal of classifying radio emission sources.
2. What hyperparameters can be modified in the network which affect the performance of the model? Combinations of hyperparameters values provided different output results. The Stochastic Gradient Descent (SGD) provided better results as the optimizer with the learning rate being constant over invscaling or adaptive. Other hyperparameters that affected the performance included the maximum number of iterations, activation function, and learning rate initialization among others.
4. Does the frequency sub-band used in different data sets have any effect on the overall performance of the classifier? Frequency sub-bands affected the performance of the model. The model accuracy, precision, recall and F1_score was

higher for the classification of sources associated with A3376 when compared to the classification of sources from various galaxy clusters although the model showed signs of overfitting.

5. Is a Multi Layer Perceptron suitable for galaxy cluster radio emission sources classification? The results from this research support the use of a Multi Layer Perceptron as a classification model of radio emission sources.
6. Can the Multi Layer Perceptron model be implemented as a predictor model for real life detection and classification of galaxy cluster sources? The Multi Layer Perceptron model is a good predictor model. It scored high on the evaluation metrics, it discovered the correlation between two of the classes which in reality would always identify the relationship between real sources and their possible classes.

5.0.6 Conclusion

An important part of solving a problem involves the evaluation of existing solutions and how they can be improved and the evaluation of potential solutions and how they could assist in solving the problem. In this case, next generation telescopes will produce vast volumes of data from observations. This presents a data analysis challenge of combing through the data to select what is important and discard anything of insignificance. The detection and classification of sources is important to achieve the scientific goals set out to provide answers about our universe. Existing approaches have concentrated on solving one of the many data challenges with varying degrees of success by using existing solutions, novel solutions or a hybrid of the two. In the area of radio astronomy source detection and classification solutions are still at an infant stage and the hunt for efficient solutions is still ongoing. This research presented an alternative solution to source detection and classification by taking into account simplicity, ease and robustness. This resulted in a Multi Layer Perceptron model capable of classifying radio emission sources associated with galaxy clusters and also capable of understanding the relationship between source attributes and the potential class labels. This was evident in the performance evaluation metrics scores.

Ideally the closer the precision and recall to one the better model. However there is a trade off between the model's precision and recall. Increasing the precision decreases the recall and maximising the recall minimises the model precision. While precision assesses the magnitude of the error that results from the false positives, recall assesses the magnitude of the error that results from the false negatives. In this case the recovery of all true positives is vital, that is all sources the model predicted as belonging to either of the three class labels should be true sources and all of them should be recovered by the model. This helps in producing a complete catalogue of sources where there are more true positives than false positives. Thus for each prediction the model makes, there should be more true positives. The precision for Class 'M' and class 'S' demonstrate that the Multi Layer Perceptron predicted 83% and 95% of the sources as true positives for each of the classes respectively. This is a good indication of the model's prediction performance in classifying the radio emission sources associated with galaxy clusters. The precision for class 'C' sources however shows the model struggles with correctly predicting sources belonging to this class indicating that less sources are correctly classified than desired by the model.

However this does not mean the recall is any less significant for this case since it is just as important to ensure that there are less false positives and more true positives. With recall demonstrating the same trend as precision where class 'S' precision outperforms the other two classes and precision for class 'C' scoring the lowest. The measure of a model's precision and recall is indicative of the successful predictive outcome when the data is unbalanced such as the case for the data sets used in this research. In this case the model performed well given the unbalanced data sets used. Unlike precision and recall, F1_score takes into consideration both the false negatives and false positives by computing the average of recall and precision in this case by computing the harmonic mean. The F1_score shows that the model performs well in predicting class 'S' sources followed by class 'M' source. while the score for class 'C' was the lowest, the model performs better than chance in the prediction of class 'C' sources. Overall the F1_score for the Multi Layer Perceptron was good which indicates that it is suitable for classifying radio emission sources associated with radio galaxy clusters. Thus proving the suitability of a Multi Layer perceptron as a radio emission classifier and data

eliminator.

5.0.7 Recommendations

The approach used in this study relied on the application of a source finding tool to create the necessary numerical data sets that were used for the training and testing on the model. This presented a limitation in the data processing. Source finding tools themselves carry a certain degree of error during computations which is carried forward into the training and testing stages. For example Gaussian fitting for single compact sources by PyBDSF is a smooth process but difficult for complex and extended sources since it involves fitting several Gaussians of varying sizes and then combining them to make one single source. This may account for the model's confusion when predicting class 'C' and class 'S' sources. In which the model confuses a single Gaussian component that is part of an extended emission source as a single compact source found in an island. An alternative approach to the recovery of extended emission sources that does not include the Gaussian fitting technique could eliminate this challenge and reduce the number of incorrectly classified sources for the two aforementioned classes.

An alternative approach to improve the model's performance when classifying radio emission sources, in particular distinguishing between class 'C' and the other two classes could be to implement a Multi Layer Perceptron fusion classifier. This can be done by implementing three models. One to distinguish between class 'C' and class 'M'. A second model to distinguish between class 'C' and class 'S'. Lastly a third model to distinguish between class 'S' and class 'M' followed by a majority voting classifier which made the final class designation.

Bibliography

- [1] Swapan Kumar Haldar. “Chapter 3 - Photogeology, Remote Sensing, and Geographic Information System in Mineral Exploration”. In: *Mineral Exploration (Second Edition)*. Ed. by Swapan Kumar Haldar. Second Edition. Elsevier, 2018, pp. 47–68. ISBN: 978-0-12-814022-2. DOI: <https://doi.org/10.1016/B978-0-12-814022-2.00003-4>. URL: <https://www.sciencedirect.com/science/article/pii/B9780128140222000034>.
- [2] National Research Council et al. *Handbook of Frequency Allocations and Spectrum Protection for Scientific Uses*. National Academies Press, 2007.
- [3] A. R. Thompson et al. “The Very Large Array.” In: 44 (Oct. 1980), pp. 151–167. DOI: 10.1086/190688.
- [4] Ron Ekers. *The History of the Square Kilometre Array (SKA) - Born Global*. 2012. arXiv: 1212.3497 [astro-ph.IM].
- [5] Peter E. Dewdney et al. “The Square Kilometre Array”. In: *Proceedings of the IEEE* 97.8 (2009), pp. 1482–1496. DOI: 10.1109/JPROC.2009.2021005.
- [6] Ray P. Norris. “Data Challenges for Next-generation Radio Telescopes”. In: *2010 Sixth IEEE International Conference on e-Science Workshops* (2010). DOI: 10.1109/esciencew.2010.13. URL: <http://dx.doi.org/10.1109/eScienceW.2010.13>.
- [7] S. Johnston et al. “Science with the Australian Square Kilometre Array Pathfinder”. In: 24.4 (Dec. 2007), pp. 174–188. DOI: 10.1071/AS07033. arXiv: 0711.2103 [astro-ph].
- [8] J. Jonas and MeerKAT Team. “The MeerKAT Radio Telescope”. In: *MeerKAT Science: On the Pathway to the SKA*. Jan. 2016, 1, p. 1.

- [9] C.J. Lonsdale et al. “The Murchison Widefield Array: Design Overview”. In: *Proceedings of the IEEE* 97.8 (2009), 1497–1506. ISSN: 1558-2256. DOI: 10.1109/jproc.2009.2017564. URL: <http://dx.doi.org/10.1109/JPROC.2009.2017564>.
- [10] Marc Postman and P Murdin. “Distribution of galaxies, clusters, and superclusters”. In: *Encyclopedia of Astronomy and Astrophysics* (2001), p. 1637.
- [11] R. G. Carlberg, H. K. C. Yee, and E. Ellingson. “The Average Mass and Light Profiles of Galaxy Clusters”. In: *The Astrophysical Journal* 478.2 (1997), 462–475. ISSN: 1538-4357. DOI: 10.1086/303805. URL: <http://dx.doi.org/10.1086/303805>.
- [12] G. W. Pratt et al. “The Galaxy Cluster Mass Scale and Its Impact on Cosmological Constraints from the Cluster Population”. In: *Space Science Reviews* 215.2 (2019). ISSN: 1572-9672. DOI: 10.1007/s11214-019-0591-0. URL: <http://dx.doi.org/10.1007/s11214-019-0591-0>.
- [13] Elisabete M. de Gouveia Dal Pino. “Cosmic Magnetic Fields”. In: *AIP Conference Proceedings* (2006). ISSN: 0094-243X. DOI: 10.1063/1.2405951. URL: <http://dx.doi.org/10.1063/1.2405951>.
- [14] Valentina Vacca et al. “Magnetic Fields in Galaxy Clusters and in the Large-Scale Structure of the Universe”. In: *Galaxies* 6.4 (2018). ISSN: 2075-4434. DOI: 10.3390/galaxies6040142. URL: <https://www.mdpi.com/2075-4434/6/4/142>.
- [15] MJ Rees. “Origin of cosmic magnetic fields”. In: *Astronomische Nachrichten: Astronomical Notes* 327.5-6 (2006), pp. 395–398.
- [16] Karsten Jedamzik, Višnja Katalinić, and Angela V Olinto. “Damping of cosmic magnetic fields”. In: *Physical Review D* 57.6 (1998), p. 3264.
- [17] Dongsu Ryu, Hyesung Kang, and Peter L Biermann. “Cosmic magnetic fields in large scale filaments and sheets”. In: *arXiv preprint astro-ph/9803275* (1998).
- [18] Ray P. Norris et al. “EMU: Evolutionary Map of the Universe”. In: *Publications of the Astronomical Society of Australia* 28.3 (2011), 215–248. ISSN: 1448-6083. DOI: 10.1071/as11021. URL: <http://dx.doi.org/10.1071/AS11021>.

- [19] Paolo Padovani. “The faint radio sky: radio astronomy becomes mainstream”. In: *The Astronomy and Astrophysics Review* 24.1 (2016). ISSN: 1432-0754. DOI: 10.1007/s00159-016-0098-6. URL: <http://dx.doi.org/10.1007/s00159-016-0098-6>.
- [20] Priyaa Thavasimani and Anna Scaife. *Square Kilometre Array : Processing Voluminous MeerKAT Data on IRIS*. 2021. arXiv: 2105.14613 [astro-ph.IM].
- [21] A. K. Aniyani and K. Thorat. “Classifying Radio Galaxies with the Convolutional Neural Network”. In: *The Astrophysical Journal Supplement Series* 230.2 (2017), p. 20. ISSN: 1538-4365. DOI: 10.3847/1538-4365/aa7333. URL: <http://dx.doi.org/10.3847/1538-4365/aa7333>.
- [22] V Lukic et al. “Radio Galaxy Zoo: compact and extended radio source classification with deep learning”. In: *Monthly Notices of the Royal Astronomical Society* 476.1 (2018), 246–260. ISSN: 1365-2966. DOI: 10.1093/mnras/sty163. URL: <http://dx.doi.org/10.1093/mnras/sty163>.
- [23] Wathela Alhassan, AR Taylor, and Mattia Vaccari. “The FIRST Classifier: compact and extended radio galaxy classification using deep Convolutional Neural Networks”. In: *Monthly Notices of the Royal Astronomical Society* 480.2 (2018), pp. 2085–2093.
- [24] Avron Barr and Emma Feigenbaum. “A Handbook of Artificial Intelligence”. In: *Computer Music Journal* 6 (Dec. 1982). DOI: 10.2307/3680201.
- [25] A. L. Samuel. “Some studies in machine learning using the game of checkers”. In: *IBM Journal of Research and Development* 44.1.2 (2000), pp. 206–226. DOI: 10.1147/rd.441.0206.
- [26] A. L. Samuel. “Some Studies in Machine Learning Using the Game of Checkers. II—Recent Progress”. In: *IBM Journal of Research and Development* 11.6 (1967), pp. 601–617. DOI: 10.1147/rd.116.0601.
- [27] Pádraig Cunningham, Matthieu Cord, and Sarah Jane Delany. “Supervised learning”. In: *Machine learning techniques for multimedia*. Springer, 2008, pp. 21–49.

- [28] Bing Liu. “Supervised learning”. In: *Web data mining*. Springer, 2011, pp. 63–132.
- [29] M Senthilkumar. “Use of artificial neural networks (ANNs) in colour measurement”. In: *Colour Measurement*. Elsevier, 2010, pp. 125–146.
- [30] BK Lavine and TR Blank. “Feed-forward neural networks”. In: (2009).
- [31] Seppo Linnainmaa. “Taylor expansion of the accumulated rounding error”. In: *BIT Numerical Mathematics* 16.2 (1976), pp. 146–160.
- [32] Stuart Dreyfus. “The computational solution of optimal control problems with time lag”. In: *IEEE Transactions on Automatic Control* 18.4 (1973), pp. 383–385.
- [33] S. Andreon et al. “Wide field imaging - I. Applications of neural networks to object detection and star/galaxy classification”. In: *Monthly Notices of the Royal Astronomical Society* 319.3 (2002), 700–716. ISSN: 1365-2966. DOI: 10.1046/j.1365-8711.2000.03700.x. URL: <http://dx.doi.org/10.1046/j.1365-8711.2000.03700.x>.
- [34] SC Odewahn et al. “Automated star/galaxy discrimination with neural networks”. In: *Digitised Optical Sky Surveys*. Springer, 1992, pp. 215–224.
- [35] AW Moore et al. *Proc. MPA/ESO/MPE Workshop, Mining the Sky*. 2001.
- [36] Philip Graff et al. “SKYNET: an efficient and robust neural network training tool for machine learning in astronomy”. In: *Monthly Notices of the Royal Astronomical Society* 441.2 (2014), pp. 1741–1759.
- [37] Raffaele D’Abrusco et al. “Mining the SDSS Archive. I. Photometric Redshifts in the Nearby Universe”. In: *The Astrophysical Journal* 663.2 (2007), 752–764. ISSN: 1538-4357. DOI: 10.1086/518020. URL: <http://dx.doi.org/10.1086/518020>.
- [38] Stefano Andreon et al. “Neural nets and star/galaxy separation in wide field astronomical images”. In: *In Proceedings of International Joint Conference on Neural Networks*. 1999, pp. 3810–3815.

- [39] Chen Wu et al. “Radio Galaxy Zoo: CLARAN—a deep learning classifier for radio morphologies”. In: *Monthly Notices of the Royal Astronomical Society* 482.1 (2019), pp. 1211–1230.
- [40] Zhixian Ma et al. “A Machine Learning Based Morphological Classification of 14,245 Radio AGNs Selected from the Best–Heckman Sample”. In: *The Astrophysical Journal Supplement Series* 240.2 (2019), p. 34.
- [41] A Vafaei Sadr et al. “DEEPSOURCE: point source detection using deep learning”. In: *Monthly Notices of the Royal Astronomical Society* 484.2 (2019), pp. 2793–2806.
- [42] Richard M. Feder et al. “Multiband Probabilistic Cataloging: A Joint Fitting Approach to Point-source Detection and Deblending”. In: *The Astronomical Journal* 159.4 (2020), p. 163. ISSN: 1538-3881. DOI: 10.3847/1538-3881/ab74cf. URL: <http://dx.doi.org/10.3847/1538-3881/ab74cf>.
- [43] A Bonaldi et al. “Square Kilometre Array Science Data Challenge 1: analysis and results”. In: *Monthly Notices of the Royal Astronomical Society* 500.3 (2021), pp. 3821–3837.
- [44] V Lukic et al. “Morphological classification of radio galaxies: capsule networks versus convolutional neural networks”. In: *Monthly Notices of the Royal Astronomical Society* 487.2 (2019), pp. 1729–1744.
- [45] V. Lukic, F. De Gasperin, and M. Brüggen. *ConvoSource: Radio-Astronomical Source-Finding with Convolutional Neural Networks*. 2019. arXiv: 1910.03631 [astro-ph.IM].
- [46] Chen Wu et al. “Radio Galaxy Zoo:Claran– a deep learning classifier for radio morphologies”. In: *Monthly Notices of the Royal Astronomical Society* 482.1 (2018), 1211–1230. ISSN: 1365-2966. DOI: 10.1093/mnras/sty2646. URL: <http://dx.doi.org/10.1093/mnras/sty2646>.
- [47] A. A. Miller. “THE SYNTHETIC-OVERSAMPLING METHOD: USING PHOTOMETRIC COLORS TO DISCOVER EXTREMELY METAL-POOR STARS”. In: *The Astrophysical Journal* 811.1 (2015), p. 30. ISSN: 1538-4357. DOI: 10.1088/0004-637x/811/1/30. URL: <http://dx.doi.org/10.1088/0004-637x/811/1/30>.

- [48] Jenni AM Sidey-Gibbons and Chris J Sidey-Gibbons. “Machine learning in medicine: a practical introduction”. In: *BMC medical research methodology* 19.1 (2019), pp. 1–18.
- [49] Mohammad Amin Fahami et al. “Detection of effective genes in colon cancer: A machine learning approach”. In: *Informatics in Medicine Unlocked* (2021), p. 100605.
- [50] Eloise Withnell et al. “XOmivAE: an interpretable deep learning model for cancer classification using high-dimensional omics data”. In: *Briefings in Bioinformatics* (2021). ISSN: 1477-4054. DOI: 10.1093/bib/bbab315. URL: <http://dx.doi.org/10.1093/bib/bbab315>.
- [51] Mehdi Hosseinzadeh et al. “A multiple multilayer perceptron neural network with an adaptive learning algorithm for thyroid disease diagnosis in the internet of medical things”. In: *The Journal of Supercomputing* 77.4 (2021), pp. 3616–3637.
- [52] Marko Mataija and Dorijan Sablić-Nemec. “Brain tumor detection based on MRI images using multilayer perceptron”. In: *Ri-STEM-2021* (2021), p. 95.
- [53] Sakshi Bhoyar et al. “Real-time Heart Disease Prediction System using Multilayer Perceptron”. In: *2021 2nd International Conference for Emerging Technology (INCET)*. IEEE. 2021, pp. 1–4.
- [54] Sarfraz Nawaz Brohi et al. “A Multilayer Perceptron Model for the Classification of Breast Cancer Cells”. In: *International Journal of Computing and Digital System* (2021).
- [55] Woo Seok Lee, Junghyo Jo, and Taegeun Song. “Machine learning for the diagnosis of early-stage diabetes using temporal glucose profiles”. In: *Journal of the Korean Physical Society* 78.5 (2021), 373–378. ISSN: 1976-8524. DOI: 10.1007/s40042-021-00056-8. URL: <http://dx.doi.org/10.1007/s40042-021-00056-8>.
- [56] Stephen K. N. Portillo et al. “Improved Point-source Detection in Crowded Fields Using Probabilistic Cataloging”. In: *The Astronomical Journal* 154.4 (2017), p. 132. ISSN: 1538-3881. DOI: 10.3847/1538-3881/aa8565. URL: <http://dx.doi.org/10.3847/1538-3881/aa8565>.

- [57] David E Jones, Vinay L Kashyap, and David A Van Dyk. “Disentangling overlapping astronomical sources using spatial and spectral information”. In: *The Astrophysical Journal* 808.2 (2015), p. 137.
- [58] Tansu Daylan, Stephen K. N. Portillo, and Douglas P. Finkbeiner. “Inference of Unresolved Point Sources at High Galactic Latitudes Using Probabilistic Catalogs”. In: *The Astrophysical Journal* 839.1 (2017), p. 4. ISSN: 1538-4357. DOI: 10.3847/1538-4357/aa679e. URL: <http://dx.doi.org/10.3847/1538-4357/aa679e>.
- [59] Brendon J. Brewer, Daniel Foreman-Mackey, and David W. Hogg. “PROBABILISTIC CATALOGS FOR CROWDED STELLAR FIELDS”. In: *The Astronomical Journal* 146.1 (2013), p. 7. ISSN: 1538-3881. DOI: 10.1088/0004-6256/146/1/7. URL: <http://dx.doi.org/10.1088/0004-6256/146/1/7>.
- [60] James O. Chibueze et al. “Jets from MRC 0600-399 bent by magnetic fields in the cluster Abell 3376”. In: *Nature* 593.7857 (2021), 47–50. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03434-1. URL: <http://dx.doi.org/10.1038/s41586-021-03434-1>.
- [61] Dharam V. Lal. “NGC 4869 in the Coma Cluster: Twist, Wrap, Overlap, and Bend”. In: *The Astronomical Journal* 160.4 (2020), p. 161. ISSN: 1538-3881. DOI: 10.3847/1538-3881/abacd1. URL: <http://dx.doi.org/10.3847/1538-3881/abacd1>.
- [62] Chris Nolting et al. “Interactions between radio galaxies and cluster shocks. I. Jet axes aligned with shock normals”. In: *The Astrophysical Journal* 876.2 (2019), p. 154.
- [63] C Gheller, F Vazza, and A Bonafede. “Deep learning based detection of cosmological diffuse radio sources”. In: *Monthly Notices of the Royal Astronomical Society* 480.3 (Aug. 2018), pp. 3749–3761. ISSN: 0035-8711. DOI: 10.1093/mnras/sty2102. eprint: <https://academic.oup.com/mnras/article-pdf/480/3/3749/25519514/sty2102.pdf>. URL: <https://doi.org/10.1093/mnras/sty2102>.

- [64] Noam I. Libeskind et al. “Tracing the cosmic web”. In: *Monthly Notices of the Royal Astronomical Society* 473.1 (2017), 1195–1217. ISSN: 1365-2966. DOI: 10.1093/mnras/stx1976. URL: <http://dx.doi.org/10.1093/mnras/stx1976>.
- [65] Wenxiao Xu et al. “Galaxy properties in the cosmic web of EAGLE simulation”. In: *Monthly Notices of the Royal Astronomical Society* 498.2 (2020), 1839–1851. ISSN: 1365-2966. DOI: 10.1093/mnras/staa2497. URL: <http://dx.doi.org/10.1093/mnras/staa2497>.
- [66] Volker Springel et al. “Simulations of the formation, evolution and clustering of galaxies and quasars”. In: *Nature* 435.7042 (2005), 629–636. ISSN: 1476-4687. DOI: 10.1038/nature03597. URL: <http://dx.doi.org/10.1038/nature03597>.
- [67] Andres C. Rodríguez et al. “Fast cosmic web simulations with generative adversarial networks”. In: *Computational Astrophysics and Cosmology* 5.1 (2018). ISSN: 2197-7909. DOI: 10.1186/s40668-018-0026-4. URL: <http://dx.doi.org/10.1186/s40668-018-0026-4>.
- [68] Ofer Metuki et al. “Galaxy properties and the cosmic web in simulations”. In: *Monthly Notices of the Royal Astronomical Society* 446.2 (2014), 1458–1468. ISSN: 0035-8711. DOI: 10.1093/mnras/stu2166. URL: <http://dx.doi.org/10.1093/mnras/stu2166>.
- [69] R. B. Larson and B. M. Tinsley. “Star formation rates in normal and peculiar galaxies.” In: 219 (Jan. 1978), pp. 46–59. DOI: 10.1086/155753.
- [70] C. Schulz, J. Pflamm-Altenburg, and P. Kroupa. “Mass distributions of star clusters for different star formation histories in a galaxy cluster environment”. In: *Astronomy & Astrophysics* 582 (2015), A93. ISSN: 1432-0746. DOI: 10.1051/0004-6361/201425296. URL: <http://dx.doi.org/10.1051/0004-6361/201425296>.
- [71] Reinout J. van Weeren et al. “The case for electron re-acceleration at galaxy cluster shocks”. In: *Nature Astronomy* 1.1 (2017). ISSN: 2397-3366. DOI: 10.1038/s41550-016-0005. URL: <http://dx.doi.org/10.1038/s41550-016-0005>.

- [72] A. N. Timokhin, F. A. Aharonian, and A. Yu. Neronov. “On the non-thermal high energy radiation of galaxy clusters”. In: *Astronomy & Astrophysics* 417.2 (2004), 391–399. ISSN: 1432-0746. DOI: 10.1051/0004-6361:20040004. URL: <http://dx.doi.org/10.1051/0004-6361:20040004>.
- [73] F Vazza. “How complex is the cosmic web?” In: *Monthly Notices of the Royal Astronomical Society* 491.4 (2019), 5447–5463. ISSN: 1365-2966. DOI: 10.1093/mnras/stz3317. URL: <http://dx.doi.org/10.1093/mnras/stz3317>.
- [74] H. Umehata et al. “ALMA DEEP FIELD IN SSA22: A CONCENTRATION OF DUSTY STARBURSTS IN A $z = 3.09$ PROTOCLUSTER CORE”. In: *The Astrophysical Journal* 815.1 (2015), p. L8. DOI: 10.1088/2041-8205/815/1/18. URL: <https://doi.org/10.1088/2041-8205/815/1/18>.
- [75] Alan Dressler. “Galaxy morphology in rich clusters-Implications for the formation and evolution of galaxies”. In: *The Astrophysical Journal* 236 (1980), pp. 351–365.
- [76] A. Weißmann et al. “Studying the properties of galaxy cluster morphology estimators”. In: *Astronomy & Astrophysics* 549 (2012), A19. ISSN: 1432-0746. DOI: 10.1051/0004-6361/201219333. URL: <http://dx.doi.org/10.1051/0004-6361/201219333>.
- [77] Caitlin M. Casey, Desika Narayanan, and Asantha Cooray. “Dusty star-forming galaxies at high redshift”. In: *Physics Reports* 541.2 (2014), 45–161. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2014.02.009. URL: <http://dx.doi.org/10.1016/j.physrep.2014.02.009>.
- [78] D. N. Hoang et al. “Radio observations of the merging galaxy cluster Abell 520”. In: *Astronomy & Astrophysics* 622 (2019), A20. ISSN: 1432-0746. DOI: 10.1051/0004-6361/201833900. URL: <http://dx.doi.org/10.1051/0004-6361/201833900>.
- [79] Avery F. Garon et al. “Radio Galaxy Zoo: The Distortion of Radio Galaxies by Galaxy Clusters”. In: *The Astronomical Journal* 157.3 (2019), p. 126. ISSN: 1538-3881. DOI: 10.3847/1538-3881/aaff62. URL: <http://dx.doi.org/10.3847/1538-3881/aaff62>.

- [80] V. Cuciti et al. “New giant radio sources and underluminous radio halos in two galaxy clusters”. In: *Astronomy & Astrophysics* 609 (2018), A61. ISSN: 1432-0746. DOI: 10.1051/0004-6361/201731174. URL: <http://dx.doi.org/10.1051/0004-6361/201731174>.
- [81] Raul Jimenez and Licia Verde. “Implications for primordial non-Gaussianity (f_{NL}) from weak lensing masses of high- z galaxy clusters”. In: *Phys. Rev. D* 80 (12 2009), p. 127302. DOI: 10.1103/PhysRevD.80.127302. URL: <https://link.aps.org/doi/10.1103/PhysRevD.80.127302>.
- [82] Ray P. Norris. “Discovering the Unexpected in Astronomical Survey Data”. In: *Publications of the Astronomical Society of Australia* 34 (2017). ISSN: 1448-6083. DOI: 10.1017/pasa.2016.63. URL: <http://dx.doi.org/10.1017/pasa.2016.63>.
- [83] R. J. van Weeren et al. “Diffuse Radio Emission from Galaxy Clusters”. In: *Space Science Reviews* 215.1 (2019). ISSN: 1572-9672. DOI: 10.1007/s11214-019-0584-z. URL: <http://dx.doi.org/10.1007/s11214-019-0584-z>.
- [84] CJJ Pearce et al. “VLA Radio Observations of the HST Frontier Fields Cluster Abell 2744: The Discovery of New Radio Relics”. In: *The Astrophysical Journal* 845.1 (2017), p. 81.
- [85] M Gendron-Marsolais et al. “Deep 230–470 MHz VLA observations of the mini-halo in the Perseus cluster”. In: *Monthly Notices of the Royal Astronomical Society* 469.4 (2017), pp. 3872–3880.
- [86] S. Riggi et al. *Astronomical source finding services for the CIRASA visual analytic platform*. 2021. arXiv: 2110.08211 [astro-ph.IM].
- [87] Ray P Norris et al. “The Evolutionary Map of the Universe pilot survey”. In: *Publications of the Astronomical Society of Australia* 38 (2021).
- [88] Paul J Hancock et al. “Compact continuum source finding for next generation radio surveys”. In: *Monthly Notices of the Royal Astronomical Society* 422.2 (2012), pp. 1812–1824.
- [89] Dario Carbone et al. “PySE: Software for extracting sources from radio images”. In: *Astronomy and computing* 23 (2018), pp. 92–102.

- [90] Paolo Serra et al. “SOFIA: a flexible source finder for 3D spectral line data”. In: *Monthly Notices of the Royal Astronomical Society* 448.2 (2015), pp. 1922–1929.
- [91] M. Whiting and B. Humphreys. “Source-Finding for the Australian Square Kilometre Array Pathfinder”. In: *Publications of the Astronomical Society of Australia* 29.3 (2012), 371–381. ISSN: 1448-6083. DOI: 10.1071/as12028. URL: <http://dx.doi.org/10.1071/AS12028>.
- [92] Luis Lucas, Tim Staley, and Anna Scaife. *Efficient Source Finding for Radio Interferometric Images*. 2019. arXiv: 1901.04956 [astro-ph.IM].
- [93] Matthew T Whiting. “duchamp: a 3D source finder for spectral-line data”. In: *Monthly Notices of the Royal Astronomical Society* 421.4 (2012), pp. 3242–3256.
- [94] Simone Riggi et al. “Caesar source finder: Recent developments and testing”. In: *Publications of the Astronomical Society of Australia* 36 (2019).
- [95] Christopher Hollitt and Melanie Johnston-Hollitt. “Feature detection in radio astronomy using the circle Hough transform”. In: *Publications of the Astronomical Society of Australia* 29.3 (2012), pp. 309–317.
- [96] Andrew M Hopkins et al. “The ASKAP/EMU Source Finding Data Challenge”. In: *Publications of the Astronomical Society of Australia* 32 (2015).
- [97] Niruj Mohan and David Rafferty. *PyBDSF: Python Blob Detection and Source Finder*. Feb. 2015. ascl: 1502.007.
- [98] AM Hopkins et al. “A new source detection algorithm using the false-discovery rate”. In: *The Astronomical Journal* 123.2 (2002), p. 1086.
- [99] C. A. Hales et al. “blobcat: software to catalogue flood-filled blobs in radio images of total intensity and linear polarization”. In: *Monthly Notices of the Royal Astronomical Society* 425.2 (2012), 979–996. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2012.21373.x. URL: <http://dx.doi.org/10.1111/j.1365-2966.2012.21373.x>.
- [100] Emmanuel Bertin and Stephane Arnouts. “SExtractor: Software for source extraction”. In: *Astronomy and astrophysics supplement series* 117.2 (1996), pp. 393–404.

- [101] RJ Sault et al. “Astronomical Data Analysis Software and Systems IV”. In: (1995).
- [102] Ray P Norris. “Astroinformatics Challenges from Next-generation Radio Continuum Surveys”. In: *Proceedings of the International Astronomical Union* 12.S325 (2016), pp. 103–113.
- [103] Marcus Frean et al. “Source detection in astronomical images by Bayesian model comparison”. In: *AIP Conference Proceedings*. Vol. 1636. 1. American Institute of Physics. 2014, pp. 55–61.
- [104] Stefan Westerlund, Christopher Harris, and Tobias Westmeier. “Assessing the Accuracy of Radio Astronomy Source-Finding Algorithms”. In: *Publications of the Astronomical Society of Australia* 29.3 (2012), pp. 301–308.
- [105] Tobias Westmeier, Attila Popping, and Paolo Serra. “Basic testing of the duchamp source finder”. In: *Publications of the Astronomical Society of Australia* 29.3 (2012), pp. 276–295.
- [106] Attila Popping et al. “Comparison of potential ASKAP HI survey source finders”. In: *Publications of the Astronomical Society of Australia* 29.3 (2012), pp. 318–339.
- [107] Batta Mahesh. “Machine Learning Algorithms-A Review”. In: *International Journal of Science and Research (IJSR).[Internet]* 9 (2020), pp. 381–386.
- [108] Arthur L Samuel. “Some studies in machine learning using the game of checkers”. In: *IBM Journal of research and development* 3.3 (1959), pp. 210–229.
- [109] DD Proctor. “Morphological annotations for groups in the first database”. In: *The Astrophysical Journal Supplement Series* 194.2 (2011), p. 31.
- [110] Deanne D Proctor. “Low-resolution pattern recognition—sorting triples in the FIRST database”. In: *Journal of Electronic Imaging* 12.3 (2003), pp. 398–409.
- [111] DD Proctor. “Comparing pattern recognition feature sets for sorting triples in the FIRST database”. In: *The Astrophysical Journal Supplement Series* 165.1 (2006), p. 95.

- [112] S Bozinovski and A Fulgosi. “The influence of pattern similarity and transfer learning upon training of a base perceptron b2”. In: *Proceedings of Symposium Informatica*. 1976, pp. 3–121.
- [113] A Vafaei Sadr et al. “DeepSource: point source detection using deep learning”. In: *Monthly Notices of the Royal Astronomical Society* 484.2 (2019), 2793–2806. ISSN: 1365-2966. DOI: 10.1093/mnras/stz131. URL: <http://dx.doi.org/10.1093/mnras/stz131>.
- [114] Ana Martinazzo, Mateus Espadoto, and Nina ST Hirata. “Deep Learning for Astronomical Object Classification: A Case Study.” In: *VISIGRAPP (5: VISAPP)*. 2020, pp. 87–95.
- [115] Burger Becker and Trienko Grobler. “Classification of Fanaroff-Riley Radio Galaxies using Conventional Machine Learning Techniques”. In: *2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*. IEEE. 2019, pp. 1–8.
- [116] Halbert White. “Learning in artificial neural networks: A statistical perspective”. In: *Neural computation* 1.4 (1989), pp. 425–464.
- [117] Frederic B Fitch. “McCulloch Warren S. and Pitts Walter. A logical calculus of the ideas immanent in nervous activity. Bulletin of mathematical biophysics, vol. 5, pp. 115–133”. In: *Journal of Symbolic Logic* 9.2 (1944).
- [118] Zhouhan Lin et al. “Neural networks with few multiplications”. In: *arXiv preprint arXiv:1510.03009* (2015).
- [119] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Advances in neural information processing systems* 25 (2012), pp. 1097–1105.
- [120] Frank Rosenblatt. “The perceptron: a probabilistic model for information storage and organization in the brain.” In: *Psychological review* 65.6 (1958), p. 386.
- [121] Paul J Werbos. “Backpropagation through time: what it does and how to do it”. In: *Proceedings of the IEEE* 78.10 (1990), pp. 1550–1560.

- [122] Sergio Luis Suárez Gómez et al. “Experience with artificial neural networks applied in multi-object adaptive optics”. In: *Publications of the Astronomical Society of the Pacific* 131.1004 (2019), p. 108012.
- [123] J Roger P Angel et al. “Adaptive optics for array telescopes using neural-network techniques”. In: *Nature* 348.6298 (1990), pp. 221–224.
- [124] Rui Liu et al. “Impacts of EUV wavefronts on coronal structures in homologous coronal mass ejections”. In: *The Astrophysical Journal* 870.1 (2018), p. 15.
- [125] Peter L Wizinowich et al. “Neural network adaptive optics for the multiple-mirror telescope”. In: *Active and adaptive optical systems*. Vol. 1542. International Society for Optics and Photonics. 1991, pp. 148–158.
- [126] Michael Lloyd-Hart. “Application of machine learning to real-time removal of atmospheric blurring of astronomical images”. In: *Proc. Eight Yale Workshop on Adaptive and Learning Systems*. Citeseer. 1994, pp. 162–167.
- [127] Mark B Jorgenson and George JM Aitken. “Neural network prediction of turbulence-induced wavefront degradations with applications to adaptive optics”. In: *Adaptive and Learning Systems*. Vol. 1706. International Society for Optics and Photonics. 1992, pp. 113–121.
- [128] AT Mello et al. “Artificial neural networks for centroiding elongated spots in Shack–Hartmann wavefront sensors”. In: *Monthly Notices of the Royal Astronomical Society* 440.3 (2014), pp. 2781–2790.
- [129] Gleb V Vdovin. “Model of an adaptive optical system controlled by a neural network”. In: *Optical Engineering* 34.11 (1995), pp. 3249–3253.
- [130] Roberto Tagliaferri et al. “Neural networks for photometric redshifts evaluation”. In: *Italian Workshop on Neural Nets*. Springer. 2003, pp. 226–234.
- [131] M Brescia et al. “A catalogue of photometric redshifts for the SDSS-DR9 galaxies”. In: *Astronomy & Astrophysics* 568 (2014), A126.
- [132] M Brescia et al. “Photometric redshifts for quasars in multi-band surveys”. In: *The Astrophysical Journal* 772.2 (2013), p. 140.

- [133] Eros Vanzella et al. “Photometric redshifts with the Multilayer Perceptron Neural Network: Application to the HDF-S and SDSS”. In: *Astronomy & Astrophysics* 423.2 (2004), pp. 761–776.
- [134] Anna M M Scaife and Fiona Porter. “Fanaroff–Riley classification of radio galaxies using group-equivariant convolutional neural networks”. In: *Monthly Notices of the Royal Astronomical Society* 503.2 (2021), 2369–2379. ISSN: 1365-2966. DOI: 10.1093/mnras/stab530. URL: <http://dx.doi.org/10.1093/mnras/stab530>.
- [135] Nihan Coskun and Tulay Yildirim. “The effects of training algorithms in MLP network on image classification”. In: *Proceedings of the International Joint Conference on Neural Networks, 2003*. Vol. 2. IEEE. 2003, pp. 1223–1226.
- [136] Hongming Tang, Anna MM Scaife, and JP Leahy. “Transfer learning for radio galaxy classification”. In: *Monthly Notices of the Royal Astronomical Society* 488.3 (2019), pp. 3358–3375.
- [137] Nicholas O Ralph et al. “Radio galaxy zoo: Unsupervised clustering of convolutionally auto-encoded radio-astronomical images”. In: *Publications of the Astronomical Society of the Pacific* 131.1004 (2019), p. 108011.
- [138] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. “Learning representations by back-propagating errors”. In: *nature* 323.6088 (1986), pp. 533–536.
- [139] S Ben Driss et al. “A comparison study between MLP and convolutional neural network models for character recognition”. In: *Real-Time Image and Video Processing 2017*. Vol. 10223. International Society for Optics and Photonics. 2017, p. 1022306.
- [140] Michael A Nielsen. *Neural networks and deep learning*. Vol. 25. Determination press San Francisco, CA, 2015.
- [141] Li Li, Miloš Doroslovački, and Murray H Loew. “Approximating the gradient of cross-entropy loss function”. In: *IEEE Access* 8 (2020), pp. 111626–111635.

- [142] Runzhou Wang and Hong-Yan Bi. “A predictive model for chinese children with developmental dyslexia—Based on a genetic algorithm optimized back-propagation neural network”. In: *Expert Systems with Applications* 187 (2022), p. 115949.
- [143] Maitrei Kohli and TV Prasad. “Identifying dyslexic students by using artificial neural networks”. In: *Proceedings of the world congress on engineering*. Vol. 1. 1. 2010, pp. 1–4.
- [144] F Izhari, HW Dhany, M Zarlis, et al. “Analysis backpropagation methods with neural network for prediction of children’s ability in psychomotoric”. In: *Journal of Physics: Conference Series*. Vol. 978. 1. IOP Publishing. 2018, p. 012085.
- [145] Fahmi Izhari, Muhammad Zarlis, et al. “Analysis of backpropagation neural neural network algorithm on student ability based cognitive aspects”. In: *IOP Conference Series: Materials Science and Engineering*. Vol. 725. 1. IOP Publishing. 2020, p. 012103.
- [146] Armand Christopher Luna et al. “License Plate Recognition for Stolen Vehicles Using Optical Character Recognition”. In: *ICT Analysis and Applications*. Springer, 2022, pp. 575–583.
- [147] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [148] Justin Jonas and MeerKAT Team. “The MeerKAT radio telescope”. In: *MeerKAT Science: On the Pathway to the SKA* (2016), p. 1.
- [149] K Knowles et al. “The MeerKAT Galaxy Cluster Legacy Survey: I. Survey Overview and Highlights”. In: *Astronomy & Astrophysics* 657 (2022), A56.
- [150] Ezz El-Din Hemdan, Marwa A. Shouman, and Mohamed Esmail Karar. *COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images*. 2020. DOI: 10.48550/ARXIV.2003.11055. URL: <https://arxiv.org/abs/2003.11055>.
- [151] D. Wells, E. Greisen, and R. Harten. “FITS—A flexible image transport system”. In: *Astronomy and Astrophysics Supplement Series* 44 (May 1981), p. 363.

- [152] William D Pence et al. “Definition of the flexible image transport system (fits), version 3.0”. In: *Astronomy & Astrophysics* 524 (2010), A42.
- [153] P Grosbol. “The FITS Data Format”. In: *Bulletin d’Information du Centre de Données Stellaires* 35 (1988), p. 7.
- [154] E W Greisen and RH Harten. “An extension of FITS for groups of small arrays of data”. In: *Astronomy and Astrophysics Supplement Series* 44 (1981), p. 371.
- [155] Partha S. Pal et al. “Solar observatory Einstein Tower: Data release of the digitized solar full-disk photographic plate archive”. In: *Astronomische Nachrichten* 341.6-7 (2020), 575–587. ISSN: 1521-3994. DOI: 10.1002/asna.202013791. URL: <http://dx.doi.org/10.1002/asna.202013791>.
- [156] R. J. Hanisch et al. “Definition of the Flexible Image Transport System (FITS)”. In: 376 (Sept. 2001), pp. 359–380. DOI: 10.1051/0004-6361:20010923.
- [157] Karl Pearson. *On the theory of contingency and its relation to association and normal correlation*. Dulau and Company, 1904.
- [158] James T Townsend. “Theoretical analysis of an alphabetic confusion matrix”. In: *Perception & Psychophysics* 9.1 (1971), pp. 40–50.
- [159] Ron Kohavi and Foster Provost. *Applications of Machine Learning and the Knowledge Discovery Process*. 1998.
- [160] Mokhairi Makhtar, Daniel C Neagu, and Mick J Ridley. “Comparing multi-class classifiers: on the similarity of confusion matrices for predictive toxicology applications”. In: *International Conference on Intelligent Data Engineering and Automated Learning*. Springer. 2011, pp. 252–261.
- [161] Miguel Altuve, Antonio J Alvarez, and Erika Severeyn. “Multiclass classification of metabolic conditions using fasting plasma levels of glucose and insulin”. In: *Health and Technology* (2021), pp. 1–10.
- [162] Kaifang Zhang, Huayou Su, and Yong Dou. “Beyond AP: a new evaluation index for multiclass classification task accuracy”. In: *Applied Intelligence* (2021), pp. 1–11.

- [163] Fushing Hsieh and Xiaodong Wang. “Coarse-and fine-scale geometric information content of Multiclass Classification and implied Data-driven Intelligence”. In: *arXiv preprint arXiv:2104.07191* (2021).
- [164] Yusuf Brima et al. “Deep Transfer Learning for Brain Magnetic Resonance Image Multi-class Classification”. In: *arXiv preprint arXiv:2106.07333* (2021).
- [165] Frank Krüger. “Activity, context, and plan recognition with computational causal behavior models”. PhD thesis. Universiät Rostock. Fakultät für Informatik und Elektrotechnik, 2018.
- [166] Alaa Tharwat. “Classification assessment methods”. In: *Applied Computing and Informatics* (2020).