UNIVERSITY OF SOUTHAMPTON

Faculty of Engineering and Physical Science
School of Electronics and Computer Science

Machine learning for high energy astronomy surveys

by

Victoria Adele Lepingwell Childress
MPhys
ORCiD: 0000-0001-8941-6675

A thesis for the degree of
Doctor of Philosophy

October 2023

University of Southampton

Faculty of Engineering and Physical Science
School of Electronics and Computer Science

Doctor of Philosophy

Machine learning for high energy astronomy surveys

by Victoria Adele Lepingwell Childress

Abstract

This thesis presents new machine learning techniques for producing high energy astronomy survey catalogues. A novel source detector is developed for application to images from the INTEGRAL satellite. This source detector utilises convolutional neural networks (CNNs) to confidently identify genuine astrophysical sources whilst rejecting instrumental artefacts. This CNN-based source detector is substantially faster than previous methods, enabling the search for sources on shorter timescales than older techniques used in the production of previous INTEGRAL catalogues. The new capabilities afforded by the CNN source detector resulted in a 5% increase in sources found from the same dataset used to produce the previous INTEGRAL catalogue.

A Bayesian source combination technique is also presented that rapidly and reliably combines excess detections into a list of distinct sources. This method is superior to previous approaches because it requires no human intervention, and thus is less prone to human bias. It also is insensitive to the order in which excesses are presented to the algorithm, thereby providing consistent source catalogues regardless of how new detections are included.

Finally, a burst detection tool built with long short-term memory (LSTM) networks is presented. This burst detector reliably detects outbursts in simulated data sets (where the ground truth is known) with the same accuracy as previous tools but operating at substantially faster speeds. The burst detector demonstrates potential for applying reliable burst detection to massive data sets like those expected to be produced by the next generation of high energy surveys.

Overall, this thesis presents a powerful set of tools that could transform the way high energy astronomy surveys operate. Whilst this thesis demonstrates the advantages of using these tools for catalogue production, they have potential applications in real-time survey operations such as followup triggers after real-time outburst detection. Tools like those presented here will be vital for high energy astrophysics in the era of big data.

Contents

List of Figures

List of Tables

Declaration of Authorship

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as: V. A. Lepingwell, A. J. Bird, and S. R. Gunn. New approaches for faint source detection in hard X-ray surveys. *Monthly Notices of the Royal Astronomical Society*, 510(3):4031–4039, March 2022.

Signed:..... Date:.....

Acknowledgements

I would like first to thank my supervisors, Professor Steve Gunn and Professor Anthony Bird, who have had unmeasurable patience with me while I finished this project. Thank you to Steve for giving me such an incredible opportunity to acquire knowledge and expertise in Machine Learning; I am so grateful for all the encouragement you have shown me through all the years. I also want to thank Tony, who has not only been there to impart his Astronomy domain knowledge to me but has provided treasured support both academically and personally. I cannot fully express how much your input has influenced me and how much fun I have had whilst being lucky enough to be your PhD student.

I have been so fortunate to have collaborated with some of the most extraordinary women. Thank you, Dr Angela Bazzano, Dr Angela Malizia, Dr Manuela Molina and Dr Mariateresa Fiocchi, for always being on the other end of an email to offer support and encouragement. I am eternally grateful for the fun memories of travelling with you all for conferences and work meetings. I will never forget how you and Dr Pietro Ubertini made me feel so welcome and the many laughs we all shared.

I cannot forget to thank Dr Sadie Jones, who I hope is now a friend for life. Thank you for all the chats, coffees, and just being your fabulous Astrowarrior self. Please never change! The astronomy group are so lucky to have you.

A big thanks to my fellow PhD candidates and Postdocs in the astronomy group, especially: Dr Rob Firth, Dr Charlotte Angus, Dr Giorgios Dimitriadis, Dr Miika Pursiainen, Dr Sam Mangham, Dr Noel Castro Segura, Dr Peter Boorman, Dr Alessandra Costantino and Dr Rory Brown. Thanks for the many beers and all the help and support.

I am always thankful for my children: Aurora, Orion, Luna and Cosmo; although they have not made this process easy, I am glad they have been with me through this PhD adventure.

And last, but by no means least, I must thank my husband, Dr Michael Childress. I do not know where I would be without his neverending love and support - and Python debugging skills! I thank my lucky stars that this PhD brought us together - I could not have finished it without you.

To Mike, thanks for everything.

Definitions and Abbreviations

AGN	Active galactic nuclei
ANN	Artificial Neural Network
BAT	Burst Alert Telescope
BeX	Be X-ray Binary
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CTA	Cherenkov Telescope Array
CV	Cataclysmic Variable
EM	Electromagnetic
FOV	Field of View
FWHM	Full-Width at Half-Max
GC	Galatic Centre
GPS	Galatic Plan Scan
GPU	Graphics Processing Unit
GRB	Gamma Ray Burst
HEALPix	Hierarchical Equal Area isoLatitude Pixelation
HMXB	High-Mass X-ray Binary
IBIS	Imager on Board the INTEGRAL Satellite
INTEGRAL	The INTErnational Gamma-Ray Astrophysics Laboratory
ISGRI	INTEGRAL Soft Gamma-Ray Imager
JEM-X	Joint European X-ray Monitor
KNN	K-Nearest Neighbour
LC	Light Curve
LMXB	Low-Mass X-ray Binary
LSTM	Long Short-Term Memory
ML	Machine Learning
NRT	Near Real TIme
OSA	Off-Line Scientific Analysis
PSF	Point Spread Function
ReLu	Rectified Linear Activation
RNN	Recurrent Neural Network
ScW	Science Window

SFXT	Super-Fast X-ray Transient
SGXB	Super Giant X-ray Binary
SNR	Signal-to-Noise Ratio
SPI	SPectrometer on INTEGRAL
SSC	Synchrotron Soft Compton
SVM	Support Vector Machine
UV	Ultraviolet
WD	White Dwarf
XMM	X-ray Multi Mirror
XRB	X-Ray Binary

Chapter 1

Introduction

High energy astrophysics is a rich and rapidly evolving sub-field of astrophysics that focuses on astrophysical phenomena involving particles with exceptionally high energy that produce photons in the X-ray and gamma-ray region of the electromagnetic spectrum. The expansive volume of data gathered by modern high energy astrophysics instruments has presented an urgent need for the simultaneous development of computational methods that can keep pace with the imminent torrent of new data. This thesis presents work on advanced computational techniques, particularly utilising machine learning (ML) applied to data gathered by the INTEGRAL satellite (see Figure 1.1).

FIGURE 1.1: The INTEGRAL spacecraft... in space!

1.1 High Energy Astrophysics

Thanks to the current generation of space telescopes with survey capabilities, the sky in the hard X-ray/soft gamma-ray band (approximately 10keV-1MeV energy[1]) has shown itself to be both well-populated and highly variable. The sources in the hard X-ray sky display a huge dynamic range, and are thus detectable on many different timescales. While the brightest sources can be detected in a single observation, the faintest sources may require 1000s of images to be co-added.

A variety of interesting astrophysical sources demonstrate variability in this band, and INTEGRAL has played an important role in observing these. Accreting compact objects (neutron stars, black holes. or white dwarfs) form a substantial component of known variable X-ray sources. Low mass X-ray binaries (LMXBs) and high-mass X-ray binaries (HMXBs) have a central neutron star or black hole which accretes material from a donor star, and these systemsare distinguished by the mass of their companion star. Accreting white dwarfs comprise a variety of transient binary systems, including cataclysmic variable (CV) systems, as well as classical and recurrent novae systems. These accreting binary systems comprise Galactic X-ray sources, but a variety of extragalactic sources also exhibit transient X-ray emission. This includes supernovae, the explosive death of massive stars; active galactic nuclei, accreting supermassive black holes at the centre of distant galaxies; gamma-ray bursts, the violent merger of binary neutron stars; and gravitational wave binaries which emit X-rays and gamam-rays in the coalescence of neutron star - black hole binaries. These systems provide a broad range of interesting sources that have been the focus of INTEGRAL observations.

LMXBs are known to have rare “superburst” events due to long and energetic thermonuclear shell flashes on neutron stars, and INTEGRAL observations have helped constrain the possible fuel and ignition condition for these flashes. HMXBs show a wide variety of variability on a broad range of timescales, and regular monitoring with facilities like INTEGRAL contribute to observations of behaviours such as giant flares from pulsars and cyclotron line emission. The new class of supergiant fast X-ray transients (SFXTs) was confidently established by INTEGRAL observations, which continue to lead towards a better understanding of whether these arise from a magnetised stellar wind with fractal structure from OB-supergiant stars. Highly absorbed persistent HMXBs have also been examined in greater detail thanks to INTEGRAL observations. INTEGRAL data is also ideally suited to help identify keV counterparts for very high energy TeV events detected by gamma ray satellites or atmospheric Cherenkov telescopes. In addition to these Galactic X-ray transients targeted by Galactic Plane observations, these data also serendipitously sample the

[1] High energy astrophysics also encompasses the soft X-ray energy range covering approximately 1eV-10keV

pseudo-isotropic populations such as CVs and extragalactic sources like AGNs (see Chapter 2 for a detailed discussion of astrophysical sources).

Surveys in the hard X-ray/soft gamma-ray band are normally carried out using coded aperture telescopes that provide a good sensitivity across a wide field of view (typically > 100 deg^2), and as such allow frequent returns to the same sky region, producing rich datasets with information in both spatial and temporal dimensions. However, the analysis of data from coded aperture telescopes is not trivial, as it is an indirect imaging method and thus separation of sources from background noise or artefacts becomes exceptionally challenging. In addition to these challenges, current techniques for source detection are expensive in human and computational time which can limit how small an observational timescale we can feasibly search on.

Catalogues generated from astronomical surveys facilitate crucial research efforts such as population studies and the opportunity to identify new (especially transient) sources or more examples of rare sources. The techniques employed to generate high-energy astrophysics catalogues are no longer adequate as they do not scale well to the ever increasing datasets. For example, 'Catalog of 1000 orbits' (cat1000) (Bird et al., 2016) is an all-sky, soft gamma-ray source catalogue which uses data from INTEGRAL's first 1000 orbits. cat1000 took the INTEGRAL/IBIS survey team 2.5 years to produce and as INTEGRAL has now exceeded 2500 orbits, and the data from other such surveys is ever increasing. As surveys, like cat1000, are not simple pipeline processing tasks that can be easily scaled up, the 'cost' and lack of scalability comes in when human decision and intervention are needed. There is a clear need for automated techniques that can scale with the data when it exceeds the capacity to be processed manually. Deploying ML methods may help to understand the basis for decisions and interventions as well as make the problem tractable with less bias and inconsistency for larger datasets.

1.2 Machine Learning

With the ever-increasing volume of data generated by modern astronomical facilities, computational tools for handling and intelligently evaluating that data are vital for the successful realisation of the scientific potential of these facilities. Machine learning promises to be a transformative tool in that endeavour as it can capture complex physical relationships within datasets without requiring explicit parameterisation of those relationships.

Machine learning is an artificial intelligence that learns from examples or data. It involves using algorithms and statistical models to analyse and identify patterns in data and then using those patterns to make predictions or decisions about new, unseen data.

Machine learning can tackle a wide range of problems in various fields, including:

- Image and speech recognition: Machine learning algorithms can classify and recognise images (LeCun et al., 2015), videos, and speech (Hannun et al., 2014). For example, facial recognition systems (Taigman et al., 2014), voice assistants (Hinton et al., 2012), and self-driving cars (Bojarski et al., 2016) use machine learning to identify and interpret images and speech.
- Natural language processing: Machine learning algorithms can be used to understand and process human language (see, e.g., Vaswani et al., 2017), which includes tasks such as sentiment analysis (Socher et al., 2013), language translation (Sutskever et al., 2014), and text summarisation.
- Fraud detection: Machine learning algorithms can detect fraudulent activity, such as credit card fraud (Varmedja et al., 2019), insurance fraud (Aslam et al., 2022), and cyber attacks (Aljabri et al., 2021). By analysing large volumes of data, machine learning algorithms can identify patterns and anomalies that indicate fraudulent behaviour (Elliott et al., 2019).
- Recommendation systems: Machine learning algorithms can analyse user behaviour and make personalised recommendations (Rendle, 2010) which include applications such as movie and music recommendations (Hu et al., 2008), e-commerce product recommendations(Alves et al., 2019), and social media content recommendations (Gairola et al., 2017).
- Financial analysis: Machine learning algorithms can analyse financial data, including stock prices (Selvin et al., 2017), market trends (Rouf et al., 2021), and consumer behaviour (Lee et al., 2021) which allows for more accurate predictions and informed decision-making (Bao et al., 2017).
- Medical diagnosis: Machine learning algorithms can analyse medical data, including patient records (Alghamdi et al., 2017), medical images (Shen et al., 2017), and genetic data (Yue and Wang, 2018), which allows for more accurate diagnoses and personalised treatment plans.

Machine learning can also be used for time series analysis and forecasting. Here are some examples of how machine learning can be used for time series analysis:

- Financial time series analysis: Machine learning algorithms can analyse financial time series data, including stock prices (Jiang et al., 2017), foreign exchange rates (Tsuji, 2022), and commodity prices, which includes trend analysis, volatility prediction, and trading signal generation (Amin, 2020).

- Energy forecasting: Machine learning algorithms can forecast energy demand and supply, including electricity demand, renewable energy production, and oil and gas production (Liu et al., 2023), which can help energy companies optimise their operations and plan for future investments (Singh et al., 2017).
- Health monitoring: Machine learning algorithms can monitor health data over time, including vital signs (Liu et al., 2014), medical records (Alghamdi et al., 2017), and biometric data, which includes tasks such as disease prediction (Mohan et al., 2019), medication adherence (Son et al., 2010), and personalised treatment plans (Miotto et al., 2016).
- Sales forecasting: Machine learning algorithms can be used to forecast sales for businesses, including retail sales, e-commerce sales, and restaurant sales which can help companies to plan inventory, optimise staffing, and forecast revenue (see Pavlyshenko, 2019, for a review).
- Climate modelling: Machine learning algorithms can analyse climate data, including temperature, precipitation, and atmospheric conditions, which includes weather forecasting (Dueben and Bauer, 2018), climate modelling (Kashinath et al., 2021), and natural disaster prediction tasks (Choi et al., 2018).

As seen in the above list Machine Learning comes in a variety of models which each have their own distinct features that favour application to particular computational problems. Machine Learning applications developed in this work focus on two distinct problems: image classification for identification of genuine sources in astrophysical imaging data, and time series data analysis for identification of periods of outburst in light curves.

One very popular flavour of machine learning utilised heavily in astronomy is a convolutional neural network (CNN). CNNs are particularly well-suited for application to imaging data, as they apply convolutions that help reveal key structure in the data. Another type of machine learning algorithm that is gaining traction is a long short-term memory (LSTM), a kind of recurrent neural network (RNN) that is excellent for characterising time series data and is becoming popular in astronomy applications.

1.3 This work: machine learning for high energy astrophysics surveys

This thesis presents three new tools for finding and characterising astrophysical sources in INTEGRAL data, making particular use of machine learning (ML) to

streamline and automate the process. These tools are: a CNN-driven source detector, a Bayesian source matching algorithm, and a LSTM-driven burst detector.

CNN Source Detector: A CNN was trained on data from the INTEGRAL/ISGRI telescope to create a source detection tool that is more sensitive than previous methods, whilst taking less time to apply to the data and reducing the human subjectivity involved in the process. This new CNN-based approach utilises all five INTEGRAL energy bands used to create the survey simultaneously – not only does this improve the accuracy and allow the network to detect sources at a lower flux threshold, it also speeds up the entire source detection and labelling process as when a search is performed on a sky map it results in a single detection for each astrophysical source instead of one from each energy band. The speed on which the CNN can be applied to the entire dataset allows inspection on a smaller timescale than previously feasible, meaning sources can be detected that only appear in a single observation but fall below the detection threshold in all-sky maps stacked from images spanning a longer time-scale. Looking on smaller time-scales is extremely useful for detecting faint and transient sources which are often the most particular and most interesting to study, as these can reveal new astrophysical mechanisms not previously observed.

Bayesian Source Merging: In addition to the CNN a method based on Bayesian reasoning was used which improved how detections from multiple observations were merged into unique sources. This method calculates a Bayesian likelihood of two sources being the same, for every pair of detections in the catalogue, then uses this likelihood to merge detections into unique sources. One main advantage of this approach is that the results are invariant to the order in which detections are presented to the algorithm, as compared to source merging algorithms which start with a catalogue of known sources.

LSTM Burst Detector: A prototype ML driven burst detector has been developed which is designed to detect and characterise all outbursts for new and known transient sources. The burst detector uses a specific type of recurrent neural network called a long-short term memory network (LSTM) – a powerful kind of recurrent neural network (RNN) that is ideally suited for application to time-series data. RNNs generally are well-suited for analysing time-ordered data, but while traditional RNNs have limited "memory" (they are less influenced by older data) LSTMs have a mechanism for retaining key information from older data. This makes it possible to not only detect important events but also to classify these events and also employ prediction capabilities allowing for time-series forecasting. The LSTM model is also exceptionally fast – orders of magnitude faster than previous burst-finding methods – allowing large volumes of data to searched very quickly.

Collectively the ML-driven approaches presented in this thesis form an exceptionally powerful toolkit for addressing the major issues expected in future surveys. This

thesis is organised as follows. Chapter 2 presents the high energy astronomy background that motivated the development of these tools, whilst Chapter 3 introduces the fundamental machine learning approaches used to address these problems and other such problems in astronomy. Chapter 4 presents the CNN-based source detector developed to identify flux excesses in INTEGRAL ISGRI images. Chapter 5 describes the Bayesian approach for merging excesses found by the source detector, and presents the outcomes for the final source catalogue produced with these new tools. Chapter 6 introduces the new LSTM-based burst detector tool used to identify burst-like activity in combined INTEGRAL+Swift light curves. Finally, Chapter 7 presents the conclusions from this thesis and a view towards its potential impact for future astronomy surveys. Parts of Chapters 4 and 5 have been published in Lepingwell et al. (2022), and Chapter 6 forms most of a paper planned for submission.

Chapter 2

Background: High Energy Astrophysics

The field of astrophysics known as “high energy” examines the far reaches of our galaxy and the cosmos. Extreme densities and temperatures, high velocities, enormous magnetic fields, and strong gravity pose persistent challenges to our understanding of the behaviour of matter in extreme environments such as black holes, neutron stars, exploding supernovae, and relativistically moving jets. Active Galactic Nuclei, pulsars, supernovae, and gamma-ray bursts all produce strange energetic phenomena, and understanding these extreme environments is crucial to making sense of them. The quality of data and the sophistication of modelling in high-energy astrophysics are both increasing at an unprecedented rate and a rare convergence of simultaneous observations from ground and space-based telescopes spanning the entire electromagnetic spectrum (see Figure 2.1) over the next few years is anticipated, including the JVLA (radio and sub-millimeter), ALMA (radio and sub-millimeter), Hubble/JWST (optical/infrared), Chandra, *Swift*, NuStar (X-rays), INTEGRAL and Fermi (gamma-rays), and HESS/MAGIC (multi-TeV gamma-rays).

High energy astrophysics is one of the most active subfields of astrophysics today. It focuses on astrophysical phenomena that involve particles of “high energy”, which in turn produce X-ray and gamma-ray photons, such as accreting super-massive and stellar-size black holes and numerous neutron star species. More generally, these astrophysical phenomena emit non-thermal photons outside of the traditional optical wavelengths. Studying non-thermal emissions in the universe makes it possible to reveal the portion of the universe that is not in a steady state, typically originating in violent environments near compact objects such as neutron stars and black holes of varying sizes. In addition to emitting broad-band non-thermal electromagnetic radiation, it is believed that these objects also emit signals outside the electromagnetic spectrum. These signals comprise cosmic rays, neutrinos, and gravitational waves.

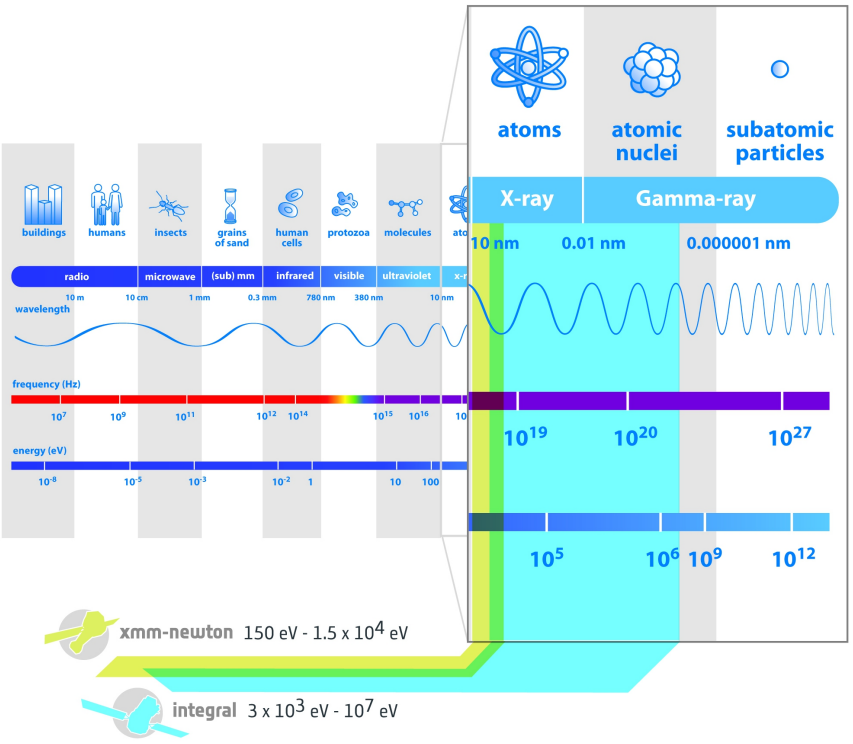

FIGURE 2.1: Overview of the electromagnetic spectrum, illustrating characteristic frequency, energy, and wavelength scale of each region of the EM spectrum. Regions of sensitivity of high energy astronomy experiments including INTEGRAL and XMM-Newton are highlighted. (Source: ESA website)

With the operation of numerous space and ground-based observational facilities, the field of high-energy astrophysics has flourished in recent decades.

In this chapter, the fundamental principles underlying the production and interaction of high-energy photons in the hard X-ray and soft gamma-ray regimes are laid out and discussed. In addition, the characteristics of sources that are known to emit at these high energies are discussed specifically in relation to X-ray binaries. Some history and background into Astrophysical surveys is also presented focusing on the INTEGRAL mission and how data collected from there is used in this thesis.

2.1 X-ray and Gamma-ray astronomy

X-ray astronomers frequently study phenomena which occur at the end of the stellar lifetimes – supernova explosions, neutron stars, and stellar black holes – and also investigate extra-galactic objects such as radio galaxies, Seyfert galaxies, quasars with accreting supermassive black holes in their centres and clusters of galaxies, the largest physical formations of our universe. Stars and galaxies can also be studied with

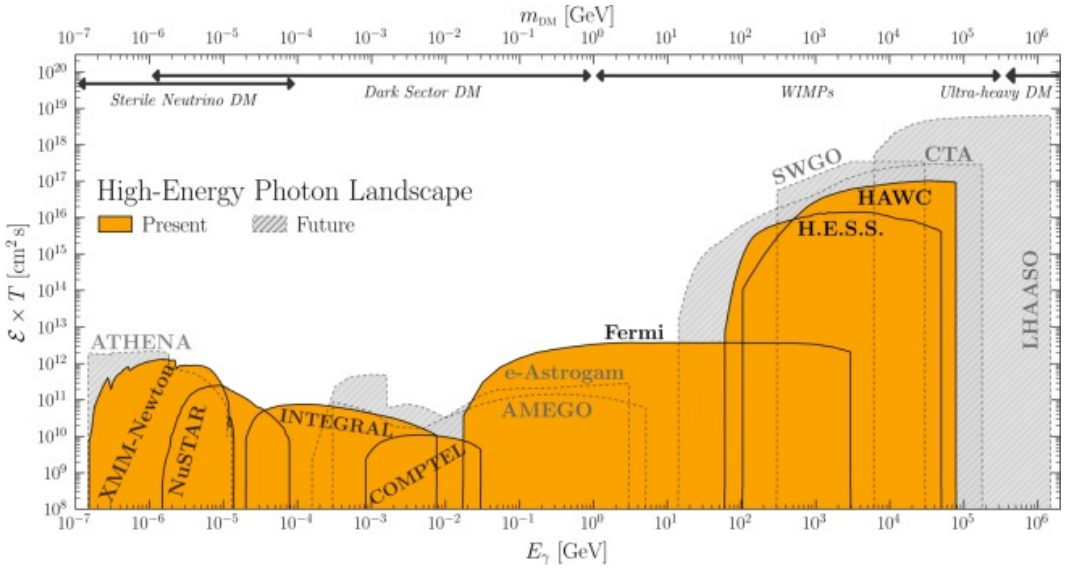

FIGURE 2.2: The exposure of existing and upcoming X-ray and gamma-ray instruments, from Boddy et al. (2022).

modern X-ray telescopes, and even comets and planets in the solar system have been observed in X-rays.

The last few decades of X-ray astronomy have provided the field with ample opportunities for observing compact X-ray sources (white dwarfs, neutron stars, black holes) and transient events, which have given great insight into the to different populations of galactic binary systems, such as low- and high-mass X-ray binaries, cataclysmic variables, symbiotic systems. A graphical overview of X-ray and gamma-ray facilities and their coverage is shown in Figure 2.2

Over 10,000 times more energetic than photons of visible light, gamma rays are the most powerful form of electromagnetic radiation. They can be produced by solar eruptions, supernovae, neutron stars, black holes, and active galaxies. Gamma-ray astronomy typically is studying the energy range from 500 keV to more than 1 TeV, as this spans over seven orders of magnitude, a wide variety of detectors are used to study smaller sub-ranges. The atmosphere of Earth absorbs the vast majority of gamma rays and X-rays. Therefore, high-altitude balloons and satellites are the usual observation platforms for both energy ranges.

2.2 Physical Mechanisms for Producing X-ray and Gamma-ray Emission

X-rays wavelengths range from 0.01 to 10 nanometers (3×10^{16} Hz to 3×10^{19} Hz), which corresponds to energies between 100 eV and 100 keV. X-ray wavelengths are shorter than UV wavelengths and longer than gamma wavelengths. In recent decades,

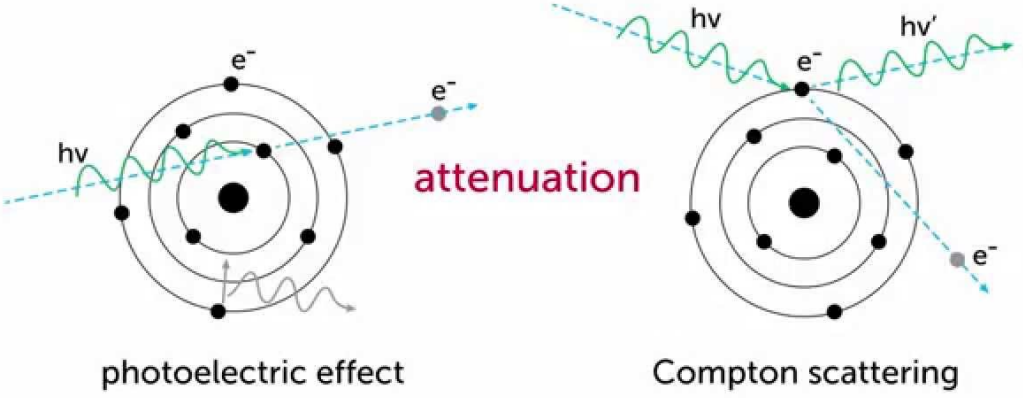

FIGURE 2.3: Comparison of some key X-ray matter interactions: in the photoelectric effect, a photon is fully absorbed and liberates a bound electron; in Compton scattering a photon scatters off an electron an imparts some of its energy to the electron. (Source: ASTRA Lab, University of Antwerp)

the distinction between X-rays and gamma rays has become more complex. X-rays are generally emitted by electrons outside the nucleus, while gamma rays are emitted by the nucleus, per the currently accepted definition. The way in which X- and gamma-rays are produced and interact is directly related to the ways in which they can be detected and what the implications are for the physical systems which generate them. There are three key processes in which they interact with matter: the photoelectric effect; Compton scattering; pair production. Which process is dominant is dependent upon the energy of the incoming photon and the atomic number, Z, of the material in which it is interacting. At low-energies of X-rays the photoelectric absorption dominates and at higher energies Compton scattering dominates. These mechanisms are illustrated in Figure 2.3 and described in greater detail as follows.

2.2.1 The photoelectric effect

A photon interacts with an atom-bound electron during the photoelectric effect. The photon's energy is transferred to the electron, ejecting it from the atom, the incident photon completely dissipates, and the atom emits an energetic photoelectron from one of its bound shells. If the photon energy, $h\nu$, is greater than the electron's binding energy, E_{bind}, then the additional energy is transferred to the electron in the form of kinetic energy, E_u with:

$$E_u = h\nu - E_{bind} \tag{2.1}$$

This process is illustrated in the left side of Figure 2.3.

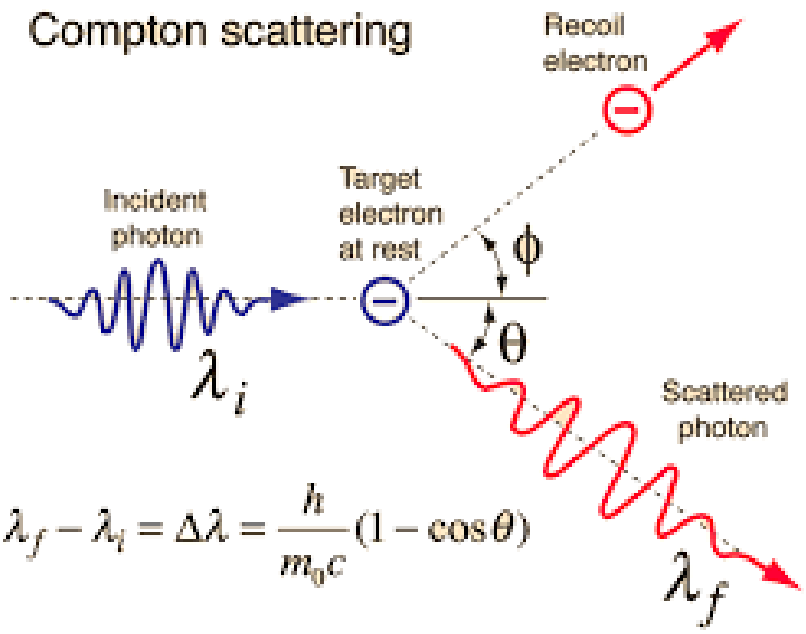

FIGURE 2.4: Compton scattering: a photon imparts some of its energy to an electron: energy and momentum conservation define the energy of the scattered photon and electron. (Source: HyperPhysics at GSU)

2.2.2 Compton Scattering

Compton scattering takes place when a photon of high energy scatters off a free electron. As shown in Figure 2.4, when a photon scatters off an electron, it transfers energy and momentum to the electron.

The energy of the scattered photon, E, is calculated using the principles of conservation of energy and momentum. The Compton scattering cross section is determined by the "Klein-Nishina formula", which reduces to the Thomson cross-section for low photon energies where conventional electron scattering is observed.

During inverse Compton scattering, it is also possible for high-energy electrons to transfer energy to photons. In this instance, an energetic electron scatters a photon. In the case of a group of particles with multiple scatterings, the up-scattered photon has an energy of γE where γ is the electron Lorentz factor and E is the photon's initial energy.

2.2.3 Pair production

At the highest energies, a gamma-ray's energy can be converted into a particle-antiparticle pair and annihilated completely. To conserve momentum, the interaction typically occurs in the Coulomb field of a nucleus. The most common

particle-antiparticle produced is an electron and positron, which requires a photon of gamma-ray energy greater than 1.02 MeV. Any surplus energy is transferred in the form of the particles' kinetic energy. At energies above 10 MeV, pair production predominates over photoelectric and Compton effects. The positron will eventually collide with an electron and destroy it, releasing gamma-rays.

2.2.4 Blackbody emission

Blackbody emission of an object is the source of both hard X-rays and soft gamma-rays, which are generated by thermal processes. Wien's displacement law allows us to determine the frequency at which the Planck law has the highest specific intensity. This is given by:

$$\nu = T \cdot 5.879 \times 10^{10} \mathrm{K}^{-1}\mathrm{Hz} \tag{2.2}$$

In order for the maximum intensity of a blackbody to exceed 20 keV, temperatures exceeding 5×10^7 K are required.These temperatures can be found in the accretion discs of stellar mass black holes and neutron stars.

2.2.5 De-excitation (line emission)

The de-excitation of nuclei results in the emission of gamma-rays. The prototypical example is the radioactive decay of unstable isotopes. However, collisions with energetic particles, such as cosmic rays, can also excite nuclei. As the rays emitted depend on the type of nucleus and the level of excitation, this results in line emission, which is used to determine the abundance of isotopes.

Electron de-excitation in high mass atoms is also a key mechanism for the production of X-rays in astrophysical sources. In hot accretion disks, the Fe Kα line is frequently observed and can be used to infer the dynamics of these systems. In supernova remnants, X-ray line emission from heavy elements can be used to infer the structure and composition of the supernova material.

2.2.6 Synchrotron radiation

By accelerating relativistic electrons using magnetic fields, synchrotron radiation is produced. As an electron moves through a magnetic field, it encounters a force perpendicular to its motion and consequently changes its direction. As a result of the electron's acceleration, it emits electromagnetic radiation.The frequency of the radiation is solely determined by the electron's energy, E, the strength of the magnetic

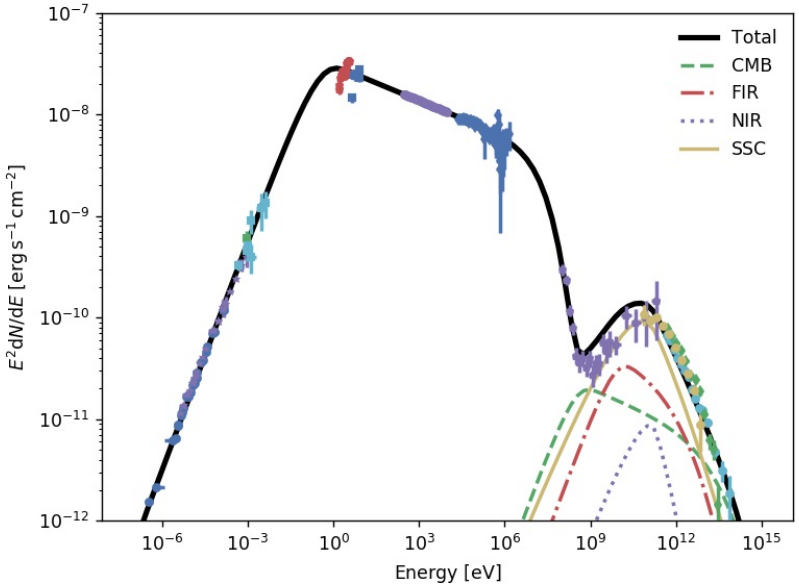

FIGURE 2.5: Observed spectral energy distribution of the Crab Nebula, with model fit including synchrotron radiation and SSC components (Zabalza, 2015)

field, B, and the direction of motion relative to the magnetic field. In a collection of particles with isotropic motion, only the field strength and the energy spectrum of the electrons are significant. The average photon energy scales as BE^2. The synchrotron spectrum is dependent on the assumed power-law energy spectrum of the electrons. If the magnetic field is aligned, the synchrotron radiation that results will be polarised.

Synchrotron radiation in most astrophysical systems peaks at radio wavelengths. These radio photons then often interact with the energetic electrons that produced them and have their energies boosted to X-ray and gamma-ray energies by inverse Compton scattering. This process is referred to as synchrotron self-Compton (SSC), and an example spectrum with SSC is shown in Figure 2.5.

2.2.7 Bremsstrahlung

Bremsstrahlung, also known as "braking radiation," arises when a charged particle travels through an electric field, such as an ion's Coulomb field. Since the rate of energy loss is proportional to m^{-2}, electrons are the predominant emission source. As the electron travels through the electric field of the ion, it is accelerated and consequently emits radiation.

Thermal Bremsstrahlung is the radiation spectrum emitted in the presence of an electric field by a body of electrons in thermal equilibrium. The higher the temperature, the faster the electrons move and the greater the energy of the photons emitted. Gamma-rays require temperatures on the order of 10^8 K to be produced. The

spectrum is flat and has a cut-off at an energy of $E \sim kT$. Consequently, the plasma temperature can be measured by locating the cut-off.

2.2.8 Accretion

In galactic and extragalactic systems, the action of matter accreting onto a compact object is the primary source of energy for high-energy emission. Accretion is the addition of material to an object over time. The origin of material transfer depends on the system in question; for instance, in some X-ray binaries, the companion star fills the Roche-lobe, and material is transferred from it to the compact object through the inner Lagrangian point. The same physical laws govern the release of energy in each of these systems regardless of the origin of the accreting material. As matter falls onto a dense object, gravitational potential energy is released and becomes available as accretion energy. For a compact object with mass M_x and radius R_x, the energy released by the accretion of a mass m can be calculated as follows:

$$E_{acc} = \frac{GM_x m}{R_x} \tag{2.3}$$

If matter is continuously added over time, then luminosity attributable to accretion is determined by the following formula:

$$L_{acc} = \frac{GM_x}{R_x}\frac{dm}{dt} \tag{2.4}$$

By heating the material within the accretion disc, gravitational potential energy is released as matter is accreted. The efficiency of the accretion process can be determined by comparing the total amount of energy possessed by the matter. Efficiency may be expressed as:

$$\eta = \frac{GM_x}{R_x c^2} \tag{2.5}$$

The Eddington limit is an additional limitation on the energy emitted through accretion. The scattering of photons by matter exerts a force on the matter, despite the fact that this force is typically very small. This radiation pressure is proportional to the radiation field flux that is incident.

$$F = \frac{\sigma_T L}{4\pi R^2 c} \tag{2.6}$$

Therefore, large fluxes have a noticeable effect on the matter they traverse. Accretion will cease if the force of radiation pressure exceeds that of gravitational attraction. This luminosity is known as the Eddington luminosity, and it is calculated as follows:

$$L_{edd} = \frac{4\pi G M_x m_p c}{\sigma_T} = 1.3 \times 10^{33} M/M_\odot \text{ erg s}^{-1} \tag{2.7}$$

where σ_T is the Thomson scattering cross section, m_p is the mass of a proton, M_x is the mass of the compact object. Additionally, the spectrum resulting from the accretion process can be estimated.

If the disc is optically thick it is expected the radiation will reach thermal equilibrium with the material of the disc, and the source would radiate as a black body with a temperature of T_{bb}. If optically thin, the radiation would escape without further interaction, and the characteristic radiation temperature would be that of the gravitational potential if each proton-electron pair were directly converted into thermal radiation.

For a neutron star or black hole system, it is expected accretion disc photons will have energies in the keV range, and a white dwarf system would have photon energies in the eV range.

2.3 Galactic Sources

Our Galaxy contains the brightest and most persistent high energy sources known (owing to the fact they are the nearest sources); most of the X-ray emitters are binary stars that produce their radiation through accretion onto a neutron star or black hole. There are two types of X-ray binaries: those with a high mass and those with a low mass companion. However, other source types, such as cataclysmic variables (CVs) and supernova remnants (Córdova, 1995) SNRs) , have been observed to emit in the hard X-ray/ gamma-ray band. The gamma ray emission is a sign of the non-thermal processes at work in these systems, which allows accretion and the dynamics of individual objects to be studied.

2.3.1 Binaries

In practice, the brightest (apparent) sources of X-ray emission in the sky are X-ray binaries, which consist of a compact star (a white dwarf, neutron star, or black hole) and a companion star (a brown dwarf, white dwarf, or main sequence star). Compact stars release gravitational energy by absorbing materials transferred from companion stars and emit radiation in the X-ray band. Many X-ray binaries within or outside of

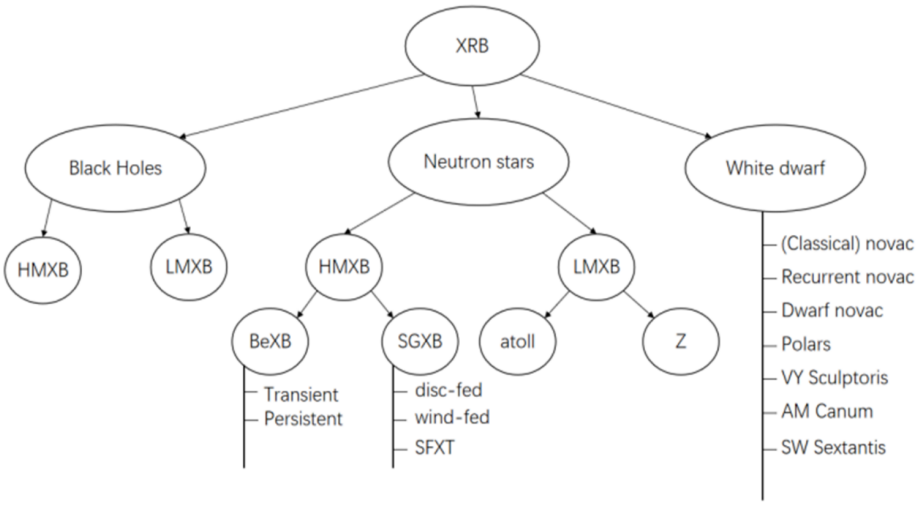

FIGURE 2.6: Classification of X-ray binaries. (Source: IOP Science)

the galaxy have been discovered by utilising and relying on various X-ray surveys and optical identification. As depicted in Figure 2.6, the classification of X-ray binary subsystems is represented by a tree diagram containing various subsystem types. When the primary star is a neutron star or a black hole, X-ray binaries can be classified according to the mass of their secondary star as either low-mass X-ray binaries (LMXBs) or high-mass X-ray binaries (HMXBs, with a companion star whose mass is greater than one solar mass).

When the primary star is a white dwarf, the X-ray system can be classified into numerous sub-types based on the observational differences. In the majority of cases, they belong to one type of essential X-ray source, namely cataclysmic variables (CVs).

Approximately 90 percent of X-ray sources with strong emissions in the Galaxy can be classified as LMXBs and HMXBs. The vast disparity in mass between LMXBs and HMXBs influences the variety of their physical characteristics. The traditional physical and observational characteristics of these two types of X-ray binary are compared in Table 2.1. It is worth noting that these characteristics reflect the traditional distinction between these classes, but subtleties of their distinction are constantly evolving. For example, some LMXBs have been found to have black holes as their central comopact object (though as noted below they cannot exhibits Type I bursts as they lack a surface for accreted material to accumulate).

2.3.2 Low Mass X-ray Binaries

Sco X-1 was the first point X-ray source observed (Giacconi et al., 1962), and later determined that it was a low mass X-ray binary. Systems in which the donor star has a mass less than $1M_{\odot}$ are considered low mass X-ray binaries (LMXBs) (Liu et al., 2001).

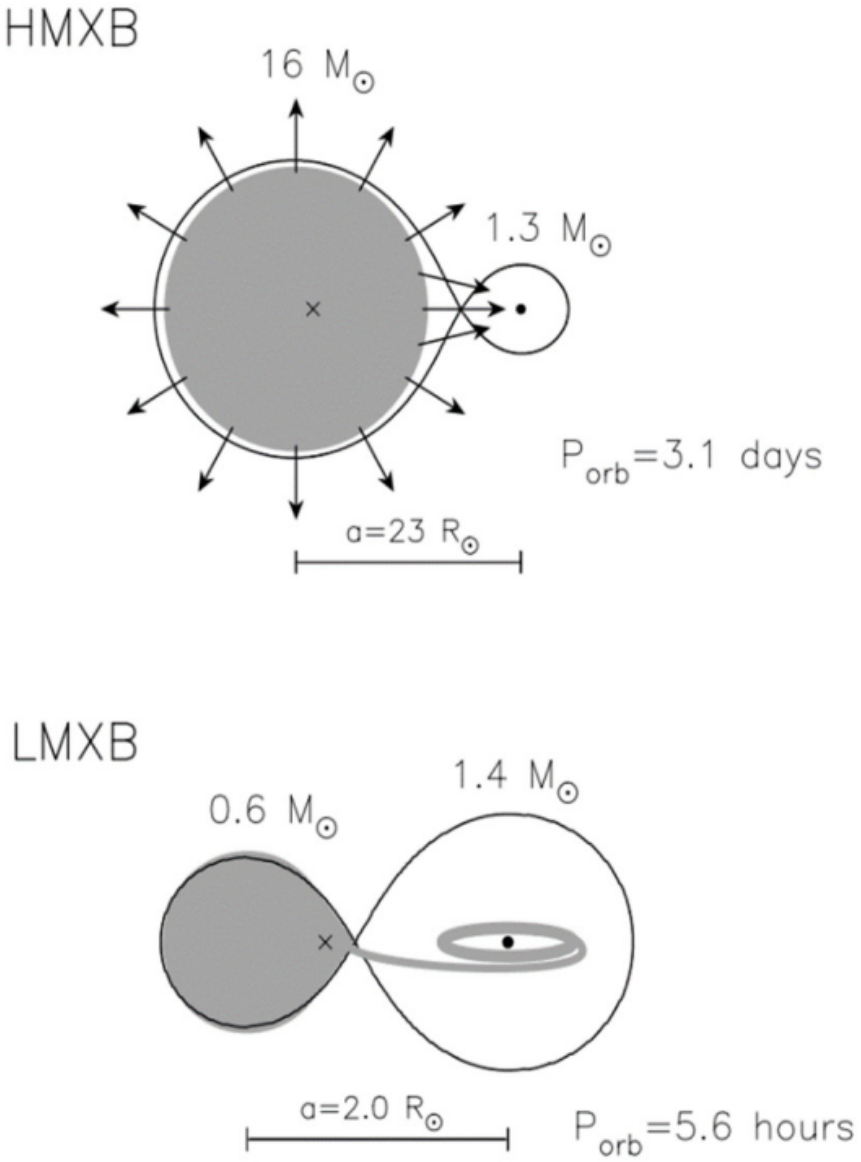

FIGURE 2.7: Typical HMXBs and LMXBs with neutron star as the accretion star. In HMXBs, the matter accreted onto the neutron stars originate from strong stellar winds or Roche Lobe overflow that passes through Lagrangian point L1. In LMXBs, neutron stars absorb matter via the accretion disc formed by accretion currents of Roche Lobe overflow which passes through Lagrangian point L1. (Source: IOP Science)

	HMXBs	LMXBs
X-ray spectra	hard spectrum, $kT \geq 15$ keV	sort spectrum $kT \leq 10$ keV
Type of time variability	X-ray pulses, no X-ray bursts	X-ray bursts but rarely pulses
Accretion process	Stellar winds (or Roche Lobe overflow)	Roch Lobe overflow
Time scale of accretion	10^5 yr	$10^7 - 10^9$ yr
Accreting compact star	Neutron stars (or black holes) with strong magnetic field	Neutron stars (or black holes) with weak magnetic field
Spatial distribution	Galactic plane	Near Galactic plane and centre
Stellar generation	Young stars, age $\leq 10^7$ yr	Old stars, age $\geq 10^9$ yr
Optical companion star	O(B) stars in early phase, $M \geq 10M_{\odot}$, brighter than X-ray: $L_{opt}/L_X \geq 1$	Blue stars, $M \leq 10M_{\odot}$, fainter than X-ray: $L_{opt}/L_X \leq 1$

TABLE 2.1: Traditionally expected features of x-ray binaries (adapted from Dong, 2012)

It's not necessary for the donor star to be a white dwarf; in fact, it can be any star with a spectral type above A. (White et al., 1995). A star cannot produce the X-ray luminosity seen in such systems if it has a weak wind. This high-energy outflow is caused by material accreting from the donor star through the inner Lagrangian point due to Roche lobe overflow. An accretion disc is formed when matter enters the compact object through the inner Lagrangian point and spirals downwards toward it. These compact objects become extremely luminous X-ray sources with luminosities $L \geq 10^{34}$ erg s^{-1} as matter is accreted onto them. In Figure 2.7 we see a diagram of a neutron star LMXB system.

Although most LMXBs in the Galaxy are found near the Galactic Centre (GC), a small number have been spotted in globular clusters. The fact that these sources are found in regions unrelated to star formation suggests that these are very old stellar systems (White et al., 1995). The sub-solar-mass companions' nature, in which they live for aeons, lends credence to this. Given the age of these systems, the neutron star is predicted to have a weak magnetic field, which is why X-ray pulsations are rarely observed. Her X-1 and 1626-673 are two notable outliers; both are classified as LMXBs, but also show X-ray pulsations. Typically the spectral type of the donor star or the system's mass function is used to determine whether or not a given system is a LMXB. In the systems with limited X and gamma ray observations a source can be classified an LMXB if either type 1 X-ray bursts (see Section 2.3.4 where these transient LMXBs are described in greater detail), an orbital period of less than 12 hours or a 1-10 keV soft spectrum with a temperature of 5-10 keV are observed. (White et al., 1995).

2.3.3 High Mass X-ray binaries

Massive, luminous OB stars are the companion stars of HMXBs, and their lifespan is determined by the star's evolutionary time scale, which is on the order of 10^{10} yr or less (10^7 yr for more massive companions). Found in the same star-forming regions in the spiral arms and galactic plane as young stellar populations, strong stellar wind can be generated because of the companion star's large mass and low density in its vicinity. The host star continues to accrete material into the Roche lobe (outer bound to the region around a star in a binary system where orbiting material is gravitationally bound to that star) once it is full, causing it to glow brightly in X-rays. This type of source is accompanied by companion stars that emit strongly in the optical. The donor star in the system is much more luminous in the optical than the X-ray source, X-ray heating is minimal and so the optical properties of the system are dominated by the donor star (White et al., 1995). Variation in X-ray luminosity can be explained by the strength and rate of change in the stellar wind velocity and the separation of binaries' constituent stars. The neutron stars in these systems are typically observed to be X-ray pulsars. This follows from the idea that HMXBs must be young systems due to the short life span of the OB stars. Hence, the neutron stars in these systems are expected to be young and so have large magnetic fields. That these are young stellar systems means that they are found in star forming regions, such as the spiral arms of the galaxy. The donor stars are of sufficient size to have a strong stellar wind. The neutron star's magnetic field remains in the high-magnetic (10^{12} G) environment at birth because the neutron star has not evolved over a long period and has only accreted less matter. The accreted material moves in the direction of the magnetic field lines at the poles of the star and is ejected as a narrow beam of radiation in the presence of a strong magnetic field. Through the process of accretion, the neutron star gains mass and energy, which can be seen as a pulsar. Supergiant X-ray binaries with an OB supergiant as the secondary star (SGXBs) and Be X-ray binaries with a Be star as the companion star are two subclasses of HMXBs distinguished by their different observational properties, in particular the type of companion. The majority of HMXBs belong to the first class (80 percent) (Kaper et al., 2004). Be systems are often seen as transient X-ray sources whilst the supergiant systems are persistent although variable.

The classification of a system as a HMXB is typically based upon the identification of the donor star spectral type or of the mass function of the system. However, in those cases where there are limited observations which are in the X-ray or gamma-ray energy regimes then the characteristic X-ray properties of traditionally identified systems can be used to classify newly detected sources. A source may be classified as an HMXB if it exhibits any of the following properties (White et al., 1995):

- Strong flaring and absorption variability on timescales of minutes
- Transient outbursts

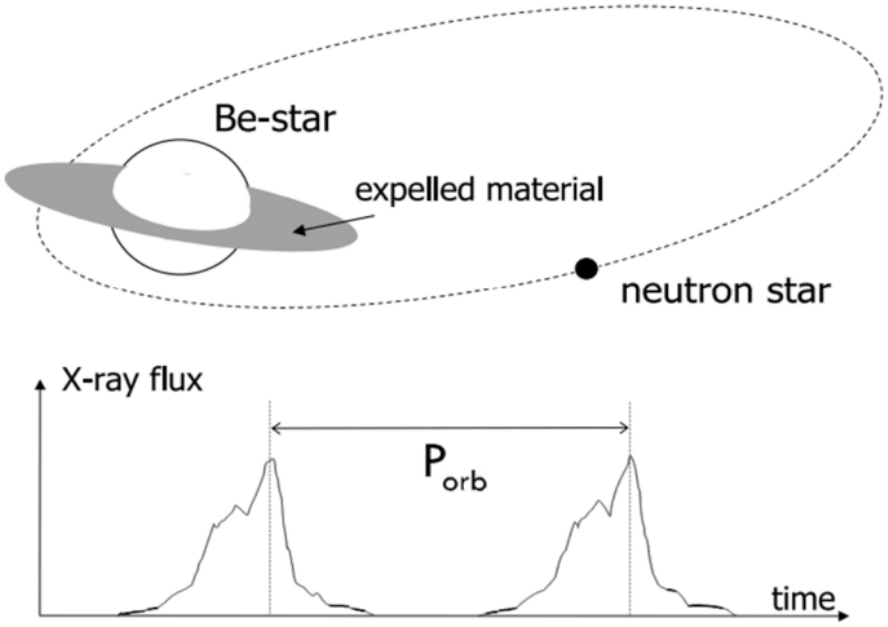

FIGURE 2.8: The diagram of orbital movements and optical variation of Be X-ray binaries. (Source: IOP Science)

- X-ray pulsations
- A hard 1-10 keV spectrum with a power-law energy index of 0-1

2.3.3.1 Be X-ray Binaries

The Be X-ray binaries are the most common type of HMXB. In neutron star binaries, the neutron star acts as the compact object while the Be star acts as the donor. As a result of their high rotational velocity, close to the break-up speed, Be stars produce a material ring around their equator. The emission lines that are the hallmark of Be stars are caused by this material. Most Be X-ray systems are short-lived but extremely intense. The orbital periods of these systems range from 12.7 to 262 days, which is considered to be moderately eccentric (Negueruela and Coe, 2002). Analysis of the arrival time of pulses is commonly used to determine the orbital period, a complementary technique to optical light curve analysis. Several distinct X-ray states can be observed in Be X-ray binaries. They might be invisible or visible at a persistently dim luminosity $L \ll 10^{36}$ erg s^{-1}.There are two main types of transient behaviour seen in these systems (Negueruela, 1998). Type I outbursts are extremely bright X-ray bursts lasting only a few seconds, with luminosities of $10^{36} - 10^{37}$ erg s^{-1}. Type II outbursts are massive X-ray bursts with luminosities greater than 10^{37} erg s^{-1} that last for a few weeks or longer. It is thought that the neutron star's passage through the Be circumstellar disc at periastron causes the periodicity of the type I outbursts. As a result, the period of orbit can be calculated from the time interval

between bursts. According to the proposed mechanism for type II outbursts, the Be star must exhibit high levels of activity in order to eject a large amount of matter, which then overwhelms the binary system. After receiving this massive amount of fuel, the neutron star becomes active for a period of weeks. As the outburst wanes, astronomers have noticed that the X-rays from Be X-ray binaries with the fastest spinning pulsars abruptly turn off. This has been blamed on the existence of a centrifugal barrier that prevents accretion below a critical mass accretion rate (Charles and Seward, 1995). The force exerted on the neutron star's magnetosphere decreases in proportion to the mass transfer rate. Because of this, the magnetosphere grows. This causes matter to be ejected from the magnetosphere because it enters at velocities greater than the orbital speed at that radius.

2.3.3.2 Supergiant X-ray binaries

Super Giant X-Ray Binaries (SGXB) are a type of HMXBs. This subclass contains an early type supergiant accreting via wind. They are bright and generally persistent and they constitute 20 percent of the known population of HMXBs and are primarily located within the Galaxy. These astronomical systems consist of an OB supergiant star and either a black hole or neutron star. The source of accreting material is typically the strong stellar wind of the donor star; however, for systems with orbital periods 2 days, Roche-lobe overflow can also occur, leading to the formation of a stable accretion disc around the compact object. In general, the orbital periods of these systems are shorter than those of Be X-ray binaries, ranging from 1.4 to 41.5 days. These orbital periods are identified through the observation of X-ray eclipses and the analysis of pulse arrival times. If the compact object is accreting material from the stellar wind, then the velocity of the stellar wind and the velocity of the compact object through the material limit the amount of material available for accretion.

Bondi-Hoyle accretion can be applied to supergiant X-ray binary stars if the stellar wind is assumed to be a uniform density material. The compact object will accrete material within a specified radius, r_{crit}. For a compact object of mass M_x, this radius is:

$$r_{crit} = \frac{2GM_{don}}{v^2} \tag{2.8}$$

where v is the velocity of the object, and M_{don} is the mass of the donor star.

Stellar winds are not homogeneous and uniform which results in large variations in the observed X-ray fluxes of HMXBs, resulting from both the variation in available accreting material and the variation in the absorption material density. This makes HMXBs a useful instrument for measuring stellar wind properties. In the case of eclipsing systems, the X-ray source can also serve as a probe of the supergiant donor star's atmosphere (Clark et al., 1994).

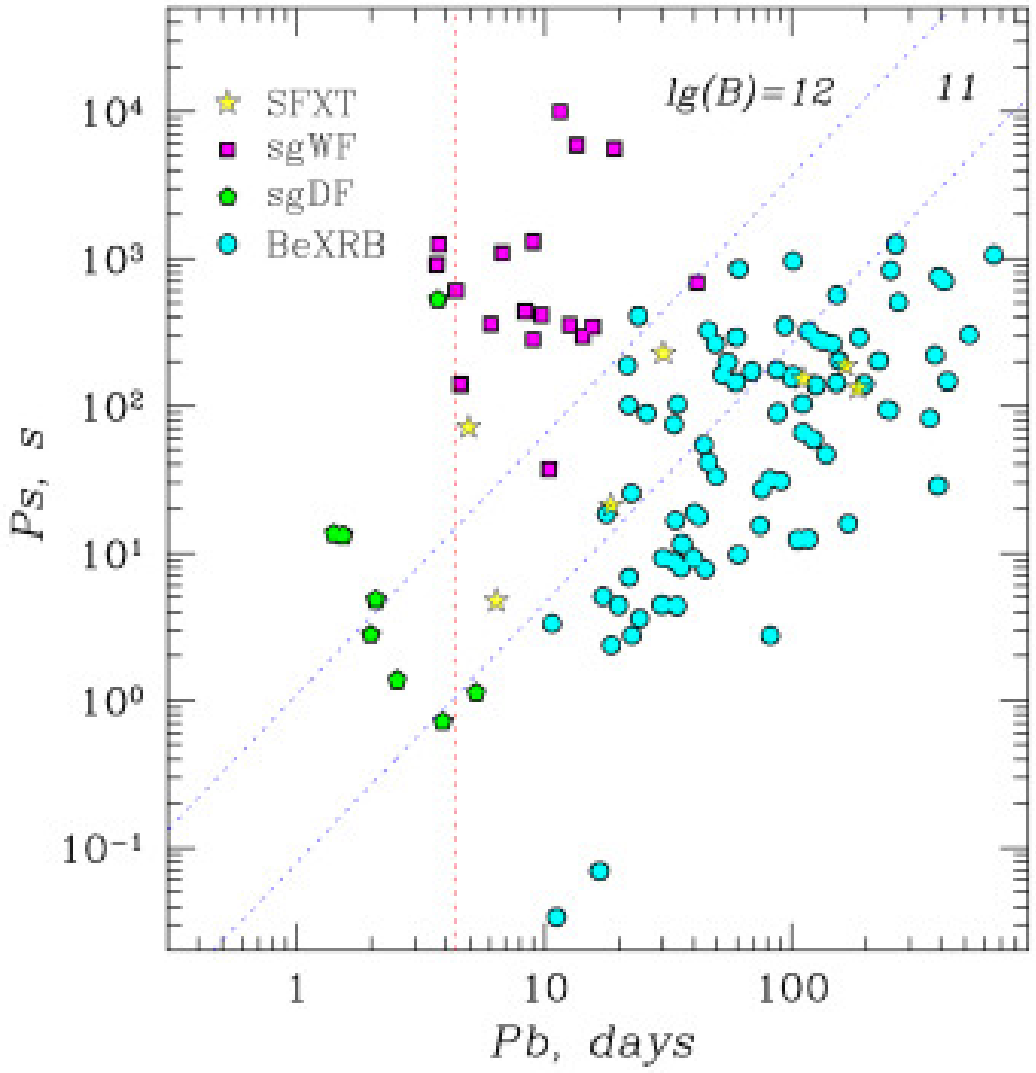

FIGURE 2.9: Spin-period (P_s) versus binary period (P_b) diagram ("Corbet diagram") for HMXBs. The different classes are distinguished by colour and symbol shape: SFXTs, disc-fed (sgDF) supergiant binaries (including ultra-luminous X-ray sources (ULXs)), wind-fed (sgWF) supergiant binaries, as well as Be X-ray Binaries (BeXRB). The vertical line indicates the binary period at which a $20R_\odot$, $22M_\odot$ supergiant fills its Roche lobe. Below the blue lines quasi-spherical accretion from the stellar wind for two different dipole magnetic field strengths B is inhibited by the centrifugal barrier, assuming a wind speed of 800 km/s (from Kretschmar et al., 2019)

The INTEGRAL mission detected new SGXBs with slow, weakly magnetised neutron star as companion (Sguera et al., 2006). These SGXBs are highly absorbed and display a mixture of persistent and variable emission. This subclass of SGXB is characterised by fast transient behaviour and have been dubbed Supergiant Fast X–ray Transients (SFXTs) – these objects are one of the most spectacular discoveries obtained by the INTEGRAL satellite. Their X–ray emission, characterised by short (typically lasting around 100-10,000 s) bright X–ray flares, is produced by a compact object (mostly a neutron star, given that X–ray pulsations have been detected in about a half of the members of the class), transiently accreting matter directly from the strong wind of the blue supergiant companion, although the mechanism producing the transient X–ray emission is still not fully understood.

INTEGRAL has also played a crucial role in the detection and characterisation of highly-absorbed persistent HMXBs (Manousakis and Walter, 2011). These systems generally have short orbital periods and long spin periods, and are characterised by strong and persistent soft X-ray absorption. INTEGRAL has been vital for establishing the orbital characteristics of these systems through timing analysis, and characterising the nature and extent of the absorbing material in these systems.

2.3.4 Transient Emissions

Type I X-ray bursts were first identified in 1975 and shortly thereafter these bursts were linked to LMXB systems and explained as thermonuclear flashes on the neutron star's surface. Due to their proximity and the neutron star's extreme gravity, the companion star's roche-lobe overflows and hydrogen is drawn into an accretion disc surrounding the neutron star.

This hydrogen is eventually accreted on the neutron star's surface, where, due to the extreme temperature and pressure, it is immediately transformed into helium. Once the helium reaches a critical mass a thin surface layer of helium ignites reaching temperatures of tens of millions of degrees and emitting a burst of X-rays. After the outburst has subsided, the binary system returns temporarily to its quiescent state, while the neutron star begins to reaccumulate the helium surface layer. The repeated process results in repeated X-ray bursts. This process is similar to recurrent novae; the compact object is a white dwarf that accretes a hydrogen-rich surface layer that undergoes explosive combustion to produce the outburst. Trasient X-ray emission is not limited to these X-ray bursts, but can also be observed in outbursts from LMXBs and HMXBs. In LMXBs, accretion continues after a burst until the accumulated layer of material reaches critical temperature and density to ignite a thermonuclear runaway to trigger the next burst.

Generally, X-ray bursts occur at regular intervals separated by hours or days. Their durations range from a few seconds to several minutes, with the burst profile displaying a rapid rise (0.3 to 10 seconds) followed by a slower decline (5 – 100 seconds). The rapid increase reflects the sudden increase in temperature caused by the explosive ignition of helium, while the slower decrease results from the slower cooling of the star's surface. All of the aforementioned pertains to Type I X-ray bursts, which constitute the vast majority of this type of object. Currently, only three sources have been observed to demonstrate Type II X-ray bursts, and one of these also exhibits Type I outbursts (van den Eijnden et al., 2022). The physical mechanisms for these burst types are described in the sections that follow.

The burst profile distinguishes a Type II burst from a Type I burst, which is believed to be the result of an increase in accretion rate from the companion star. Type I bursts exhibit a rapid increase followed by a gradual decline, whereas Type II bursts begin and end abruptly, with no gradual decay from the peak. They may also exhibit rapid burst successions separated by a few minutes.

2.3.4.1 Type I X-ray bursts

LMXBs have demonstrated Type I X-ray bursts.The maximum intensity of the bursts is observed between fractions of a second and ten seconds after the burst begins, and the bursts decay in 10 seconds to minutes. During this time, 10^{39} ergs of X-ray energy are typically emitted (Charles and Seward, 1995). Depending on the energy band in which the burst is observed, its profile will appear differently. The initial rise is concurrent across all bands, but the decay is more pronounced at higher energy regimes.This is a characteristic of thermal X-ray emission, in which the temperature of the burst begins high and decreases over time. When the neutron star accretes hydrogen from the accretion disc, a layer of hydrogen accumulates on its surface. Under the steady influence of fusion, this hydrogen layer begins to expand, forming a layer of helium beneath the hydrogen layer. This helium layer eventually begins to convert to carbon, but the process is rapid and unstable, resulting in a type I X-ray burst and a thermonuclear flash on the neutron star. In 10 seconds, the surface of the neutron star cools from an average of 30 million degrees to 15 million degrees (Charles and Seward, 1995). The system then begins to accrete matter once more, eventually leading to a second thermonuclear flash.

During the decay of the burst, the X-ray spectrum represents a cooling blackbody (Swank et al., 1977). If the distance to the system is known, the radius of the neutron star can be estimated from the flux and temperature of the blackbody emission. Using the X-ray flux, F_x, from a type I burst at a distance d due to the blackbody emission of

a sphere with a radius R and a temperature T Stefan's law yields:

$$4\pi d^2 F_x = 4\pi R^2 \sigma T^4 \tag{2.9}$$

Therefore, the radius of a neutron star can be calculated as:

$$R = d \left(\frac{F_x}{\sigma T^4} \right)^{1/2} \tag{2.10}$$

This results in a radius of 10 km for sources with reasonable distance estimates, the typical value associated with neutron stars. A careful examination of the X-ray spectrum during a burst reveals that the temperature of the blackbody and, consequently, the radius change. This is explained by an initial expansion followed by a contraction of the neutron star photosphere. Radiation pressure forces the hydrogen layer to expand as the X-ray luminosity reaches the Eddington limit. The luminosity can then continue to rise because the Eddington limit for helium is 1.75 times greater (Charles and Seward, 1995). The Eddington Limit for a neutron star of mass $1.4 M_\odot$ is 1.8×10^{38} erg s^{-1}; therefore, the luminosity of a type I X-ray burst must not exceed this value. Therefore, type I X-ray bursts can provide distance estimations. This is, however, always an upper limit, as the peak flux of a burst need not have reached the Eddington limit. Type I bursts cannot occur in LMXBs where the compact object is a black hole because hydrogen must accumulate on a surface for them to occur. Type I X-ray bursts are therefore conclusive proof that the compact object is a neutron star.

2.3.4.2 X-ray type II bursts

An X-ray source exhibiting rapidly repeating X-ray bursts was identified early on in the study of X-ray bursting sources. On timescales as short as 10 seconds, bursts were observed to recur, with the brightest bursts being 1,000 times brighter than the faintest. Initially, this behaviour posed a threat to the thermonuclear flash model of type I bursts, until it became apparent that it had a different origin. These bursts do not display random behaviour. There are longer intervals between the more luminous bursts than the less luminous ones. The X-ray spectra of these type II bursts show no indication of a cooling blackbody. As the time-averaged emission of these bursts is comparable to or greater than that of the persistent emission, it is believed that the bursts are caused by a sudden release of material onto the neutron star's surface as a result of an unstable accretion flow. This resembles the behaviour of a relaxation oscillator. Despite many years of observations and numerous theoretical models, the actual cause of the bursts remains a mystery. The lack of identification of the gating mechanism that dumps matter onto the neutron star constitutes the primary obstacle to understanding these sources (Lewin et al., 1997). Numerous models have been

developed to explain the rapid burster behaviour and the origin of accretion instability.

The presence of a neutron star with a strong enough magnetic field to produce a magnetosphere, which acts as a gate through which the accreting material must pass, is one of the models proposed to explain the origin of these bursts (Lamb, 1977). At the gate, accreted matter accumulates until the magnetosphere can no longer withstand the pressure, at which point the material falls through onto the neutron star. Explaining the correlation between burst strength and time between bursts, when a large amount of material passes through, resulting in a bright type II burst, it takes time for a sufficient amount of material to build up again, which explains the relationship between burst strength and time between bursts. The accretion of matter via a gating mechanism does not, however, prevent a layer of hydrogen from eventually accumulating on the neutron star's surface and causing a thermonuclear flash. Consequently, these rapid burster sources are capable of producing both type I and type II X-ray bursts.

2.3.4.3 X-ray pulsations

Typically, X-ray pulsations are observed from HMXBs when the compact object is a neutron star. Due to the young age of these binary systems, neutron stars invariably possess powerful magnetic fields. Consequently, as matter accretes onto the neutron star, there will come a time when the magnetic field predominates; this point is the magnetospheric radius. At this point, an accretion disc will be disrupted, and material will begin to move along magnetic field lines. The material will reach the magnetic poles of the neutron star, where it will generate a hot shock upon impacting the surface. This shock is responsible for the X-ray emission in a beam. If the magnetic axis of the neutron star is not aligned with its spin axis, then this beam will pass in and out of our line of sight, causing us to observe pulsations.

It was discovered that the spin period, P_{pulse}, and orbital period, P_{orb}, of various HMXB system types were correlated in various ways (Corbet, 1986). This is depicted in the Corbet diagram depicted in Figure 2.9. The accretion discs that may form in these systems are capable of applying a torque to the neutron star and altering its spin frequency. There is a positive correlation between the spin and orbital periods in Be X-ray binaries. There is an anticorrelation between spin and orbital period in supergiant systems where the donor is filling its Roche-lobe. Finally, it is observed that all supergiants with an underfilled Roche lobe have long spin periods, but there appears to be no correlation with orbital period. This is not surprising, as stellar wind accreting systems have much less available angular momentum for transfer.

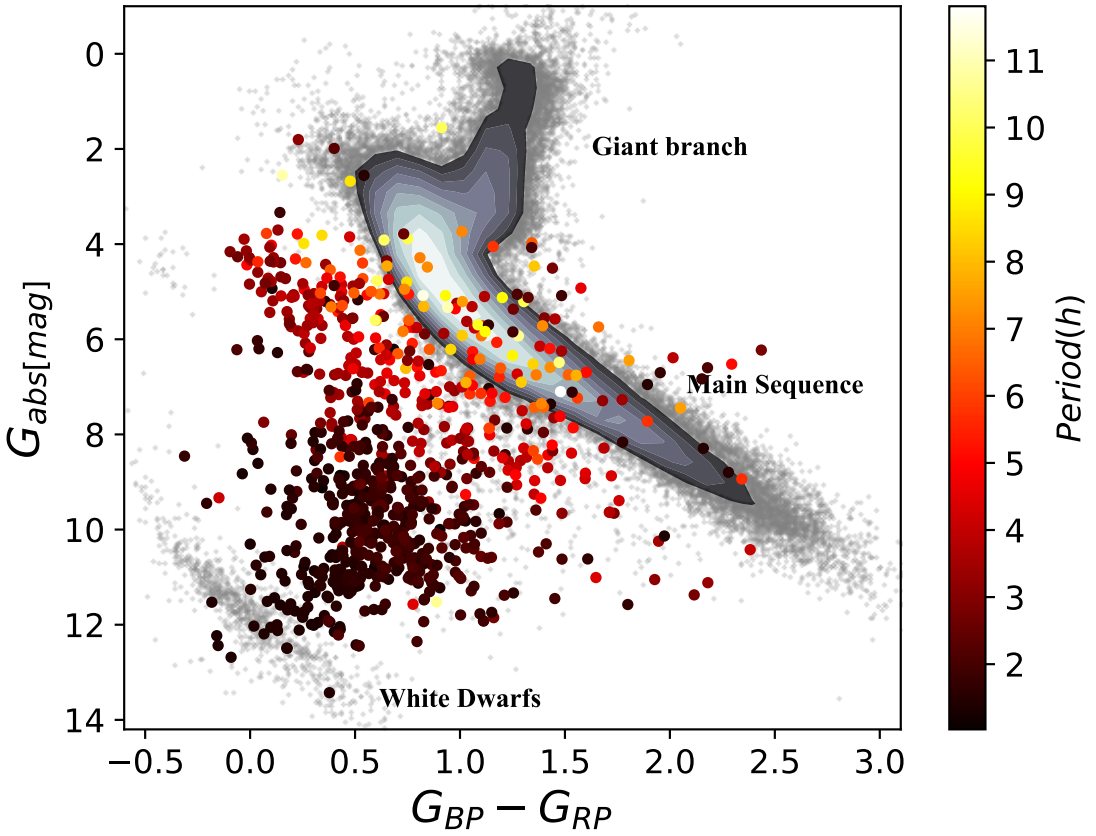

FIGURE 2.10: CVs in the *Gaia* colour-magnitude diagram, colour coded by orbital period, from Abril et al. (2020)

As the accretion rate increases, the angular momentum flow from the accretion material is anticipated to increase. As a result, the pulsar is spun faster. At very low mass accretion rates, it is believed that neutron stars lose angular momentum as magnetic field lines that couple to the outer, slower accretion flow disc remove angular momentum (Psaltis and Norman, 2000).

2.3.5 Cataclysmic Variables / Accreting White Dwarfs

In terms of composition, cataclysmic variables (CVs) are very similar to X-ray binaries. The primary distinction between CVs and XRBs is that the compact object is a white dwarf (WD), leading to the newly preferred label of "accreting white dwarfs". In these systems, the donor star is usually a late-type star on or close to the main sequence whose Roche-lobe is being filled. Typically, these systems have extremely close separations, with orbital periods ranging from 80 minutes to a few hours (Córdova, 1995). Figure 2.10 shows the location of known CVs in the colour-magnitude diagram from *Gaia*, illustrating the variety of donor stars and orbital periods (Abril et al., 2020). CVs exhibit a vast array of behaviours that are mainly controlled by the rate of mass transfer from the donor and the magnetic moment of the WD.

It has been observed that all classes of CVs emit X-rays, although the properties of the emission vary between classes (see, e.g., Mukai, 2017, for a review). For WDs with a magnetic field of less than 10^4 Gauss, the system is nonmagnetic. In these systems, as in the case of LMXBs, the donor star fills its Roche-lobe and an accretion disc forms around the WD. The magnetic field has minimal influence on the accretion flow. For low accretion rates through the disc, the disc temperature and viscosity are too low to transfer mass through the disc at the same rate as it arrives. Consequently, the disc serves as a gas reservoir (Kuulkers et al., 2004).

When the disc reaches the critical density, it becomes optically thick, gets hot, and rapidly transfers mass to the surface of the WD. This causes Dwarf Nova explosions that can last weeks. In magnetic systems, the magnetic field is strong enough to disrupt the donor accretion flow. Polars have around 10^6 Gauss magnetic fields, and the rotation of the WD is phase-locked with the binary system (Córdova, 1995) Consequently, material flows from the donor star to the WD along the field lines.

In the case of Intermediate Polars, the WD spins faster than the orbital period of the binary, and a partial accretion disc may form. Variations in their mass accretion rates are reflected in the luminosities of both types of magnetic CVs. It is believed that nearly all CVs build up a layer of hydrogen on their surface, which eventually undergoes thermonuclear runaway and results in a Classical Nova explosion. In some binary systems this process repeats itself resulting in Recurrent Novae. Similar to the type I X-ray bursts observed in LMXBs, this process produces radioactive isotopes with distinctive gamma-ray line signatures.

2.4 Extragalactic sources

Besides all of the galactic X-ray sources there are numerous sources of high energy emission which are external to the Galaxy.

2.4.1 Supernovae

Supernova explosions are the most energetic stellar phenomena known, with an average energy output of 10^{51} erg. The majority of this energy is released as kinetic energy, with initial velocities between 10,000 and 15,000 km s^{-1} (Charles and Seward, 1995). This material's propagation causes the ejected material and surrounding medium to reach extremely high temperatures, resulting in X-ray emission.

2.4.1.1 Type Ia Supernovae

All Type Ia supernovae exhibit nearly identical light curves, indicating a similar progenitor star and mechanism. They are found in binary systems containing a white dwarf. The WD accumulates matter until the electron degeneracy pressure can no longer support its mass. The collapse raises the core's temperature to the point where carbon and oxygen begin to fuse. This causes an explosive wave to propagate through the WD, shattering it. Numerous tonnes of radioactive substances, including ^{56}Ni, are produced in the supernova explosion, and this has been observationally confirmed with gamma ray nuclear decay lines detected by INTEGRAL (Churazov et al., 2015). X-ray emission from Type Ia supernovae is traditionally not expected, but may arise if there is a substantial amount of circumstellar material. To date, no X-rays have been detected from unambiguous Typa Ia, but have been detected from SN2012ca (Bochenek et al., 2018) which is unclear as to whether it is a Type Ia supernova or a core collapse supernova. It is worth noting that Galactic (and Magellanic Cloud) supernova remnants from very old supernovae do emit X-rays from electron de-excitation in heavy elements (e.g. Fe).

2.4.1.2 Type II Supernovae

Type II supernovae reach their maximum luminosity more slowly and are typically less bright than Type Ia supernovae. There appear to be more variations in the light curves of these systems, indicating a progenitor with greater individuality. The optical spectra are dominated by broad hydrogen emission lines, indicating that these supernovae originate from young, massive stars with hydrogen-rich envelopes (Charles and Seward, 1995). The star is typically a red giant or supergiant and has undergone multiple levels of fusion, with Fe at its core, surrounded by Si, which is surrounded by lighter and lighter elements. Eventually, the core reaches a pressure at which the Fe begins to decay into lighter nuclei, lowering the core's pressure and causing it to contract. This decrease in pressure eventually causes the core to collapse due to its own gravity, a phenomenon known as a runaway effect. The energy release generates an outward-propagating shock wave that rips the star apart.

X-ray emission has been detected in numerous Type II supernovae, and arises from shock emission from dense circumstellar material surrounding the exploding star. Figure 2.11 (from Quirola-Vásquez et al., 2019) shows the X-ray light curves for numerous Type II supernovae, which can be used to place constraints on the dying star's circumstellar material. Galactic remnants of ancient Type II supernovae (e.g., the Crab and Cas A) also emit X-rays via electron de-excitation, as do some of their compact remnants (e.g., the Crab pulsar).

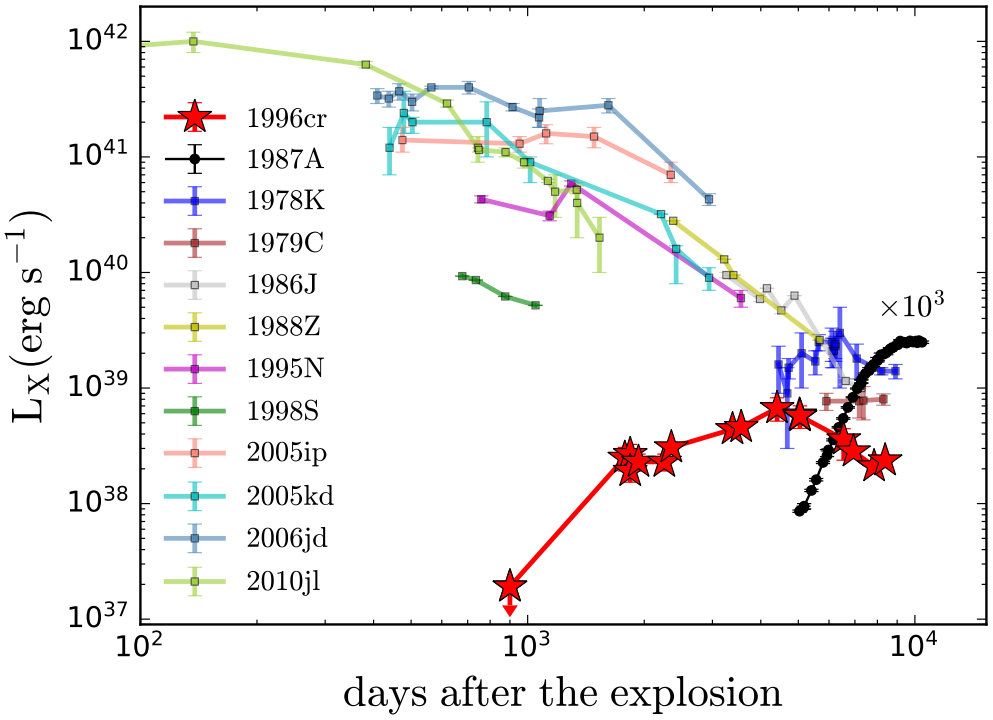

FIGURE 2.11: X-ray light curves (covering 0.5-10 keV) of Type II supernovae (from Quirola-Vásquez et al., 2019).

2.4.2 Active Galactic Nuclei

Active Galactic Nuclei (AGN) are extremely luminous and variable galaxies in which a significant amount of energy is released from the galactic nucleus. It is believed that the source of this emission is the accretion of matter onto a supermassive ($10^8 M_\odot$) black hole in the centre of the galaxy. Seyfert galaxies, Quasars, Blazars, and radio galaxies are among the many AGN subclasses that have been derived based on their optical and radio properties. There are two classifications for Seyfert galaxies: type 1 and type 2. Seyfert galaxies exhibit optical emission lines with widths of one thousand kilometres per second. In addition, Seyfert 1 galaxies display broad lines with widths 10,000 km s-1 (Antonucci, 1993). Quasar optical spectra are very similar to those of Seyfert galaxies, indicating a common emission mechanism; in fact, a distant, bright Seyfert galaxy would appear to be a quasar (Charles and Seward, 1995). In contrast, the spectra of blazars are completely devoid of optical features.

Looking at these objects in the high energy band reveals a unique set of characteristics. Seyfert galaxies are observed to emit the majority of their high energy emission up to 100 keV, indicating that the source of this emission is most likely thermal processes and providing insight into the AGN's physical environment. Quasars can be significantly brighter at higher energies and are primarily detected at energies greater

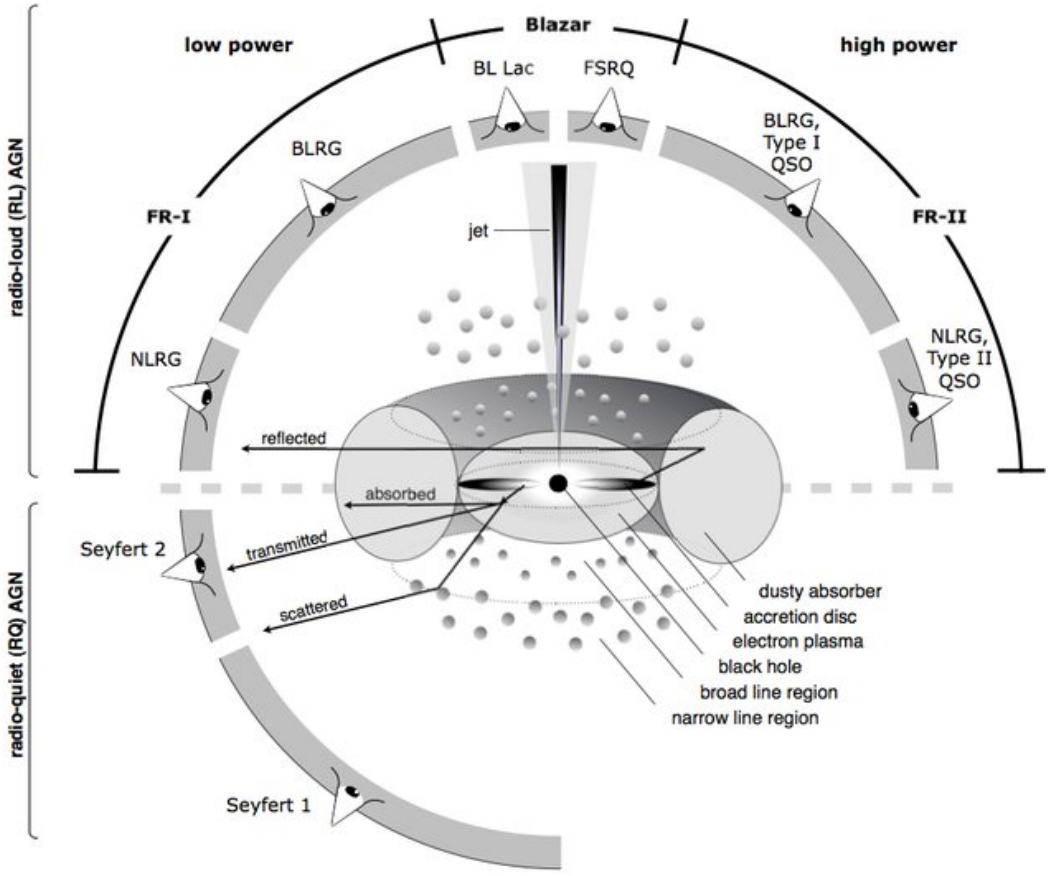

FIGURE 2.12: The unified model of Active Galactic Nuclei (AGNs). Credit: Beckmann and Shrader (2012)

than 100 MeV. It is believed that the source of this emission is a powerful jet visible along our line of sight to the AGN. In the case of Blazars, it is believed that we are looking directly down the AGN jet, resulting in the jet emission dominating what we observe.

It has been proposed that the properties of each AGN class are a result of observing the same object from different vantage points, despite the fact that each class appears to be observationally distinct (Antonucci, 1993). This concept is represented graphically in Figure 2.12. Thus, it is believed that all AGN are composed of a supermassive black hole surrounded by an accretion disc and a hot corona. Existing beyond this is a cold molecular torus of material. When viewed through the molecular torus, the object appears to be a Seyfert 2, but when viewed directly down the jet, it appears to be a Blazar. If there is no jet in the system, the source will appear radio silent.

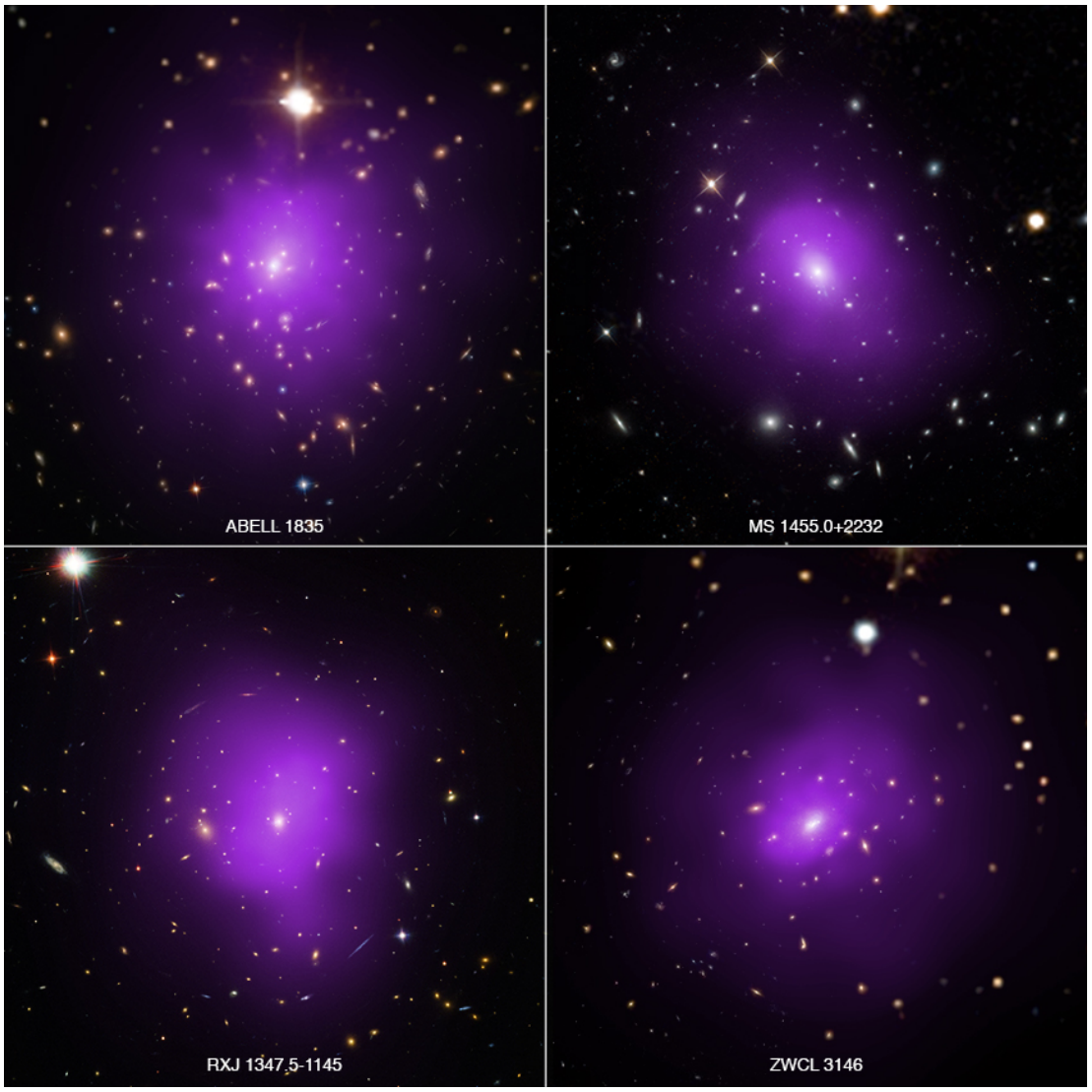

FIGURE 2.13: X-ray emission (purple) for several galaxy clusters observed by *Chandra* (Source: Chandra photo album at chandra.harvard.edu).

2.4.3 Galaxy Clusters

Galaxy clusters vary in size from a few to several thousand galaxies. These galaxies are gravitationally bound, and the space between them is filled with extremely hot gas. This gas has temperatures greater than 10^7 Kelvin and emits diffuse high-energy radiation (Westmore, 2002). The gas emits a thermal Bremsstrahlung spectrum with kT ranging from 1 to 10 keV (see examples in Figure 2.13). Several clusters have demonstrated an excess of emission at energies greater than 20 keV. There are two possible explanations for this excess: the inverse Compton scattering of cosmic microwave background photons and the presence of an active galactic nucleus within the cluster (Sarazin, 1986).

2.4.4 Gamma-Ray Bursts

Gamma-ray bursts (GRBs) are transient events that are incredibly bright and brief. When a GRB explodes, it is brighter than the entire gamma-ray sky. The duration of a GRB ranges from milliseconds to minutes, and its light curve and spectrum are unique for each event. They were initially discovered by the Vela-B military satellites in the 1960s and were not well understood until recently. The BATSE experiment aboard the Compton Gamma-ray Observatory made the first significant breakthrough in determining the origins of these sources. BATSE was an all-sky monitor whose primary scientific objective was to detect and locate GRBs (Fishman et al., 1985).

The distribution of the bursts across the sky was found to be isotropic regardless of the selection criteria. This indicated that GRBs originated outside the galaxy. BeppoSAX's 1997 detection of X-ray afterglows, which lasted for weeks after the initial burst, was another significant advance (Piro et al., 1995). Observations of X-ray afterglows have enabled the identification of more precise source positions and optical band follow-ups. Since then, the host galaxies of GRBs have been detected. GRB origins are still not completely understood. Nonetheless, two fundamental theoretical models exist to explain these objects. The first scenario involves the collision of two compact objects, such as a neutron star and a black hole. This event is believed to release sufficient energy to account for the observed GRB properties. The second example is a hyper-nova. This is the collapse of a colossal star, which leads to the formation of a black hole. Recent studies have identified the characteristics of supernovae in the optical afterglows of GRBs, making this an excellent candidate for GRBs.

The use of multiple astronomical observatories to gather information about the universe is becoming increasingly important. It is anticipated that a number of well-known astrophysical sources will also emit other types of 'messengers' in addition to photons. These 'messengers' include cosmic rays, gravitational waves, and neutrinos; bringing more information in addition to that which is carried by electromagnetic radiation. Together, they help to draw a complete phenomenological picture of a number of astrophysical events and measure key cosmological parameters. It is widely considered that multi-messenger astronomy is the only method that will be capable of shedding light on a number of fundamental facets of both physics and cosmology. The most recent groundbreaking discoveries of a gravitational wave source associated with a short gamma-ray burst and of a neutrino event found to be spatially consistent with a flaring blazar have already demonstrated the major role that high-energy sources will play in multi-messenger observations.

2.5 Introduction to Astrophysics surveys

An astronomy survey is a map or image of a portion of the sky (or the entire sky) for which there is no particular observation target. Alternatively, an astronomy survey may consist of a collection of images, spectra, or other observations of objects with a shared kind or characteristic. Due to instrumentation restrictions, electromagnetic spectrum surveys are frequently limited to a single band; however, multiwavelength surveys can be conducted by employing numerous detectors, each sensitive to a distinct bandwidth. Generally, surveys have been conducted as part of the development of astronomy catalogues. They may also look for ephemeral astronomical phenomena. They frequently employ wide-field astrographs. An astronomy catalogue is a list or tabulation of celestial objects, often grouped together because they share a common type, morphology, origin, technique of detection, or method of discovery. Astronomical catalogues are usually the outcome of an astronomical survey of some kind.

2.5.1 Data Products

The main deliverables of all astrophysics surveys are data products which can be utilised for scientific analyses. These products begin with low-level data products like images, which in turn can be used to produce higher level data products like mosaic images, spectra, and light curves. High-level products like source catalogues can also be produced from the lower level data products. The production of these products and their uses are described below.

Images: Astronomical images present flux (and typically also flux error) as a function of position on the sky. Astronomical imaging instruments typically collect raw flux in many pixels on a distributed detector, and these raw flux values are processed to correct for instrument sensitivity. For focusing telescopes the position on the detector is directly related to position on the sky, but for coded aperture masks telescopes the flux received on the detector is related to the 'shadow' of the coded aperture mask (see for example INTEGRAL described in the next section). An example INTEGRAL image is shown in Figure 2.14.

Spectra: Spectra present specific flux (flux per unit energy / frequency / wavelength) as a function of photon energy for a single source. For detectors that can measure the energy of a detected photon, this consists of counting photons in each energy bin at a fixed point on the sky, and correcting for any instrument sensitivity versus energy. Object spectra are extremely useful because they can be directly compared to models for physical mechanisms that produce a spectrum for a source. An example spectrum using both ISGRI and SPI on INTEGRAL is shown in Figure 2.15.

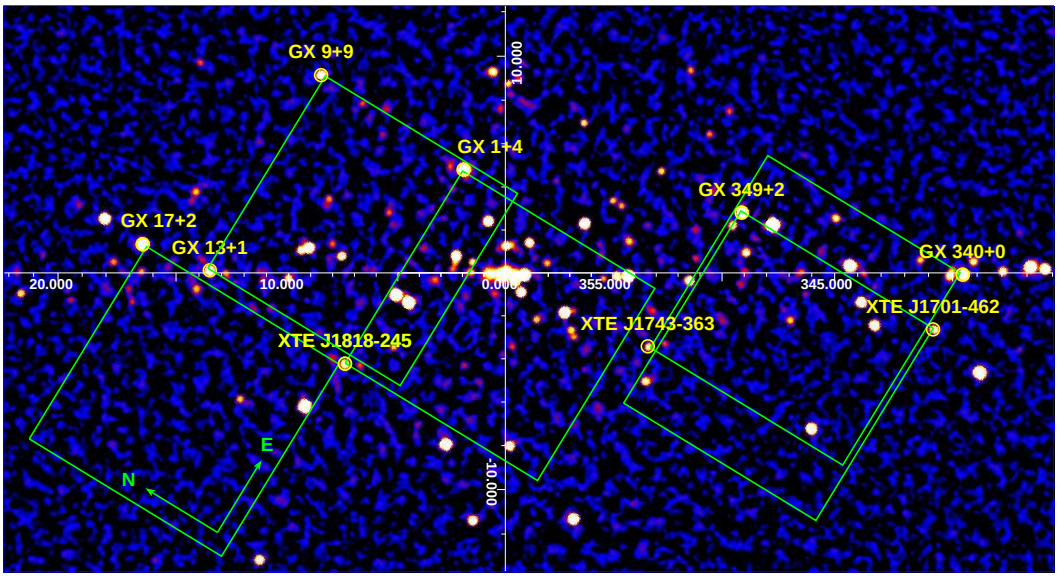

FIGURE 2.14: Example INTEGRAL image of sources on the Galactic Plane. spanning approximatley $20^o \times 80^o$. Green squares note locations of "ghosts" of bright sources caused by the INTEGRAL coded mask. From Krivonos et al. (2010)

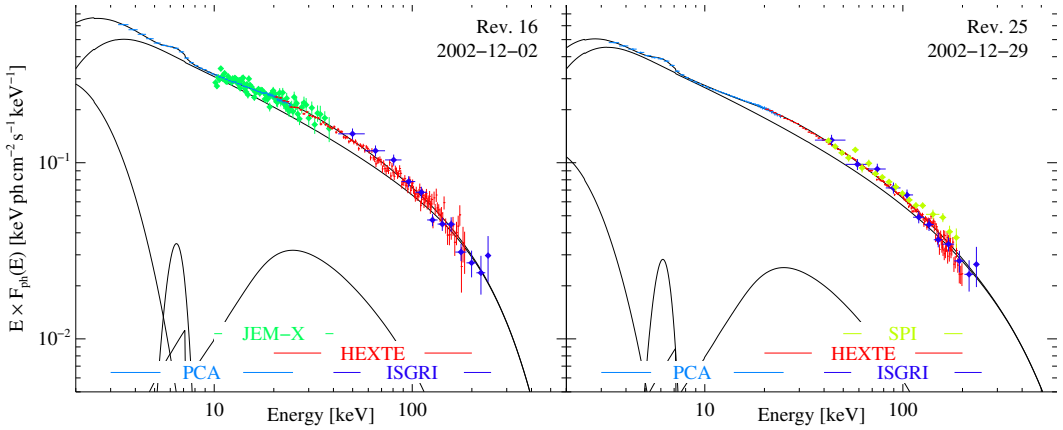

FIGURE 2.15: Spectrum of Cygnus-1 using numerous instruments, including IBIS and SPI onboard INTEGRAL. From Wilms et al. (2004)

Light Curves: Light curves are the flux of the same source as a function of time. These are produced by collecting flux measurements for the same source across images taken at different times (once you make sure it's the same source). Light curves are useful for comparison to models that predict how a source's flux should evolve, and also for analysing patterns of emission that can reveal characteristic timescales of a system (e.g. accreting binary systems). An example INTEGRAL light curve for the 2015 outburst of V404 Cyg is show in Figure 2.16.

Catalogues: catalogues are a higher level data product produced by astrophysics surveys built on low-level image products. Spatial and temporal searches are applied to the images to measure properties of sources in the images for inclusion in the

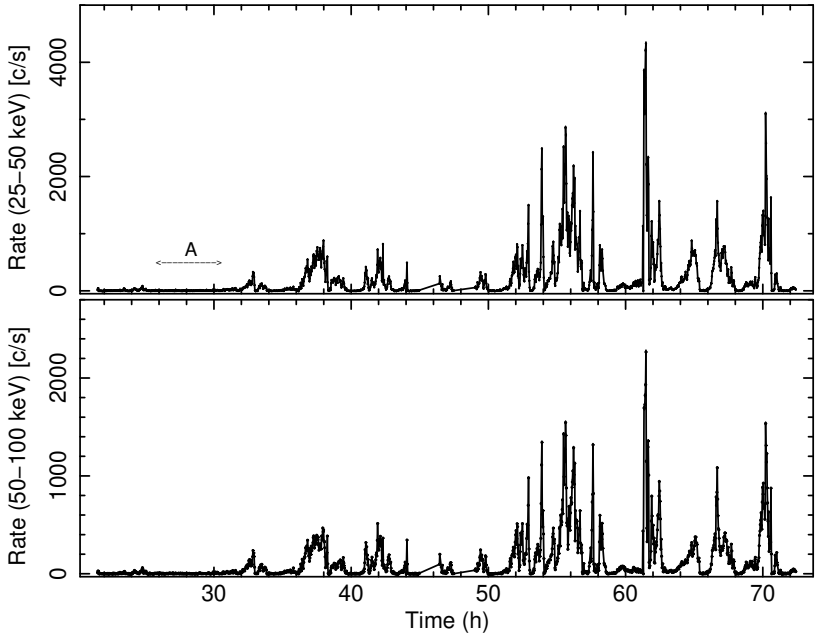

FIGURE 2.16: Light curve of V404 Cyg 2015 outburst. From Natalucci et al. (2015)

catalogue. Typical source properties included in catalogue are source position, flux, error (or significance), and potentially source classification.

2.5.2 Background to HE surveys

Since the beginning of X-ray astronomy, numerous X-ray surveys have been successfully carried out to discover new types of X-ray emitters and investigate the nature of the Cosmic X-Ray Background.

The first ever all-sky survey was carried out by Uhuru, also known as the Small Astronomical Satellite 1 (SAS-1); it had a sensitivity of 103 times the Crab intensity and contained 339 X-ray sources with observed energy between 2 and 6 keV (Forman et al., 1978). One of the most prevalent types of discovered objects was a binary star system, followed by supernova remnants, Seyfert galaxies, and clusters of galaxies.

Detailed investigations of the cosmic X-ray sky were conducted by Markert et al. (1979) using the MIT 1–40 keV X-ray detectors aboard the OSO 7 satellite from October 1971 to May 1973. The authors identified the intensity limits or upper bounds for 3rd Uhuru and OSO 7 listed sources in multiple energy bands, one of which being the 15–40 keV range, making this the first major hard X-ray catalogue (Giacconi et al., 1974; Markert et al., 1976, 1977).

The first attempt to survey the sky at MeV energy range (26 - 1200 keV) was accomplished with the Sky Survey Instrument aboard Ariel V during 1974 -1979 (Coe et al., 1982). This survey also discovered the first galaxy LogN-LogS relation at energies above 100 keV.

Skinner et al. (1987) produced the first observations of the Galactic Center in high-energy X-rays up to 30 keV by utilising a coded mask X-ray telescope launched on the Spacelab 2 mission.

An all-sky survey in the hard X-ray band (2.8 to 30 keV) was carried out with the assistance of the BeppoSAX Wide Field Camera (Jager et al., 1997), which observed twenty-one transients in the galactic centre region and more than fifty transients and recurring sources along the galaxy's plane.

Levine et al. (1984) published the findings of the first comprehensive analysis of X-ray sources at high X-ray energy (13–180 keV) that was conducted across the whole sky, which were derived from data that was acquired between August 1977 and January 1979 by the UCSD/MIT Hard X-Ray and Low-Energy Gamma-Ray Instrument A4 aboard the HEAO 1 satellite. The survey includes 72 sources with flux sensitivities between 10 and 15 mCrab.

TTM/COMIS, a coded-mask imaging spectrometer, was mounted on the Kvant module of the MIR orbiting station in 1989 and used to take observations of the Galactic Center region (Sunyaev et al., 1991b; Sunyaev, 1991). In order to ascertain the component of these sources that is responsible for the emission of hard X-rays, several measurements of galactic sources were carried out (Sunyaev et al., 1989, 1991a; Kaniovsky et al., 1997; Borkous et al., 1997).

Hard X-ray observations of the Large Magellanic Cloud (LMC) were also obtained with the TTM/COMIS instrument. These observations included monitoring and spectral observations of LMC X–1, LMC X–2, LMC X–3, LMC X–4, and PSR 0540-693. These observations were obtained at energies ranging from 2–30 keV. A database of 67 X-ray sources with a confidence level of greater than four was published. These X-ray sources were observed by the TTM/COMIS telescope between 1988 and 1998 (Emelyanov et al., 2000).

Between 1990 and 1992, the ART-P coded-mask telescope aboard the KVANT observatory made observations of more than 400 sky fields spanning an energy range of 2.5–60 keV (Sunyaev et al., 1990).

Pavlinsky et al. (1994) surveyed the GC using a 5x5 detector array (Art-P/Granat) and found it possible to detect 12-point X-ray sources with a sensitivity of 1 mCrab in the 3-17 keV energy range.

At higher energies, the SIGMA telescope aboard GRANAT could observe more than a fourth of the sky with a sensitivity greater than 100 mCrab between 1990 to 1998 (Revnivtsev et al., 2004). The SIGMA telescope (Paul et al., 1991) made it possible to recreate the first images of the hard X-ray sky with an angular resolution of 15 deg in the energy range of 35-1300 keV, which was accomplished using the SIGMA telescope and identified 37 hard X-ray sources with energy ranging from 40 to 100 keV (Revnivtsev et al., 2004).

2.6 Hard X-ray surveys in the 2000s

During the 2000s, significant progress was made in this field due to the launch of two space-based X-ray telescopes: INTEGRAL and *Swift*.

INTEGRAL was launched in 2002 and is specifically designed for hard X-ray imaging and spectroscopy. It has a large field of view and high sensitivity to hard X-rays, making it ideal for studying various sources, including active galactic nuclei, binary stars, and supernova remnants.

The *Swift* satellite was launched in 2004 and is primarily designed to study GRBs, the most luminous explosions in the universe. It also has the capability to perform X-ray surveys of the sky, allowing astronomers to study a wide range of high-energy astronomical sources.

One notable survey performed by INTEGRAL is the *Swift*/INTEGRAL All-Sky Hard X-ray Survey (SIX), described in a paper by Bottacini et al. (2012). This survey used data from the *Swift* and INTEGRAL satellites to perform a deep, wide-field survey of the hard X-ray sky. The survey covered an energy range of 20-100 keV. It was able to detect more than 100 previously unknown hard X-ray sources, providing a wealth of new information on high-energy astronomical objects.

2.6.1 INTEGRAL

The European Space Agency's (ESA) International Gamma-Ray Astrophysics Laboratory (INTEGRAL) was launched in 2002 and has performed over 20 years of observations in the energy range five keV to 10 MeV. It has a complement of three primary high-energy instruments: the SPectrometer on INTEGRAL (SPI), the Joint European X-ray Monitor (JEM-X), and the Imager on Board INTEGRAL (IBIS). IBIS comprises the INTEGRAL Soft Gamma Ray Imager (ISGRI) and Pixellated Imaging CaeSium Iodide Telescope (PICsIT) detectors surrounded by an active veto. IBIS/ISGRI uses a tungsten-coded mask to determine source astrometry through de-convolution of the mask pattern (shadow) projected onto the detectors.

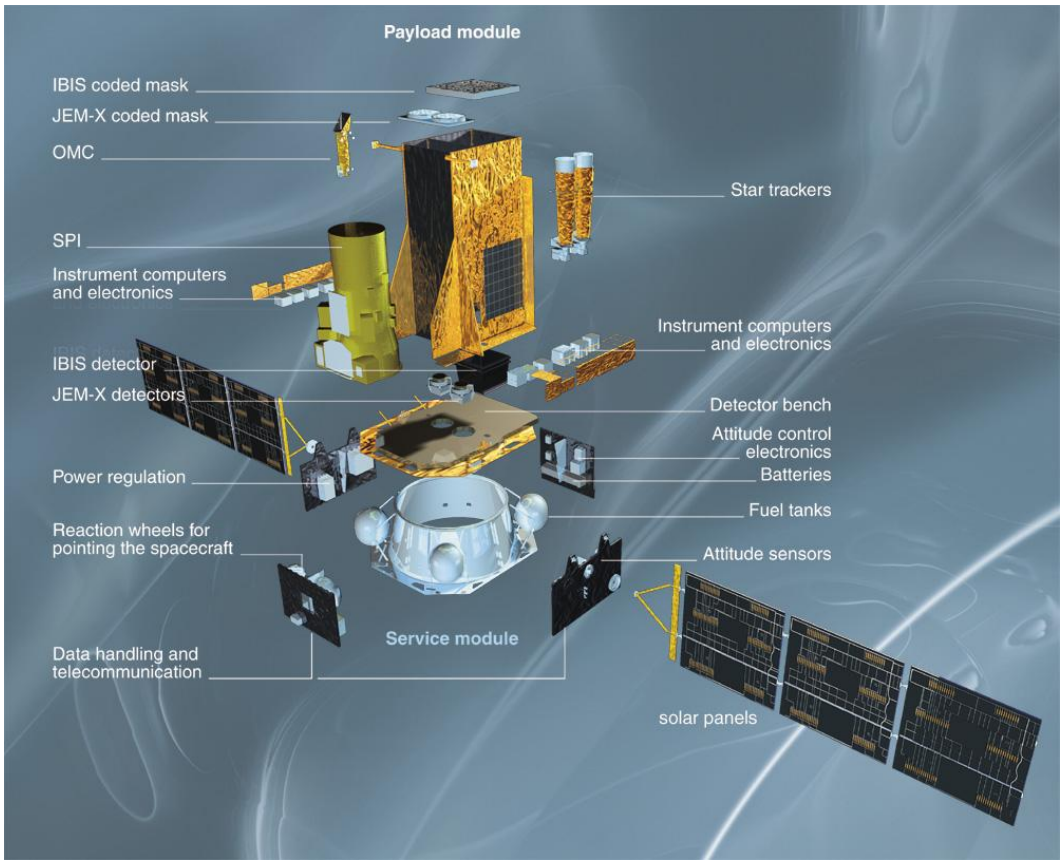

FIGURE 2.17: The INTEGRAL spacecraft with its four scientific instruments: INTEGRAL spacecraft model The gamma-ray spectrometer SPI, The gamma-ray imager IBIS, The two X-ray monitors JEM-X, and the optical monitoring camera OMC. (Source: ESA)

The IBIS telescope (Ubertini et al., 2003) is the most suitable device on board INTEGRAL for imaging surveys in the hard X-ray range. This device has the optimal field of view, sensitivity, and angular resolution for a rapid wide-angle scan of the sky, which is optimised with the scientific objective of regularly monitoring a large portion of the Galactic plane and discovering the majority of expected transient sources, whose existence was predicted by X-ray missions such as BeppoSAX and RXTE operating in the 1990s at lower energies and coarser spatial resolutions. ISGRI (INTEGRAL Soft Gamma Ray Imager; Lebrun et al., 2003) is a low-energy detector layer comprised of a pixelated 128×128 CdTe solid-state detector that sees the sky through a coded aperture mask. IBIS/ISGRI produces images of the sky with an FWHM resolution of 12 arcmin over a 28×28 degree field of view and a working energy range of 15 to 1000 keV. In addition to ISGRI, IBIS is also equipped with the Pixellated Imaging Caesium Iodide Telescope (PICsIT Labanti et al., 2003) which is sensitive to a higher energy range of 175 keV to 20 MeV. This thesis focuses on data obtained by IBIS/ISGRI.

The IBIS telescope employs coded-aperture imaging. Because gamma-rays cannot be deflected like lower energy photons, taking images of the gamma rays requires a

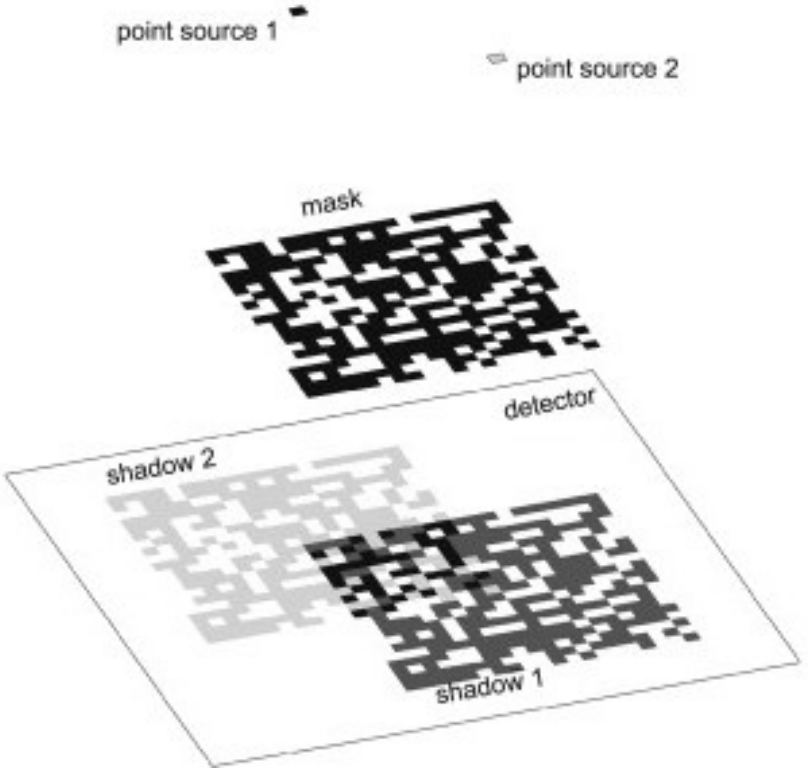

FIGURE 2.18: Demonstration of how coded masks can be used for source localisation by using source shadows on the detector. (Source: Schultz et al., 2009)

different approach. Coded aperture masks achieve this using a distinct pattern of material that blocks gamma rays, so that gamma ray light from a source produces a shadow in the pattern of the mask. This enables source localisation (see Figure 2.18). The coded mask pattern for INTEGRAL is shown in Figure 2.19.

The IBIS telescope incorporates a very large fully-coded FOV of 9×9 deg (all source radiation is controlled by the mask) and a partially-coded FOV of 28×28 deg as a result of its coded-aperture design (only a fraction of source flux is modulated by the mask). In addition, the "dithering" pattern around the nominal target position, a controlled and systematic spacecraft dithering manoeuvre introduced to minimise systematic effects due to spatial and temporal background variations in the spectrometer's (SPI) detectors, results in a sky coverage that is even more extensive. The effective latitude coverage of the Galactic plane is $|b| \leq 17.5$ due to the combination of the normal 5×5 dithering grid and multiple INTEGRAL pointings made possible through the approved Guest Observer Program at the Galactic X-ray sources (see Krivonos et al., 2012). Consequently, INTEGRAL is able to do time-resolved mapping of the Galactic plane on a scale of one year. This makes it possible to take snapshots of the entire Galaxy in hard X-rays, which is not possible with grazing X-ray telescopes with a small field of view.

The IBIS telescope's energy response to X-rays with energy over 20 keV enables the

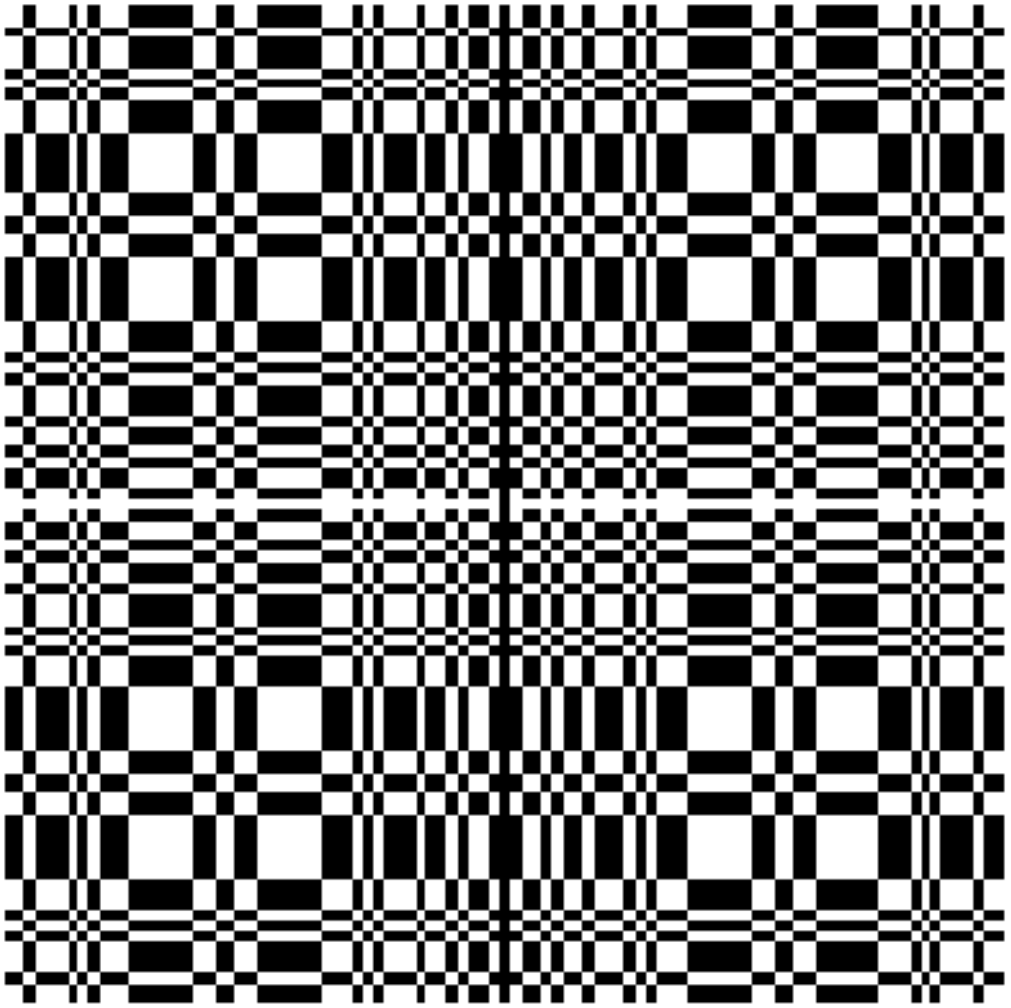

FIGURE 2.19: Drawing depicting the final coded-mask pattern used for the Imager on-board the INTEGRAL Satellite (IBIS). (Source: ESA)

detection of extremely obscured objects rendering the INTEGRAL/IBIS Galaxy survey independent of line-of-sight X-ray photon attenuation. The ISGRI detector of INTEGRAL/IBIS can offer a census of Galactic X-ray emitters with more than 100 keV energy. (Bazzano et al., 2006; Krivonos et al., 2015)

The timescale of one INTEGRAL revolution (orbit) is around 3 days, and would consist of approximately 100 Science Windows (ScW). A ScW is a basic piece of INTEGRAL data which is either a pointing or a slew. A pointing is a period during which the spacecraft axis pointing direction remains stable. Due to the dithering strategy of INTEGRAL, the nominal pointing duration is of order of 2 ks. A slew is a period during which the spacecraft is maneuvered from one stable position to another, i.e. moving from one pointing to the next. INTEGRAL will continuously follow one dithering pattern throughout one observation unless scientific requirements exist to observe sources over a longer exposure period, in this case the dithering modes can be switched off allowing INTEGRAL to be in staring mode. An observation group is a

number of ScWs used in the data analysis. During the scientific analysis, all the ScWs belonging to the same observation are grouped together to form an Observation Group (OG). Images can be created at a ScW level, or they can be stacked together to create a mosaic image. This can be done for several ScWs, whole revolutions, or for bigger OG groups.

2.6.2 INTEGRAL Surveys

INTEGRAL was guaranteed observations of the Galactic Plane to locate transient sources and conduct time-resolved mapping of the galactic plane (Winkler, 2001) during the mission's first years, which resulted in the first deep survey of the Galactic Plane, listing 100 known and ten new X-ray sources. During the first year of observations, INTEGRAL's ability to detect novel transient sources was demonstrated for four months when ten new sources were discovered.

Deep surveys of the Galactic centre region were also conducted; sixty sources with a flux exceeding 1.5 mCrab were identified, including two sources detected for the first time at energies greater than 20 keV, highlighting the remarkable capability of the INTEGRAL telescopes to create sensitive sky maps (Revnivtsev et al., 2004).

Then INTEGRAL hard X-ray cartography of the Galactic Plane was extended by thorough investigations of the Sagittarius spiral arm (Molkov et al., 2004), announcing the detection of 28 X-ray sources with a flux level exceeding 1.4 mCrab, including three new sources.

In 2005, further observations on the Galactic Plane (Revnivtsev et al., 2006) observed 47 hard X-ray sources in total, 15 of which were new. The classifications of the sources were 12 AGN and 11 and 6 galactic binary systems with high-mass and low-mass optical companions.

Monitoring the source activity in the vicinity of the Galactic bulge began in 2005, which provided a comprehensive analysis of a homogenous (hard) X-ray sample of 76 sources in the Galactic bulge area from data gathered in the 1.5 years of the programme and was demonstrated that nearly all Galactic bulge sources are variable whilst identifying six new hard X-ray sources (Kuulkers et al., 2007).

INTEGRAL observations of the Cassiopeia region spanning 1.6 Ms revealed 11 sources with energies more than 20 keV (Den Hartog et al., 2006).

The first systematic study of hard X-ray sources found with the IBIS telescope, based on 5 Ms of total exposure time, as a result of ongoing observations and, subsequently, fast increasing exposure duration, was published by Bird et al. (2004). CAT1 listed 120 sources (28 unidentified objects) detected with a sensitivity of 1 mCrab in the energy range of 20 to 100 keV.

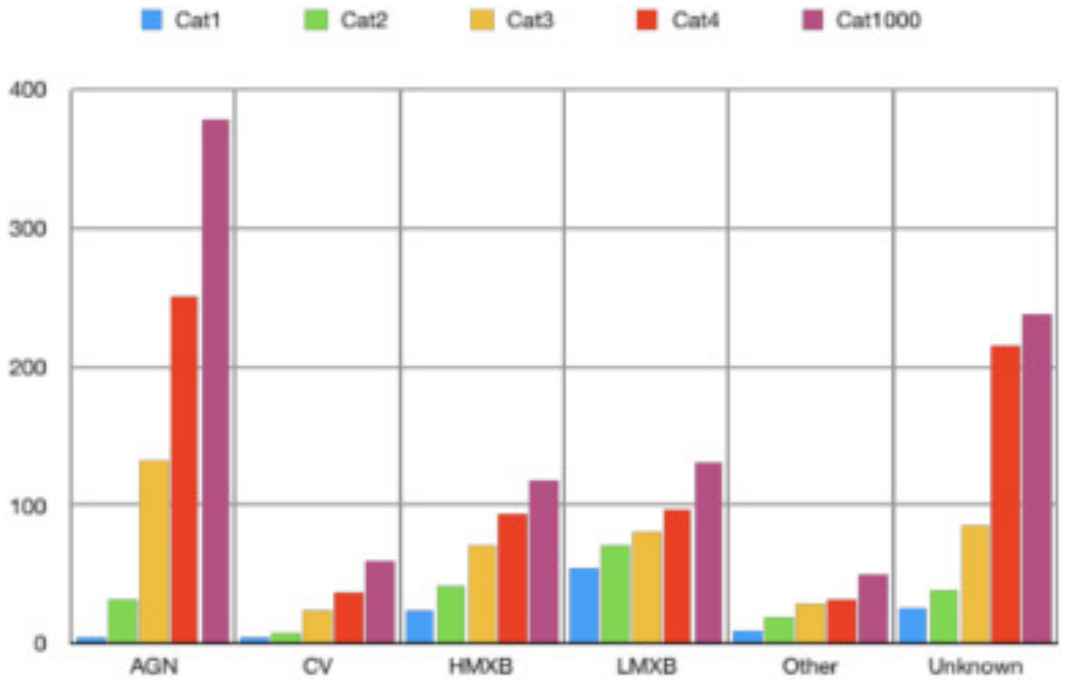

FIGURE 2.20: Evolution of source type and number through the five INTEGRAL IBIS/ISGRI catalogues produced to date (Source: HEASS proposal 2022, PI: Angela Malizia).

The second INTEGRAL/IBIS/ISGRI catalogue (Bird et al., 2006) with 10 Ms of total exposure time consisting of 209 sources - 25 percent of new sources.

A third INTEGRAL survey, an all-sky study based on more than 40 Ms of INTEGRAL exposure, found more than 400 high-energy sources in the energy range 17 - 100 keV, including both transients and faint persistent objects, as seen on time-averaged maps. (Bird et al., 2007).

The first IBIS catalogue generated using alternate sky reconstruction algorithms (Krivonos et al., 2007) contains 403 objects, 316 of which above a 5 detection threshold on the time-averaged map of the sky, with the remaining items detected in various exposure subsamples. Using an improved method of sky image reconstruction for the IBIS telescope to conduct the (at the time) most sensitive survey of the Galaxy above 20 keV - contains 521 objects, 449 of which exceed a 5 detection threshold on the time-averaged sky map, and 53 of which were spotted during distinct observation periods. 262 of the detected sources are galactic, while 221 are extragalactic (Krivonos et al., 2007).

The fourth catalogue of soft gamma-ray sources obtained with IBIS/ISGRI based on 70 Ms of high-quality observations made during the first five and a half years of the mission and public observations with more than 700 high-energy sources found in the 17 100 keV energy range (Bird et al., 2010). This work carefully analysed IBIS data using the most recent official OSA software and source detection methods paying

particular attention to optimising the identification of common transient sources in order to detect both transients and faint persistent objects that can only be revealed with long exposure times.

An all-sky catalogue of sources based on IBIS data during the initial one thousand orbits of INTEGRAL which contained 110 Ms of public scientific observations with a focus on the Galactic Plane and extragalactic deep exposures presented 939 sources detected in the 17 100 keV energy band that surpass a 4.5 significance level with 120 new sources (Bird et al., 2016). The evolution of source types and counts through these INTEGRAL catalogues is illustrated in Figure 2.20.

An INTEGRAL Galactic Plane study based on nine years of observations (Krivonos et al., 2012), the catlog contains a total of 72 hard X-ray sources identified at $S/N > 4{,}7$ that were previously unknown to INTEGRAL surveys.

Krivonos et al. (2015) presented an INTEGRAL all-sky study at energies exceeding 100 keV with greatly increased exposure time. The catalogue of discovered sources contains 132 objects - dominated by 97 hard X-ray sources of Galactic origin (mostly Low-Mass X-ray Binaries and High-Mass X-ray Binaries – LMXBs and HMXBs, 83 in total) compared to the extragalactic source population, which is composed of 35 AGNs.

The most recent INTEGRAL survey is INTEGRAL/IBIS 17-yr hard X-ray all-sky survey, providing flux information from 17 to 290 keV and the catalogue includes 929 objects, 890 of which exceed a detection threshold of 4.5 sigma. (Krivonos et al., 2022)

In addition to imaging surveys and catalogues, INTEGRAL has also produced a continuous stream of transient detections via the INTEGRAL Burst Alert System (IBAS, Mereghetti et al., 2003). The automated software in IBAS was designed to provide automated alerts with precise localisations (4 arcminutes) to the high energy community within tens of seconds of the detection of GRBs with IBIS/ISGRI. When first introduced, IBAS was a marked improvement over the speed and localisation capabilities of previous facilities, and typically produces 5-10 GRB detections each year.

2.6.3 *Swift*

The *Swift* satellite (Gehrels et al., 2004), launched in 2004, is a multi-wavelength observatory designed primarily to study GRBs. It has three main instruments onboard: the Burst Alert Telescope (BAT, Barthelmy et al., 2005), the X-Ray Telescope (XRT, Burrows et al., 2005), and the UV/Optical Telescope (UVOT, Roming et al., 2005). The BAT instrument is sensitive to gamma-rays and X-rays and is used to detect GRBs in real-time. Once a GRB is detected, the satellite can rapidly repoint to observe

the afterglow with the XRT and UVOT instruments. This allows astronomers to study the evolution of GRBs and their afterglows, providing valuable information on the physics of these powerful explosions (see, e.g., Berger, 2014, for a review). In addition to GRB observations, *Swift* also performs X-ray sky surveys. For example, the *Swift*/BAT Hard X-ray Survey (Baumgartner et al., 2013) is a deep, all-sky survey in the hard X-ray band (14-195 keV) that provides a comprehensive view of the high-energy sky. The survey has detected over a thousand new X-ray sources. It has been used to study a variety of astronomical objects, including active galactic nuclei, binary stars, and supernova remnants.

The *Swift* satellite orbits the Earth in a sun-synchronous, low-Earth orbit. This means it follows a nearly polar orbit that keeps the same local time of day as it passes over a given point on the Earth. This allows the satellite to observe the same region of the sky at the same time of day, making it easier to perform long-term monitoring of astronomical objects. The orbit of *Swift* has a period of about 96 minutes and an altitude of about 500 km, which provides it with a good balance between observing time and exposure to the Earth's radiation environment. The satellite is equipped with a pointing system that allows it to rapidly repoint to new targets within about 90 seconds, making it highly efficient for observing fast-evolving astronomical events, such as GRBs. The data generated by the *Swift* satellite is diverse and includes X-ray, ultraviolet, and optical observations, as well as gamma-ray burst detections. X-ray and ultraviolet data are used to study X-ray sources' spectral and temporal properties. In contrast, optical data is used to study the afterglows of GRBs and the properties of their host galaxies. The gamma-ray burst data is used to study the properties of these powerful explosions, including their energetics, spectra, and temporal evolution.

The Burst Alert Telescope (BAT) on the *Swift* satellite is a coded-mask imaging instrument designed to detect GRBs and perform follow-up observations in the X-ray and ultraviolet bands. The BAT instrument has a wide field of view (about 1/6 of the sky) and a large energy range (15-150 keV), making it well-suited for the study of high-energy astronomical objects. Figure 2.21 shows the arrangement of the BAT detector and its coded mask which utilises a random pattern (Barthelmy et al., 2005).

The BAT data reduction process involves several steps to convert the raw data into scientifically useful information. The first step is to perform event extraction, in which individual gamma-ray photons are identified and their arrival times and energies are recorded. Next, a mask-weighting algorithm is applied to the data to correct for the effects of the coded mask and to reconstruct an image of the sky. This image is then processed to identify sources and to perform source characterisation, including source position, flux, and spectrum.

In the case of GRB detections, the BAT data is used to trigger an alert to the astronomical community, allowing follow-up observations to be performed with other

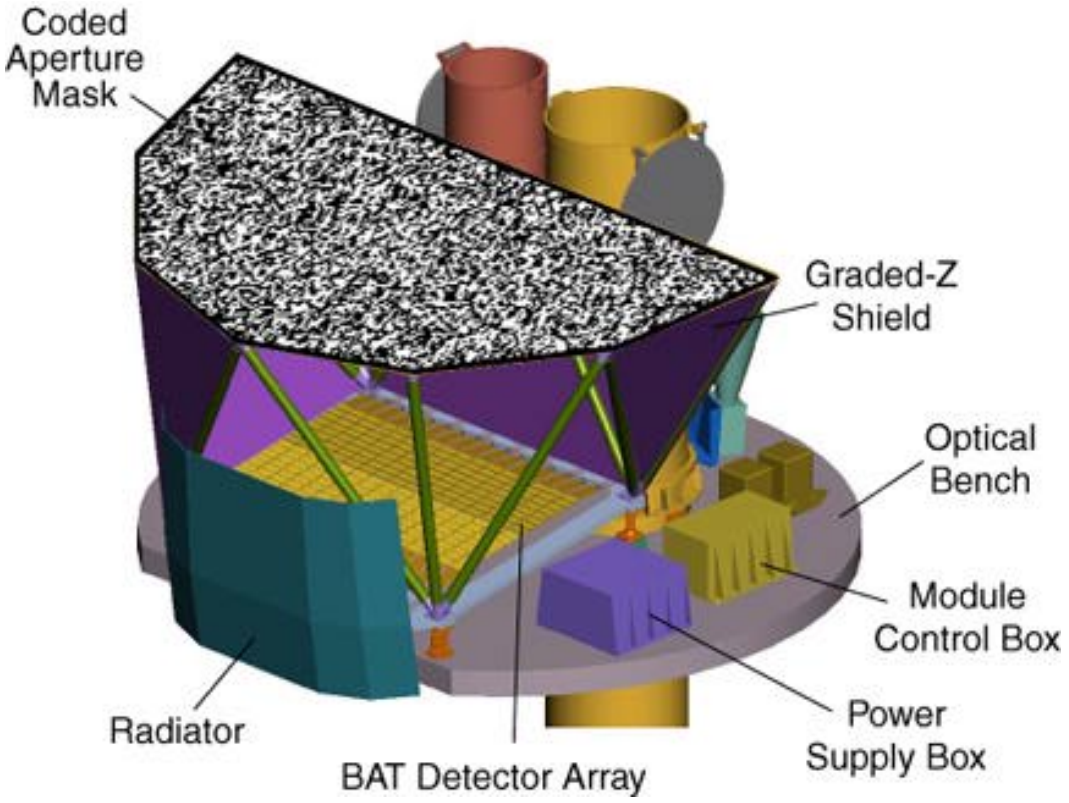

FIGURE 2.21: Diagram of *Swift* BAT instrument and its coded mask. (Source: nasa.gov)

telescopes. The BAT data is also used to study the properties of GRBs, including their temporal and spectral evolution, energetics, and spectral lags.

Overall, the data reduction process for the BAT instrument on the *Swift* satellite is a crucial step in converting the raw data into scientifically useful information. The resulting data is used to study a wide range of high-energy astronomical objects and to advance our understanding of the universe and the processes that drive it.

2.6.4 The SIX Survey

One of the first surveys to combine hard X-ray data from *Swift* and INTEGRAL was the *Swift* – INTEGRAL X-ray (SIX) survey (Bottacini et al., 2012).

The SIX survey performed a careful reprocessing of data from both *Swift* and INTEGRAL to obtain sky maps that combined all available observations. The sky maps from the two instruments were then combined by resampling and cross-calibrating them, with careful attention paid to interpolation during the resampling. The final combined mosaic achieved greater sensitivity than the individual surveys and covered an area of over 6000 square degrees.

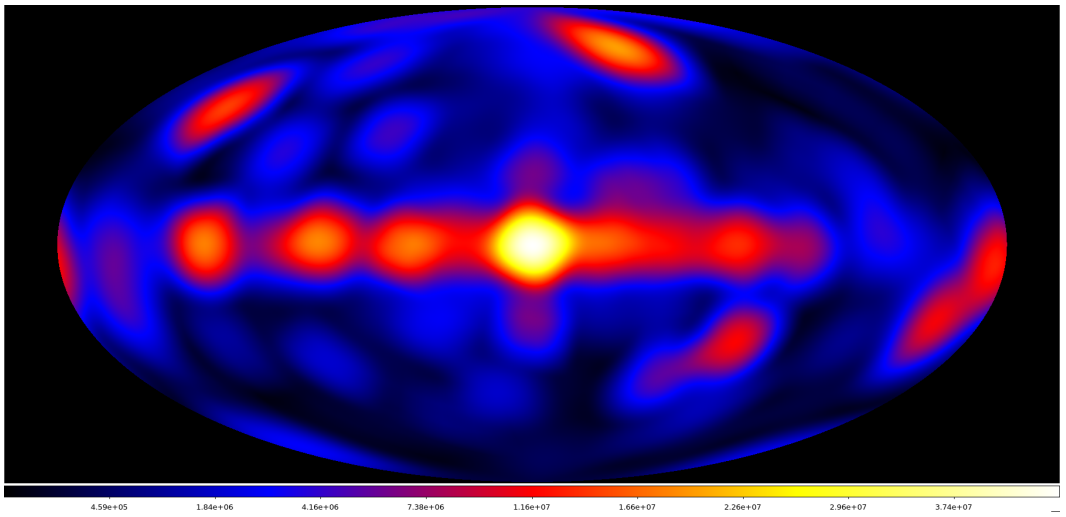

FIGURE 2.22: INTEGRAL Exposure maps (Source: Bird et al., 2016)

This survey identified 113 sources of various types including AGN, blazars, X-ray galaxies and galaxy clusters and some Galactic X-ray sources. This facilitated measurement of X-ray source density distribution (the $\log N - \log S$ relation), obscured AGN fractions (versus both luminosity and redshift), and the AGN X-ray luminosity function.

2.6.5 cat1000

'Catalog of 1000 orbits' (hereafter cat1000, Bird et al., 2016), is an all-sky, soft gamma-ray source catalogue which uses data from INTEGRAL's first 1000 orbits. This survey had a total exposure time of approximately 3.5 years and included 939 sources – the exposure map for cat1000 is shown in Figure 2.22. This catalogue used a light-curve based method to search for transient events, but did not attempt to search for sources on the shortest timescales due to the size of the dataset that would have created. The shortest timescale mapped corresponded to the $\sim$3 day orbit of INTEGRAL.

The data for cat1000 was taken from all publicly available data from revolution number 12 to revolution 1000, unless flagged as Bad Time Intervals (when known data errors make the data unusable). This amounted to approx. 73000 ScWs. Pipeline processing was done by using the standard OSA software (Goldwurm et al., 2003) and includes all the processing of the data up to and including generating sky images for each individual ScW. These images were created in 5 energy bands (17-30,30-60,18-60,20-40 and 20-100 keV).

A reference catalogue is created, this can be considered as the 'prior' knowledge of the data. This is a catalogue of the sources that have been previously detected by INTEGRAL and is used in the pipeline processing for image processing. To calculate

detector noise, each ScW image is sigma clipped (which removes the sources) and the RMS is calculated. This allows any ScWs that have a low image quality (below a certain sigma threshold) to be removed (for example, ScWs that were collected during high solar activity or soon after a perigee passage). Also staring data is removed at this stage as these exhibit much higher systematic noise than dithering observations - this left 67000 ScWs.

The next stage is to create the mosaic images. This process statistically averages the ScWs images and can be used to create all-sky maps, or maps of specific regions of the sky. Maps were created on different time-scales, e.g. for a whole revolution, part of a revolution or whole-archive timescales. In total, for cat1000, 60 all-sky maps (and variants) and over 19000 revolution maps were constructed and searched (Bird et al., 2016).

The mosaic maps were then searched for potential sources using two different algorithms. From this an initial candidate list was made by iteratively merging the excess lists from each map to the reference catalogue. Any new potential sources were then added to a base list if they had a detection threshold above a certain number (4.5 sigma for whole-archive and 6 sigma for a revolution map). At the end of the process these were manually inspected.

Light curves (LC) were then made for every candidate source in five energy bands. Using what is known as the 'bursticity method' (Bird et al., 2016), the sources were searched for variable source emission. This method allows optimisation of detection of any known or suspected sources that are present on timescales longer than single ScW.

Finally a final source list is created, this is done manually by domain experts. This task is the most time intensive and where the vast majority biases get included. A number of steps are performed to minimise the possibility of false catalogue entries. Firstly each source is manually inspected by a number of domain experts. They inspect each source's point spread function (PSF) shape, consistency across multiple energy bands, and the significance of the source relative to the local noise levels in the map. There must be unanimous agreement before a potential source is declared a true source.

Figure 2.23 shows the catalogue process which historically has taken on average around 18 months (though this increases as the data volume increases), and many of the tasks require skilled domain experts spending many hours to manually check and classify objects. For missions like INTEGRAL this is just tenable, but for the next-generation telescopes many of these tasks will need to be redesigned and automated. Another disadvantage of the current method is that of human bias which can only be partly addressed by blind voting and "citizen science". As well as removing human subjectivity, ML methods would allow published catalogues to be pre-filtered to a lesser degree and labelled with meta-data that will allow the datasets to be re-used in unanticipated ways. Three areas where machine learning can be

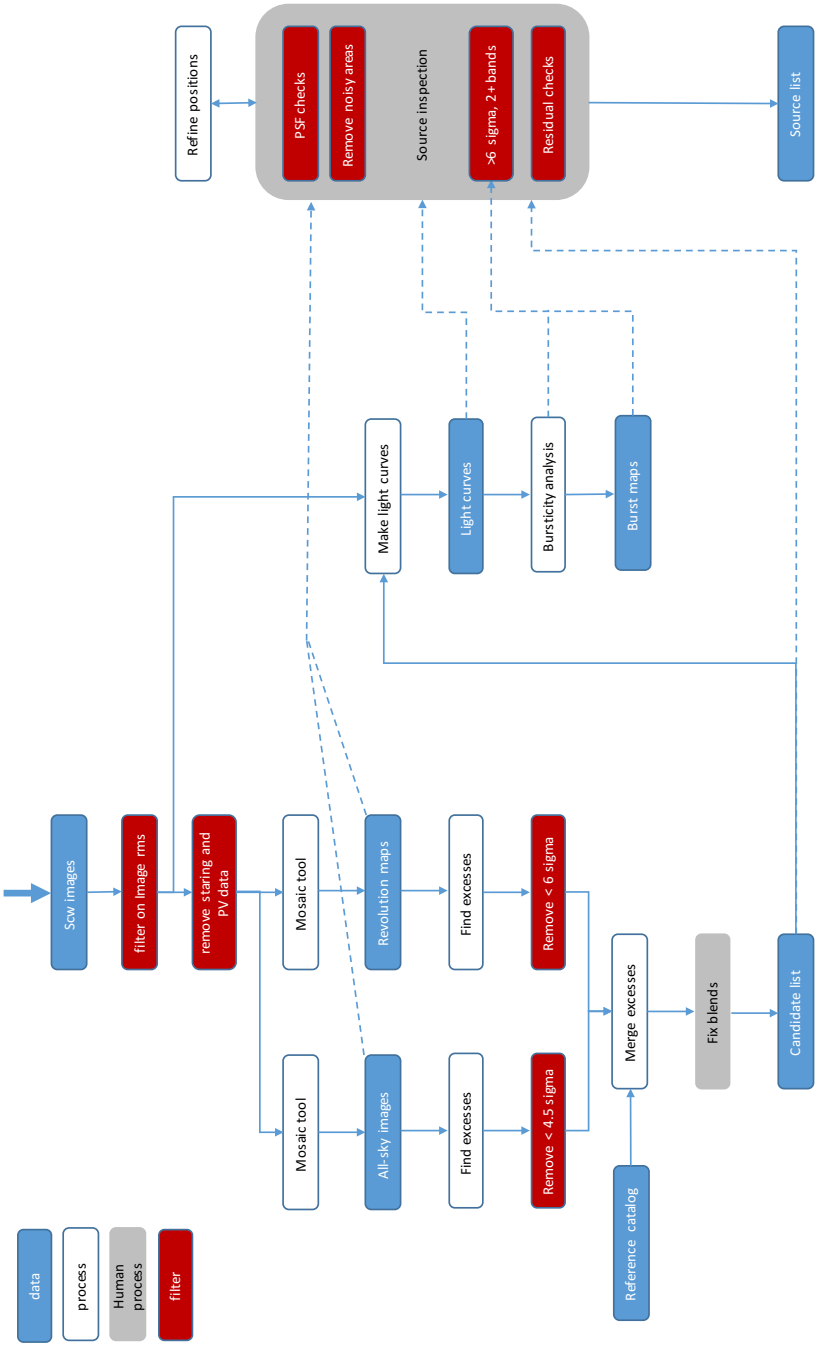

FIGURE 2.23: Flowchart showing ISGRI data catalogue processing pipeline Bird et al. (2016).

applied to streamline the catalogue process are in source detection, source classification and outlier detection.

The techniques employed to generate ISGRI catalogues (such as cat1000) are no longer adequate as they do not scale well to the ever increasing dataset. cat1000 took 9 domain experts to spend 2.5 years to produce and as INTEGRAL has now exceeded

2500 orbits there is a clear need to improve the techniques and tools used. This motivated the development of the tools presented in this thesis.

Chapter 3

Background: Machine Learning in Astronomy

Currently, astronomy is dependent on humans to play the main role in data analysis, interpretation, and classification of celestial objects, but this will be unsustainable for the huge volume of data collected from the next generation of telescopes. These massive datasets have huge potential to generate new knowledge and understanding but being able to make full use of the data and extract the important information will require the data to be processed computationally.

As astronomy enters the big data era, the field is seeing a rapid rise in the volume and complexity of data and will continue to do so as technological advances continue to allow for the production of ever more powerful detectors, telescopes, and computers. The raw data harvested will have the potential to form massive multi-temporal and multi-wavelength datasets of high level data products, so long as the astronomy community can harness the correct tools that allow us the efficient handling of such large and complex data sets. Current techniques for exploiting large datasets in astronomy are already becoming inadequate in dealing with the continually increasing volume of data that is produced, this will only become exacerbated as the next generation of telescopes are launched.

Future instruments will increase the number of observed objects and associated measured properties by more than an order of magnitude. Figure 3.1 illustrates how data volumes in astronomy surveys have grown over the past few decades and are poised to grow substantially more in the future. For example, every image taken by the LSST will be several Gigabytes in size. Every night more than 10 Terabytes of data will be collected, which will accumulate to 60 Petabytes of data over its lifetime (LSST Science Collaboration et al., 2009). The Cherenkov Telescope Array (CTA, Actis et al., 2011) will be an array of over 100 telescopes designed to observe Cherenkov radiation from particle showers produced in the upper atmosphere by the passage of high

energy photons and cosmic rays. When fully operational, the CTA is predicted to produce 27 Petabytes of data per year (Arrabito and Bregeon, 2020). The Square Kilometer Array (SKA, Dewdney et al., 2009) will consist of over 100,000 radio antennas in Australia (SKA-Low) and 200 radio telescopes in South Africa (SKA-Mid). The Australian site alone is expected to produce 8 terabits of raw data per second (according to SKA mission site: skao.int) and expects to archive over 300 Petabytes of data per year.

To keep up with this accelerated growth, astronomers are beginning to look to automated methods to detect, characterise, and classify astronomical objects, and many machine learning algorithms can be applied to a variety of tasks in this pursuit. Machine Learning (ML) provides computers with the ability to learn without being explicitly programmed (Samuel, 1969). ML algorithms, when applied correctly, are excellent tools for finding patterns in data. Not only do ML techniques allow larger datasets to be analysed quickly, but they also have the potential to remove the problem of human bias and, if applied well, can even outperform humans at certain tasks (LeCun et al., 2015).

A broad variety of ML techniques were explored for the scientific objectives of this thesis. The outcomes of these explorations are presented in great detail in Chapter 4, where the superior performance of deep learning algorithms is demonstrated thereby justifying their selection for the applications achieved in this thesis.

3.1 Introduction to Foundational Concepts in Machine Learning

3.1.1 Definition of Machine Learning

Machine learning (ML) is a subset of the broader field of artificial intelligence (AI). AI encompasses all computing capabilities designed to mimic the decision-making ability of humans. This can include rule-based systems, problem solving and pattern recognition, as well as machine learning approaches including deep learning. ML specifically refers to AI applications which utilise algorithms that allow a computer to learn from and make decisions based on data. Specifically, ML uses statistical techniques to learn patterns in data and encode them in a model which can be used to make predictions for new unseen data. Whilst AI refers to the broader goal of simulating human-like intelligence, ML specifically focuses on the development of algorithms that can learn from data to achieve this goal.

Deep Learning is an even more focused subset of ML comprising neural networks with three or more layers, and is meant to mimic the learning behaviour of the human

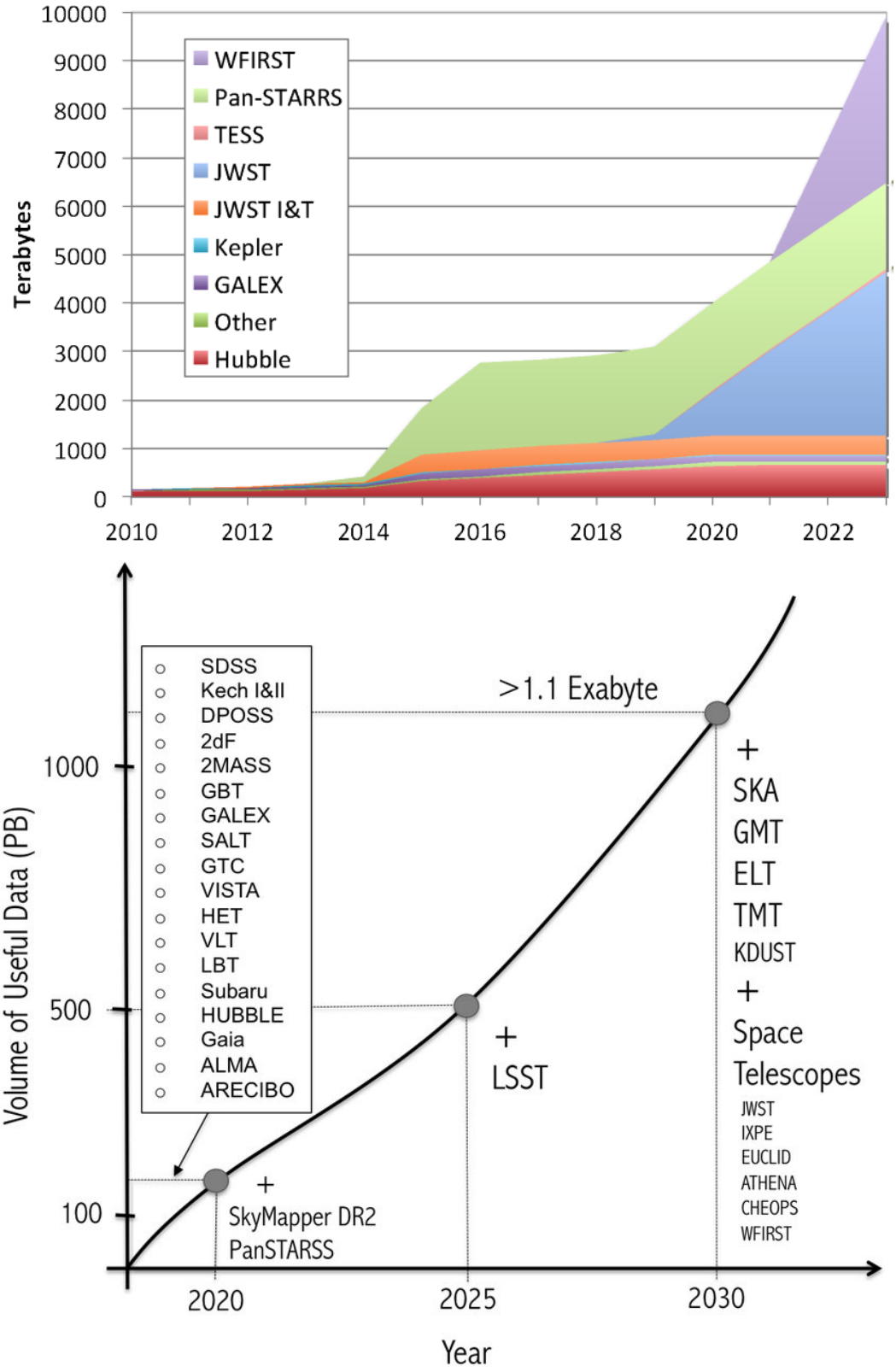

FIGURE 3.1: Representative curves of growth for data volumes in astronomy experiments for historical surveys (top:, from MAST https://www.cosmos.esa.int/web/machine-learning-group/why-use-ml-in-astronomy-) and planned surveys (bottom, from Rosa et al. 2020, https://doi.org/10.1590/0001-3765202020200861)

brain. A neural network is a computational model inspired by human neural networks, and it contains processing elements known as ”neurons” organised into layers. Neurons transmit information between one another in a similar style as human neurons, with information being received from neurons in upstream layers then processed by neurons before being passed to neurons in the downstream layers. Deep learning neural networks typically have an input layer that receives data, one or more hidden layers that performs calculations on the data and learns the underlying data relationships, and an output layer that produces the final model prediction.

3.1.2 Terms and Concepts in ML

ML models universally process a set of input data composed of numerous ”features” to compute an estimation of the desired output of the model, its ”prediction”. The canonical ML dataset typically used as a training example is the UCI Iris dataset (Fisher, 1988), which uses properties of iris flowers as inputs to models that predict the category of iris (See Figure 3.2). In ML models, features are measurable quantities that describe input data, so for example in the iris dataset the features are the length and width of the petals and sepals of the irises. Features can be continuous such as the iris part dimensions, or discrete (categorical) such as real-world applications that might process data containing day of the week for time series data (an ordinal feature where adjacent categories have meaning) or geographical data like country (where adjacency in name or even location might not be meaningful for the target model).

Model outputs, or ”predictions”, also are either continuous or categorical: ML models that produce a continuous predicted variable are known as regression models, whilst models that produce a discrete categorical prediction are known as classification models. In classification models, the different possible predicted categories are called ”classes”, and in the iris dataset example the classes are the three possible categories of iris plant. A regression model might attempt to predict a continuous variable, so for example in the iris dataset a model could be trained to predict sepal length from sepal width and petal width and length.

ML models learn how to make more accurate predictions by undergoing a ”training” process where data is presented to the model and it the model ”learns” how to make better predictions through the use of sophisticated algorithms. A detailed description of ML model training approaches is provided in Section 3.3, but fundamentally ML models learn by evaluating the merit of predictions from subtly different weights on input features. For example in the iris dataset, if a simple linear model was used to predict iris classes from the four input features and its performance evaluated by measuring the fraction of correct predictions (”accuracy”), an ML model would use the difference in accuracy for different linear weights to learn how to adjust those

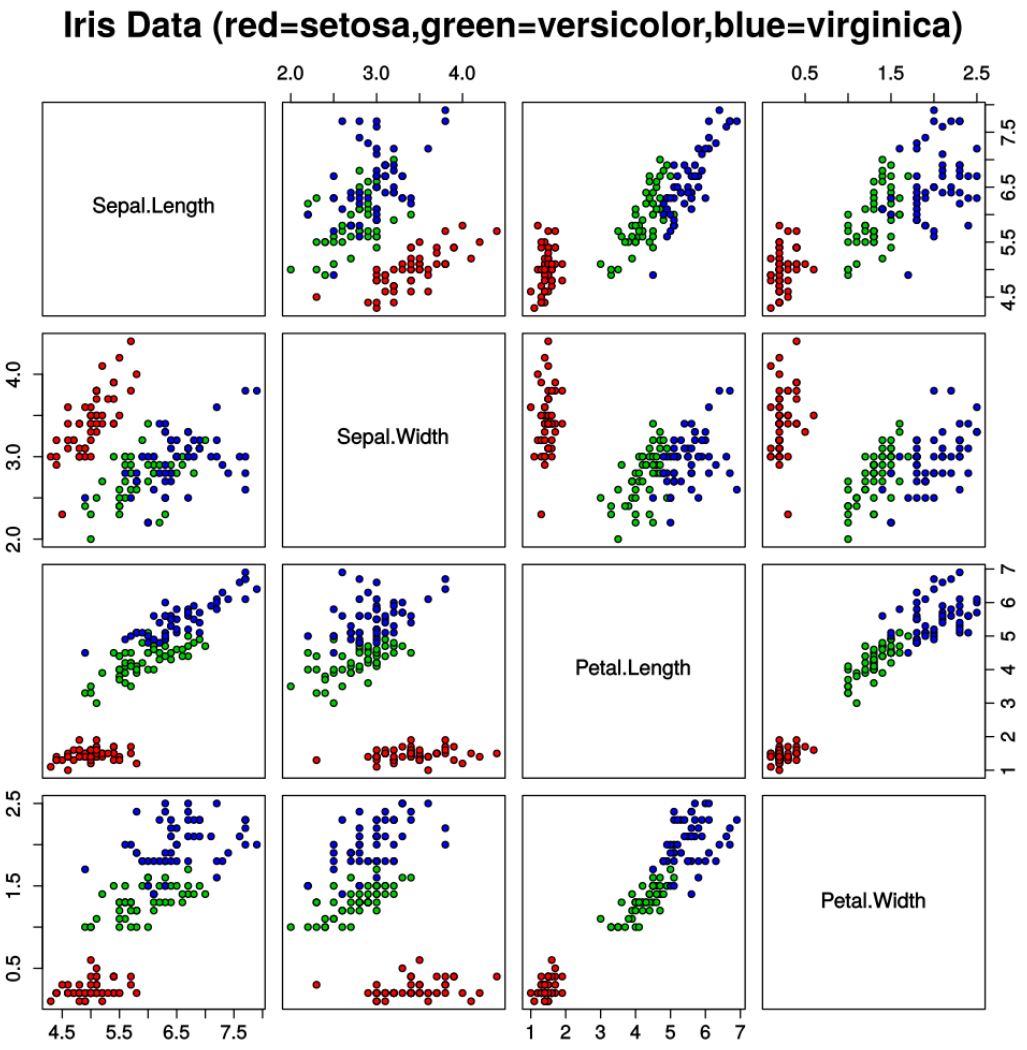

FIGURE 3.2: Scatter plot of features of members of the iris dataset, colour-coded by iris category (Source: wikimedia commons)

weights to achieve a more accurate prediction. This process is repeated until an optimal performance is achieved.

3.2 Applications of Machine Learning in Astronomy

3.2.1 Supervised Classification

Like all ML algorithms, supervised learning algorithms typically have model hyperparameters. Hyperparameters are values that control the learning process and determine the values of model parameters the algorithm learns, and these hyperparameter values can be tuned or optimised. The algorithm uses hyperparameters while learning; they are not part of the resulting model. After the

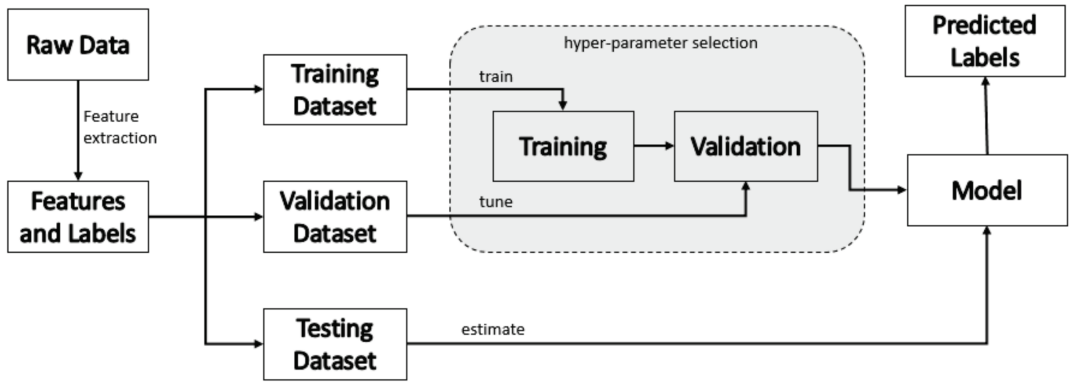

FIGURE 3.3: General workflow diagram of machine learning algorithms (Source: Basavaraju et al., 2020)

learning process, the model parameters are estimated or learned, constituting the model itself. The parameters are the values an algorithm can change independently as it learns and these values, and by extension the model's performance, are directly impacted by how the hyperparameters are tuned.

Setting the model hyperparameters is an important part of the training stage of supervised learning. Before training begins, the labelled dataset is split into three subsets: training, validation and testing. The training set is used to train and make the model learn the hidden features/patterns in the data, as initially the model is fit to the training set. Then the validation set is used to validate our model performance during training, allowing for unbiased evaluation of the model's performance. Not only can this stage be used to tune the hyperparameters but also to avoid overfitting. Overfitting occurs when a classifier models the training data too well but has not learned to generalise, so does not perform well on new data. This happens when a model over-trains on the detail and noise in the input data rather than being able to generalise and be applied to unseen new data and make accurate predictions. Finally, the test set is used to test the model after training, providing an unbiased final model performance by the trained model predicting the target variable of this unseen data and can be used to compare performances of different supervised learning algorithms. This workflow is summarised graphically in Figure 3.3.

The performance of supervised algorithms strongly depends on the input dataset. The majority of supervised ML algorithms are sensitive to the number of features in the training set and what is referred to as "the curse of dimensionality" arises when analysing data in high-dimensional spaces resulting in the feature selection being a crucial element of the process.

Classification assigns new data objects (which could be observations, sources, light curves etc) into pre-defined classes. This is done by using a training set, which consists of observations where the class is known and then training a computer

program with this data. Binary classification deals with problems where there are only two classes than can be assigned (e.g., star-galaxy classification), and multi class classification when there are more than two classes (e.g., astrophysical transient classification). It can be imagined that the input space is separated into distinct areas, known as decision regions, with decision boundaries being the borders of these regions. The data is known as being linearly separable when decision boundaries allow for all the data to be perfectly split into their respective classes.

Presented below is a brief description of some supervised classification algorithms.

3.2.1.1 Discriminant Analysis

Discriminant analysis is well understood and simple to implement with a number of variants. Linear Discriminant Analysis and and Quadratic Discriminant Analysis are two types of ML classifiers that produce linear and quadratic decision boundaries respectively. They use linear or quadratic combinations of the input variables to model the degree to which data points belong to the target classes. It can then make predictions on new input data depending on where the new data point sits in the feature space in relation to the decision boundaries. These two techniques are relatively simple and can therefore be unsuitable if there are nonlinear and complex interactions between the variables (Feigelson and Babu, 2012). Discriminant analysis has been applied to astronomy problems such as the classification of Gaia eclipsing binary and multiple systems (Ferrari et al., 2015), morphological classification of galaxies (Süveges et al., 2017), and prediction of solar flares (Leka et al., 2023).

3.2.1.2 Naive Bayes

Naive Bayes is one of the simplest methods of ML classification (Ripley, 1996) and is based on Bayes' theorem, and is used in applications such as spam filtering and sentiment prediction. This algorithm takes the naive assumption of class conditional independence, essentially assuming that the presence of a feature in a class is unrelated to the presence of any other feature. This assumption allows for a small computation time, but also limits the performance of the classifier in more complex datasets as conditional independence is seldom true for problems in the real world. Although generally, it is incorrect to assume the independence between variables, Naive Bayes algorithms have been shown to work well for many datasets. Naive Bayes has been used in morphological galaxy classification (Bazell and Aha, 2001; Vavilova et al., 2021), classifying X-ray Sources (Broos et al., 2011; Tranin et al., 2022), estimating photometric redshifts (Carrasco Kind and Brunner, 2014) and classification of variable objects and transients (Mahabal et al., 2009).

3.2.1.3 K-Nearest Neighbour

K-Nearest Neighbour (KNN) is a simple classifier and it is most suited to be applied to data when there is minimal or no prior knowledge about the distribution. In KNN an object is assigned a class by calculating the majority vote of its K nearest neighbours in the multidimensional parameter space and then assigning that class (James et al., 2013)

It is possible to weight the voting by the proximity of the neighbours. This would mean the very nearest neighbour are more influential, contributing more than neighbours that are further away. KNN has the advantage that it can work well with non-linear data, it is simple to use and there is no separate training phase (model training and final classification are simultaneous). It is important to choose a K that is large enough to avoid over-fitting but also small enough to avoid oversimplifying the distribution. KNN can be computationally intensive in large datasets as all the data points need to be examined in order to decide which points are the nearest neighbours. KNN has been used in solar flare forecasting (Li et al., 2007), galaxy density estimation (Ferdosi et al., 2011), quasar-star classification (Peng et al., 2013), and classification of black holes and neutron stars (de Beurs et al., 2022, see Figure 3.4).

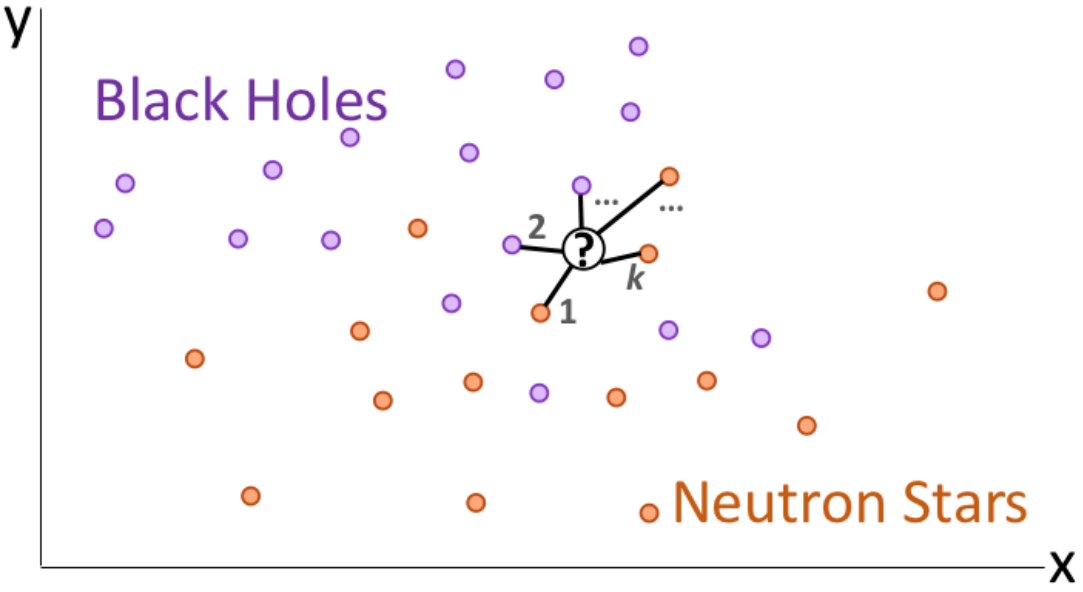

FIGURE 3.4: An example of KNN classification of black holes and neutron stars (Source: de Beurs et al., 2022).

3.2.1.4 Support Vector Machine

Support Vector Machines (SVM) are more straight forward to apply to binary classification problems. A SVM works by finding the hyperplane in the feature space that allows all data points to be kept separate from all the other class types by making

the separation as large as possible. If there is no hyperplane that separates the data points, then this method can be used to form a soft margin, which is a hyperplane that separates as many of the data points into their respective classes as possible while minimising the distance of the misclassified points (Bishop, 2006). SVM can be used with the kernel trick (Schoölkopf and Smola, 2002) which projects the data onto a higher-dimensional feature space which makes the data more sparse and allows a well-separated hyperplane to be found and produces a non-linear classifier. SVMs can take a long time to train and it can be hard to incorporate domain knowledge, it is sometimes also hard to extend this method to problems with more than two classes. SVMs have been used in photometric redshift estimation (Jones and Singal, 2016), prediction of solar flares (Tamayo et al., 2016), prediction of instability timescales of compact planetary systems (Kurcz et al., 2016), source classification (Hui and Gao, 2015), blazar classificataion (Arsioli and Dedin, 2020) and star-galaxy separation (see Figure 3.5 from Małek et al., 2013; Raboonik et al., 2016; Wang et al., 2022).

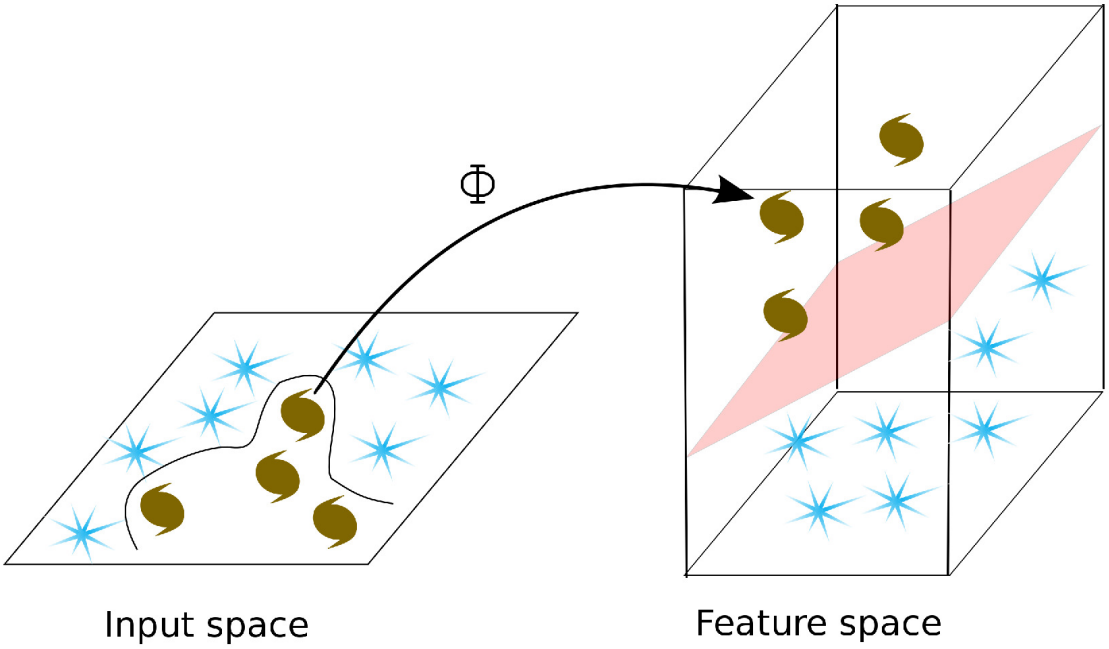

FIGURE 3.5: An example of SVMs being used to perform star-galaxy separation (Source: Małek et al., 2013).

3.2.1.5 Decision Trees

Decision Trees build models in the form of a tree structure. This method can be applied to both regression and classification problems, but here it will be discussed in the context of classification (Breiman et al., 1984). A Decision Trees is incrementally constructed by using recursive binary partitioning to divide up the feature space, breaking down the data into increasingly smaller subgroups. This results in layers of nodes that form the tree structure (Esposito et al., 1997). Decision Trees are simple and

powerful classifiers. They work well even when the feature space is complicated, and generally yield results with very small bias. On the downside, Decision Trees produce results with high variance. Making small changes to the features in the training set can produce tree structures that vary a lot. This is a product of the hierarchical nature of the Decision Trees, in other words making a small difference in the highest nodes of a tree can then have the knock-on effect of large changes to the structure as we move further down the tree. To overcome these problems, there are ensemble methods that can be used. Bootstrap aggregation (Hothorn and Lausen, 2003), or bagging is one method and is the basis for a technique called Random Forest (RF Liaw and Wiener, 2002). Another is boosting where multiple Decision Trees are made and whose combined performance is significantly better than that of any of one the single Decision Trees (Dietterich, 2000). Even so, simple decision tree models have been impactful in astronomy, such as their use in star-galaxy classification (see Figure 3.6 from Vasconcellos et al., 2011).

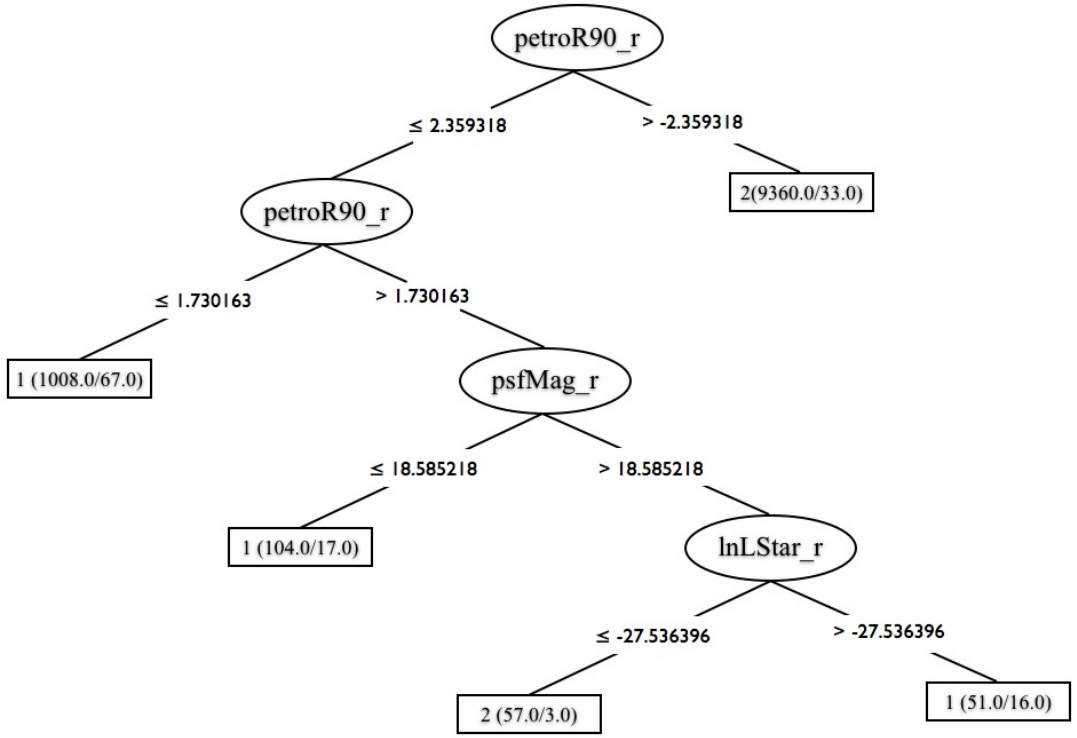

FIGURE 3.6: An example of decision tree being used to perform star-galaxy classification in SDSS data (Source: Vasconcellos et al., 2011)

3.2.1.6 Random Forest

A Random Forest is constructed from many Decision Trees. Classification of a new object occurs by evaluating the output of each Decision Tree in the Random Forest. Each tree will assign a class to the input variable ("votes" for that class). The Random

Forest then makes a final classification based on which class has achieved the most votes from all the trees in the Random Forest. Random Forests are very popular in Astronomy because they can handle thousands of input variables, but are relatively easy to understand and build. Random Forests can be trained to classify data from a particular mission and then applied to new data as it comes in. Additionally, Random Forests have the potential to estimate which variables (or features) are important when classifying by counting the fraction of branches that depend on each feature. Random Forests have been used in photometric classification of supernovae (Dai et al., 2017; Richards et al., 2012), redshift estimation (Dai et al., 2017), source classification (Richards et al., 2012; Lo et al., 2014; Fotopoulou et al., 2016; Jayasinghe et al., 2018; Mistry et al., 2022), photometric lightcurve classification (Möller et al., 2016; Zaidi and Narayan, 2016), star-galaxy classification (Clarke et al., 2020), transiting planet classification (Jenkins et al., 2013; Schanche et al., 2019), source identification (Scaringi et al., 2008) and automated artifact rejection (Luther et al., 2016).

3.2.1.7 Boosted Decision Trees

Boosted Decision Trees are an ensemble approach to ML. Boosting aggregates simple rules to create a predictive model. In this method many weak learners are used, which on their own only perform marginally better than chance, but they can then be boosted to improve performance. A sequence of simple Decision Trees can be created, and then over each iteration the training observations can be reassigned a weight so that any misclassified data points are given greater influence when the next weak learner is created. This effectively means the algorithm pays more attention to the most difficult to classify points, which overall improves the performance of every subsequent decision tree. Boosted Decision Trees have been used in signal detection (Vinciguerra et al., 2017; IceCube Collaboration et al., 2016), photometric redshift estimation (Beck et al., 2016; Wolf et al., 2016), to predict the stability of planetary systems (Tamayo et al., 2016), star-galaxy classification (Golob et al., 2021), and object classification (Lochner et al., 2016; Sevilla-Noarbe and Etayo-Sotos, 2015; Sahakyan et al., 2023)

3.2.1.8 Artificial Neural Networks

Artificial neural networks (ANN), which are loosely inspired by the structure and workings of a biological brain are increasingly popular in astronomy. They are made up of a group of interconnected nodes, each of which processes the data that it receives and then passes this information on to other nodes via weighted connections. ANNs are a non-linear statistical data modelling tool, which may be used to model complex relationships between a set of inputs and outputs.

If we look at just one artificial neuron, we can see it has multiple inputs and one output. The individual neuron will fire if its output is above a certain threshold. If we connect neurons into a network, then the resulting structures can learn very complex hypotheses. ANNs have been used in signal detection (Vinciguerra et al., 2017), object classification (Bass and Borne, 2016; Salvetti, 2016; Mukund et al., 2016; Charnock and Moss, 2016) as also illustrated in Figure 3.7 (from de Dios Rojas Olvera et al., 2021), photometric redshift estimations (Oyaizu et al., 2008; Bilicki et al., 2018), galaxy morphology (Kuminski and Shamir, 2016), source detection (Dong, 2012), stellar cluster characterisation (Cantat-Gaudin et al., 2020) and is also used in the Source Extractor (SExtractor) software (Bertin and Arnouts, 1996).

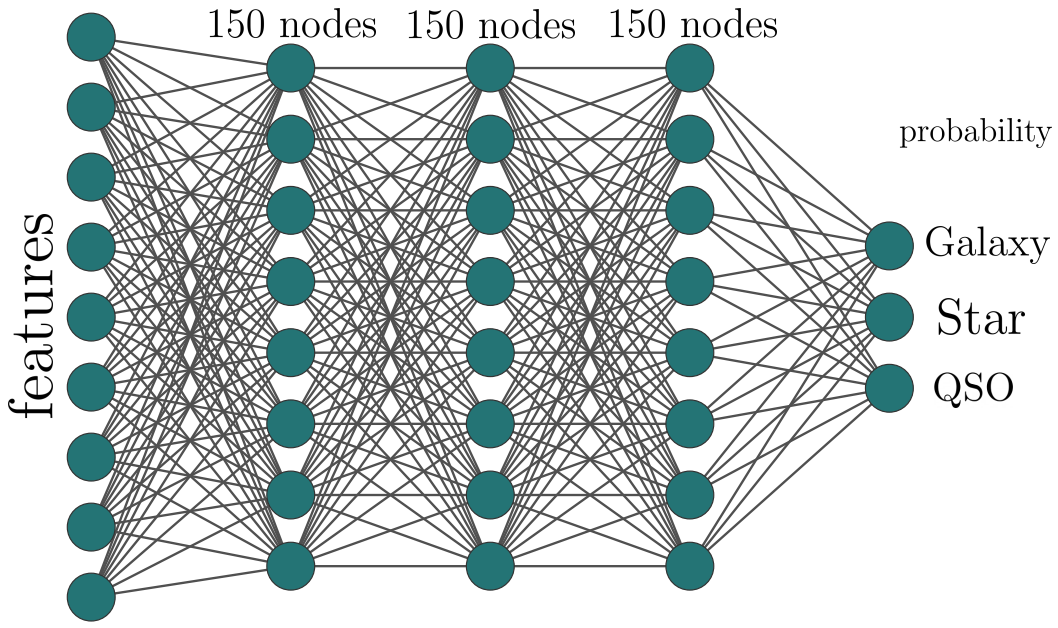

FIGURE 3.7: An example model architecture that utilises artificial neural networks to perform classify objects from a large optical survey as stars, galaxies, or QSOs (de Dios Rojas Olvera et al., 2021).

3.2.2 Unsupervised Classification

Unsupervised machine learning is a type of machine learning in which the training data are not labelled and the algorithm is not provided with any guidance or feedback on the correct output for a given input (Ghahramani, 2004). Instead, the goal of unsupervised learning is to discover patterns or structures in the data that can be used to classify or cluster the data points into different groups.

Unsupervised learning has been used in a variety of applications in astronomy, including the classification and clustering of astronomical objects such as stars, galaxies, and quasars (Clarke et al., 2020). For example, unsupervised learning

algorithms have been used to identify patterns in the distribution of stars in the Milky Way galaxy, to classify different types of galaxies based on their properties, and to identify clusters of quasars in large astronomical surveys (Ethiraj and Bolla, 2022; Mahalakshmi et al., 2022; Logan and Fotopoulou, 2020).

Unsupervised learning algorithms have also been used to extract features or patterns from astronomical data that domain experts may miss. For example, unsupervised learning algorithms have been used to identify features in the spectra of astronomical objects that can be used to classify them or to measure their physical properties (Fraix-Burnet et al., 2021; Garcia-Dias et al., 2018).

3.2.2.1 K-means

K-means clustering is a type of unsupervised machine learning algorithm that is used to partition a dataset into a specified number (k) of clusters (Lloyd, 1982; MacQueen, 1965). The algorithm works by iteratively assigning each data point to the cluster with the nearest mean, and then re-computing the means of the clusters based on the newly assigned data points. This process is repeated until the assignments of data points to clusters converge, at which point the clusters are considered to be stable.

K-means clustering has been used in a many of applications in astronomy, including the classification of astronomical objects such as stars, galaxies, and quasars. For example, k-means clustering has been used to identify patterns in the distribution of stars in the Milky Way galaxy, to classify different types of galaxies based on their properties, to identify clusters of quasars in large astronomical surveys (Turner et al., 2019; Ordovás-Pascual and Sánchez Almeida, 2014; Kyrychenko and Kolomiyets, 2021; Wagstaff and Laidler, 2005) and to classify stellar spectra (Garcia-Dias et al., 2018).

K-means clustering has also been used to extract features or patterns from astronomical data that may not be immediately apparent to the human eye. For example, k-means clustering has been used to identify features in the spectra of astronomical objects that can be used to classify them or to measure their physical properties (Sánchez Almeida et al., 2010).

Overall, k-means clustering has proven to be a useful tool for the analysis of astronomical data, and it has the potential to enable the discovery of new patterns and relationships in the universe that may not have been previously known.

3.3 Fundamentals of Machine Learning

3.3.1 Varieties of Machine Learning Models

ML comes in three flavours: supervised, unsupervised, and reinforcement learning. The type of learning that is suitable will depend on the type of problem being undertaken, and if the use of prior information can be incorporated fully, partially or not at all. Most astronomy uses of ML employ supervised learning when datasets consist of input variables (X) and an output variable (Y). The aim is to obtain an approximation of the mapping function ($Y = f(X)$), so the model can be applied to previously-unseen data and predict the target variable with high accuracy.

In practice, this means a labelled training set is required where each data point will have a number of attributes, individual measurable properties, or characteristics (features) with a known target property. An ML algorithm can then be applied to the dataset which will learn the mapping function from the input to the output. The aim of this type of ML is to try to compute a very accurate approximation of the mapping function so that it can be used with new input data to predict new output data (Kotsiantis et al., 2007).

Supervised learning problems are either regression or classification problems; the difference between the two types is that the dependent attribute is numerical for regression and categorical for classification. In regression models, the output variable is a real or continuous value, and the algorithms model the relationship between the input variables and the output variable to be able to predict output values for new inputs.In classification models, the output variable is typically a most likely category – a discrete value from a finite list – but could also be a set of probabilities for each category. Classification models attempt to draw some conclusions from observed values and, if given a set of one or more input variables, will try to predict the value of one or more output labels.

Supervised learning models are constructed according to the input dataset, allowing supervised learning algorithms to approximate very complex non-linear relations in the models, allowing for better accuracy and performance compared to traditional methods that typically adopt a parametric approach to data modelling.

Unsupervised ML does not require a training set, there are only features and no target. Unsupervised learning algorithms allow for complex relationships that exist in the dataset to be discovered, without labels provided by an expert. This allows for the data to be processed by the ML algorithms without prejudice, presumptions or preconceptions. However, important and useful information might be left out and this potentially presents a greater challenge for unsupervised approaches. This type of ML can be used for clustering and dimensionality reduction, and anomaly detection – in

general trying to simplify complex data by allowing a model to find underlying patterns. The aim here is to create a model based on the inherent structure of the data and, as there is no training set, the ML algorithms are free to find potentially previously unseen patterns and structure and assign target labels, unlike in supervised ML where the target labels are already provided (Gentleman and Carey, 2008).

Semi-supervised methods fall somewhere between the previous two methods. They are typically employed where labelled data is relatively expensive to obtain. Although they can be more involved and challenging to use, they do allow for some prior information to be added to the algorithm whilst also allowing the freedom to interpret the data objectively and help draw useful conclusions (Zhu and Goldberg, 2009).

Reinforcement learning (Sutton et al., 1998) is a machine learning technique that trains an artificial agent to make a sequence of decisions in a dynamic environment. The agent receives feedback in the form of rewards or penalties for each action taken, and its goal is to maximise its long-term cumulative reward (Watkins and Dayan, 1992). The reinforcement learning framework includes an agent, environment, and reward function. The agent is responsible for taking actions in the environment, and the environment provides feedback on the agent's actions. The reward function maps the agent's actions to a scalar value, the reward or penalty the agent receives for taking that action in the environment. The agent learns to make optimal decisions by exploring the environment and updating its policy based on the received rewards. This process is called "trial-and-error" learning because the agent tries different actions to see which one yields the highest reward. Over time, the agent learns to make better decisions using the experience gained from previous trials (see, e.g., Mnih et al., 2013; Silver et al., 2016).

Reinforcement learning has been used in several applications in astronomy and astrophysics. One example is the use of reinforcement learning to optimise the pointing of telescopes (Naghib et al., 2019). Telescopes need to be pointed accurately at specific targets in the sky, and the accuracy of the pointing depends on various factors, such as weather conditions, telescope mechanics, and the target's movement. Reinforcement learning algorithms can be used to learn the optimal pointing strategy by considering these various factors and maximising the reward, which can be defined as the amount of light collected by the telescope. Another example is the use of reinforcement learning to detect gravitational waves.

In Astronomy, the datasets these algorithms are applied to will contain objects such as stars or galaxies and the associated features, the input variables, are measured properties, such as spectra or light- curves. An example of a classification task would be the algorithm predicting output variables of either "star" or "quasar" where the objects spectra was the input variable. Whereas an example of regression would be the algorithm estimating a redshift value.

3.3.2 Model construction and performance

The development and evaluation of a machine learning model involves training, validation, and testing. The training comprises learning the model's parameters (weights) using a training dataset. A validation dataset is used to adjust the model's hyperparameters, such as the learning rate and the regularisation coefficient, as part of the validation phase. The testing technique involves evaluating the performance of the final model with a fixed hyperparameter using an independent test dataset (Hastie et al., 2009). In practice, where there are limitations on dataset size there is usually a trade-off between estimation accuracy in the training phase and testing accuracy depending on how the dataset is split.

Quantitative performance metrics are numerical values that are used to evaluate the performance of a machine learning model. These metrics can be used to compare different models or to compare the performance of a single model under different conditions. Some common quantitative performance metrics include accuracy, precision, recall, and F1 score. These metrics are commonly used in classification tasks, although they can also be applied to other types of tasks such as regression or clustering.

Accuracy is the proportion of correct predictions made by the model out of all the predictions made. It is calculated as the number of correct predictions divided by the total number of predictions. Precision is the proportion of positive predictions that are actually correct. It is calculated as the number of true positive predictions divided by the sum of true positive and false positive predictions. Recall is the proportion of positive cases that were correctly predicted. It is calculated as the number of true positive predictions divided by the sum of true positive and false negative predictions. F1 score is a metric that combines precision and recall. It is calculated as the harmonic mean of precision and recall, with a higher score indicating a better balance between the two. The F1 score is often used as a single metric to compare the performance of different models, as it takes into account both the precision and recall of the model.

A confusion matrix, as shown in figure 3.8, is a table that is used to visualise the performance of a machine learning model on a classification task (Devroye et al., 2013). The rows of the table represent the actual classes, and the columns represent the predicted classes. The entries in the table show the number of samples that were predicted to belong to each class. The diagonal elements of the confusion matrix represent the number of samples that were correctly classified, and the off-diagonal elements represent the number of samples that were misclassified. The confusion matrix can be used to compute various performance metrics, such as precision, recall, and F1 score.

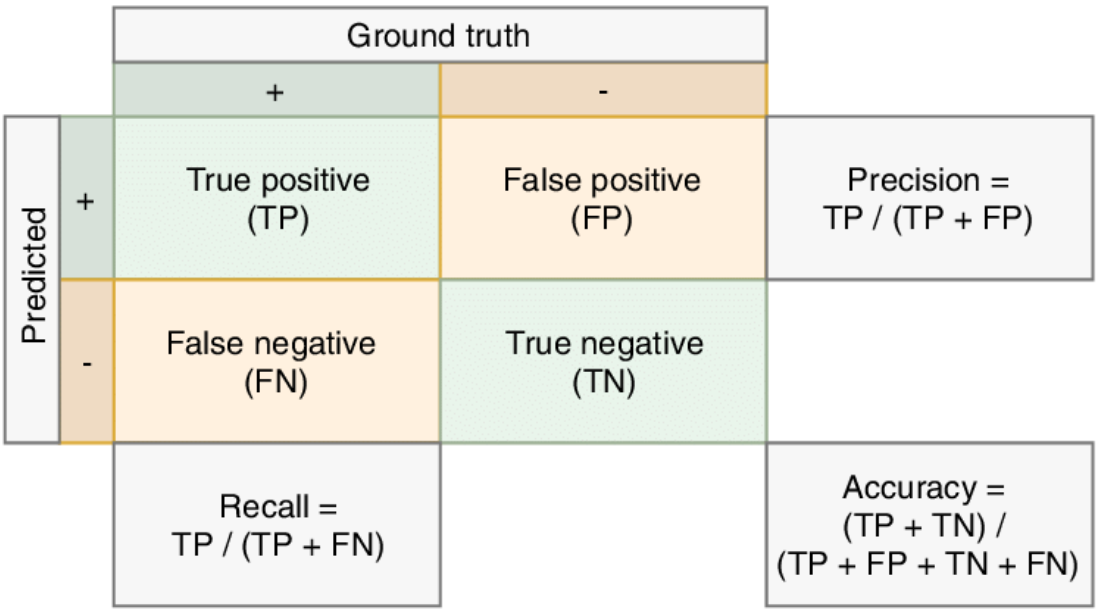

FIGURE 3.8: Confusion matrix and evaluation metrics for two-class classification (Jeppesen et al., 2019)

3.4 Machine Learning Techniques Applied in Astronomy

3.4.1 Signal and Object Detection

Being able to detect and analyse signals quickly and efficiently is an important task for an astronomer. It is often required that follow-up observations, particularly in other wavebands that can paint a clearer picture of the source's nature, need to occur on very short time scales in order to be able to produce the best scientific results. Some areas where signal detection is essential are transient events, variable sources, gravitational wave signals as well as in the fields of space science and planetary exploration and environmental monitoring. Often the first stage in classification of astronomical objects is detecting if an object is present above a particular signal-to-noise ratio (SNR). In many of these cases it is essential to process the data quickly to allow rapid discovery, but even today this can be computationally intensive, so a challenge for future datasets will be to perform this task automatically, accurately and quickly. Machine learning techniques are already being used in signal detection, including using ANNs for source detection in images (Bertin and Arnouts, 1996), and gravitational wave signal classification in time series data (Vinciguerra et al., 2017). These methods will become essential when the next generations missions begin to deliver data.

3.4.2 Object Classification

Classifying astronomical objects is an important initial stage in the scientific process. By placing objects into different classes it enables us to make generalisations about the classes and also allows us to state hypotheses and to make comparisons with models. It also allows quick identification of rare or new objects for follow-up observations optimised to account for object classes. The next generation of telescope will require more automated techniques to search a single or combined database for particular classes of interesting or rare objects. Also ML methods can be used to classify new objects automatically; ML classifiers can be constructed using data from old missions or from an earlier stage in the mission and used as training sets, and then the classifier can be applied to new data to automatically determine what class an object is. Astronomers are already using several ML methods to perform object classification, including:

- Solar flare classification with SVMs (Raboonik et al., 2016)
- Classification of high-energy (X-ray and gamma-ray) sources using Naive Bayes (Broos et al., 2011), random forests (Scaringi et al., 2008; Lo et al., 2014), or ANNs (Salvetti, 2016)
- Classification of optical transients, including variable stars and supernovae, using decision trees (Mahabal et al., 2009), boosted decision trees (Lochner et al., 2016), random forests (Bass and Borne, 2016), and deep neural networks (Charnock and Moss, 2016)
- Quasar-star classification with SVMs and KNN (Peng et al., 2013)
- Photometric redshift template selection with random forests (Fotopoulou et al., 2016)
- Eclipsing binary light curve classification with LDA and RF (Süveges et al., 2017)
- Gravitational wave signal classification with ANNs (Mukund et al., 2016)

Though individual results vary, these classification methods show high accuracy (typically 90-95% successful classification) and completeness.

3.4.3 Star-Galaxy Separation

Almost all stars are unresolved in photometric datasets and so appear as point sources because of their relatively small size compared to the distance between us. Other sources such as supernova and quasars for example, are also seen as point sources, whereas galaxies are further away but generally subtend a larger angle so appear to us

as extended sources. Due to this, it is an important task for astronomers to be able to categorise what objects are by the separation of photometric catalogues into stars, galaxies and other astrophysical objects. Already work has been done to automate this process as the number of sources in a typical survey is already large. For example, Sevilla-Noarbe and Etayo-Sotos (2015) use boosted decision trees for star-galaxy separation to achieve star impurity levels less than 1% in their galaxy sample.

3.4.4 Galaxy Morphology

Galaxies can be classified by their morphology, which is grouping them by their size, shape or visual appearance. The most well known galaxy morphological classification scheme is the Hubble sequence - which broadly puts galaxies into four classes, ellipticals, lenticulars, spirals and irregulars. ML classifiers can be usefully applied here. Using supervised methods features can be extracted from the images of galaxies with known morphology and used to build classifiers which can then be applied to new data. Ferrari et al. (2015) used linear discriminant analysis to achieve greater than 90% successful morphological classifications, and Bazell and Aha (2001) used both naive bayes and ANNs to classify morphology of galaxies. Unsupervised methods can also be applied to datasets where the initial distribution of classes is unknown.

There has also been use of what has come to be known as citizen science, particularly made popular by Galaxy Zoo (Lintott et al., 2008). This is where members of the general public can view images of different galaxies and assign them to particular classes. This labels the images which then can be used to form a training set to build ML classifiers. The full final sample of citizen-science-labelled galaxy morphologies have allowed Galaxy Zoo to build ML classifiers that achieve optical galaxy morphological classification with 97% accuracy (Domínguez Sánchez et al., 2018) and radio galaxy classification with greater than 90% accuracy (Wu et al., 2019).

3.4.5 Time Series

Time series analysis looks at an ordered sequence of values of a variable at time intervals. In an astronomy context this could mean investigating periodic processes such as orbital and rotational times of astrophysical objects, or exploring transient data related to supernovae, gamma-ray bursts and other sources from detection, identification and classification or looking at stochastic processes relating to accreting systems such a neutron stars and black holes. ML classification of light curves (LC) can be used in astronomical time series analysis. A LC is a graph showing the variation in the light received over a period of time from an object. They can aid the rapid detection and discovery of events in real time, which then allows quick decisions to be made about which objects are to be followed up for detailed study.

Some work in supernova classification using time series photometric data has employed random forests (Zaidi and Narayan, 2016) and boosted decision trees (Möller et al., 2016) achieving 95% purity.

3.4.6 Photometric Redshifts

It is known that each chemical element emits photons at specific wavelengths and these are seen as either absorption or emission lines in the spectra of astronomical objects. By observing the position of these lines, we can tell which elements are present in a source. In astronomy, a phenomena known as cosmological redshift occurs, which sees these spectral lines moved towards the red end of the spectrum. The wavelength at which the radiation is originally emitted is lengthened as it travels through our expanding universe. Astronomers are able to determine how far away distant objects are by measuring this wavelength expansion. High resolution spectroscopy is the most reliable approach to measure redshift, but is observationally expensive compared to broadband photometry. Reliable measurements of redshift from imaging data covering a few different wavebands, known as photometric redshifts, is a more desirable option for obtaining redshifts for large quantities of sources.

During recent years, additional work done has been done using photometric data to estimate redshift. There is much more photometric data available than spectral data, so although photometric redshifts are less accurate, they allow for much larger sample sizes for statistical studies. Astronomers have been using a variety of ML to determine photometric redshifts for galaxies including Naive Bayes (Carrasco Kind and Brunner, 2014), SVMs (Jones and Singal, 2016), KNN (Beck et al., 2016), and boosted decision trees and ANNs (Wolf et al., 2016). These methods have shown great success in estimating photometric redshifts for normal galaxies with errors in $\Delta z/(1+z)$ of order $0.01 - 0.03$. In active galactic nuclei and quasars, it can be a more difficult task to estimate photometric redshift, as bright and broad emission lines dominate the sources spectra. ML techniques have been applied to this area to solve this problem, such as a CNN-based approach (Pasquet-Itam and Pasquet, 2018) that yields quasar redshift accuracy of 0.3 (for redshifts up to around $z = 6$).

3.4.7 Outliers

One of the greatest desires of astronomers is to discover and learn about unusual, rare or unknown types of astronomical objects or phenomena. One way of finding said objects is to detect and identify outliers as they often provide useful information. An outlier can be described as an observation that deviates so much from other observations so that it arouses suspicions that it is generated by a different mechanism (Hawkins, 1980). Outliers are often ignored, considered noise or removed from

datasets, but these objects could lead to the discovery of totally unexpected new classes of objects. As well as the potential to discover new types of objects, outlier detection is also useful in enabling good data analysis. Outliers in data can impact results and distort the statistical significance of a sample. Therefore developing rigorous methods to identify outliers and what impact they have on results is an important task in large scale datasets (Garg and Kalai, 2016). Astronomers have employed a variety of techniques to find outliers: unsupervised random forests to detect unusual galaxies (Baron and Poznanski, 2017) or stellar systems (Reis et al., 2018); unusual quasar selection using self-organising maps (Meusinger et al., 2012); anomalous stellar system identification in Kepler data using unsupervised random forests coupled to manifold-learning algorithms (Martínez-Galarza et al., 2021); and unusual optical transient identification using a combination of unsupervised random forests, a KNN-based density approach, Gaussian mixture models, and support vector machines (Malanchev et al., 2021).

3.4.8 Limitations of ML in astro

Despite their utility in various astronomical activities, many Machine Learning techniques were not created for noisy and incomplete astronomical datasets. In particular, measurable features usually contain a wide range of uncertainty values, and these uncertainties are frequently disregarded during model training. In actuality, the performance of Machine Learning algorithms is heavily dependent on the signal-to-noise ratio of the objects in the sample, and a model designed for a dataset with specific noise characteristics will fail when applied to a dataset with different noise characteristics. In addition, whereas in computer vision the labels provided to the algorithm are regarded as "ground truth" (for example, the categorisation of cats and dogs in photographs), in astronomy the labels may be ambiguous. For example, in a classification task of "authentic" versus "fake" in transient detection on multiple images (Bloom et al., 2012) the labels in the training set are taken from the manual categorisation of scientists and citizen-scientists. Some may regard a certain event to be "authentic," while others may label it "fake." In addition, labels inside the training set may be the output of a different algorithm that generates labels accompanied by uncertainty. Similarly, the bulk of Machine Learning algorithms ignore such uncertainty.

Recently a modified Random Forest was deployed to account for measurement (feature) and class uncertainty (Reis et al., 2018). The Probabilistic Random Forest method considers features and labels as random variables rather than deterministic values, with each random variable represented by a probability distribution function whose mean is the reported measurement and whose variance is the stated uncertainty. Tests indicated that the Probabilistic Random Forest outperforms the

traditional Random Forest when applied to datasets with varied noise levels, with up to a 10 percent improvement in classification accuracy when using noisy features and a 30 percent improvement when using noisy labels. Probabilistic Random Forest accommodates missing values in the data naturally and outperforms Random Forest when applied to a dataset with distinct noise characteristics in the training and test sets.

When applying supervised learning algorithms to astronomical datasets, uncertainty treatment, transfer of information, and interpretability of the resulting models are the major challenges. As previously stated, the majority of supervised learning algorithms are not built for astronomical data; they implicitly assume that all the dataset's features are of a comparable quality and that the provided labels can be considered ground truth (Baron, 2019). However, astronomical records are noisy and insufficient, and the designations provided by human professionals are frequently confusing. Therefore, supervised learning techniques function best when applied to datasets with a high signal-to-noise ratio or stable noise characteristics. The efficiency of supervised learning algorithms is largely dependent on the noise characteristics of the objects in the sample; hence, an algorithm trained on a dataset with particular noise characteristics will not generalise to a similar dataset with other noise characteristics. Consequently, it is essential to tweak existing tools and build new algorithms that account for data uncertainty during model development. In addition, these algorithms must provide prediction uncertainties based on the intrinsic properties of the sample items and their respective measurement uncertainties.

The second hurdle when applying supervised learning algorithms to astronomical datasets is the transfer of knowledge. In other words, an algorithm trained on a particular survey with a particular instrument, cadence, and object targeting selection will typically fail to generalise to a different instrument with different characteristics, even if the object's intrinsic properties are comparable in both the surveys. Consequently, machine learning algorithms are deployed largely to finished surveys and infrequently to ongoing polls that have not yet amassed sufficient labelled data.

Knowledge transfer is of particular significance when searching for uncommon phenomena where supervised learning methods trained on simulated data cannot generalise effectively to real datasets. This challenge can be overcome through transfer learning strategies (see, e.g., Weiss et al., 2016, for a review). Although these methods are discussed in computer science literature, they are rarely employed in astronomy. The interpretation of the resulting models is the third problem related with the use of supervised learning techniques to astronomical datasets. Even though supervised learning algorithms provide an extraordinarily flexible and versatile framework for the building of complex decision functions, beating conventional algorithms in classification and regression tasks, the resulting models are typically challenging to

analyse. Thus, it is not always possible to comprehend what the model has learned and why it reaches particular results.

Typically, as scientists, we wish to comprehend the created model and the decision-making procedure, as this understanding can teach us something new about the underlying physics. This challenge is especially significant for cutting-edge deep learning systems, which have been shown to succeed in a variety of tasks. As we continue to develop more advanced classification and regression tools, it is essential that we also develop methods for understanding their outcomes. Interpreting the results and comparing multiple unsupervised learning methods are the most challenging components of applying unsupervised learning algorithms to astronomical datasets. Because unsupervised learning algorithms frequently optimise an internal cost function that does not necessarily correspond with our scientific motivation, and because these algorithms are not trained according to a definition of "ground truth", their results may lead to incorrect interpretations of trends and patterns in our datasets. Numerous cutting-edge algorithms are modular, allowing us to construct a cost function more suited to the current task. Consequently, it is essential to develop cost functions that better fit with our scientific goals. Analysing and comparing the outputs of an unsupervised learning algorithm with those of other algorithms still requires domain expertise; the process cannot be automated.

Chapter 4

A deep learning approach to source detection in hard X-ray surveys

Thanks to the current generation of space telescopes with survey capabilities, the sky in the hard X-ray / soft gamma-ray band (approximately 10keV-1MeV energy) has shown itself to be both well-populated and highly variable. Surveys in this band are normally carried out using coded aperture telescopes that provide a good sensitivity across a wide field of view (typically $>$ 100 square degrees), and as such allow frequent returns to the same sky region, producing rich datasets with information in both spatial and temporal dimensions.

However, the analysis of data from coded aperture telescopes is not trivial, as it is an indirect imaging method and sky images can contain systematic noise as a result of an imperfect instrument model, and also when an adequate description of the source distribution cannot be determined, as is sometimes the case in crowded regions where sources cannot be fully resolved. The sources in the hard X-ray sky display a huge dynamic range, and are thus detectable on many different timescales. While the brightest sources can be detected in a single observation, the faintest sources may require 1000s of images to be co-added. More recent surveys (Bird et al., 2010, 2016) have searched for ways to detect sources on all timescales in an efficient way, but these are generally expensive in operator effort, and as the data from such surveys is ever increasing, there is a need for automated techniques that can scale with the data when it exceeds the capacity to be processed manually.

This chapter presents a new method to efficiently search for excesses in sky maps that may also contain systematic artefacts of the imaging process. Following on from this, the next chapter presents how to intelligently combine lists of excesses found in multiple maps. These maps may be generated in different energy bands, in different sky orientations, or at different times - but all may be considered potential information concerning each putative source. Due to the dynamic range across the source

population, a bright persistent source may be detected 1000s of times in individual observations (as well as in any co-added maps) whereas information on a fainter source may be determined from just co-added images. Conversely a bright transient may be seen in just a few observations, and be completely undetectable in a co-added image mosaic. The challenge is to combine all the excess detections into a coherent source catalogue, suppressing statistical noise in the presence of systematic effects, efficiently and without introducing the subjective biases that human intervention can produce.

Specifically in this chapter, images produced by the INTEGRAL/IBIS telescope in the ∼18-100 keV band are used (though it is worth noting that the low energy sensitivity limit for IBIS/ISGRI has degraded over time to now approximately 25 keV), but the methods would be equally applicable to the Swift/BAT telescope images as well as other instruments.

As previously discussed both the Swift/BAT and INTEGRAL/IBIS use coded aperture masks to create images of the X-rays that are detected as already described. In Swift/BAT, the coded mask is made up of a grid of 52,960 square apertures, each of which is 4.1 mm on a side. The detector is made up of a grid of 32,768 individual detectors, each of which measures 4 mm by 4 mm. The resulting images have a resolution of ∼22 arcminutes (∼0.37 degrees) (Barthelmy et al., 2005).

In INTEGRAL/IBIS, the coded mask is made up of a pattern of 16384 square apertures, each of which is 3.7 mm on a side. The IBIS/ISGRI detector is made up of a grid of 4096 individual detectors, each of which measures 4.5 mm by 4.5 mm. The resulting images have a resolution of ∼12 arcminutes (∼0.2 degrees) (Ubertini et al., 2003).

There are other instruments that also use coded aperture masks to create X-ray images, such as the All-Sky Monitor (ASM, Levine et al., 1996) on board the Rossi X-ray Timing Explorer (RXTE). These instruments have similar technical data relationships to Swift/BAT and INTEGRAL/IBIS, but they also have some differences in their designs and resolutions.

These similarities mean it is possible to apply the method presented in this chapter to these different instruments, although the differences mean that retraining of the method presented in this chapter with data from the individual missions would have to occur.

This chapter introduces new tools that have been developed using deep learning techniques to improve how multiple IBIS/ISGRI maps are searched for sources. First the specific dataset for this analysis will be introduced, along with a brief description of previous methods for source detection. A description of early tests exploring different ML models follows, where traditional methods are compared to a more

modern and more effective deep learning model. The final architecture of the deep learning method and the training process are then discussed, including how the test and training set were generated. A comparison of this new method to the source detection tools used in producing recent catalogues can be found at the conclusion of the next chapter, where the crucial technique of excess merging for final catalogue production is described.

4.1 How were sources identified in previous catalogues?

Cat1000 used both the standard astronomy package SExtractor (Bertin and Arnouts, 1996) and a custom-made piece of software called *peakfind* to detect sources in stacked ISGRI maps. This chapter presents a new tool using a deep learning technique – specifically, a convolutional neural network (CNN) – to search ISGRI maps to detect sources. Cat1000 used a custom-built piece of software called *megamerge* to combine the detections found in multiple maps. This thesis presents a new method using Bayesian matching (Budavári and Szalay, 2008) that removes some of the bias that was inherently part of the *megamerge* process (see Chapter 5). Together source detection and excess merging form the key steps for survey catalogue production, and this section presents further detail on source detection approaches for cat1000.

In cat1000 approximately 67000 ScWs were used to create stacked maps using a purpose-built image mosaic tool which was developed to statistically average the images from multiple input maps. This allowed all-sky maps (these all-sky maps have a pixel size of 2.4 arcminutes) to be created from a large number of input ScWs (recall Figure 2.23 from Chapter 2 shows the entire cat1000 process from the ScW images to the final sources list). Mosaics were constructed for five energy bands (17-30,30-60,18-60,20-40 and 20-100 keV) and in four sky projections: centred on the Galactic Center, on the Galactic anti-center, north Galactic polar and south Galactic polar. These projections were chosen in order to reduce PSF distortion which impedes the source detection algorithms used.

60 all-sky maps and over 19000 revolution maps were constructed and searched for flux excesses indicating an astrophysical source to produce an initial excess list. Two different techniques were used to search the mosaics maps, the standard SExtractor tool, and *peakfind* which takes into account the varying levels of systematic background. SExtractor has sophisticated algorithms for pre-filtering the image to enhance detection of specific PSFs, and is capable of de-blending some complex source regions, both of which are important for ISGRI maps. However, it does not work so well with local variations in systematic noise levels, which are a feature of the ISGRI maps. *peakfind* uses a recursive search around the peak position to detect excesses, maintaining some de-blending capability but performing only limited tests on the

shape of the PSF detected. The main benefit of *peakfind* is that it performs a local assessment of the image noise and assesses excesses relative to that local background. As such it is much less vulnerable to over-detection in noisy areas of the maps.

The use of multiple excess detection algorithms was valuable in cat1000 production as the different methods used could be compared and an excess appearing in both lists could be treated with higher confidence. However the different underlying approaches meant that all excesses still needed manual checking as complex regions were often interpreted differently by the two algorithms, and this was a natural path for operator bias to be introduced.

During cat1000 production, both SExtractor and *peakfind* had to be run on a large number of maps in five individual energy bands, for different sky projections and on different timescales - revolution level, observation sequences and whole-archive. Although the performance of the methods was similar, the time taken for this and the subsequent combination of the excesses found in these maps was substantial because this complex task placed a high demand on both computing and human resources.

For future catalogues, it is hoped that new techniques can be developed to make this task more tractable - and indeed extend it to the ScW timescale data which had not been attempted for previous catalogues. The use of HEALPix (Hierarchical Equal Area isoLatitude Pixelation) based maps will reduce the number of sky projections needed, and an image search that combines multiple energy bands will not only save time but may also provide a more robust detection as the energy bands are not completely independent and it is expected that a source appearing in one energy band to appear in some, but not necessarily all, others. Such combination logic was part of the manual inspection of the excesses, and an automated method which took the same approach should be less vulnerable to random image noise. Unfortunately this is still not fool-proof, as systematic noise and ghost sources appear at the same position in every energy band. In principle, an automated method could also recognise sensible ratios of flux in different energy bands as the typical X-ray source spectra give rise to fairly predictable flux ratios across the energy bands.

These complex requirements naturally motivate the exploration of machine learning models for source detection. In the sections that follow, the results of initial exploration of traditional ML methods will be presented and compared to the more promising results achieved with deep learning.

4.2 Preliminary exploration of ML models

Several different machine learning algorithms were explored and their performances measured whilst trying to ascertain which would he most effective in developing a

source detector. During this exploration stage, a small subset of the dataset, revolution 60, was used. This revolution is generally accepted amongst the INTEGRAL collaboration as a good example of high-quality data suitable for testing new analysis techniques. The data were tested with both traditional ML approaches (Section 4.2.1) and deep learning techniques (Section 4.2.2). A final architecture based on CNNs was selected for deployment on the full INTEGRAL/ISGRI dataset and is described in greater detail in the section that follows (Section 4.3).

4.2.1 Machine Learning Methods

For initial exploration of ML methods, ScW images were broken down into 11x11 pixel windows and a domain expert was asked to suggest several features that could be extracted from the windows to enable the algorithms to classify if a source was present. This window size, which corresponds to roughly 26 arminutes across, was chosen to be large enough to contain a typical source but small compared to the average distance between adjacent sources.

The first feature was generated by fitting a 2D Gaussian over a 11x11 pixel window. The peak of the Gaussian was calculated to a sub-pixel level and then the maximum amplitude was calculated. The second feature was local std from the intensity maps. The third feature was the maximum value taken from the significance map in the 11x11 window.

A training set was created by first using the OSA generated source list to determine what the pixel coordinates were for each known source in the revoultion. A process called sigma clipping was then used. This had the effect of removing the sources from the image, and produced a masked array where anything above a certain threshold was excluded. This allows the local noise to be calculated on the intensity maps. The other features were extracted for each source, which were: (i) the peak flux for a gaussian fit to the flux profile, (ii) the hardness ratio (20-40 keV compared to 60-100 keV) of the source, and (iii) its galactic coordinates. Then examples of non-sources were generated - everything else in the image that is not an actual source. This ranges from zero-valued areas to noise and other excesses in the images that are not sources. The mask-array that was generated in the initial sigma clipping stage was used as this had already removed all the sources. Then a sliding 11x11 pixel window, was moved over each image extracting the same features as for the sources. 70 percent of both the sources and the non-sources were used for the training set, and the remaining 30 percent were held back to be used to test the classifier's performance.

Using the training set, six classifiers were built: Linear Discriminant Analysis, KNN, Decision Trees, Random Forest, a BDT and an ensemble classifier called RUSBoost (Random Under Sampling Seiffert et al., 2010). RUSBoost works very well when

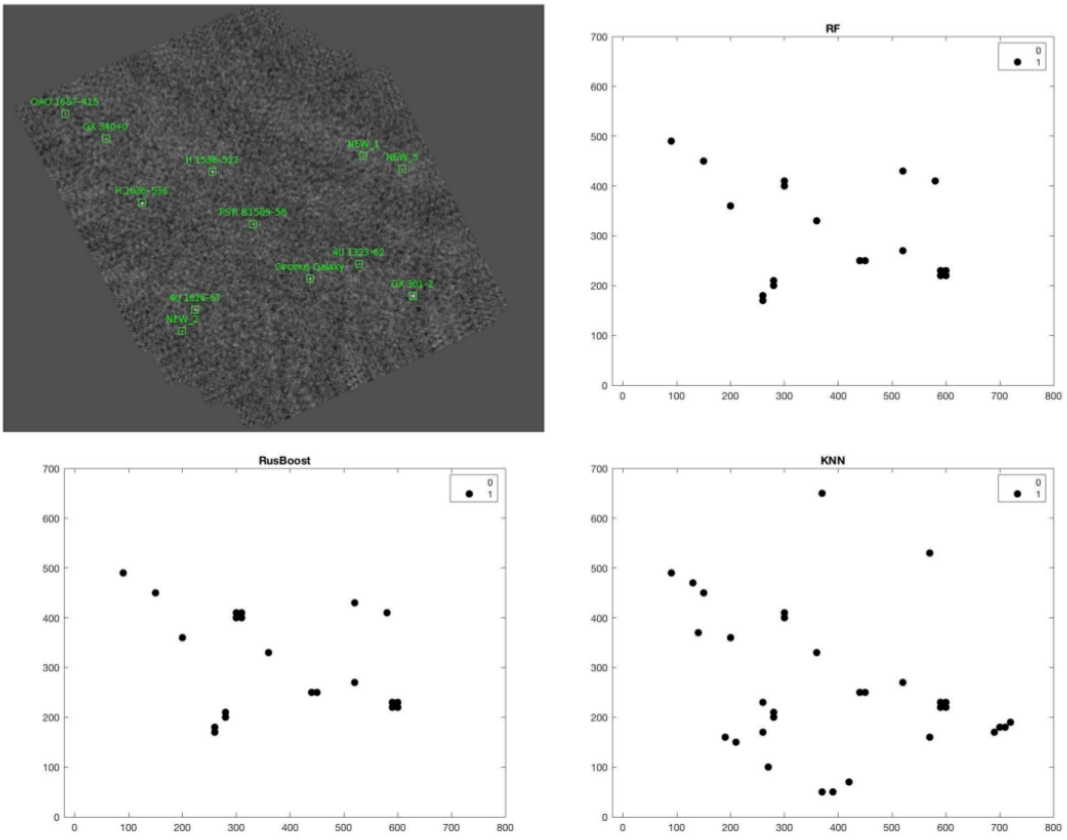

FIGURE 4.1: Images of sources in rev 1786 from OSA (top left) compared to locations (pixel coordinates) of sources from from three ML methods: Random Forest (top right), RUSBoost (bottom left) and KNN (bottom right).

dealing with imbalanced datasets. This means some classes in the training data have fewer members than others. In this case there are many more examples of non-sources than sources. The algorithm takes N (the number of members in the class with the fewest members in the training data) as the basic unit for sampling. Classes with more members are under-sampled by taking only N observations of each class. In other words, if there are K classes, then, for each weak learner in the ensemble, RUSBoost takes a subset of the data with N observations from each of the K classes. The boosting procedure follows the procedure in AdaBoostM2 (Eibl and Pfeiffer, 2005) for reweighting and constructing the ensemble.

As Table 4.1 shows, several of the classifiers performed well (Decision Tree, Random Forest and RUSboost) on the test data. For RUSBoost this is most probably to do with the fact it works extremely well on unbalanced datasets. RUSboost only misclassified four of the sources, and on further inspection, these sources were only present in a single ScW, so would be very weak sources. Decision Tree methods often work well when dealing with classification problems and in this case works well as the data was non-linearly separable. This maybe why linear discriminate analysis had the lowest performance as it works best when the data can be linearly separated. NRT data for

LDA - Training Set	Ground Truth Source	Ground Truth Background
Predicted Source	123	207
Predicted Background	176	50169

LDA - Test Set	Ground Truth Source	Ground Truth Background
Predicted Source	65	84
Predicted Background	63	21506

kNN - Training Set	Ground Truth Source	Ground Truth Background
Predicted Source	299	19
Predicted Background	0	50357

kNN - Test Set	Ground Truth Source	Ground Truth Background
Predicted Source	126	12
Predicted Background	2	21578

DT - Training Set	Ground Truth Source	Ground Truth Background
Predicted Source	299	18
Predicted Background	0	50358

DT - Test Set	Ground Truth Source	Ground Truth Background
Predicted Source	123	6
Predicted Background	5	21584

abKD - Training Set	Ground Truth Source	Ground Truth Background
Predicted Source	299	633
Predicted Background	0	49743

abKD - Test Set	Ground Truth Source	Ground Truth Background
Predicted Source	122	286
Predicted Background	6	21304

RUSBoost - Train. Set	Ground Truth Source	Ground Truth Background
Predicted Source	295	8
Predicted Background	4	50368

RUSBoost - Test Set	Ground Truth Source	Ground Truth Background
Predicted Source	124	6
Predicted Background	4	21584

RF - Training Set	Ground Truth Source	Ground Truth Background
Predicted Source	299	4
Predicted Background	0	50372

RF - Test Set	Ground Truth Source	Ground Truth Background
Predicted Source	123	4
Predicted Background	5	21586

TABLE 4.1: Confusion matrices for various classifiers applied to ISGRI sources and background, for both training and test sets.

revolution 1786 (the recent revolution at time of this early analysis) was then download and run through some classifiers. Random Forest and RUSBoost performed very well, and identified all the OSA software sources. Figure 4.1 shows this and the result for KNN is also included as a comparison. Although for this example (only using revolution 60) our dataset was fairly small, both Random Forest and RUSBoost can deal with much larger datasets.

4.2.2 Deep Learning Method

4.2.2.1 Convolutional Neural Networks

Convolutional neural networks (CNN) developed in recent years are most commonly applied in the field of image processing because they perform well at dealing with image recognition and classifications tasks and are considered to be one of the leading techniques in the field (LeCun et al., 1995). In the absence of domain knowledge they can work well with raw features; a CNN automatically learns the underlying features required to detect when a source is present (Schmidhuber, 2015). These distinct advantages make a CNN the ideal choice for source detection in high energy astronomy images. A CNN is a supervised method and thus requires a training set to be run through a CNN many times, adjusting the CNN's parameters using backpropagation to minimise a loss function (LeCun et al., 1988).

In image classification, pixels that are near each other are quite likely to be more related than two pixels that are further away. This means that the pixels' proximity to one another is an important factor whilst classifying and CNNs specifically take advantage of this fact (LeCun et al., 2015). In a standard neural network, every pixel is linked to every single neuron, in the case of image classification this added computational load makes training more difficult and resulting models are often less accurate. A CNN removes a lot of these less significant connections, and makes the image processing computationally manageable through filtering the connections by proximity. In a given layer, rather than linking every input to every neuron, CNNs restrict the connections intentionally so that any one neuron accepts the inputs only from a small subsection of the previous layer. Therefore, each neuron is responsible for processing only a certain portion of an image. Combined with effective training methods for deep layer networks this provides a powerful approach.

As demonstrated in Figure 4.2 an image in a CNN is passed through a combination of successive layers:

- *Convolutional layer*: here the filters can be thought of as feature identifiers.

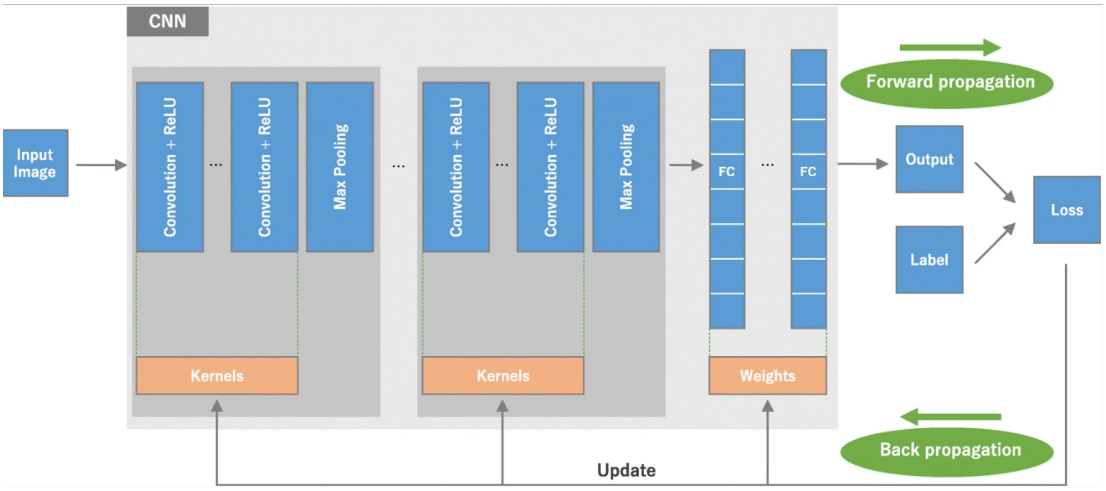

FIGURE 4.2: This image shows a typical architecture of a CNN, including the input layer, convolutional layers, pooling layers, fully connected layers, and output layer (Yamashita et al., 2018).

- *Nonlinear layer*: the CNN uses ReLu (negative inputs are zeroed) (Hahnloser et al., 2000). This allows the network to approximate arbitrary functions by introducing non-linearities. (Nair and Hinton, 2010)
- *Max Pooling layer*: this down-samples the input, thereby increasing the network's efficiency and allowing the network to train quicker (Nagi et al., 2011)
- *Fully connected layers*: neurons in a fully connected layer have connections to all activations in the previous layer. The CNN can then use a softmax activation function to produce the final output: a probability of the input being a source and not a source (and ensure a partition of unity) (Bishop, 2006).
- *Categorical crossentropy* : This is a loss function that is used in multi-class classification tasks where an example can only belong to two or more label classes, and the model must decide which one. It computes the loss between the labels and predictions. (Zhang and Sabuncu, 2018).

A CNN is a very powerful and efficient model which, unlike some other machine learning methods, performs automatic feature extraction. The network picks out the important features in an image in order for it to make highly accurate classifications. In fact CNNs can outperform humans in image classification due to the networks' ability to pick out underlying patterns and structures that domain experts can be unaware existed (LeCun et al., 2015).

4.2.2.2 CNN results

A CNN was trained using examples of sources and background extracted in 11×11 pixel windows from 3 channels (intensity, significance and exposure maps) in the energy band 18-60 keV. It was important when selecting these examples that the sources included had a very high confidence of their detection and classification. This was to try and keep any human biases out of the network. The sources that were used were all from INTEGRAL's 60th revolution. This revolution included 106 ScW and occurred in 2003, allowing plenty of time for followup observations and further analysis to have been undertaken.

A dataset with examples of sources and background in 11×11 pixel window was generated; this size was picked as it was large enough to contain the whole source, but not too large that other sources would be in the window. The windows were then split into a training set (70%) and a test set (30%). The test set would not be used in training the CNN but would be used to produce a measure of the network's perform.

The dataset was very unbalanced, there were many more examples of background (71,966) than of sources (427). In ML, an unbalanced dataset is one in which the number of examples in each class is significantly different from the others. For example, a dataset for a binary classification problem might have 90% of the examples in one class and only 10% in the other. This is a common problem in real-world applications where certain classes might be much rarer than others.

When dealing with an unbalanced dataset, it's important to use an appropriate loss function to train the model. The choice of loss function can have a significant impact on the model's ability to learn the minority class. One common approach is to use a loss function that is inversely proportional to the class priors, which reflects the relative importance of each class in the dataset (He et al., 2008). This method was used in this scenario to stop the network just classifying everything as background and still producing a high performance percentage.

Fundamentally this is an image classification task where the CNN learns the features which separate a window containing a source from the background. The training set is run through CNN many times adjusting the CNN's parameters to minimise a loss function. The window is passed through a series of layers as already described.

Once the network was fully trained, a 11×11 pixel window could be moved across an entire astronomical image which could contain many sources. Figure 4.3 shows how the network takes the input window image, passes it through a series of convolution, nonlinear activation (ReLu), pooling (downsampling), and fully-connected layers and then returns an output for each window, with the assumption that each window contains only one source. The network was trained using 50376 examples of

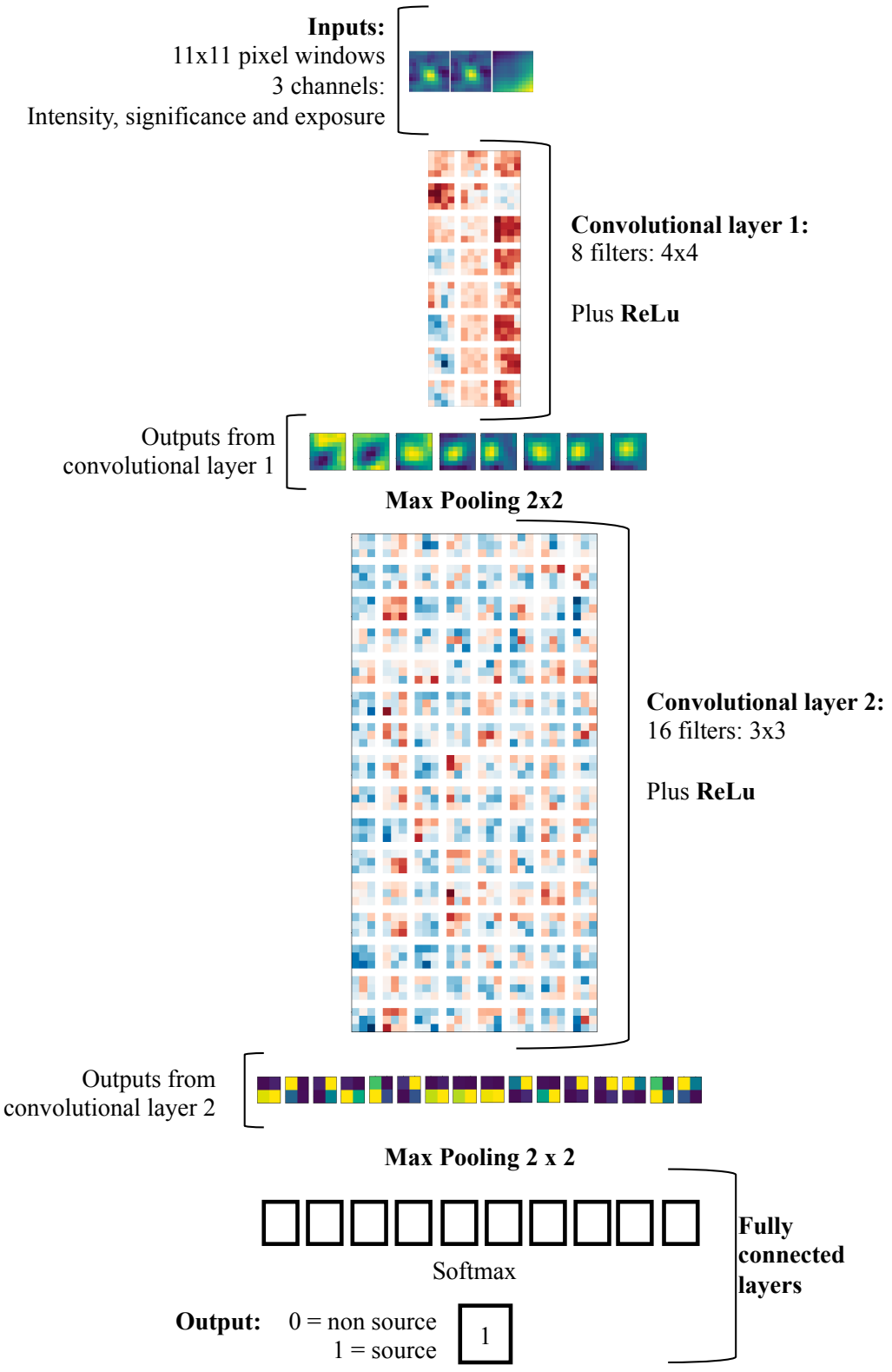

FIGURE 4.3: The architecture of the source detector CNN. This shows how a 11×11 pixel window with a source present is processed through the CNN resulting in its detection.

TABLE 4.2: A selection of the different CNN architectures tried when producing the ISGRI source detector

CNN Number	Convolution layer 1	Max Pooling	Convolution layer 2	Max Pooling
1	10 filters 4x4	2x2	5 filters 3x3	2x2
2	8 filters 3x3	2x2	8 filters 3x3	2x2
3	32 filters 4x4	2x2	16 filters 3x3	2x2
4	16 filters 3x3	2x2	8 filters 3x3	2x2
5	15 filters 4x4	2x2	8 filters 3x3	2x2

background and 299 examples of sources. A test set of unseen data consisted of 21560 background examples and 128 source examples.

The preliminary results, as seen in Table 4.3, show this CNN has excellent performance and indicates that this CNN could not only decrease the time taken to detect sources in the maps, but could also remove the human bias which is sometime seen in the decision making process of source detection.

As previously discussed CNNs are excellent tools at finding the underlying features in images which allow the network to classify with a high accuracy. Other methods were tried, such as random forest, boosted decision trees (RUSBoost) and other traditional ML methods. Although they were quicker to train, they did not perform as well (see results from Section 4.2.1). All these methods performed reasonably well on finding the sources, although they struggled on the noise that the OSA software had labelled as sources and these methods also picked these examples out as sources too. As seen in Table 4.5 the CNN performed well on not misclassifying noise as sources.

One of the challenges of observing X-rays is that they are difficult to focus, as they are absorbed and scattered by the materials they pass through. This can lead to the formation of ”ghosts” in the images produced by the telescope. Ghosts are faint, blurred copies of the real sources that are shifted and distorted from their true positions as seen in Figure 4.4.

The cause of ghosts in IBIS/ISGRI images is related to the way the telescope is designed and its coded mask. The X-rays that pass through the mask create an image on the detector, which is then reconstructed using a mathematical algorithm called deconvolution. However, the deconvolution process is complicated by the fact that the X-rays are scattered and absorbed by the materials they pass through, causing distortions in the image. These distortions can create ghost images of the real sources, which are shifted and blurred versions of the original source. Additional ghosts are caused by the physical structures (screws and glue) affixing the mask to its support structure, which results in ghosts when not accounted for in the mask model

Arch. 1 - Training Set	Ground Truth Source	Ground Truth Background
Predicted Source	297	4
Predicted Background	2	50372

Arch. 1 - Test Set	Ground Truth Source	Ground Truth Background
Predicted Source	123	8
Predicted Background	5	21582

Arch. 2 - Training Set	Ground Truth Source	Ground Truth Background
Predicted Source	298	3
Predicted Background	1	50373

Arch. 2 - Test Set	Ground Truth Source	Ground Truth Background
Predicted Source	126	1
Predicted Background	2	21589

Arch. 3 - Training Set	Ground Truth Source	Ground Truth Background
Predicted Source	299	0
Predicted Background	0	50376

Arch. 3 - Test Set	Ground Truth Source	Ground Truth Background
Predicted Source	125	2
Predicted Background	3	21588

Arch. 4 - Training Set	Ground Truth Source	Ground Truth Background
Predicted Source	298	1
Predicted Background	1	50375

Arch. 4 - Test Set	Ground Truth Source	Ground Truth Background
Predicted Source	127	2
Predicted Background	1	21588

Arch. 5 - Training Set	Ground Truth Source	Ground Truth Background
Predicted Source	298	1
Predicted Background	1	50375

Arch. 5 - Test Set	Ground Truth Source	Ground Truth Background
Predicted Source	127	4
Predicted Background	1	21586

TABLE 4.3: Confusion matrices for different CNN architectures

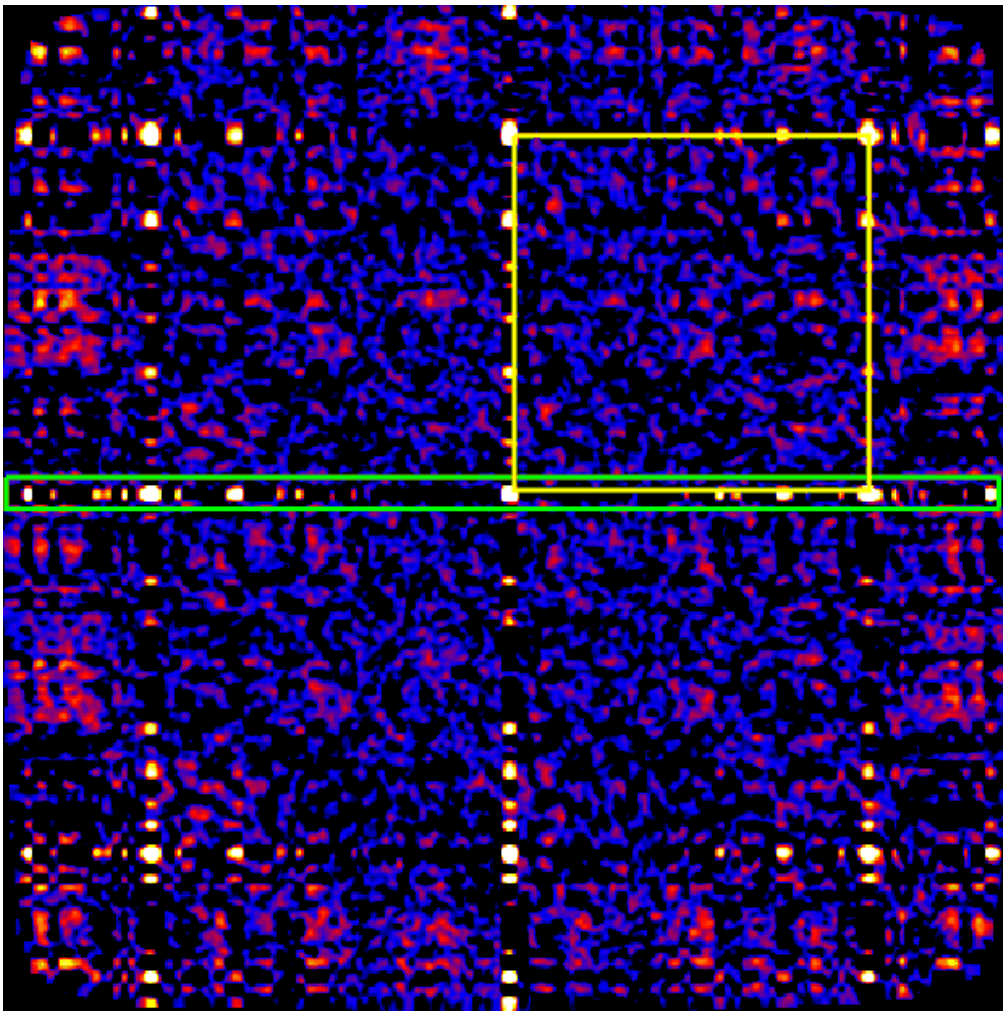

FIGURE 4.4: A sky image of a 2.7 ks observation of the bright source Crab Nebula (spanning $\sim$25°). The source is in the center of the fov. The most significant ghosts are located in a grid of a $\sim$10.7° around the source (Krivonos et al., 2010).

Ghosts have the same PSF and surrounding features as real sources. A ghost can only be recognised as such by studying the context of the whole image and looking for the expected pattern of ghosts (which are at least 10.7 degrees away). In INTEGRAL's case, extensive use of the prior catalogue is used to choose between ambiguous ghost positions (Krivonos et al., 2010).

4.3 Final ML Model for Source Detector

In the previous section, the merits of different ML approaches to source detection were investigated. Traditional ML techniques were outperformed by a deep learning model utilising 3 input channels: exposure map plus flux and significance from a single energy band. The final source detector architecture evolved from this initial proof-of-concept model to include 11 channels: flux and significance from five energy

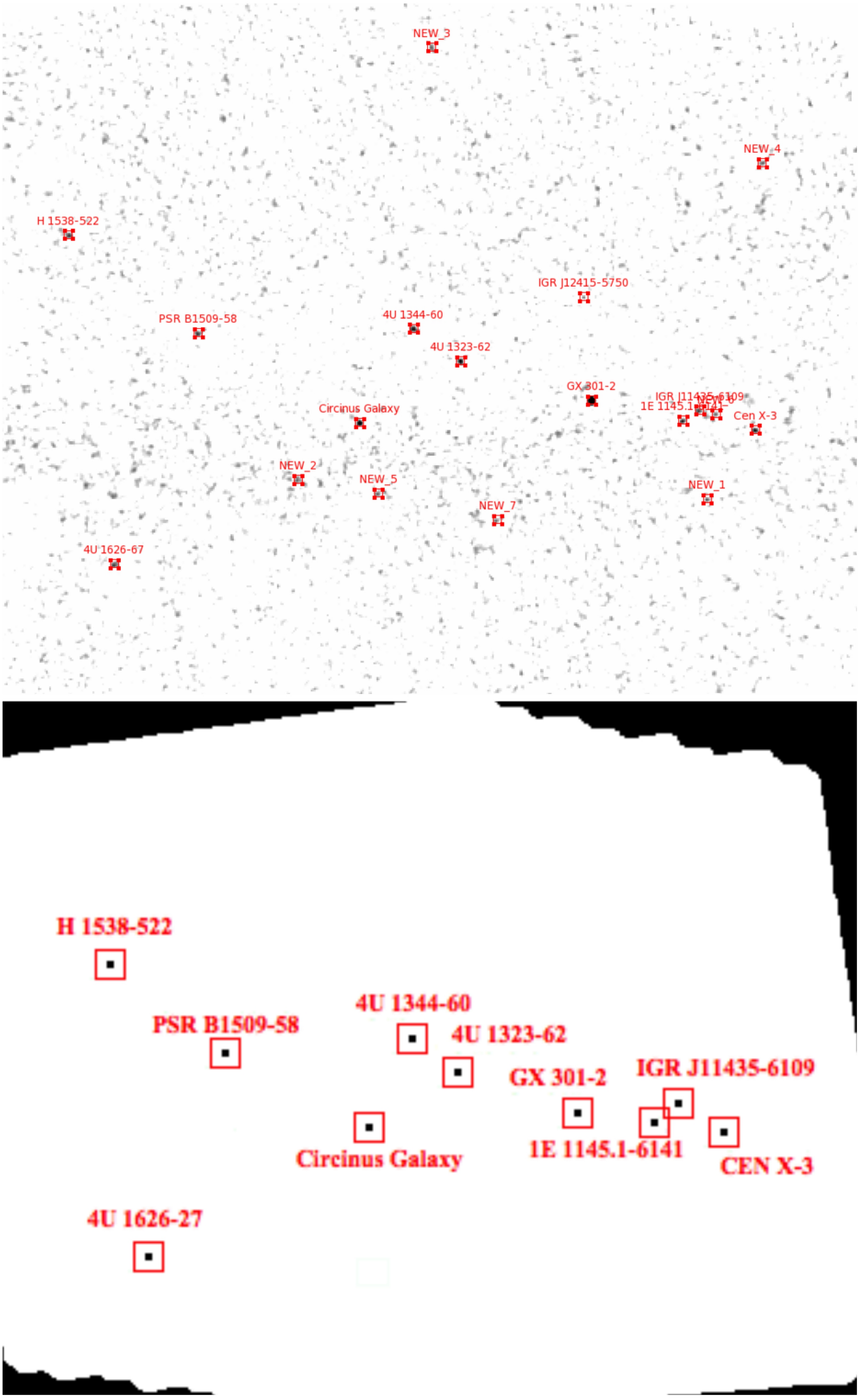

FIGURE 4.5: The top image shows an ISGRI mosaic significance map (spanning approximately 10 degrees) for revolution 1839 with an exposure time of 57.6 ks created using the standard OSA software (Goldwurm et al., 2003). The bottom image was produced by using the CNN. Each 11×11 pixel window was passed through the classifier. Sources detected are in red. OSA detects some noise as sources, which the CNN does not. The CNN does not detect IGR J12415-5750 although this was only a marginal detection in OSA.

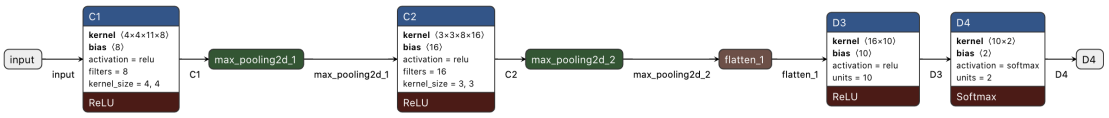

FIGURE 4.6: The architecture of the final source detector CNN. The CNN uses two convolutional layers with kernels of various sizes, followed by a multi-layer perceptron and softmax function to determine final classification of the image as containing a source or not. A visual representation of the CNN applied to an example image is presented in Figure 4.7.

bands (17-30,30-60,18-60,20-40 and 20-100 keV), as well as the exposure map. This model was the basis of final source detector, whose development and deployment on INTEGRAL data is presented in this section.

4.3.1 Training the CNN

The CNN was trained using ∼25,000 examples of both sources and background extracted in 11 × 11 pixel windows from 11 channels: intensity and significance across five energy bands (17-30, 30-60, 18-60, 20-40 and 20-100 keV), plus the exposure map. All the images from the five energy bands are stacked and sent through the network at once. This allows the CNN to be able to see a whole spectrum of information about the source at once. The exposure map was also included as an 11th channel. This gives the CNN information about how near to the centre of the map a window is as in an ISGRI ScW image, as the centre of the image has the highest exposure and the edges of the map the least. The training and test data for this final CNN architecture was intentionally chosen to have an even balance of sources and background to avoid the complications of imbalanced training sets encountered in the early models that were tested (Sections 4.2.1 and 4.2.2.2).

Once the network was fully trained, a 11 × 11 pixel window could be moved across an entire ISGRI ScW image which could contain many sources. Figure 4.7 shows how the network takes the input window image, passes it through a series of convolution, nonlinear activation (ReLu), pooling (downsampling), and fully-connected layers and then returns an output for each window, with the assumption that each window contains only one source. In all but the galactic centre, sources are sparse enough that there exists a very low chance of detecting more than one in a 11 × 11 pixel window.

To generate the training set ∼200 ScWs (∼a fifth of the ScWs in cat1000) were used and in each ScW the results file generated from the IBIS pipeline (Goldwurm et al., 2003) was used to select any sources that were present with a significance of over 5 sigma. Any machine learning model trained on a human-labelled dataset could potentially learn the biases of the labeller. The IBIS pipeline uses The General Reference Catalog as the master table listing all known high-energy sources of relevance to INTEGRAL. These objects are all those that have been detected by INTEGRAL or that are known to

Training Set	Ground Truth Source	Ground Truth Background
Predicted Source	17289	0
Predicted Background	2	17291

Test Set	Ground Truth Source	Ground Truth Background
Predicted Source	7408	0
Predicted Background	2	7410

TABLE 4.4: Confusion matrices for datasets used to train and test the final CNN. Correctly classified sources dominate over misclassifications.

be brighter than 1 mCrab in the 1 keV to 10 MeV band. The objects in the catalogue were compiled from several sources to ensure confidence in the validity of these training examples and that any individuals' bias would have limited impact.

For every source a random window of background was also selected to include in the training set. The resulting training set contained $\sim$ 25,000 examples of source and background. 70 percent of this was used to train the CNN while 30 percent was left out of the training processes and used as a test set to measure performance.

4.3.2 CNN architecture

Several CNN networks were trained with different architectures with the aim of designing the simplest architecture as possible without effecting the network's performance. It was found that by increasing the number of filters and layers from the architecture chosen and shown in Figure 4.6 there was no improvement in performance but when the number of filters was reduced there was a noticeable change in performance. For example a network with six filters in the first convolutional layers and 14 in the second convolutional layer found 6 false negatives and 4 false positives when applied across the test set, this is in contrast to just 2 false negative and no false positives from the final network. Figure 4.7 breaks down the final network layer by layer and illustrates how an 11x11 pixel window that includes a source would be passed through the network.

4.3.3 CNN Testing and Performance

Table 4.4 shows the confusion matrices for the CNN for both the training and the test set. From all these 11x11 pixel windows none of the background sources were classified incorrectly and only 4 sources were not detected.

Another five CNNs were also trained with the same architecture, with similar performances, each using a different fifth of the ScWs used in cat1000. This was to ensure a stable network had been produced that would have a similar high level of

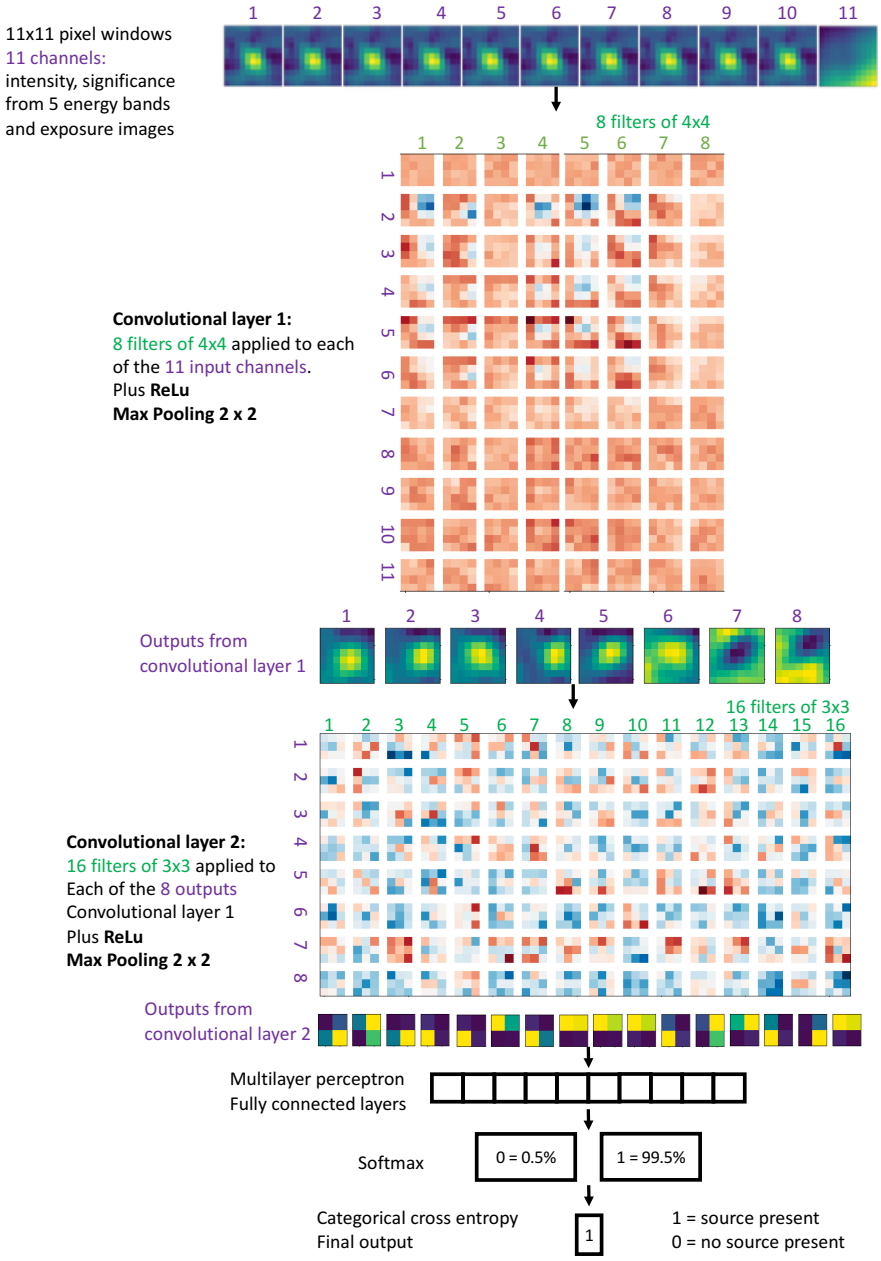

FIGURE 4.7: The architecture of the source detector CNN, which shows how a 11×11 pixel window with a source present is processed through the CNN resulting in its detection. The first convolutional layer applies eight 4x4 filters to the pixels windows from each of the 11 energy bands producing the feature map shown. A max pooling layer and ReLu function applied to these maps produce the eight outputs from the first convolutional layer shown as the centre row of images. The next convolutional layer applies a similar sequence of processing, as do any subsequent layers in a CNN. The outputs from the convolutional layers are passed through a multilayer perceptron and a softmax function to decide the final classification of the image as containing a source or not.

performance regardless of which ScW were used in the training stage and that the model was not overfitting.

The CNN takes $\sim$ 6 hours to train using Keras with a TensorFlow backend on a NVIDIA TITAN Xp with 12 gigabits of RAM and took $\sim$1 day to apply to the whole cat1000 dataset. This is significantly faster than the process in cat1000: running *peakfind* and SExtractor on the mosaic maps took days, then manual source inspection took weeks or months. Similarity of the final source catalogues from the CNN method and cat1000 is assessed in the following chapter.

4.4 Summary and Next Steps

This chapter presented the development of a source detection tool for INTEGRAL data using deep learning with CNNs. The rationale for choosing this model was presented, along with a comparison of the performance of deep learning models to that of traditional ML approaches. It was found that the CNN outperformed the older methods, so the CNN was expanded to include inputs from multiple INTEGRAL energy bands. Additionally, the model architecture was tuned to achieve optimal performance with the minimal model complexity.

With this rapid and reliable source detection tool in hand, it is possible to search through substantial volumes of INTEGRAL data. Specifically, this tool now facilitates the ability to search at the shortest possible observational timescales, the ScW level. As will be shown in the next chapter, this has the advantage of being able to find sources that are bright in very few ScWs and might be washed out in co-added image stacks. Additionally, this allows variable sources in crowded fields to be detected separately in any images where not all sources are bright.

The ability to search images at the shortest timescales does present the additional challenge of identifying which excesses arise from the same astrophysical source. This is a nontrivial problem that increases in complexity as data volumes increase. To address this problem, a Bayesian excess merging process was employed, and this is the focus of the following chapter.

Chapter 5

A Bayesian approach to merging excesses in hard X-ray surveys

Whilst the source detector presented in the previous chapter is excellent at finding sources in individual images, substantial work must follow to reduce a list of detections across numerous images into a single unique source list. Excess merging is a common requirement in astronomy, particularly in the areas of catalogue matching and time domain studies.

Multi-wavelength studies offer a powerful tool for constraining physical processes in astronomical sources. This has been substantially more accessible due to the availability of a broad array of publicly released source catalogues from major surveys covering large portions of the sky (e.g. SDSS; York et al., 2000) or deep views into smaller regions of sky (e.g. Hubble Deep / Ultra-Deep Fields). These rich survey catalogues motivate the need to match catalogues in a systematically reliable fashion, giving rise to sophisticated approaches for merging catalogues using probabilistic reasoning (e.g., Budavári and Szalay, 2008; Nguyen et al., 2022).

Source matching is particularly crucial for time domain studies where telescopes return to the same part of the sky on numerous visits. This has been a common strategy for transient searches and variability studies, including the Galactic Plane Scan (GPS) and Galactic bulge monitoring program using INTEGRAL. Additionally, many large surveys now adopt the strategy of repeated visits to the same part of the sky as a method for achieving greater sensitivity to faint sources (e.g. the Rubin Observatory, formerly LSST; LSST Science Collaboration et al., 2009). Matching of sources in individual images in such surveys enables rich time domain analyses, but again requires a robust approaches for source association.

These astrophysical research objectives motivate the need to develop a reliable tool for associating disparate detections arising from the same astrophysical source. Such a

technique should ideally be symmetric, i.e. not sensitive to the order in which catalogues are introduced to the merging process. Additionally, the approach should use probabilistic reasoning rather than simple cuts in order to disentangle potentially ambiguous matches. Toward that end, this chapter presents the development and testing of a new excess merging tool founded in Bayesian reasoning which allows the final compilation of a unique source catalogue from the excesses identified with the CNN source detector.

5.1 Previous approaches to excess merging: *megamerge*

In cat1000, excesses were merged using a custom built algorithm called *megamerge* that iteratively merged the excess lists from each map into a base list which took the cleaning catalog (only the sources in the ISGRI reference catalog that had been previously detected by IBIS) as a starting point. It should be noted that SExtractor and *peakfind* were applied to each of the five energy bands individually, meaning *megamerge* was needed to merge detections across the five bands. This illustrates an additional advantage of the CNN source detector, as it identifies a single excess using all five energy bands simultaneously.

The first drawback from using *megamerge* was that it used the reference catalog as a starting point for the merging, so it added a bias into the process. This would be compounded when the source was originally discovered by INTEGRAL. A further problem was that the result was dependent upon the order in which the excesses to merge were presented to the algorithm. New excesses were presented to the merging algorithm and a decision on whether to merge with an existing excess in the database was made purely on position - on whether the two excesses fell within a given merge radius. The 'catalog' position of the excess would be updated if the newly merged excess had higher significance than those previously used - the point source location accuracy of a coded mask imager depends strongly on the detection significance. The inherent risk though was that a strong new detection at the limit of the merger radius could not only "steal" an existing detection but also cause its coordinates to change. Clearly the order in which the excesses were presented could have a significant influence on the outcome of the process. No attempt was made to optimise or mitigate this during cat1000 production, so as a result significant effort was required to check results manually.

The *megamerge* algorithm made no use of pre-existing information (beyond the position) to decide if two sources separated by the merger limit were likely matches. As a simple example, two 50-sigma excesses separated by 8' were very unlikely to be the same source, but would be erroneously merged by *megamerge*. On the other hand, two 4.5-sigma excesses separated by 8' could be the same source, since the point

source location accuracy at 50σ and 5σ were $\sim$0.5′ and $\sim$5′, respectively (Scaringi et al., 2010).

As a result of these limitations, and concerns over how robust the outcome was, it was decided there was an urgent need for a better approach to merging the source lists.

5.2 Bayesian excess merging approach

This section introduces the final method for merging excesses in INTEGRAL catalogues founded on Bayesian reasoning and built on the methods developed by Budavári and Szalay (2008) and Rosen et al. (2016). Section 5.2.1 presents a very brief review of Bayes' Theorem to motivate its application to astronomical object pair matching presented in Section 5.2.2. Finally, the way this Bayesian match probability is applied to the full INTEGRAL excess list is described in Section 5.2.3.

5.2.1 Introduction to Bayesian reasoning

Bayesian reasoning is a method to calculate the likelihood of an event by incorporating prior knowledge about other conditions or features that impact the probability of that event. This approach is founded on application of Bayes' Theorem:

$$P\left(A|B\right) = \frac{P\left(B|A\right)P(A)}{P(B)} \tag{5.1}$$

where the key elements of the equation are:

- $P\left(A|B\right)$ – the probability of event A given that we have observed event B, this is known as the **posterior** and is the targeted probability (or probability distribution) that is to calculated
- $P\left(B|A\right)$ – the probability of event B given that we have observed event A, this is known as the **likelihood** and captures knowledge of the dependent relationship of B on A
- $P(A)$ – the global probability that A has occurred amongst all possible scenarios, this is the **prior** that captures the global belief that A will occur
- $P(B)$ – similar to the prior this is the global probability that B will occur amongst all possible scenarios, and is sometimes referred to as **evidence**

Bayes' Theorem can be demonstrated with a simple easily understood example. Suppose two coins are flipped and it is known that one of the coins landed as heads,

what is the probability that *both* coins landed as heads? Let the Bayesian terms for this problem be explicitly defined as:

- A is the event of both coins landing as heads
- B is the event of at least one coin landing as heads (which is known to be true)

Without prior knowledge of the outcomes of the coin flips, there are four possible results: heads-heads, heads-tails, tails-heads, tails-tails. One can then calculate the three components of the right hand side of Bayes' Theorem:

- $P(A)$ is the probability that two coins have been flipped and both landed as heads – this is $1/4$
- $P(B)$ is the probability that two coins have been flipped and *at least one* has landed as heads – this is $3/4$
- $P\left(B|A\right)$ is the probablity that at least one coin lands heads (B) given that it was observed that both coins landed as heads (A) – this is 1

Placing these values into Bayes' Theorem shows:

$$P\left(A|B\right) = \frac{1 \times 1/4}{3/4} = \frac{1}{3} \tag{5.2}$$

This makes intuitive sense: there are four possible scenarios for flipping two coins and if it is known that at least one coin has landed heads then the one case with no heads can be eliminated, leaving three equally possible scenarios for which only one meets the desired outcome of both coins being heads.

This simple example illustrates the power of Bayesian reasoning: by incorporating all prior knowledge about events that affect the event of interest, a more accurate probablity of the event of interest happening can be calculated. For this work, the goal is to determine the probability that two excesses found in INTEGRAL data arise from the same astrophysical source.

5.2.2 Matching astronomical sources using Bayesian reasoning

The primary objective of this work is to determine a set of unique sources from a set of excesses detected by the CNN source detector. The foundation of the technique here begins by establishing the Bayesian probability that two excesses arise from the same astrophysical sources, and this section describes in detail the mathematics behind that calculation.

Following Budavári and Szalay (2008) let there be two excesses with measured positions on the sky x_1 and x_2 with positional uncertainties σ_1 and σ_2 (it is assumed that this uncertainty has no directionality). Let the following events (in the Bayesian sense) be defined as:

- H – the hypothesis that the two observed excesses arise from the same source
- K – the complementary hypothesis that the two observed excesses arise from different sources
- D – the event that the two excesses have been observed

Bayesian reasoning will be used then to determine the target probability $P\left(H|D\right)$ which is the (posterior) probability that both excesses arise from the same source (i.e., the hypothesis H) given the observed locations and uncertainties of the excesses (i.e., the data D).

Because H and K are complementary hypotheses (the excesses must either be from the same source or from different ones), the total probability from these two must sum to one:

$$P(H) + P(K) = 1 \tag{5.3}$$

as must the posterior probabilities:

$$P\left(H|D\right) + P\left(K|D\right) = 1 \tag{5.4}$$

This equation can be rearranged then as:

$$P\left(H|D\right)\left[1 + \frac{P\left(K|D\right)}{P\left(H|D\right)}\right] = 1 \tag{5.5}$$

meaning it will be useful to calculate the relative probability of the two hypotheses given the observed data. This can be further re-arranged by invoking Bayes' Theorem for the two posteriors:

$$P\left(H|D\right) = \frac{P\left(D|H\right)P(H)}{P(D)} \tag{5.6}$$

$$P\left(K|D\right) = \frac{P\left(D|K\right)P(K)}{P(D)} \tag{5.7}$$

$$\Rightarrow \frac{P\left(H|D\right)}{P\left(K|D\right)} = \frac{P\left(D|H\right)P(H)}{P\left(D|K\right)P(K)} \tag{5.8}$$

Finally, combining Equations 5.8, 5.5 and 5.3 results in a new expression for the posterior probability of the two excesses arising from the same source:

$$P\left(H|D\right) = \left[1 + \frac{1 - P(H)}{B \cdot P(H)}\right]^{-1} \tag{5.9}$$

which now depends only on the "Bayes factor" $B \equiv P(D|H)/P(D|K)$ introduced as the ratio of posterior probabilities of the data given the two hypotheses, and $P(H)$ which is the probability of two observations arising from the same source with no data to constrain the hypothesis. Both of these terms are examined in more detail as follows.

First $P(H)$, the probability of two excesses arising from the same source when no data is available to constrain those excesses, can be estimated as the pure probability of any two excesses from two observations arising from the same source. Suppose two observations detect N_1 and N_2 excesses, respectively, and it is known that N_* sources appear in both sets. Then the probability of any one excess chosen from N_* giving rise to the one source out of N_1 from the first observation and the one source out of N_2 from the second observation is:

$$p_0 \equiv P(H) = \frac{N_*}{N_1 N_2} \tag{5.10}$$

where this probability is now labelled p_0 following the convention introduced in Rosen et al. (2016) and adopted by Lepingwell et al. (2022).

Next it is necessary to derive the Bayes factor $B \equiv P(D|H)/P(D|K)$, the ratio of posterior probability of the data (the two observed excess locations and uncertainties) arising from the same source to the probability of the data coming from different sources. Whilst a more rigorous derivation of this factor can be found in Budavári and Szalay (2008), this can be understood as the estimation found by treating the locations and errors as Gaussian probability distributions. Under this assumption the Bayes factor becomes:

$$B = \frac{2}{\sigma_1^2 + \sigma_2^2} \exp\left[-\frac{\psi^2}{2(\sigma_1^2 + \sigma_2^2)}\right] \tag{5.11}$$

where ψ is the angular separation between the excess locations.

5.2.3 Final method for merging INTEGRAL excesses

The final approach for merging INTEGRAL excesses was a new tool built on the method used by XMM (Rosen et al., 2016) and further developed to work on ISGRI data. This method first merges individual excesses into unique sources using the Bayesian probability approach described above in Section 5.2.2 before matching the merged point to a source in the reference catalogue. This removes the bias noted with the method employed in cat1000 (*megamerge*) as the reference catalogue is no longer being used as a starting point and the order the excesses are presented to the algorithm are now irrelevant as it uses Bayesian probability to decide which order to merge in.

First the algorithm searches for any pairs of ScW detections that are less than 8 arcminutes away. Any detections found within the same ScW are excluded from this

matching process as they will be different sources. Each pair has a Bayesian match probability (p_{match}) assigned to it using Equation 5.13 (derived in Section 5.2.2) where σ_1 and σ_2 are the position error radii of each detection in the pair (radians), ψ is the angular separation between the pair, $p_o = N^*/N_1N_2$ where N_1 and N_2 are the number of objects in the sky based on the surface densities in the two fields. Each N value is derived from the number of detections in the two observations and then scaled to the whole sky and N^* is the number of objects common between them.

$$p_{match} = \left[1 + \frac{1 - p_0}{B \cdot p_0}\right]^{-1} \tag{5.12}$$

$$B = \frac{2}{\sigma_1^2 + \sigma_2^2} \exp - \left[\frac{\psi^2}{2(\sigma_1^2 + \sigma_2^2)}\right] \tag{5.13}$$

This allows the algorithm to make a first cut and remove any pairs with $p_{match} < 0.5$ then determine the order to match the detections. This cut on p_{match} is important for rejecting poor matches, as the number of good matches goes into calculation of a goodness of cluster that is used to set the order for final excess merging.

For each excess, a goodness of cluster (GoC) is calculated as the ratio of the number of good matches it has with other excesses to the area of its positional error. This GoC is useful because it prioritises the detections that have the smallest error radius and most amount of pair matches. As noted by Rosen et al. (2016), this means the merging order begins with excesses that lie at the centre of a group of detections, which are most likely to be associated with a genuine unique source. The algorithm sorts the excesses in order of ascending GoC and iterates down the list. For each excess the algorithm sorts that excess's matching pairs in ascending order. The algorithm then progresses down this GoC-sorted list to merge sources by assigning a common source ID to sources matched in a pair. The process concludes when each cluster has a unique source ID. A visual representation of this process can be seen in the flowchart shown in Figure 5.1.

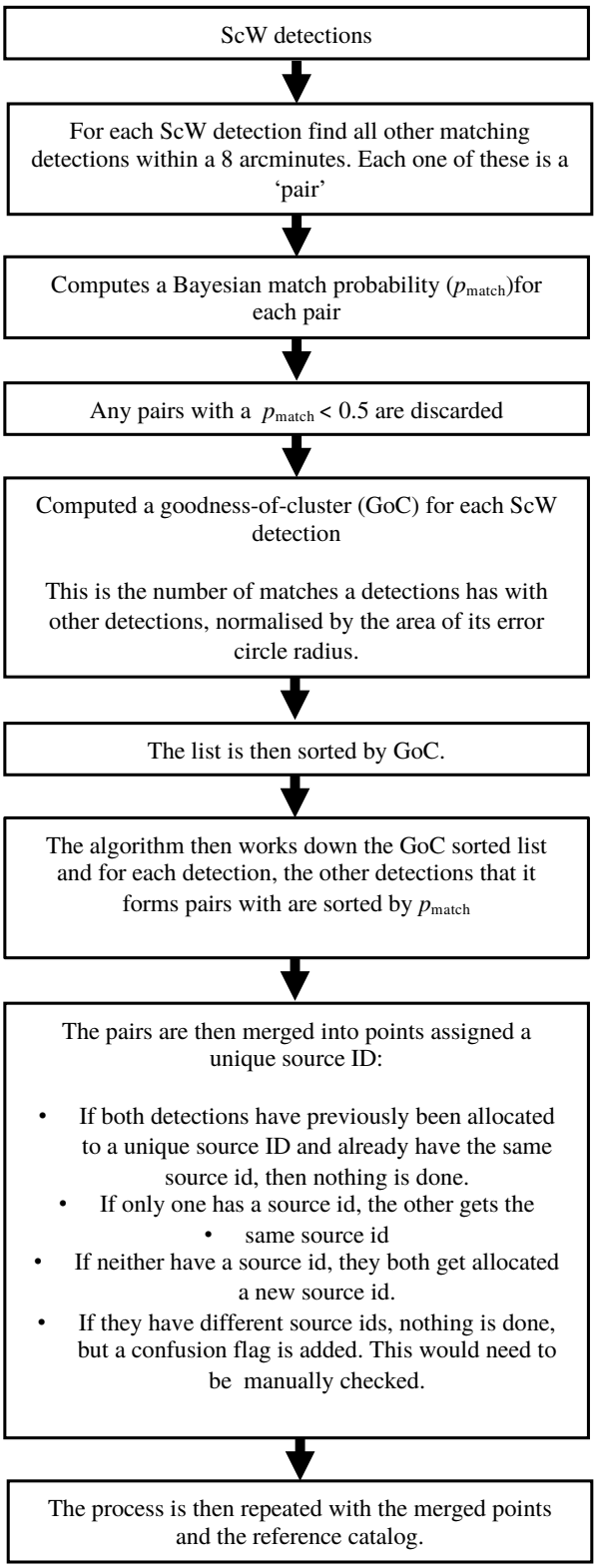

FIGURE 5.1: A flowchart showing how the BM algorithm works.

Detection method	Excesses found over 5 sigma	Merging method	cat1000 sources	non cat1000 sources
CNN	∼100,000	Megamerge	434	14
		Bayesian Matching	448	25
SExtractor	∼200,000	Megamerge	93	1
		Bayesian Matching	96	2
Peakfind	∼500,000	Megamerge	175	0
		Bayesian Matching	179	5

TABLE 5.1: Sources detected by each method. CNN detections utilise all five energy bands simultaneously and produce a single detection for each source, whilst SExtractor and *peakfind* search each energy band individually meaning a single source could be detected separately in all five energy bands (which ultimately requires significantly more time to perform the merging process).

5.3 Performance

To measure the performance of the new Bayesian matching tool, it was applied to the cat1000 dataset and compared to results when the traditional tools were applied. This section discusses the outputs from each method and how they compare. A list of detections was also provided to domain experts for manual inspection. A detailed inspection was also applied to the galactic centre as the most difficult region to catalogue in cat1000.

It is important to note that comparisons between this work and cat1000, while instructive, are not meant to be direct comparisons of exactly parallel approaches. Indeed, the ScW-based search used here is ideal for finding significant transient events on small timescales and strong persistent sources, while weaker persistent sources are better suited for discovery in stacked images such as those analysed for cat1000.

5.3.1 Method Comparison

Each of the three source detection methods was applied across the entire dataset to generate an excess list. While *peakfind* and SExtractor were applied to each energy band separately, the CNN source detector utilised the five bands simultaneously. Thus the CNN finds about 1/5 as many excesses as *peakfind*, which makes similar judgements about the PSF as the CNN (SExtractor does not perform as well in this regard). These three excess lists were then passed through both merging methods to generate a list of merged sources. Table 5.1 breaks down each method and shows how many cat1000 sources were found and also how many sources were found that were not included in cat1000. Not only do the new CNN and Bayesian matching methods recover more sources, but figure 5.2 shows how these methods have a lower flux threshold compared to the traditional tools. Overall the CNN can detect sources at a lower flux than the other two methods.

All of the methods presented here find fewer sources than the full cat1000 due to the fact they are being applied on single ScW images, whereas cat1000 was applied to stacked images. These stacked images were ideal for finding faint persistent sources, which is not possible on a ScW level. *ScW-level analysis finds bright persistent sources or fainter transient sources, and the CNN performed best at finding such sources missed in cat1000.*

It is noteworthy that 25 sources that were not included in cat1000 were found using the new CNN and Bayesian matching methods. These sources were presented to domain experts that were part of the cat1000 survey team for them to manually inspect the sources and provide an evaluation of the performance of the CNN. Out of the 25 sources the domain experts agreed that all but two were genuine sources. These two

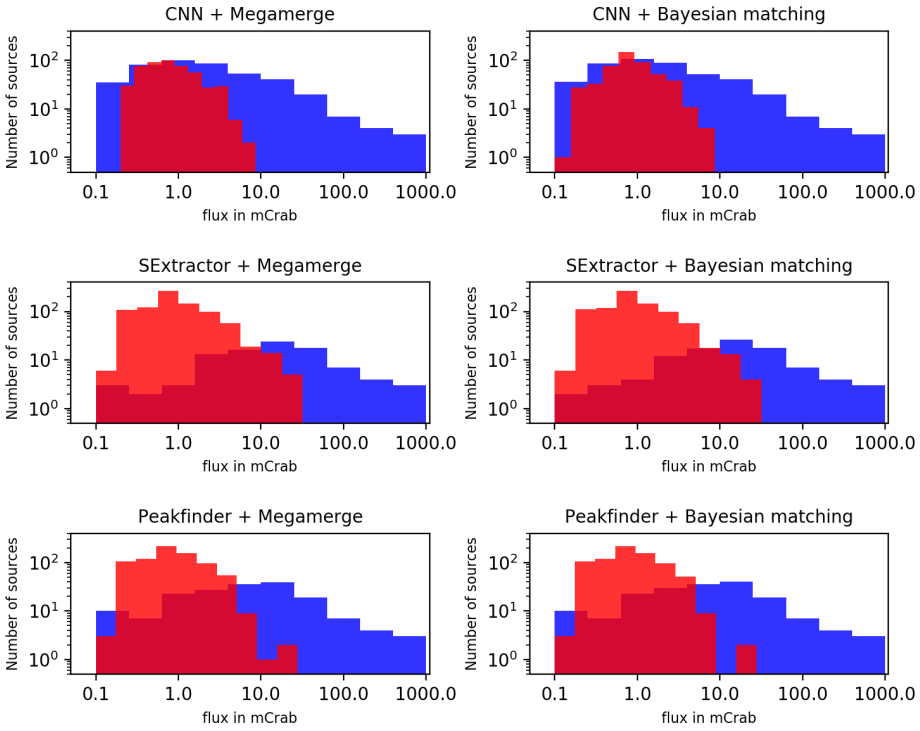

FIGURE 5.2: Histograms showing the proportion of cat1000 sources found (blue) vs not found (red) and their fluxes in cat1000 for every combination of the three source detection techniques and the two merging techniques. While other source detection approaches produce a detection efficiency that decreases (generally) monotonically as source brightness decreases, our method is able to detect fainter sources by leveraging the subtler source signatures captured by the CNN. These sources can have exceptionally low average fluxes (0.1 mCrab) in the stacked images used for cat1000, but will be bright enough in some ScW images to be detected using these methods.

sources were both found in the noisy borders of the ISGRI images which exist due to the coded-mask of the instrument. A member of the survey team would reject this on manual inspection but due to the localised nature of the CNN it would have been hard for the CNN to have picked these out as noise. To avoid this problem in the future, the CNN could either be just applied to a small section of the ScW image that avoided this area, or these detections could be flagged for manual inspection. While the addition of the exposure map as the 11th channel significantly reduces this effect, it does not entirely remove all artefacts from the extreme edges of the image. The 23 genuine new sources will be presented in a forthcoming analysis by the INTEGRAL team.

5.3.2 Blended Sources and Galactic Centre

22 of the sources in cat1000 were labelled as blended, in other words their positions were considered unreliable due to nearby sources within the angular resolution of the

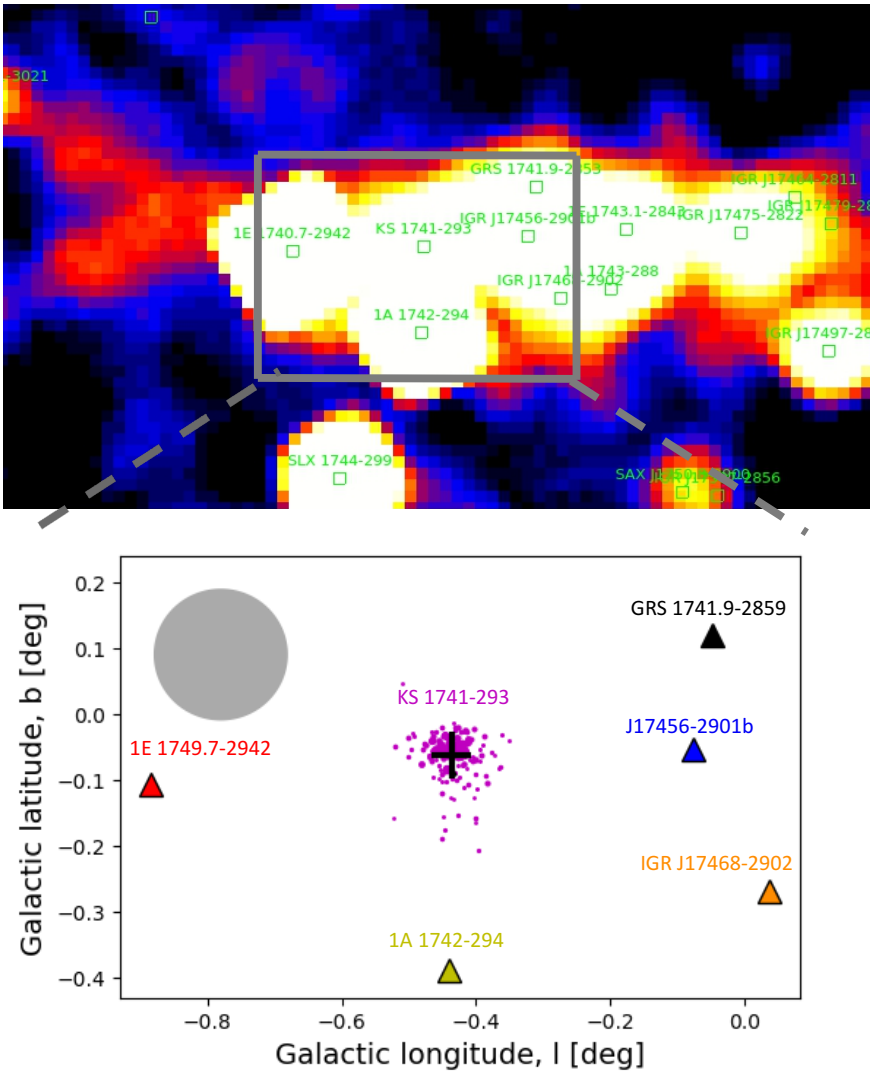

FIGURE 5.3: Stacked INTEGRAL/IBIS image of the Galactic Centre (top), and coordinate plot (bottom) of science window detections of a specific source (magenta) and final merged source locations for its neighbours (coloured triangles), with the ISGRI resolution shown in grey.

telescope. In some cases there was ambiguity in the identification of what appeared as one source, whereas other pairs of sources showed clear extension beyond a single point source. All cases were subject to lengthy visual inspection to determine their best representation in the final catalogue. In addition to simple blended pairs, 13 sources in the crowded Galactic Centre region were also listed as blended - in this case the true determination of the emitters was impossible due to the crowded and highly variable nature of the region. Some of these regions may prove useful test cases to understand if the Bayesian matching method is better able to untangle these complex regions, but unfortunately all of the blended sources in cat1000 are faint, with

persistent fluxes below 10 mCrab. Nevertheless, some pairs are of sources that are variable in nature, so de-blending may be possible as precise positions with small uncertainties can be obtained from science windows during which the sources are bright. As the CNN method used here searches for sources at the ScW-level, it frequently recover sources that may appear blended in a stacked image (see Figure 5.3). If the point source location is sufficiently robust, these sources should not be merged during subsequent analysis.

5.4 Conclusion and prospects for generating future catalogues

The CNN-based approach to source detection in IBIS/ISGRI has yielded several key advantages. Firstly, this method utilises all five energy bands simultaneously – not only does this improve the accuracy and allow the network to detect sources at a lower flux threshold, it also speeds up the merging process as it results in a single excess for each astrophysical source instead of one from each energy band. Secondly, the speed on which the CNN can be applied to the entire dataset makes it possible now to look on a ScW timescale and detect sources that only appear in a single ScW but fall below the detection threshold in all-sky maps stacked from images spanning a revolution time-scale. Finally, the CNN-driven approach removes human biases from the source detection process, making the list of sources here more impartial than previous approaches.

Bayesian matching recovers marginally more sources than *megamerge* - but has the "right" set of answers (i.e., agnostic to where the excesses come from) and has removed a bias from the process by not using the reference catalogue as a starting point. Another limitation removed is that the order in which excesses are presented to the algorithm no longer impacts the end result.

Looking on a ScW level allows the detection of sources that have outbursts on smaller timescales than previous studies of the IBIS/ISGRI dataset. This approach also helps to resolve the emission from the GC - detection at ScW level is easier to do than with stacked images as sources are not all "on" at the same time.

The combination of the CNN and Bayesian matching produces a very accurate merged list of detections with very few detections needing to be manually checked - compared to the old method which took 2.5 years for 9 people to manually check each source for inclusion into the catalogue.

If these tools are to be applied to generate future catalogues this work would need to be extended to include revolution maps and all-sky maps, enabling the source detector to find the weak persistent sources that are not detected at a high enough significance in a single ScW. When the ScW images are stacked currently using a

mosaic tool, the point spread functions (PSF) of the sources become distorted and the CNN's performance may drop when applied to these images. One way to overcome this would be to use HEALPix, which produces a subdivision of a spherical surface in which each pixel covers the same surface area as every other pixel. This would maintain the PSF of the sources and allow the CNN to be used in order to detect sources in stacked images, although care would need to be taken to maintain the same resolution.

One possible setback of this method is that because the CNN looks on a local 11x11 pixel window it does not know how noisy the entire map is. In most cases this should not be a shortcoming, as the local noise level will be more important for informing the CNN of the likelihood of a source being present. However there are some isolated cases where the global noise map is useful. When domain experts make a decision about a source the "flatness" of the map is taken into account. In addition to this, two false positives were still detected in the extremities of the partially coded FOV, although such excesses would normally be suppressed by the low exposure in that area. This problem could be overcome by not applying the CNN in the border region of the ScW maps, or a flag applied to any detections in this area for a manual inspect.

The newly developed source detection and merging method is reliable, scalable, removes need for continuous human intervention and eliminates some of the human subjectively that previously existed. This will be ideal for application to help generate future ISGRI catalogues.

Chapter 6

Detecting outbursts in hard X-ray light curves using deep learning

Many X-ray sources are transient in nature and the ability to automatically detect when a source is in outburst would be a useful tool for providing a more comprehensive view of outburst behaviour in these sources. Similarly, a tool which can find a single period of intense emission within an otherwise extended period of quiescence would enable detection of otherwise elusive transient events. All the sources in the X-ray sky show a degree of variability in their flux, although the magnetically-powered pulsars (such as the Crab pulsar) come close to being constant emitters - the Crab is used as a calibration standard and the flux unit of '1 Crab' is used extensively [1]. Astronomers classify sources as 'persistent' - meaning they are visible most of the time when observed, or 'transient' - meaning they are usually below the threshold of detectability, but then become visible during times of increased emission. However, these simplistic definitions are based on detectability rather than the intrinsic source behaviour - a source which shows flares of emission above persistent lower-level quiescent emission would be 'persistent' if nearby and hence detectable in quiescence, while it would be 'transient' if moved to a greater distance such that only the peak emission were detectable. Furthermore, sources classified as transient in one waveband may be persistent in others or demonstrate different emission mechanisms, or could be found as persistent sources with the availability of more sensitive instrumentation. A clear example of this is the group of 'burst-only' sources seen by BeppoSAX (In't Zand et al., 1999; Cornelisse et al., 2004), which are seen as persistent emitters (with flares of stronger emission) by INTEGRAL.

Different x-ray sources undergo outburst on different time scales. In X-ray binaries with a neutron star as the central compact object, gas from a nearby companion star

[1] However recent studies have shown significant long-term changes in the Crab's flux (see, e.g., Jourdain and Roques, 2020).

can accrete onto the neutron star until a critical condition triggers a thermonuclear runaway. Such thermonuclear bursts are observed as brief, bright flashes of X-ray emission that last for seconds to hours and repeat on a timescale of hours to months. Be/X-ray binaries (BeX) are a class of HMXB binaries that consist of a Be star and a neutron star. The variability time scales in BeX range from seconds to years.

The hard X-ray sky is intrinsically variable on a large range of timescales from 1s (short Gamma Ray Bursts, type 1 bursts from Neutron Stars) to 100d (X-ray novae, Type 2 bursts from Be-XRBs). Localising outbursts in lightcurves makes it possible to automatically highlight the most interesting and transient behaviour of sources. Not only would this give insight into the nature of the sources, their outbursts and aid in the classification of the object, but it could yield information on such parameters as the sources' periodicity. As well as applying a burst detector tool to lightcurves of known sources, this method could be extended and applied to every point on the sky allowing the detection and exploration of new transient objects. This would be useful in detecting sources that are missed when searching stacked maps or mosaics for sources. For example, a source may not be detected in stack maps if it had been in outburst in the early mission but not later in the mission: as more data accumulates when the source was not in outburst, its detection in stacked maps becomes of lower and lower significance. Having a robust method to search lightcurves for bursts would allow the recoverery of these sources.

At present a source's light curve is investigated by manually inspecting it for unusual activity or by calculating its "bursticity". Domain experts in this field find it difficult to define exactly what an outburst is. Attempts have been made to describe how 'bursty' a source is (e.g. tags included in cat1000; Bird et al., 2016); by measuring a source's bursticity the approach was designed to determine if the source becomes detectable on shorter timescales than the whole dataset. If the source is normally "off" then the mean flux is reduced by adding more null observations, and a source that is bright for a short time can become undetectable. Sources that have multiple outbursts on top of detectable quiescent emission will have a bursticity of around 1. Although bursticity is a useful tool it does not tell us how many bursts have been detected in a source or the duration of these bursts. Brute force methods can be applied to answer these questions, but they can take many hours and can also produce inaccurate results. As each hard X-ray source behaves differently, there is no "normal" behaviour for a source. Due to this fact, what is actually of interest in each source's lightcurve is when the source is behaving unusually or anomalously.

This Chapter presents a new approach to detecting outbursts in hard X-ray light curves using deep learning. Specifically, a burst detector that employs a Long Short-Term Memory (LSTM) network is developed that operates substantially faster than previous approaches. The burst detector is trained and tested on simulated light curves, and applied to select test cases of INTEGRAL+BAT light curves.

6.1 Current process for burst detection: "bursticity".

The previous method for detecting bursts in INTEGRAL light curves involved calculating a bespoke metric entitled "bursticity". Light curves for each source were generated in the 18-60 keV band on ScW time sampling. A variable-sized time window was scanned along each source light curve, with the window length being varied from short timescales (0.5 days, $\sim$ 10 ScW) to the longest possible (i.e. the full light curve duration). All points within each window are used to calculate a significance (flux divided by error), and the maximum significance (and its associated time window) are recorded, along with the significance across the whole light curve. Bursticity is then defined as the ratio of this maximum significance to the significance for the whole light curve.

This definition means a persistent source should have bursticity of 1, as the greatest significance for a persistent source comes from combining all the data points in the light curve. Any source with bursticity above 1 must have some subset of its light curve where the observations are of higher significance than the entire light curve, likely due to a sharp increase in flux due to an outburst. One limitation of this method is that multiple repeated bursts would not have their significances combined to make an even higher bursticity, however this is a robust way to find sources with intermittent high significance which would be expected for a burst.

This bursticity method was highly impactful in the previous search for sources in cat1000 (Bird et al., 2016), aiding in the discovery of around 100 sources that would have otherwise been missed. The method additionally identifies the time interval for maximum bursticity, a key property of the source variability. The cat1000 burst analysis also produced mosaic maps on the timescale of maximum significance, enabling precise determination of key source properties such as its position by combining the highest significance observations to minimise measurement errors.

6.2 Dataset

The initial dataset intended to be analysed with the new burst detector consisted of the 18-60 keV IBIS/ISGRI light curves for sources identified in cat1000. Each light curve has flux values extracted from ScW-level ISGRI flux maps at the candidate source position covering the first 1500 orbits of INTEGRAL. In general these ISGRI light curves consist of numerous groups of consecutive ScW data points, which are separated in time by the 20-minute ScW duration, with groups of data points separated by long gaps of a few to ten days.

To provide more comprehensive temporal coverage, the ISGRI observations are also augmented by fluxes from the *Swift* BAT daily light curve data. These fluxes are the daily average (and respective error) flux for the source in bins of one day. As such, these BAT flux points have a broader range of sensitivities depending on the effective observing time for the source.

The ISGRI-only light curves can be useful for directly testing the performance of the new burst detector as compared to the previous approach used for cat1000. In most cases the BAT data is helpful for filling in observational gaps which can reveal some outburst activity that might be otherwise missed with only ISGRI data (see, e.g., 1A0535+262 in the lower panel of Figure 6.1). Whilst BAT provides good temporal coverage, the ISGRI data are generally more sensitive and so enable detection of outburst activity that might fall below the BAT detection threshold. This balanced dataset forms the main sample of light curves intended to be analysed with the new burst detector.

6.3 Clustering (unsupervised)

One of the earliest attempts that was made to explore burst detection techniques involved unsupervised learning where light curves were broken into sub-sequences and a clustering algorithm applied to look for groups with similar features. This approach was intended to separate quiescent segments of light curves from those segments containing burst-like activity.

This method requires choosing a window size and stride (separation between subsequent window starts) size, so allows the algorithm to decide what is an anomaly. A sliding window of size 50 data points was used so as to compare with the current software ('bursticity') as this is the window size it uses. Each one of these 50 points of the light curve was projected into a 50-dimensional space and then the window was moved along (with a stride of 10) and this was repeated for the whole lightcurve. Each point in this space represents a subsequence of the lightcurve. Similar segments will cluster together. The middle of each cluster (the centroid - this can be considered the average of the cluster) will provide some measure of the prototypical lightcurve pattern that all those segments are specific instances of. The cluster centroids provide us with a set of "normal" lightcurve segments.

Then the algorithm tries to use the set of "normal" segments to reconstruct a set of data to be tested. If the data is similar in shape to the original lightcurve, it will be able to manage a good reconstruction. However, if the data contains some abnormal shape, it will not be able to reconstruct it using our normal shape library, and it will get a reconstruction error. This error will indicate an anomaly.

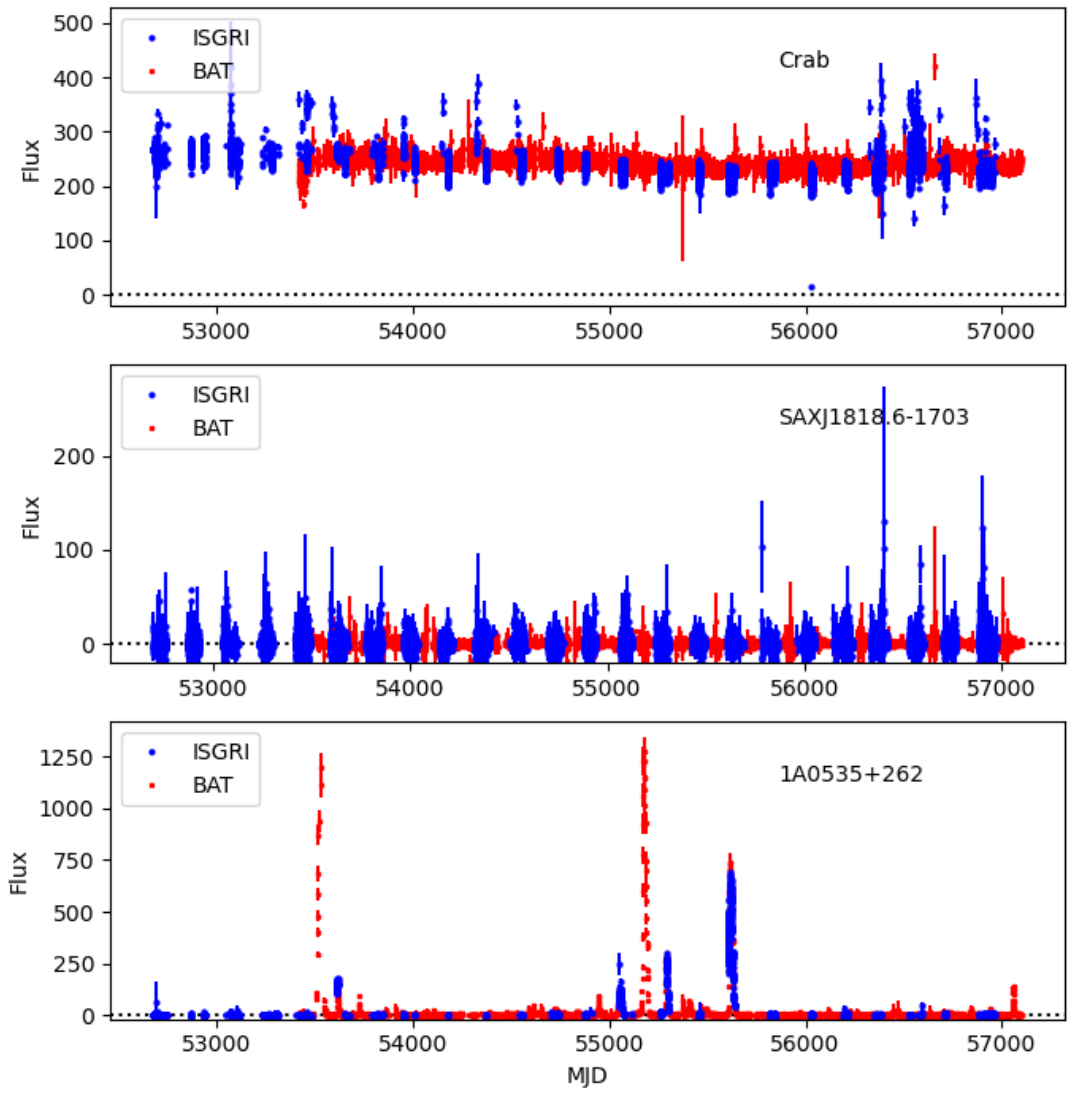

FIGURE 6.1: Example (real) light curves including ISGRI and BAT daily flux measurements.

As INTEGRAL is not always observing all areas of the sky there are gaps in the lightcurves. These gaps were removed (i.e. lightcurves are presented to this algorithm as a sequence of fluxes without time labels), but a suitable size of sliding window and stride were selected to ensure that the algorithm would not be confused with parts of the lightcurve that may have a big gap in observation times. As the fluxes from ISGRI have error bars, sometimes very large ones, significance light curves were used as the significance is just the ratio of the flux with its error. This approach is illustrated in Figure 6.2.

This approach had some key advantages: it was insensitive to gaps in the light curve, and also did not require an explicitly labeled dataset for training. However, this method is not robust to poorer quality data where bursts may have substantial overlap with noise. In these cases, removing temporal gaps can be disadvantageous

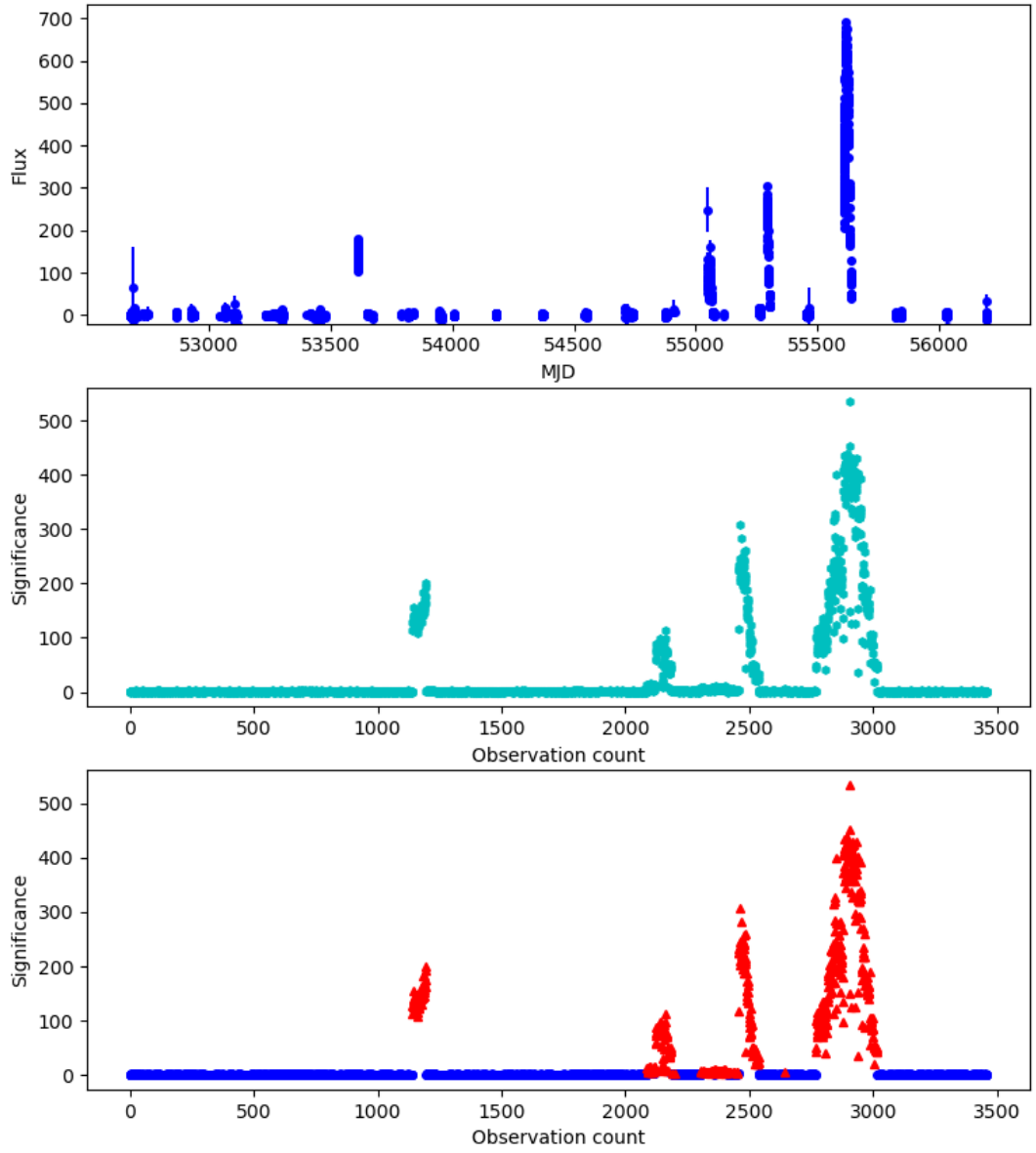

FIGURE 6.2: Outlier detection tests for ISGRI light curve of 1A0535+262. Top: full light curve, Middle: significance light curve with gaps removed, Bottom: significance light curve classified by outlier detection (outliers shown as red triangles).

because that temporal evolution of flux contains a key signal that is washed out. Though it might be possible to fill temporal gaps with data from other instruments (e.g., BAT data used in the later analyses of this chapter), this would produce a heterogeneous dataset that is not well suited to a clustering approach. For these reasons it was decided that outlier detection would not be a suitable approach for burst detection on the full dataset.

6.4 Supervised classification

Outburst detection in light curves could also be approached as a time-series classification problem. Time series classification uses supervised machine learning to analyse multiple labelled classes of time series data and then predict or classify the class that a previously unlabelled time-series belongs to. Specifically, this problem involves training a classifier on a large sample of light curves that either have a burst or do not have a burst, and applying that classifier to a new light curve to determine whether a burst is present.

Similarly to the source detector CNN, supervised time series classification in all of the data needs a class label for all of the training data. Another challenge with real-world data collection is that the lengths of the samples in a time series are typically different, which can be problematic for some classifiers. Gaps in the data, or data values that are null, are also not handled well by most time series classification algorithms.

6.4.1 Long Short Term Memory

Recurrent neural networks (RNN)/long short term memory (LSTM; Hochreiter and Schmidhuber, 1997) networks have been seen to be be particularly useful for learning sequences containing longer term patterns of unknown length, due to their ability to maintain long term memory. Recurrent hidden layers can be stacked to form networks that allow the learning of higher level temporal features.

A RNN is a type of artificial neural network which uses sequential data or time series data and they are often applied to ordinal or temporal problems, such as language translation, natural language processing, speech recognition, and image captioning (Rumelhart et al., 1985). Similarly to CNNs, RNNs need to be supplied with training data although they differ by having a “memory”; they take information from previous inputs which influences the current input and output. Unlike traditional deep neural networks RNNs presume the output of recurrent neural networks depend on the prior elements within the sequence.

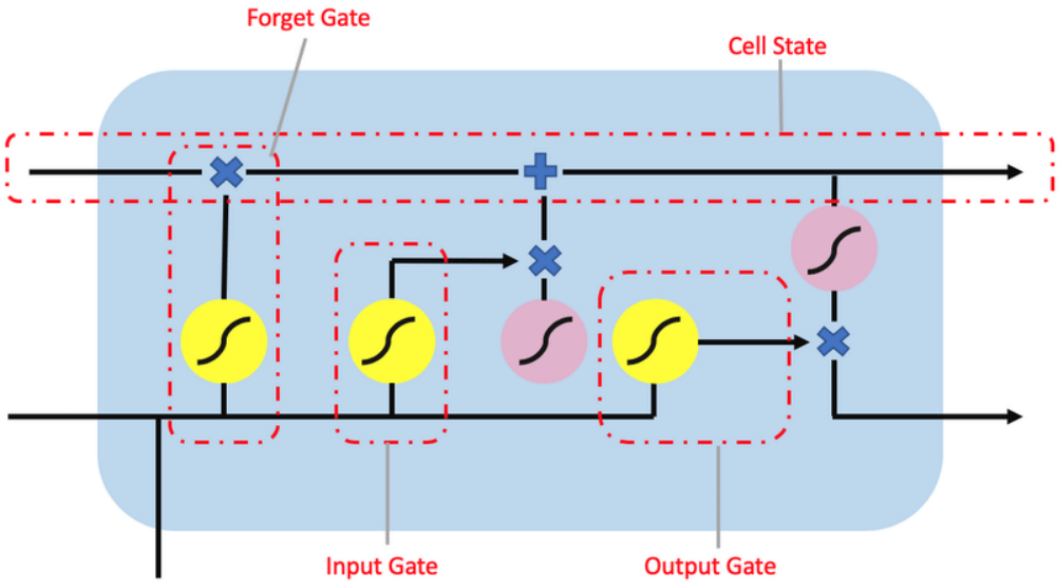

FIGURE 6.3: The LSTM unit is composed of a forget gate, an output gate, and an input gate. The yellow circle depicts the sigmoid activation function, whereas the pink circle depicts a tanh activation function. The "x" and "+" symbols are the element-wise multiplication and addition operator (Rengasamy et al., 2020).

RNNs suffer from short-term memory, which in practise means sequences over a certain length find it difficult to carry information from earlier time steps to later ones. Also during back propagation, RNNs suffer from the vanishing gradient problem, meaning if a layer gets a small gradient update then it will stop learning. This is usually seen in earlier layers of the network resulting in the network "forgetting" what it had previously seen in longer sequences.

The benefit of using LSTMs (Hochreiter and Schmidhuber, 1997) for sequence classification is that they can learn from the raw time series data directly, and in turn do not require domain expertise to manually engineer input features. The model can learn an internal representation of the time series data and ideally achieve comparable performance to models fit on a version of the dataset with engineered features (Hochreiter and Schmidhuber, 1997).

A LSTM network includes at least one LSTM cell. LSTM cells are designed to retain information for extended periods, making them well-suited for tasks requiring the model to remember past events. As shown in figure 6.3 LSTM cells have three gates (input, output, and forget gates) that control the flow of information into and out of the cell, as well as an internal state, called the cell state (Lu and Salem, 2017). The gates and cell state allow LSTM cells to selectively remember or forget information from the past, making them more effective at modelling long-term dependencies than other types of RNNs (Schäfer et al., 2006).

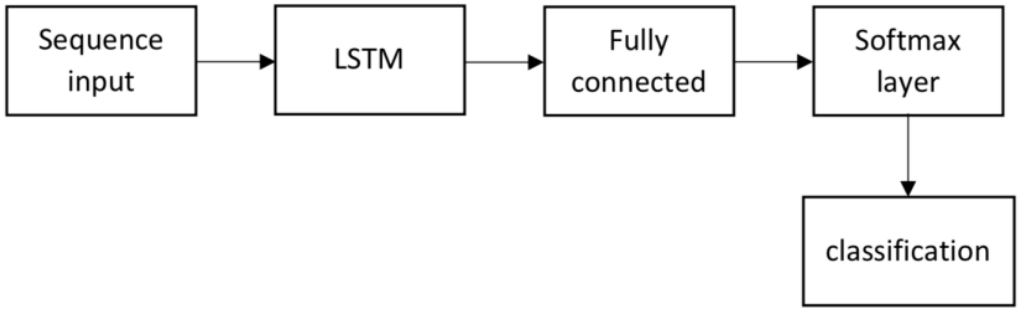

FIGURE 6.4: Architecture diagram of simple LSTM network for classification

In a typical LSTM model, the LSTM cells are stacked on top of each other, and the output of one cell is fed as input to the next cell. The final output of the LSTM layer is then fed into a fully connected layer and a softmax output layer to produce the final predictions.

As we have previously stated, LSTMs are a robust artificial neural network for classifying, processing and predicting time series data inspired by recurrent neural networks (RNN). LSTMs are particularly well-suited to time series classification tasks Korstanje (2021). A typical LSTM architecture for time series classification consists of one or more LSTM cells followed by a fully connected layer and a softmax output layer as seen in Figure 6.4.

When an LSTM is used for classification, the general process starts with the sequence input flow into the LSTM layer; then, it goes through the fully connected layer and softmax layer. Finally, the process ends with the classification layer to assign the class label. The LSTM layer is the core unit of the whole process. The architecture diagram of an LSTM layer is demonstrated in Figure 6.5. The key idea of LSTM is the previous cell state, and the current sequence of time steps is used to compute the current output and update the current cell state (Karim et al., 2017).

In an LSTM layer (Figure 6.5), there are the learning weights W, recurrent weights R, and the bias b. At each time step t, the LSTM network processes the input vector x_t and the hidden state from the previous time step h_{t-1} to update the current hidden state h_t. The calculation of the current hidden state h_t is done through the following processes:

- The forget gate f_t: decides what information to discard from the previous memory cell state c_{t-1}. It is computed using a sigmoid activation function (σ) applied to a linear combination of the previous hidden state h_{t-1} and the current input x_t. $f_t = \sigma(W_f h_{t-1} + W_f x_t + b_f)$

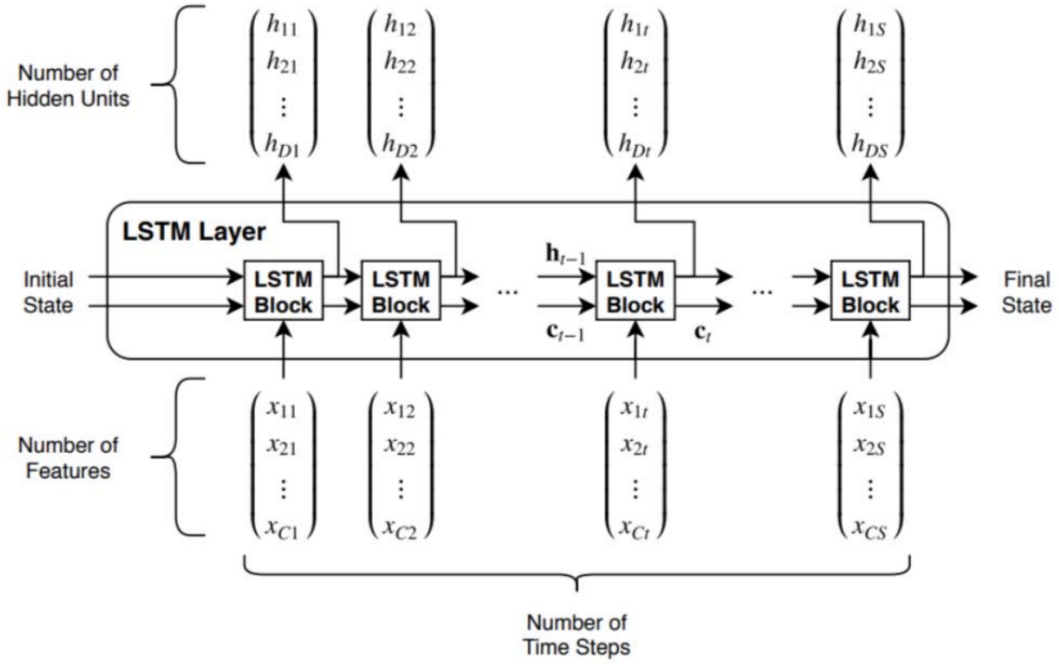

FIGURE 6.5: Architecture diagram of LSTM layer

- The input gate i_t: decides what new information to add to the current memory cell state c_t. It is computed using a sigmoid activation function applied to a linear combination of the previous hidden state h_{t-1} and the current input x_t.
  $i_t = \sigma(W_i h_{t-1} + W_i x_t + b_i)$

- The memory cell state c_t: stores information from the current and previous time steps. It is updated using the previous memory cell state c_{t-1}, the forget gate f_t, and the input gate i_t with a tanh activation function.
  $c_t = f_t c_{t-1} + i_t \tanh(W_c h_{t-1} + W_c x_t + b_c)$

- The output gate o_t: decides what information to output and use for the prediction. It is computed using a sigmoid activation function applied to a linear combination of the current hidden state h_{t-1} and the current input x_t.
  $o_t = \sigma(W_o h_{t-1} + W_o x_t + b_o)$

- The current hidden state h_t: is the output of the LSTM at time step t. It is computed using the memory cell state c_t and the output gate o_t with a tanh activation function. $h_t = o_t \tanh(c_t)$

In a classification task, the input sequence is processed one time step at a time, and the LSTM gates continuously update the memory cell state. The final hidden state of the LSTM is then used to make a prediction using a fully connected layer with a softmax activation function. The prediction is a probability distribution over the possible classes, and the class with the highest probability is considered the final prediction

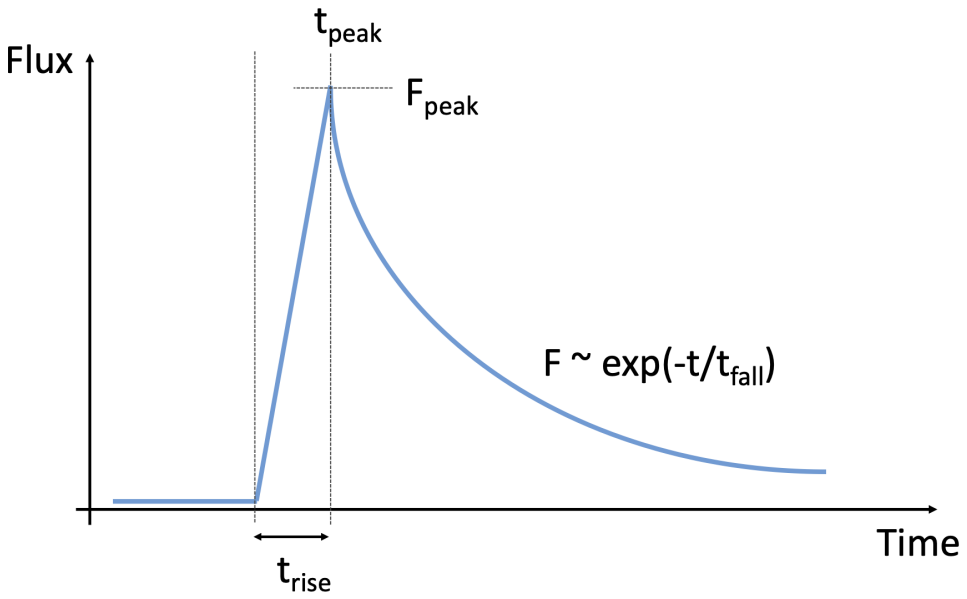

FIGURE 6.6: Representative FRED (Fast Rise Exponential Decay) burst flux profile: the burst flux rises quickly in a time t_{rise} to a peak flux F_{peak} (at time t_{peak}) then decays exponentially with a characteristic decay time of t_{fall}.

$p = softmax(W_p h_t + b_p)$ where p is the predicted probability distribution, $W_p h_t$ is the weight matrix connecting the final hidden state h_t to the prediction, and b_p is the bias term.

6.4.2 Simulated Bursts

Supervised learning requires a very large reliably-labelled training dataset, which would require an untenable amount of expert human time to label such a large real dataset. Thus it was decided that simulated light curves containing bursts should be used to train and assess the potential performance of the model. In addition to providing a large training set, simulated bursts enable us to have a ground truth for training the model.

6.4.2.1 Properties of Simulated Bursts

The type of bursts generated in these simulations are what are commonly referred to as "FRED"s – Fast Rise Exponential Decay. The shape of a FRED and its key parameters are illustrated in Figure 6.6: the burst has a rapid linear rise on a timescale of t_{rise} to a peak flux f_{peak} and then decays exponentially with a characteristic decay timescale t_{decay}. To be explicit the decaying flux goes like $\exp(-t/t_{decay})$ where $t = 0$ is the time of peak flux. The range of values for these three parameters used in the simulations is shown in Table 6.1.

Parameter	Range
t_{rise}	0.01 – 2 days
t_{fall}	1 – 100 days
F_{peak}	0.1 – 100 mCrab

TABLE 6.1: Range of FRED parameters in simulated bursts (values chosen from uniform distribution in this range).

It is worth noting that not all X-ray bursts will take on a FRED shape, and the decision to model all simulated bursts as FREDs is not intended to represent all possible X-ray burst shapes. Instead, this was chosen as a sensible starting point for training a model to identify bursts whose shape was known (and could be simply parametrised). Inclusion of diverse burst shapes is an intended goal for future development of the burst detector.

The final goal of the simulated bursts is to produce simulated equivalents of real (noisy and perhaps sparsely sampled) light curves. This was done in two steps: first simulate a "perfect" light curve, then use it to simulate what a real instrument would have observed from it. The former task is described here whilst the latter task is described in the next subsection.

A simulated light curve can have one or several bursts in it. Each light curve consists of flux values and associated time values spanning the duration of the real light curves (about 3000 days). In practice the time values are an array on intervals of 0.001 days, much shorter than ISGRI ScW separations, and flux values start at zero and increase when a simulated burst is added to the light curve. Each bursty light curve gets a random number of bursts drawn uniformly between 1 and 8 bursts, with the time of peak flux chosen randomly from all available time values and FRED parameters chosen uniformly from the ranges shown in Table 6.1 A few example simulated "perfect" light curves are shown in Figure 6.7.

6.4.2.2 Mimicking Real Observations

Once simulated "perfect" light curves were generated, there was then the need to simulate what real observations of this "perfect" light curve would look like. This means simulating the observational gaps like those in real ISGRI light curves, and simulating the noise as seen by the instrument. For most sources, ISGRI observations were augmented with the Swift BAT daily flux light curves, meaning each light curve is a combination of ISGRI and BAT observation dates.

Each final simulated light curve had time sampling drawn from the sample of observed light curves, i.e. a random source was chosen from the list of known ISGRI sources and the simulated light curve used the time sampling of that source. The flux

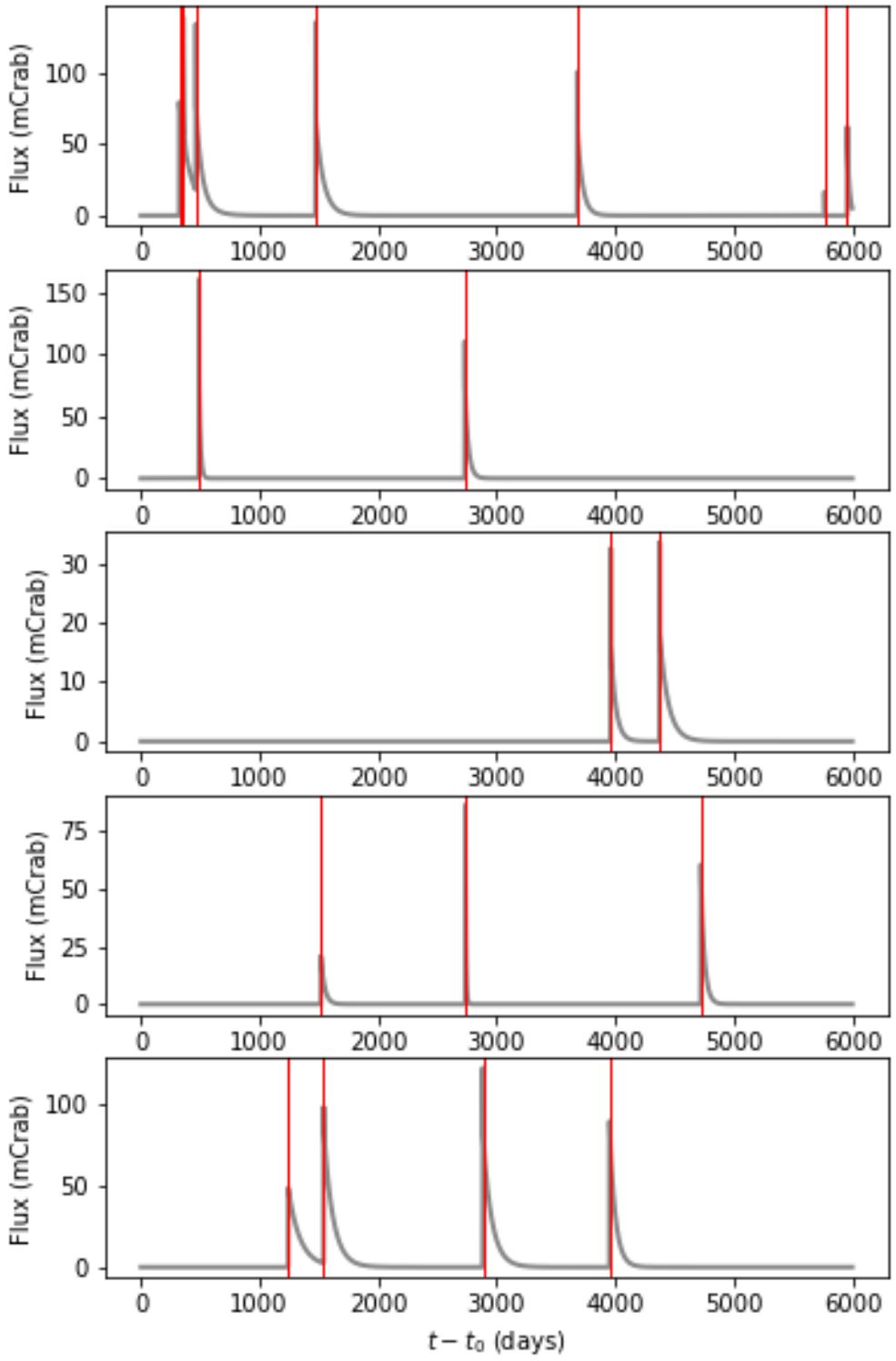

FIGURE 6.7: Examples of "perfect" light curves produced in the simulations (grey) with peak time for each burst highlighted (red vertical lines).

values for those times were then calculated by interpolating the over-sampled "perfect" light curve.

Errors for the simulated fluxes were calculated as follows. First a relationship between flux and error was examined for each instrument (ISGRI and BAT). A simple second order polynomial relationship was fitted between all log error and log flux values for each instrument, enabling the calculation of flux error for each simulated flux value. With this error value in hand, a random noise value was calculated and added to the "perfect" flux value for each epoch thus yielding the final simulated flux. Some examples of simulated light curves with real noise and observing gaps are shown in Figure 6.8.

6.4.3 Training and validating the burst detector model

Windowing the data: With both real and simulated light curves in hand, the model data format needed to be established to look for bursts in real data using a model trained on the simulated light curves. It was decided to train a model that would take as input flux and error values covering a window period t_{window} and uniformly sampled on a timescale of 20 minutes (one standard ScW duration). These evenly-sampled values were calculated by interpolating a light curve on the window time values, where for each light curve numerous windows were evaluated in steps of t_{step} typically 10% of the window size t_{window} (for example, with a 500-day window the data would have windows stepped by 50 days).

Setting the training set: The training set for the model was generated by applying this windowing process to the simulated light curves. Each window was then assigned a label depending on whether a burst was present in the window. For this analysis, a label of "1" meant the peak of a burst was inside the window whilst a label of "0" meant no burst peak was present, even if a significant fraction of the exponential decay of the burst is present.

6.4.4 Model Architecture

When designing a LSTM network typically one would consider the number of LSTM cells and the size of the fully connected layer and also the input and output dimensions of the LSTM cells. There is no fixed rule for how many LSTM layers that can be used in a time series classification model. The different numbers of layers can be experimented with to see how the model performance is affected. One consideration is the complexity of the time series data and the task. If the data is relatively simple and the task is straightforward, a single LSTM layer may be sufficient. On the other hand, if the data is very complex or the task is more difficult,

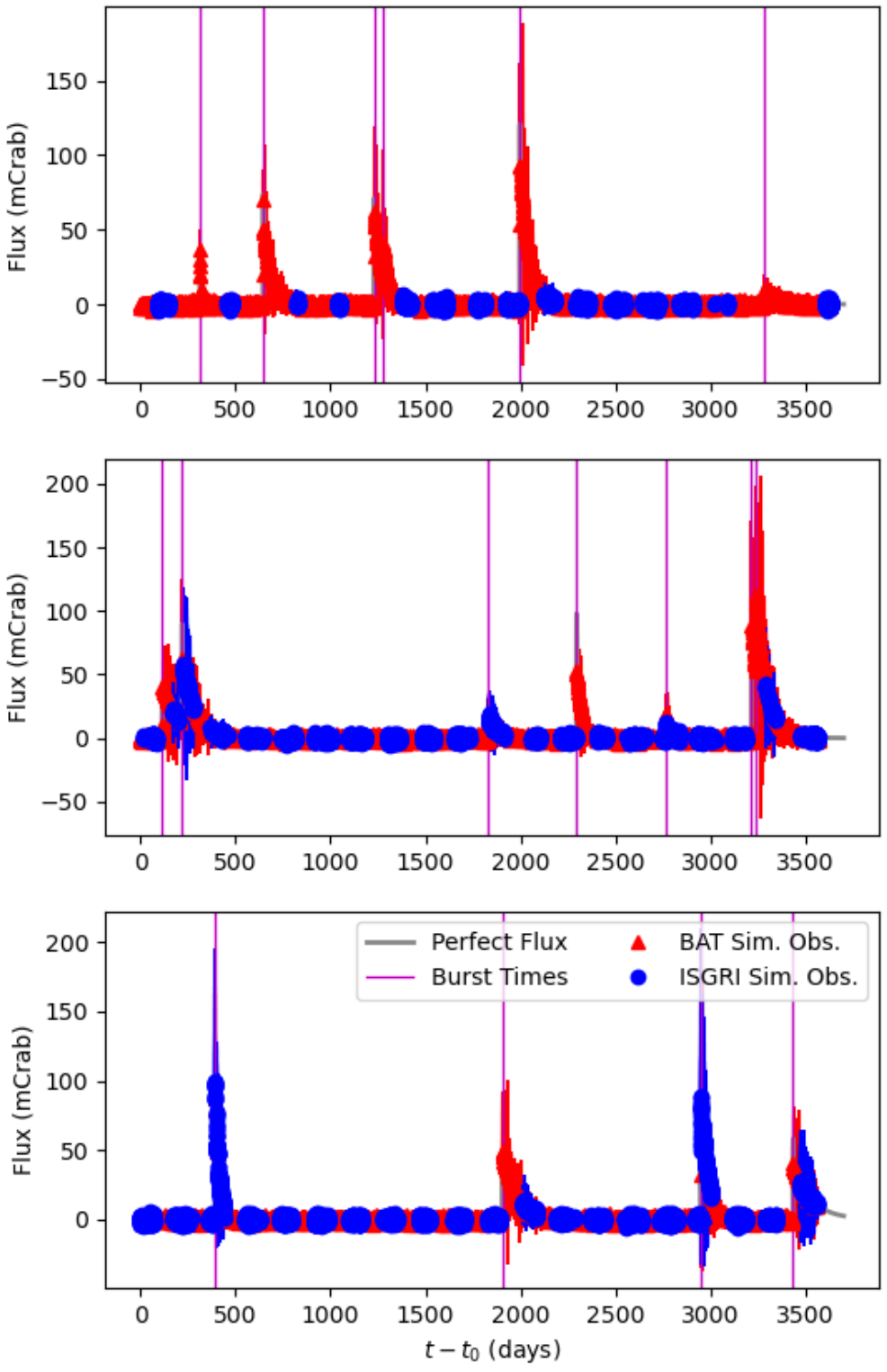

FIGURE 6.8: Examples of simulated light curves with realistic noise and gaps.

multiple LSTM layers may be needed to achieve a good performance (Pascanu et al., 2013).

A stacked LSTM is a deep learning model composed of multiple LSTM layers stacked on top of each other. The output of one layer is fed as input to the next layer. This allows the model to capture and process more complex and higher-level features from the input sequence (Hermans and Schrauwen, 2013). A stacked LSTM improves the ability of a deep learning model to capture and process complex relationships in input sequences. Some reasons for using a stacked LSTM include (Shinde et al., 2021):

- Handle long-term dependencies: Each LSTM layer can capture dependencies between elements in the input sequence that are separated by a large distance, allowing the model to handle long-term dependencies.
- Increase model capacity: Adding more LSTM layers to the model increases its capacity to learn and process complex relationships, allowing it to perform better on difficult tasks.
- Reduce overfitting: By stacking multiple LSTM layers, the model is able to learn more abstract and high-level features from the input, reducing overfitting to the training data.
- Handle high-dimensional input: Stacking multiple LSTM layers allows the model to process high-dimensional input data by gradually reducing the dimensionality through each layer.

In addition to the number of layers, the number of LSTM cells in each layer can be altered. The number of cells can affect the model's ability to remember past information, which can be important for time series data.

Graves et al. (2013) introduced Stacked LSTMs or Deep LSTMs in an application of LSTMs to speech recognition, outperforming a benchmark on a difficult standard problem. It was discovered that the depth of the network was more important than the number of memory cells in a given layer. Stacking LSTMs is now a reliable method for solving complex time series problems.

This motivated the exploration of different LSTM network architectures to achieve a model that could suitably capture the complexity of the light curve data. Figure 6.9 shows the final architecture for the LSTM network used in this work, which uses stacked LSTMs followed by a multi-layer perceptron for final light curve classification.

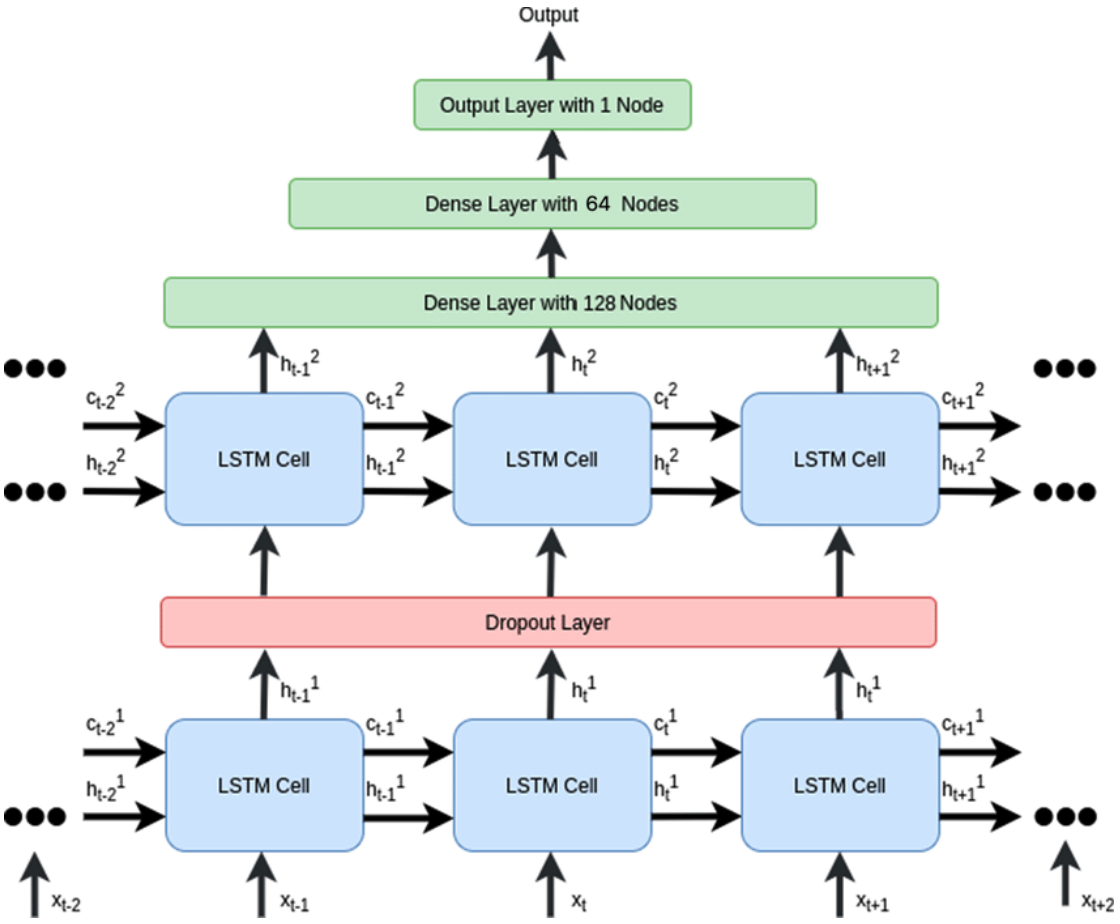

FIGURE 6.9: LSTM burst detector final architecture. The network consists of two LSTM layers, with 100 LSTM cells, then a multi-layer perceptron.

6.4.5 Validating model on simulated data

To establish the burst detector's performance as a proof of concept, its model parameters were refined by training and testing on simulated data for a variety of window sizes.

Model Training: Model training was performed by simulating 10,000 light curves containing a single burst with duration chosen randomly from the range listed above, with the integrated signal-to-noise of the burst chosen randomly between 3 and 9. In practice, the relationship between flux and error from ISGRI data was used along with the burst duration and sampling to set the peak flux of the burst for the chosen signal-to-noise. For a given window size W, segments of the simulated light curves were isolated (after re-sampling to a uniform time sampling as mentioned above) including the burst peak and labelled as 1. The training set was augmented with an equal number of samples of non-burst windows drawn from 2,000 light curves with zero flux but realistic noise and temporal sampling, and labelled these as 0. A model was trained with the above architecture for 1000 epochs, and retained both the final

trained model and a checkpointed version from when the loss function was minimised.

Model Testing: The model was tested for each window size W on a set of 10,000 simulated light curves as follows. First, the light curves were re-sampled to a common uniform time sampling, then used a running window of width w and step size $s = w/5$ (e.g. the 10-day window model was applied to windows of width $w = 10$ for every step of size $s = 2$) to collect windows of the light curve to be fed into the model. Windows where the model detected a burst (where it assigned label of 1 to the window) were collected into unique ranges by grouping overlapping windows. An example run of the burst detector on a simulated light curve is demonstrated in Figure 6.10, where the windows the LSTM flags as containing bursts are highlighted in purple. The success of the model was determined in two parts:

- *Fraction of bursts found*: did one of the regions labelled by the model as containing a burst contain the actual (simulated) burst peak?
- *Rate of false positives*: did the model label a range of the light curve as having a burst when there wasn't one?

Performance: Our LSTM models ran at substantially faster speeds than the bursticity method. Whilst bursticity took nearly four days to run on the full set of 10,000 simulated light curves, each of our models took approximately two hours to run on the same dataset. This run time was consistent across the four models, which makes sense as the product of the number of data points and number of windows analysed by the LSTM was the same for each model. It is worth noting that the run time for bursticity does have some dependence on the minimum and maximum burst durations allowed. Additionally, each LSTM took roughly five minutes to train on our GPU, whilst bursticity requires no training time. These performance figures correspond to being run on the same GPU-capable machine, and demonstrate how GPU-accelerated deep learning models can result in a substantial improvement in computation time.

Results: For this exercise four models were trained with window sizes of w =10, 25, 50, and 100 days. The fractions of found bursts as a function of signal-to-noise for these four models were compared to that of bursticity, and plot the results in Figure 6.11. It can be seen that the proof-of-concept LSTM-driven burst detector performs comparably, or better, at all signal-to-noise values. Completeness at low signal-to-noise improves as the window size increases, but as shown below this comes at the expense of more false detections.

Another important check to perform is to determine whether the burst detector is capable of finding bursts whose duration is shorter than its window size. Figure 6.12

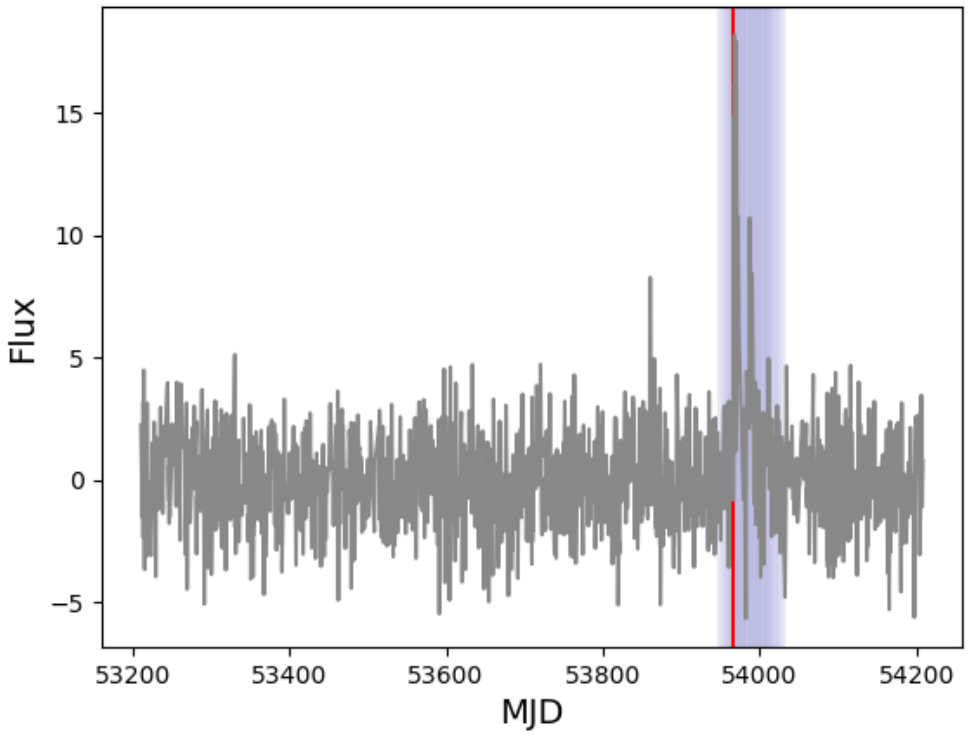

FIGURE 6.10: Example results of running our burst detector (in this case, the $w = 25$ days window size model) on a simulated light curve. The raw light curve is shown in grey, the peak of the simulated burst is highlighted by the vertical red line, and windows where the burst detector found a burst are shown as (partially transparent) purple regions. In this instance the burst detector reliably finds the burst.

shows a narrow simulated burst ($t_{fall} = 5$ days) run through all four window sizes – the burst detector successfully finds the burst peak in all four cases.

Figure 6.13 shows the fraction of light curves for which a false positive detection occurred, again as a function of signal-to-noise for the four models for our different window sizes. There is very little dependence of false detection fraction on signal-to-noise of the burst, as would be expected if the burst detector is finding false bursts in regions of pure noise. The fraction of false detections becomes significant as the window size increases to 50 days or higher. Specifically, the 50-day window model suffers from false detections in roughly 25% of light curves, whilst the 100-day window model has false detections in 60% of light curves. It is unclear what drives this increase in false detections, particularly given its lack of dependence on signal-to-noise, but ultimately it indicates the broader windows have a higher likelihood of being mistaken as containing a burst. Thus the increased detection success at low signal-to-noise comes at the expense of increased false detections at all signal-to-noise values.

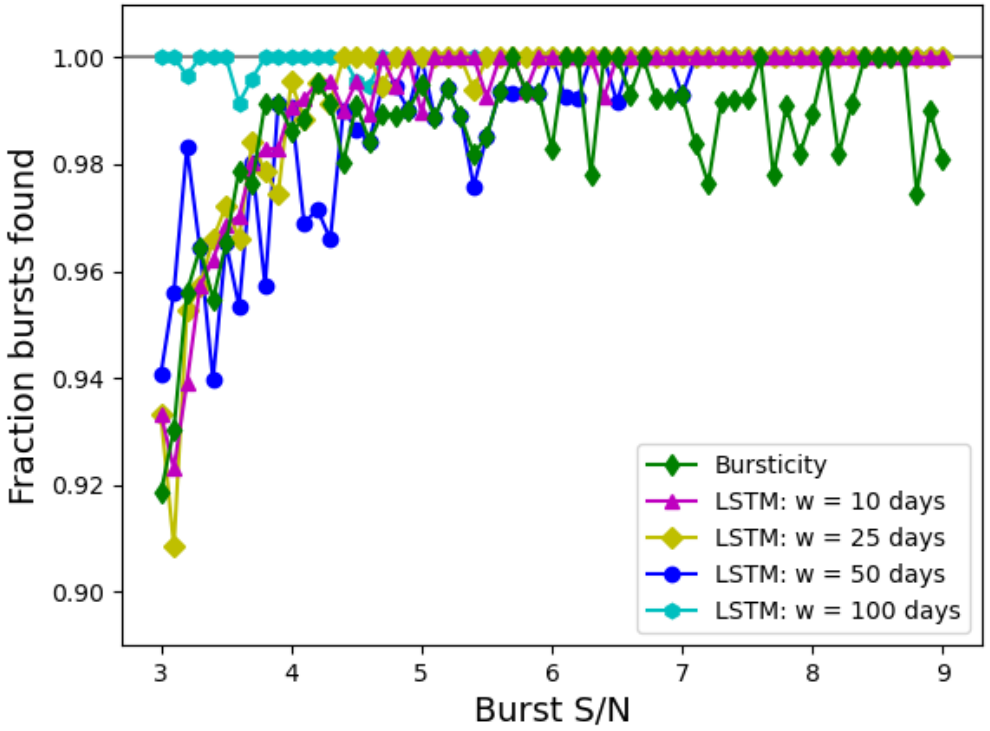

FIGURE 6.11: Fraction of simulated light curves where the ground truth burst was successfully detected. Results from the four LSTM models with different window sizes are shown, as well as the traditional "bursticity" method.

	Actual Negative	Actual Positive
Model Negative	6937	63
Model Positive	5	6995

TABLE 6.2: Confusion matrix for training set for 10-day window LSTM burst detector.

	Actual Negative	Actual Positive
Model Negative	2963	37
Model Positive	6	2994

TABLE 6.3: Confusion matrix for test set for 10-day window LSTM burst detector.

Tables 6.2 and 6.3 present the confusion matrix for the 10-day window burst detector model. As can be seen, the model performs very well overall, with only 1% mis-classifications. The consistent performance in training and test datasets gives reassurance that the model has not been over-fitted.

Results for light curves without bursts (false positive rates): The LSTM burst detector was applied to 2000 light curves with no bursts but realistic noise and time sampling (as described above). The LSTM was very successful at rejecting these noisy light curves: a burst was identified in 0 cases for the 10-day LSTM, 0 cases for the 25-day LSTM, 4 cases (0.2%) for the 50-day LSTM, and 1 case (0.05%) for the 100-day

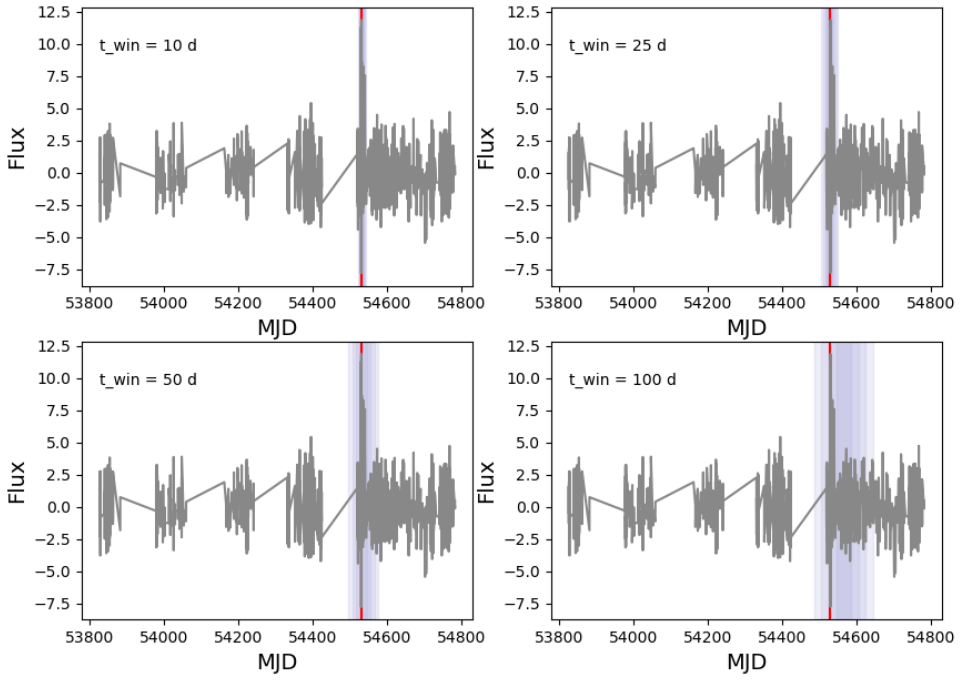

FIGURE 6.12: Representative test of whether the LSTM burst detector is capable of finding extremely narrow bursts (narrower than the search window). Here the same narrow burst (with $t_{fall} = 5$ days) was tested on all four window size models, and it can be seen that the burst can be successfully found by the LSTM burst detector.

	Actual Negative	Actual Positive
Model Negative	6981	19
Model Positive	0	7000

TABLE 6.4: Confusion matrix for training set for 25-day window LSTM burst detector.

	Actual Negative	Actual Positive
Model Negative	2987	13
Model Positive	0	3000

TABLE 6.5: Confusion matrix for test set for 25-day window LSTM burst detector.

	Actual Negative	Actual Positive
Model Negative	6997	3
Model Positive	0	7000

TABLE 6.6: Confusion matrix for training set for 50-day window LSTM burst detector.

	Actual Negative	Actual Positive
Model Negative	2997	3
Model Positive	0	3000

TABLE 6.7: Confusion matrix for test set for 50-day window LSTM burst detector.

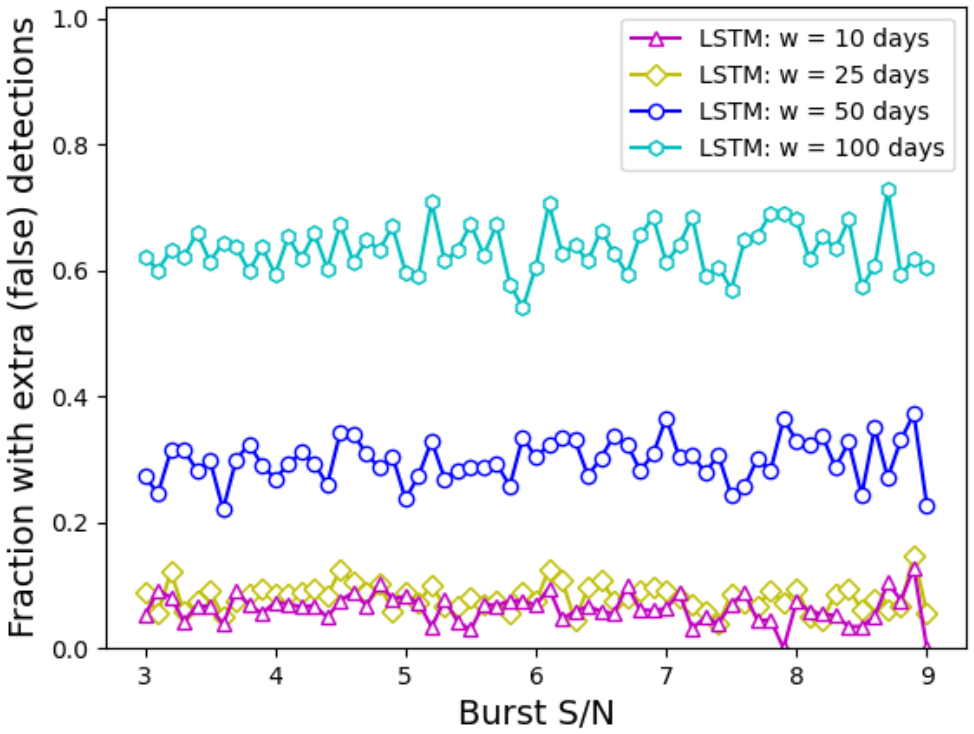

FIGURE 6.13: Fraction of simulated light curves with a region of the light curve that doesn't contain the actual burst peak erroneously flagged by our burst detector as containing a burst. This is plotted as a function of the burst peak signal-to-noise for the four models with different window sizes.

	Actual Negative	Actual Positive
Model Negative	6988	12
Model Positive	0	7000

TABLE 6.8: Confusion matrix for training set for 100-day window LSTM burst detector.

	Actual Negative	Actual Positive
Model Negative	2994	6
Model Positive	0	3000

TABLE 6.9: Confusion matrix for test set for 100-day window LSTM burst detector.

LSTM burst detector models, respectively. This makes sense (and is statistically consistent with results from the confusion matrices) as the LSTM was explicitly trained to reject purely noisy light curves.

In contrast, the previous burst finding method "bursticity" found a significant burst in these same noisy (burst-free) light curves about 2.5% of the time (with little dependence on the light curve duration). This makes sense because bursticity is looking for increased significance in a given window, which will always occur in at least some cases of pure noise. The LSTM, however, has been trained to reject these cases where the remainder of the light curve is inconsistent with an actual burst.

While the LSTM burst detector does a better job of rejecting purely noisy light curves, it does still show some false detections in other regions of the simulated light curves that contain bursts. This is undoubtedly a result of the training set presented to the burst detector, which has labelled burst peaks and pure noise. A more diverse training set with other regions of bursty light curves would likely improve this by helping the model learn how to identify regions with nonzero average flux but not a burst.

6.4.6 Initial tests on real light curves

Finally, the LSTM burst detector was tested on the actual light curves of select known sources from cat1000 to assess its readiness for application to real data.

As a requisite example, the 10-day window burst detector was applied to the light curve of IGRJ17464-3213, a black hole candidate x-ray transient. Figure 6.14 shows the LSTM applied to this light curve, with a zoomed view of some clear outburst activity successfully identified by the LSTM burst detector. Whilst a quantitative assessment of the burst detector's performance is impossible with real data where the underlying ground truth burst behaviour is inherently unknowable, such examples where experts would universally agree the behaviour is "burst-like" are instructive for demonstrating the burst detector's capabilities. It is also reassuring that the burst detector can identify clear outburst activity even though this source does not exhibit bursts with the same shape as the simulated training set (FREDs, which would not be expected from a source of this type). However the full duration of more extended bursts is not entirely identified by the burst detector, indicating there still exists some room for refinement of the burst detector if we want to use it for fully characterising burst durations.

A full assessment of the burst detector's ability to detect genuine outbursts in light curve data is planned as an immediate followup objective to this work. This requires substantial human effort to inspect and label outburst activities in the light curves, so this is planned as part of upcoming INTEGRAL survey team efforts. However, anecdotal evidence from light curve analyses like the one presented here show the

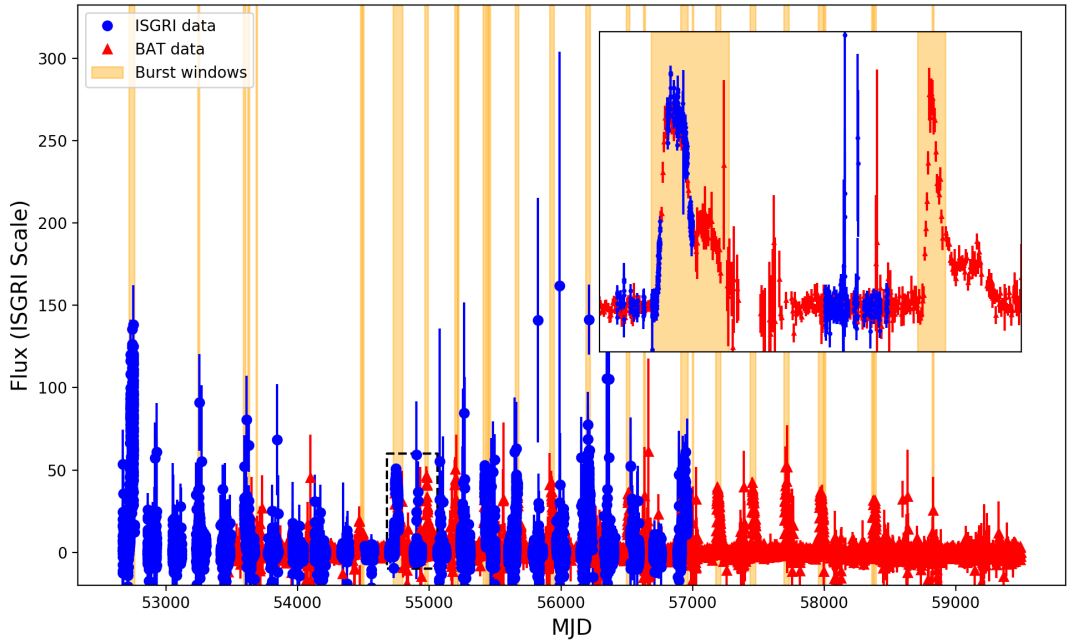

FIGURE 6.14: Light curve of IGRJ17464-3213 as analysed by the LSTM burst detector. The inset shows a zoom-in of some prominent activity (highlighted in the main plot by the dashed black box) where outbursts are successfully detected by the burst detector.

promise of this LSTM burst detector to become a powerful tool for outburst detections in light curves.

6.5 Case Study: SAX J1818.6-1703

SAX J1818.6-1703 is a SFXT high-mass binary detected in outburst by both SAX and INTEGRAL missions. INTEGRAL has detected this source many times, and established the 30-day orbital period of the binary system. Outbursts typically occur at periastron when the neutron star encounters the supergiant stellar wind, and are typically 1-2 days in duration. While SAX J1818.6-1703 is one of the SFXT systems with a relatively high duty cycle, only ∼20 significant outbursts have been seen by IBIS/ISGRI (see, e.g., Boon et al., 2016).

In this section the INTEGRAL and BAT light curve is used as a case study for testing the timing and burst detection capabilities of the previous burst detection technique *burstfind* and our new LSTM burst detector.

6.5.1 Performance of *burstfind*

A series of test runs have been performed with *bf3* (CPU version) and *bf4* (GPU version), varying the maximum burst duration being searched. Other run parameters

	search time (s)		
Max duration (d)	bf3 (CPU)	bf4 (GPU)	bursts found
10	24	1.1	23
100	146	2.6	22
1000	6840	51.5	18
10000	9660	52	2

TABLE 6.10: Performance for CPU and GPU versions of burstfind for test light curve

used were: (a) minimum window duration = 0.05d, (b) minimum science windows = 3, (c) off-axis angle limit = 20°, (d) residual limit = 0, (e) minimum burst significance = 5.

The SAX J1818.6-1703 test light curve contains 15580 data points. After application of filters (in this case only the *scw_exposure* $>$ *100s* filter as off-axis angle and residual filters were disabled), 15525 data points remain. The overall light curve has a detection significance of 22.3 sigma based on a mean flux of 0.335 ± 0.015 c/s.

It can be seen that the maximum burst duration has a major effect on run-time for the burstfind algorithm. The number of trial scw subsets considered increases rapidly as the maximum duration increases, and hence the run-time does too. This would encourage the user to select a short maximum duration, but unless the outbursts are all short, this carries risk of bias. As an example, BeX binaries typically exhibit Type 1 bursts of a few days, but then occasionally a much longer ($\sim$100d) outburst so selecting only short bursts would not be ideal. On the other hand, the run-times for the CPU-based code quickly become untenable for survey production where $\sim$5000 putative sources might need to be checked. Even selecting 1000d max duration would imply more than 1 CPU-year. The GPU-based version overcomes this problem with a typical speed-up of 50-100 times, and the problem becomes tractable even for larger numbers of light curves. Another effect of the burst duration parameter that can be seen is a sudden drop in the number of bursts identified when the maximum burst length is too great - in this case, once the brightest outburst has been identified and removed, the next optimised significance comes from including the whole remaining light curve. This is related to the original purpose of *burstfind* - the algorithm is designed to detect when/if a source is significantly detected, i.e. when the flux is significantly above a zero level. It is not designed to detect when a source is significantly above a *non-zero* baseline level, although in principle removing the source mean level between each iteration could be used to improve this aspect of the performance.

For reference, the bursts identified by the *bf4* code when run with a 10-day maximum duration is shown in Figure 6.15 below, compared to the set found by the LSTM burst detector. It should be noted that some of the lower-significance bursts are of longer duration (approaching the 10-day limit) and may be spurious. In practice,

LC Fraction	No. Windows	Runtime (s)
1/1000	2	0.037
1/100	22	0.084
1/10	220	0.154
1	2205	0.329

TABLE 6.11: Timing of LSTM burst detector for application to different subsets of SAXJ1818.6-1703 light curve.

bootstrapping techniques should be used to determine the true threshold level for a light curve. The threshold will depend upon the number of trials performed, and hence on both the length of the light curve and the analysis parameters.

6.5.2 Performance of LSTM burst detector

The search for bursts in the light curve of SAX J1818.6-170 was repeated using the new LSTM burst detector. Specifically, the version of the burst detector with 10-day windows was used, as this showed a low incidence of false positives and is probably best suited for analysing this source due to its short outburst durations.

The run time for the LSTM burst detector is not dependent on any input parameters but does depend on the volume of data presented to it. As was done above, the light curve for the source was resampled to an even time sampling, and 10-day windows with a stride of 2 days were generated (in total this produced $\sim$ 2200 windows covering the epochs of ISGRI data and contemporaneous BAT daily fluxes). The burst detector was run on different subsets of the resampled light curve to test its run time, and the results are shown below in Table 6.11.

It can be seen that the timing of executing the LSTM burst detector scales logarithmically with the volume of data presented to it. Even when presented with the full resampled light curve (which has over 1 million data points in all the resampled windows presented to the LSTM, a factor of nearly 100 greater than the number of points presented to *burstfind*), the burst detector takes well under a second to execute.

Figure 6.15 presents the regions of found bursts from both algorithms. There are some important details to note here: whilst *burstfind* is designed to isolate exact subsets of data containing bursts, the LSTM burst finder identifies which of the 10-day subsamples of the resampled light curve are likely to contain bursts (and overlapping windows have been combined into unique regions with burst-like activity). This means the minimum time region for burst-like activity is 10 days, and any time distinct episodes of outburst occur less than 10 days apart these will appear in the same region of burst-like activity.

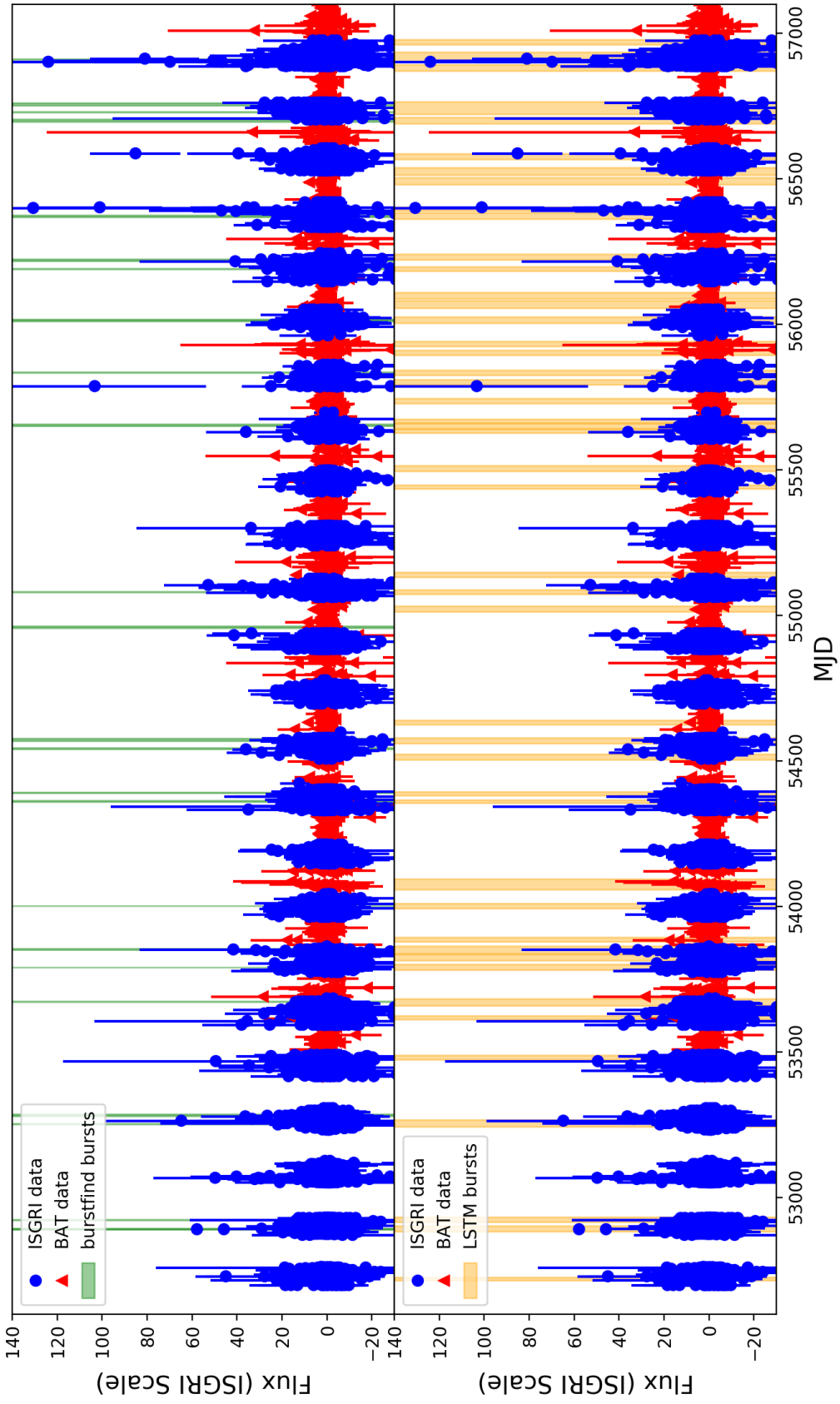

FIGURE 6.15: Burst detection algorithm comparison for the light curve of SAX J1818.6-1703 searched by *burstfind* (top) and the LSTM burst detector (bottom).

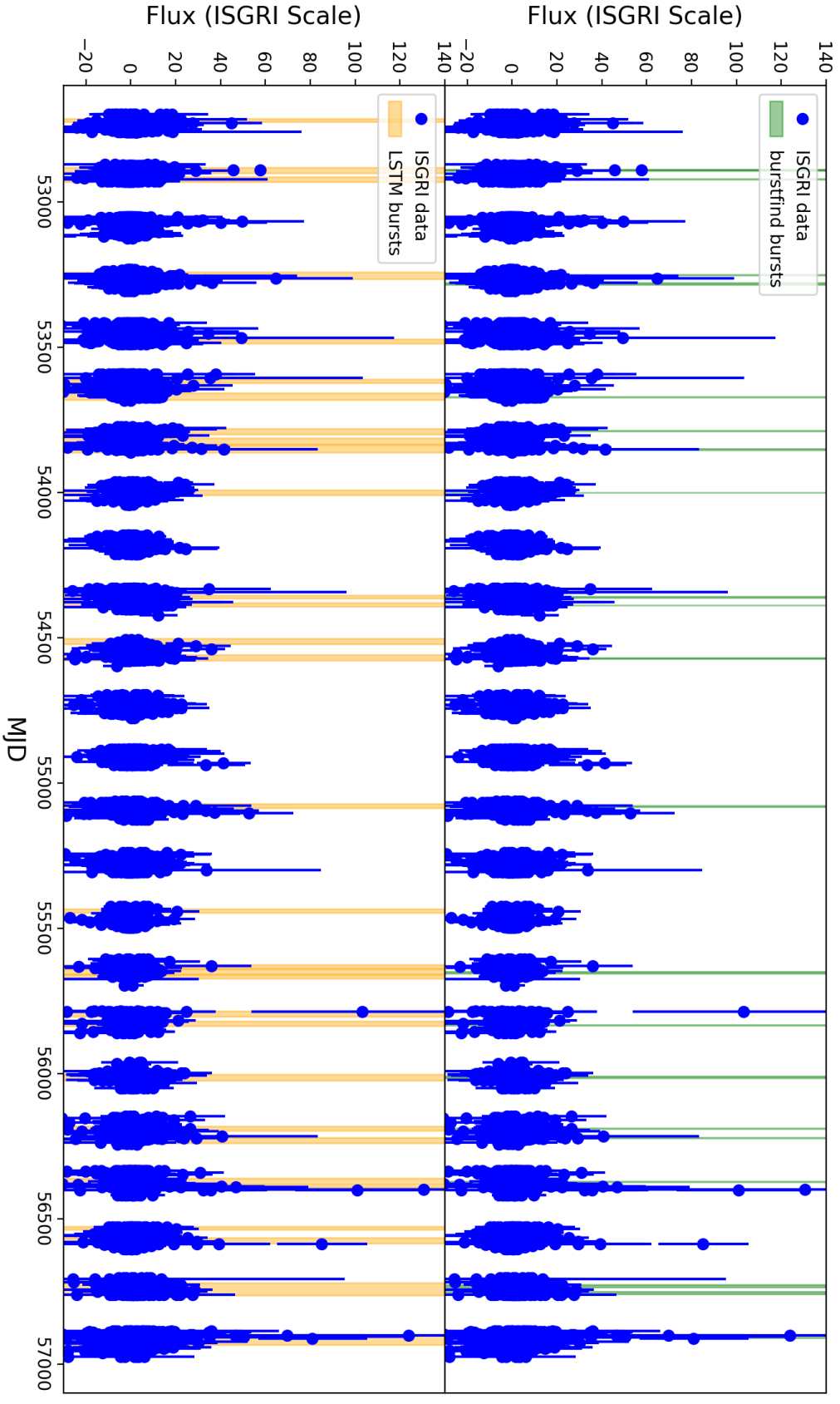

FIGURE 6.16: Same as Figure 6.15 but using only ISGRI data.

It is worth noting that the LSTM burst detector finds nearly all of the bursts identified by *burstfind* (for the combined ISGRI+BAT light curve) except two (which may be spurious), plus additional regions of burst-like activity. Initial visual inspection indicates most of the detected "bursts" appear legitimate, though it is difficult to assess the model's accuracy since the ground truth of the source's burst activity is unknown.

As an additional comparison, the ISGRI light curve alone was also tested with both burst finding algorithms, and the results are printed graphically in Figure 6.16. The results of this comparison are similar to that of the full ISGRI+BAT light curve analysis above: the LSTM finds nearly all of the *burstfind* bursts, plus some additional regions of burst-like activity. It is worth noting that the inclusion of BAT data provides additional sampling of the light curve in times between ISGRI observations, revealing the burst-like activity not observed by ISGRI. Conversely, the finer sampling of ISGRI means BAT data does not reveal substantially more bursts during periods where ISGRI observations are available.

6.6 Conclusions and next steps

This chapter presented a proof-of-concept test of a new ML-driven approach to detecting bursts in high energy time series data. The final model settled on was one whose core component is a LSTM which captures the temporal structure of outbursts.

It was demonstrated that the LSTM model trained on simulated bursty light curves was successful at finding bursts and rejecting noise. The fraction of bursts that were successfully found by our LSTM is comparable to or better than the previous method "bursticity". More importantly, it was found that the LSTM was able to operate substantially faster than the bursticity approach.

This increase in burst detection speed could be transformative for the search for transients in INTEGRAL data and potentially future missions as well. One key goal for the legacy INTEGRAL dataset is to derive light curves at every point on the sky through HEALpix resampling of all INTEGRAL/IBIS and Swift/BAT observations to the same pixel grid on the sky. HEALpix provides a non-orthogonal grid with quasi-uniform spacing across the whole sky. When map-based data is required, a switch to HEALpix-based mapping ensures compatibility across formats, and distortion-free maps for optimised source searching. This also ensures that instrument specific details such as resolution and detector discretisation can be treated appropriately for both telescopes together.

With the speed of the LSTM burst detector, it becomes feasible to search through all of these light curves at every point on the sky. Thus it will not only be possible to find

additional bursts in known transient sources, but also potentially feasible to find new transients with short burst activity not previously detected. The speed of the LSTM burst detector also means it could have applications in analysis of live on-board data from the next generation of high energy satellites. Whilst currently outburst detection in satellites is performed at the image level, the LSTM burst detector could potentially facilitate outburst detection at pixel level.

This analysis has also identified some ways in which the burst detector could be refined in the future. In some simulated light curves the LSTM burst detector labelled additional regions of the light curve as containing a burst when it did not. This means the burst detector succeeds at detecting real bursts and rejecting noise, but might sometimes flag additional (false) bursts. This can be remedied by supplying a more diverse training set that includes regions of bursty (simulated) light curves that do not contain the burst peak. For application to live data, this will require a training set that includes a broad range of light curve behaviours that are not classified as bursty – this is a time consuming challenge as it requires human labelling of data, or alternatively the ability to simulate all "expected" non-bursty light curve behaviours. Additionally, an ideal training set would also include a suitable variety of outburst shapes in the training sample, though this requires formal definitions of what constitutes a burst.

One additional feature planned for the burst detector is to combine results from LSTMs with varying window sizes. In a similar spirit to the manner in which many image classification algorithms combine convlutional kernels of multiple sizes (e.g., UNet), a burst detector that combines outputs from LSTMs with varying window sizes. This not only facilitate the combination of information about temporal behaviour on multiple timescales, but would remove any need for human selection of model hyperparameters such as window size.

Another improvement for the burst detector could be its expansion to operate on multi-band data. It was found for the source detector CNN that this increased the fidelity of the source detector because the presence of the source in multiple bands. This work starts developing the burst detector in a single band (18-60 keV) as a proof of concept, but the INTEGRAL data is already on hand to include multiple bands. If the analogous data from *Swift*/BAT can be prepared, then a multi-band burst detector could potentially be trained in a similar way. Whilst INTEGRAL data alone can still be sufficient to detect outbursts, it was shown above that the augmentation of *Swift* is advantageous for temporal coverage in INTEGRAL gaps.

Finally, another potential future development of the burst detector could be to explore a semi-supervised approach, particularly by treating this as an outlier detection problem. To do so would require the labelling of a large volume of "non-burst" data and training a model to detect data that isn't consistent with this. One key advantage of this approach is that it could have the potential to identify all anomalous light curve

behaviour without having to be explicitly trained to identify it (as the LSTM was trained to find FREDs). Once a burst has been detected, additional ML models could be deployed to classify the burst. This sort of "all-purpose" outburst detector would be extremely useful for both legacy datasets and future live satellite missions.

Chapter 7

Conclusions

Next generation astronomy missions will produce orders of magnitude greater volumes of data than current instruments. As high energy astronomy enters the era of big data, this community needs smarter methods for data analysis to keep pace with torrential accumulation of data. This need motivated the work in this thesis, where new tools were presented that leverage machine learning and Bayesian reasoning to efficiently analyse data from the INTEGRAL satellite. These tools were utilised to efficiently produce a new catalogue for the first 1000 orbits of INTEGRAL data. This new catalogue found 5% more sources than the catalogue produced using older methods on the same dataset, particularly sources near the Galactic centre which were blended in stacked images used for source detection in the previous catalogue. Furthermore, machine learning tools substantially reduce the bias introduced by human intervention. With such capabilities, these tools have the potential to transform the way high energy astronomy surveys are done.

7.1 Outcomes for new high energy astrophysics survey tools

This thesis presented three new tools for efficiently producing high energy astronomy surveys:

1. *CNN source detector:* A tool for detecting sources in INTEGRAL images that utilises a convolutional neural network (CNN) to reliably identify windows within the image that contain a genuine astrophysical source. The speed and reliability of this tool enables searching for sources at the science window level (individual images), recovering transient sources missed in stack image source catalogues. The source detector also removes the need for human intervention, consequently reducing human-induced bias.

2. *Bayesian matching:* A method for combining excesses detected in individual INTEGRAL images into a list of unique sources using Bayesian reasoning. The most critical benefit of this method is that it operates on a full list of excesses simultaneously and thus does not produce results dependent on the order of excess combination.

3. *Burst detector:* A tool for detecting outbursts in INTEGRAL+BAT light curves that utilises a long short-term memory (LSTM) model to capture time series behaviour and thereby reliably identify regions of light curves containing an outburst. This can be useful not only for source verification but also deeper scientific investigation into source variability.

7.2 Future applications of new tools

The source detector presented here has potential to revolutionise not only future surveys with INTEGRAL data, but also surveys produced by other instruments. The burst detector has broad applicability in time series data, particularly light curves for transient astrophysical sources. Both of these tools are already planned to be deployed for a new legacy survey HEASS (Section 7.2.1). The source detector has great potential for adaptation to future high energy astronomy missions, particularly the CTA (Section 7.2.2) which is rapidly approaching active operations. Finally, the burst detector could be transformative for future high energy satellite missions (Section 7.2.3).

7.2.1 A High Energy All-Sky Survey (HEASS)

The machine learning tools for high energy survey catalogue production presented in this thesis have already been earmarked to be integral components of a new project: the High Energy All-Sky Survey (HEASS) – a collaboration between the University of Southampton and INAF Bologna (including PI Dr. Angela Malizia) and INAF Roma in Italy. HEASS will generate the most sensitive, unbiased and complete survey of the hard X-ray sky (20-300keV) ever produced by combining over 20 years of INTEGRAL/ISGRI images with contemporaneous Swift/BAT observations. The legacy source catalogue for HEASS will be built with the ML source detector presented in this thesis. Additionally, the burst detector presented here will be crucial for the analysis of the light curves produced as part of HEASS.

The major backbone of HEASS will be the 20+ years of data collected by INTEGRAL, which currently consists of over 2500 orbits. This full dataset will first be processed with the latest version of the INTEGRAL software, OSA11. To make the final imaging data products for HEASS, all INTEGRAL images will be resampled to the sample

HEALPix-based grid on the sky. To augment and enhance this rich dataset, HEASS will resample and cross-calibrate Swift/BAT data covering the same mission period.

With the newly processed, resampled and cross-calibrated images on hand, the CNN source detector will then be retrained on the newly processed dataset so that it can capture the new subtleties of the data. The catalogue produced with the CNN source detector applied to this dataset will be transformative for high energy astronomy. The most recent BAT survey (Oh et al., 2018) lists approximately 1600 sources – comparing the sensitivity of that survey to the expected sensitivity of HEASS and using the $\log(N) - \log(S)$ relationship from (Bottacini et al., 2012) HEASS should yield a catalogue of 4000 sources.

With all INTEGRAL and BAT data resampled to the same HEALPix-based map, this will enable a creation of a light curve for every pixel on the sky. These light curves will be ideal for inspection with the burst detector to search for outbursts – both to determine a burst history for all known sources as well as potential discovery of new extremely short duration outbursts formerly hidden within long periods of otherwise quiescent emission. This effort could only be made possible by the burst detector's speed of execution, which facilitates the search for bursts in exceptional volumes of data. This volume of data would be untenable for examination by slower methods used in the past.

7.2.2 ML Source Detection in Future High Energy Surveys

Following on from the rich legacy generated with HEASS, the next natural evolution of the source detector will be to apply it to data from other high energy missions. One ideal candidate mission is the Cherenkov Telescope Array (CTA, Actis et al., 2011), as it will generate volumes of data that cannot be reasonably inspected by humans and will also have source signals with complex shapes that traditional source detection tools (e.g., SExtractor) are not well suited for. In the next few years CTA's data analysis preparation will be in an advanced state and will have a substantial body of simulations suitable for presenting to the ML source detector.

7.2.3 Live Burst Detection for High Energy Satellite Missions

The newly developed burst detector presented here has the potential to be greatly impactful in future high energy astrophysics missions. The burst detector operates on time series data to determine epochs when an astrophysical source is in outburst. It is designed to automatically detect bursts without the need for human inspection. This tool has the potential to be transformative for high energy transient astrophysics; if suitably tuned for new missions, this burst detector could be deployed on future space

missions to enable real-time decision making about target followup prioritisation. The versatility of the burst detector's deep-learning networks also means that with some fine tuning it can be retrained to look for other interesting time-series events.

7.3 Summary

The suite of tools presented in this thesis constitute a powerful toolkit for conducting high energy astronomy survey catalogue production. The CNN source detector can rapidly process a large volume of imaging data to reliably detect excesses – this means *every* image taken in a survey can be analysed for astrophysical sources. The Bayesian matching approach is able to then group those detections into a list of unique sources – these grouped excesses can then be combined to calculate average fluxes and positions (and their uncertainties) for a final catalogue of sources. This same collection of grouped excesses can be examined in the time domain as light curves – and these light curves can be presented to the burst detector to identify epochs of outburst, yielding outburst histories for each source. Combined, these tools form a comprehensive approach for detecting sources in high energy astronomy surveys and producing source catalogues and outburst histories. Most importantly, these tools are fast, reliable, and can operate without human input.

These tools promise to be a powerful toolkit for application to other high energy astronomy surveys. Some additional work is required to adapt these tools to other surveys. The source detector CNN must be trained on labelled examples of sources and background, as it is a supervised learning approach to image classification. The Bayesian matching approach does not require any specific tuning for new instruments, as it operates on the excess list produced by the source detector. The burst detector, like the source detector, is trained in a supervised learning approach so it requires labelled examples of bursts and quiescent periods from light curves from the survey. It was found that this step requires particular care to prepare a training set with a suitable variety of labelled light curves. The instrument-specific elements of the development of these tools was a subdominant component of the work of this thesis, so it is expected that all of these model tuning efforts are tractable for new surveys.

There already exists a wealth of potential applications of the machine learning driven tools presented in this thesis. The next major INTEGRAL catalogue is already slated to make use of these tools, and they will serve as vital instruments for a new survey HEASS being conducted in collaboration between Southampton and several Italian institutions. The source detector is well-poised to be adapted to future high energy astronomy missions, including potentially the CTA. The burst detector can also be adapted to detect and potentially classify different types of outbursts. In general, data gathered from any astronomy survey can be considered a sampling of the sky in two

spatial dimensions and one temporal dimensions (i.e., a data cube), plus potentially a spectral dimension. The tools presented here facilitate searching in all axes of survey parameter space: source detection in the spatial dimensions, burst detection in the temporal, and capability to integrate spectral information into both. Most crucially these tools are exceptionally fast and produce results with high fidelity without the need for human intervention – this makes them ideally suited to be applied to the torrent of data promised to be produced by the next generation of facilities as high energy astrophysics enters the era of big data.

Appendix A

Appendix: CNN Sources missing from cat1000

The following sources were detected with the CNN source detector but not in the stacked cat1000 images:

Coordinates and source type are taken from the INTEGRAL reference catalog at http://www.isdc.unige.ch/integral/catalog/latest/catalog.html

Source Name	RA	DEC	Description
1H 1746-370	17:50:13	-37:03:08	LMXB
1RXS J174607.8-21333	17:46:03	-21:33:27	Symbiotic star
AX J1735.8-3207	17:35:48	-32:07:12	Unclassified
AX J1739.3-2923	17:39:19	-29:23:54	Unclassified
HESS J1825-137	18:25:54	-13:47:00	Gamma-ray source
IGR J05253+6447	05:24:29	+64:44:44	AGN (Sey-2)
IGR J06074+2205	06:07:26.6	+22:05:48	HMXB (Be)
IGR J13057+2036	13:05:42.0	+20:36:00	Suspected AGN
IGR J16285-4630	16:28:27.1	-46:30:25	Unclassified
IGR J16336-4733	16:33:29.9	-47:33:33	Unclassified
IGR J16442-5548	16:44:38.2	-55:50:46	Unclassified
IGR J16582-2937	16:58:11.5	-29:37:34	Unclassified
IGR J17198-3020	17:19:48.7	-30:17:27	Unclassified
IGR J17219-1509	17:21:55.7	-15:09:40	Suspected AGN
IGR J17394-3007	17:39:26.6	-30:07:16	Unclassified
IGR J17407-2808	17:40:42.1	-28:07:26	LMXB
IGR J17461-2204	17:46:08.0	-22:03:32	Unclassified
IGR J17467-2848	17:46:43.9	-28:48:18	Unclassified
IGR J18159-3353	18:15:54.0	-33:53:00	Possible SFXT
IGR J18193-2542	18:19:17.0	-25:42:11	Unclassified
IGR J19254-3901	19:25:25.9	-39:01:44	Unclassified
IGR J19284+0107	19:28:29.6	+01:06:42	Unclassified
NGC 5899	15:15:23.6	+42:02:30	AGN (Sey-2)
RX J0137.7+5814	01:37:50.0	+58:14:11	Unclassified
Swift J1658.2-4242	16:58:12.7	-42:41:56	BH

TABLE A.1: Sources found by the CNN source detector but not included in cat1000.

References

J. Abril, L. Schmidtobreick, A. Ederoclite, et al. Disentangling cataclysmic variables in Gaia's HR diagram. , 492(1):L40–L44, February 2020. .

M. Actis, G. Agnetta, F. Aharonian, et al. Design concepts for the Cherenkov Telescope Array CTA: an advanced facility for ground-based high-energy gamma-ray astronomy. *Experimental Astronomy*, 32(3):193–316, December 2011. .

M. Alghamdi, M. Al-Mallah, S. Keteyian, et al. Predicting diabetes mellitus using smote and ensemble machine learning approach: The henry ford exercise testing (fit) project. *PloS one*, 12(7):e0179805, 2017.

M. Aljabri, S. S. Aljameel, R. M. A. Mohammad, et al. Intelligent techniques for detecting network attacks: review and research directions. *Sensors*, 21(21):7070, 2021.

L. G. A. Alves, J. S. Andrade Jr, Q. S. Hanley, et al. The hidden traits of endemic illiteracy in cities. *Physica A: Statistical Mechanics and its Applications*, 515:566–574, 2019.

M. N. Amin. Predicting price of daily commodities using machine learning. In *2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT)*, pages 1–6. IEEE, 2020.

R. Antonucci. Unified models for active galactic nuclei and quasars. , 31:473–521, January 1993. .

L. Arrabito and J. Bregeon. A DIRAC-based Prototype for the Cherenkov Telescope Array Data Management, Processing and Simulations. In R. Pizzo, E. R. Deul, J. D. Mol, J. de Plaa, and H. Verkouter, editors, *Astronomical Data Analysis Software and Systems XXIX*, volume 527 of *Astronomical Society of the Pacific Conference Series*, page 339, January 2020.

B. Arsioli and P. Dedin. Machine learning applied to multifrequency data in astrophysics: blazar classification. , 498(2):1750–1764, October 2020. .

F. Aslam, A. I. Hunjra, Z. Ftiti, et al. Insurance fraud detection: Evidence from artificial intelligence and machine learning. *Research in International Business and Finance*, 62:101744, 2022.

W. Bao, J. Yue, and Y. Rao. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PloS one*, 12(7):e0180944, 2017.

D. Baron. Machine learning in astronomy: a practical overview. 04 2019.

Dalya Baron and Dovi Poznanski. The weirdest SDSS galaxies: results from an outlier detection algorithm. , 465(4):4530–4555, March 2017. .

S. D. Barthelmy, L. M. Barbier, J. R Cummings, et al. The burst alert telescope (bat) on the swift midex mission. *Space Science Reviews*, 120:143–164, 2005.

A. Basavaraju, J. Du, F. Zhou, et al. A machine learning approach to road surface anomaly assessment using smartphone sensors. *IEEE Sensors Journal*, 20(5): 2635–2647, 2020. .

G. Bass and K. Borne. Supervised ensemble classification of Kepler variable stars. *Monthly Notices of the Royal Astronomical Society*, 459(4):3721–3737, jul 2016. ISSN 13652966. . URL https://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/stw810.

W. H. Baumgartner, J. Tueller, C. B. Markwardt, et al. The 70 Month Swift-BAT All-sky Hard X-Ray Survey. , 207(2):19, August 2013. .

D. Bazell and David W. Aha. Ensembles of Classifiers for Morphological Galaxy Classification. *The Astrophysical Journal*, 548(1):219–223, feb 2001. ISSN 0004-637X. . URL http://stacks.iop.org/0004-637X/548/i=1/a=219.

A. Bazzano, J. B. Stephen, M. Fiocchi, et al. INTEGRAL IBIS Census of the Sky Beyond 100 keV. , 649(1):L9–L12, September 2006. .

R. Beck, L. Dobos, T. Budavári, et al. Photometric redshifts for the SDSS Data Release 12. *Monthly Notices of the Royal Astronomical Society*, 460(2):1371–1381, aug 2016. ISSN 0035-8711. . URL https://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/stw1009.

V. Beckmann and C. R. Shrader. *Active Galactic Nuclei*. 2012.

E. Berger. Short-Duration Gamma-Ray Bursts. , 52:43–105, August 2014. .

E. Bertin and S. Arnouts. SExtractor: Software for source extraction. *Astronomy and Astrophysics Supplement Series*, 117(2):393–404, jun 1996. ISSN 0365-0138. . URL http://aas.aanda.org/10.1051/aas:1996164http://adsabs.harvard.edu/abs/1996A{&}AS..117..393B.

M. Bilicki, H. Hoekstra, M. J. I. Brown, et al. Photometric redshifts for the Kilo-Degree Survey. Machine-learning analysis with artificial neural networks. , 616:A69, August 2018. .

A. J. Bird, A. Malizia, A. Bazzano, et al. The first IBIS/ISGRI soft γ-ray galactic plane survey catalog. *The Astrophysical Journal Letters*, 28(revolution 46):33–37, may 2004. ISSN 0004-637X. . URL http://stacks.iop.org/1538-4357/607/i=1/a=L33http://dx.doi.org/10.1086/421772.

A. J. Bird, A. Malizia, A. Bazzano, et al. The Second IBIS/ISGRI Soft Gamma-Ray Survey Catalog. *The Astrophysical Journal*, 636(2):765–776, jan 2006. ISSN 0004-637X. . URL http://stacks.iop.org/0004-637X/636/i=2/a=765.

A. J. Bird, A. Malizia, A. Bazzano, et al. The Third IBIS/ISGRI Soft Gamma-Ray Survey Catalog. , 170(1):175–186, May 2007. .

A. J. Bird, A. Malizia, A. Bazzano, et al. THE FOURTH IBIS/ISGRI SOFT GAMMA-RAY SURVEY CATALOG. *The Astrophysical Journal Supplement Series*, 186 (1):1–9, jan 2010. ISSN 0067-0049. . URL http://stacks.iop.org/0067-0049/186/i=1/a=1?key=crossref.d0b329ddbf6566398fa64587d1c2e1cf.

A. J. Bird, A. Malizia, A. Bazzano, et al. THE IBIS SOFT GAMMA-RAY SKY AFTER 1000 INTEGRAL ORBITS. *The Astrophysical Journal Supplement Series*, 223(1):15, mar 2016. ISSN 1538-4365. . URL http://stacks.iop.org/0067-0049/223/i=1/a=15?key=crossref.74bd82155bf529793772d68868f420c1.

C. M. Bishop. Pattern recognition. *Machine Learning*, 2006. URL http://www.academia.edu/download/30428242/bg0137.pdf.

J. S. Bloom, J. W. Richards, P. E. Nugent, et al. Automating Discovery and Classification of Transients and Variable Stars in the Synoptic Survey Era. , 124(921): 1175, November 2012. .

C. D. Bochenek, V. V. Dwarkadas, J. M. Silverman, et al. X-ray emission from SN 2012ca: A Type Ia-CSM supernova explosion in a dense surrounding medium. , 473 (1):336–344, January 2018. .

K. K. Boddy, M. Lisanti, S. D. McDermott, et al. Snowmass2021 theory frontier white paper: Astrophysical and cosmological probes of dark matter. *Journal of High Energy Astrophysics*, 35:112–138, August 2022. .

M. Bojarski, D. Del Testa, D. Dworakowski, et al. End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*, 2016.

C. M. Boon, A. J. Bird, and A. B. others Hill. Spectral variation in the supergiant fast X-ray transient SAX J1818.6-1703 observed by XMM-Newton and INTEGRAL. , 456 (4):4111–4120, March 2016. .

V. Borkous, A. Kanionskii, M. Pavlinsky, et al. Imaging Capabilities of X-Ray All Sky Monitor for Alpha Space Station. In M. Matsuoka and N. Kawai, editors, *All-Sky X-Ray Observations in the Next Decade*, page 265, January 1997.

E. Bottacini, M. Ajello, and J. Greiner. The deep look at the hard x-ray sky: the swift–integral x-ray (six) survey. *The Astrophysical Journal Supplement Series*, 201(2): 34, 2012.

L. Breiman, J. Friedman, C.J. Stone, et al. Classification and regression trees. 1984.

P. S. Broos, K. V. Getman, M. S. Povich, et al. a Naive Bayes Source Classifier for X-Ray Sources. *The Astrophysical Journal Supplement Series*, 194(1):4, may 2011. ISSN 0067-0049. . URL http://stacks.iop.org/0067-0049/194/i=1/a=4?key=crossref.abbd67488593a3f5ba4c95bcb7e24de2.

T. Budavári and A. S. Szalay. Probabilistic Cross-Identification of Astronomical Sources. , 679(1):301–309, May 2008. .

D. N. Burrows, J. E. Hill, J. A. Nousek, et al. The Swift X-Ray Telescope. , 120(3-4): 165–195, October 2005. .

T. Cantat-Gaudin, F. Anders, A. Castro-Ginard, et al. Painting a portrait of the Galactic disc with its stellar clusters. , 640:A1, August 2020. .

M. Carrasco Kind and R. J. Brunner. Exhausting the information: Novel Bayesian combination of photometric redshift PDFs. *Monthly Notices of the Royal Astronomical Society*, 442(4):3380–3399, jun 2014. ISSN 13652966. . URL https://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/stu1098.

P. A. Charles and F. D. Seward. *Exploring the X-ray Universe*. CUP Archive, 1995.

T. Charnock and A. Moss. Deep Recurrent Neural Networks for Supernovae Classification. *The Astrophysical Journal*, 837(2):6, mar 2016. ISSN 20418213. . URL http://stacks.iop.org/2041-8205/837/i=2/a=L28?key=crossref.388bd920b1f2254a71d451bb3caa5e4dhttp://arxiv.org/abs/1606.07442.

C. Choi, J. Kim, J. Kim, et al. Development of heavy rain damage prediction model using machine learning based on big data. *Advances in meteorology*, 2018, 2018.

E. Churazov, R. Sunyaev, J. Isern, et al. Gamma-rays from Type Ia Supernova SN2014J. , 812(1):62, October 2015. .

A. O. Clarke, A. M. M. Scaife, R. Greenhalgh, et al. Identifying galaxies, quasars, and stars with machine learning: A new catalogue of classifications for 111 million sdss sources without spectra. *Astronomy & Astrophysics*, 639:A84, 2020.

M. J. Coe, G. F. Carpenter, A. R. Engel, et al. Journal of high energy X-ray observations from the Ariel V satellite. , 200:385–390, August 1982. .

R. H. D. Corbet. The three types of high-mass X-ray pulsator. , 220:1047–1056, June 1986. .

F. A. D. Córdova. Cataclysmic variable stars. In *X-ray Binaries*, pages 331–389, January 1995.

R. Cornelisse, J. J. M. in't Zand, E. Kuulkers, et al. Burst-only sources: probing type I X-ray bursters at low persistent luminosities. *Nuclear Physics B Proceedings Supplements*, 132:518–523, June 2004. .

M. Dai, S. Kuhlmann, Y. Wang, et al. Photometric classification and redshift estimation of LSST Supernovae. jan 2017. URL http://arxiv.org/abs/1701.05689.

Z. L. de Beurs, N. Islam, G. Gopalan, et al. A Comparative Study of Machine-learning Methods for X-Ray Binary Classification. , 933(1):116, July 2022. .

J. de Dios Rojas Olvera, I. Gómez-Vargas, and J. A. Vázquez. Observational cosmology with Artificial Neural Networks. *arXiv e-prints*, art. arXiv:2112.12645, December 2021.

P. R. Den Hartog, W. Hermsen, L. Kuiper, et al. Integral survey of the cassiopeia region in hard x rays. *Astronomy & Astrophysics*, 451(2):587–602, 2006.

L. Devroye, L. Györfi, and G. Lugosi. *A probabilistic theory of pattern recognition*, volume 31. Springer Science & Business Media, 2013.

P. E. Dewdney, P. J. Hall, R. T. Schilizzi, et al. The Square Kilometre Array. *IEEE Proceedings*, 97(8):1482–1496, August 2009. .

T. G. Dietterich. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees. *Machine Learning*, 40(2):139–157, 2000. ISSN 0885-6125. . URL http://link.springer.com/10.1023/A:1007607513941http://en.scientificcommons.org/42637098{%}5Cnuuid/7906280C-AEF8-405A-9A94-6BAA1DDAED1E.

H. Domínguez Sánchez, M. Huertas-Company, M. Bernardi, et al. Improving galaxy morphologies for SDSS with Deep Learning. , 476(3):3661–3676, February 2018. .

X. Dong. *Detecting AGNs using multi-filter imaging data*. PhD thesis, 2012. URL https://ui.adsabs.harvard.edu/{#}abs/2012PhDT.....97D/abstracthttp://search.proquest.com.ezproxy.library.yorku.ca/docview/1095136625?accountid=15182{%}5Cnhttp://sfx.scholarsportal.info/york?url{_}ver=Z39.88-2004{&}rft{_}val{_}fmt=info:ofi/fmt:kev:mtx:dissertation.

P. D. Dueben and P. Bauer. Challenges and design choices for global weather and climate models based on machine learning. *Geoscientific Model Development*, 11(10):3999–4009, 2018.

G. Eibl and K. P. Pfeiffer. Multiclass boosting for weak classifiers. *Journal of Machine Learning Research*, 6:189–210, 2005. ISSN 15337928. URL http://www.jmlr.org/papers/volume6/eibl05a/eibl05a.pdfhttp://jmlr.csail.mit.edu/papers/volume6/eibl05a/eibl05a.pdf.

A. Elliott, M. Cucuringu, M. Martinez Luaces, et al. Anomaly detection in networks with application to financial transaction networks. *arXiv preprint arXiv:1901.00402*, 2019.

A. N. Emelyanov, N. L. Aleksandrovich, and R. A. Sunyaev. A Catalog of X-ray Sources as Observed by the TTM/COMIS Telescope Onboard the Mir-Kvant Observatory in 1988-1998. *Astronomy Letters*, 26:297–308, May 2000. .

F. Esposito, D. Malerba, G. Semeraro, and J. Kay. A comparative analysis of methods for pruning decision trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):476–493, may 1997. ISSN 01628828. . URL http://ieeexplore.ieee.org/document/589207/.

S. Ethiraj and Bharath K. Bolla. Classification of quasars, galaxies, and stars in the mapping of the universe multi-modal deep learning. *arXiv preprint arXiv:2205.10745*, 2022.

E. D. Feigelson and G. J. Babu. Big data in astronomy. *Significance*, 9(4):22–25, aug 2012. ISSN 17409705. . URL http://doi.wiley.com/10.1111/j.1740-9713.2012.00587.x.

B. J. Ferdosi, H. Buddelmeijer, S. C. Trager, et al. Comparison of density estimation methods for astronomical datasets. *Astronomy & Astrophysics*, 531:A114, jul 2011. ISSN 0004-6361. . URL http://www.aanda.org/10.1051/0004-6361/201116878.

F. Ferrari, R. R. de Carvalho, and M. Trevisan. Morfometryka—a New Way of Establishing Morphological Classification of Galaxies. *The Astrophysical Journal*, 814 (1):55, nov 2015. ISSN 1538-4357. . URL http://stacks.iop.org/0004-637X/814/i=1/a=55?key=crossref.d82405d4abc89446493730b535f3dfa9.

R. A. Fisher. Iris. UCI Machine Learning Repository, 1988. DOI: https://doi.org/10.24432/C56C76.

G. J. Fishman, C. A. Meegan, T. A. Parnell, et al. Burst and Transient Source Experiment (BATSE) for the Gamma Ray Observatory (GRO). In *19th International Cosmic Ray Conference (ICRC19), Volume 3*, volume 3 of *International Cosmic Ray Conference*, pages 343–346, August 1985.

W. Forman, C. Jones, L. Cominsky, et al. The fourth Uhuru catalog of X-ray sources. , 38:357–412, December 1978. .

S. Fotopoulou, F. Pacaud, S. Paltani, et al. The XXL Survey VI. The 1000 brightest X-ray point sources. *Astronomy & Astrophysics*, 592:A5, aug 2016. ISSN 0004-6361. . URL http://www.aanda.org/10.1051/0004-6361/201527402.

D. Fraix-Burnet, C. Bouveyron, and J. Moultaka. Unsupervised classification of sdss galaxy spectra. *Astronomy & Astrophysics*, 649:A53, 2021.

S. Gairola, Y. K. Lal, V. Kumar, et al. A neural clickbait detection engine. *arXiv preprint arXiv:1710.01507*, 2017.

R. Garcia-Dias, C. Allende Prieto, J. Sánchez Almeida, et al. Machine learning in apogee-unsupervised spectral classification with k-means. *Astronomy & Astrophysics*, 612:A98, 2018.

V. K. Garg and A. T. Kalai. Meta-Unsupervised-Learning: A supervised approach to unsupervised learning. dec 2016. URL http://arxiv.org/abs/1612.09030.

N. Gehrels, G. Chincarini, P. Giommi, et al. The Swift Gamma-Ray Burst Mission. , 611(2):1005–1020, August 2004. .

R. Gentleman and V. J. Carey. Unsupervised Machine Learning. In *Bioconductor Case Studies*, pages 137–157. Springer New York, New York, NY, 2008. . URL http://link.springer.com/10.1007/978-0-387-77240-0{_}10.

Z. Ghahramani. Unsupervised learning. *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2-14, 2003, Tübingen, Germany, August 4-16, 2003, Revised Lectures*, pages 72–112, 2004.

R. Giacconi, H. Gursky, F. R. Paolini, et al. Evidence for x rays from sources outside the solar system. *Phys. Rev. Lett.*, 9:439–443, Dec 1962. . URL https://link.aps.org/doi/10.1103/PhysRevLett.9.439.

R. Giacconi, S. Murray, H. Gursky, et al. The Third UHURU Catalog of X-Ray Sources. , 27:37, February 1974. .

A. Goldwurm, P. David, L. Foschini, et al. The INTEGRAL/IBIS Scientific Data Analysis. *Astronomy and Astrophysics, v.411, p.L223-L229 (2003)*, 58:7, nov 2003. ISSN 0004-6361. . URL http://arxiv.org/abs/astro-ph/0311172http://dx.doi.org/10.1051/0004-6361:20031395.

A. Golob, M. Sawicki, A. D. Goulding, et al. Classifying stars, galaxies, and AGNs in CLAUDS + HSC-SSP using gradient boosted decision trees. , 503(3):4136–4146, May 2021. .

A. Graves, A.-R. Mohamed, and G. Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 6645–6649. Ieee, 2013.

R. H. R. Hahnloser, R. Sarpeshkar, M. A. Mahowald, et al. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, 405(6789): 947–951, 2000.

A. Hannun, C. Case, J. Casper, et al. Deep speech: Scaling up end-to-end speech recognition. *arXiv preprint arXiv:1412.5567*, 2014.

T. Hastie, R. Tibshirani, and J. H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

D. M. Hawkins. Introduction. In *Identification of Outliers*, pages 1–12. Springer Netherlands, Dordrecht, 1980. ISBN 978-94-015-3994-4. . URL http://link.springer.com/10.1007/978-94-015-3994-4{_}1.

H. He, Y. Bai, E. A. Garcia, et al. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, pages 1322–1328. IEEE, 2008.

M. Hermans and B. Schrauwen. Training and analysing deep recurrent neural networks. *Advances in neural information processing systems*, 26, 2013.

G. Hinton, L. Deng, D. Yu, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine*, 29(6):82–97, 2012.

S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997.

T. Hothorn and B. Lausen. Double-bagging: Combining classifiers by bootstrap aggregation. *Pattern Recognition*, 36(6):1303–1309, 2003. ISSN 00313203. . URL http://www.sciencedirect.com/science/article/pii/S0031320302001693.

Y. Hu, Y. Koren, and C. Volinsky. Collaborative filtering for implicit feedback datasets. In *2008 Eighth IEEE international conference on data mining*, pages 263–272. Ieee, 2008.

W. Hui and W. Gao. Stacked denoising autoencoders applied to clinical diagnose and classification. *Acta Astronomica Sinica*, 57:7–10, 2015. URL https://ui.adsabs.harvard.edu/{#}abs/2016AcASn..57..344Q/abstract.

IceCube Collaboration, M. G. Aartsen, K. Abraham, et al. First search for dark matter annihilations in the Earth with the IceCube Detector. *The European Physical Journal C*, 77(2):1–13, feb 2016. ISSN 14346052. . URL http://link.springer.com/10.1140/epjc/s10052-016-4582-yhttp://arxiv.org/abs/1609.01492.

J. J. M. In't Zand, J. Heise, J. M. Muller, et al. Discovery of SAX J1753.5-2349 and SAX J1806.5-2215: two X-ray bursters without detectable steady emission. *Nuclear Physics B Proceedings Supplements*, 69(1-3):228–231, January 1999. .

R. Jager, W. A. Mels, A. C. Brinkman, et al. The Wide Field Cameras onboard the BeppoSAX X-ray Astronomy Satellite. , 125:557–572, November 1997. .

G. James, D. Witten, T. Hastie, et al. *An Introduction to Statistical Learning*, volume 103 of *Springer Texts in Statistics*. Springer New York, New York, NY, 2013. ISBN 978-1-4614-7137-0. . URL http://link.springer.com/10.1007/978-1-4614-7138-7.

T. Jayasinghe, C. S. Kochanek, K. Z. Stanek, et al. The ASAS-SN catalogue of variable stars I: The Serendipitous Survey. , 477(3):3145–3163, July 2018. .

J. M. Jenkins, S. D. McCauliff, J. H. Catanzarite, et al. Likely Planet Candidates Identified by Machine Learning Applied to Four Years of Kepler Data. *AAS/Division for Planetary Sciences Meeting Abstracts*, 45, 2013. URL https://ui.adsabs.harvard.edu/{#}abs/2013DPS....4520406J/abstract.

J. H. Jeppesen, R. H. Jacobsen, F. Inceoglu, et al. A cloud detection algorithm for satellite imagery based on deep learning. *Remote sensing of environment*, 229:247–259, 2019.

Z. Jiang, D. Xu, and J. Liang. A deep reinforcement learning framework for the financial portfolio management problem. *arXiv preprint arXiv:1706.10059*, 2017.

E. Jones and J. Singal. A custom support vector machine analysis of the efficacy of galaxy shape information in photometric redshift estimation. *Astronomy & Astrophysics*, 600:A113, apr 2016. ISSN 0004-6361. . URL http://www.aanda.org/10.1051/0004-6361/201629558.

E. Jourdain and J. P. Roques. 2003-2019 Monitoring of the Crab Emission through INTEGRAL SPI, or Vice Versa. , 899(2):131, August 2020. .

A. S. Kaniovsky, V. A. Arefiev, N. L. Aleksandrovich, et al. Three hard X-ray transients: GRO J0422+32, GRS 1716-24, GRS 1009-45. Broad band observations by Roentgen-MIR-KVANT observatory. *Advances in Space Research*, 19(1):29–34, May 1997. .

L. Kaper, A. Van der Meer, and A. H. Tijani. High-mass x-ray binaries and ob runaway stars. In *International Astronomical Union Colloquium*, volume 191, pages 128–131. Cambridge University Press, 2004.

F. Karim, S. Majumdar, H. Darabi, et al. Lstm fully convolutional networks for time series classification. *IEEE access*, 6:1662–1669, 2017.

K. Kashinath, M. Mustafa, A. Albert, et al. Physics-informed machine learning: case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A*, 379(2194):20200093, 2021.

J. Korstanje. *Advanced forecasting with Python*. Springer, 2021.

S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas. Supervised Machine Learning : A Review of Classification Techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, pages 3–4, 2007. ISSN 09226389. . URL https://books.google.co.uk/books?hl=en{&}lr={&}id=vLiTXDHr{_}sYC{&}oi=fnd{&}pg=PA3{&}dq=supervised+machine+learning{&}ots=CXsAAwXDnq{&}sig=Qrkk{_}HGDVjavHOAXwEtxGXWMWOo.

P. Kretschmar, F. Fürst, L. Sidoli, et al. Advances in Understanding High-Mass X-ray Binaries with INTEGRALand Future Directions. , 86:101546, December 2019. .

R. Krivonos, M. Revnivtsev, A. Lutovinov, et al. INTEGRAL/IBIS all-sky survey in hard X-rays. , 475(2):775–784, November 2007. .

R. Krivonos, S. Tsygankov, M. Revnivtsev, et al. Integral/ibis 7-year all-sky hard x-ray survey-ii. catalog of sources. *Astronomy & Astrophysics*, 523:A61, 2010.

R. Krivonos, S. Tsygankov, A. Lutovinov, et al. Integral/ibis nine-year galactic hard x-ray survey. *Astronomy & Astrophysics*, 545:A27, 2012.

R. Krivonos, S. Tsygankov, A. Lutovinov, et al. INTEGRAL 11-year hard X-ray survey above 100 keV. , 448(4):3766–3774, April 2015. .

R. A. Krivonos, S. Y. Sazonov, E. A. Kuznetsova, et al. INTEGRAL/IBIS 17-yr hard X-ray all-sky survey. , 510(4):4796–4807, March 2022. .

E. Kuminski and L. Shamir. a Computer-Generated Visual Morphology Catalog of 3,000,000 Sdss Galaxies. *The Astrophysical Journal Supplement Series*, 223(2):20, apr 2016. ISSN 1538-4365. . URL http://stacks.iop.org/0067-0049/223/i=2/a=20?key=crossref.fb3e5bbda07ce60d4360adee26a49ee6.

A. Kurcz, M. Bilicki, A. Solarz, et al. Towards automatic classification of all WISE sources. *Astronomy & Astrophysics*, 592:A25, aug 2016. ISSN 0004-6361. . URL http://www.aanda.org/10.1051/0004-6361/201628142http://arxiv.org/abs/1604.04229{%}5Cnhttp://dx.doi.org/10.1051/0004-6361/201628142{%}5Cnhttp://www.aanda.org/10.1051/0004-6361/201628142.

E. Kuulkers, J. in't Zand, J. Homan, et al. X-ray spectral and timing properties of the 2001 superburst of 4U 1636-536. In Philip Kaaret, Frederick K. Lamb, and Jean H. Swank, editors, *X-ray Timing 2003: Rossi and Beyond*, volume 714 of *American Institute of Physics Conference Series*, pages 257–260, July 2004. .

E. Kuulkers, S. E. Shaw, A. Paizis, et al. The INTEGRAL Galactic bulge monitoring program: the first 1.5 years. , 466(2):595–618, May 2007. .

I. Kyrychenko and S. Kolomiyets. K-means method for clustering parameters of meteor data orbits. *43rd COSPAR Scientific Assembly. Held 28 January-4 February*, 43: 302, 2021.

C. Labanti, G. Di Cocco, G. Ferro, et al. The Ibis-Picsit detector onboard Integral. , 411: L149–L152, November 2003. .

F. K. Lamb. Knowledge of Neutron Stars from X-Ray Observations. In Michael D. Papagiannis, editor, *Eighth Texas Symposium on Relativistic Astrophysics*, volume 302, page 482, December 1977. .

F. Lebrun, J. P. Leray, P. Lavocat, et al. ISGRI: The INTEGRAL Soft Gamma-Ray Imager. , 411:L141–L148, November 2003. .

Y. LeCun, D. Touresky, G. Hinton, et al. A theoretical framework for back-propagation. In *Proceedings of the 1988 connectionist models summer school*, volume 1, pages 21–28, 1988.

Y. LeCun, Y. Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.

Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

J. Lee, O. Jung, Y. Lee, O. Kim, et al. A comparison and interpretation of machine learning algorithm for the prediction of online purchase conversion. *Journal of Theoretical and Applied Electronic Commerce Research*, 16(5):1472–1491, 2021.

K. D. Leka, K. Dissauer, G. Barnes, et al. Properties of Flare-imminent versus Flare-quiet Active Regions from the Chromosphere through the Corona. II. Nonparametric Discriminant Analysis Results from the NWRA Classification Infrastructure (NCI). , 942(2):84, January 2023. .

V. A. Lepingwell, A. J. Bird, and S. R. Gunn. New approaches for faint source detection in hard X-ray surveys. *Monthly Notices of the Royal Astronomical Society*, 510 (3):4031–4039, March 2022. .

A. M. Levine, F. L. Lang, W. H. G. Lewin, et al. The HEAO1 A-4 catalog of high-energy X-ray sources. , 54:581–617, April 1984. .

A. M. Levine, H. Bradt, W. Cui, et al. First Results from the All-Sky Monitor on the Rossi X-Ray Timing Explorer. , 469:L33, September 1996. .

W. H. G. Lewin, E. P. J. van den Heuvel, and J. van Paradijs. *X-ray Binaries*, volume 26. Cambridge University Press, 1997.

R. Li, H.-N. Wang, H. He, et al. Support Vector Machine combined with K-Nearest Neighbors for Solar Flare Forecasting. *Chinese Journal of Astronomy and Astrophysics*, 7(3):441–447, jun 2007. ISSN 1009-9271. . URL http://stacks.iop.org/1009-9271/7/i=3/a=15?key=crossref.b964824415a641ea943420598244bac1.

A. Liaw and M. Wiener. Classification and Regression by randomForest. *R news*, 2 (December):18–22, 2002. ISSN 16093631. . URL http://ai2-s2-pdfs.s3.amazonaws.com/6e63/3b41d93051375ef9135102d54fa097dc8cf8.pdf.

C. J. Lintott, K. Schawinski, A. Slosar, et al. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. , 389(3):1179–1189, September 2008. .

H. Liu, J. Liang, Y. Liu, et al. A review of data-driven building energy prediction. *Buildings*, 13(2):532, 2023.

N. T Liu, J. B. Holcomb, C. E Wade, et al. Utility of vital signs, heart rate variability and complexity, and machine learning for identifying the need for lifesaving interventions in trauma patients. *Shock*, 42(2):108–114, 2014.

Q. Z. Liu, J. Van Paradijs, and E. P. J. Van Den Heuvel. A catalogue of low-mass x-ray binaries. *Astronomy & Astrophysics*, 368(3):1021–1054, 2001.

S. Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28 (2):129–137, 1982.

K. K. Lo, S. Farrell, T. Murphy, et al. Automatic classification of time-variable X-ray sources. *arXiv.org*, astro-ph.I(1):20, apr 2014. ISSN 0004-637X. . URL http://stacks.iop.org/0004-637X/786/i=1/a=20?key=crossref.5e26019539d411614a486b25f2440881http://stacks.iop.org/0004-637X/786/i=1/a=20?key=crossref.5e26019539d411614a486b25f2440881{%}5Cnpapers3://publication/doi/10.1088/0004-637X/786/1/20.

M. Lochner, J. D. McEwen, H. V. Peiris, et al. PHOTOMETRIC SUPERNOVA CLASSIFICATION WITH MACHINE LEARNING. *The Astrophysical Journal Supplement Series*, 225(2):31, aug 2016. ISSN 1538-4365. . URL http://stacks.iop.org/0067-0049/225/i=2/a=31?key=crossref.e8c866186c9c1b4aec03a7ecf12ccd9bhttp://arxiv.org/abs/1603.00882.

C. H. A. Logan and S. Fotopoulou. Unsupervised star, galaxy, qso classification-application of hdbscan. *Astronomy & Astrophysics*, 633:A154, 2020.

LSST Science Collaboration, P. A. Abell, J. Allison, et al. LSST Science Book, Version 2.0. *arXiv.org*, astro-ph.I(November):201, dec 2009. ISSN 1740-9713. . URL http://adsabs.harvard.edu/cgi-bin/nph-data{_}query?bibcode=2009arXiv0912.0201L{&}link{_}type=ABSTRACT{%}5Cnpapers2://publication/uuid/573030D0-721D-4390-8DF5-437EE75384A0{%}5Cnhttp://arxiv.org/abs/0912.0201.

Y. Lu and F. M. Salem. Simplified gating in long short-term memory (lstm) recurrent neural networks. In *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*, pages 1601–1604. IEEE, 2017.

K. Luther, K. Boone, B. Hayden, et al. Automated Artifact Rejection for Transient Identification in WFC3 IR Image Subtractions. *American Astronomical Society Meeting Abstracts*, 227, 2016. URL https://ui.adsabs.harvard.edu/{#}abs/2016AAS...22723703L/abstract.

J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proc. 5th Berkeley Symposium on Math., Stat., and Prob*, page 281, 1965.

A. A. Mahabal, S. G. Djorgovski, A. J. Drake, et al. Towards the Automated Classification of Variable Objects and Transients. *American Astronomical Society Meeting Abstracts #213*, 213, 2009. URL https://ui.adsabs.harvard.edu/{#}abs/2009AAS...21342706M/abstract.

G. S. Mahalakshmi, B. Swadesh, R. R. V. Aswin, et al. Classification and feature prediction of star, galaxies, quasars, and galaxy morphologies using machine learning. 2022.

K. L. Malanchev, M. V. Pruzhinskaya, V. S. Korolev, et al. Anomaly detection in the Zwicky Transient Facility DR3. , 502(4):5147–5175, April 2021. .

K. Małek, A. Solarz, A. Pollo, et al. The VIMOS Public Extragalactic Redshift Survey (VIPERS). A support vector machine classification of galaxies, stars, and AGNs. , 557:A16, September 2013. .

A. Manousakis and R. Walter. X-ray wind tomography of the highly absorbed HMXB IGR J17252-3616. , 526:A62, February 2011. .

T. H. Markert, C. R. Canizares, G. W. Clark, et al. OSO-7 observations of high galactic latitude X-ray sources. , 206:265–272, May 1976. .

T. H. Markert, C. R. Canizares, G. W. Clark, et al. Observations of galactic X-ray sources by OSO-7. , 218:801–814, December 1977. .

T. H. Markert, P. F. Winkler, F. N. Laird, et al. The MIT/OSO 7 catalog of X-ray sources: intensities, spectra, and long-term variability. , 39:573–632, April 1979. .

J. R. Martínez-Galarza, F. B. Bianco, D. Crake, et al. A method for finding anomalous astronomical light curves and their analogues. , 508(4):5734–5756, December 2021. .

S. Mereghetti, D. Götz, J. Borkowski, et al. The INTEGRAL Burst Alert System. , 411: L291–L297, November 2003. .

H. Meusinger, P. Schalldach, R. D. Scholz, et al. Unusual quasars from the Sloan Digital Sky Survey selected by means of Kohonen self-organising maps. , 541:A77, May 2012. .

R. Miotto, L. Li, B. A. Kidd, et al. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Scientific reports*, 6 (1):1–10, 2016.

D. Mistry, C. M. Copperwheat, M. J. Darnley, et al. Machine learning-based search for cataclysmic variables within Gaia Science Alerts. , 517(3):3362–3376, December 2022. .

V. Mnih, K. Kavukcuoglu, D. Silver, et al. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.

S. Mohan, C. Thirumalai, and G. Srivastava. Effective heart disease prediction using hybrid machine learning techniques. *IEEE access*, 7:81542–81554, 2019.

S. V. Molkov, A. M. Cherepashchuk, A. A. Lutovinov, et al. A Hard X-ray Survey of the Sagittarius Arm Tangent with the IBIS Telescope of the INTEGRAL Observatory: A Catalog of Sources. *Astronomy Letters*, 30:534–539, August 2004. .

A. Möller, V. Ruhlmann-Kleider, C. Leloup, et al. Photometric classification of type Ia supernovae in the SuperNova Legacy Survey with supervised learning. page 27, aug 2016. ISSN 14757516. . URL http://arxiv.org/abs/1608.05423http://dx.doi.org/10.1088/1475-7516/2016/12/008.

K. Mukai. X-Ray Emissions from Accreting White Dwarfs: A Review. , 129(976): 062001, June 2017. .

N. Mukund, S. Abraham, S. Kandhasamy, et al. Transient Classification in LIGO data using Difference Boosting Neural Network. sep 2016. URL http://arxiv.org/abs/1609.07259.

Elahesadat Naghib, Peter Yoachim, Robert J. Vanderbei, Andrew J. Connolly, and R. Lynne Jones. A Framework for Telescope Schedulers: With Applications to the Large Synoptic Survey Telescope. , 157(4):151, April 2019. .

J. Nagi, F. Ducatelle, G. A. Di Caro, et al. Max-pooling convolutional neural networks for vision-based hand gesture recognition. In *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pages 342–347. IEEE, 2011.

V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Icml*, 2010.

L. Natalucci, M. Fiocchi, A. Bazzano, et al. High Energy Spectral Evolution of V404 Cygni during the 2015 June Outburst as Observed by INTEGRAL. , 813(1):L21, November 2015. .

I. Negueruela. On the nature of be/x-ray binaries. *Arxiv preprint astro-ph/9807158*, 1998.

I. Negueruela and M. J Coe. The population of massive x-ray binaries-i. the large magellanic cloud. *Astronomy & Astrophysics*, 385(2):517–532, 2002.

T. Nguyen, A. Basu, and T. Budavári. Globally Optimal and Scalable N-way Matching of Astronomy Catalogs. , 163(6):296, June 2022. .

K. Oh, M. Koss, C. B. Markwardt, et al. The 105-month swift-bat all-sky hard x-ray survey. *The Astrophysical Journal Supplement Series*, 235(1):4, 2018.

I. Ordovás-Pascual and J. Sánchez Almeida. A fast version of the k-means classification algorithm for astronomical applications. *Astronomy & Astrophysics*, 565:A53, 2014.

H. Oyaizu, M. Lima, C. E. Cunha, et al. Photometric Redshift Error Estimators. *The Astrophysical Journal*, 689(2):709–720, dec 2008. ISSN 0004-637X. . URL http://stacks.iop.org/0004-637X/689/i=2/a=709.

R. Pascanu, C. Gulcehre, K. Cho, et al. How to construct deep recurrent neural networks. *arXiv preprint arXiv:1312.6026*, 2013.

J. Pasquet-Itam and J. Pasquet. Deep learning approach for classifying, detecting and predicting photometric redshifts of quasars in the Sloan Digital Sky Survey stripe 82. , 611:A97, April 2018. .

J. Paul, P. Mandrou, J. Ballet, et al. SIGMA: The hard X-ray and soft gamma-ray telescope on board the GRANAT space observatory. *Advances in Space Research*, 11 (8):289–302, January 1991. .

M. N. Pavlinsky, S. A. Grebenev, and R. A. Sunyaev. X-Ray Images of the Galactic Center Obtained with Art-P/Granat: Discovery of New Sources, Variability of Persistent Sources, and Localization of X-Ray Bursters. , 425:110, April 1994. .

B. M. Pavlyshenko. Machine-learning models for sales time series forecasting. *Data*, 4 (1):15, 2019.

N. B. Peng, Y. X. Zhang, and Yo. H. Zhao. A SVM-kNN method for quasar-star classification. *Science China: Physics, Mechanics and Astronomy*, 56(6):1227–1234, jun 2013. ISSN 16747348. . URL http://link.springer.com/10.1007/s11433-013-5083-8.

L. Piro, L. Scarsi, and R. C. Butler. SAX: the wideband mission for x-ray astronomy. In Silvano Fineschi, editor, *X-Ray and EUV/FUV Spectroscopy and Polarimetry*, volume 2517 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 169–181, October 1995. .

D. Psaltis and C. Norman. On the origin of quasi-periodic oscillations and broad-band noise in accreting neutron stars and black holes. *Arxiv preprint astro-ph/0001391*, 2000.

J. Quirola-Vásquez, F. E. Bauer, V. V. Dwarkadas, et al. The exceptional X-ray evolution of SN 1996cr in high resolution. , 490(4):4536–4564, December 2019. .

A. Raboonik, H. Safari, N. Alipour, et al. PREDICTION OF SOLAR FLARES USING UNIQUE SIGNATURES OF MAGNETIC FIELD IMAGES. *The Astrophysical Journal*, 834(1):11, dec 2016. ISSN 1538-4357. . URL http://stacks.iop.org/0004-637X/834/i=1/a=11?key=crossref.9063074623075ee047a40e03d51eed0a.

I. Reis, D. Baron, and S. Shahaf. Probabilistic random forest: A machine learning algorithm for noisy data sets. *The Astronomical Journal*, 157:16, 12 2018. .

Itamar Reis, Dovi Poznanski, Dalya Baron, Gail Zasowski, and Sahar Shahaf. Detecting outliers and learning complex structures with large spectroscopic surveys - a case study with APOGEE stars. , 476(2):2117–2136, May 2018. .

S. Rendle. Factorization machines. In *2010 IEEE International conference on data mining*, pages 995–1000. IEEE, 2010.

D. Rengasamy, M. Jafari, B. Rothwell, et al. Deep learning with dynamically weighted loss function for sensor-based prognostics and health management. *Sensors*, 20(3): 723, 2020.

M. G. Revnivtsev, R. A. Sunyaev, D. A. Varshalovich, et al. A hard x-ray survey of the galactic-center region with the ibis telescope of the integral observatory: a catalog of sources. *Astronomy Letters*, 30(6):382–389, 2004.

M. G. Revnivtsev, S. Y. Sazonov, S. V. Molkov, et al. Hard X-ray survey of the Galactic plane region in Crux: A catalog of sources. *Astronomy Letters*, 32(3):145–148, March 2006. .

J. W. Richards, D. Homrighausen, P. E. Freeman, et al. Semi-supervised learning for photometric supernova classification. *Monthly Notices of the Royal Astronomical Society*, 419(2):1121–1135, jan 2012. ISSN 00358711. . URL https://academic.oup.com/mnras/article-lookup/doi/10.1111/j.1365-2966.2011.19768.x.

B. D. Ripley. Pattern Recognition and Neural Networks. *Technometrics*, 39(2):233, 1996. ISSN 00401706. . URL http://www.jstor.org/stable/1270922?origin=crossref.

P. W. A. Roming, T. E. Kennedy, K. O. Mason, et al. The Swift Ultra-Violet/Optical Telescope. , 120(3-4):95–142, October 2005. .

S. R. Rosen, N. A. Webb, M. G. Watson, et al. The xmm-newton serendipitous survey-vii. the third xmm-newton serendipitous source catalogue. *Astronomy & Astrophysics*, 590:A1, 2016.

N. Rouf, M. B. Malik, T. Arif, et al. Stock market prediction using machine learning techniques: a decade survey on methodologies, recent developments, and future directions. *Electronics*, 10(21):2717, 2021.

D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

N. Sahakyan, V. Vardanyan, and M. Khachatryan. Gradient boosting decision trees classification of blazars of uncertain type in the fourth Fermi-LAT catalogue. , 519 (2):3000–3010, February 2023. .

D. Salvetti. Classifying Unidentified Gamma-ray Sources. *ArXiv e-prints*, mar 2016. URL http://arxiv.org/abs/1603.00231.

A. L. Samuel. Some studies in machine learning using the game of checkers. II—Recent progress. *Annual Review in Automatic Programming*, 6(November):1–36, 1969. ISSN 00664138. . URL http://researcher.watson.ibm.com/researcher/files/us-beygel/samuel-checkers.pdf.

J. Sánchez Almeida, J. A. L Aguerri, C. Munoz-Tunón, et al. Automatic unsupervised classification of all sloan digital sky survey data release 7 galaxy spectra. *The Astrophysical Journal*, 714(1):487, 2010.

C. L. Sarazin. X-ray emission from clusters of galaxies. *Reviews of Modern Physics*, 58 (1):1, 1986.

S. Scaringi, A. J. Bird, D. J. Clark, et al. ISINA: INTEGRAL Source Identification Network Algorithm. *Monthly Notices of the Royal Astronomical Society*, 390(4): 1339–1348, 2008. ISSN 00358711. . URL http://onlinelibrary.wiley.com/store/10.1111/j.1365-2966.2008.13765.x/asset/j.1365-2966.2008.13765.x.pdf;jsessionid=43432E6EB5DFD5E346813CCF66F18FE3.f04t01?v=1{&}t=j1roigyd{&}s=cbae2670164aa12b2af83e3f93e78962ff158b26.

S. Scaringi, A. J. Bird, A. B. Hill, et al. A new determination of the integral/ibis point source location accuracy. *Astronomy & Astrophysics*, 516:A75, 2010.

A. M. Schäfer, S. Udluft, and H. G. Zimmermann. Learning long term dependencies with recurrent neural networks. In *Artificial Neural Networks–ICANN 2006: 16th International Conference, Athens, Greece, September 10-14, 2006. Proceedings, Part I 16*, pages 71–80. Springer, 2006.

N. Schanche, A. Collier Cameron, G. Hébrard, et al. Machine-learning approaches to exoplanet transit detection and candidate validation in wide-field ground-based surveys. , 483(4):5534–5547, March 2019. .

J. Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61: 85–117, 2015.

Bernhard. Schoölkopf and Alexander J. Smola. *Learning with kernels : support vector machines, regularization, optimization, and beyond*. MIT Press, 2002. ISBN 0262194759.

URL https://books.google.co.uk/books?hl=en{&}lr={&}id=y8ORL3DWt4sC{&}oi=fnd{&}pg=PR13{&}dq=kernel+trick+support+vector+machines{&}ots=bKAZduO3FF{&}sig=s8Nh0cSS83kayvY{_}0uL-k6{_}arKc{#}v=onepage{&}q=kerneltricksupportvectormachines{&}f=false.

L. J. Schultz, M. S. Wallace, M. C. Galassi, et al. Hybrid coded aperture and Compton imaging using an active mask. *Nuclear Instruments and Methods in Physics Research A*, 608(2):267–274, September 2009. .

C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, et al. RUSBoost: A hybrid approach to alleviating class imbalance. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, 40(1):185–197, 2010. ISSN 10834427. . URL http://sci2s.ugr.es/keel/pdf/algorithm/articulo/2010-IEEETSMCpartA-RUSBoostAHybridApproachtoAlleviatingClassImbalance.pdf.

S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, et al. Stock price prediction using lstm, rnn and cnn-sliding window model. In *2017 international conference on advances in computing, communications and informatics (icacci)*, pages 1643–1647. IEEE, 2017.

I. Sevilla-Noarbe and P. Etayo-Sotos. Effect of training characteristics on object classification: An application using Boosted Decision Trees. *Astronomy and Computing*, 11:64–72, jun 2015. ISSN 22131337. . URL http://linkinghub.elsevier.com/retrieve/pii/S2213133715000347.

V. Sguera, A. Bazzano, A. J. Bird, et al. Unveiling supergiant fast x-ray transient sources with integral. *The Astrophysical Journal*, 646(1):452, 2006.

D. Shen, G. Wu, and H.-I. Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017.

N. Shinde, S. Chandana, S. A. Patil, et al. Stacked lstm based wafer classification. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 5786–5790. IEEE, 2021.

D. Silver, A. Huang, C. J. Maddison, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016.

S. Singh, S. Hussain, and Mohammad A. Bazaz. Short term load forecasting using artificial neural network. In *2017 Fourth International Conference on Image Information Processing (ICIIP)*, pages 1–5. IEEE, 2017.

G. K. Skinner, A. P. Willmore, C. J. Eyles, et al. Hard X-ray images of the galactic centre. , 330(6148):544–547, December 1987. .

R. Socher, A. Perelygin, J. Wu, et al. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642, 2013.

Y.-J. Son, H.-G. Kim, E.-H. Kim, et al. Application of support vector machine for prediction of medication adherence in heart failure patients. *Healthcare informatics research*, 16(4):253–259, 2010.

R. Sunyaev. Recent Gamma-Ray Observations of the Galactic Center. In *Bulletin of the American Astronomical Society*, volume 23, page 1395, September 1991.

R. Sunyaev, E. Churazov, M. R. Gilfanov, et al. X-Ray Observations of the Large Magellanic Cloud Field by the TTM Instrument on Board the KVANT Module - 1988NOV - 1989JUN. In J. Hunt and B. Battrick, editors, *Two Topics in X-Ray Astronomy, Volume 1: X Ray Binaries. Volume 2: AGN and the X Ray Background*, volume 1 of *ESA Special Publication*, page 633, November 1989.

R. Sunyaev, S. Grebenev, A. Kaniovsky, et al. Hard x-rays from supernova 1987A: results of Mir-Kvant and Granat in 1987-1990 and expectations. In Philippe Durouchoux and Nikos Prantzos, editors, *Gamma-Ray Line Astrophysics*, volume 232 of *American Institute of Physics Conference Series*, pages 211–217, August 1991a. .

R. Sunyaev, P. Mandrou, and J. Paul. Galactic Center. , 5245:2, April 1991b.

R. A. Sunyaev, E. Churazov, V. Efremov, et al. Highlights from the KVANT mission. *Advances in Space Research*, 10(2):41–46, January 1990. .

I. Sutskever, O. Vinyals, and Q. V. Le. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27, 2014.

R. S. Sutton, A. G. Barto, et al. *Introduction to reinforcement learning*, volume 135. MIT press Cambridge, 1998.

M. Süveges, F. Barblan, I. Lecoeur-Taïbi, et al. Gaia eclipsing binary and multiple systems. Supervised classification and self-organizing maps. feb 2017. URL http://arxiv.org/abs/1702.06296.

J. H. Swank, R. H. Becker, E. A. Boldt, et al. Spectral evolution of a long x-ray burst. *The Astrophysical Journal*, 212:L73–L76, 1977.

Y. Taigman, M. Yang, M. Ranzato, et al. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1701–1708, 2014.

D. Tamayo, M. Ali-Dib, R. Cloutier, et al. Machine Learning Algorithms For Predicting the Instability Timescales of Compact Planetary Systems. *AAS/Division for Planetary Sciences Meeting Abstracts*, 48, 2016. URL https://ui.adsabs.harvard.edu/{#}abs/2016DPS....4830907T/abstract.

H. Tranin, O. Godet, N. Webb, et al. Probabilistic classification of X-ray sources applied to Swift-XRT and XMM-Newton catalogs. , 657:A138, January 2022. .

C. Tsuji. Exchange rate forecasting via a machine learning approach. *iBusiness*, 14(3): 119–126, 2022.

S. Turner, L. S. Kelvin, I. K. Baldry, et al. Reproducible k-means clustering in galaxy feature data from the gama survey. *Monthly Notices of the Royal Astronomical Society*, 482(1):126–150, 2019.

P. Ubertini, F. Lebrun, G. Di Cocco, et al. IBIS: The Imager on-board INTEGRAL. , 411: L131–L139, November 2003. .

J. van den Eijnden, N. Degenaar, T. D. Russell, et al. Radio monitoring of transient Be/X-ray binaries and the inflow-outflow coupling of strongly magnetized accreting neutron stars. , 516(4):4844–4861, November 2022. .

D. Varmedja, M. Karanovic, S. Sladojevic, et al. Credit card fraud detection-machine learning methods. In *2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)*, pages 1–5. IEEE, 2019.

E. C. Vasconcellos, R. R. de Carvalho, R. R. Gal, et al. Decision Tree Classifiers for Star/Galaxy Separation. , 141(6):189, June 2011. .

A. Vaswani, N. Shazeer, N. Parmar, et al. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

I. B. Vavilova, D. V. Dobrycheva, M. Y. Vasylenko, et al. Machine learning technique for morphological classification of galaxies from the SDSS. I. Photometry-based approach. , 648:A122, April 2021. .

S. Vinciguerra, M. Drago, G. A. Prodi, et al. Enhancing the significance of gravitational wave bursts through signal classification. *Classical and Quantum Gravity*, 34(9):1–22, may 2017. ISSN 0264-9381. . URL http://stacks.iop.org/0264-9381/34/i=9/a=094003?key=crossref.1dd4e3625f5348861ac8f9431bc2275c.

K. L. Wagstaff and V. G Laidler. Making the most of missing values: Object clustering with partial data in astronomy. In *Astronomical data analysis software and systems XIV*, volume 347, page 172, 2005.

C. Wang, Y. Bai, C. López-Sanjuan, et al. J-PLUS: Support vector machine applied to STAR-GALAXY-QSO classification. , 659:A144, March 2022. .

C. J. C. H. Watkins and P. Dayan. Q-learning. *Machine learning*, 8:279–292, 1992.

K. Weiss, T. M. Khoshgoftaar, and D. Wang. A survey of transfer learning. *Journal of Big data*, 3(1):1–40, 2016.

M. J. Westmore. *Dynamic flat-fielding of BATSE data and the BATSE all sky survey*. PhD thesis, University of Southampton, UK, January 2002.

N. E. White, F. Nagase, and A. N. Parmar. The properties of X-ray binaries. In *X-ray Binaries*, pages 1–57, January 1995.

J. Wilms, K. Pottschmidt, M. A. Nowak, et al. INTEGRAL/RXTE Observations of Cygnus X-1. In Philip Kaaret, Frederick K. Lamb, and Jean H. Swank, editors, *X-ray Timing 2003: Rossi and Beyond*, volume 714 of *American Institute of Physics Conference Series*, pages 116–119, July 2004. .

C. Winkler. The INTEGRAL Core Observing Programme. In A. Gimenez, V. Reglero, and C. Winkler, editors, *Exploring the Gamma-Ray Universe*, volume 459 of *ESA Special Publication*, pages 471–478, September 2001.

C. Wolf, A. S. Johnson, M. Bilicki, et al. The 2-degree Field Lensing Survey: photometric redshifts from a large new training sample to r $<$ 19.5. *Mon. Not. R. Astron. Soc*, 000(0000):0–0, apr 2016. ISSN 0035-8711. . URL https://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/stw3151.

C. Wu, O. I. Wong, L. Rudnick, et al. Radio Galaxy Zoo: CLARAN - a deep learning classifier for radio morphologies. , 482(1):1211–1230, January 2019. .

R. Yamashita, M. Nishio, R. K. G. Do, et al. Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9:611–629, 2018.

D. G. York, J. Adelman, Jr. Anderson, J. E., et al. The Sloan Digital Sky Survey: Technical Summary. , 120(3):1579–1587, September 2000. .

T. Yue and H. Wang. Deep learning for genomics: A concise overview. *arXiv preprint arXiv:1802.00810*, 2018.

V. Zabalza. naima: a python package for inference of relativistic particle energy distributions from observed nonthermal spectra. *Proc. of International Cosmic Ray Conference 2015*, page 922, 2015.

T. Zaidi and G. Narayan. Supernova Photometric Lightcurve Classification. *American Astronomical Society, AAS Meeting #227, id.348.20*, 227, 2016. URL http://adsabs.harvard.edu/abs/2016AAS...22734820Z.

Z. Zhang and M. R. Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. In *32nd Conference on Neural Information Processing Systems (NeurIPS)*, 2018.

X. Zhu and A. B. Goldberg. Introduction to Semi-Supervised Learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 3(1):1–130, jan 2009. ISSN 1939-4608. . URL http://www.morganclaypool.com/doi/abs/10.2200/S00196ED1V01Y200906AIM006.