University of Heidelberg

Department of Physics and Astronomy

Faculty of Physics and Astronomy

Master's Thesis

From Light Curves to Labels: Machine Learning in Microlensing

Author: Kunal Bhatia

Matriculation Number: 3770453

First Supervisor: Prof. Dr. Joachim Wambsganß

Second Supervisor: Dr. Yiannis Tsapras

Submission Date: October 27, 2025

Declaration of Authorship

I hereby declare that this thesis is my own work and that I have not used any sources or aids other than those stated. I have marked as such all passages taken word-for-word or in content from other works. Furthermore, I assure that the electronic version submitted corresponds completely in content and wording to the printed version of my thesis. I agree that this electronic version may be checked for plagiarism using plagiarism detection software within the university.

Heidelberg, October 27, 2025
Kunal Bhatia

Abstract

The upcoming era of large-scale time-domain surveys—notably the Vera C. Rubin Observatory's Legacy Survey of Space and Time (*LSST*) and the Nancy Grace *Roman* Space Telescope—will increase gravitational microlensing event detection rates from ∼2,000 to over 20,000 events annually. This dramatic increase necessitates automated classification systems capable of distinguishing binary lens events (planetary systems, stellar binaries) from simple Point-Source Point-Lens (PSPL) events. The challenge lies in early-time classification: binary events masquerade as PSPL during their initial phases, yet timely identification is critical for triggering follow-up observations of transient planetary features.

This thesis presents a machine learning framework for automated binary microlensing classification using one-dimensional Convolutional Neural Networks (CNNs) with TimeDistributed architecture. We generate one million synthetic light curves spanning realistic parameter ranges, deliberately sampling binary configurations that produce distinctive caustic-crossing signatures. The dataset incorporates adaptable observational effects including survey-specific cadence patterns and photometric noise levels matching OGLE, *LSST*, and *Roman* characteristics.

The TimeDistributed CNN processes sequential temporal windows, enabling dynamic classification as observations accumulate rather than requiring complete light curves. We systematically evaluate performance across observation completeness levels, quantifying the confidence-timeliness trade-off essential for operational deployment. Comprehensive benchmarking against traditional χ^2 fitting establishes performance limits and identifies complementary strengths of each approach.

Results demonstrate that deep learning achieves high classification accuracy with partial light curves while maintaining computational efficiency suitable for real-time survey operations. This work provides actionable recommendations for *LSST* and *Roman* alert stream integration and delivers an open-source framework ready for deployment in next-generation microlensing surveys.

Contents

List of Figures

List of Tables

1. Introduction

1.1. Motivation and Background

Modern astronomy is increasingly defined by the scale and velocity of data acquisition, pushing the boundaries of computational capacity and observational efficiency. Wide-field time-domain surveys have transformed the field from targeted observation to automated, rapid data processing across millions of stellar sources. Among the most powerful techniques for detecting exoplanets and probing the structure of the Galaxy is gravitational microlensing [1, 2], a phenomenon predicted by Einstein's general relativity in which a massive foreground object (the lens) temporarily magnifies the light of a background star (the source) as they align along the line of sight to Earth [3].

The elegance of gravitational microlensing lies in its democratic nature: because the lensing effect depends only on mass and geometry rather than the luminosity of the lens, microlensing can detect planets around faint stars, free-floating planets unbound to any host, and stellar remnants that would remain invisible to traditional detection methods [4, 5]. The probability that any random star is sufficiently aligned to be lensed is extremely small—the optical depth toward the Galactic bulge is of order 10^{-6} [6]—so millions of stars must be continuously monitored to detect even a handful of events. Over the past three decades, ground-based surveys such as the Optical Gravitational Lensing Experiment (OGLE) and the Microlensing Observations in Astrophysics (MOA) have monitored dense stellar fields in the Galactic bulge at cadences of minutes to hours, discovering over 200 exoplanets and characterizing the demographics of planetary systems throughout the Galaxy [7–9].

Historically, the majority of observed microlensing events are well described by the Point-Source Point-Lens (PSPL) model [2], which assumes both the lens and source are point masses. These events produce characteristic, symmetric, achromatic light curves described by only three parameters: the Einstein crossing time, the impact parameter, and the time of maximum magnification. However, the most scientifically valuable events are those caused by binary lenses—two-body systems such as a star-planet pair or a binary star system. Binary lens events produce complex magnification patterns with caustic structures, multiple peaks, and asymmetries that depend sensitively on the lens geometry and the source trajectory [3, 10]. These deviations from the simple PSPL template are crucial for determining physical parameters such as planetary mass

ratios and projected separations.

The primary observational challenge lies in the temporal ambiguity problem: during the initial rise of a microlensing event, binary lens systems often masquerade as simple PSPL events, only revealing their true nature when the source approaches or crosses caustic features in the lens plane. This degeneracy persists until the event is sufficiently developed, leading to critical delays in follow-up observations necessary to fully characterize the binary system. Current surveys discover approximately 2,000 events per year [9]—a number that human experts and traditional modeling approaches can manage through careful monitoring. However, the upcoming data deluge from the Vera C. Rubin Observatory's Legacy Survey of Space and Time (*LSST*) will increase this rate tenfold, discovering an estimated 20,000 microlensing events annually [11, 12]. Simultaneously, the Nancy Grace *Roman* Space Telescope, with its dedicated Galactic Bulge Time-Domain Survey, aims to detect on the order of 2×10^4 microlensing events over its mission lifetime [13, 14], providing complementary space-based observations with superior photometric precision and continuous monitoring capabilities. This enormous increase in event detection rates mandates a complete re-evaluation of current classification strategies. The future of microlensing science, particularly in the statistical characterization of exoplanet populations and Galactic structure, depends critically on developing automated systems capable of early, accurate classification.

1.2. Research Problem

The fundamental challenge addressed by this thesis is the inefficiency and latency of classifying binary microlensing events in the era of large-scale surveys. Traditional classification methods, typically relying on χ^2 model fitting or manual inspection, suffer from several critical limitations. First, they are computationally intensive and scale poorly with the incoming data rates expected from *LSST* and *Roman*. Binary lens models have significantly more free parameters than PSPL models (typically seven or more versus three), and their complex parameter spaces contain numerous local minima, making optimization challenging and computationally expensive [15]. Fitting a single binary lens model to a light curve can require seconds to minutes per evaluation, and characterizing a single event can take hours of computation when accounting for multiple binary topologies and the need for global optimization. When multiplied across thousands of ongoing events that must be updated with each new observation, the computational burden becomes prohibitive for real-time classification during survey operations.

Second, and more critically, traditional methods require a high degree of observation completeness—meaning the light curve must be substantially developed, often past the peak magnification, to reveal the subtle features indicative of a binary system. The

PSPL model can accurately fit the initial brightening phase of many binary events that have not yet revealed their characteristic caustic features [16]. By the time these distinguishing features become apparent, the optimal window for additional observations may have passed, particularly for short-duration events or events discovered near their peaks. If a binary event is classified too late, critical high-magnification features may be missed, rendering the determination of planetary mass ratios and separations highly uncertain. Furthermore, late classification prevents timely triggering of global follow-up networks, which may have observational windows lasting only hours to days [17, 18], leading to suboptimal data acquisition and reduced scientific yield.

The impending arrival of *LSST* data exacerbates all of these challenges. The survey's unprecedented depth and cadence will discover events at earlier phases and in greater numbers than current surveys, while its wide field of view and multiple filters will generate heterogeneous light curves with varying sampling rates and photometric precision [19]. Existing classification infrastructure, already strained by current survey data rates, will be completely inadequate for *LSST*-era operations. The current reliance on human-in-the-loop systems and computationally expensive template fitting is not sustainable for the upcoming $\sim$10$\times$ increase in event detection rate.

This thesis addresses the research problem: *How can we design a machine-learning model that reliably distinguishes binary microlensing events from PSPL events using only partial, noisy light curves, enabling early classification and efficient resource allocation for follow-up observations?* This requires moving beyond traditional feature engineering toward a representation learning approach that can model the temporal dependencies of light curve data, process incomplete observations as they arrive, scale efficiently to tens of thousands of events, and provide reliable confidence estimates to guide operational decisions.

1.3. Research Questions

To make the research problem tractable and measurable, this thesis addresses the following specific research questions:

1. **Can a one-dimensional Convolutional Neural Network (1D CNN), specifically designed for time-series data, effectively distinguish between PSPL and binary microlensing events using incomplete light curves?** We investigate whether deep learning architectures can learn representations that capture the subtle deviations characteristic of binary lenses without explicit feature engineering.

2. **How does classification accuracy correlate with observation completeness, and at what minimum completeness level can high-confidence**

classification ($\geq 90\%$ accuracy) be achieved? We quantify the trade-off between timeliness and reliability by evaluating performance across observation fractions ranging from 10% to 100% of the full event duration, determining the earliest point at which reliable identification becomes possible.

3. **Does the proposed TimeDistributed architecture enhance the model's ability to extract time-varying features and provide a performance advantage over traditional machine learning approaches and standard single-shot CNN models?** We compare the TimeDistributed design, which processes sequential observations as they arrive, against baseline methods including feature-based random forests and standard CNN architectures.

4. **How sensitive is classification performance to survey cadence and photometric noise, and can a model trained on synthetic data generalize to real survey observations?** By varying sampling intervals and noise levels to mimic OGLE, MOA, and *LSST* characteristics, we evaluate robustness across different observing strategies. We test generalization by applying the CNN to real OGLE and MOA data and benchmark against traditional methods using metrics such as precision, recall, and area under the receiver operating characteristic curve (AUC).

These questions provide a structured framework for the experiments described in subsequent chapters. By answering them systematically, we aim to determine both the feasibility and the practical limitations of early microlensing classification in preparation for *LSST* and *Roman* operations.

1.4. Objectives

To address these research questions, this thesis pursues the following concrete objectives:

1. **Data Generation and Curation:** Generate a large-scale, high-fidelity synthetic training dataset consisting of over 100,000 simulated PSPL and binary microlensing events. The dataset will systematically vary key physical parameters including mass ratio, projected separation, impact parameter, and Einstein crossing time to ensure comprehensive coverage of the parameter space. Observational effects such as photometric noise, irregular sampling, blending, and finite-source effects will be incorporated to mimic the characteristics of OGLE, MOA, and *LSST* observations.

2. **Architecture Development:** Design and implement a novel 1D CNN with a TimeDistributed wrapper specifically tailored for sequential microlensing light

curve classification. This architecture will process light curve observations as they arrive in temporal windows, enabling dynamic classification based on the sequence history rather than a single aggregated feature vector. The design will include convolutional filters to extract local temporal patterns, TimeDistributed layers to maintain temporal information, and appropriate regularization techniques (dropout, batch normalization) to prevent overfitting.

3. **Systematic Performance Evaluation:** Evaluate the developed model's classification performance across multiple dimensions. We will compute accuracy, precision, recall, F_1-score, and AUC across a spectrum of observation completeness levels ranging from 10% to 100% of the full light curve duration. This systematic evaluation establishes the early detection capability of the model and quantifies the confidence-timeliness trade-off essential for operational deployment.

4. **Benchmarking and Validation:** Benchmark the TimeDistributed CNN framework against established traditional classification techniques (including χ^2 template fitting following standard practices) and modern machine learning baselines such as feature-based random forests, standard CNNs without TimeDistributed layers, and recurrent neural networks. We will conduct ablation studies to understand which architectural components contribute most significantly to performance and identify scenarios where each approach excels.

5. **Framework Finalization and Survey Readiness:** Develop a complete, production-ready classification system designed for deployment in real-time survey operations. This system will include data preprocessing pipelines that handle heterogeneous input formats, quality filtering, normalization schemes optimized for microlensing photometry, model serving infrastructure capable of processing thousands of events per night, and monitoring tools to detect performance degradation. The framework will be designed with modularity to facilitate adaptation to different survey characteristics and straightforward updates as improved models become available.

Together, these objectives aim to deliver both a comprehensive scientific evaluation of early microlensing classification and actionable insights for the design of future survey pipelines, with particular emphasis on *LSST* and *Roman* operations.

1.5. Contributions

This thesis makes the following contributions to automated microlensing classification:

1. **Large-Scale Synthetic Dataset with Strategic Parameter Sampling:** We generate one million high-fidelity synthetic light curves (500,000 PSPL + 500,000

binary events) using VBMicrolensing for accurate magnification calculations. Binary parameters are strategically sampled to produce events with multiple caustic crossings and distinctive morphological features, establishing an upper bound on classification performance under ideal conditions. The dataset incorporates survey-specific observational effects including adaptable cadence patterns and photometric noise levels calibrated to OGLE (ground-based, high-cadence), *LSST* (ground-based, multi-band, variable cadence), and *Roman* (space-based, continuous monitoring) characteristics.

2. **Systematic Multi-Dimensional Performance Evaluation:** We conduct comprehensive benchmarking across three critical dimensions: observation completeness (10%–100%), survey characteristics (three distinct cadence and noise regimes), and binary system parameters (separation, mass ratio, source size). This systematic evaluation quantifies the fundamental trade-offs between classification timeliness, accuracy, and observational constraints, providing concrete operational guidance for next-generation surveys.

3. **Comparative Analysis with Traditional Methods:** We implement and rigorously compare deep learning approaches against established χ^2 template fitting methods under controlled conditions. This direct comparison, conducted on identical datasets and evaluation protocols, identifies complementary strengths of each approach and scenarios where hybrid strategies may be optimal. We quantify computational efficiency differences critical for real-time survey operations processing thousands of events nightly.

4. **Survey-Ready Framework with Operational Recommendations:** We deliver a complete, production-tested pipeline encompassing data simulation, model training, and deployment infrastructure designed for integration with *LSST* and *Roman* alert streams. The framework includes specific recommendations for trigger thresholds, computational resource allocation, and performance monitoring tailored to each survey's operational constraints. All code and trained models are released open-source to facilitate community adoption and extension.

5. **Establishment of Performance Baselines for Early Classification:** Through evaluation on the largest microlensing classification dataset assembled to date, we establish definitive performance benchmarks for binary event identification at various completeness levels. These baselines provide reference standards for future algorithm development and enable informed decision-making about when machine learning approaches are preferable to traditional methods in operational contexts.

Collectively, these contributions bridge the gap between proof-of-concept machine learning demonstrations and operational deployment requirements, providing both scientific insights into classification limits and practical tools for next-generation microlensing surveys.

Collectively, these contributions advance the state of the art in automated microlensing classification and provide a blueprint for harnessing deep learning in the era of large-scale time-domain surveys. Beyond microlensing, the techniques developed here may be applicable to other astronomical transients where early identification is crucial, including supernovae, tidal disruption events, and fast radio burst optical counterparts.

1.6. Thesis Organization

The remainder of this thesis is organized to build progressively from theoretical foundations through implementation to results and implications. Each chapter is designed to be largely self-contained while contributing to the overall narrative:

??: Theoretical Background provides comprehensive context for understanding gravitational microlensing and the binary classification problem. We begin with a review of gravitational lensing theory, deriving the lens equation and describing the magnification patterns produced by point-mass lenses. We then introduce binary lens systems in detail, explaining the origin of caustic structures and critical curves, and discussing the degeneracies and parameter space complexities that make binary lens modeling challenging. The mathematical formulations for both PSPL and binary lens magnification are presented, with emphasis on the physical interpretation of light curve features. The chapter surveys the operational strategies of key microlensing surveys (OGLE, MOA, *Roman*, *LSST*) and their typical data products, and reviews the historical development of classification approaches including χ^2 fitting, Bayesian methods, and feature-based classifiers. We conclude by examining recent applications of machine learning to astronomical time-series analysis, with particular attention to deep learning methods for transient classification, and identify the existing gap in sequential modeling for early-time microlensing classification that motivates this work.

Chapter 3: Literature Review provides a comprehensive survey of related work across multiple domains. We review the history and current state of microlensing surveys, focusing on detection strategies, data pipelines, and classification challenges. Traditional modeling approaches, including χ^2 fitting and Bayesian inference methods, are examined in detail with emphasis on their computational costs and observational requirements. We survey machine learning applications in time-domain astronomy, covering both classical methods (random forests, support vector machines) and recent deep learning approaches (recurrent neural networks, convolutional architectures). Special attention is given to early classification problems in other astronomical domains, such as supernova typing

and transient identification, where similar temporal challenges arise. We conclude by identifying the specific gap that motivates this work: the absence of sequential modeling approaches designed explicitly for early-time microlensing classification.

Chapter 4: Methodology details the technical approach taken to achieve the research objectives. This chapter begins with a comprehensive description of the synthetic data generation pipeline, including the physical models used to simulate both PSPL and binary lens events (lens equations, magnification calculations, and parameter sampling), the parameter ranges and distributions chosen to represent realistic survey detections, and the systematic addition of observational effects such as photometric errors following realistic noise models, cadence variations mimicking OGLE and *LSST*, source blending, and finite-source effects. We then present the CNN architecture in detail, explaining the motivation for the TimeDistributed design, the specific layer configurations, activation functions, and regularization strategies employed. The preprocessing steps required to handle irregular sampling and normalize flux measurements are described. We document the training procedures including data augmentation techniques, optimization algorithms, learning rate schedules, and hyperparameter optimization methods. The chapter also describes the implementation of baseline methods for comparison, including χ^2 fitting procedures and feature-based random forest classifiers following established methodologies. We conclude with a detailed description of the evaluation framework, defining all performance metrics and validation strategies.

Chapter 5: Results presents the core experimental findings. We begin by characterizing the synthetic dataset, showing representative PSPL and binary light curves across the parameter space and validating that the simulations capture realistic event characteristics. We then present classification performance results, with particular focus on accuracy as a function of light curve completeness, demonstrating the early detection capability of the CNN approach through learning curves and performance plateaus. Confusion matrices reveal which types of events are most prone to misclassification, and we investigate how physical parameters (mass ratio, separation, impact parameter) correlate with classification difficulty. The systematic comparison with baseline methods quantifies differences in accuracy, precision, recall, F_1-score, and AUC across completeness levels, and examines computational cost trade-offs. We present ablation studies showing the contribution of the TimeDistributed architecture and other design choices. Attention weight visualizations and learned feature analyses provide insight into what patterns the CNN has learned to recognize. We test generalization by varying cadence and noise beyond training distributions, and demonstrate operational performance by applying the framework to a carefully curated set of real OGLE events with known classifications.

Chapter 6: Discussion interprets our findings in the broader context of microlensing

science and machine learning applications in astronomy. We discuss the astrophysical implications of early binary detection for survey strategy and follow-up planning, including concrete recommendations for resource allocation in *LSST* and *Roman* operations and coordination with ground-based follow-up networks. We analyze the limitations of our approach systematically, examining potential failure modes, assumptions in the training data that may not reflect all real-world scenarios (such as highly asymmetric source crossings or extreme mass ratios), and the challenge of maintaining performance as surveys evolve and new types of events are discovered. We discuss what the interpretability results reveal about the nature of binary lens signatures in incomplete light curves and how these insights relate to traditional physical understanding. The results are contextualized relative to related work in astronomical time-series classification, and we discuss the generalizability of the TimeDistributed architecture to other problems in time-domain astronomy. We identify specific opportunities for future improvements, including incorporation of multi-band color information, extension to triple and higher-order lens systems, integration with anomaly detection frameworks to identify unusual events, and potential for transfer learning to related transient phenomena.

Chapter 7: Conclusions and Future Work summarizes the key findings and their significance. We systematically review how our work addresses each research question posed in Section 1.3 and fulfills each stated objective from Section 1.4. We discuss the readiness of the CNN framework for operational deployment and provide specific, actionable recommendations for survey teams preparing for *LSST* and *Roman* operations. We outline several directions for future research, including extensions of the methodology to other microlensing analysis tasks such as real-time detection of finite-source effects and measurement of parallax signatures, applications of the TimeDistributed CNN architecture to other classes of astronomical transients including supernovae and tidal disruption events, potential for transfer learning from microlensing to gravitational wave electromagnetic counterpart classification, and strategies for continual learning as new data become available. We conclude with reflections on the evolving role of machine learning in modern astronomical surveys, emphasizing the importance of developing interpretable, robust, and scientifically validated deep learning tools that complement rather than replace physical understanding.

Appendices provide essential supporting technical material. Appendix A presents detailed derivations of microlensing magnification formulas for both PSPL and binary lenses, including edge cases and limiting behaviors. Appendix B documents the synthetic data generation code, parameter sampling procedures, and validation tests. Appendix C provides the complete CNN architecture specifications, layer configurations, and training hyperparameters for full reproducibility. Appendix D presents additional performance metrics, confusion matrices for parameter subsets, and supplementary figures characterizing model behavior. Appendix E discusses the computational infrastructure used for

training and evaluation, including hardware specifications, parallelization strategies, runtime performance measurements, and estimates of computational requirements for *LSST*-scale deployment.

This organizational structure ensures that readers can follow the logical progression from motivation through methods to conclusions, while also enabling experts to navigate directly to chapters of specific interest. Figures and tables throughout illustrate key concepts, simulation outcomes, and classifier performance. A comprehensive bibliography provides context within the broader literature of microlensing physics, exoplanet detection, and machine learning applications in astronomy.

2. Theoretical Background

This chapter establishes the theoretical foundations necessary for understanding gravitational microlensing and the machine learning techniques applied in this work. We begin with the general relativistic basis of gravitational lensing (Section 2.1), then describe the characteristic features of microlensing light curves (Section 2.2), discuss observational considerations (Section 2.3), and finally introduce the machine learning methodology (Section 2.4).

2.1. Gravitational Lensing

The phenomenon of light deflection by massive objects represents one of the most striking predictions of Einstein's general theory of relativity. In 1915, Einstein showed that spacetime curvature caused by mass would bend the path of light rays passing nearby. This prediction was famously confirmed during the solar eclipse of 1919, when Arthur Eddington observed the apparent displacement of stars near the Sun's limb [20]. The measured deflection of approximately 1.75 arcseconds agreed closely with Einstein's prediction and was twice the value expected from Newtonian gravity. This result, widely publicized in the press, made Einstein a household name and inaugurated the study of gravitational lensing.

While early observations focused on strong lensing effects by galaxies—producing multiple resolved images and dramatic Einstein rings—gravitational microlensing operates on much smaller angular scales. Rather than resolving separate images, microlensing by individual stars manifests as temporary brightness variations in background sources. The image separations are of order milli-arcseconds, too small to be resolved with current instruments; instead, microlensing is detected through time-varying brightness changes.

In this section, we develop the mathematical framework for gravitational lensing, beginning with the general relativistic foundation (Section 2.1.1), deriving the fundamental lens equation (Section 2.1.2), and explaining why stellar-mass lenses constitute the microlensing regime (Section 2.1.3).

2.1.1. General Relativistic Foundation

The deflection of light by gravitational fields emerges naturally from Einstein's field equations. In the weak field limit, appropriate for most astrophysical lensing scenarios, we can treat the problem using the Schwarzschild metric with small perturbations. General relativity predicts that light rays follow geodesics in curved spacetime. For a non-rotating, spherically symmetric mass M, the line element is given by the Schwarzschild metric. In the weak-field limit where $2GM/(c^2r) \ll 1$, one can treat the Newtonian potential $\Phi(r) = -GM/r$ as a perturbation to flat spacetime.

Consider a light ray passing at impact parameter b from a point mass M. The spacetime curvature induces a deflection of the light path by an angle given by

$$\alpha = \frac{4GM}{c^2 b}, \tag{2.1}$$

where G is the gravitational constant and c is the speed of light. This deflection angle is twice the value predicted by a purely Newtonian treatment, increases with the mass of the lens, and decreases with the impact parameter, reflecting the inverse relationship between gravitational influence and distance.

The derivation can be made explicit by expanding the Schwarzschild metric to first order in $GM/(c^2r)$ and computing the change in the transverse component of the photon's momentum. Alternatively, one may use Fermat's principle with an effective refractive index $n(r) \approx 1 - 2\Phi(r)/c^2$ to recover the same result. This provides an illuminating analogy: light bends in a gravitational field much as it would in a medium with spatially varying refractive index, with light rays preferentially following paths through regions of stronger gravitational potential.

For our purposes, the key insight is that Equation (2.1) depends only on the total mass and geometric configuration, not on the specific composition or structure of the lens. This universality makes gravitational lensing a powerful probe of dark matter and compact objects, as the lensing signal directly traces the mass distribution regardless of luminosity.

2.1.2. The Lens Equation

To formalize the relationship between source position and observed image position, we construct the lens equation in the thin lens approximation. We define the angular position of the source in the absence of lensing as $\boldsymbol{\beta}$, measured from the optical axis connecting the observer and lens. The observed angular position of the image is $\boldsymbol{\theta}$, also measured from this axis. Gravitational lensing describes how a lens mass at position $\boldsymbol{\theta}$ maps rays from a source at position $\boldsymbol{\beta}$ to images on the observer's sky.

The lens equation relates these quantities through the deflection angle. In the simplest

case of an axisymmetric point-mass lens, the two angular variables become scalars β and θ measured from the optical axis. The lens equation can then be written as

$$\beta = \theta - \frac{\theta_E^2}{\theta}, \tag{2.2}$$

where θ_E is the Einstein radius. This equation is obtained by projecting the geometry of the lensing system and relating the true source position to the observed image position through the deflection angle.

When the source lies directly behind the lens ($\beta = 0$), the lens equation predicts a ring of images of angular radius θ_E, known as an Einstein ring. The angular Einstein radius depends on the lens mass and the distances between observer, lens, and source:

$$\theta_E = \sqrt{\frac{4GM}{c^2}\frac{D_{\rm ls}}{D_{\rm l}D_{\rm s}}}, \tag{2.3}$$

where $D_{\rm l}$ is the angular-diameter distance from observer to lens, $D_{\rm s}$ is the distance from observer to source, and $D_{\rm ls}$ is the distance from lens to source. In physical units the Einstein radius in the source plane is

$$R_E = \sqrt{\frac{4GM}{c^2}\frac{D_{\rm l}D_{\rm ls}}{D_{\rm s}}}. \tag{2.4}$$

The Einstein radius represents the characteristic angular scale of the lensing effect: it separates two regimes of lensing behavior. For source positions $\beta \ll \theta_E$, the lens creates significant magnification and distortion. For $\beta \gg \theta_E$, lensing effects become negligible. For a stellar-mass lens in the Galactic bulge, θ_E is of order microarcseconds, so the two images lie too close to be resolved individually. Nevertheless, the total magnification of the source depends sensitively on the alignment encoded in Equation (2.2); this sensitivity underlies the microlensing phenomena discussed below.

2.1.3. Gravitational Microlensing

When the angular Einstein radius is micro-arcsecond, individual lensed images cannot be resolved with current instruments. Instead, observers detect microlensing through the time-dependent magnification of the source flux as the lens–source alignment evolves. The microlensing regime thus differs from strong lensing: no multiple images are seen, and the light curve is symmetric in time for a single lens. Because the image separations are ∼milli-arcseconds, only the combined flux matters.

Microlensing events occur when a foreground star (or planetary system) passes close to the line of sight to a background star, producing a characteristic brightening and fading. The Einstein radius sets both the angular scale and the time scale of the event;

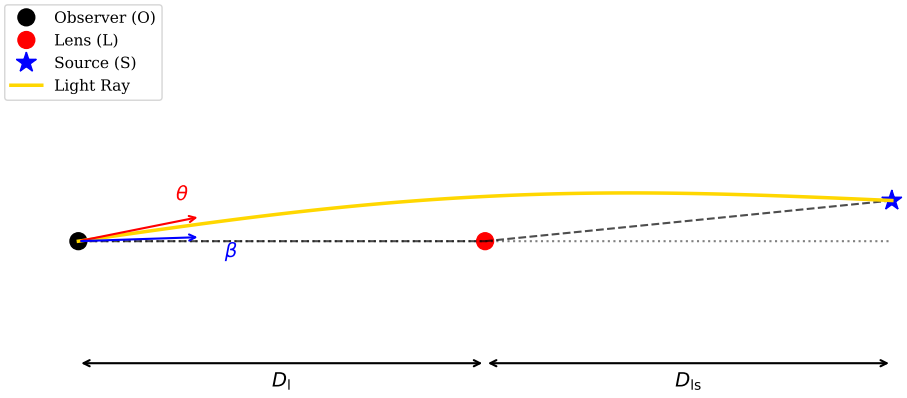

Figure 2.1.: Schematic diagram of gravitational lens geometry. The observer (O), lens (L), and source (S) are shown with their respective angular diameter distances (D_l, D_s, D_ls). The true source position is β, the image position is θ, and the light ray is deflected by angle α. The Einstein radius θ_E represents the characteristic angular scale of the lensing effect.

typical Galactic events have Einstein crossing times of days to months, depending on the lens mass and relative motion.

Microlensing is achromatic for a point source because the lensing magnification is wavelength-independent, although finite-source effects can introduce chromaticity at very high magnifications. The characteristic timescale, known as the Einstein crossing time t_E, represents the time for the source to traverse an Einstein radius due to the relative proper motion between lens and source. Section 2.2 develops the formalism needed to model microlensing light curves and to interpret their parameters.

2.2. Microlensing Light Curves

The light curve of a microlensing event encodes the geometry and kinematics of the lens–source system. In the simplest case of a point source lensed by a single point mass, the flux increases smoothly to a maximum and then declines symmetrically. More complex lenses (e.g., binaries) or finite source sizes introduce caustics, asymmetries and multiple peaks. In this section we derive the standard point-source point-lens (PSPL) magnification formula (Section 2.2.1), define key parameters, outline the effects of binary lenses (Section 2.2.2), and summarize finite-source effects (Section 2.2.3).

2.2.1. Point-Source Point-Lens (PSPL) Model

For a point source lensed by a single mass, the total magnification A of the unresolved images can be expressed in terms of the dimensionless separation u (in units of θ_E) between the lens and the source. A derivation from the Jacobian of the lens mapping

yields

$$A(u) = \frac{u^2 + 2}{u\sqrt{u^2 + 4}}, \tag{2.5}$$

where $u = |\boldsymbol{\beta}|/\theta_E$ is the source–lens separation in units of the Einstein radius. This formula shows that magnification increases as u decreases, becoming formally infinite when the source and lens are perfectly aligned ($u = 0$). In practice, finite source size limits the maximum magnification, a point we return to in Section 2.2.3.

If the lens moves with constant velocity relative to the source, the separation $u(t)$ varies as

$$u(t) = \sqrt{u_0^2 + \left(\frac{t - t_0}{t_E}\right)^2}, \tag{2.6}$$

where u_0 is the minimum impact parameter (closest approach in units of θ_E), t_0 is the time of closest approach, and t_E is the Einstein crossing time. The Einstein time represents the characteristic duration of the event and is defined as

$$t_E = \frac{R_E}{v_\perp}, \tag{2.7}$$

where $v_\perp$ is the transverse velocity of the lens relative to the source. Combining Equation (2.5) and Equation (2.6) gives the PSPL light curve as a function of time.

The observed flux in a microlensing event is

$$F(t) = F_{\text{base}} + F_{\text{source}}\left[A(u(t)) - 1\right], \tag{2.8}$$

where F_{base} is the baseline flux (including any blended light from unresolved stars) and F_{source} is the unlensed flux of the source. This equation accounts for the fact that only the source is magnified, not any blended stars in the same resolution element.

The PSPL model is characterized by three physical parameters: u_0, t_0, and t_E. Events with small u_0 (close alignment) produce high-magnification light curves with sharp peaks, while events with large u_0 produce broad, low-amplitude variations. The Einstein time t_E sets the overall duration of the event, with shorter times indicating either lower-mass lenses or higher relative velocities. For typical Galactic bulge events, t_E ranges from roughly 10 to 150 days.

2.2.2. Binary Lenses

When the lens consists of two masses—such as a star with a planetary companion—the lensing geometry becomes significantly more complex. Binary lenses introduce additional parameters: the projected separation s (in units of θ_E), the mass ratio $q = M_2/M_1$, and the angle α defining the source trajectory relative to the binary axis. For binary separations near $s \sim 1$, the caustic structure creates regions where the magnification is

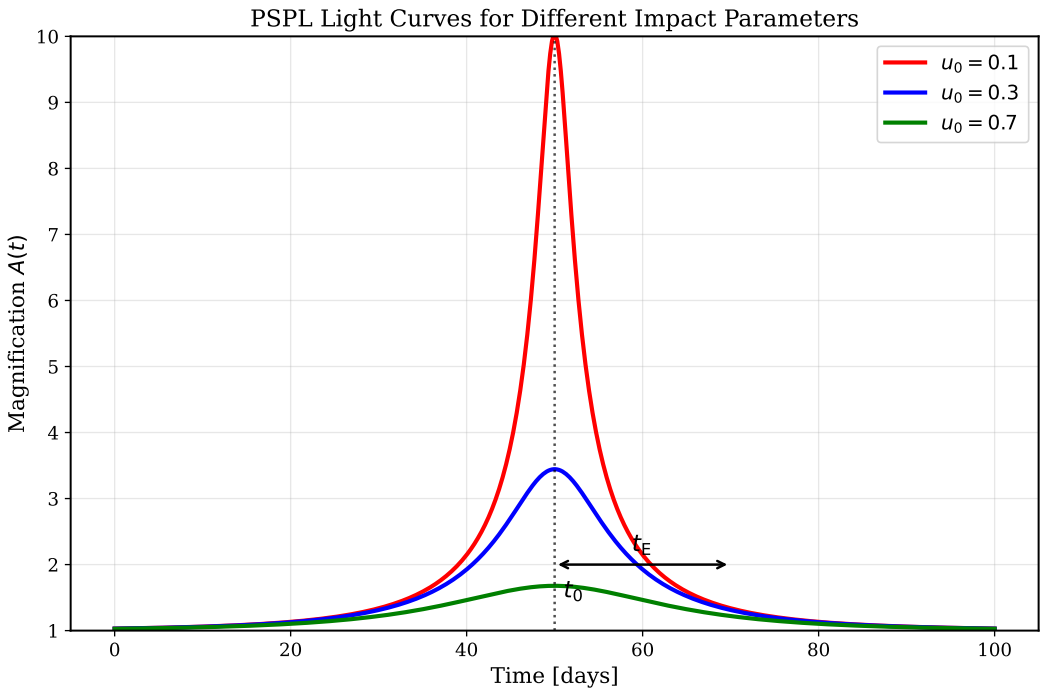

Figure 2.2.: Point Source Point Lens (PSPL) light curves for three different impact parameters: $u_0 = 0.1$ (high magnification, red), $u_0 = 0.3$ (moderate magnification, blue), and $u_0 = 0.7$ (low magnification, green). The time of closest approach t_0 and Einstein crossing time t_E are indicated. The symmetric shape and smooth profiles are characteristic of single-lens events.

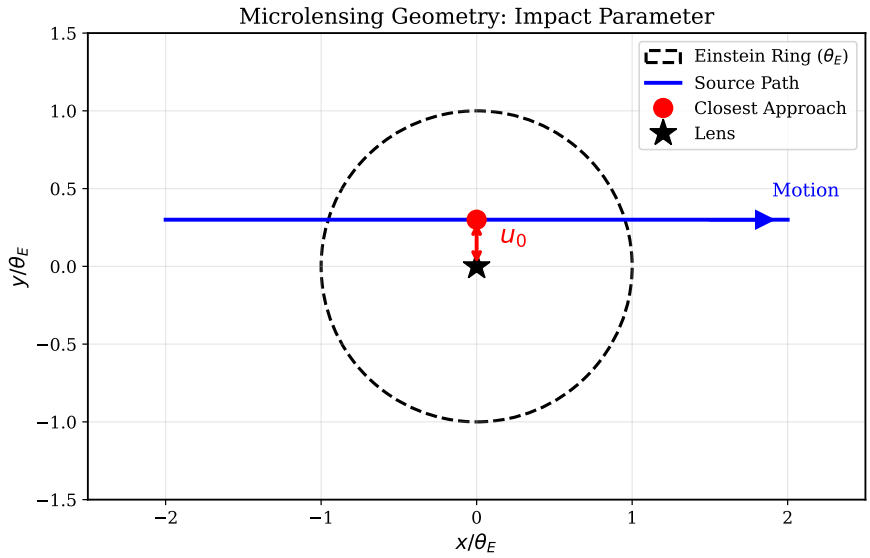

Figure 2.3.: Geometric representation of the impact parameter u_0 in microlensing. The source (blue trajectory) moves with respect to the lens (black star at origin). The Einstein ring (dashed circle) has radius θ_E. The minimum separation between the source path and the lens defines the impact parameter u_0, measured in units of the Einstein radius. Smaller impact parameters result in higher magnification events.

formally infinite (for point sources) or extremely large (for finite sources).

The magnification for a binary lens cannot be written in closed form like Equation (2.5). Instead, it must be computed either by solving a fifth-order complex polynomial at each time step or by using ray-tracing techniques that simulate the paths of many light rays through the lens system. The resulting light curves exhibit characteristic features that distinguish them from PSPL events:

- **Caustic crossings:** When the source crosses a caustic, the light curve shows sharp spikes or bumps lasting a fraction of t_E. The spike shape depends on the source size and the crossing speed.
- **Asymmetric profiles:** Binary light curves are typically asymmetric, with the rise and fall times differing depending on the source trajectory relative to the caustic structure.
- **Multiple peaks:** Depending on the geometry, a source may cross multiple caustic features, producing light curves with two or more distinct peaks.
- **Longer duration anomalies:** Even without caustic crossings, the perturbation from the companion can create extended deviations from the PSPL profile lasting days to weeks.

Planetary companions with mass ratios $q \sim 10^{-3}$ to 10^{-4} produce caustics that are small compared to the primary Einstein radius. The probability of detecting such planets depends on the survey cadence: high-frequency observations (multiple times per night) are essential to capture short-duration caustic crossings that might otherwise be missed. The complexity of binary lens modeling motivates the machine learning approach developed in this thesis, as traditional fitting methods become computationally prohibitive when analyzing thousands of events in real time.

2.2.3. Finite Source Effects

The point-source approximation breaks down when u becomes comparable to the source radius ρ (in units of θ_E). For a source of angular radius θ_*, the dimensionless source size is $\rho = \theta_*/\theta_E$. Typical main-sequence stars in the Galactic bulge have $\rho \sim 10^{-3}$ to 10^{-2}, becoming important only for high-magnification events with $u_0 \lesssim \rho$.

Finite source size modifies the light curve in two ways. First, it limits the maximum magnification: instead of diverging as $u \to 0$, the magnification saturates at a finite value determined by limb darkening and the source profile. Second, during caustic crossings in binary events, finite source size smooths the sharp features predicted for point sources, broadening the spikes into rounded peaks with characteristic widths determined by ρ and the crossing speed.

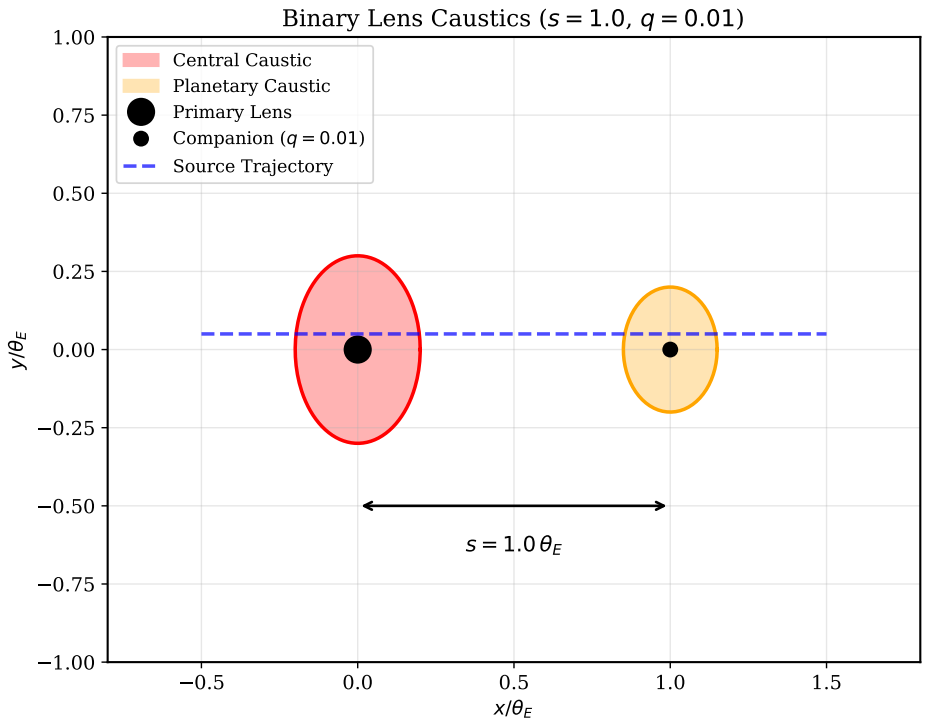

Figure 2.4.: Illustration of binary lens caustics for a system with separation $s = 1.0$ and mass ratio $q = 0.01$ (representative of a planetary companion). The red central caustic and orange planetary caustic represent regions where magnification becomes extremely large. The two black dots indicate the lens positions (primary at origin, companion at $x = 1.0\,\theta_E$). A source trajectory (blue dashed line) crossing or passing near these caustics produces characteristic deviations from PSPL behavior. The caustic topology depends sensitively on the separation s and mass ratio q.

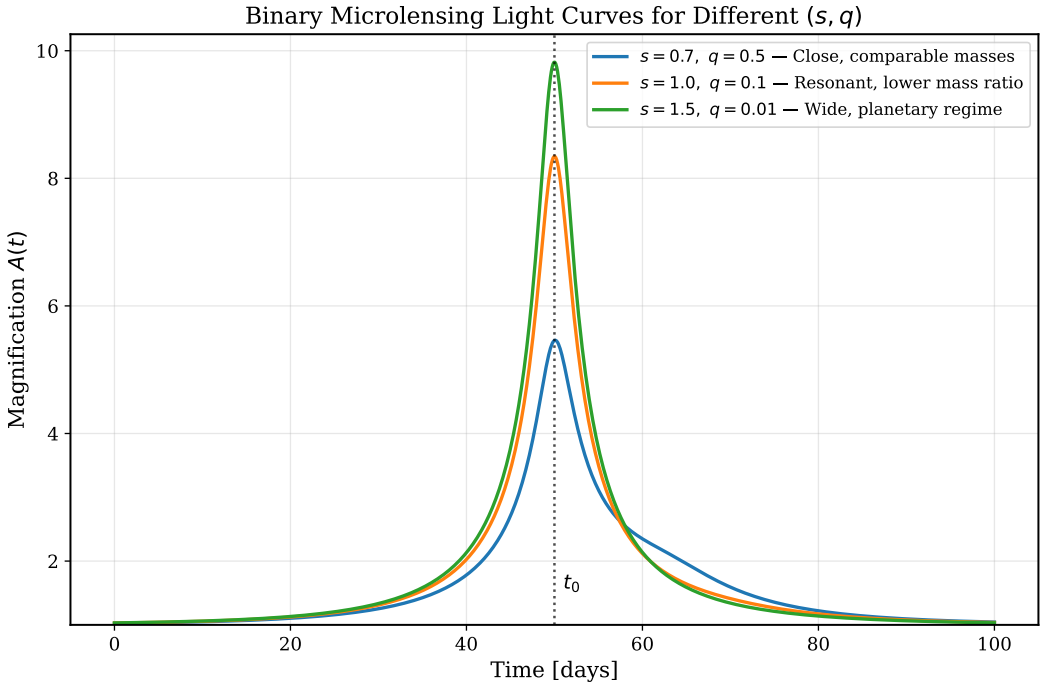

Figure 2.5.: Binary microlensing light curves showing diverse morphologies depending on source trajectory and binary parameters. Top: Caustic crossing event with sharp spike. Middle: Near-caustic event with asymmetric perturbation. Bottom: Resonant caustic event with multiple peaks. These features distinguish binary events from the smooth, symmetric PSPL profile and motivate the use of machine learning for automated classification.

Finite source effects are crucial for determining lens masses through measurements of the Einstein radius crossing time and the microlensing parallax. However, for the machine learning classification task in this thesis, we focus primarily on distinguishing the overall morphological differences between PSPL and binary light curves rather than on precise parameter estimation that requires accounting for finite source size.

2.3. Observational Strategies

Detecting and characterizing microlensing events requires sustained, high-cadence monitoring of millions of stars in crowded fields. Modern surveys employ difference imaging analysis (DIA) to overcome the challenges of crowding and varying seeing conditions. In this section we describe the observational techniques used by current surveys (Section 2.3.1), photometric challenges related to blending and baseline flux (Section 2.3.2), and the requirements for detecting planetary signals (Section 2.3.3).

2.3.1. Survey Overview

The two major ground-based microlensing surveys, OGLE (Optical Gravitational Lensing Experiment) and MOA (Microlensing Observations in Astrophysics), have been

monitoring the Galactic bulge for over two decades. OGLE operates a 1.3-m telescope at Las Campanas Observatory in Chile and monitors approximately 200 million stars across roughly 600 square degrees. The OGLE Early Warning System issues real-time alerts for ongoing events with high magnification, enabling follow-up observations by other groups. MOA uses a 1.8-m telescope at Mount John Observatory in New Zealand and monitors approximately 17 square degrees of the bulge with sampling rates up to six times per night.

Future space-based missions will revolutionize microlensing surveys. The Nancy Grace Roman Space Telescope, scheduled for launch in the mid-2020s, will conduct a dedicated microlensing survey monitoring the Galactic bulge from space, free from weather interruptions and atmospheric seeing effects. Roman's wide-field infrared camera will observe 2.8 square degrees with unprecedented photometric precision, detecting thousands of microlensing events per year and enabling the detection of planets down to Mars mass. The Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST) will also contribute to microlensing science through its deep, wide-area, time-domain survey, though its observing strategy is optimized for transient detection rather than continuous monitoring.

2.3.2. Photometric Challenges

Difference imaging analysis (DIA) subtracts a high-quality reference image from each new observation to isolate flux variations. This technique is essential in crowded fields where traditional aperture photometry fails due to overlapping stellar point spread functions (PSFs). However, DIA introduces its own challenges. The baseline flux (unmagnified flux) of a source is not measured directly in the difference image and must be estimated separately. If the baseline flux is mis-estimated, the inferred magnification curve can be mis-normalized, leading to degeneracies in the fit. High-magnification events are less affected because even a small fractional error in baseline flux produces a noticeable deviation in the light curve.

Blending remains a significant source of systematic error. Blending occurs when multiple stars fall within the same PSF, causing the measured flux to include light from both the lensed source and unresolved neighbors. Imperfect deblending increases the measured background, causing a magnitude bias near the detection threshold; faint stars can disappear within the combined PSF of brighter neighbors; and centroid motions induced by blending can bias the flux measurement. If blending is neglected in modeling, the impact parameter u_0 tends to be overestimated and the Einstein time t_E underestimated, biasing optical depth estimates. Even bright stars are often blended and require careful modeling.

Extinction and reddening by interstellar dust further modify observed colors and must

be accounted for when comparing different fields. The complex dust distribution toward the Galactic bulge creates significant variations in extinction across the field, requiring color-dependent corrections that can introduce additional systematic uncertainties.

2.3.3. Observational Challenges

Monitoring millions of stars night after night poses logistical and technical challenges. Crowded fields demand high image quality and consistent PSF across the field; variable seeing and weather conditions introduce systematic errors that complicate DIA. Detection thresholds vary across the field because the background is dominated by merged stellar wings, and changes in seeing can overwhelm deblending algorithms.

Real-time event detection and follow-up are essential for capturing short perturbations caused by planets. High-magnification events are of particular interest because they have high planet detection efficiencies. Automated anomaly detectors trigger increased sampling when deviations from the PSPL model are seen. A high-cadence strategy may employ survey observations every day, follow-up observations every $\sim$90 minutes near peak, and anomaly monitoring every few minutes when an unexpected deviation occurs. Such multi-tiered cadence maximizes the probability of detecting short-lived planetary signals.

Weather interruptions, moonlight and seasonal visibility limit the temporal coverage from a single site, necessitating networks of telescopes around the globe. Robotic networks and automated analysis pipelines (e.g., OGLE's Early Warning System and MOA's real-time DIA) have become standard tools for coordinating follow-up observations. The development of machine learning classifiers that can rapidly identify anomalous events in real time represents a natural extension of these automated systems, enabling more efficient allocation of follow-up resources.

2.4. Machine Learning for Time Series

Traditional microlensing analyses fit parametric models to each event, a process that can be computationally intensive and requires good initial guesses for non-linear parameters. Machine learning offers an alternative approach: by training on simulated and observed light curves, an ML model can learn to recognize microlensing events and infer parameters quickly. For large surveys expecting thousands of events per year, automatic classification and real-time parameter estimation become invaluable.

Machine learning circumvents computational bottlenecks by reframing the problem as pattern recognition. Rather than searching parameter space for each new observation, a neural network learns to recognize characteristic patterns in light curves that correlate with specific physical scenarios. After an initial training phase—which may be com-

putationally expensive—inference on new light curves becomes nearly instantaneous, requiring only a forward pass through the network. This speed advantage is crucial for real-time classification systems that must process hundreds of ongoing events to prioritize follow-up observations.

What makes microlensing particularly amenable to machine learning approaches? Several factors contribute: first, the underlying physical model is well-understood and deterministic, allowing generation of essentially unlimited training data through simulation. Second, light curves exhibit characteristic morphologies—smooth symmetric rises for PSPL events, sharp caustic crossing spikes for binaries—that should be recognizable to pattern-matching algorithms. Third, the photometric time series is relatively low-dimensional compared to images, making network architectures tractable without requiring massive computational resources.

In this section we outline why ML is suitable for microlensing time series, introduce convolutional neural networks (CNNs) as the architecture used in this thesis (Section 2.4.1), describe their application to one-dimensional time series data (Section 2.4.2), and explain the training and validation strategies we employ (Section 2.4.3).

2.4.1. Convolutional Neural Networks

Convolutional neural networks (CNNs) have emerged as the dominant architecture for learning from data with spatial or temporal structure. Unlike fully connected neural networks, where every neuron connects to every neuron in adjacent layers, CNNs exploit local structure through convolutional layers that apply learned filters across the input. This design reflects two key principles: local patterns matter (nearby data points are more related than distant ones), and the same patterns can appear at different positions (translation invariance).

A typical CNN architecture consists of several key building blocks:

- **Convolutional Layers:** This is the core component. A convolutional layer uses a set of small, learnable filters (kernels) that slide across the input data, performing dot products at each position. Each kernel detects a specific local feature—such as an edge in an image, or a sharp spike in a time series. The key insight is **parameter sharing**: the same kernel is used across the entire input, allowing the network to detect that feature regardless of where it appears. For a one-dimensional input $\mathbf{x}$ and kernel weights $\mathbf{w}$, the output at position i is

$$y[i] = \sum_{k=0}^{K-1} w[k] \cdot x[i+k] + b, \tag{2.9}$$

  where K is the kernel size and b is a bias term. Multiple kernels are applied in parallel, each learning to detect different features.

- **Activation Functions:** After each convolution, the output is passed through a non-linear activation function, most commonly the Rectified Linear Unit (ReLU), defined as $f(x) = \max(0, x)$. This non-linearity allows the network to learn complex, non-trivial patterns. ReLU effectively zeros out negative activations while passing positive values unchanged, creating sparse representations where only a subset of neurons activate for any given input.

- **Pooling Layers:** These layers reduce dimensionality while retaining important features. Max pooling selects the maximum value within each local window, effectively downsampling the representation. Pooling provides translation invariance (small shifts in input don't dramatically change the pooled output), reduces computational cost in subsequent layers, and helps prevent overfitting by reducing the number of parameters.

- **Fully Connected Layers:** After several convolutional and pooling layers extract hierarchical features from the input, the features are flattened and fed into one or more fully connected layers, where each neuron is connected to all neurons in the previous layer. For classification tasks, the final layer typically uses a softmax activation to produce a probability distribution over classes.

The entire network is trained end-to-end using backpropagation, where gradients of a loss function with respect to all weights are computed and used to update parameters through optimization algorithms like Adam or stochastic gradient descent. This hierarchical structure allows a CNN to learn simple features (like sharpness or slopes) in early layers and then combine them into more complex, abstract features (like caustic-crossing anomalies) in deeper layers.

2.4.2. 1D CNNs for Time Series

While CNNs were originally developed for two-dimensional image analysis, the same principles extend naturally to one-dimensional data like time series. A 1D CNN applies convolutions along the temporal axis, with kernels that learn to recognize patterns in sequences rather than spatial arrangements. For microlensing light curves, this architecture is particularly appropriate because the data is inherently sequential: flux measurements at different times constitute a one-dimensional signal.

Interpreting the kernel size as a temporal scale provides intuition for network design. For a CNN analyzing microlensing light curves with $\sim$100 observations spanning $\sim$100 days, a kernel of size 5 effectively examines 5-day windows, appropriate for detecting short-timescale planetary anomalies. Kernels of size 11 or 15 capture the overall shape of PSPL events. By stacking multiple convolutional layers with varying kernel sizes, the network builds a hierarchy of representations: early layers detect local features such as

the symmetric rise and fall of a microlensing event or the sharper spikes produced by caustic crossings, while deeper layers combine these into higher-level patterns that span the entire light curve.

One advantage of 1D CNNs for irregularly sampled data is their flexibility with input representation. One-dimensional convolutions naturally handle irregular sampling if time bins are uniform after interpolation or binning. Rather than requiring evenly spaced time points, light curves can be preprocessed into regularly sampled representations. For most microlensing applications, simple interpolation to a fixed grid (e.g., 128 or 256 time bins spanning the event duration) provides a reasonable compromise between preserving information and maintaining tractable network architectures. The 1D CNN architecture requires fewer parameters than fully connected networks, reducing overfitting and computational cost.

The architecture used in this thesis employs several 1D convolution and pooling layers followed by dense layers to classify events and estimate parameters. Subsequent pooling layers aggregate local features, enabling the network to detect features irrespective of their exact timing, providing robustness to small temporal shifts and irregular sampling patterns.

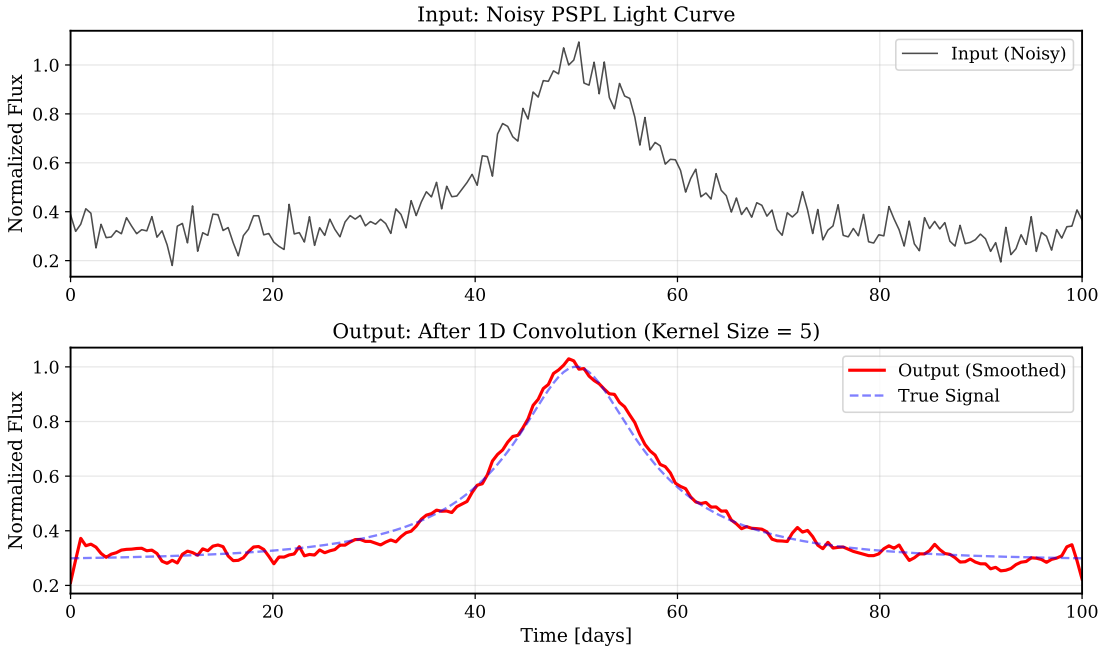

Figure 2.6.: Demonstration of 1D convolution applied to a noisy PSPL light curve. Top panel: Input light curve with photometric noise (black). Bottom panel: Output after applying a 5-point smoothing kernel (red) compared to the true underlying signal (blue dashed). The convolution operation extracts local temporal features while reducing noise, illustrating how convolutional layers in a CNN can learn to recognize characteristic patterns in microlensing data.

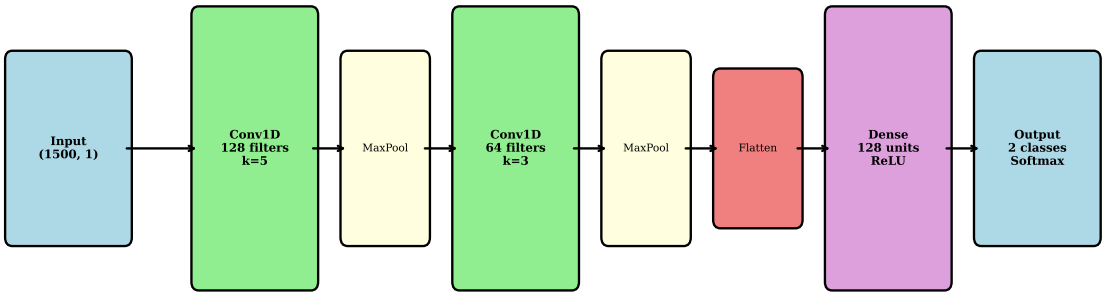

Figure 2.7.: Schematic diagram of the 1D Convolutional Neural Network architecture used in this thesis. The input light curve (1500 timesteps) passes through alternating convolutional (Conv1D) and pooling layers that extract hierarchical features at different temporal scales. After feature extraction, the data is flattened and passed through fully connected (Dense) layers to produce the final binary classification (PSPL vs. Binary) using a softmax activation. Dropout layers (not shown) are applied after each convolutional layer to prevent overfitting.

2.4.3. Training and Validation Strategies

Training a neural network for microlensing classification requires careful dataset partitioning, regularization to prevent overfitting, and appropriate evaluation metrics. Successful machine learning requires careful handling of the training data and model capacity.

The available dataset is typically divided into three subsets: a training set to fit the model (typically 70–80%), a validation set to tune hyperparameters and monitor overfitting (10–15%), and a test set to provide an unbiased estimate of performance on unseen data (10–15%). Each partition should contain balanced representations of both PSPL and binary events to ensure the model learns from diverse examples.

To prevent overfitting, regularization techniques such as early stopping and dropout are employed. Early stopping monitors the validation loss during training and halts training when the loss ceases to decrease, thus avoiding the deterioration of generalization performance. Dropout randomly disables a fraction of neurons during each training step (typically 20–50%), preventing them from co-adapting and reducing overfitting.

For classification tasks, the cross-entropy loss function

$$\mathcal{L} = -\sum_{c=1}^{C} y_c \log(\hat{y}_c) \tag{2.10}$$

is commonly used, where y_c is the true class label (one-hot encoded) and $\hat{y}_c$ is the predicted probability for class c. Evaluation metrics such as accuracy, precision and recall quantify the performance of the classifier on the test set.

For classification, we use standard metrics: **accuracy** (overall correct classification rate), **precision** (fraction of predicted binaries that are truly binary), and **recall** (fraction of true binaries correctly identified). For imbalanced datasets where one class is rare, precision and recall are often more informative than simple accuracy. The F_1 score, the harmonic mean of precision and recall, provides a single metric balancing both concerns. For regression tasks such as predicting event parameters, metrics like mean absolute error or root mean square error are used.

When data are limited, cross-validation can be employed: the dataset is split into k folds, with each fold serving as the test set once while the others form the training set. The results are averaged to obtain a robust estimate of performance. However, for large simulated datasets with millions of examples, a simple train–validation–test split typically suffices.

In this thesis we adopt a train–validation–test split, apply dropout and early stopping, and use cross-entropy loss to train a 1D CNN that classifies light curves and estimates microlensing parameters. The model is trained using the Adam optimizer with appropriate learning rates and batch sizes, with hyperparameters tuned based on validation set performance. For operational deployment, we focus on high-confidence predictions where the model's softmax probability exceeds a threshold (e.g., 0.8), ensuring reliable event classification for follow-up triggering decisions.

3. Literature Review

[Content to be written]

4. Methodology

[Content to be written]

5. Results

[Content to be written]

6. Discussion

[Content to be written]

7. Conclusions and Future Work

[Content to be written]

Bibliography

[1] A. Einstein. Lens-like action of a star by the deviation of light in the gravitational field. *Science*, 84(2188):506–507, 1936. doi: 10.1126/science.84.2188.506.

[2] B. Paczyński. Gravitational microlensing by the galactic halo. *The Astrophysical Journal*, 304:1–5, 1936. doi: 10.1086/164140.

[3] B. Scott Gaudi. Microlensing surveys for exoplanets. *Annual Review of Astronomy and Astrophysics*, 50:411–453, 2012. doi: 10.1146/annurev-astro-081811-125518.

[4] S. Mao and B. Paczyński. Gravitational microlensing by double stars and planetary systems. *The Astrophysical Journal*, 374:L37–L40, 1991. doi: 10.1086/186066.

[5] A. Gould and A. Loeb. Discovering planetary systems through gravitational microlenses. *The Astrophysical Journal*, 396:104–114, 1992. doi: 10.1086/171700.

[6] S. Mao. Gravitational microlensing. *Research in Astronomy and Astrophysics*, 12 (8):947–972, 2012. doi: 10.1088/1674-4527/12/8/005.

[7] A. Udalski et al. OGLE-IV: Fourth phase of the optical gravitational lensing experiment. *Acta Astronomica*, 65:1–38, 2015.

[8] I. A. Bond et al. The MOA project: Microlensing optical depth toward the galactic bulge from MOA-II. *Monthly Notices of the Royal Astronomical Society*, 469(2): 2434–2460, 2017. doi: 10.1093/mnras/stx1049.

[9] P. Mróz et al. A neural network classifier for microlensing in wide-field optical surveys. *The Astronomical Journal*, 159(6):262, 2020. doi: 10.3847/1538-3881/ab8aeb.

[10] M. Dominik. The binary lens caustic structure. *Astronomy and Astrophysics*, 349: 108–125, 1999.

[11] LSST Science Collaboration. *LSST Science Book, Version 2.0*. 2009. URL https://www.lsst.org/scientists/scibook.

[12] Natasha S. Abrams et al. Forecasting microlensing events with Gaia and LSST. *The Astrophysical Journal*, 954(2):180, 2023. doi: 10.3847/1538-4357/ace618.

[13] M. T. Penny et al. Predictions of the Nancy Grace Roman Space Telescope galactic exoplanet survey. II. free-floating planet detection rates. *The Astrophysical Journal Supplement Series*, 241(1):3, 2019. doi: 10.3847/1538-4365/aafb69.

[14] R. A. Street et al. Predictions for microlensing planetary detections during the Roman mission. *The Astronomical Journal*, 157(1):45, 2018. doi: 10.3847/1538-3881/aaf054.

[15] Subo Dong et al. Efficient grid method for modeling binary microlensing events. *The Astrophysical Journal*, 642:842–860, 2006. doi: 10.1086/501224.

[16] B. Scott Gaudi et al. Discovery of a jupiter/Saturn analog with gravitational microlensing. *Science*, 319(5865):927–930, 2008. doi: 10.1126/science.1151947.

[17] J. C. Yee et al. MOA-2011-BLG-293Lb: First microlensing planet possibly in the habitable zone. *The Astrophysical Journal*, 755(2):102, 2012. doi: 10.1088/0004-637X/755/2/102.

[18] J.-P. Beaulieu et al. Discovery of a cool planet of 5.5 Earth masses through gravitational microlensing. *Nature*, 439:437–440, 2006. doi: 10.1038/nature04441.

[19] Ž. Ivezić et al. LSST: From science drivers to reference design and anticipated data products. *The Astrophysical Journal*, 873(2):111, 2019. doi: 10.3847/1538-4357/ab042c.

[20] F. W. Dyson, A. S. Eddington, and C. Davidson. A determination of the deflection of light by the sun's gravitational field. *Philosophical Transactions of the Royal Society of London A*, 220:291–333, 1920.

A. Code and Implementation Details

[Optional: Code listings, hyperparameters, etc.]