

# Table of Content

<b>1. Introduction</b>	<b>1</b>
1.1. Motivation . . . . .	1
1.1.1. Previous Works . . . . .	2
1.1.2. General Problem and Associated Difficulties . . . . .	4
1.2. Hypotheses . . . . .	6
1.3. General Objective . . . . .	6
1.4. Specific Objectives . . . . .	6
1.5. Main Contributions . . . . .	7
1.6. Organization of this Work . . . . .	7
<b>2. Theoretical Background</b>	<b>9</b>
2.1. Notation . . . . .	9
2.1.1. Multi-Band Light-Curve . . . . .	9
2.1.2. First and Last Sequence Steps . . . . .	10
2.1.3. Time Difference . . . . .	10
2.2. Data Science . . . . .	11
2.2.1. Linear Algebra . . . . .	11
2.2.1.1. Vector Concatenation . . . . .	11
2.2.1.2. Linear Projection . . . . .	11
2.2.1.3. Eigenvectors and Eigenvalues . . . . .	12
2.2.2. Probability Theory . . . . .	12

2.2.2.1.	Random Variable and Probability Distribution . . . . .	12
2.2.2.2.	Cumulative Distribution Function (CDF) . . . . .	13
2.2.2.3.	Expectation Operator . . . . .	13
2.2.2.4.	Covariance and Variance . . . . .	13
2.2.2.5.	Covariance Matrix . . . . .	14
2.2.3.	Information Theory . . . . .	14
2.2.3.1.	Entropy . . . . .	15
2.2.3.2.	Cross-Entropy . . . . .	15
2.2.4.	Machine Learning . . . . .	15
2.2.4.1.	Supervised Learning . . . . .	17
2.2.4.2.	Unsupervised Learning . . . . .	18
2.2.4.3.	Semi-Supervised Learning . . . . .	19
2.2.4.4.	Models and Algorithms . . . . .	19
2.2.5.	Bayesian Learning . . . . .	20
2.2.5.1.	Bayes Theorem . . . . .	20
2.2.5.2.	Maximum A Posteriori (MAP) Estimation . . . . .	21
2.2.5.3.	Maximum Likelihood Estimation (MLE) . . . . .	22
2.2.6.	Iterative Optimization Algorithms . . . . .	24
2.2.6.1.	Gradient Descent (GD) . . . . .	24
2.2.6.2.	Stochastic Gradient Descent (SGD) . . . . .	25
2.2.6.3.	Other Optimization Algorithms . . . . .	26
2.2.7.	Monte Carlo Method . . . . .	26
2.2.8.	Markov Chain Model . . . . .	27
2.2.8.1.	Stationary Distribution . . . . .	28
2.2.9.	Monte Carlo Inference . . . . .	29
2.2.9.1.	Markov Chain Monte Carlo (MCMC) . . . . .	29
2.2.9.2.	Metropolis Hastings (MH) . . . . .	29

2.2.10. Artificial Neural Networks (ANNs) and Deep Learning . . . . .	31
2.2.10.1. Perceptron Model . . . . .	31
2.2.10.2. Fully-Connected Layer . . . . .	32
2.2.10.3. Multi-Layer Perceptron (MLP) . . . . .	33
2.2.11. Recurrent Neural Networks (RNNs) . . . . .	33
2.2.11.1. Vanilla RNNs . . . . .	34
2.2.11.2. Long Short-Term Memory Model (LSTM) . . . . .	34
2.2.11.3. Gated Recurrent Unit (GRU) Model . . . . .	36
2.2.12. Multi-Head Self-Attention Mechanisms . . . . .	38
2.2.12.1. Multi-Head Dot-Attention . . . . .	38
2.2.12.2. Multi-Head Self-Attention . . . . .	42
2.2.12.3. Additional Operations . . . . .	42
2.2.12.4. Extra Properties and Limitations . . . . .	43
2.2.13. Sequence Encoding . . . . .	44
2.2.13.1. Positional Encoding . . . . .	44
2.2.13.2. Juxtaposed Positional Encoding . . . . .	45
2.2.13.3. Temporal Encoding . . . . .	45
2.2.13.4. Juxtaposed Temporal Encoding . . . . .	46
2.2.13.5. Time2Vec Encoding . . . . .	46
2.2.14. Fourier Decomposition . . . . .	47
2.2.15. Feature-wise Linear Modulation (FiLM) . . . . .	48
2.2.16. Deep Learning Regularization and Training Techniques . . . . .	48
2.2.16.1. Dropout . . . . .	49
2.2.16.2. Early Stopping . . . . .	49
2.2.16.3. Batch Normalization . . . . .	49
2.2.16.4. Linear Warm-Up Schedule . . . . .	50
2.2.17. Decision Trees . . . . .	51

2.2.17.1. Classification and Regression Trees (CART) . . . . .	52
2.2.18. Random Forest (RF) . . . . .	53
2.2.18.1. Balanced Random Forest (BRF) . . . . .	54
2.2.19. Dimensionality Reduction . . . . .	54
2.2.19.1. Principal Component Analysis (PCA) . . . . .	55
2.2.19.2. Uniform Manifold Approximation and Projection (UMAP) .	56
2.3. Astronomy . . . . .	57
2.3.1. Optical Spectrometry . . . . .	57
2.3.2. Optical Photometry . . . . .	58
2.3.2.1. Photometric Flux . . . . .	58
2.3.2.2. Apparent Magnitude . . . . .	59
2.3.2.3. Magnitude-Flux Transformation . . . . .	59
2.3.2.4. Multi-Band Light-Curve . . . . .	60
2.3.3. Supernovae (SNe) . . . . .	61
2.3.3.1. Thermonuclear SNe . . . . .	61
2.3.3.2. Core-Collapse SNe . . . . .	62
2.3.4. Astronomical Surveys, Datasets, and Brokers . . . . .	62
2.3.5. Supernova Parametric Model (SPM) . . . . .	63
2.3.6. Computation of Astrophysical Features for Light-Curves . . . . .	64
<b>3. Methodology</b>	<b>67</b>
3.1. Dataset and Pre-Processing . . . . .	67
3.1.1. Dataset . . . . .	67
3.1.2. Pre-Processing . . . . .	68
3.1.2.1. Stratified 5-Fold Cross-Validation . . . . .	68
3.1.2.2. Simultaneous Multiple Observations . . . . .	68
3.1.2.3. Sigma Clipping Error Filter . . . . .	68
3.2. Generation of Supernova Synthetic Multi-Band Light-Curves . . . . .	69

3.2.1.	Estimation of the Posterior Distribution of SPM Parameters . . . . .	69
3.2.1.1.	MCMC parameters and settings . . . . .	70
3.2.2.	Sampling Time Window . . . . .	72
3.2.3.	Generation of Synthetic Observations . . . . .	72
3.2.4.	Synthetic Training-Set Generation . . . . .	74
3.3.	BRF Baseline Classifier . . . . .	74
3.3.1.	Photometric and Astrophysical Features . . . . .	74
3.3.2.	Balanced Random Forest . . . . .	75
3.4.	TimeModAttn Model . . . . .	75
3.4.1.	Loss Functions . . . . .	76
3.4.2.	Proposed Model Formulation . . . . .	77
3.4.3.	Serial Encoder . . . . .	77
3.4.3.1.	Temporal Modulation (TimeFiLM) . . . . .	78
3.4.3.2.	Multi-Head Self-Attention Mechanism (MHSelfAttn) . . . . .	80
3.4.3.3.	Temporal Offsets in Multi-Band Light-curves . . . . .	80
3.4.4.	Parallel Encoder . . . . .	81
3.4.4.1.	Temporal Offsets in Multi-Band Light-curves . . . . .	82
3.4.5.	Parallel Decoder . . . . .	82
3.4.6.	Classifier . . . . .	83
3.4.7.	Optimization Problem . . . . .	83
3.4.7.1.	Imbalance Learn and Regularization . . . . .	84
3.4.8.	RNN Baseline Models . . . . .	85
3.4.8.1.	Temporal Offsets in Multi-Band Light-curves . . . . .	85
<b>4.</b>	<b>Results and Analyses</b>	<b>87</b>
4.1.	Results and Analyses . . . . .	87
4.1.1.	Late-Classification Scenario . . . . .	88
4.1.2.	Early-Classification Scenario . . . . .	90

4.1.3.	Confusion Matrices and Operational Curves . . . . .	93
4.1.4.	Multi-band Effect in Classification . . . . .	95
4.1.5.	Interpretability Experiments . . . . .	95
4.1.5.1.	Attention Scores . . . . .	96
4.1.5.2.	Attention-Based Statistics . . . . .	97
4.1.5.3.	Temporal Modulation Variability . . . . .	99
4.1.6.	Empirical Computational Cost . . . . .	101
4.1.7.	Additional Experiments . . . . .	101
<b>5.</b>	<b>Conclusions</b>	<b>103</b>
5.1.	Conclusions . . . . .	103
5.1.1.	Future Work . . . . .	106
5.1.2.	Limitations . . . . .	108
	<b>Bibliography</b>	<b>116</b>
	<b>Appendices</b>	<b>117</b>
<b>A.</b>	<b>Theoretical Background Concepts</b>	<b>118</b>
A.1.	Additional Literature Summary Table . . . . .	118
A.2.	Balanced Multi-Class Performance Metrics . . . . .	118
<b>B.</b>	<b>Methodology Details</b>	<b>121</b>
B.1.	Feature Extraction Implementation Details . . . . .	121
B.2.	SPM Bounds and MCMC Prior Distribution . . . . .	122
B.3.	Conditional Observation-Error Distribution Estimation . . . . .	122
B.4.	Reconstruccion Loss Issue . . . . .	124
B.5.	Model Input Normalization . . . . .	125
B.6.	Class Balance Strategy . . . . .	125
B.7.	Multi-Band Light-Curve Data-Augmentation . . . . .	126

B.8. Repository Brief Roadmap and Used Software . . . . .	127
<b>C. Experiments and Results</b>	<b>129</b>
C.1. Relevance Ranking of the BRF model . . . . .	129
C.2. 2D Projections for Synthetic Light-Curves . . . . .	129
C.2.1. Limitations . . . . .	132
C.3. Additional Synthetic Light-Curves Generation Methods . . . . .	132
C.3.1. Performance Comparison . . . . .	134
C.3.1.1. Comparison Against the TimeModAttn Model . . . . .	138
C.4. Multi-band Light-Curve Reconstruction . . . . .	142
C.5. 2D Projections for TimeModAttn Model . . . . .	143
C.6. Ablation Studies and Additional Models . . . . .	146
C.6.1. Number of Fourier Decomposition Harmonics and Attention Heads .	146
C.6.2. TimeModRNN and CatTimeAttn Models . . . . .	149
C.6.2.1. TimeModRNN Model Discussion . . . . .	149
C.6.2.2. CatTimeAttn Model Discussion . . . . .	151
C.6.3. Number of MHSelfAttn’s Layers in the Encoder . . . . .	152
<b>D. Alternative Abstracts</b>	<b>155</b>
D.1. Extended 250 Words Abstract . . . . .	155
D.2. Extended 300 Words Abstract . . . . .	156