

2–4 August 2022
Busan, Republic of Korea

Machine Learning in Astronomy: Possibilities and Pitfalls

Edited by

Jess McIver
Ashish Mahabal
Christopher Fluke

ISSN 1743-9213

International Astronomical Union



CAMBRIDGE
UNIVERSITY PRESS



MACHINE LEARNING IN ASTRONOMY: POSSIBILITIES AND PITFALLS
IAU SYMPOSIUM 368

COVER ILLUSTRATION:

‘Word cloud generated from symposium abstracts’
(C) Jess McIver, Ashish Mahabal, Christopher Fluke

IAU SYMPOSIUM PROCEEDINGS SERIES

Chief Editor

JOSÉ MIGUEL RODRIGUEZ ESPINOSA, General Secretariat

Instituto de Astrofísica de Andalucía

Glorieta de la Astronomía s/n

18008 Granada

Spain

IAU-general.secretary@iap.fr

Editor

DIANA WORRALL, Assistant General Secretary

HH Wills Physics Laboratory

University of Bristol

Tyndall Avenue

Bristol

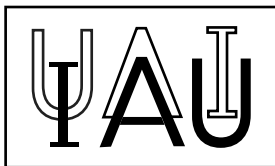
BS8 1TL

UK

IAU-assistant.general.secretary@iap.fr

INTERNATIONAL ASTRONOMICAL UNION
UNION ASTRONOMIQUE INTERNATIONALE

International Astronomical Union



MACHINE LEARNING IN ASTRONOMY: POSSIBILITIES AND PITFALLS

PROCEEDINGS OF THE 368th SYMPOSIUM OF
THE INTERNATIONAL ASTRONOMICAL UNION
BUSAN, REPUBLIC OF KOREA
2–4 AUGUST, 2022

Edited by

Jess McIver

University of British Columbia, Canada

Ashish Mahabal

California Institute of Technology, USA

and

Christopher Fluke

Swinburne University of Technology, Australia



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom
1 Liberty Plaza, Floor 20, New York, NY 10006, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

© International Astronomical Union 2025

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of the International Astronomical Union.

First published 2025

Printed in Great Britain by Henry Ling Limited, The Dorset Press, Dorchester, DT1 1HQ

Typeset in System L^AT_EX 2 ϵ

*A catalogue record for this book is available from the British Library Library of Congress
Cataloguing in Publication data*

This journal issue has been printed on FSCTM-certified paper and cover board. FSC is an
independent, non-governmental, not-for-profit organization established to promote the
responsible management of the world's forests. Please see www.fsc.org for information.

ISBN 9781009345194 hardback
ISSN 1743-9213

Table of Contents

Preface	vii
Editors	ix
List of Participants	x
Enhancing Exoplanet Surveys via Physics-informed Machine Learning <i>Eric B. Ford</i>	1
How do we design data sets for Machine Learning astronomy? <i>Renée Hložek</i>	11
Deep Machine Learning in Cosmology: Evolution or Revolution? <i>Ofer Lahav</i>	28
An Astronomers Guide to Machine Learning <i>Sara A. Webb and Simon R. Goode</i>	40
Panel Discussion: Practical Problem Solving for Machine Learning <i>Guillermo Cabrera, Sungwook E. Hong, Lilianne Nakazono, David Parkinson and Yuan-Sen Ting</i>	50
Panel discussion: Methodology for fusion of large datasets <i>Nikhita Madhanpall, Kai Polsterer, Mike Walmsley and Shay Zucker</i>	56
The entropy of galaxy spectra <i>I. Ferreras, O. Lahav, R. S. Somerville and J. Silk</i>	58
Unsupervised classification: a necessary step for Deep Learning? <i>Didier Fraix-Burnet</i>	63
Spectral identification and classification of dusty stellar sources using spectroscopic and multiwavelength observations through machine learning <i>Sepideh Ghaziasgar, Amirhossein Masoudnezhad, Atefeh Javadi, Jacco Th. van Loon, Habib G. Khosroshahi and Negin Khosravaninezhad</i>	67
Simulating transient burst noise with gengli <i>Melissa Lopez, Vincent Boudart, Stefano Schmidt and Sarah Caudill</i>	73
Detecting complex sources in large surveys using an apparent complexity measure <i>David Parkinson and Gary Segal</i>	80
Machine Learning in the study of Star Clusters with Gaia EDR3 <i>Priya Hasan, Md Mahmudunnobe, Mudasir Raja, Md Saifuddin and S N Hasan</i>	86
Assessing the Quality of Massive Spectroscopic Surveys with Unsupervised Machine Learning <i>John F. Suárez-Pérez and Jaime Forero-Romero</i>	91

Neural Networks for Meteorite and Meteor Recognition	95
<i>Aisha Al-Owais, Maryam Sharif, Ilias Fernini, Antonios Manousakis</i>	
Unsupervised clustering visualisation tool for Gaia DR3	98
<i>Marco Álvarez, Carlos Dafonte, Minia Manteiga, Daniel Garabato, Raúl Santoveña and Lara Pallas</i>	
Kinematic Planetary Signature Finder (KPSFinder): convolutional neural network-based tool to search for exoplanets in ALMA data	101
<i>Jaehan Bae</i>	
Predicting Physical Parameters of Cepheid and RR Lyrae variables in an Instant with Machine Learning	104
<i>A. Bhardwaj, E. P. Bellinger, S. M. Kanbur and M. Marconi</i>	
Bayesian deconvolution of a rotating spectral line profile to a non-rotating one	107
<i>M. Curé, P. Escarate, L. Celedon, J. Cavieres, E. Olivares, I. Araya, C. Arcos, R. Pezoa, G. Farias and N. Machuca</i>	
A short study on the representation of Gravitational Waves Data for Convolutional Neural Network	110
<i>M. Grespan</i>	
Search for Microlensing Signature in Gravitational Waves from Binary Black Hole Events	113
<i>Kyungmin Kim</i>	
Deep learning and numerical simulations to infer the evolution of MaNGA galaxies	116
<i>Regina Sarmiento, Johan H. Knapen, Marc Huertas-Company, Annalisa Pillepich, Sebastián F. Sánchez, Héctor Ibarra-Medel and Eduardo Lacerda</i>	
Data pre-extraction for better classification of galaxy mergers	119
<i>W. J. Pearson, L. E. Suelves, NEP Team and GAMA Team</i>	
Stellar spectra classification and clustering using deep learning	122
<i>Tomasz Rózański</i>	
Is GMM effective in membership determination of Open Clusters?	125
<i>S. N. Hasan, Md Mahmudunnobe, Priya Hasan and Mudasir Raja</i>	
Deep Radio Image Segmentation	127
<i>Hattie Stewart, Mark Birkinshaw and Jason Yeung</i>	
Computational techniques for high energy astrophysics and medical image processing	129
<i>Nicolás Vásquez, Jennifer Ortega, David Erazo, Ricardo Caiza and Orlando Gutiérrez</i>	
Deep learning proves to be an effective tool for detecting previously undiscovered exoplanets in Kepler data	132
<i>Amelia M. Yu</i>	
Author Index	137

Preface

Symposium 368 of the IAU GA in Busan on “Machine Learning - Possibilities and Pitfalls” was held from Aug 2 to 4 2022. The theme revolved around machine learning in astronomy, in particular the potential it has created to solve many open challenges, but also the possible problems that can be encountered when applying these often black-boxy techniques without proper care.

In the symposium there were eight invited and 13 contributed talks. There were also four panel discussions during which 15 total panelists shared the stage. For want of time another ~ 70 abstracts were converted to virtual talks and posters available throughout the meeting.

The symposium started with two detailed tutorials: “ML tutorial for the broader community” and “Classic Machine Learning vs Deep Learning: when, why and how?”. These provided a gentle introduction to those who were not steeped in the details of machine learning (ML), giving them the impetus to be able to follow the other talks later. For those already well-versed they served as a quick revision.

The plenary talks addressed the trends and challenges in deep learning, addressing how deep learning contrasts with shallow learning, the challenges of explainability and interpretability as more deep learning techniques are embraced, and the path to collaborative Human-AI learning addressing how with modern ML/AI methods for the first time we are starting to discover patterns from data-driven approaches and how this leads to areas not investigated before.

Other invited talks included reviews of various areas, including an overview of machine learning in astronomy, existing and forthcoming large and diverse datasets that form staple inputs to ML algorithms, the role of physics-informed machine learning in exoplanet characterization, and citizen science aspects for creating labeled datasets and for outreach and education.

The discussion sessions covered various aspects of the advances in ML, and the possible unintended misuse thereof. The first one discussed the breadth of ML applications in astronomy, highlighting the fact that for many problems classical ML techniques may be the first refuge, but also that recent improvements in techniques mean that one does not need huge datasets to do effective machine learning. Another discussed the important topic about combining diverse datasets, in particular archival datasets with real-time data. A third took on practical problem solving, including interpretability. Finally, the fourth discussion handled more recent trends in multi-messenger astronomy. These open discussions saw strong participation by students and postdocs.

The contributed talks - chosen from a large pool of abstracts - provided more extensive coverage of the field, touching upon more surveys, outliers and anomalies, unsupervised methods, incorporation of errors, non-Gaussianity, Generative Adversarial Networks, etc. The speakers included not only senior researchers but also undergraduate and graduate students. Many of the topics the invited and contributed talks did not cover were covered by the virtual talks and posters, providing a well-rounded symposium with science topics ranging from our neighborhood to cosmology.

The symposium received overwhelming support with close to 450 astronomers wanting to attend the symposium in person, and about 100 submitting abstracts for presentations. We tried to select representative presentations from across the spectrum, with the talks covering theory, simulations, applications, interpretations and on datasets from the Solar System to Galactic to extragalactic astronomy to cosmology. There were many deserving abstracts that had to be converted to virtual talks and posters.

We ensured that there was ample time for discussion through various panels on topics like Gravitational Wave/Multi-Messenger Astronomy, broader ML, and fusion of large datasets, in which all attendees were able to participate. For example, the challenges that ML is facing include the lack of interpretability and explainability. On another level, not many techniques allow for proper uncertainty quantification. While we did not expect the symposium to solve these big problems, we did manage to grow awareness amongst a larger set of users and practitioners of the issues, moving the needle towards better ML practices.

The broad hope and expectation was that attendees gain exposure to the breadth of available datasets and techniques, and the expertise on display during the symposium allows them to step out of their comfort zone to take on bigger problems while leveraging best practices. In particular, we hope to see more population studies incorporating more publicly available datasets (including transfer learning across datasets) rather than specialized studies involving smaller private and proprietary datasets, and we believe the symposium has helped the field move in that direction.

Since the pandemic had not yet abated, and the corresponding health safety measures were still in place, we likely lost many interested attendees (even from the 15 SOC members, only five could attend in person). Also, there were many sessions in parallel further dividing the attendees. Yet, throughout the symposium over 100 attendees were present plus many online (the exact number was not reported). There were a few complaints from those online about not being able to follow everything in the room - this is the unfortunate reality of hybrid meetings.

There were ample opportunities for discussions which the attendees took advantage of. There was dedicated time for poster viewing; since these were virtual posters it is likely that they got less visibility during the meeting from those in person, however, having the posters available to view online for an extended period may have resulted in more overall visibility in the field.

Overall the symposium set things up well to invite novices into the world of machine learning with datasets large and small, and provided the experts with more fodder to explore new problems, datasets, and techniques while providing all attendees with the resource to take steps towards avoiding the associated pitfalls.

Editors

Jess McIver
University of British Columbia, Canada

Ashish Mahabal
California Institute of Technology, USA

Christopher Fluke
Swinburne University of Technology, Australia

Scientific Organising Committee

Ashish Mahabal, California Institute of Technology, USA
Christopher Fluke, Swinburne University of Technology, Australia
Tara Murphy, University of Sydney, Australia
Jess McIver, University of British Columbia, Canada
Arman Shafieloo, Korea Astronomy and Space Science Institute, Republic of Korea
Andrew J Connolly, University of Washington, USA
Dalya Baron, Carnegie Observatories, USA
Gwendolyn Eadie, University of Toronto, Canada
Francisco Förster Burón, Universidad de Chile, Chile
Kai Polsterer, Heidelberg Institute for Theoretical Studies, Heidelberg, Germany
Pavlos Protopapas, Harvard School of Engineering, Cambridge, USA
Nadia Blagorodnova, University of Barcelona, Spain
David Parkinson, Korea Astronomy and Space Science Institute, Republic of Korea
Vanessa McBride, Office of Astronomy for Development, Cape Town, South Africa
G.C. Anupama, Indian Institute of Astrophysics, India

Local Organising Committee

Arman Shafieloo, Korea Astronomy and Space Science Institute, Republic of Korea
David Parkinson, Korea Astronomy and Space Science Institute, Republic of Korea
Ashish Mahabal, California Institute of Technology, USA

List of Participants

Aisha Alowais
Alisher S. Hojaev
Amelia Yu
Annalisa Pillepich
Anupam Bhardwaj
Arman Shafieloo
Ashish Mahabal
Atila Poro
Atrideb Chatterjee
Bekdaulet Shukirgaliyev
Boon Kiat Oh
Connor Bottrell
Cristiano Sabiu
Da Eun Kang
David Parkinson
Dennis Crake
Didier Fraix-Burnet
Djordje Savic
Dominic Adams
Dongjin Lee
Eric Ford
Erik Rodrigues de Lima
Eunsu Park
Garreth Martin
George Djorgovski
Georgios Vernardos
Gor Mikayelyan
Gordian Edenhofer
Guillermo Torralba Elipe
Guillermo Cabrera
Hanwool Koo
Ignacio Ferreras
Jaehan Bae
Jakob Knollmüller
Jeroen Audenaert
Jinkoo Yim
Joeri Van Leeuwen
Johan Knapen
John Suárez-Pérez
Josef Durech
Joshua Speagle
Juan Rafael Martinez Galarza
Kai Polsterer
Krisztián Vida
Kyungmin Kim
Laurent Eyer
Lilianne Nakazono
Lukasz Wyrzykowski

Marc Huertas-Company
Marco Álvarez
Margherita Grespan
Maria Luisa Buzzo
Marian Douspis
Mark Birkinshaw
Médéric Boquien
Melissa Lopez
Michel Cure
Michelle Lochner
Mike Walmsley
Monika Soraisam
Nadejda Blagorodnova Mujortova
Nahathai Tanakul
Nicolás Vásquez
O. Ivy Wong
Ofer Lahav
Oliver Boersma
Pablo Manuel Sánchez Alarcón
Priya Shah
Rania Hamdani
Raquel Ruiz Valença
Renee Hlozek
Rodrigo Calderon
Ryan Keeley
Ryohtaroh Ishikawa
Syed Najamul Hasan
Sandor Pinter
Sara Webb
Seong-Heon Lee
Sepideh Ghaziasgar
Shay Zucker
Simon Anghel
Sirinrat Sithajan
Stefan Wagner
Steffani Grondin
Suhyun Shin
Suk Yee Yong
Sumi Kim
Sungwook E Hong
Tomasz Rózański
Vanessa McBride
Vishal Upendran
Vitor Cernic
William Pearson
Wuhyun Sohn
Yuan-Sen Ting