Characterizing Stellar Streams with Error-Aware Machine Learning

Alexandros Pratsos[1], **Biprateep Dey**[1,2,*], **Ting S. Li**[1]

[1]University of Toronto [2]Vector Institute

Abstract

Stellar streams are thin, elongated collections of stars formed by gravitational disruption of orbiting star clusters or dwarf galaxies and are highly sensitive probes of the Milky Way's dark matter distribution and formation history. We present SCREAM (**S**tream **C**ha**R**acterization with **E**rror **A**ware **M**achine Learning), a weakly-supervised framework to identify member stars of stellar streams. Building on the CATHODE method originally developed for particle physics, SCREAM identifies streams as localized feature-space over-densities, avoiding rigid physical priors like assumed gravitational potentials or strict isochrone filtering. Crucially, SCREAM is the first machine learning (ML) framework in this domain to directly incorporate observational uncertainties into the neural network training objective. Using astrometric and photometric data from Gaia Data Release 3 and the Dark Energy Spectroscopic Instrument (DESI) Legacy imaging survey, we demonstrate our algorithm's performance on the prominent GD-1 stream. Validated against independent labels, SCREAM achieves an F1 score of 0.745, substantially outperforming existing ML methods in both precision and recall. Furthermore, SCREAM recovers the physically expected diffuse "cocoon" of GD-1 and faint main-sequence members that classical physics-based algorithms (e.g., STREAMFINDER) miss. Our results highlight the transformative potential of uncertainty-aware, weakly-supervised ML to uncover complex galactic structures.

1 Introduction

Stellar streams are narrow, extended structures formed when gravitationally bound progenitors (e.g., globular clusters, dwarf galaxies) are disrupted by their host's gravitational field (Johnston, 1998). The Milky Way hosts over a hundred known streams, typically containing $\mathcal{O}(100 - 1000)$ observable stars (Mateu, 2023). Their physical narrowness and low velocity dispersions make them sensitive probes of dark matter interactions within the galactic halo (Lynden-Bell and Lynden-Bell, 1995).

Stream identification from large astronomical surveys has historically relied on both physics-based and data-driven methods (Bonaca and Price-Whelan, 2025). While approaches like STREAMFINDER (Malhan and Ibata, 2018) are highly successful, they depend on assumed models of the Milky Way's gravitational potential and rigid stellar evolution priors. To avoid these dependencies, recent frameworks such as VIA MACHINAE (Shih et al., 2022, 2024; Hallin et al., 2025), Pettee et al. (2024), and SKYCURTAINS (Sengupta et al., 2025) treat stream detection as weakly-supervised anomaly detection. They adapt particle physics techniques like ANODE (Nachman and Shih, 2020), CWoLA (Metodiev et al., 2017), and CATHODE (Hallin et al., 2022) to model streams as localized over-densities in the feature space. However, these methods often ignore uncertainties in the input measurements, employ heavy post-processing (e.g., line-finding algorithms), and focus primarily on identifying new stream candidates over finding member stars of an already known stream (also called characterization).

*Corresponding author: Biprateep Dey (biprateep@pitt.edu)

Conference on Physics and AI at Stanford University (PAI 2026).

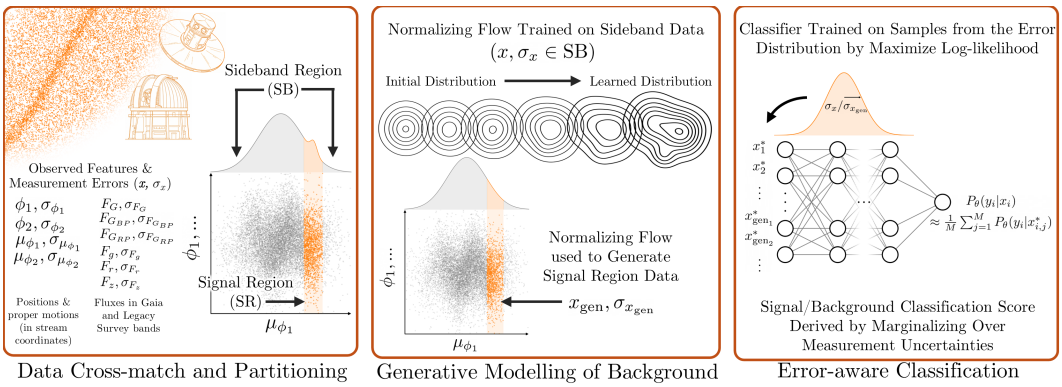

Figure 1: **Schematic representation of the SCREAM methodology**. **Left (Data Cross-match and Partitioning):** Kinematic and photometric features (x) and associated measurement uncertainties (σ_x), are compiled from Gaia and the DESI Legacy Survey. Data is partitioned into a Signal Region (SR) and a background-dominated Sideband Region (SB) based on along-stream proper motion (μ_{ϕ_1}). **Middle (Generative Modeling of Background):** A conditional Normalizing Flow trained on the SB data learns the joint distribution of background features and errors. It acts as an interpolator to generate realistic background data ($x_{\text{gen}}, \sigma_{x_{\text{gen}}}$) in the SR. **Right (Error-aware Classification):** A neural network classifier distinguishes true SR data from synthetic background. The training marginalizes over observational uncertainties by Monte Carlo sampling from the Gaussian error distributions, yielding a robust classification score.

We present SCREAM (**S**tream **C**ha**R**acterization with **E**rror **A**ware **M**achine learning), an anomaly detection framework designed to account for observational uncertainties in stellar stream searches and characterization. Building on CATHODE, SCREAM marginalizes over feature uncertainties within the loss function to more accurately capture the underlying distribution. Our approach requires minimal physical priors, accommodates arbitrary datasets, integrates uncertainties during training, and eliminates post-processing steps. This preserves anomaly score statistics and enables the discovery of members that rigid models might miss. While extensible to blind detection, this work focuses on characterizing known streams. We demonstrate SCREAM on the GD-1 stream (Grillmair and Dionatos, 2006), validating predictions against independent reliable labels (Jarvis et al., 2026) to confirm the algorithm's ability to recover the true stellar stream distribution.

2 Data and Methods

Data Sets Used. We utilize astrometric and photometric data from Gaia Data Release 3 (DR3) (Gaia Collaboration et al., 2023), cross-matched with the DESI Legacy Imaging Survey DR9 (Dey et al., 2019) for additional photometric features. We query objects present in both catalogs within a 3° padding of the GD-1 stream track defined in `galstreams` (Mateu, 2023). Gaia provides sky positions and proper motions (angular velocities) transformed into a local, stream-aligned coordinate frame: ϕ_1 aligns with the projected path, ϕ_2 is the perpendicular transverse angle, and μ_{ϕ_1} and μ_{ϕ_2} denote their respective proper motions. To complement these kinematic data, we include broadband flux measurements (F_G, $F_{G_{BP}}$, $F_{G_{RP}}$, F_g, F_r, F_z), corrected for Milky Way dust extinction (Schlegel et al., 1998). Finally, we explicitly incorporate the observational uncertainties (denoted as σ with the corresponding subscript) for all features directly into our framework.

The CATHODE Framework. We build our algorithm by extending CATHODE (Hallin et al., 2022), a weakly-supervised anomaly detection method originally developed to identify anomalous populations in high-dimensional particle physics data. We first partition the data into two non-random subsets, a signal region (SR) and its complement, the sideband region (SB), using the value of a single feature. We define SR using the along-stream proper motion, μ_{ϕ_1}, as streams are tightly kinematically constrained in this dimension. For GD-1, we define the SR as the 5th to 95th percentile range of μ_{ϕ_1} from the STREAMFINDER catalog. Every observation falls into either the SR or SB based strictly on its μ_{ϕ_1} value.

Next, we perform conditional density estimation of the background in the SB, assuming $P_{\text{bg}}(\mathbf{x}, \sigma_{\mathbf{x}}|\mu_{\phi_1} \in \text{SB}) \approx P_{\text{data}}(\mathbf{x}, \sigma_{\mathbf{x}}|\mu_{\phi_1} \in \text{SB})$. We implement this via conditional Normal-

izing Flows (NF) using `pzflow` (Crenshaw et al., 2024) with Rational-Quadratic Rolling Spline Couplings (Dinh et al., 2014; Durkan et al., 2019). We append measurement uncertainties to the NF's input space, jointly modeling features and errors so generated backgrounds possess realistic uncertainties. Leveraging the NF's interpolation capabilities, we generate background-like data within the SR. We fit a Kernel Density Estimate (KDE) to the μ_{ϕ_1} distribution of the observed SR data, sample $\mu_{\phi_1} \sim P_{\mathrm{KDE}}(\mu_{\phi_1})$, and apply the inverse NF transformation to generate synthetic background features $(\mathbf{x}_{\mathrm{gen}}, \sigma_{\mathbf{x}_{\mathrm{gen}}})$. To prevent class imbalance, we generate the same number of samples as the true SR observations.

A neural network then classifies observed versus generated background data in the SR. Since the generated data models the background, the classifier cannot differentiate the two without a signal. Thus, any learned separation is directly attributable to the stellar stream signal. Although these class probabilities are miscalibrated, they preserve a monotonic relationship with the true underlying signal probabilities (Metodiev et al., 2017).

Extending `CATHODE` to Incorporate Measurement Uncertainties. When training a standard classifier, the objective is to maximize the likelihood of the true labels given deterministic inputs. In `SCREAM`, we treat the observed features $\mathbf{x}_i$ for an object i as a single realization drawn from a distribution centered around the true, unobserved features $\mathbf{x}_i^{\mathrm{true}}$. The likelihood for a single observation is:

$$P_\theta(y_i|\mathbf{x}_i) = \int P_\theta(y_i|\mathbf{x}_i^{\mathrm{true}}) P(\mathbf{x}_i^{\mathrm{true}}|\mathbf{x}_i)\, d\mathbf{x}_i^{\mathrm{true}}. \tag{1}$$

Using Bayes' theorem and assuming a uniform prior over the true features, we obtain:

$$P_\theta(y_i|\mathbf{x}_i) \propto \int P_\theta(y_i|\mathbf{x}_i^{\mathrm{true}}) P(\mathbf{x}_i|\mathbf{x}_i^{\mathrm{true}})\, d\mathbf{x}_i^{\mathrm{true}}, \tag{2}$$

where $P(\mathbf{x}_i|\mathbf{x}_i^{\mathrm{true}})$ is the distribution from which our observed features are drawn, modeled here as an uncorrelated multidimensional Gaussian. Exploiting the symmetry of the Gaussian distribution, we approximate this analytically intractable integral using Monte Carlo (MC) integration. We draw M feature samples, $\mathbf{x}_{i,j}^*$, from the noise distribution centered on the observation $\mathbf{x}_i$ with a diagonal covariance matrix defined by $\sigma_{\mathbf{x}_i}$. The integral is then approximated by the summation:

$$P_\theta(y_i|\mathbf{x}_i) \approx \frac{1}{M} \sum_{j=1}^{M} P_\theta(y_i|\mathbf{x}_{i,j}^*). \tag{3}$$

For a binary classification problem, $P_\theta(y_i|\mathbf{x}_{i,j}^*)$ follows a Bernoulli distribution. Therefore, the approximated negative log-likelihood (loss) for a single observation is:

$$\mathcal{L}_{\mathrm{MC}} = -\ln\left(\frac{1}{M}\sum_{j=1}^{M} p_{i,j}^{y_i}(1-p_{i,j})^{1-y_i}\right). \tag{4}$$

where, $p_{i,j}$ is the predicted class probability. This loss is averaged over the training set and minimized to find the optimal model parameters. Though novel for this domain, a similar approach to incorporate input measurement uncertainties have been used in other contexts in previous work such as Wang et al. (2023) and Sun et al. (2023). A schematic overview of the whole algorithm is shown in Figure 1.

Our classification network is a Multi-Layer Perceptron (Rosenblatt, 1958) with GELU non-linearities (Hendrycks and Gimpel, 2016), optimized via Adam (Kingma and Ba, 2015). Full architectural details are in our code repository[2]. To increase stochastic regularization, we set $M = 10$ during training. During inference, we draw $M = 100$ samples and average them to obtain the classification score for each object. Predictions for the entire data set are generated by $k = 5$-fold cross-validation.

3 Results and Discussion

We evaluate the performance of `SCREAM` against two distinct reference datasets: `STREAMFINDER` (SF), serving as a proxy for an algorithmic baseline and the highly reliable labels derived from DESI radial velocity measurements from Jarvis et al. (2026).

[2] `https://github.com/pratsosa/SCREAM`

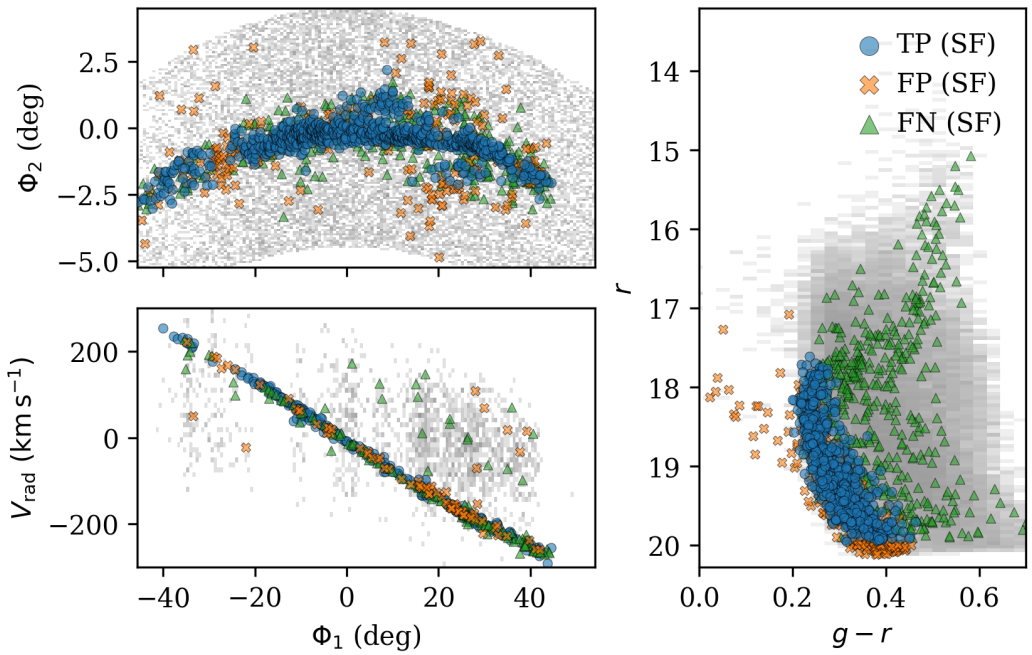

Figure 2: **Visualization of SCREAM performance relative to the STREAMFINDER (SF) baseline.** The background distribution is shown in gray. **Top Left (ϕ_2 vs. ϕ_1):** SCREAM recovers stars along the full spatial extent of GD-1. Many apparent False Positives (orange crosses) located off the narrow central track are confirmed members via independent radial velocity labels. **Right (r vs. $g-r$):** Predictions trace the dense main sequence. While missing brighter RGB stars ($r < 18.5$), SCREAM recovers more faint main sequence stars than SF. **Bottom Left ($V_{\rm rad}$ vs. ϕ_1):** SCREAM False Positives align with the true kinematic track, demonstrating high precision identifying genuine members SF missed. False Negatives (green triangles) indicate recall can be improved.

At a 0.878 confidence threshold, SCREAM achieves an F1 score of 0.66 (Precision: 0.69, Recall: 0.62) against SF labels. Evaluated against true spectroscopic labels, performance improves to an F1 of 0.745 (Accuracy: 0.959, Precision: 0.871, Recall: 0.650). This indicates our approach captures the true physical distribution more accurately than the SF baseline. While differing datasets complicate direct comparisons with other machine learning-based methods, SCREAM outperforms algorithms like SKYCURTAINS (Sengupta et al., 2025) by 12 percentage points in precision and 28 in recall, while also accounting for input measurement uncertainties in the classifications.

Figure 2 visualizes the algorithm performance with respect to STREAMFINDER. Spatially (top-left), SCREAM uniquely recovers the stream's diffuse "cocoon", a physically expected feature often missed by SF's rigid priors. Thus, many apparent False Positives relative to SF (orange crosses) are genuine members. Photometrically (right), minimizing physical priors allows SCREAM to recover fainter member stars that SF misses. However, the model struggles with brighter Red Giant Branch (RGB) stars as their lower feature-space density causes the anomaly detector to miss them. Kinematically (bottom-left), apparent False Positives lie almost entirely on the true kinematic track, reinforcing high true precision, though False Negatives on the track highlight room for recall improvement.

To our knowledge, SCREAM is among the first ML-based stellar stream identification and characterization frameworks to incorporate observational uncertainties directly into the training objective. Because SF labels were never used during training, SCREAM is not restricted by previous assumptions. Future work will address current limitations. We will exploit the model's probabilistic nature to handle missing data and feature scaling and density estimation improvements will enable the detection of low-density RGB stars. A thorough vetting of our algorithm on mock data will also be performed in a future work. Finally, implementing a scanning signal-region window and sampling from full, non-diagonal error covariance matrices will enable blind, all-sky searches and improved classification across multiple Milky Way streams.

Acknowledgments and Disclosure of Funding

The authors acknowledge the support of the Data Sciences Institute, University of Toronto via the 2025 Summer Undergraduate Data Science (SUDS) Opportunities Program. B. Dey is a postdoctoral fellow at the University of Toronto in the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship Program, a program of Schmidt Sciences.

The authors advocate for the judicious and ethical use of artificial intelligence in scientific discourse. In accordance with the American Physical Society's guidelines on appropriate use of AI-based writing tools, we confirm that AI assistance was utilized in the preparation of this statement solely to polish, condense, and edit our original ideas. The authors retain full responsibility for the accuracy, integrity, and content of this document.

This research used data obtained with the Dark Energy Spectroscopic Instrument (DESI). DESI construction and operations is managed by the Lawrence Berkeley National Laboratory. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High-Energy Physics, under Contract No. DE–AC02–05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract. Additional support for DESI was provided by the U.S. National Science Foundation (NSF), Division of Astronomical Sciences under Contract No. AST-0950945 to the NSF's National Optical-Infrared Astronomy Research Laboratory; the Science and Technology Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the French Alternative Energies and Atomic Energy Commission (CEA); the National Council of Humanities, Science and Technology of Mexico (CONAHCYT); the Ministry of Science and Innovation of Spain (MICINN), and by the DESI Member Institutions: www.desi.lbl.gov/collaborating-institutions. The DESI collaboration is honored to be permitted to conduct scientific research on I'oligam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. National Science Foundation, the U.S. Department of Energy, or any of the listed funding agencies.

The Legacy Surveys consist of three individual and complementary projects: the Dark Energy Camera Legacy Survey (DECaLS; Proposal ID #2014B-0404; PIs: David Schlegel and Arjun Dey), the Beijing-Arizona Sky Survey (BASS; NOAO Prop. ID #2015A-0801; PIs: Zhou Xu and Xiaohui Fan), and the Mayall z-band Legacy Survey (MzLS; Prop. ID #2016A-0453; PI: Arjun Dey). DECaLS, BASS and MzLS together include data obtained, respectively, at the Blanco telescope, Cerro Tololo Inter-American Observatory, NSF's NOIRLab; the Bok telescope, Steward Observatory, University of Arizona; and the Mayall telescope, Kitt Peak National Observatory, NOIRLab. Pipeline processing and analyses of the data were supported by NOIRLab and the Lawrence Berkeley National Laboratory (LBNL). The Legacy Surveys project is honored to be permitted to conduct astronomical research on Iolkam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation.

NOIRLab is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation. LBNL is managed by the Regents of the University of California under contract to the U.S. Department of Energy.

This project used data obtained with the Dark Energy Camera (DECam), which was constructed by the Dark Energy Survey (DES) collaboration. Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundacao Carlos Chagas Filho de Amparo, Financiadora de Estudos e Projetos, Fundacao Carlos Chagas Filho de Amparo a Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Cientifico e Tecnologico and the Ministerio da Ciencia, Tecnologia e Inovacao, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energeticas, Medioambientales

y Tecnologicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenossische Technische Hochschule (ETH) Zurich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciencies de l'Espai (IEEC/CSIC), the Institut de Fisica d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig Maximilians Universitat Munchen and the associated Excellence Cluster Universe, the University of Michigan, NSF's NOIRLab, the University of Nottingham, the Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University.

BASS is a key project of the Telescope Access Program (TAP), which has been funded by the National Astronomical Observatories of China, the Chinese Academy of Sciences (the Strategic Priority Research Program "The Emergence of Cosmological Structures" Grant # XDB09000000), and the Special Fund for Astronomy from the Ministry of Finance. The BASS is also supported by the External Cooperation Program of Chinese Academy of Sciences (Grant # 114A11KYSB20160057), and Chinese National Natural Science Foundation (Grant # 12120101003, # 11433005).

The Legacy Survey team makes use of data products from the Near-Earth Object Wide-field Infrared Survey Explorer (NEOWISE), which is a project of the Jet Propulsion Laboratory/California Institute of Technology. NEOWISE is funded by the National Aeronautics and Space Administration.

The Legacy Surveys imaging of the DESI footprint is supported by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under Contract No. DE-AC02-05CH1123, by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract; and by the U.S. National Science Foundation, Division of Astronomical Sciences under Contract No. AST-0950945 to NOAO.

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (`https://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, `https://www.cosmos.esa.int/web/gaia/dpac/consortium`). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

References

Ana Bonaca and Adrian M. Price-Whelan. Stellar streams in the Gaia era. New A Rev., 100:101713, June 2025. doi: 10.1016/j.newar.2024.101713.

John Franklin Crenshaw, J. Bryce Kalmbach, Alexander Gagliano, Ziang Yan, Andrew J. Connolly, Alex I. Malz, Samuel J. Schmidt, and The LSST Dark Energy Science Collaboration. Probabilistic Forward Modeling of Galaxy Catalogs with Normalizing Flows. AJ, 168(2):80, August 2024. doi: 10.3847/1538-3881/ad54bf.

Arjun Dey, David J. Schlegel, Dustin Lang, Robert Blum, Kaylan Burleigh, Xiaohui Fan, Joseph R. Findlay, Doug Finkbeiner, David Herrera, Stéphanie Juneau, Martin Landriau, Michael Levi, Ian McGreer, Aaron Meisner, Adam D. Myers, John Moustakas, Peter Nugent, Anna Patej, Edward F. Schlafly, Alistair R. Walker, Francisco Valdes, Benjamin A. Weaver, Christophe Yèche, Hu Zou, Xu Zhou, Behzad Abareshi, T. M. C. Abbott, Bela Abolfathi, C. Aguilera, Shadab Alam, Lori Allen, A. Alvarez, James Annis, Behzad Ansarinejad, Marie Aubert, Jacqueline Beechert, Eric F. Bell, Segev Y. BenZvi, Florian Beutler, Richard M. Bielby, Adam S. Bolton, César Briceño, Elizabeth J. Buckley-Geer, Karen Butler, Annalisa Calamida, Raymond G. Carlberg, Paul Carter, Ricard Casas, Francisco J. Castander, Yumi Choi, Johan Comparat, Elena Cukanovaite, Timothée Delubac, Kaitlin DeVries, Sharmila Dey, Govinda Dhungana, Mark Dickinson, Zhejie Ding, John B. Donaldson, Yutong Duan, Christopher J. Duckworth, Sarah Eftekharzadeh, Daniel J. Eisenstein, Thomas Etourneau, Parker A. Fagrelius, Jay Farihi, Mike Fitzpatrick, Andreu Font-Ribera, Leah Fulmer, Boris T. Gänsicke, Enrique Gaztanaga, Koshy George, David W. Gerdes, Satya Gontcho A. Gontcho, Claudio Gorgoni, Gregory Green, Julien Guy, Diane Harmer, M. Hernandez, Klaus Honscheid, Lijuan Wendy Huang, David J. James, Buell T. Jannuzi, Linhua Jiang, Richard Joyce, Armin Karcher, Sonia Karkar, Robert Kehoe, Jean-Paul Kneib, Andrea Kueter-Young, Ting-Wen Lan, Tod R. Lauer, Laurent Le Guillou, Auguste Le Van Suu, Jae Hyeon Lee, Michael Lesser, Laurence Perreault Levasseur, Ting S. Li, Justin L. Mann, Robert Marshall, C. E. Martínez-Vázquez, Paul Martini, Hélion du Mas des Bourboux, Sean McManus, Tobias Gabriel Meier, Brice Ménard, Nigel Metcalfe, Andrea Muñoz-Gutiérrez, Joan Najita, Kevin Napier, Gautham Narayan,

Jeffrey A. Newman, Jundan Nie, Brian Nord, Dara J. Norman, Knut A. G. Olsen, Anthony Paat, Nathalie Palanque-Delabrouille, Xiyan Peng, Claire L. Poppett, Megan R. Poremba, Abhishek Prakash, David Rabinowitz, Anand Raichoor, Mehdi Rezaie, A. N. Robertson, Natalie A. Roe, Ashley J. Ross, Nicholas P. Ross, Gregory Rudnick, Sasha Safonova, Abhijit Saha, F. Javier Sánchez, Elodie Savary, Heidi Schweiker, Adam Scott, Hee-Jong Seo, Huanyuan Shan, David R. Silva, Zachary Slepian, Christian Soto, David Sprayberry, Ryan Staten, Coley M. Stillman, Robert J. Stupak, David L. Summers, Suk Sien Tie, H. Tirado, Mariana Vargas-Magaña, A. Katherina Vivas, Risa H. Wechsler, Doug Williams, Jinyi Yang, Qian Yang, Tolga Yapici, Dennis Zaritsky, A. Zenteno, Kai Zhang, Tianmeng Zhang, Rongpu Zhou, and Zhimin Zhou. Overview of the DESI Legacy Imaging Surveys. AJ, 157(5):168, May 2019. doi: 10.3847/1538-3881/ab089d.

Laurent Dinh, David Krueger, and Yoshua Bengio. NICE: Non-linear Independent Components Estimation. *arXiv e-prints*, art. arXiv:1410.8516, October 2014. doi: 10.48550/arXiv.1410.8516.

Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural Spline Flows. *arXiv e-prints*, art. arXiv:1906.04032, June 2019. doi: 10.48550/arXiv.1906.04032.

Gaia Collaboration, A. Vallenari, A. G. A. Brown, T. Prusti, J. H. J. de Bruijne, F. Arenou, C. Babusiaux, M. Biermann, O. L. Creevey, C. Ducourant, D. W. Evans, L. Eyer, R. Guerra, A. Hutton, C. Jordi, S. A. Klioner, U. L. Lammers, L. Lindegren, X. Luri, F. Mignard, C. Panem, D. Pourbaix, S. Randich, P. Sartoretti, C. Soubiran, P. Tanga, N. A. Walton, C. A. L. Bailer-Jones, U. Bastian, R. Drimmel, F. Jansen, D. Katz, M. G. Lattanzi, F. van Leeuwen, J. Bakker, C. Cacciari, J. Castañeda, F. De Angeli, C. Fabricius, M. Fouesneau, Y. Frémat, L. Galluccio, A. Guerrier, U. Heiter, E. Masana, R. Messineo, N. Mowlavi, C. Nicolas, K. Nienartowicz, F. Pailler, P. Panuzzo, F. Riclet, W. Roux, G. M. Seabroke, R. Sordo, F. Thévenin, G. Gracia-Abril, J. Portell, D. Teyssier, M. Altmann, R. Andrae, M. Audard, I. Bellas-Velidis, K. Benson, J. Berthier, R. Blomme, P. W. Burgess, D. Busonero, G. Busso, H. Cánovas, B. Carry, A. Cellino, N. Cheek, G. Clementini, Y. Damerdji, M. Davidson, P. de Teodoro, M. Nuñez Campos, L. Delchambre, A. Dell'Oro, P. Esquej, J. Fernández-Hernández, E. Fraile, D. Garabato, P. García-Lario, E. Gosset, R. Haigron, J.-L. Halbwachs, N. C. Hambly, D. L. Harrison, J. Hernández, D. Hestroffer, S. T. Hodgkin, B. Holl, K. Janßen, G. Jevardat de Fombelle, S. Jordan, A. Krone-Martins, A. C. Lanzafame, W. Löffler, O. Marchal, P. M. Marrese, A. Moitinho, K. Muinonen, P. Osborne, E. Pancino, T. Pauwels, A. Recio-Blanco, C. Reylé, M. Riello, L. Rimoldini, T. Roegiers, J. Rybizki, L. M. Sarro, C. Siopis, M. Smith, A. Sozzetti, E. Utrilla, M. van Leeuwen, U. Abbas, P. Ábrahám, A. Abreu Aramburu, C. Aerts, J. J. Aguado, M. Ajaj, F. Aldea-Montero, G. Altavilla, M. A. Álvarez, J. Alves, F. Anders, R. I. Anderson, E. Anglada Varela, T. Antoja, D. Baines, S. G. Baker, L. Balaguer-Núñez, E. Balbinot, Z. Balog, C. Barache, D. Barbato, M. Barros, M. A. Barstow, S. Bartolomé, J.-L. Bassilana, N. Bauchet, U. Becciani, M. Bellazzini, A. Berihuete, M. Bernet, S. Bertone, L. Bianchi, A. Binnenfeld, S. Blanco-Cuaresma, A. Blazere, T. Boch, A. Bombrun, D. Bossini, S. Bouquillon, A. Bragaglia, L. Bramante, E. Breedt, A. Bressan, N. Brouillet, E. Brugaletta, B. Bucciarelli, A. Burlacu, A. G. Butkevich, R. Buzzi, E. Caffau, R. Cancelliere, T. Cantat-Gaudin, R. Carballo, T. Carlucci, M. I. Carnerero, J. M. Carrasco, L. Casamiquela, M. Castellani, A. Castro-Ginard, L. Chaoul, P. Charlot, L. Chemin, V. Chiaramida, A. Chiavassa, N. Chornay, G. Comoretto, G. Contursi, W. J. Cooper, T. Cornez, S. Cowell, F. Crifo, M. Cropper, M. Crosta, C. Crowley, C. Dafonte, A. Dapergolas, M. David, P. David, P. de Laverny, F. De Luise, and R. De March. Gaia Data Release 3. Summary of the content and survey properties. A&A, 674:A1, June 2023. doi: 10.1051/0004-6361/202243940.

C. J. Grillmair and O. Dionatos. Detection of a 63° Cold Stellar Stream in the Sloan Digital Sky Survey. ApJ, 643(1):L17–L20, May 2006. doi: 10.1086/505111.

Anna Hallin, Joshua Isaacson, Gregor Kasieczka, Claudius Krause, Benjamin Nachman, Tobias Quadfasel, Matthias Schlaffer, David Shih, and Manuel Sommerhalder. Classifying anomalies through outer density estimation. Phys. Rev. D, 106(5):055006, September 2022. doi: 10.1103/PhysRevD.106.055006.

Anna Hallin, David Shih, Claudius Krause, and Matthew R. Buckley. Via Machinae 3.0: A search for stellar streams in Gaia with the CATHODE algorithm. *arXiv e-prints*, art. arXiv:2509.08064, September 2025. doi: 10.48550/arXiv.2509.08064.

Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.

Emma Jarvis, Ting S. Li, Sergey E. Koposov, Raymond G. Carlberg, Monica Valluri, Nasser Mohammed, J. Aguilar, S. Ahlen, Carlos Allende Prieto, Leandro Beraldo e Silva, D. Bianchi, D. Brooks, Amanda Byström, T. Claybaugh, A. P. Cooper, A. Cuceu, A. de la Macorra, Arjun Dey, Biprateep Dey, P. Doel, J. E. Forero-Romero, E. Gaztañaga, Oleg Y. Gnedin, Satya Gontcho A Gontcho, G. Gutierrez, K. Honscheid, R. Joyce, R. Kehoe, T. Kisner, Namitha Kizhuprakkat, A. Kremin, Mika Lambert, M. Landriau, L. Le Guillou, Gustavo E. Medina, A. Meisner, R. Miquel, S. Nadathur, Joan Najita, N. Palanque-Delabrouille, W. J. Percival, F. Prada, I. Pérez-Ràfols, Tian Qiu, Alexander H. Riley, Constance M. Rockosi, G. Rossi, E. Sanchez, Nathan Sandford, E. F. Schlafly, D. Schlegel, J. Silber, D. Sprayberry, G. Tarlé, B. A. Weaver, R. Zhou, and H. Zou. Characterizing the GD-1 Stream with DESI DR2 Data: Thin Stream and Hot Cocoon. *arXiv e-prints*, art. arXiv:2604.20958, April 2026. doi: 10.48550/arXiv.2604.20958.

Kathryn V. Johnston. A Prescription for Building the Milky Way's Halo from Disrupted Satellites. ApJ, 495(1):297–308, March 1998. doi: 10.1086/305273.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015. URL `https://arxiv.org/abs/1412.6980`.

D. Lynden-Bell and R. M. Lynden-Bell. Ghostly streams from the formation of the Galaxy's halo. MNRAS, 275(2):429–442, July 1995. doi: 10.1093/mnras/275.2.429.

Khyati Malhan and Rodrigo A. Ibata. STREAMFINDER - I. A new algorithm for detecting stellar streams. MNRAS, 477(3):4063–4076, July 2018. doi: 10.1093/mnras/sty912.

Cecilia Mateu. galstreams: A library of Milky Way stellar stream footprints and tracks. MNRAS, 520(4):5225–5258, April 2023. doi: 10.1093/mnras/stad321.

Eric M. Metodiev, Benjamin Nachman, and Jesse Thaler. Classification without labels: learning from mixed samples in high energy physics. *Journal of High Energy Physics*, 2017(10):174, October 2017. doi: 10.1007/JHEP10(2017)174.

Benjamin Nachman and David Shih. Anomaly detection with density estimation. Phys. Rev. D, 101 (7):075042, April 2020. doi: 10.1103/PhysRevD.101.075042.

Mariel Pettee, Sowmya Thanvantri, Benjamin Nachman, David Shih, Matthew R. Buckley, and Jack H. Collins. Weakly supervised anomaly detection in the Milky Way. MNRAS, 527(3): 8459–8474, January 2024. doi: 10.1093/mnras/stad3663.

Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

David J. Schlegel, Douglas P. Finkbeiner, and Marc Davis. Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. ApJ, 500 (2):525–553, June 1998. doi: 10.1086/305772.

Debajyoti Sengupta, Stephen Mulligan, David Shih, John Andrew Raine, and Tobias Golling. SKY-CURTAINS: model-agnostic search for stellar streams with Gaia data. MNRAS, 536(2):1104–1114, January 2025. doi: 10.1093/mnras/stae2570.

David Shih, Matthew R. Buckley, Lina Necib, and John Tamanas. VIA MACHINAE: Searching for stellar streams using unsupervised machine learning. MNRAS, 509(4):5992–6007, February 2022. doi: 10.1093/mnras/stab3372.

David Shih, Matthew R. Buckley, and Lina Necib. VIA MACHINAE 2.0: Full-sky, model-agnostic search for stellar streams in Gaia DR2. MNRAS, 529(4):4745–4767, April 2024. doi: 10.1093/mnras/stae446.

Zechang Sun, Joshua S. Speagle, Song Huang, Yuan-Sen Ting, and Zheng Cai. Zephyr : Stitching Heterogeneous Training Data with Normalizing Flows for Photometric Redshift Inference. *arXiv e-prints*, art. arXiv:2310.20125, October 2023. doi: 10.48550/arXiv.2310.20125.

Bingjie Wang, Joel Leja, V. Ashley Villar, and Joshua S. Speagle. SBI^{++}: Flexible, Ultra-fast Likelihood-free Inference Customized for Astronomical Applications. ApJ, 952(1):L10, July 2023. doi: 10.3847/2041-8213/ace361.