universe

MDPI

Article

Exploring Late Stellar Evolution in the Era of Large Surveys: Machine Learning Prospects for Hot Subdwarfs and White Dwarfs

Princy Ranaivomanana [1] and Murat Uzundag [2,*]

[1] Department of Physics, University of Antananarivo, P.O. Box 906, Antananarivo 101, Madagascar; rtprincy@gmail.com
[2] Instituut voor Sterrenkunde, KU Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium
* Correspondence: muratuzundag.astro@gmail.com

Abstract

The rapid growth of large-scale astronomical surveys and advances in data-driven analysis techniques have transformed the study of late-stage stellar evolution. Modern facilities are producing large volumes of photometric, spectroscopic, and astrometric data, enabling systematic investigations of compact stellar populations across the Milky Way. Among the most important tracers of these advanced evolutionary phases are hot subdwarfs and white dwarfs: hot subdwarfs are core-helium-burning tracers of late, binary-driven stellar evolution, while white dwarfs represent the final evolutionary endpoint of low- and intermediate-mass stars. These compact objects provide important laboratories for studying stellar interiors, binary evolution, and the long-term fate of planetary systems. This paper explores how recent advances in machine learning are being applied to the detection, characterization, and, when combined with follow-up spectroscopy and modeling, the physical interpretation of hot subdwarfs and white dwarfs. By combining photometric, spectroscopic, and time-domain observations with these computational tools, it is now possible to efficiently discover rare objects, detect stellar variability, and probe the internal structure and evolutionary pathways of compact stars. Ultimately, these developments highlight the growing role of advanced algorithms in supporting the study of the final stages of stellar evolution, provided their outputs are validated against physical observables.

Keywords: stellar evolution; white dwarfs; hot subdwarfs; astronomical surveys; machine learning; time-domain astronomy; stellar variability

Check for updates

Academic Editor: Haijun Tian

Received: 15 June 2026
Revised: 17 August 2026
Accepted: 21 August 2026
Published: 26 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Stars are the fundamental building blocks of galaxies and play a central role in driving the physical and chemical evolution of the Universe. Through their life cycles, stars synthesize heavy elements, regulate the energetics of the interstellar medium, and ultimately shape the formation and evolution of planetary systems. Although the general framework of stellar evolution is well established, important uncertainties remain in our understanding of the internal physical processes that govern stellar structure and evolution. Key mechanisms, including rotation, internal mixing, magnetic fields, and chemical stratification, are particularly difficult to constrain observationally, especially during the late stages of stellar evolution when stars evolve rapidly toward compact remnants [1].

Over the past decade, astronomy has entered an unprecedented era of data-rich discovery driven by large photometric, spectroscopic, and astrometric surveys. Space missions such as Gaia [2] have revolutionized stellar astrophysics by providing precise distances,

photometry, and astrometry for more than a billion stars in the Milky Way. At the same time, ground-based spectroscopic programs including the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) [3], currently one of the largest spectroscopic surveys of stars, Sloan Digital Sky Survey-V (SDSS-V) [4], the William Herschel Telescope Enhanced Area Velocity Explorer (WEAVE) [5] and the 4-metre Multi-Object Spectroscopic Telescope (4MOST) [6] are delivering large-scale stellar spectroscopy across wide areas of the sky. In addition, the Legacy Survey of Space and Time, now underway at the Vera C. Rubin Observatory, is expected to dramatically expand the discovery space for variable and transient phenomena over the coming decade [7]. Together, these facilities are generating massive and heterogeneous datasets that enable systematic studies of stellar populations across the Galaxy, but also require new computational and statistical methods to extract meaningful physical insights.

Among the most important objects for understanding the final stages of stellar evolution are hot subdwarfs and white dwarfs. Hot subdwarfs (sdO/B stars) are low-mass, core-helium-burning stars located at the extreme blue end of the horizontal branch in the Hertzsprung–Russell (HR) diagram [8–10]. Their positions in an observational HR diagram (or color-magnitude diagram) are illustrated in Figure 1, using Gaia DR3 data for all objects within 1 kpc (see Appendix A.1 in [11] for data query details). They are widely believed to form primarily through binary evolution channels involving significant mass loss, such as stable Roche-lobe overflow, common-envelope ejection, or the merger of helium white dwarfs, although single-star formation channels (e.g., via enhanced mass loss on the red giant branch) have not been completely excluded and may account for a minority of the population [9]. These stars are also thought to be responsible for the ultraviolet excess observed in elliptical galaxies and provide important constraints on mass-loss processes during the red giant phase [12–14]. A subset of hot subdwarfs exhibits pulsations [15–17] that enable detailed probing of their interiors through asteroseismology, offering unique insights into stellar interiors and evolutionary pathways [18,19].

White dwarfs represent the final evolutionary stage of the vast majority of stars, including most stars with initial masses below roughly eight solar masses, accounting for more than 95% of the stellar population in the Milky Way [20]. As compact stellar remnants supported by electron degeneracy pressure, white dwarfs preserve valuable information about the evolutionary histories of their progenitor stars. Many white dwarfs exhibit photometric variability caused by pulsations, binarity, rotation, or the presence of orbiting debris from disrupted planetary bodies. Pulsating white dwarfs in particular serve as powerful laboratories for asteroseismology, enabling detailed investigations of their internal composition, thermal structure, and evolutionary cooling rates [21–23].

Although hot subdwarfs and white dwarfs differ substantially in evolutionary stage, formation channel, and observational selection effects, they are considered jointly in this work because they share a common set of data-analysis challenges: both are intrinsically rare compared to main-sequence contaminants and are easily confused with other source classes in photometric surveys. In addition, hot subdwarfs and white dwarfs increasingly rely on the same family of machine learning (ML) tools (automated candidate selection, variability classification, and parameter inference) to be efficiently identified and characterized in current and forthcoming large-scale surveys. Hot subdwarfs additionally represent a relevant evolutionary precursor to a fraction of white dwarfs formed through binary mass-loss channels, further motivating a combined discussion of their machine-learning-driven study.

Despite significant progress in stellar astrophysics, several fundamental questions remain unresolved. The formation channels of hot subdwarfs and the role of binary interactions in shaping their evolution remain active areas of investigation. Similarly, the interpretation of pulsation spectra in pulsating compact stars requires detailed theoretical modeling to

connect observed frequencies with internal structural properties. In addition, the discovery of planetary debris disks and metal-polluted atmospheres around many white dwarfs has opened a new window into the long-term fate of planetary systems after the death of their host stars. These systems provide valuable opportunities to study the composition and dynamical evolution of planetary material long after the main-sequence phase [24,25].

The large volume and diversity of data produced by modern astronomical surveys are reshaping the study of compact stellar remnants. Large-scale datasets from missions such as Gaia, combined with continuous observations from space-based photometric surveys, such as Kepler/K2 [26,27] and Transiting Exoplanet Survey Satellite (TESS) [28], are enabling the discovery and characterization of large populations of hot subdwarfs and white dwarfs. At the same time, the scale and complexity of these datasets necessitate the use of data-driven methodologies, including ML techniques, automated classification pipelines, and advanced statistical analysis [29]. In this context, data-driven approaches are emerging as effective tools for identifying rare stellar populations, classifying variable stars, and extracting subtle astrophysical signals from large survey datasets. By combining large photometric and spectroscopic surveys with modern computational methods, it is now possible to investigate the late stages of stellar evolution with improved statistical power and precision. In this review, we examine how ML and data-driven analysis techniques are being applied to large astronomical surveys to advance our understanding of hot subdwarfs, white dwarfs, and the complex physical processes that govern the final stages of stellar evolution.

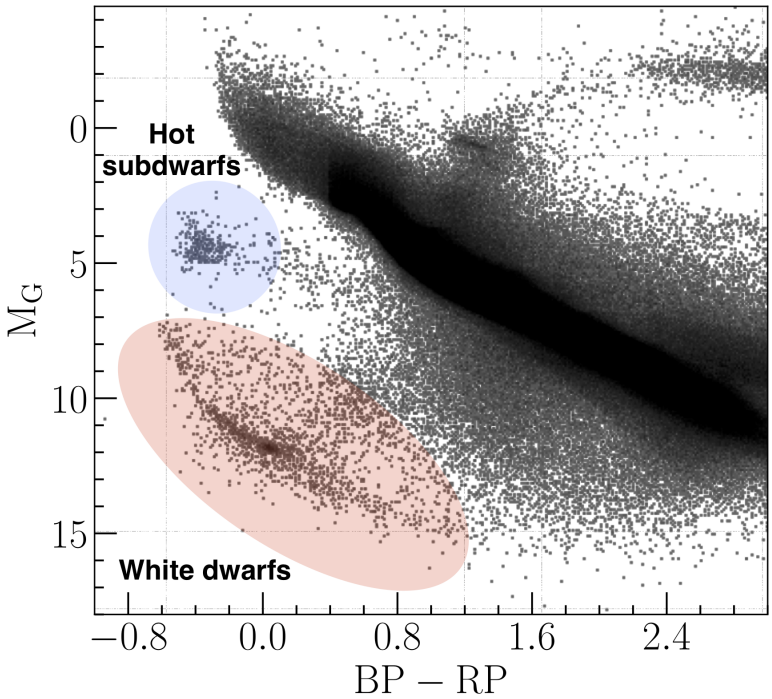

Figure 1. Observational HR diagram depicting hot subdwarfs (blue-shaded area) and white dwarfs (red-shaded area) from the Gaia DR3 catalog. The sample is restricted to sources within 1 kpc with reliable parallaxes (relative parallax uncertainty $< 20\%$) and a re-normalized unit weight error (RUWE) < 1.4, and no extinction correction has been applied. The blue- and red-shaded regions are illustrative loci intended to guide the eye to the approximate location of hot subdwarfs and white dwarfs, respectively, rather than formal selection boundaries; see Appendix A.1 of [11] for the full query and selection criteria. Note that this 1 kpc distance cut is the reason the plotted white dwarf sample appears comparatively small: the full Gaia DR3 white dwarf catalog contains over 360,000 candidates [30], and the population shown here is not intended to be representative of that full sample size, but only of the region it occupies in the HR diagram.

We stress that this article is positioned as a narrative, topical review rather than a systematic or fully quantitative meta-analysis: it does not introduce a new algorithm, benchmark dataset, or quantitative comparison of methods, but instead synthesizes recent applications of ML to hot subdwarfs and white dwarfs, following an explicit (though not exhaustive) selection of the literature described below, and identifies the key methodological gaps and future directions that we consider most relevant for the community.

Scope and Selection of Literature

Given the breadth of both the ML literature and the surveys relevant to late stellar evolution, we restrict the scope of this review as follows. In terms of ML tasks, we focus on candidate identification, variable-source and spectral classification, binary detection, atmospheric-parameter inference, and unsupervised discovery of rare subclasses, since these represent the tasks most extensively addressed for hot subdwarfs and white dwarfs to date. In terms of surveys, we emphasize Gaia, TESS, ZTF, SDSS, LAMOST, and the forthcoming Rubin/LSST, WEAVE, and 4MOST facilities, as these dominate the current and near-future data volume relevant to compact stellar remnants. We selected hot subdwarfs and white dwarfs, rather than other compact or evolved stellar populations, as representative case studies of late stellar evolution because they are the two classes for which ML applications have matured most rapidly over the past five years, span the full range of supervised and unsupervised techniques discussed in Section 2, and illustrate, in a directly comparable way, the challenges of rare-object identification in large, contaminated survey datasets. The literature reviewed here was primarily selected from works published over the last decade, with an emphasis on the last five years, and does not aim at an exhaustive literature coverage.

The paper is organized as follows: Section 2 provides an overview of data-driven methods and ML techniques commonly used in stellar astrophysics, with an emphasis on their application to large and heterogeneous survey data. Section 3 focuses on the application of these approaches to hot subdwarfs, including their identification, variability classification, and characterization in the context of large photometric and spectroscopic datasets. Section 4 presents similar methodologies applied to white dwarfs, highlighting advances in variability detection, spectral analysis, and the study of planetary debris and binary systems. Finally, Section 5 summarizes the main conclusions and findings from the literature and discusses future prospects for the study of late stellar evolution in the era of next-generation astronomical surveys and data-driven discovery.

2. Data-Driven Methods in Stellar Astrophysics

The integration of ML techniques into stellar astronomy is a rapidly evolving domain. This is because astronomy has entered the big data era, with facilities such as Vera Rubin producing tens of terabytes nightly, while astrometric and spectroscopic instruments such as Gaia, WEAVE, and 4MOST generate complex downstream catalogs (see Table 1). Scalable, data-driven solutions are therefore required to leverage the wealth of data generated by these large surveys. The past decade has seen the rapid development of a broad spectrum of ML architectures in stellar astronomy, ranging from classical algorithms to advanced deep learning models. These algorithms were specifically designed to address common challenges driven by the large volume of data produced by modern astronomy facilities. These challenges are summarized in Table 2 for each facility, while standard ML methods applied in stellar astronomy are reviewed in the following sections.

Table 1. Approximate data volume and acquisition rates for six modern optical astronomy facilities dedicated to stellar science. Values represent typical raw science data or operational averages and may vary with observing mode. The "data volume" column reports the processed or downlinked science product per day/night, while the "instrumental rate" column reports the underlying camera readout or downlink rate; these two quantities are of a different physical nature (integrated product vs. instantaneous rate) and should not be compared directly across facilities.

Facility	Stellar Science Role	Data Volume	Instrumental Rate	Refs.
Gaia	Astrometry, photometry, spectroscopy	∼40 GB day^{-1} (downlink)	1–8.5 Mbps (telemetry)	[2]
TESS	High-cadence stellar photometry	∼14 GB day^{-1} (downlink)	∼100 Mbps (telemetry)	[28]
ZTF	Time-domain stellar variability	∼1.0–1.4 TB night^{-1} (science product)	∼150 MB s^{-1} (readout)	[31,32]
LSST	Large-scale time-domain survey	∼15 TB night^{-1} (science product)	∼3.2 GB s^{-1} (readout)	[7,33]
WEAVE	Multi-object stellar spectroscopy	15–25 GB night^{-1} (science product)	2–5 MB s^{-1} (readout)	[5,34]
4MOST	Galactic stellar spectroscopy	20–30 GB night^{-1} (science product)	3–6 MB s^{-1} (readout)	[6,35]

Table 2. Representative machine learning challenges associated with major modern stellar astronomy facilities.

Facility	Dominant Data Type	ML Challenge	Refs.
Gaia	Time series + catalog	Anomaly detection, object classification, XP-spectra parameterization, astrometric-binary identification, and variability characterization	[36–38]
TESS	Light curves	Transit detection, denoising, stellar-variability classification, asteroseismic mode extraction, rotation-period recovery, and flare detection	[39–41]
ZTF	Imaging + light curves	Real-time transient classification	[42,43]
LSST	Massive imaging streams	Alert filtering and event prioritization	[44]
WEAVE	Spectra	Atmospheric-parameter and abundance inference	[45]
4MOST	Spectroscopic surveys	Atmospheric-parameter and abundance inference	[46]

2.1. Supervised Learning in Stellar Astronomy

Supervised ML methods are widely applied in stellar astronomy for classification or prediction tasks where data labels are available. Classical algorithms such as Random Forests (RF), Support Vector Machines (SVMs), and Gradient Boosting methods (e.g., XGBoost, LightGBM, CatBoost) frequently perform these tasks. Specifically, they have been applied to classify variable sources, derive stellar parameters, and predict both stellar ages and surface rotation periods [37,47–50]. Furthermore, ensemble and gradient boosting methods, including RF and XGBoost, offer high interpretability through feature importance rankings and remain computationally efficient for large-scale classification tasks. However, these traditional ML methods require careful manual feature extraction from raw data to build training datasets. In time-domain astronomy, these features are typically descriptive statistics derived from time-series data. To extract these statistical parameters from light curves, researchers routinely rely on specialized pipelines, such as the Feature Extractor for Time Series (FEETS) [51], the Time Series Feature Extraction Library (TSFEL) [52], and Superphot+ [53].

While manual feature extraction is a statistically robust method for representing tabular data, it struggles to capture the complex, non-linear patterns hidden within unstructured data, such as astronomical images, light curves, and raw stellar spectra. In contrast, deep learning algorithms bypass this limitation by performing automatic representation learning, extracting hierarchical features directly from the raw data through deep neural network architectures. Among these algorithms, Convolutional Neural Networks (CNNs) are widely used for spectral analysis and identifying physical/chemical labels [46,54,55], recurrent Neural Networks (RNNs), such as Long Short-Term Memory (LSTM) and Gated Recurrent

Units (GRUs), have been deployed for stellar parameter prediction [56] and RR Lyrae stars metallicity estimation [57] using light curve data from TESS and Gaia DR3, respectively.

It is important to note that the performance of these supervised methods is frequently reported using overall classification accuracy, which can be misleading when applied to hot subdwarfs and white dwarfs. Both classes are strongly outnumbered by main-sequence and giant contaminants in typical survey volumes, so a classifier can achieve high accuracy while still missing a large fraction of true positives or, conversely, admitting substantial contamination. Class-imbalance-aware metrics, such as precision, recall, the F1 score, or the area under the precision–recall curve, are therefore preferable to accuracy alone when these methods are applied to rare compact-object searches. Similarly, the choice of training set itself can propagate biases: models trained on spectroscopically confirmed subsamples (e.g., from SDSS or LAMOST) may not generalize to the full photometric candidate population probed by Gaia, WEAVE, or 4MOST.

2.2. Unsupervised Learning in Stellar Astronomy

Supervised ML algorithms require extensive collections of accurately labeled data, such as expert-labeled light curves, to achieve adequate performance. However, generating these representative training sets is a highly resource-intensive, manual, and time-consuming effort, which is a significant bottleneck given that modern surveys produce unlabeled data at a much faster rate. To this end, unsupervised ML techniques have become essential in astronomy, as they can capture hidden structures and patterns without relying on labeled datasets. Additionally, with unsupervised learning, it is possible to discover new classes of objects or anomalies in the datasets, which is not possible with supervised learning.

Unsupervised learning methods often apply dimensionality reduction algorithms to find meaningful representations of the data in the lower-dimensional feature space, followed by clustering algorithms to identify potential clusters. Dimensionality reduction techniques, such as t-distributed stochastic neighbor embedding (t-SNE) [58], uniform manifold approximation and projection (UMAP) [59], and principal component analysis (PCA) [60] have been used to cluster variable stars and discover anomalies in massive datasets [11,29]. UMAP and t-SNE are both non-linear techniques that map high-dimensional data into a two- or three-dimensional space while trying to preserve the local neighborhood structure of the data (i.e., points that are similar in the original feature space remain close together in the projection). The two methods differ in how they build this projection: t-SNE converts pairwise distances between points into similarity probabilities, using a Gaussian distribution in the high-dimensional space and a Student's t-distribution in the low-dimensional space, and then minimizes the divergence between the two probability distributions, whereas UMAP relies on a graph-based construction rooted in topological data analysis to preserve the neighborhood structure of the data. Neither method is universally superior to the other, and their relative performance depends on the dataset and the specific clustering task at hand; in practice, it is therefore advisable to apply both algorithms to a given dataset and adopt whichever yields the more informative or better-separated embedding. Clustering algorithms, including density-based spatial clustering of applications with noise (DBSCAN) [61] and Gaussian Mixture Models (GMMs) [62,63], are in turn used to search for coherent stellar clusters and group members from astrometric parameters [64–67].

Caution is required when interpreting the resulting low-dimensional embeddings. Neither t-SNE nor UMAP preserves global distances in a physically meaningful way: the relative size, shape, and separation of clusters in the projected space do not necessarily correspond to genuine differences in the underlying physical parameter space and can instead be strongly affected by the choice of hyperparameters (e.g., perplexity, number

of neighbors), the input feature set, and the specific sample selection used to build the embedding. Consequently, cluster boundaries identified in t-SNE or UMAP maps should be regarded as a qualitative guide for candidate selection and hypothesis generation, rather than as a quantitative or physically calibrated classification, unless independently validated against spectroscopically confirmed samples or physical models.

A specific application of this unsupervised pipeline to the clustering of hot subdwarf light curves, together with the resulting t-SNE embeddings, is presented in Section 3.2, in the context of the detection of binarity and variability in this class of stars.

2.3. Advanced Architectures

The field is increasingly adopting advanced deep learning models, including autoencoders (standard and variational) for data compression [68,69], Generative Adversarial Networks (GANs) [70] for disentangling physical and chemical spectral properties [71], and Transformer-based models to capture long-range dependencies in light curves [41,72–74].

Advanced ML architectures, such as Transformer [75] and Variational Autoencoders (VAEs) [76], offer transformative advantages in stellar astronomy by automatically extracting informative representations directly from raw, multimodal data, thereby eliminating the need for computationally expensive manual feature engineering [69]. Transformers excel at capturing long-range dependencies and broad-timescale patterns, such as stellar granulation, that traditional networks often miss, all while leveraging GPU parallelization to achieve inference speeds hundreds of times faster than classical models like Random Forests [72]. Concurrently, generative models and self-supervised autoencoders facilitate unsupervised anomaly discovery and the creation of realistic synthetic spectra, providing a vital alternative strategy when labeled observational data is scarce [68]. However, building and training these advanced architectures introduces significant challenges, most notably the high computational resource cost of self-supervised pre-training, which demands massive datasets and weeks of processing time before the models can even be fine-tuned for specific tasks [73]. Additionally, these representations can lose specific global context if their pre-training auxiliary tasks are not comprehensive enough. For generative models like VAEs, relying on standard unimodal Gaussian priors introduces statistical imprecision, making it difficult to properly model known labels with non-normal distributions or to accurately represent the inherent physical degeneracy of complex stellar spectra [68].

2.4. Machine Learning Candidates vs. Physically Confirmed Objects

A recurring point throughout this review is the distinction between objects flagged as candidates by an ML pipeline and objects that are physically confirmed through dedicated follow-up. For hot subdwarfs, physical confirmation typically requires spectral classification together with constraints on effective temperature, surface gravity, and helium abundance. For white dwarfs, confirming a magnetic field, metal pollution, or a binary companion similarly relies on spectroscopy, radial velocity monitoring, multi-epoch photometry, or polarimetry. We therefore regard ML primarily as a tool for discovery and priority ranking within large surveys, rather than as a final confirmation of physical classification. As such, dedicated follow-up spectroscopy and time-domain observations remain necessary to confirm the physical nature of a machine learning candidate.

3. Applications to Hot Subdwarfs

As previously mentioned, hot subdwarfs offer an ideal testbed for the study of binary stellar evolution. While their exact formation and evolution remain subjects of ongoing research, the systematic observation and identification of these stars are crucial for improving our understanding. Historically, the identification of hot subdwarf candidates relied on

traditional color-cut methods followed by resource-intensive visual inspection. However, the exponential growth of multi-epoch photometric and spectroscopic data from large-scale astronomical surveys has driven the integration of ML techniques into the identification and study of these objects.

To date, most ML applications have focused on identifying hot subdwarf candidates and classifying their variability. This is primarily because hot subdwarf candidate selections in observational HR diagrams are heavily contaminated by cataclysmic variables (CVs), low-mass main-sequence stars, and white dwarfs, as shown in Figure 1. Consequently, robust and scalable statistical and ML methods are required to isolate hot subdwarfs from these contaminants. In addition to candidate selection, photometric variability detection has become a computationally demanding task given the large volume of time-series data generated by modern photometric and spectroscopic surveys. We categorize recent progress in ML applications to hot subdwarf studies into three primary domains: (i) the automated identification and variability classification of candidates within massive datasets; (ii) the detection and analysis of binarity and transient phenomena; and (iii) the inference of atmospheric parameters.

3.1. Automated Hot Subdwarf Candidate Selection in Large-Scale Surveys

The primary challenge in hot subdwarf research is isolating these rare candidates from overwhelmingly large survey datasets. For photometric surveys, novel multiscale object detection algorithms, such as the Hot Subdwarf Detector (HsdDet), have been deployed to locate candidates directly within Sloan Digital Sky Survey (SDSS) images [77]. This work has identified 263 new hot subdwarfs, among 29,695 candidates detected in 216,014 images, using a spectral analysis tool XTGRID, e.g., [78]. Additionally, advanced unsupervised learning and dimensionality reduction algorithms, such as UMAP, have been used to reduce high-dimensional Gaia BP/RP (XP) spectra coefficients into 2D similarity maps [79]. These maps reveal that stellar BP-RP color is the dominant parameter driving the segregation of these stars from main sequence contaminants.

Spectroscopic data processing has seen equally significant advancements. The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) data has been extensively exploited using Convolutional Neural Networks (CNNs), Support Vector Machines (SVMs), and Hierarchical Extreme Learning Machines (HELM) to detect hot subdwarfs [80–84]. These deep learning frameworks exhibit robust classification performance; for example, CNN-based hybrid models applied to LAMOST spectra have achieved testing accuracies of 96.17% [83]. Similarly, kernel SVMs using Asymmetric Least Squares (ALS) baseline corrections achieved 87.0% accuracy by identifying unique spectral patterns [84]. These ML pipelines systematically expand the known populations of both He-rich and He-poor hot subdwarfs.

3.2. Detection of Binarity, Variability, and Transient Phenomena

Because most models suggest a common-envelope binary evolution scenario for hot subdwarfs, detecting binarity and transient phenomena (e.g., eclipses, stellar flares) is a high priority. ML approaches using Self-Organizing Maps (SOMs) and CNN ensembles trained on normalized Gaia XP spectra have provided profound statistical insights into binary fractions [85]. Recent models predict that the binary ratio is approximately 18% among hot subdwarfs showing no astrometric or photometric variability [79]. However, this fraction rises dramatically to over 60% for targets with pure photometric variability, approximately 77% for targets exhibiting both astrometric and photometric variability, and peaks at roughly 82% for stars showing only astrometric variability.

Furthermore, ML algorithms leveraging summary statistics of sparsely sampled Gaia multi-band photometry (particularly the G-band amplitude) have successfully clustered distinct variable classes [11,29], yielding over 89 new variable hot subdwarfs from combined Gaia and TESS observations. Most of these newly identified variables consist of reflection-effect and HW Vir-type binary systems. This approach efficiently separates subdwarfs from CVs without requiring expensive follow-up spectroscopy, providing an efficient way to assemble large, homogeneously selected samples of pulsating and non-pulsating hot subdwarfs. Such samples can, in turn, help empirically map the observed boundaries of the pulsation instability strip, keeping in mind that the instability strip itself is a theoretical construct and that the detection probability of pulsations, most of which are exhibited by sdB stars, can vary across the strip depending on pulsation amplitude and survey cadence.

This unsupervised clustering approach has been applied specifically to cluster Gaia DR3 light curves of confirmed and candidate hot subdwarfs [29]. Figure 2 presents an overview of the corresponding ML pipeline applied to these data. The flowchart on the left panel of Figure 2 summarizes the unsupervised learning steps. Approximately 1500 Gaia light curves were extracted, followed by a feature extraction phase that yielded 49 statistical and physical parameters after the removal of highly correlated ones. An initial t-SNE visualization was produced to explore the structure of the resulting feature space, followed by Gaussian Mixture Model (GMM) clustering to assign preliminary labels to the identified clusters. The pipeline then underwent iterative hyperparameter optimization, guided by the silhouette score and visual inspection of the graphical output. The silhouette score is a standard clustering-quality metric, ranging from -1 to $+1$, that measures how similar each point is to the other members of its own cluster relative to the nearest neighboring cluster; values close to $+1$ indicate compact, well-separated clusters, while values near zero or negative indicate overlapping or poorly defined clusters. Uninformative features were progressively removed using Random Forest feature importance ranking to further improve the clustering visualization. The optimal feature subset, which yielded the best silhouette score, was then used to generate the final t-SNE visualization shown alongside the flowchart.

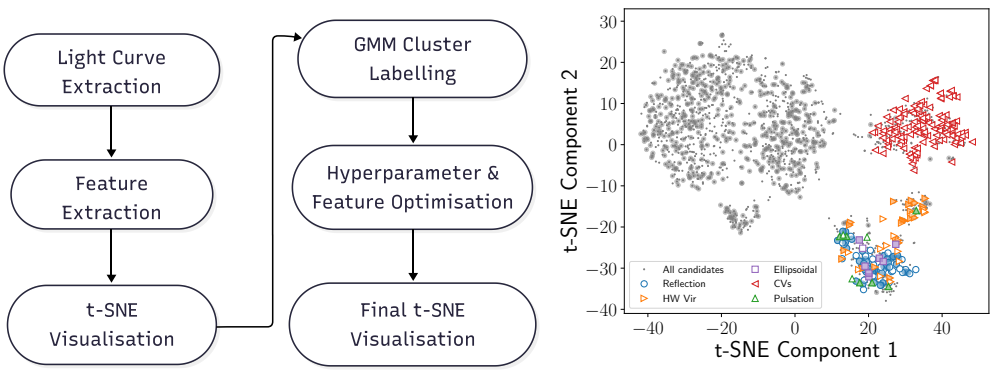

Figure 2. (**Left**) panel: Overview of the unsupervised machine learning pipeline applied to Gaia DR3 light curves of hot subdwarf candidates. (**Right**) panel: Resulting t-SNE embeddings (figure taken from Ranaivomanana et al. [29]), showing three main clusters that reflect three primary variability types. The three labeled clusters represent reflection-effect binaries, pulsators, and sources with dubious variability; the separation between clusters reflects primarily the light-curve morphology and amplitude used as input features, and the physical class assignment of individual sources within each cluster was subsequently verified against the literature classifications where available, rather than inferred from the embedding alone.

In addition to baseline variability, ML has been pivotal in identifying high-frequency transient events. Massive light-curve surveys from the Transiting Exoplanet Survey Satellite (TESS) and Evryscope have been leveraged using random forests, XGBoost, and neural

networks to identify brief transit signals and flares [86,87]. Analyses of flares associated with compact stars strongly suggest that these events likely originate from unresolved cool main-sequence companions rather than the hot subdwarfs themselves [88]. This confirms that binarity and environmental density heavily dictate the observed evolutionary paths of these systems.

3.3. Atmospheric Parameter Inference

Beyond classification, advanced deep learning algorithms were recently used to predict hot subdwarf atmospheric parameters: effective temperature, surface gravity, and helium abundance. For instance, a CNN model with channel and spatial attention layers has been designed specifically for this task [89]. The model was trained on both 11,396 synthetic spectra and 945 observed spectra from LAMOST, achieving comparable prediction performance to traditional methods, with mean absolute errors of 730 K, 0.09 dex, and 0.03 dex in effective temperature, surface gravity, and helium abundance predictions, respectively. A similar CNN-based model had previously been developed for hot subdwarf identification and atmospheric parameter prediction [90], though its prediction performance was slightly lower than that of the recent work. Despite this successful milestone in predicting physical parameters, the scarcity of observed spectra for these stars remains a hindrance to fully exploiting deep-learning-based prediction approaches, which require significantly large training samples.

3.4. Outstanding Challenges Specific to Hot Subdwarf Classification

Beyond the general contamination problem discussed above, several difficulties are specific to hot subdwarfs and deserve more explicit attention. First, the machine learning separability among hot subdwarf spectral subtypes (sdB, sdO, He-sdO, He-sdB, and sdOB) remains only partially explored: most existing classifiers are optimized to separate hot subdwarfs from contaminants rather than to resolve the finer subtype sequence, and the extent to which photometric or low-resolution spectroscopic features alone can recover this sequence is not yet well quantified. Second, an open question is whether classification outputs from ML pipelines can be used to indirectly constrain the dominant formation channel of a given hot subdwarf, such as common-envelope ejection, stable Roche-lobe overflow, or the merger of two helium white dwarfs. For instance, population-level correlations between inferred binary status, orbital period, and companion type may provide additional constraints on possible formation channels. This remains largely unexplored territory for future work. Third, Gaia XP low-resolution spectra, while extremely valuable for large-scale candidate selection [79], have intrinsic limitations in disentangling helium abundance, surface gravity, and contamination from a main-sequence companion, since these effects can produce degenerate signatures in the low-resolution BP/RP coefficients; higher-resolution spectroscopy remains necessary to break this degeneracy. Finally, distinguishing genuine HW Vir-type eclipsing systems, reflection effects, ellipsoidal variations, and pulsations from CV contamination in TESS, Gaia, and ZTF light curves remains challenging when only sparse or single-band photometry is available, and misclassification between these variability types can bias derived binary fractions and instability-strip boundaries.

4. Applications to White Dwarfs

The rapid growth of large spectroscopic and photometric surveys has fundamentally transformed the study of white dwarfs, enabling population-scale analyses that were previously impractical. However, the volume, complexity, and heterogeneity of these data sets require the development of automated, data-driven methodologies for efficient classification, characterization, and discovery. In recent years, ML and advanced statistical techniques

have been widely applied to white dwarf research, broadly falling into several key categories: (i) automated classification and identification, (ii) discovery of rare and peculiar subclasses, (iii) analysis of binary systems, (iv) spectral modeling and parameter inference, and (v) exploration of planetary debris and chemical composition. These categories can equivalently be viewed as addressing a sequence of scientific questions that recurs throughout this section: how the overall white dwarf sample is first assembled from Gaia HR-diagram selection and confirmed with SDSS/LAMOST spectroscopy (Section 4.1); how confirmed white dwarfs are subsequently sorted into spectral classes such as DA, DB, DQ, DZ, DC, magnetic, and ultracool white dwarfs (Sections 4.1 and 4.2); how physical parameters such as effective temperature, surface gravity, mass, cooling age, and atmospheric composition are then estimated (Section 4.4); how special physical processes, including magnetic fields, metal pollution, debris disks, and binary interaction, are identified (Sections 4.3 and 4.5); and how time-domain phenomena such as pulsation, rotational modulation, eclipsing binarity, and outburst or accretion events are detected. We follow this thematic thread within the subsections below, while noting that a dedicated systematic treatment of time-domain white dwarf phenomena, comparable to Section 3 (binarity, variability, and transient phenomena) for hot subdwarfs, is still comparatively underdeveloped in the literature and represents a natural direction for future dedicated reviews.

Selection effects also need to be kept in mind when interpreting the results summarized below. Gaia color-absolute magnitude selection does not affect all white dwarf subpopulations equally. Specifically, the selection tends to be least complete for the coolest, faintest white dwarfs; for composite-spectrum white dwarf–main-sequence (WD–MS) binaries whose photometry is diluted by the companion; for strongly magnetic white dwarfs whose broadband colors can be shifted by Zeeman splitting; and for lightly polluted white dwarfs whose metal features are only detectable at higher spectral resolution. Moreover, because many of the ML models discussed below are trained on spectroscopically confirmed subsamples from SDSS or LAMOST, their training data may not be representative of the full population of photometrically selected Gaia white dwarf candidates, so the derived class fractions and occurrence rates should be interpreted as conditional on the specific selection function of the training set rather than as unbiased estimates for the underlying population.

4.1. Automated Classification and Identification

Efficiently classifying white dwarf spectra is a primary challenge for modern wide-field surveys, as the sheer volume of data has made manual visual inspection impossible. To address this, several ML pipelines have been developed to automate the process. For instance, models that integrate Gaia astrometry with SDSS spectroscopy achieve over 90% accuracy, enabling the creation of catalogs containing hundreds of thousands of objects [91]. More advanced convolutional neural networks, in particular multimodal versions that process both 1D spectra and 2D images, have pushed classification accuracy toward 99%, significantly reducing the requirement for human intervention [92]. In addition to spectral analysis, new object-detection frameworks can now identify white dwarf candidates directly from imaging data. By bypassing traditional photometric color-cut limitations, these methods have successfully uncovered tens of thousands of new candidates within massive survey datasets [93].

4.2. Unsupervised Learning and Discovery of Rare Subclasses

Unsupervised ML techniques have proven particularly powerful for identifying rare or previously unknown subclasses of white dwarfs. Dimensionality reduction methods such as UMAP and t-SNE have been widely used to map high-dimensional spectral data into lower-dimensional spaces, revealing distinct clusters corresponding to different spectral

types and physical properties [11,94–96]. These methods have enabled the discovery of new polluted white dwarf candidates, the identification of magnetic white dwarfs, and the detection of rare spectral features that may be missed by supervised classification. For example, clustering techniques combined with density-based algorithms have successfully separated magnetic and non-magnetic populations and enabled the estimation of magnetic field strengths in cases where direct measurements are unavailable [97]. Similarly, self-organizing maps and other unsupervised approaches have identified new metal-polluted white dwarfs in Gaia data, demonstrating their effectiveness in mining large, sparsely labeled datasets [98,99].

4.3. Binary Systems and Population Studies

Data-driven approaches have also been widely applied to the identification and characterization of white dwarf binary systems, particularly WD–MS binaries. Neural network-based approaches combined with Gaussian process classification have enabled the identification of tens of thousands of WD–MS candidates in Gaia data, including systems in which the main-sequence companion dominates the observed spectrum [100]. In addition, classification techniques have been extended to simulated datasets to prepare for future observations. For example, ML models trained on synthetic spectra have demonstrated strong performance in classifying WD–MS systems in low-resolution Gaia spectra [101]. Beyond electromagnetic observations, ML is also being applied to gravitational-wave astronomy, where it is used to identify compact binaries detectable by future missions such as LISA, including double white dwarf systems and rarer neutron star–white dwarf binaries [102].

4.4. Spectral Modeling and Parameter Inference

Another major application of ML in white dwarf studies is the inference of stellar parameters and chemical abundances from spectroscopic data. Traditional spectral fitting techniques are computationally expensive and difficult to scale to large datasets. To address this, ML-based spectral modeling frameworks have been developed to generate synthetic spectra and perform parameter estimation. For example, neural-network-based models trained on large grids of atmosphere models can infer effective temperature, surface gravity, and abundances with accuracies comparable to classical methods [103]. These approaches have also been successfully applied to detailed abundance analyses of metal-polluted white dwarfs, enabling efficient characterization of multiple chemical species from spectroscopic observations, e.g., [104].

4.5. Planetary Debris and Chemical Composition

White dwarfs polluted by planetary debris provide a unique opportunity to study the composition of exoplanetary material. ML techniques have significantly improved the identification and characterization of these systems. Unsupervised clustering methods applied to Gaia spectra have revealed distinct groups of metal-polluted white dwarf candidates and increased the number of known systems after confirming their physical nature through spectroscopic follow-up observations [99]. Comparative studies of different ML approaches have shown that both supervised and unsupervised methods play complementary roles in identifying polluted white dwarfs. While supervised models benefit from labeled training data, unsupervised techniques are particularly effective at discovering new candidates and rare subclasses [95]. These developments are enabling large-scale studies of planetary system evolution and the chemical diversity of exoplanetary material.

5. Conclusions and Future Prospects

In this review, we have summarized how the interplay between large-scale data from modern astronomical facilities and advanced ML techniques is reshaping the identification

and characterization of hot subdwarfs and white dwarfs. Various ML models have been implemented in stellar astronomy, ranging from classical algorithms to deep learning architectures, using both supervised and unsupervised approaches. We have also seen that astronomical facilities generate a wide variety of data, including tabular catalogs, images, spectra, and time series, at rates reaching up to 15 TB per night (as an integrated nightly science product; see Table 1). Consequently, various ML tasks emerge, such as object classification, transit detection, and astrophysical parameter estimation.

Table 3 summarizes, side by side, the main similarities and differences between hot subdwarfs and white dwarfs that motivate their joint treatment in this review, together with the machine learning tasks and open challenges specific to each class.

Table 3. Comparison of hot subdwarfs and white dwarfs as case studies for machine-learning-driven discovery in late stellar evolution.

Aspect	Hot Subdwarfs	White Dwarfs
Evolutionary stage	Core-helium-burning, late but not final	Final compact remnant
Dominant formation channel	Binary mass loss (CE, RLOF, white dwarf merger)	Single- and binary-star envelope loss
Main contaminants	BHB stars, CVs, low-mass main-sequence stars, white dwarfs	Faint main-sequence/subdwarf stars, WD–MS composites
Dominant ML tasks	Candidate selection, variability/binary classification	Spectral classification, parameter and abundance inference
Main open ML challenge	Subtype separability (sdB/sdO/He-sdB/He-sdO/sdOB); scarce labeled spectra	Selection-function bias from spectroscopic training sets; rare-subclass discovery

In hot subdwarf research, ML techniques are still largely applied to candidate identification and variability classification, while little work has been done regarding physical parameter estimation, such as predicting atmospheric parameters and abundances. This limitation is mainly due to the scarcity of labeled data and the high cost of acquiring the spectral data needed to confirm genuine candidates. Nevertheless, the integration of unsupervised ML methods has significantly improved the detection of hot subdwarf candidates, enabling robust statistical studies and physical inferences about these stars. In particular, unsupervised ML approaches have been shown to efficiently categorize large datasets of hot subdwarfs and contaminants using photometric data alone, which were traditionally difficult to separate without additional spectroscopy.

The application of ML techniques and advanced statistical methods has similarly reshaped white dwarf research across multiple fronts. Automated classification pipelines now enable the processing of vast spectroscopic and photometric surveys while maintaining consistent accuracy at scale, though, as discussed in Section 4, this accuracy is conditioned on the completeness and representativeness of the underlying training samples. Unsupervised learning techniques have been particularly effective in revealing rare and previously unknown subclasses, including magnetic and metal-polluted white dwarfs, by uncovering hidden structure in high-dimensional survey data. In parallel, ML techniques have improved the identification and characterization of binary systems, supported scalable spectral modeling and parameter inference, and enabled large-scale studies of chemically polluted atmospheres that trace the composition of exoplanetary material. Together, these developments demonstrate that data-driven approaches are becoming essential tools for extracting physical insight from the rapidly expanding white dwarf datasets when combined

with the follow-up validation discussed above, and they will continue to play a central role in future survey-driven discoveries.

However, despite these substantial advances, important limitations remain. Specifically, the performance of current classification and inference methods remains sensitive to data quality. In low signal-to-noise regimes, where spectral features become ambiguous and photometric variability is poorly constrained, current ML models can produce incomplete or uncertain classifications. This reveals gaps in the parameter space where certain classes of objects are systematically underrepresented or misidentified due to selection effects.

To address these issues, future research might explore combining a data-centric approach with ML architecture improvements. That is, rather than focusing solely on optimizing ML models, the community would benefit from actively curating higher-quality training sets. This can be done, for instance, through active learning strategies that prioritize targeted spectroscopic follow-up of rare or ambiguous sources [105], synthetic data generation for underrepresented classes [106], or multimodal learning approaches that combine heterogeneous data via cross-survey cross-matching to mitigate selection biases [107,108]. At the same time, model-side advances remain essential; in particular, uncertainty-aware methods such as Bayesian inference and conformal prediction [109,110], as well as physics-informed architectures [111,112] that incorporate domain knowledge as structural constraints, offer complementary paths toward more interpretable and well-calibrated classifications.

More concretely, we identify the following routes as particularly promising over the coming years: (i) the establishment of public, community-agreed benchmark datasets for hot subdwarfs and white dwarfs, with standardized train/validation/test splits and documented selection functions, to enable direct comparison between competing ML methods; (ii) a systematic shift toward calibrated probabilistic outputs, rather than hard class labels, so that downstream population studies can propagate classification uncertainty; (iii) the incorporation of stellar atmosphere models, evolutionary tracks, and color-magnitude diagram priors, together with observational uncertainties, directly into the loss functions of physics-informed architectures, so that model outputs remain consistent with known stellar physics; and (iv) the design of low-latency, real-time classification pipelines capable of triaging compact stellar objects directly within the Rubin/LSST alert stream, where the volume of nightly alerts will preclude purely manual vetting.

Beyond these technical and methodological improvements, it is worth emphasizing the fundamental astrophysical questions that ML-enabled samples may help address as survey volumes continue to grow. As Rubin/LSST, WEAVE, 4MOST, and future Gaia-like missions deliver homogeneously selected samples of hot subdwarfs and white dwarfs that are orders of magnitude larger than those available today, ML-based approaches could enable population-level physical inference from large, homogeneous samples of compact stars. For example, sufficiently large and well-characterized samples of binary hot subdwarfs, classified and parameterized at scale, could provide new empirical constraints on the common-envelope efficiency parameter and on the relative importance of different binary mass-loss channels, which are currently poorly constrained by the comparatively small samples available from targeted spectroscopic follow-up [13,14,113,114]. Similarly, large, homogeneously selected samples of pulsating compact stars could refine the empirical boundaries of the instability strips and improve understanding of pulsation-mode selection of the hot subdwarf and white dwarf instability strips, complementing and testing current theoretical pulsation-driving models [17,19,20,115]. For white dwarfs, statistically robust, ML-derived samples of cooling ages, masses, and atmospheric compositions could sharpen the empirical initial–final mass relation [20,116–118] and allow white dwarf cooling sequences to be used as reliable chronometers for Galactic archaeology on a much larger scale than is currently possible [119,120]. Larger samples of metal-polluted white dwarfs, efficiently

identified through ML pipelines, could likewise enable population-level statistical studies of the bulk composition of exoplanetary material [121,122], opening the possibility of investigating whether the composition of planetary material varies with Galactic environment or progenitor stellar properties. Realizing this potential, however, will require the methodological advances discussed above, in particular robust uncertainty quantification and careful control of selection effects, since population-level physical inferences drawn from ML-classified samples are only as reliable as the completeness and purity of the underlying catalogs.

Furthermore, degeneracies between different astrophysical scenarios can make the interpretation of the data more difficult. A notable example is the confusion between short-period binary systems and long-period pulsating hot subdwarfs or white dwarfs, which can produce similar photometric signatures in single-epoch or sparsely sampled data, as demonstrated in Ranaivomanana et al. [11,29]. These challenges underscore the critical importance of dedicated follow-up observations. Multi-epoch spectroscopy and time-series photometry are essential to confirm candidate classifications, break parameter degeneracies, and accurately disentangle variability mechanisms. Such observations provide the temporal baseline required to distinguish orbital motion from pulsational behavior and to improve parameter constraints derived from ML models. Therefore, while ML methods are indispensable for efficiently mining large surveys, their full scientific potential can only be realized when combined with systematic follow-up campaigns that ensure robust validation and physical interpretation of the identified sources.

Author Contributions: Conceptualization, P.R. and M.U.; methodology, P.R. and M.U.; writing—original draft preparation, P.R. and M.U.; writing—Sections 2 and 3, P.R.; writing—abstract, Sections 1 and 4, M.U.; writing—Section 5, P.R. and M.U.; visualization, P.R.; software, P.R.; investigation, P.R. and M.U.; writing—review and editing, P.R. and M.U.; project administration, P.R. and M.U. All authors have read and agreed to the published version of the manuscript.

Funding: M.U. was supported by the Research Foundation Flanders (FWO) through a Postdoctoral Fellowship (grant agreement No. 1247624N). P.R. received no specific external funding for this work.

Data Availability Statement: No new data were generated in this review. All data discussed and summarized in this manuscript are publicly available in the original publications cited throughout the text and in the references therein. The Gaia DR3 data used to produce Figure 1 are publicly available through the Gaia Archive (https://gea.esac.esa.int/archive/).

Acknowledgments: The authors acknowledge the use of Mermaid Live Editor (https://mermaid.live), accessed on 15 June 2026, for the purposes of creating the diagram in the left panel of Figure 2.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Aerts, C. Probing the interior physics of stars through asteroseismology. *Rev. Mod. Phys.* **2021**, *93*, 015001. [CrossRef]
2. Prusti, T. et al. [Gaia Collaboration] The Gaia mission. *Astron. Astrophys.* **2016**, *595*, A1. [CrossRef]
3. Cui, X.Q.; Zhao, Y.H.; Chu, Y.Q.; Li, G.P.; Li, Q.; Zhang, L.P.; Su, H.J.; Yao, Z.Q.; Wang, Y.N.; Xing, X.Z.; et al. The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST). *Res. Astron. Astrophys.* **2012**, *12*, 1197–1242. [CrossRef]
4. Kollmeier, J.A.; Zasowski, G.; Rix, H.W.; Johns, M.; Anderson, S.F.; Drory, N.; Johnson, J.A.; Pogge, R.W.; Bird, J.C.; Blanc, G.A.; et al. SDSS-V: Pioneering Panoptic Spectroscopy. *arXiv* **2017**, arXiv:1711.03234. [CrossRef]
5. Jin, S.; Trager, S.C.; Dalton, G.B.; Aguerri, J.A.L.; Drew, J.E.; Falcón-Barroso, J.; Gänsicke, B.T.; Hill, V.; Iovino, A.; Pieri, M.M.; et al. The wide-field, multiplexed, spectroscopic facility WEAVE: Survey design, overview, and simulated implementation. *Mon. Not. R. Astron. Soc.* **2024**, *530*, 2688–2730. [CrossRef]
6. de Jong, R.S.; Agertz, O.; Berbel, A.A.; Aird, J.; Alexander, D.A.; Amarsi, A.; Anders, F.; Andrae, R.; Ansarinejad, B.; Ansorge, W.; et al. 4MOST: Project overview and information for the First Call for Proposals. *Messenger* **2019**, *175*, 3–11. [CrossRef]
7. Ivezić, Ž.; Kahn, S.M.; Tyson, J.A.; Abel, B.; Acosta, E.; Allsman, R.; Alonso, D.; AlSayyad, Y.; Anderson, S.F.; Andrew, J.; et al. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *Astrophys. J.* **2019**, *873*, 111. [CrossRef]
8. Heber, U. Hot Subdwarf Stars. *Annu. Rev. Astron. Astrophys.* **2009**, *47*, 211–251. [CrossRef]

9. Heber, U. Hot Subluminous Stars. *Publ. Astron. Soc. Pac.* **2016**, *128*, 082001. [CrossRef]
10. Heber, U. Hot Subdwarf Stars. *arXiv* **2024**, arXiv:2410.11663. [CrossRef]
11. Ranaivomanana, P.; Johnston, C.; Iorio, G.; Groot, P.J.; Uzundag, M.; Kupfer, T.; Aerts, C. Unsupervised learning for variability detection with Gaia Data Release 3 photometry: The main sequence–white dwarf valley. *Astron. Astrophys.* **2025**, *704*, A70. [CrossRef]
12. Ivanova, N.; Justham, S.; Chen, X.; De Marco, O.; Fryer, C.L.; Gaburov, E.; Ge, H.; Glebbeek, E.; Han, Z.; Li, X.D.; et al. Common envelope evolution: Where we stand and how we can move forward. *Astron. Astrophys. Rev.* **2013**, *21*, 59. [CrossRef]
13. Han, Z.; Podsiadlowski, P.; Maxted, P.F.L.; Marsh, T.R. The origin of subdwarf B stars—II. *Mon. Not. R. Astron. Soc.* **2003**, *341*, 669–691. [CrossRef]
14. Han, Z.; Podsiadlowski, P.; Maxted, P.F.L.; Marsh, T.R.; Ivanova, N. The origin of subdwarf B stars—I. The formation channels. *Mon. Not. R. Astron. Soc.* **2002**, *336*, 449–466. [CrossRef]
15. Charpinet, S.; Fontaine, G.; Brassard, P.; Chayer, P.; Rogers, F.J.; Iglesias, C.A.; Dorman, B. A Driving Mechanism for the Newly Discovered Class of Pulsating Subdwarf B Stars. *Astrophys. J.* **1997**, *483*, L123–L126. [CrossRef]
16. Kilkenny, D.; Koen, C.; O'Donoghue, D.; Stobie, R.S. A new class of rapidly pulsating star—I. EC 14026-2647, the class prototype. *Mon. Not. R. Astron. Soc.* **1997**, *285*, 640–644. [CrossRef]
17. Green, E.M.; Fontaine, G.; Reed, M.D.; Callerame, K.; Seitenzahl, I.R.; White, B.A.; Hyde, E.A.; Østensen, R.; Cordes, O.; Brassard, P.; et al. Discovery of A New Class of Pulsating Stars: Gravity-Mode Pulsators among Subdwarf B Stars. *Astrophys. J.* **2003**, *583*, L31–L34. [CrossRef]
18. Uzundag, M.; Córsico, A.H.; Kepler, S.O.; Althaus, L.G.; Werner, K.; Reindl, N.; Bell, K.J.; Higgins, M.; da Rosa, G.O.; Vučković, M.; et al. Pulsating hydrogen-deficient white dwarfs and pre-white dwarfs observed with TESS. II. Discovery of two new GW Vir stars: TIC 333432673 and TIC 095332541. *Astron. Astrophys.* **2021**, *655*, A27. [CrossRef]
19. Uzundag, M.; Krzesinski, J.; Pelisoli, I.; Németh, P.; Silvotti, R.; Vučković, M.; Dawson, H.; Geier, S. A comprehensive search for hot subdwarf stars using Gaia and TESS. I. Pulsating hot subdwarf B stars. *Astron. Astrophys.* **2024**, *684*, A118. [CrossRef]
20. Althaus, L.G.; Córsico, A.H.; Isern, J.; García-Berro, E. Evolutionary and pulsational properties of white dwarf stars. *Astron. Astrophys. Rev.* **2010**, *18*, 471–566. [CrossRef]
21. Winget, D.E.; Kepler, S.O. Pulsating white dwarf stars and precision asteroseismology. *Annu. Rev. Astron. Astrophys.* **2008**, *46*, 157–199. [CrossRef]
22. Fontaine, G.; Brassard, P. The Pulsating White Dwarf Stars. *Publ. Astron. Soc. Pac.* **2008**, *120*, 1043. [CrossRef]
23. Córsico, A.H.; Althaus, L.G.; Miller Bertolami, M.M.; Kepler, S.O. Pulsating white dwarfs: New insights. *Astron. Astrophys. Rev.* **2019**, *27*, 7. [CrossRef]
24. Veras, D. Post-main-sequence planetary system evolution. *R. Soc. Open Sci.* **2016**, *3*, 150571. [CrossRef]
25. Veras, D.; Leinhardt, Z.M.; Bonsor, A.; Gänsicke, B.T. Formation of planetary debris discs around white dwarfs—I. Tidal disruption of an extremely eccentric asteroid. *Mon. Not. R. Astron. Soc.* **2014**, *445*, 2244–2255. [CrossRef]
26. Borucki, W.J.; Koch, D.; Basri, G.; Batalha, N.; Brown, T.; Caldwell, D.; Caldwell, J.; Christensen-Dalsgaard, J.; Cochran, W.D.; DeVore, E.; et al. Kepler Planet-Detection Mission: Introduction and First Results. *Science* **2010**, *327*, 977. [CrossRef]
27. Haas, M.R.; Barclay, T.; Batalha, N.M.; Bryson, S.; Caldwell, D.A.; Campbell, J.; Coughlin, J.; Howell, S.B.; Jenkins, J.M.; Klaus, T.C.; et al. The Kepler Mission on Two Reaction Wheels is K2. In Proceedings of the American Astronomical Society Meeting Abstracts #223, Washington, DC, USA, 5–9 January 2014; Volume 223, p. 228.01.
28. Ricker, G.R.; Winn, J.N.; Vanderspek, R.; Latham, D.W.; Bakos, G.Á.; Bean, J.L.; Berta-Thompson, Z.K.; Brown, T.M.; Buchhave, L.; Butler, N.R.; et al. Transiting Exoplanet Survey Satellite (TESS). *J. Astron. Telesc. Instrum. Syst.* **2015**, *1*, 014003. [CrossRef]
29. Ranaivomanana, P.; Uzundag, M.; Johnston, C.; Groot, P.J.; Kupfer, T.; Aerts, C. Variability in hot sub-luminous stars and binaries: Machine-learning analysis of Gaia DR3 multi-epoch photometry. *Astron. Astrophys.* **2025**, *693*, A268. [CrossRef]
30. Gentile Fusillo, N.P.; Tremblay, P.E.; Cukanovaite, E.; Vorontseva, A.; Lallement, R.; Hollands, M.; Gänsicke, B.T.; Burdge, K.B.; McCleery, J.; Jordan, S. A catalogue of white dwarfs in Gaia EDR3. *Mon. Not. R. Astron. Soc.* **2021**, *508*, 3877–3896. [CrossRef]
31. Bellm, E.C.; Kulkarni, S.R.; Graham, M.J.; Dekany, R.; Smith, R.M.; Riddle, R.; Masci, F.J.; Helou, G.; Prince, T.A.; Adams, S.M.; et al. The Zwicky Transient Facility: System Overview, Performance, and First Results. *Publ. Astron. Soc. Pac.* **2019**, *131*, 018002. [CrossRef]
32. Graham, M.J.; Kulkarni, S.R.; Bellm, E.C.; Adams, S.M.; Barbarino, C.; Blagorodnova, N.; Bodewits, D.; Bolin, B.; Brady, P.R.; Cenko, S.B.; et al. The Zwicky Transient Facility: Science Objectives. *Publ. Astron. Soc. Pac.* **2019**, *131*, 078001. [CrossRef]
33. Jurić, M.; Kantor, J.; Lim, K.T.; Lupton, R.H.; Dubois-Felsmann, G.; Jenness, T.; Axelrod, T.S.; Aleksić, J.; Allsman, R.A.; AlSayyad, Y.; et al. The LSST Data Management System. In *Proceedings of the Astronomical Data Analysis Software and Systems XXV*; Lorente, N.P.F., Shortridge, K., Wayth, R., Eds.; Astronomical Society of the Pacific Conference Series; Astronomical Society of the Pacific: San Francisco, CA, USA, 2017; Volume 512, p. 279. [CrossRef]

34. Dalton, G.; Trager, S.C.; Abrams, D.C.; Carter, D.; Bonifacio, P.; Aguerri, J.A.L.; MacIntosh, M.; Evans, C.; Lewis, I.; Navarro, R.; et al. WEAVE: The next generation wide-field spectroscopy facility for the William Herschel Telescope. In *Proceedings of the Ground-Based and Airborne Instrumentation for Astronomy IV*; McLean, I.S., Ramsay, S.K., Takami, H., Eds.; Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series; SPIE: Bellingham, WA, USA, 2012; Volume 8446, p. 84460P. [CrossRef]
35. de Jong, R.S.; Bellido-Tirado, O.; Chiappini, C.; Depagne, É.; Haynes, R.; Johl, D.; Schnurr, O.; Schwope, A.; Walcher, J.; Dionies, F.; et al. 4MOST: 4-metre multi-object spectroscopic telescope. In *Proceedings of the Ground-Based and Airborne Instrumentation for Astronomy IV*; McLean, I.S., Ramsay, S.K., Takami, H., Eds.; Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series; SPIE: Bellingham, WA, USA, 2012; Volume 8446, p. 84460T. [CrossRef]
36. Rimoldini, L.; Holl, B.; Audard, M.; Mowlavi, N.; Nienartowicz, K.; Evans, D.W.; Guy, L.P.; Lecoeur-Taïbi, I.; Jevardat de Fombelle, G.; Marchal, O.; et al. Gaia Data Release 2. All-sky classification of high-amplitude pulsating stars. *Astron. Astrophys.* **2019**, *625*, A97. [CrossRef]
37. Rimoldini, L.; Holl, B.; Gavras, P.; Audard, M.; De Ridder, J.; Mowlavi, N.; Nienartowicz, K.; Jevardat de Fombelle, G.; Lecoeur-Taïbi, I.; Karbevska, L.; et al. Gaia Data Release 3. All-sky classification of 12.4 million variable sources into 25 classes. *Astron. Astrophys.* **2023**, *674*, A14. [CrossRef]
38. Eyer, L.; Audard, M.; Holl, B.; Rimoldini, L.; Carnerero, M.I.; Clementini, G.; De Ridder, J.; Distefano, E.; Evans, D.W.; Gavras, P.; et al. Gaia Data Release 3. Summary of the variability processing and analysis. *Astron. Astrophys.* **2023**, *674*, A13. [CrossRef]
39. Sullivan, P.W.; Winn, J.N.; Berta-Thompson, Z.K.; Charbonneau, D.; Deming, D.; Dressing, C.D.; Latham, D.W.; Levine, A.M.; McCullough, P.R.; Morton, T.; et al. The Transiting Exoplanet Survey Satellite: Simulations of Planet Detections and Astrophysical False Positives. *Astrophys. J.* **2015**, *809*, 77. [CrossRef]
40. Yu, L.; Vanderburg, A.; Huang, C.; Shallue, C.J.; Crossfield, I.J.M.; Gaudi, B.S.; Daylan, T.; Dattilo, A.; Armstrong, D.J.; Ricker, G.R.; et al. Identifying Exoplanets with Deep Learning. III. Automated Triage and Vetting of TESS Candidates. *Astron. J.* **2019**, *158*, 25. [CrossRef]
41. Morvan, M.; Nikolaou, N.; Yip, K.; Waldmann, I. Don't Pay Attention to the Noise: Learning Self-supervised Representations of Light Curves with a Denoising Time Series Transformer. In Proceedings of the Machine Learning for Astrophysics, Catania, Italy, 30 May–1 June 2022; p. 11. [CrossRef]
42. Mahabal, A.; Rebbapragada, U.; Walters, R.; Masci, F.J.; Blagorodnova, N.; van Roestel, J.; Ye, Q.Z.; Biswas, R.; Burdge, K.; Chang, C.K.; et al. Machine Learning for the Zwicky Transient Facility. *Publ. Astron. Soc. Pac.* **2019**, *131*, 038002. [CrossRef]
43. Duev, D.A.; Mahabal, A.; Masci, F.J.; Graham, M.J.; Rusholme, B.; Walters, R.; Karmarkar, I.; Frederick, S.; Kasliwal, M.M.; Rebbapragada, U.; et al. Real-bogus classification for the Zwicky Transient Facility using deep learning. *Mon. Not. R. Astron. Soc.* **2019**, *489*, 3582–3590. [CrossRef]
44. Narayan, G. et al. [ANTARES Collaboration] Machine-learning-based Brokers for Real-time Classification of the LSST Alert Stream. *Astrophys. J. Suppl. Ser.* **2018**, *236*, 9. [CrossRef]
45. Guiglion, G.; Matijevič, G.; Queiroz, A.B.A.; Valentini, M.; Steinmetz, M.; Chiappini, C.; Grebel, E.K.; McMillan, P.J.; Kordopatis, G.; Kunder, A.; et al. The RAdial Velocity Experiment (RAVE): Parameterisation of RAVE spectra based on convolutional neural networks. *Astron. Astrophys.* **2020**, *644*, A168. [CrossRef]
46. Nepal, S.; Guiglion, G.; de Jong, R.S.; Valentini, M.; Chiappini, C.; Steinmetz, M.; Ambrosch, M.; Pancino, E.; Jeffries, R.D.; Bensby, T.; et al. The Gaia-ESO Survey: Preparing the ground for 4MOST and WEAVE galactic surveys. Chemical evolution of lithium with machine learning. *Astron. Astrophys.* **2023**, *671*, A61. [CrossRef]
47. Richards, J.W.; Starr, D.L.; Butler, N.R.; Bloom, J.S.; Brewer, J.M.; Crellin-Quick, A.; Higgins, J.; Kennedy, R.; Rischard, M. On Machine-learned Classification of Variable Stars with Sparse and Noisy Time-series Data. *Astrophys. J.* **2011**, *733*, 10. [CrossRef]
48. Anders, F.; Gispert, P.; Ratcliffe, B.; Chiappini, C.; Minchev, I.; Nepal, S.; Queiroz, A.B.A.; Amarante, J.A.S.; Antoja, T.; Casali, G.; et al. Spectroscopic age estimates for APOGEE red-giant stars: Precise spatial and kinematic trends with age in the Galactic disc. *Astron. Astrophys.* **2023**, *678*, A158. [CrossRef]
49. Gomes, N.R.C.; Del Sordo, F.; Torgo, L. Predicting stellar rotation periods using XGBoost. *Astron. Astrophys.* **2024**, *690*, A131. [CrossRef]
50. Machado-Pereira, E.; Daflon, S.; Placco, V.M.; Cruz, P.; Borges Fernandes, M.; Jiménez-Esteban, F.; Caballero, J.A.; Anguiano, B.; Yuan, H.; Bonatto, C.; et al. J-PLUS: Stellar parameter determination of members of open clusters using machine learning. *Mon. Not. R. Astron. Soc.* **2026**, *546*, stag132. [CrossRef]
51. Cabral, J.B.; Sánchez, B.; Ramos, F.; Gurovich, S.; Granitto, P.M.; Vanderplas, J. From FATS to feets: Further improvements to an astronomical feature extraction tool based on machine learning. *Astron. Comput.* **2018**, *25*, 213–220. [CrossRef]
52. Barandas, M.; Folgado, D.; Fernandes, L.; Santos, S.; Abreu, M.; Bota, P.; Liu, H.; Schultz, T.; Gamboa, H. TSFEL: Time Series Feature Extraction Library. *SoftwareX* **2020**, *11*, 100456. [CrossRef]
53. de Soto, K.M.; Villar, V.A.; Berger, E.; Gomez, S.; Hosseinzadeh, G.; Branton, D.; Campos, S.; DeLucchi, M.; Kubica, J.; Lynn, O.; et al. Superphot+: Real-time Fitting and Classification of Supernova Light Curves. *Astrophys. J.* **2024**, *974*, 169. [CrossRef]

54. Sharma, K.; Kembhavi, A.; Kembhavi, A.; Sivarani, T.; Abraham, S.; Vaghmare, K. Application of convolutional neural networks for stellar spectral classification. *Mon. Not. R. Astron. Soc.* **2020**, *491*, 2280–2300. [CrossRef]
55. Guiglion, G.; Nepal, S.; Chiappini, C.; Khoperskov, S.; Traven, G.; Queiroz, A.B.A.; Steinmetz, M.; Valentini, M.; Fournier, Y.; Vallenari, A.; et al. Beyond Gaia DR3: Tracing the [α/M]–[M/H] bimodality from the inner to the outer Milky Way disc with Gaia-RVS and convolutional neural networks. *Astron. Astrophys.* **2024**, *682*, A9. [CrossRef]
56. Hinners, T.A.; Tat, K.; Thorp, R. Machine Learning Techniques for Stellar Light Curve Classification. *Astron. J.* **2018**, *156*, 7. [CrossRef]
57. Monti, L.; Muraveva, T.; Garofalo, A.; Clementini, G.; Valentini, M.L. Unified deep learning approach for estimating the metallicities of RR Lyrae stars using light curves from Gaia Data Release 3. *Astron. Astrophys.* **2025**, *702*, A148. [CrossRef]
58. van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605. [CrossRef]
59. McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* **2018**, arXiv:1802.03426. [CrossRef]
60. Jolliffe, I.T. *Principal Component Analysis*, 2nd ed.; Springer Series in Statistics; Springer: New York, NY, USA, 2002. [CrossRef]
61. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining KDD'96*; AAAI Press: Washington, DC, USA, 1996; pp. 226–231.
62. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. R. Stat. Soc. Ser. B Methodol.* **1977**, *39*, 1–38. [CrossRef]
63. Reynolds, D. Gaussian Mixture Models. In *Encyclopedia of Biometrics*; Springer: Boston, MA, USA, 2009; pp. 659–663. [CrossRef]
64. Castro-Ginard, A.; Jordi, C.; Luri, X.; Álvarez Cid-Fuentes, J.; Casamiquela, L.; Anders, F.; Cantat-Gaudin, T.; Monguió, M.; Balaguer-Núñez, L.; Solà, S.; et al. Hunting for open clusters in Gaia DR2: 582 new open clusters in the Galactic disc. *Astron. Astrophys.* **2020**, *635*, A45. [CrossRef]
65. Logan, C.H.A.; Fotopoulou, S. Unsupervised star, galaxy, QSO classification. Application of HDBSCAN. *Astron. Astrophys.* **2020**, *633*, A154. [CrossRef]
66. Prisinzano, L.; Damiani, F.; Sciortino, S.; Flaccomio, E.; Guarcello, M.G.; Micela, G.; Tognelli, E.; Jeffries, R.D.; Alcalá, J.M. Low-mass young stars in the Milky Way unveiled by DBSCAN and Gaia EDR3: Mapping the star forming regions within 1.5 kpc. *Astron. Astrophys.* **2022**, *664*, A175. [CrossRef]
67. Noormohammadi, M.; Khakian Ghomi, M.; Haghi, H. The membership of stars, density profile, and mass segregation in open clusters using a new machine learning-based method. *Mon. Not. R. Astron. Soc.* **2023**, *523*, 3538–3554. [CrossRef]
68. Cvrček, V.; Romaniello, M.; Šára, R.; Freudling, W.; Ballester, P. Stellar parameter prediction and spectral simulation using machine learning: A systematic comparison of methods with HARPS observational data. *Astron. Astrophys.* **2025**, *693*, A256. [CrossRef]
69. Huijse, P.; De Ridder, J.; Eyer, L.; Rimoldini, L.; Holl, B.; Chornay, N.; Roquette, J.; Nienartowicz, K.; Jevardat de Fombelle, G.; Fritzewski, D.J.; et al. Learning novel representations of variable sources from multi-modal Gaia data via autoencoders. *Astron. Astrophys.* **2025**, *701*, A150. [CrossRef]
70. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *arXiv* **2014**, arXiv:1406.2661.
71. Manteiga, M.; Santoveña, R.; Álvarez, M.A.; Dafonte, C.; Penedo, M.G.; Navarro, S.; Corral, L. Disentangling stellar atmospheric parameters in astronomical spectra using generative adversarial neural networks: Application to Gaia/RVS parameterisation. *Astron. Astrophys.* **2025**, *694*, A326. [CrossRef]
72. Pan, J.S.; Ting, Y.S.; Yu, J. Astroconformer: The prospects of analysing stellar light curves with transformer-based deep learning models. *Mon. Not. R. Astron. Soc.* **2024**, *528*, 5890–5903. [CrossRef]
73. Donoso-Oliva, C.; Becker, I.; Protopapas, P.; Cabrera-Vives, G.; Vishnu, M.; Vardhan, H. ASTROMER. A transformer-based embedding for the representation of light curves. *Astron. Astrophys.* **2023**, *670*, A54. [CrossRef]
74. Cabrera-Vives, G.; Moreno-Cartagena, D.; Astorga, N.; Reyes-Jainaga, I.; Förster, F.; Huijse, P.; Arredondo, J.; Muñoz Arancibia, A.M.; Bayo, A.; Catelan, M.; et al. ATAT: Astronomical Transformer for time series and Tabular data. *Astron. Astrophys.* **2024**, *689*, A289. [CrossRef]
75. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762. [CrossRef]
76. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. *arXiv* **2013**, arXiv:1312.6114. [CrossRef]
77. Zhang, J.; Bu, Y.; Wu, H.; Zhang, Y.; Li, S.; Lei, Z.; Yi, Z.; Kong, X.; Liu, M. Finding New Hot Subdwarf Stars in SDSS Images. *Astron. J.* **2025**, *170*, 94. [CrossRef]
78. Németh, P.; Kawka, A.; Vennes, S. A selection of hot subluminous stars in the GALEX survey—II. Subdwarf atmospheric parameters. *Mon. Not. R. Astron. Soc.* **2012**, *427*, 2180–2211. [CrossRef]

79. Ambrosch, M.; Viscasillas Vázquez, C.; Solano, E.; Ulla, A.; Pérez-Couto, X.; Pérez-Fernández, E.; Medžiūnas, A.; Manteiga, M.; Dafonte, C.; Drazdauskas, A.; et al. Detection of hot subdwarf binaries and He-poor hot subdwarf stars using machine learning methods and a large sample of Gaia XP spectra. *Astron. Astrophys.* **2026**, *708*, A23. [CrossRef]
80. Bu, Y.; Lei, Z.; Zhao, G.; Bu, J.; Pan, J. Searching for Hot Subdwarf Stars from the LAMOST Spectra. I. Method. *Astrophys. J. Suppl. Ser.* **2017**, *233*, 2. [CrossRef]
81. Bu, Y.; Zeng, J.; Lei, Z.; Yi, Z. Searching for Hot Subdwarf Stars from the LAMOST Spectra. III. Classification of Hot Subdwarf Stars in the Fourth Data Release of LAMOST Using a Deep Learning Method. *Astrophys. J.* **2019**, *886*, 128. [CrossRef]
82. Lei, Z.; Bu, Y.; Zhao, J.; Németh, P.; Zhao, G. Searching for hot subdwarf stars from the LAMOST Spectra. II. Pure spectroscopic identification method for hot subdwarfs. *Publ. Astron. Soc. Jpn.* **2019**, *71*, 41. [CrossRef]
83. Tan, L.; Mei, Y.; Liu, Z.; Luo, Y.; Deng, H.; Wang, F.; Deng, L.; Liu, C. A Robust Identification Method for Hot Subdwarfs Based on Deep Learning. *Astrophys. J. Suppl. Ser.* **2022**, *259*, 5. [CrossRef]
84. Tahir, M.; Yude, B.; Mehmood, T.; Bashir, S.; Zhenping, Y.; Awais, M. Classification of LAMOST spectra of B-type and hot subdwarf stars using kernel support vector machine. *Sci. Rep.* **2024**, *14*, 16815. [CrossRef]
85. Viscasillas Vázquez, C.; Solano, E.; Ulla, A.; Ambrosch, M.; Álvarez, M.A.; Manteiga, M.; Magrini, L.; Santoveña-Gómez, R.; Dafonte, C.; Pérez-Fernández, E.; et al. Advanced classification of hot subdwarf binaries using artificial intelligence techniques and Gaia DR3 data. *Astron. Astrophys.* **2024**, *691*, A223. [CrossRef]
86. Ratzloff, J.K.; Barlow, B.N.; Németh, P.; Corbett, H.T.; Walser, S.; Galliher, N.W.; Glazier, A.; Howard, W.S.; Law, N.M. Hot Subdwarf All Southern Sky Fast Transit Survey with the Evryscope. *Astrophys. J.* **2020**, *890*, 126. [CrossRef]
87. Xing, K.; Zong, W.; Silvotti, R.; Fu, J.N.; Charpinet, S.; Cang, T.; Hermes, J.J.; Ma, X.Y.; Wang, H.; Wang, X.; et al. Flare Hunting in Hot Subdwarf and White Dwarf Stars from Cycles 1–5 of TESS Photometry. *Astrophys. J. Suppl. Ser.* **2024**, *271*, 57. [CrossRef]
88. Lin, C.L.; Apai, D.; Giampapa, M.S.; Ip, W.H. Scalable, Advanced Machine Learning Based Approaches for Stellar Flare Identification: Application to TESS Short-cadence Data and Analysis of a New Flare Catalog. *Astron. J.* **2024**, *168*, 234. [CrossRef]
89. Lei, Z.; Dong, Y.; Kou, B.; Feng, M.; Hu, K.; Bu, Y.; Zhao, J. Deep Learning–driven Atmospheric Parameter Prediction for Hot Subdwarf Stars with Synthetic and Observed Spectra. *Astrophys. J. Suppl. Ser.* **2026**, *283*, 16. [CrossRef]
90. Cheng, Z.; Kong, X.; Wu, T.; Zhang, A.; Liu, B.; Bu, Y.; Lei, Z.; Zhang, Y.; Yi, Z.; Liu, M. Se-ResNet+SVM Model: An Effective Method of Searching for Hot Subdwarfs from LAMOST. *Astrophys. J. Suppl. Ser.* **2024**, *274*, 2. [CrossRef]
91. Vincent, O.; Bergeron, P.; Dufour, P. Data-driven selection and spectral classification of white dwarf stars. *Mon. Not. R. Astron. Soc.* **2023**, *521*, 760–771. [CrossRef]
92. Zhang, S.; Zhang, Y.; Liu, C. A White Dwarf Catalog from LAMOST DR11 Using Deep Learning. *Astrophys. J. Suppl. Ser.* **2025**, *279*, 36. [CrossRef]
93. Zhang, J.; Bu, Y.; Zhang, M.; Xie, D.; Yi, Z. A Multiple-detection-heads Machine Learning Algorithm for Detecting White Dwarfs. *Astrophys. J. Suppl. Ser.* **2025**, *276*, 53. [CrossRef]
94. Byrne, X.; Bonsor, A.; Rogers, L.K.; Manser, C.J. Semi-supervised spectral classification of DESI white dwarfs by dimensionality reduction. *Mon. Not. R. Astron. Soc.* **2024**, *535*, 2246–2259. [CrossRef]
95. Byrne, X.; Bonsor, A.; Rogers, L.K.; Badenas-Agusti, M. Finding rare classes in large data sets: The case of polluted white dwarfs from Gaia XP spectra. *RAS Tech. Instrum.* **2025**, *4*, rzaf044. [CrossRef]
96. Munday, J.; Tremblay, P.E.; Pelisoli, I.; Killestein, T.; Martikainen, J.; Jones, D.; Bédard, A.; Sowicka, P. Classifying white dwarfs from multi-object spectroscopy surveys with machine learning. *Mon. Not. R. Astron. Soc.* **2026**, *546*, stag246. [CrossRef]
97. Kalita, S.; Uniyal, A.; Bulik, T.; Mizuno, Y. Identifying highly magnetized white dwarfs: A dimensionality reduction framework for estimating magnetic fields. *J. High Energy Astrophys.* **2026**, *53*, 100598. [CrossRef]
98. Pérez-Couto, X.; Pallas-Quintela, L.; Manteiga, M.; Villaver, E.; Dafonte, C. Identifying New High-confidence Polluted White Dwarf Candidates Using Gaia XP Spectra and Self-organizing Maps. *Astrophys. J.* **2024**, *977*, 31. [CrossRef]
99. Kao, M.L.; Hawkins, K.; Rogers, L.K.; Bonsor, A.; Dunlap, B.H.; Sanders, J.L.; Montgomery, M.H.; Winget, D.E. Hunting for Polluted White Dwarfs and Other Treasures with Gaia XP Spectra and Unsupervised Machine Learning. *Astrophys. J.* **2024**, *970*, 181. [CrossRef]
100. Li, J.; Ting, Y.S.; Rix, H.W.; Green, G.M.; Hogg, D.W.; Ren, J.J.; Müller-Horn, J.; Seeburger, R. Identification of 30,000 White Dwarf–Main-sequence Binary Candidates from Gaia DR3 BP/RP (XP) Low-resolution Spectra. *Astrophys. J. Suppl. Ser.* **2025**, *279*, 47. [CrossRef]
101. Echeverry, D.; Torres, S.; Rebassa-Mansergas, A.; Ferrer-Burjachs, A. Random Forest classification of Gaia DR3 white dwarf-main sequence spectra: A feasibility study. *Astron. Astrophys.* **2022**, *667*, A144. [CrossRef]
102. Tay, I.K.C.; Korol, V.; Lechien, T. Disentangling the Galactic binary zoo: Machine learning classification of stellar remnant binaries in LISA data. *arXiv* **2026**, arXiv:2603.06341. [CrossRef]
103. Badenas-Agusti, M.; Viaña, J.; Vanderburg, A.; Blouin, S.; Dufour, P.; Xu, S.; Sha, L. cecilia: A machine learning-based pipeline for measuring metal abundances of helium-rich polluted white dwarfs. *Mon. Not. R. Astron. Soc.* **2024**, *529*, 1688–1714. [CrossRef]

104. Badenas-Agusti, M.; Xu, S.; Vanderburg, A.; De, K.; Dufour, P.; Rogers, L.K.; Hoyos, S.; Blouin, S.; Viaña, J.; Bonsor, A.; et al. A machine-learning compositional study of exoplanetary material accreted onto five helium-atmosphere white dwarfs with cecilia. *Mon. Not. R. Astron. Soc.* **2025**, *540*, 746–773. [CrossRef]
105. Ishida, E.E.O. et al. [COIN Collaboration] Optimizing spectroscopic follow-up strategies for supernova photometric classification with active learning. *Mon. Not. R. Astron. Soc.* **2019**, *483*, 2–18. [CrossRef]
106. García-Jara, G.; Protopapas, P.; Estévez, P.A. Improving Astronomical Time-series Classification via Data Augmentation with Generative Adversarial Networks. *Astrophys. J.* **2022**, *935*, 23. [CrossRef]
107. Shah, V.G.; Rehemtulla, N.; Miller, A.A.; Sharma Chaudhary, S.; Coughlin, M.W.; Le Calloch, A.; Graham, M.J.; Castaneda Jaimes, J.; Jegou du Laz, T.; Mahabal, A.A.; et al. Leveraging Multimodality for Real-Time Classification of Transients and Variables found by the Zwicky Transient Facility. *arXiv* **2026**, arXiv:2607.00228. [CrossRef]
108. Ghosh, A.; Oldag, D.; Tauraso, M.; Connolly, A.J.; Ferguson, P.; Jones, D.; Khullar, G.; Sasli, A.; Venkatesh, S.; Wang, G.; et al. Hyrax: An Extensible Framework for Rapid ML Experimentation and Unsupervised Discovery in the Era of Rubin, Roman, and Euclid. *arXiv* **2026**, arXiv:2605.18959. [CrossRef]
109. Walmsley, M.; Smith, L.; Lintott, C.; Gal, Y.; Bamford, S.; Dickinson, H.; Fortson, L.; Kruk, S.; Masters, K.; Scarlata, C.; et al. Galaxy Zoo: Probabilistic morphology through Bayesian CNNs and active learning. *Mon. Not. R. Astron. Soc.* **2020**, *491*, 1554–1574. [CrossRef]
110. Bonnet-Guerrini, R.; Sanchez, B.; Fouchez, D.; Racine, B.; Guy, M.; Sabalbal, M.; Yassine, M.; Piuri, V. Interpretable Human-Label-Free Deep Learning for Real-Bogus Classification with Uncertainty Quantification. *arXiv* **2026**, arXiv:2607.05393. [CrossRef]
111. Baty, H. A hands-on introduction to Physics-Informed Neural Networks for solving partial differential equations with benchmark tests taken from astrophysics and plasma physics. *arXiv* **2024**, arXiv:2403.00599. [CrossRef]
112. Athalathil, J.J.; Vaidya, B.; Kundu, S.; Upendran, V.; Cheung, M.C.M. Surface Flux Transport Modeling Using Physics-informed Neural Networks. *Astrophys. J.* **2024**, *975*, 258. [CrossRef]
113. Zorotovic, M.; Schreiber, M.R.; Gänsicke, B.T.; Nebot Gómez-Morán, A. Post-common-envelope binaries from SDSS. IX: Constraining the common-envelope efficiency. *Astron. Astrophys.* **2010**, *520*, A86. [CrossRef]
114. Ge, H.; Tout, C.A.; Chen, X.; Kruckow, M.U.; Chen, H.; Jiang, D.; Li, Z.; Liu, Z.; Han, Z. The Common Envelope Evolution Outcome-A Case Study on Hot Subdwarf B Stars. *Astrophys. J.* **2022**, *933*, 137. [CrossRef]
115. Østensen, R.H.; Oreiro, R.; Solheim, J.E.; Heber, U.; Silvotti, R.; González-Pérez, J.M.; Ulla, A.; Pérez Hernández, F.; Rodríguez-López, C.; Telting, J.H. A survey for pulsating subdwarf B stars with the Nordic Optical Telescope. *Astron. Astrophys.* **2010**, *513*, A6. [CrossRef]
116. Romero, A.D.; Campos, F.; Kepler, S.O. The age-metallicity dependence for white dwarf stars. *Mon. Not. R. Astron. Soc.* **2015**, *450*, 3708–3723. [CrossRef]
117. Cummings, J.D.; Kalirai, J.S.; Tremblay, P.E.; Ramirez-Ruiz, E.; Choi, J. The White Dwarf Initial-Final Mass Relation for Progenitor Stars from 0.85 to 7.5 $M_{\odot}$. *Astrophys. J.* **2018**, *866*, 21. [CrossRef]
118. El-Badry, K.; Rix, H.W.; Weisz, D.R. An Empirical Measurement of the Initial-Final Mass Relation with Gaia White Dwarfs. *Astrophys. J.* **2018**, *860*, L17. [CrossRef]
119. Gaensicke, B.; Tremblay, P.E.; Barstow, M.; Bono, G.; Burleigh, M.; Casewell, S.; Dhillon, V.; Farihi, J.; Garcia-Berro, E.; Geier, S.; et al. Stellar archaeology with Gaia: The Galactic white dwarf population. *arXiv* **2015**, arXiv:1506.02653. [CrossRef]
120. Kilic, M.; Bergeron, P.; Dame, K.; Hambly, N.C.; Rowell, N.; Crawford, C.L. The age of the Galactic stellar halo from Gaia white dwarfs. *Mon. Not. R. Astron. Soc.* **2019**, *482*, 965–979. [CrossRef]
121. Farihi, J. Circumstellar debris and pollution at white dwarf stars. *New Astron. Rev.* **2016**, *71*, 9–34. [CrossRef]
122. Harrison, J.H.D.; Bonsor, A.; Kama, M.; Buchan, A.M.; Blouin, S.; Koester, D. Bayesian constraints on the origin and geology of exoplanetary material using a population of externally polluted white dwarfs. *Mon. Not. R. Astron. Soc.* **2021**, *504*, 2853–2867. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.