Unsupervised White Dwarf Spectral Classification by Dimensionality Reduction

Xander Byrne ,[1]★ Amy Bonsor ,[1] Laura K. Rogers [1]
[1]Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

As a new generation of large-sky spectroscopic surveys comes online, the enormous data volume poses challenges in classifying spectra. Although machine-learning-based spectral classifiers exist, they are usually trained in a supervised way, and their performance thus relies heavily on training data. We explore the use of dimensionality reduction to transform the high-dimensional space of white dwarf spectra, from the Dark Energy Spectroscopic Instrument's Early Data Release, into a two-dimensional map. This provides a new view of the dataset which groups together white dwarfs of different spectral classes. By cropping the spectra to windows around particular spectral lines, prior knowledge about the wavelengths of important spectral features can be exploited to isolate particular spectral classes. Although dimensionality reduction is a fully unsupervised technique, it can also be used in a supervised way, allowing spectra from other sources to be classified by visualising their similarity to the spectra in the dataset. With upcoming surveys promising tens of millions of spectra, this work highlights the potential for unsupervised techniques as efficient means of classification and dataset visualisation.

Key words: methods: data analysis – stars: white dwarfs – surveys

1 INTRODUCTION

A suite of Stage IV large-sky spectroscopic surveys promises to enormously increase the number of sources with intermediate-resolution spectroscopy, enabling advances across all areas of astronomy. These surveys include:

- **4MOST** (4-metre Multi-Object Spectroscopic Telescope; de Jong et al. 2016): nine surveys which aim to obtain accurate radial velocities and chemical abundances for millions of Galactic stars, probe the history of supermassive black hole accretion through a spectroscopic sample of about 1 million active galactic nuclei, and constrain cosmological parameters by obtaining redshifts for 10-20 million emission line galaxies;
- **DESI** (Dark Energy Spectroscopic Instrument; DESI Collaboration et al. 2016a,b), which plans to study the growth of structure in the Universe through observations of baryon acoustic oscillations and redshift-space distortions, and trace the dark matter distribution by obtaining redshifts for 30 million galaxies and quasars at $1.0 \lesssim z \lesssim 3.5$. When these faint targets are obscured by moonlight over Iolkam Du'ag / Kitt Peak, the targeting switches to surveys of Milky Way stars and bright galaxies;
- **SDSS-V** (Sloan Digital Sky Survey V; Kollmeier et al. 2017): an all-sky survey from two 2.5 m telescopes (one in each hemisphere), this aims to record multi-epoch spectra for over six million targets, including five million Galactic stars;
- **WEAVE** (William Herschel Telescope Enhanced Area Velocity Explorer; Dalton et al. 2012): a northern-hemisphere complement to 4MOST, WEAVE will record accurate spectroscopic velocities of Galactic disk and halo stars, IFU H I maps of $\sim 10^4$ low-redshift galaxies, and wide-area views of galaxy clusters.

All four of these surveys have achieved first light, with SDSS-V's first data release (SDSS DR19; Almeida et al. 2023) and DESI's early data release (DESI EDR; DESI Collaboration et al. 2023) delivered last year. As these spectroscopic surveys begin to return enormous quantities of data, and with the exabyte-scale Legacy Survey of Space and Time (LSST; Ivezić et al. 2019) on the horizon, automated techniques will become absolutely necessary in aiding the extraction of scientific results from the vast wealth of data being collected.

Entries in astronomical datasets often have high dimensionality. They may be images of hundreds of pixels, light curves with thousands of measurements, or spectra with thousands of fluxes. Any *structure* in a dataset – e.g., clusters or sequences – can be investigated quantitatively by representing each entry as a vector in a very high-dimensional space: a dataset of N spectra with D flux bins can be thought of as a set of N points in D-dimensional data space. Visualisation of datasets in more than two or three dimensions is challenging, but whereas a D-dimensional data point might ostensibly live in $\mathbb{R}^D$, it is very likely that every point in the dataset is located on or near a much smaller submanifold within this space. In a spectrum, the flux in a given bin will usually be similar to the flux in adjacent bins. Similarly, two different sources of the same astrophysical type (e.g. two white dwarfs of the same spectral classification) will have similar spectra, and will thus be nearby in the data space. This massively reduces the effective dimensionality of the dataset, compared to the dimensionality of the raw data points themselves. The observation that many datasets exist near low-dimensional submanifolds of the data space has motivated the development of several *dimensionality reduction* (DR) techniques, including Principal Com-

★ E-mail: xbyrne@ast.cam.ac.uk

© 2024 The Authors

ponent Analysis (PCA; Pearson 1901), diffusion maps (Coifman & Lafon 2006; Lafon & Lee 2006), Locally Linear Embedding (LLE; Roweis & Saul 2000), t-distributed Stochastic Neighbour Embedding (tSNE; van der Maaten & Hinton 2008), and Uniform Manifold Approximation and Projection (UMAP; McInnes et al. 2018). These methods calculate a two-dimensional map of the dataset – known as an *embedding* – in which the distances between pairs of nearby points in the data space are preserved as far as possible.

DR techniques have found widespread use in a variety of areas of astronomy, including the analysis of low-resolution spectra. Boroson & Green (1992) use PCA to explore correlations between the physical parameters of 87 low-redshift quasars observed in a small spectroscopic campaign. Richards et al. (2009) employ PCA alongside diffusion maps to predict redshifts for several thousand low-redshift galaxy spectra observed by SDSS DR6. Hawkins et al. (2021) use tSNE to reduce the dimensionality of a set of low-resolution ($R \sim 750$) optical spectra of stars in the Hobby-Eberly Telescope Dark Energy Experiment (HETDEX; Gebhardt et al. 2021) to estimate a star's $T_{\rm eff}$, $\log g$, and [Fe/H], finding 416 metal-poor candidate stars. Kao et al. (2024) apply UMAP to low-resolution ($R \sim 50$) Gaia DR3 XP coefficients of white dwarf spectra, isolating a group of 465 objects of which 90 are known *polluted* WDs, showing metal absorption features; the remaining 375 candidates are subject to an ongoing high-resolution spectroscopic campaign.

White dwarfs (WDs) are the final state of the $\gtrsim 97$ per cent of stars with a zero-age mass of less than 9–12 $M_\odot$ (Lauffer et al. 2018; Althaus et al. 2010, 2021). Without internal nuclear fusion, these stellar cinders gradually cool on well-characterised timescales, making them important tracers of the evolution and assembly of the Galaxy (e.g., Winget et al. 1987; Tremblay et al. 2014). The spectrum of a white dwarf usually deviates from that of a black body due to absorbing species in its atmosphere. With a typical mass of $\sim 0.6\,M_\odot$ and radius of $\sim R_E$, the extreme gravitational field gives WDs a strongly stratified structure. Metals diffuse down through WD atmospheres on timescales much shorter than their cooling age (Schatzman 1945; Paquette et al. 1986; Koester 2009; Wyatt et al. 2014), leaving thin atmospheres dominated by H and/or He. Most WD spectra show absorption features due to these light elements, and a WD is classified respectively as DA or DB if spectral features of H or He are present. As these WDs cool below particular temperatures, atomic transitions are no longer excited ($\lesssim 5000$ K for H; $\lesssim 11000$ K for He) and the absorption lines gradually fade away, giving a featureless spectrum (DC). In certain WDs, the convection zone reaches deep into the interior, dredging up carbon from the inert core and imprinting molecular C_2 features in the spectrum (DQ; Fontaine et al. 1984; Koester et al. 1982, 2020; Blouin et al. 2023). In other cases, instabilities in remnant planetary systems can cause planetesimals to be tidally disintegrated and accreted by WDs, lending metal absorption features to the spectrum (Bonsor et al. 2011; Frewen & Hansen 2014; Mustill et al. 2018; Maldonado et al. 2020). These so-called *polluted* WDs (DZ) provide rare insights into the bulk composition and geology of rocky exoplanets.

Following the release of the DESI EDR (DESI Collaboration et al. 2023), the spectra of 3673 WD candidates targeted in this data release were visually inspected and classified by Manser et al. (2024), spectroscopically confirming 2706 WDs, of which 1400 had not been identified by previous spectroscopic surveys. In particular, 152 polluted WDs were found, of which 121 were newly discovered or previously not classified as polluted: the higher resolution of DESI spectra enables the identification of metal lines not visible in lower-resolution spectra. However, visual classification requires significant amounts of expert time, with substantial increases in target numbers in the future ($\sim 70\,000$ WD candidates from the DESI first data release; Cooper et al. 2023), increasing the benefits of automation.

One class of methods for automatically classifying and analysing WD spectra that has achieved some success is supervised machine learning (for example, Yang et al. 2020; Tan et al. 2023; García-Zamora et al. 2023; Vincent et al. 2023, 2024). However, a drawback of supervised machine learning techniques is the reliance on the training set used to train models. The 'imbalanced learning problem' describes the difficulties of training supervised classifiers on datasets with large class imbalances (He & Garcia 2009; Johnson & Khoshgoftaar 2019). Such class imbalances are inevitable in WD datasets as some spectral types are much rarer than others. Additionally, incorrect labels in the training set limit the accuracy of the resulting model (e.g., Frenay & Verleysen 2014). Unsupervised procedures such as dimensionality reduction do not rely on an external training set; they merely reveal structures present within the dataset itself.

In this work, we apply dimensionality reduction – specifically, tSNE – to white dwarf spectra from the DESI EDR, exploring the method's success in classifying and characterising WD spectra in an unsupervised way. We emphasise that these methods could equally be applied to sets of main-sequence stars, quasars, or galaxies, all of which are targeted by the aforementioned spectroscopic surveys. The paper proceeds as follows. In Section 2, we describe the dataset and tSNE method in more detail. In Section 3, we demonstrate this method's ability to identify clusters and sequences in the DESI EDR WD catalogue, a way to incorporate prior knowledge about the location of distinctive spectral lines, and a method of estimating the classification of spectra external to the catalogue. Section 4 discusses the application of this method to the bulk classification of large spectroscopic WD datasets. Section 5 summarises our work.

2 METHODS

2.1 Data

The DESI EDR contains $N = 3673$ candidate WDs from the catalogue compiled by Gentile Fusillo et al. (2019) whose exposures include a median signal-to-noise ratio > 0.5 in at least one of DESI's three spectral arms (Manser et al. 2024). These publicly available[1] exposures were obtained and restacked. The spectra span 3600–9824 Å, with a wavelength spacing of $\Delta\lambda = 0.8$ Å, giving $D = 7781$ flux bins for each spectrum[2].

2.2 Dimensionality reduction

A spectrum can be represented by a high-dimensional vector, where each component corresponds to the flux in a given wavelength bin. Spectra produced by the DESI pipeline have 7781 flux bins, so each spectrum can be mapped to a 7781-dimensional vector. The vectors are not uniformly-distributed in $\mathbb{R}^{7781}$; spectra with shared spectral features are clustered together, as their vector components corresponding to wavelength bins near shared spectral features will take similar values. For example, component 3703 corresponds to a wavelength of 6562.4 Å, very close to the H α line. The normalised spectra of DA WDs will have lower values in their 3703rd component than spectra *without* any H absorption features DAs will hence – at least

[1] https://data.desi.lbl.gov/public/index.html

[2] Note that the full-width-half-maximum line width of the DESI spectrographs is ≈ 1.8 Å (DESI Collaboration et al. 2022); these data are therefore slightly oversampled.

in this component, and by extension other components – be close together. This inherent structure in the data means that each point in the dataset lies on or close to a submanifold of the data space with much lower dimensionality. DR techniques seek to find a low- (usually two-) dimensional representation of the structure of the high-dimensional dataset.

A popular non-linear dimensionality reduction technique is *t*-distributed Stochastic Neighbour Embedding (*t*SNE; van der Maaten & Hinton 2008). Briefly, *t*SNE attempts to map a set of high-dimensional vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N \in \mathbb{R}^D$ into a set of two-dimensional embedding vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N \in \mathbb{R}^2$, in such a way as to preserve the 'similarity' between each pair of data points. The details of how this is achieved are given in Appendix A (see also van der Maaten & Hinton 2008). Many other DR techniques, such as Uniform Manifold Approximation and Projection (UMAP; McInnes et al. 2018) and LargeVis (Tang et al. 2016) can be shown to belong to the same family of procedures, but with differing definitions of similarity (see Appendix C of McInnes et al. 2018).

2.3 Data preprocessing

Before applying DR, it is of paramount importance to preprocess the spectra, as shown in Fig. 1. Some spectra show large spikes, due to cosmic rays or detector artefacts. Interpreted as vectors, this corresponds to some components of the vector being very large, and as a result the vector would be very far away in data space from where it would be without the artefacts. These spikes usually correspond to pixels with very low signal-to-noise: pixels with $\mathrm{S/N} < 0.2$ are linearly interpolated. Fig. 1 shows that major, though not all, artefacts are removed by this step. Overzealous removal of artefacts could remove genuine spectral features, such as the sharp emission lines of cataclysmic variables (see Section 3.2.2): we found that the interpolation described improved the effectiveness of the subsequent dimensionality reduction as compared to, say, median boxcar smoothing.

Additionally, the spectra show a range of absolute scales: the median fluxes span three orders of magnitude. As we are interested in intrinsic properties of the WDs in this sample, rather than their distance from Earth, the spectra are normalised to zero mean and unit variance (see Fig. 1).

In its simplest form, dimensionality reduction treats all components of the vectors equally. However, some wavelengths are naturally of more significance than others, namely those of atomic absorption and emission features. To 'focus' the technique on a particular spectral line, we crop the spectrum to a window containing that line (as well as some continuum) before applying DR. This additional preprocessing step is explored in Section 3.2.

3 RESULTS

3.1 Applying *t*SNE to the full spectra

Dimensionality reduction was applied to the DESI EDR WD catalogue, using *t*SNE. The dimensionality reduction takes an average of 5.5 s on an Intel 3.60 GHz CPU. The resulting embedding is shown in Fig. 2(a). Each point in this embedding corresponds to an individual spectrum, and several clusters and sequences appear. There is a long V-shaped sequence stretching from the left, to the bottom, to the upper right of the embedding; it turns off into a cluster towards the centre right. A shorter second sequence extends from the top of the embedding towards the centre. Several strings and clumps are found in between and around these two sequences.

The nature of these structures is elucidated in Fig. 2(b) and (c), showing respectively the primary spectral class (acc. Manser et al. 2024) and effective temperature (acc. H-atmosphere model fitting of Gentile Fusillo et al. 2019). Several interesting features include:

- The V-shaped sequence consists of primarily DA white dwarfs;
- The second sequence in the upper part of the frame contains DBs, DZs, DQs, and DCs;
- DZs are found in various places[3]: (i) an island near to the second sequence, (ii) towards the cool end of the DA sequence, (iii) amongst the second sequence;
- Extragalactic sources (crosses) are mostly located in strings and small clumps around the second sequence;
- The majority of main-sequence stars (black star icons) are found towards the centre right, though their cluster merges somewhat with the cool end of the DA sequence. A few are also scattered elsewhere;
- The WDs in the DA sequence have been sorted approximately according to effective temperature, with hotter WDs towards the left of the sequence[4];

We emphasise that although the data are in this case labelled by class identified by visual inspection, the method would nonetheless provide a valuable analysis if the classes were *a priori* unknown, as will be the case with upcoming spectroscopic surveys. For example, it is well-known that approximately 80 per cent of WDs are DAs. As such, when applying DR to a WD dataset one would implicitly identify the largest cluster as mostly DAs.

That DR is capable of separating different classes of WD into groups in this way is ultimately due to similarities between spectra of the same spectral class, and differences between different classes. Expressed as vectors, DA spectra (say) are much closer to vectors of other DA spectra of similar H abundance and effective temperature, than to DB spectra (say). Note that this would not be true if the spectra had not been preprocessed (see Section 2.3): closer WDs would have brighter spectra than more distant WDs of the same spectral type, and hence spectral vectors with larger magnitudes. The spectral vectors would therefore be very far away from each other, and hence embedded far away from each other, despite being similar astrophysical objects. The same would be true of spectra with very large artefacts. The preprocessing procedure is thus crucial in removing any irrelevant aspects of the data.

The H-atmosphere model fitting of Gentile Fusillo et al. (2019) suggests that the WDs in the V-shaped sequence transition smoothly from $T_{\mathrm{eff}} \gtrsim 80000$ K through to $T_{\mathrm{eff}} \lesssim 4000$ K. We see that the well-sampled variation that must exist in the tilt of the spectral black-body continuum is preserved in two dimensions. The hottest WD in this sequence has a very similar spectrum to the second-hottest, so the two are embedded close together; the second-hottest has a similar spectrum to the third-hottest; etc. Anomalies of intermediate temperature at opposite ends of the sequence (see Fig. 2(c)) are main-sequence stars or other objects erroneously assigned a WD temperature. That T_{eff} "spans" the embedding identifies this parameter as that which describes a spectrum to first order: the parameter which primarily determines the spectrum is the temperature. If T_{eff} were not of interest,

[3] Kao et al. (2024) also find that DZs with different characteristics (primarily T_{eff}) are scattered by dimensionality reduction into multiple regions.

[4] This trend is also apparent in the second sequence, but because many of these are DBs and DQs, and the temperatures shown here are fitted assuming H-atmosphere models, the trend is less meaningful. However, the trend is borne out in the fitted temperatures of He-atmosphere models.

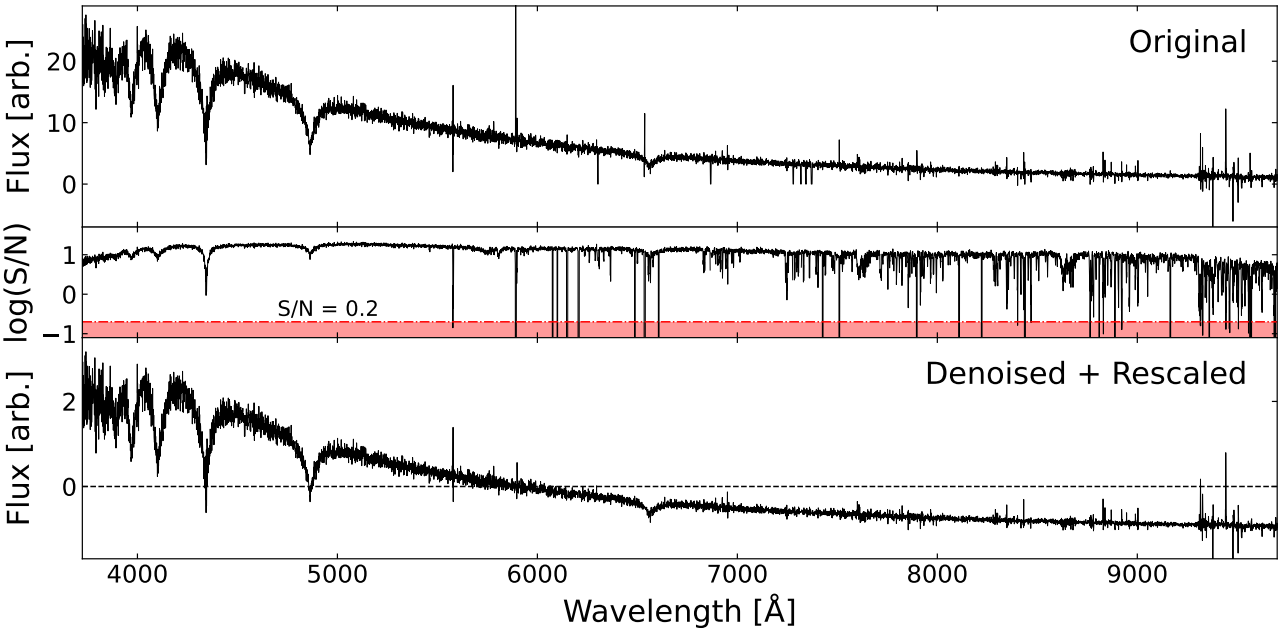

Figure 1. Preprocessing stages, as illustrated on a cherry-picked DESI EDR spectrum. The upper panel shows the raw spectrum, with several artefacts. The second panel shows that the signal-to-noise ratio is very low near many of these artefacts; where it falls below 0.2, the pixels are interpolated. The spectra are then rescaled to zero mean and unit variance, as shown in the lower panel. Major artefacts have been removed, though some smaller artefacts remain.

one could include preprocessing steps to erase any temperature information in the spectra (e.g., dividing by the best-fitting black body) before applying DR. Unable to pick up on this information, DR would then arrange the embedding according to second-order parameters, such as hydrogen abundance. In such a scenario, the spectrum of a very hot DA would appear similar to that of a very cool DA and the sequence would thus be collapsed in the embedding.

The transitions undergone by WDs as they gradually cool are also borne out in the spectral classifications of the embedding, shown in Fig. 2(b), though some subtleties emerge as a result of the classification system itself. For example, as the temperature of a DA falls below around $T_{\rm eff} \approx 5000$ K, hydrogen transitions are no longer excited, the Balmer lines gradually fade away, and the DA becomes a DC. However, the transition is gradual: the Balmer lines do not suddenly disappear below some sharp temperature cutoff. As such, the distinction between DA and DC is a fuzzy one, and whether a cool hydrogen-atmosphere WD spectrum is visually classified as DA or DC depends whether the classifier can subjectively identify H features against noise. This task is not trivial, but we argue that it is ultimately arbitrary: there is negligible physical difference between a DA with very weak Balmer lines, and a hydrogen-atmosphere DC.

Similar trends are borne out in the secondary sequence, which shows DBs at the top of the embedding transitioning into a mixture of DCs and DQs. The transition from DB to DC has a similar physical origin to the DA-to-DC transition discussed in the previous paragraph; as expected this transition occurs at a higher temperature as shown in Fig. 2(c). The DC transition temperatures appear roughly in line with theoretical predictions of ≈ 5000 K and ≈ 11000 K for H- and He-dominated atmospheres respectively. Aside from forming a DC, a cooling He-atmosphere WD may begin dredging carbon from the core, giving rise to C_2 Swan bands in the spectrum, which is then classified DQ (Fontaine et al. 1984). As this cooling occurs, He transitions fail to be excited and the He lines of a DB fade away. As with Balmer lines, He lines and Swan bands can be arbitrarily weak, creating a degree of blurriness between DBs, DQs, and DCs. That these three spectral types are all projected together in the second sequence at the top of the embedding reflects the smooth transition between them. Indeed, Fig. 2(c) shows that the DCs in this part of the embedding have temperatures[5] well in excess of the ≈ 5000 K DA-to-DC transition. These DCs must therefore have He-dominated atmospheres; if they were H-dominated then Balmer features would be visible. These observations therefore suggest that the two main clusters in the embedding correspond to H- and He-dominated atmospheres.

Briefly, we discuss the presence of a small number of DAs in the cool end of the He-atmosphere sequence (see Fig. 2(b), just above centre), of which there are 43. Of these, 24 are classified by Manser et al. (2024) as DAH, DAe, DAP, or some combination of these secondary classifications. We believe that these are found in the He-atmosphere sequence as the Swan bands of the DQs are mimicked by the Zeeman-split Balmer features of a DAH, emission features of a DAe, or other unidentified features in DAPs. The remaining 19 DAs here are either pure DAs, DABs, DAZs, or some combination of these. These DAs are invariably warm ($T_{\rm eff} \gtrsim 7000$ K) and have weak Balmer lines, making them appear similar to DCs of a similar temperature. It is thus understandable that these DAs are projected near to these warm (thus He-atmosphere) DCs.

It is somewhat surprising that there is a gap between the two sequences. Based on the observations outlined above, this region should be populated with DCs of

The automated nature of DR can protect against human error in the visual classification of WD spectra. Fig. 3 shows a zoom-in towards the top of the embedding around an island populated mostly by WDs with primary classification DZ. According to the extensive visual

[5] Although the temperatures shown in Fig. 2(c) are based on H-atmosphere models, the He-atmosphere model temperatures show the same results.

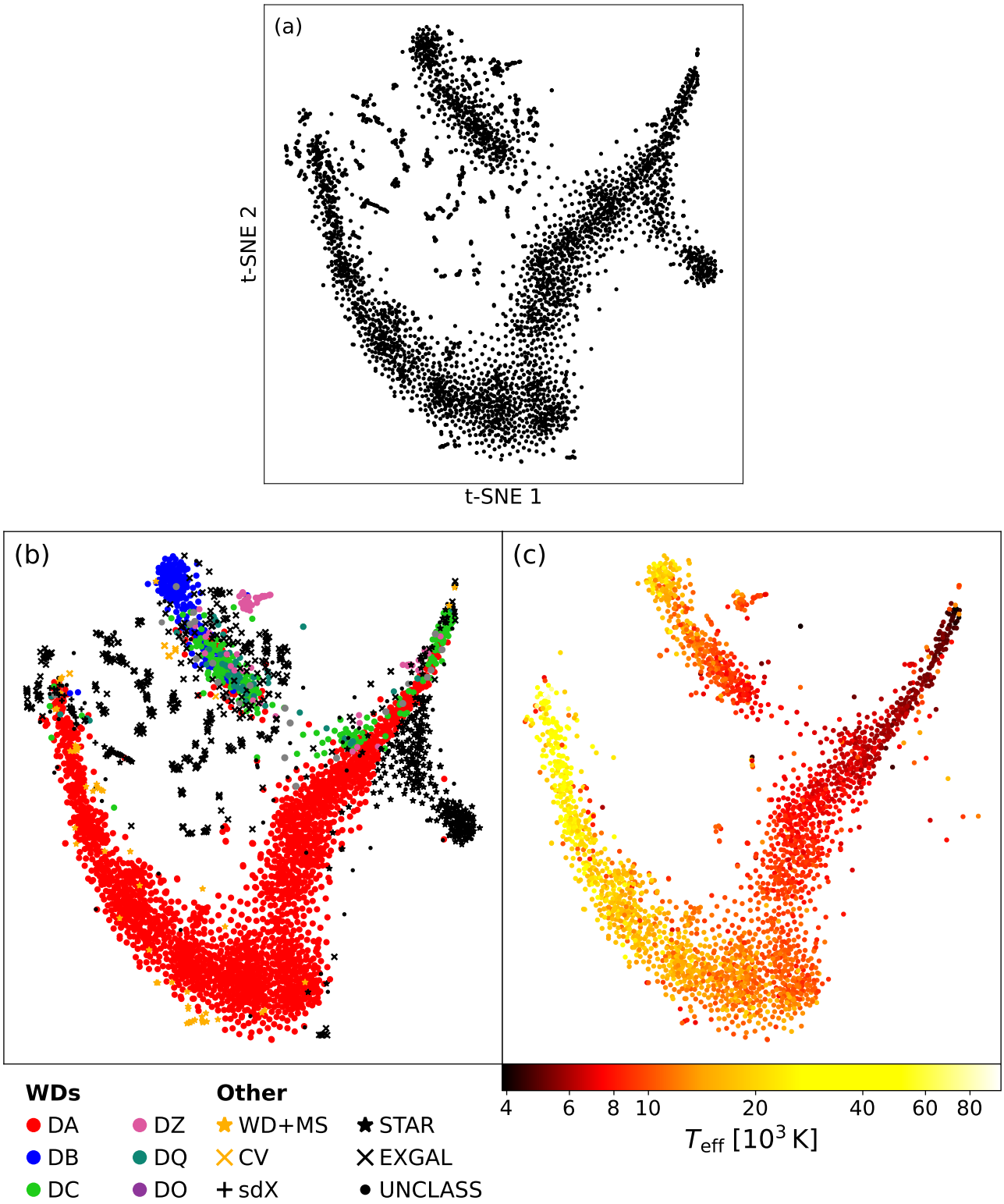

Figure 2. (a) *t*SNE embedding of DESI EDR spectra. The dimension-reduced embedding aims to reflect the pairwise distances between the high-dimensional spectra in a two-dimensional space; as such the axes are arbitrary. (b) The embedding is colour-coded according to the visual spectral classification of Manser et al. (2024). For WDs, the colour corresponds to primary spectral type: DAs (as well as DAZs, etc.) in red; DBs (and DBAs, etc.) in blue, and so on. Other sources not corresponding to individual WDs have different symbols (see key). The main feature of the embedding is the sequence of DAs (red), though several other clusters are clear (see text). (c) The embedding is colour-coded by effective temperature, according to the hydrogen-atmosphere WD model which best fits the sources' Gaia photometry (Gentile Fusillo et al. 2019). The DA sequence extends from hotter WDs on the left around to cooler WDs at the top right. Note that objects with a low probability of being a WD ($P_{\rm WD} < 0.75$; see Gentile Fusillo et al. 2015) are not assigned a temperature (Gentile Fusillo et al. 2019).

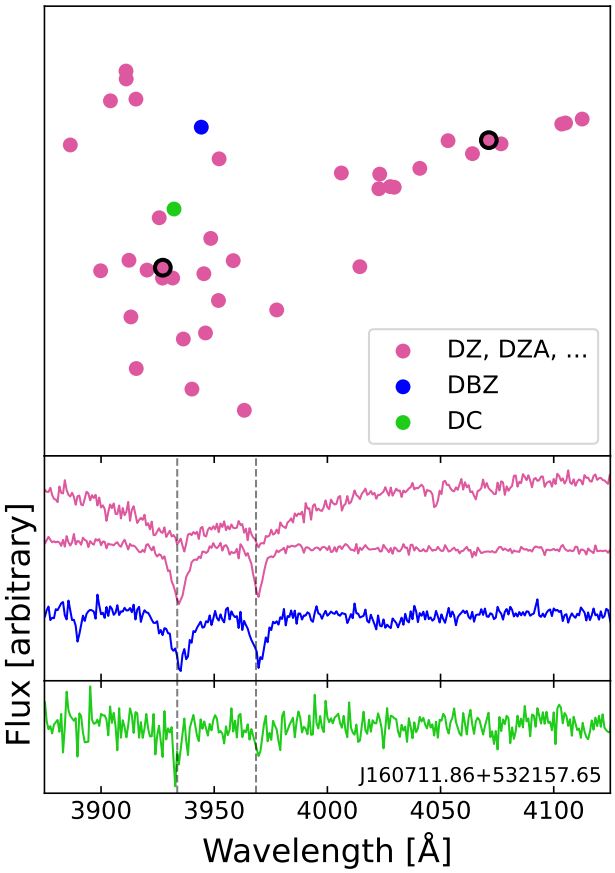

Figure 3. Zoomed-in view of the upper centre of the embedding shown in Fig. 2; spectra of four WDs. Except for two objects classified by Manser et al. (2024) as DBZ and DC, all the objects on this island have primary classification DZ. Two of these DZs (highlighted) have their spectra shown in the second panel, along with the DBZ spectrum. All three spectra show strong absorptions due to Ca II (dashed vertical lines). The final panel shows the putative DC, which also shows absorption features here.

classifications of Manser et al. (2024), this island also contains a DBZ (J133305.34+325400.11) and a DC (J160711.86+532157.65). These objects appear to have been grouped together owing to the presence of Ca H and K lines in all the spectra, including the putative DC. The presence of Ca absorption lines in the DC's spectrum suggests that this WD is in fact a DZ; indeed Kleinman et al. (2013) classify this source as a DZ based on its SDSS DR7 spectrum. The bottom panel of Figure 3 shows the absorption lines to be quite weak, and would probably not have been noticeable without zooming in. While large visual classification campaigns are reliable, this illustrates that errors are nonetheless possible, and that such errors can be easily identified with automated methods such as DR.

3.2 Classification using specific spectral regions

As mentioned in Section 2.3, DR treats all components of a vector – i.e., all wavelengths – equally, but wavelengths including spectral features are of particular interest. In this subsection we explore the use of cropping the spectrum, to a window around some particular spectral line, to see if DR can better separate individual spectral classes from the rest.

3.2.1 DBs

We first crop the spectra to the range $\lambda \in [5500, 6100]$ Å, a window which includes a He line at 5876 Å. The dimensionality of the spectra cropped to this wavelength range is $D = 750$, much lower than $D = 7781$ for the full spectra. The result of applying DR to this cropped dataset is shown in Fig. 4.

DR separates an island of around 200 objects, containing the vast majority ($\approx 180/200 = 90$ per cent) of DBs. The second panel of Fig. 4 demonstrates that the DBs (as well as DABs, etc.) which dominate the island are isolated primarily to their strong 5876 Å absorption. This shared feature causes the spectral vectors to be close together in data space; this is preserved by DR in two dimensions.

The third panel of this figure shows three examples of false positives that are projected onto this island despite not being DBs. For some of these spectra there is perhaps a very shallow or broad absorption feature, but it is difficult to tell by eye. These spectra also appear to be relatively noisy, so these spectra may have been projected nearby due to coincidentally-shared noise. The distinction between a genuine weak spectral feature and noise would be difficult for dimensionality reduction to ascertain, as it does not account for variations in flux between different spectral bins of the same spectrum, but rather differences in flux in the same bin between different spectra.

The bottom panel of Fig. 4 shows false negatives which are not on the island. The first two have very weak He absorption lines, and may have been identified as DBs through spectral features in other parts of the spectrum. The third object has a large noise feature at likely due to lower sensitivity near the edge of the blue arm of the DESI spectrograph.

3.2.2 Cataclysmic variables

Cataclysmic variables (CVs) are binary systems in which a donor star overfills its Roche lobe and transfers matter into an accretion disk around a white dwarf. The characteristic spectral features of a CV are strong emission lines from the disk, which depending on the orientation of the system may be double-peaked (Smak 1969; Huang 1972). As these unique features are most commonly seen in the Balmer series, we isolate CVs from the DESI EDR by cropping the spectra to three windows around H α, H β and H γ: the wavelength range selected is $\lambda \in [6500, 6600]$ Å $\cup$ $[4800, 4900]$ Å $\cup$ $[4300, 4400]$ Å. For each spectrum, a vector is created by cropping to each of these ranges, and concatenating the three 'sub-vectors'. Cropping the spectral vectors in these ranges gives $D = 375$-dimensional vectors. Following preprocessing and dimensionality reduction as above, the resulting embedding is shown in Fig. 5. All 12 of the CVs identified by Manser et al. (2024) are located on a small island at the top right of the embedding, no doubt as a result of their shared emission features, most of which are double-peaked. By zooming in to characteristic wavelength ranges, DR is therefore able to identify CVs from this sample with 100 per cent efficiency.

3.3 Classifying new white dwarf spectra against DESI EDR

In addition to bulk classification of a large dataset of N spectra, DR can aid the automatic classification of individual spectra external to said dataset. This can be achieved as follows.

(i) Interpolate the external spectrum to the wavelength grid of the DESI spectra, and append it, giving a dataset of $N + 1$ spectra: $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N, \mathbf{x}^* \in \mathbb{R}^D$.

(ii) Apply DR, giving $N + 1$ two-dimensional points: $\mathbf{y}_1, \ldots,$

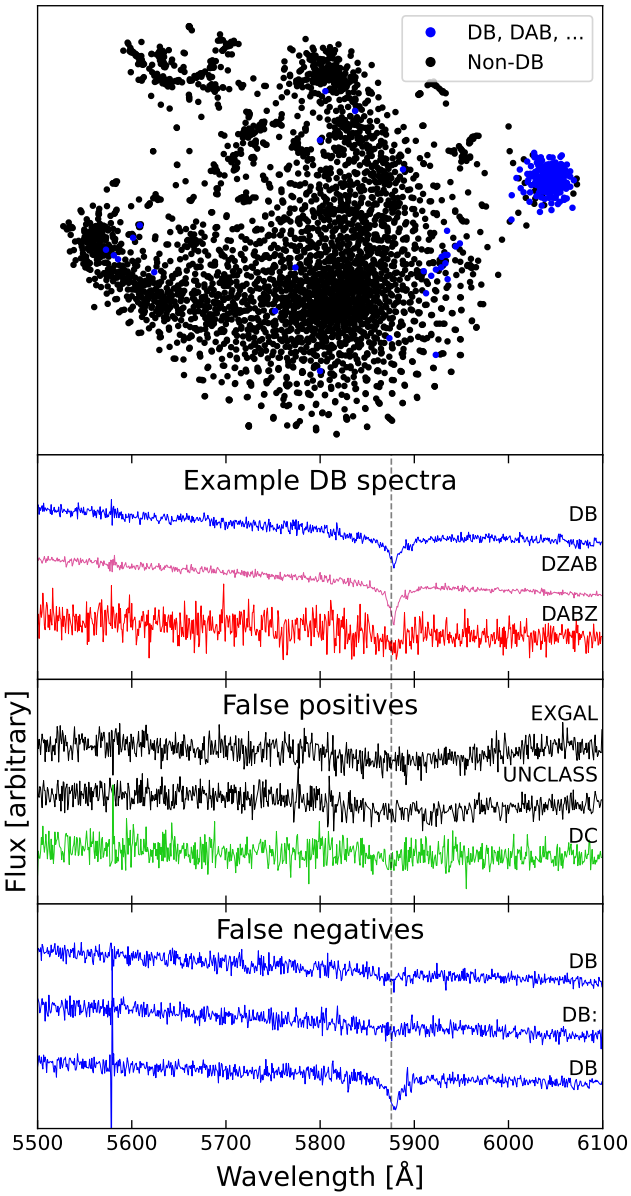

Figure 4. Top panel: dimensionality-reduced embedding of DESI spectra cropped between 5500–6100 Å. An island is isolated containing about 200 objects, of which about 180 are DBs (inc. DABs, DBZs etc.); some DBs are not on the island. Second panel: three spectra on the island are shown, each featuring a strong He absorption at 5876 Å (dashed line). Third panel: three objects on the island which are however not classified as DBs. Shallow absorptions are just about visible. Bottom panel: three DBs *not* located on the island. The first two spectra show very weak He absorption features; the third spectrum shows a large artefact.

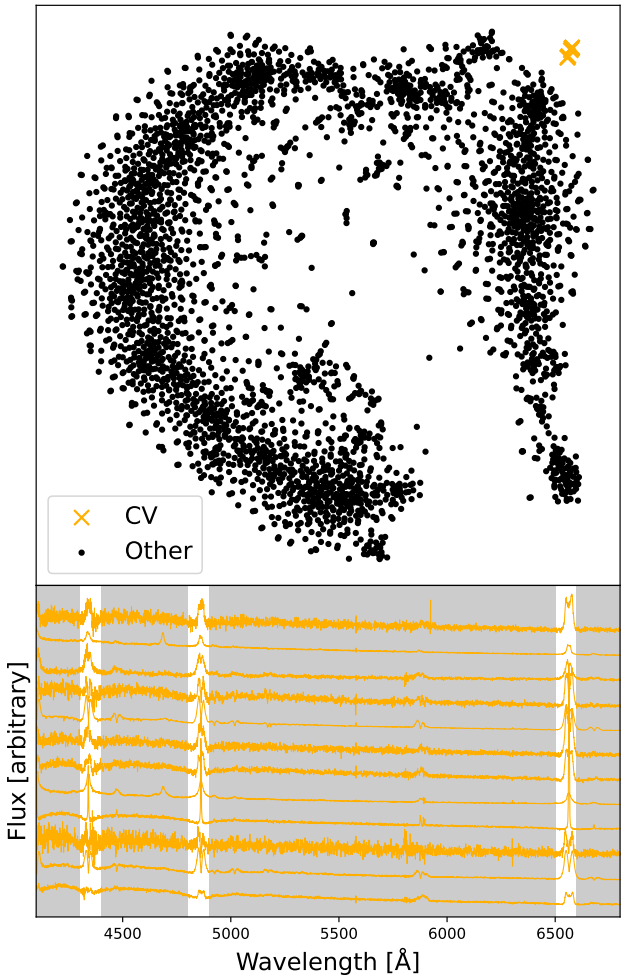

Figure 5. Top panel: dimensionality-reduced embedding of DESI spectra cropped to three regions around Hα, Hβ, and Hγ. An isolated island to the top right shows all 12 CVs in the sample, identified as similar to each other by shared emission features in these regions. Lower panel: spectra of the CVs. The ranges over which the spectra were cropped before DR is highlighted, and surrounds the strong (and in some cases double-peaked) emission features.

$\mathbf{y}_N, \mathbf{y}^* \in \mathbb{R}^2$. Since $N \gg 1$, the resulting embedding will not be noticeably different from the embedding produced from the original set of N spectra shown in Fig 2, though with an extra point $\mathbf{y}^*$

(iii) Identify the location of the external spectrum $\mathbf{y}^*$ in the new embedding. The spectrum will be projected near to similar spectra, and its spectrum $\mathbf{x}^*$ will thus likely be classified the same as the spectra projected near to it.

To demonstrate this application, a spectrum from the SDSS WD catalogue compiled by Gentile Fusillo et al. (2019) was selected at random from each of the following spectral classes: DA, DB, DZ, DC, DQ, DAO. Each was interpolated to the same wavelength grid and then (individually) appended to the DESI dataset described above. *t*SNE was again applied to these six sets of $N + 1$ spectra; the dimensionality reduction still only takes a few seconds. The results of this procedure are shown in Fig. 6.

The SDSS DA is projected within the main DA sequence. As such, if one did not *a priori* know the classification of this spectrum, one could apply the above procedure, note that this spectrum is projected among other DAs, and be confident that this spectrum is also of a DA, having been identified as a very similar spectrum by DR. Similarly, the SDSS DB and DZ are also projected into regions of the embedding with WDs with the same respective classification. The DCs, DQs and DAOs are also projected amongst other WDs in their respective classes, but these classes are not found in isolated islands even in the original embedding (Fig. 2); rather they are blended into other regions. For example, the DC is projected near to one end of the second sequence, which is occupied by not only DCs but also DQs, DZs, and even extragalactic sources. Classifying them in the

manner suggested above would therefore be difficult. Additional data would be necessary to distinguish it as a DC, including perhaps visual inspection. However, the location in the embedding nonetheless informs the classification. For example, for the DQ, it is projected far away from any DBs. Thus although it may not be immediately clear whether it is a DQ, DC, or DZ, when visually classifying it DR suggests that a human classifier need not look for helium features.

4 DISCUSSION

4.1 An aid for spectral classification

Approximately 70 000 WDs are targeted in the Milky Way Survey as part of the full DESI data release, expected in 2025 (Cooper et al. 2023). Whereas the entire DESI EDR WD catalogue has been classified thanks to an extensive visual classification campaign (Manser et al. 2024), the full data release will contain a factor of 20 more WD candidate spectra. We argue that automated methods such as DR could significantly reduce the expert time expended in classifying large spectroscopic surveys such as the DESI data release.

As the DESI EDR WD catalogue consists of commissioning and survey validation observations, it is unlikely to be precisely representative of the WD population. However, under the first-order assumption that it *is* representative, applying DR to any large WD spectroscopic survey with a similar selection function would likely give similar maps to the one presented here. Such maps would feature a large swathe of DAs (organised roughly by temperature), a second sequence of DBs, DCs, and DQs, and various islands of DZs, MS stars, and extragalactic sources. Spectral classification campaigns could therefore begin by applying DR to the dataset; the location of a spectrum in the embedding would then act as a useful prior when classifying it visually. Although a spectrum being on the DA sequence of the embedding may not add much value in the classification of a spectrum with obvious Balmer features, classifying a more ambiguous spectrum would benefit from an initial automated suggestion. For example, if located towards the very hot end of the DA sequence, a human classifier should look for DA features as well as DO and subdwarf features. If near the lower end of the secondary sequence, one should look for features distinguishing between a DQ and a DC, while bearing in mind that the spectrum could also be an extragalactic source or a DZ. Locating spectra on a map offers an intuitive – and crucially, automated – assessment of the likely classification of a spectrum. This could be used in conjunction with other, non-spectral information, such as the source's *Gaia* magnitudes, or the source's 'probability of being a white dwarf' ($P_{\rm WD}$; Gentile Fusillo et al. 2015), when assessing spectral class.

For some classes with little representation in the EDR, the full release may contain enough similar objects for them to form their own cluster. For example, the EDR contains a single white dwarf classified DO (J171600.53+422131.17; Manser et al. 2024), along with 10 DAOs, near the hot (left-hand) end of the DA sequence of Fig. 2. It is possible that with more DOs and DAOs in the full data release, DR would be able to separate them out from the DA sequence based on their common He II lines.

Such poorly-represented classes also highlight advantages of unsupervised techniques over, for example, supervised machine learning. The accuracy of supervised models is dependent on their training data, and strongly imbalanced training data makes recognising under-represented classes more difficult (e.g., He & Garcia 2009; Johnson & Khoshgoftaar 2019; Das et al. 2022). As an extreme example, if one class constitutes 99 per cent of the training set, then a trivial model which predicts that class for every data point would be 99 per cent accurate. Additionally, although multiclass classifiers usually allow for uncertainty in their predictions by outputting a set of numbers summing to one (interpreted as a discrete probability distribution over the possible classes), supervised classifiers often erroneously make very high-confidence predictions, especially for data not well-represented in the training set (e.g., Nguyen et al. 2015; Guo et al. 2017; Hein et al. 2019). The unsupervised technique presented in this work automatically accounts for such uncertain classifications: spectra with minuscule Balmer features would be located in the overlap between DA and DC regions of the embedding, naturally informing a human classifier of some uncertainty in the automated classification. Such a classifier would then devote more effort to accurately identifying distinguishing features in the spectrum; unambiguous spectra located solidly in the DA region would merely require a confirmatory glance at most.

4.2 Limitations

It is clear that dimensionality reduction alone is not capable of classifying the spectra of all white dwarf candidates. Though the technique is quickly able to classify a large fraction of the dataset, particularly when focused on a wavelength range around some distinctive spectral features, we outline drawbacks to DR in this subsection.

As outlined in the previous subsection, DR does not simply take a spectrum as input and produce a set of pseudo-probabilities of various classes as output, as with supervised multiclass classifiers. While such a classification could be biased and may not reflect the nuances of WD spectra, the outputs are readily interpretable and requires no further human intervention. Locating a spectrum on a map relative to others requires some judgement, which while less time-consuming than studying the overall spectrum, is less trivial than simply receiving a set of probabilities.

There are multiple regions of the embedding shown in Fig. 2 where spectra of different classes are embedded close together. These include:

- The second sequence, at the upper centre of the embedding, contains DBs, DCs, DQs, DZs, and extragalactic sources. Although focusing on helium absorption features aids in the identification of DBs (see Section 3.2.1), the breadth of characteristic spectral features of DQs (Swan bands), cool DZs (deep Ca H and K lines, overlapping with H ϵ), and quasars (Lyman α forest) has made separating these objects with DR difficult.
- Near the cooler end of the DA sequence, there is some overlap between the DAs, the main-sequence stars, DCs, and some DZs. As mentioned in Section 3, there are physical reasons for the continuous transition between DAs and DCs, but the presence of main-sequence stars and some DZs is somewhat unsatisfactory.
- The hotter end of the DA sequence also contains DAOs, subdwarfs, main-sequence stars, and a smaller number of very hot DCs, DQs, and DBs.

These overlapping regions would make the automated classification of the dataset less clear-cut than in the 'purer' regions. However, even in these regions the classification is narrowed down: although at the cooler end of the DA sequence DR methods may not be able to distinguish a DA from a DC, it would at least inform a human spectral classifier that there is no need to look for the helium lines of a DB. To distinguish between such spectral classes, supervised machine learning techniques such as those demonstrated by Vincent et al. (2023) have shown an accuracy of $\approx$ 95 per cent on SDSS DR17 spectra.

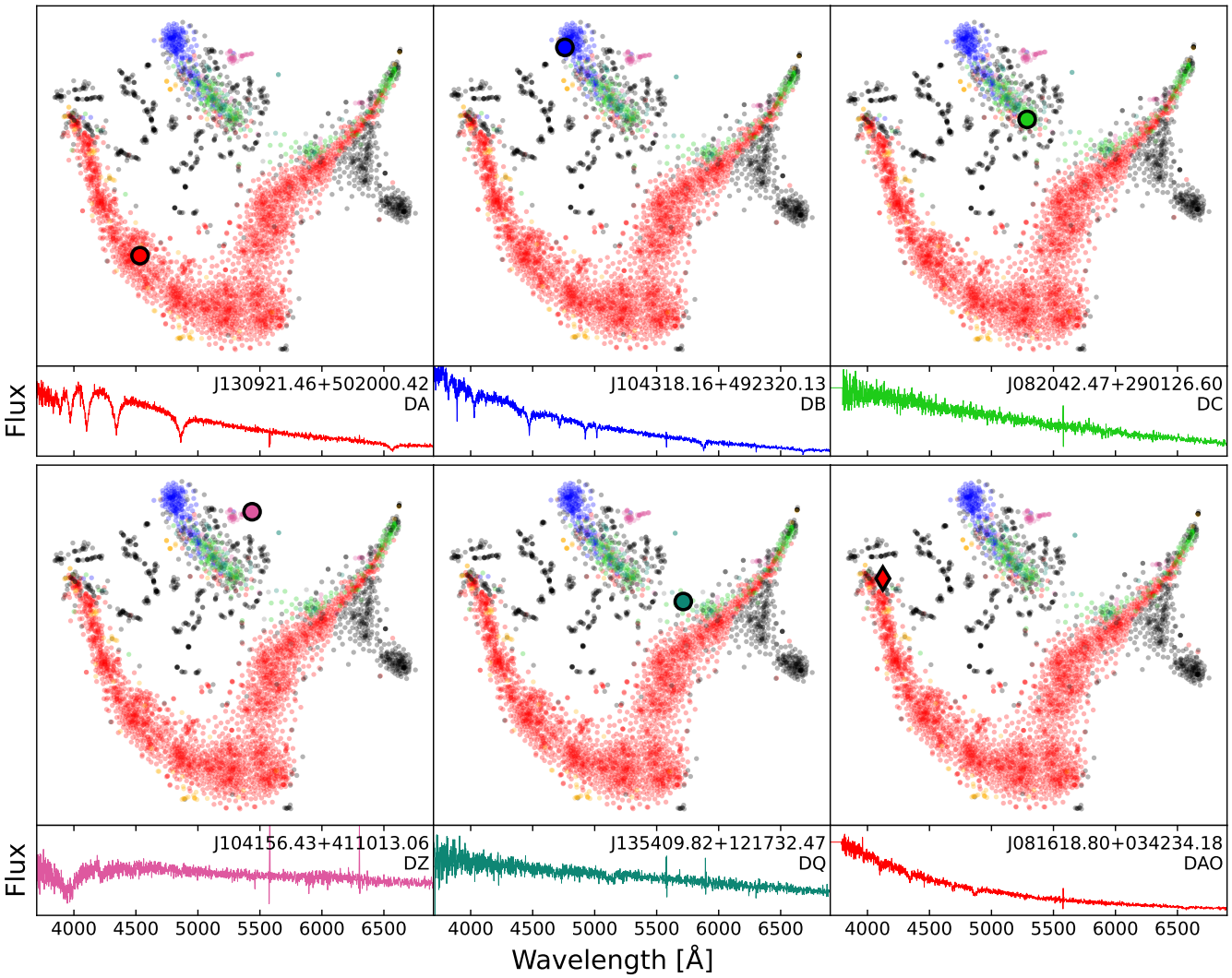

Figure 6. Projection of external spectra appended to DESI EDR WD dataset. Each embedding is almost identical to the embedding of N spectra shown in Fig. 2, but with one additional point. The projection of the external spectrum is highlighted in each case, using the same symbol as Fig. 2. The DA, DB, and DZ are projected near to other objects classified as such, so these objects could reliably have been classified using DR as well as visual inspection. The DC, DQ, and DAO are also projected near to objects with the same class, but these regions of the embedding are more ambiguous. The external spectrum appended is shown in each case, together with its SDSS name and spectral classification (according to Gentile Fusillo et al. 2019).

Secondly, although computation time has not been a significant issue in the dimensionality reduction of the DESI EDR WD catalogue ($N = 3673$), this may become a consideration when applied to full spectroscopic data releases. The dimensionality reductions conducted above take an average of 5.5 s; with a larger dataset of 70 000 WDs anticipated in the full data release (Cooper et al. 2023) and a complexity of $O(N \log N)$ (Barnes & Hut 1986; van der Maaten 2014), an application of tSNE would take 140 s. Although this by no means precludes the application of this technique (visual classification of the dataset would take far longer), the longer computation time could limit the number of experiments with different normalisations or spectral windows (see Section 3.2) that could be reasonably attempted.

5 CONCLUSIONS

We outline the use of dimensionality reduction as an aid in the classification of large-sky medium-resolution spectroscopic surveys. Providing a proof-of-concept through the application of tSNE to the DESI EDR WD catalogue, we demonstrate the method's ability to map out the dataset structure of a spectroscopic survey, identifying spectra of various classes, in a way which naturally indicates uncertainty between classes. By focusing on spectral windows containing particular features, sources with spectral features in these windows can be identified with high accuracy: CVs and DBs can be identified with respectively 100 per cent and 85 per cent accuracy. The technique identifies absorption lines that have been missed even by visual classification, and takes only a few seconds to indicate spectral classifications for the entire catalogue. Additionally, dimensionality reduction can be used to help classify individual spectra external to the original dataset. There is no reason that the techniques outlined here

could not be similarly applied to the subclassification of other subsets of spectroscopic surveys, such as main-sequence stars, quasars, or galaxies. As the coverage, depth, and resolution of spectroscopic surveys improve over the next few years, our work highlights the necessity and ease with which automated techniques can be exploited to maximise the scientific returns from these enormous datasets.

ACKNOWLEDGEMENTS

The authors thank Chris Manser for informative conversations regarding the formatting and access to DESI data. We thank Keith Hawkins and Mariona Badenas-Agusti for discussions on the normalisation of spectra. We thank also Siyi Xu for advice on the interpretation of the embeddings.

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. DEâĂŞAC02âĂŞ05CH11231, 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 NSFâĂŹs 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 Science and Technology of Mexico (CONACYT); 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 Iolkam DuâĂŹag (Kitt Peak), a mountain with particular significance to the Tohono OâĂŹodham Nation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors 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.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High Performance Computing at the University of Utah. The SDSS website is www.sdss4.org. SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics | Harvard & Smithsonian, the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

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.

This research has made use of the SIMBAD database (Wenger et al. 2000) and the VizieR catalogue access tool (Ochsenbein et al. 2000), CDS, Strasbourg Astronomical Observatory, France. In addition to Python packages referenced in the text, we also acknowledge the use of NumPy (Harris et al. 2020), Matplotlib (Hunter 2007), PyVO (Graham et al. 2014), pandas (pandas development team 2020; Wes McKinney 2010), SciPy (Virtanen et al. 2020), Astropy (Astropy Collaboration et al. 2013, 2018, 2022), and bokeh (Bokeh Development Team 2018).

DATA AVAILABILITY

The data used in this work were obtained from the public archives of DESI, *Gaia*, and SDSS. Data products and Python scripts will be made available upon acceptance of the manuscript at https://github.com/xbyrne/desi_edr_wd_tsne.

REFERENCES

Almeida A., et al., 2023, ApJS, 267, 44
Althaus L. G., Córsico A. H., Isern J., García-Berro E., 2010, A&ARv, 18, 471
Althaus L. G., et al., 2021, A&A, 646, A30
Astropy Collaboration et al., 2013, A&A, 558, A33
Astropy Collaboration et al., 2018, AJ, 156, 123
Astropy Collaboration et al., 2022, ApJ, 935, 167
Barnes J., Hut P., 1986, nature, 324, 446
Blouin S., Bédard A., Tremblay P.-E., 2023, MNRAS, 523, 3363
Bokeh Development Team 2018, Bokeh: Python library for interactive visualization. https://bokeh.pydata.org/en/latest/
Bonsor A., Mustill A. J., Wyatt M. C., 2011, MNRAS, 414, 930
Boroson T. A., Green R. F., 1992, ApJS, 80, 109
Coifman R. R., Lafon S., 2006, Applied and computational harmonic analysis, 21, 5
Cooper A. P., et al., 2023, ApJ, 947, 37
DESI Collaboration et al., 2016a, arXiv e-prints, p. arXiv:1611.00036
DESI Collaboration et al., 2016b, arXiv e-prints, p. arXiv:1611.00037
DESI Collaboration et al., 2022, AJ, 164, 207
DESI Collaboration et al., 2023, arXiv e-prints, p. arXiv:2306.06308
Dalton G., et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. p. 84460P, doi:10.1117/12.925950
Das S., Mullick S. S., Zelinka I., 2022, IEEE Transactions on Artificial Intelligence, 3, 973
Fontaine G., Villeneuve B., Wesemael F., Wegner G., 1984, ApJ, 277, L61
Frenay B., Verleysen M., 2014, IEEE Transactions on Neural Networks and Learning Systems, 25, 845
Frewen S. F. N., Hansen B. M. S., 2014, MNRAS, 439, 2442
García-Zamora E. M., Torres S., Rebassa-Mansergas A., 2023, A&A, 679, A127
Gebhardt K., et al., 2021, ApJ, 923, 217

Gentile Fusillo N. P., Gänsicke B. T., Greiss S., 2015, MNRAS, 448, 2260
Gentile Fusillo N. P., et al., 2019, MNRAS, 482, 4570
Graham M., Plante R., Tody D., Fitzpatrick M., 2014, PyVO: Python access to the Virtual Observatory, Astrophysics Source Code Library, record ascl:1402.004 (ascl:1402.004)
Guo C., Pleiss G., Sun Y., Weinberger K. Q., 2017, in Precup D., Teh Y. W., eds, Proceedings of Machine Learning Research Vol. 70, Proceedings of the 34th International Conference on Machine Learning. PMLR, pp 1321–1330, https://proceedings.mlr.press/v70/guo17a.html
Harris C. R., et al., 2020, Nature, 585, 357
Hawkins K., et al., 2021, ApJ, 911, 108
He H., Garcia E. A., 2009, IEEE Transactions on knowledge and data engineering, 21, 1263
Hein M., Andriushchenko M., Bitterwolf J., 2019, in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 41–50
Huang S.-S., 1972, ApJ, 171, 549
Hunter J. D., 2007, Computing in Science & Engineering, 9, 90
Ivezić Ž., et al., 2019, ApJ, 873, 111
Johnson J. M., Khoshgoftaar T. M., 2019, Journal of Big Data, 6, 1
Kao M. L., Hawkins K., Rogers L. K., Bonsor A., Dunlap B. H., Sanders J. L., Montgomery M. H., Winget D. E., 2024, arXiv e-prints, p. arXiv:2405.17667
Kleinman S. J., et al., 2013, ApJS, 204, 5
Koester D., 2009, A&A, 498, 517
Koester D., Weidemann V., Zeidler E. M., 1982, A&A, 116, 147
Koester D., Kepler S. O., Irwin A. W., 2020, A&A, 635, A103
Kollmeier J. A., et al., 2017, arXiv e-prints, p. arXiv:1711.03234
Kullback S., Leibler R. A., 1951, The annals of mathematical statistics, 22, 79
Lafon S., Lee A., 2006, IEEE Transactions on Pattern Analysis and Machine Intelligence, 28, 1393
Lauffer G. R., Romero A. D., Kepler S. O., 2018, MNRAS, 480, 1547
Maldonado R. F., Villaver E., Mustill A. J., Chavez M., Bertone E., 2020, MNRAS, 499, 1854
Manser C. J., et al., 2024, arXiv e-prints, p. arXiv:2402.18641
McInnes L., Healy J., Melville J., 2018, arXiv preprint arXiv:1802.03426
Mustill A. J., Villaver E., Veras D., Gänsicke B. T., Bonsor A., 2018, MNRAS, 476, 3939
Nguyen A., Yosinski J., Clune J., 2015, in Proceedings of the IEEE conference on computer vision and pattern recognition. pp 427–436
Ochsenbein F., Bauer P., Marcout J., 2000, A&AS, 143, 23
Paquette C., Pelletier C., Fontaine G., Michaud G., 1986, ApJS, 61, 177
Pearson K., 1901, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2, 559
Pedregosa F., et al., 2011, Journal of Machine Learning Research, 12, 2825
Richards J. W., Freeman P. E., Lee A. B., Schafer C. M., 2009, ApJ, 691, 32
Roweis S. T., Saul L. K., 2000, Science, 290, 2323
Schatzman E., 1945, Annales d'Astrophysique, 8, 143
Smak J., 1969, Acta Astron., 19, 155
Tan L., Liu Z., Wang F., Mei Y., Deng H., Liu C., 2023, ApJS, 268, 28
Tang J., Liu J., Zhang M., Mei Q., 2016, in Proceedings of the 25th international conference on world wide web. pp 287–297
Tremblay P. E., Kalirai J. S., Soderblom D. R., Cignoni M., Cummings J., 2014, ApJ, 791, 92
Vincent O., Bergeron P., Dufour P., 2023, MNRAS, 521, 760
Vincent O., Barstow M. A., Jordan S., Mander C., Bergeron P., Dufour P., 2024, A&A, 682, A5
Virtanen P., et al., 2020, Nature Methods, 17, 261
Wenger M., et al., 2000, A&AS, 143, 9
Wes McKinney 2010, in Stéfan van der Walt Jarrod Millman eds, Proceedings of the 9th Python in Science Conference. pp 56 – 61, doi:10.25080/Majora-92bf1922-00a
Winget D. E., Hansen C. J., Liebert J., van Horn H. M., Fontaine G., Nather R. E., Kepler S. O., Lamb D. Q., 1987, ApJ, 315, L77
Wyatt M. C., Farihi J., Pringle J. E., Bonsor A., 2014, MNRAS, 439, 3371
Yang Y., Zhao J., Zhang J., Ye X., Zhao G., 2020, AJ, 160, 236
de Jong R. S., et al., 2016, in Evans C. J., Simard L., Takami H., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI. p. 99081O, doi:10.1117/12.2232832
pandas development team T., 2020, pandas-dev/pandas: Pandas, doi:10.5281/zenodo.3509134, https://doi.org/10.5281/zenodo.3509134
van der Maaten L., 2014, The journal of machine learning research, 15, 3221
van der Maaten L., Hinton G., 2008, Journal of machine learning research, 9

APPENDIX A: MATHEMATICAL DETAILS OF *T*SNE

The goal of *t*SNE (and dimensionality reduction in general) is to map a set of N high-dimensional vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ into two-dimensional vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$, in such a way that the 'similarity' between each pair of vectors is approximately preserved under the map. Different DR methods use different definitions of similarity (McInnes et al. 2018); for *t*SNE, similarity between high-dimensional vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ is defined by:

$$p_{ij} = \frac{1}{2N}\left(p_{i|j} + p_{j|i}\right), \tag{A1}$$

where $p_{i|j}$ is defined by a normal distribution:

$$p_{i|j} = \frac{\exp\left(-\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2/2\sigma^2\right)}{\sum_{k\neq i}^{N}\exp\left(-\left\|\mathbf{x}_k - \mathbf{x}_j\right\|^2/2\sigma^2\right)}, \tag{A2}$$

where $\|\cdot\|^2$ is the L2 norm and σ is a hyperparameter known as the perplexity. In this work, the perplexity is set to $\sigma = 30$, the default value in the SCIKIT-LEARN implementation used (Pedregosa et al. 2011).

Similarity between low-dimensional points $\mathbf{y}_i$ and $\mathbf{y}_j$ is defined by a student's *t*-distribution with one degree of freedom:

$$q_{ij} = \frac{\left(1 + \left\|\mathbf{y}_i - \mathbf{y}_j\right\|^2\right)^{-1}}{\sum_{k\neq i}^{N}\left(1 + \left\|\mathbf{y}_k - \mathbf{y}_j\right\|^2\right)^{-1}}. \tag{A3}$$

To faithfully maintain the structure of the dataset as far as possible when reducing the dimensionality, the difference between the pairwise similarity distributions is minimised. The difference between the distributions is quantified by the Kullback-Leibler divergence (Kullback & Leibler 1951):

$$\mathcal{KL}(p||q) \equiv \sum_{i,j} p_{ij}\log\left(\frac{p_{ij}}{q_{ij}}\right), \tag{A4}$$

and is minimised by optimising the positions $\mathbf{y}_i$ of the low-dimensional points, for example by gradient descent. The result is a set of two-dimensional vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_N$ which are separated from each other by similar distances as $\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_N$ are from each other.

The SCIKIT-LEARN implementation of *t*SNE used here (Pedregosa et al. 2011) makes use of the Barnes-Hut algorithm (Barnes & Hut 1986), speeding up the calculation of the embedding from $O(N^2)$ to $O(N \log N)$ at the expense of a very small reduction in accuracy (van der Maaten 2014).

This paper has been typeset from a TEX/LATEX file prepared by the author.