https://doi.org/10.15388/vu.thesis.636
https://orcid.org/0000-0003-0804-6938

VILNIUS UNIVERSITY
CENTER FOR PHYSICAL SCIENCES AND TECHNOLOGY

Markus Ambrosch

Investigating Stellar Atmospheres with Convolutional Neural Networks

DOCTORAL DISSERTATION

Natural Sciences,
Physics (N 002)

VILNIUS 2024

This dissertation was written between 2019 and 2023 at Vilnius University, Institute of Theoretical Physics and Astronomy.

Academic supervisor – Dr. Šarūnas Mikolaitis (Vilnius University, Natural Sciences, Physics, N 002)

This doctoral dissertation will be defended in a public meeting of the Dissertation Defence Panel

Chair – Dr. Donatas Narbutis (Vilnius University, Natural Sciences, Physics, N 002)

Members:

Dr. Friedrich Anders (University of Barcelona, Natural Sciences, Physics, N 002)

Dr. Vidas Dobrovolskas (Vilnius University, Natural Sciences, Physics, N 002)

Prof. Dr. Paulius Miškinis (Vilnius Gediminas Technical University, Natural Sciences, Physics, N 002)

Dr. Vidas Regelskis (Vilnius University, Natural Sciences, Physics, N 002)

at 14:00 on September 18, 2024, in the National Center for Physical Sciences and Technology, room A101. Address: Saulėtekio al. 3, LT–10257, Vilnius, Lithuania; tel. +370 5 2234637; email tfai@tfai.vu.lt

The text of this dissertation can be accessed at the libraries of Vilnius University and Center for Physical Sciences and Technology, as well as on the website of Vilnius University: www.vu.lt

https://doi.org/10.15388/vu.thesis.636
https://orcid.org/0000-0003-0804-6938

VILNIAUS UNIVERSITETAS
FIZINIŲ IR TECHNOLOGIJOS MOKSLŲ CENTRAS

Markus Ambrosch

Žvaigždžių atmosferų tyrimas su konvoliuciniais neuroniniais tinklais

DAKTARO DISERTACIJA

gamtos mokslai,
fizika (N 002)

VILNIUS 2024

Disertacija rengta 2019 – 2023 metais Vilniaus universiteto Teorinės fizikos ir astronomijos institute.

Mokslinis vadovas – dr. Šarūnas Mikolaitis (Vilniaus universitetas, gamtos mokslai, fizika, N 002)

Gynimo taryba:

Pirmininkas – **dr. Donatas Narbutis** (Vilniaus universitetas, gamtos mokslai, fizika, N 002)

Nariai:

Dr. Friedrich Anders (Barselonos universitetas, gamtos mokslai, fizika, N 002)

Dr. Vidas Dobrovolskas (Vilniaus universitetas, gamtos mokslai, fizika, N 002)

Prof. dr. Paulius Miškinis (Vilniaus Gedimino technikos universitetas, gamtos mokslai, fizika, N 002)

Dr. Vidas Regelskis (Vilniaus universitetas, gamtos mokslai, fizika, N 002)

Disertacija ginama viešame Gynimo tarybos posėdyje 2024 m. rugsėjo mėn. 18 d. 14 val. Nacionalinio fizinių ir technologijos mokslų centro posėdžių salėje A101. Adresas: Saulėtekio al. 3, LT–10257, Vilnius, Lietuva; tel. tel. +370 5 2234637; el. paštas tfai@tfai.vu.lt

Disertaciją galima peržiūrėti Vilniaus universiteto ir Fizinių ir technologijos mokslų centro bibliotekose ir Vilniaus universiteto interneto svetainėje adresu: www.vu.lt

Contents

Acronyms and abbreviations

Term	Abbreviation
Apache Point Observatory Galactic Evolution Experiment	APOGEE
Convolutional neural network	CNN
(internal) data release	(i)DR
European Southern Observatory	ESO
Fiber Large Array Multi-Element Spectrograph	FLAMES
Gaia-ESO survey	GES
Gaia Radial Velocity Spectrometer	*Gaia*-RVS
Model Atmospheres with Radiative and Convective Scheme	MARCS
(non-)local thermodynamic equilibrium	(N)LTE
Rectified Linear Unit	ReLU
Radial velocity	RV
RAdial Velocity Experiment	RAVE
Signal-to-noise ratio	S/N
t-distributed stochastic neighbor embedding	t-SNE
Ultraviolet and Visual Echelle Spectrograph	UVES
Very Large Telescope	VLT
William Herschel Telescope Enhanced Area Velocity Explorer	WEAVE
4-metre Multi-Object Spectrograph Telescope	4MOST

Aims and novelty

The main aim of this work is to build and train convolutional neural networks that can efficiently parameterize large numbers of stellar spectra. The networks predict multiple stellar atmospheric parameters and chemical abundances simultaneously, accurately, and precisely. This thesis aims to show that novel machine learning methods offer reliable tools that will aid in harnessing the full potential of upcoming large-scale spectroscopic surveys. It will be shown that the quality of the network predictions depends strongly on the quality of the training spectra and the associated labels. Training labels are determined with classical spectroscopic methods. To be able to judge the performance of the network, it is important to know how the training set was built and what its limitations are. The methods developed for this thesis improve on previous similar studies in terms of network architecture complexity, training strategies, and the assessment of the accuracy and precision of the network predictions. It is shown that convolutional neural networks are able to identify relevant spectral lines. The network predicts labels from these lines in a way that is intuitive and consistent with how classical methods analyze stellar spectra. Relying on existing spectral features provides important explainability of the network predictions and is key for finding chemically peculiar stars. The publications related to this thesis are part of the preparations for the upcoming WEAVE and 4MOST surveys. Each of these surveys will provide spectra of several million stars located in all parts of the Milky Way galaxy. The usefulness of the here developed methods has been confirmed in the recent literature, where they have been extended to investigate more than 800 000 *Gaia*-RVS spectra.

Statements of this thesis

1. Stars in the globular cluster NGC 1851 can be separated into two subsamples, with a difference of 0.07 dex in the mean metallicity. The subsamples also differ in their mean carbon-to-nitrogen ratios and abundances of the s-processed neutron-capture elements Y, Zr, Ba, La, Ce, and Nd. The average difference of these *s*-processed elements in the two subpopulations is 0.35 dex. We found no notable distinction in the average abundance-to-iron ratios of carbon, europium, α- and iron-peak elements.

2. We can determine if a newly observed spectrum is similar in shape to a set of training set spectra with the unsupervised machine-learning algorithm t-SNE. The results for a training-like test set, with spectra of similar shape and S/N ranges as the training set, are in good agreement with the GES input values. We find that the quality of the CNN results degrades for spectra with S/N < 30, especially for abundance predictions. Observed spectra that are different from the training set spectra are not parameterized accurately.

3. After training, convolutional neural networks can simultaneously predict stellar atmospheric parameters $T_{\rm eff}$ and $\log(g)$, and the chemical abundances [Mg/Fe], [Al/Fe], and [Fe/H] from $\sim$ 35 000 stellar spectra in a few minutes. The training precision is 37 K for $T_{\rm eff}$, 0.06 dex for $\log(g)$, and < 0.08 dex for all predicted chemical abundances.

   We investigated the Mg-Al anti-correlation in globular clusters, ranging from -0.92 to -1.40 dex in metallicity. In the most metal-poor clusters, where our training set contains only a few stars, our CNN mainly recovers the Al spread in the clusters. The match between GES Mg-Al anti-correlation and CNN anti-correlation is improving for clusters with higher [Fe/H], where the training data is denser.

   Based on the CNN label predictions for [Mg/Fe] and [Fe/H], we separated our sample stars into a thin- and a thick-disk population. We find that the thick disk stars are on average 1 Gyr older than the thin disk stars. The orbit eccentricities of the thick disk stars show a negative trend with [Fe/H] ($\Delta e/\Delta$[Fe/H] = -0.26). These CNN-based results are consistent with similar studies in the literature.

4. We found 31 previously unidentified lithium-rich giants in the Gaia-ESO iDR6 data set by training a convolutional neural network to predict Li abundances. The predicted A(Li) of all of these giants exceeds 2 dex, their log(g) is lower than 3.5 dex. The precision of the network label predictions for these 31 stars is high, with e$T_{\rm eff}$ < 50 K and elog(g), e[M/H], and eA(Li) < 0.1 dex. None of these 31 Li-rich giants have previously been reported in any Gaia-ESO paper, and they do not appear in the GALAH survey catalog of Li-rich giants in the southern hemisphere. Lithium-rich giants are rare. Machine learning methods will play a crucial role in finding these objects in the observations of the upcoming large-scale spectroscopic surveys.

5. Network gradients can be used to demonstrate the sensitivity of our network to various parts of the input spectra. These gradients show that the network is able to identify absorption features that are relevant for the predicted stellar parameters and abundances. Our CNN associates the strengths of the found features to the values of these stellar labels. Caution must be applied when choosing input labels, because strongly correlated input labels lead to strongly correlated network gradients. The network then predicts labels based on unrelated spectral features (for example, absolute Al abundance from Mg absorption lines). Inferring stellar parameters from correlations like these can lead to satisfying results for some spectra. However, stars with exotic chemical compositions will not be parameterized well.

6. Our training, validation, and test sets have in common that the differences between CNN predictions and GES values increase at the edges of the parameter space. At the edges, the number of available training spectra is small. Increasing the number of training spectra in these parameter regimes would allow increasing the accuracy (mean bias) of the CNN predictions, as well as precision as the number of sample observations increases.

   The scatter between predictions from the multiple different CNN models can be used to assess the internal precision of our network. This scatter is small: On average 27 K for $T_{\rm eff}$, 0.04 for log(g), and 0.03 dex for [Mg/Fe], [Al/Fe], and [Fe/H] alike. We find that the uncertainties increase at the edges of the parameter space. The uncertainties also increase as the spectra S/N decreases. Therefore, the spectra S/N and the

position of the predicted labels in the parameter space should also be considered when estimating the label precision for individual spectra.

Publications related to this thesis

- **Gaia-ESO Survey: Detailed elemental abundances in red giants of the peculiar globular cluster NGC 1851**
  Tautvaišienė, G.; Drazdauskas, A.; Bragaglia, A.; Martell, S. L.; Pancino, E.; Lardo, C.; Mikolaitis, Š.; Minkevičiūtė, R.; Stonkutė, E.; **Ambrosch, M.** ; Bagdonas, V.; Chorniy, Y.; Sanna, N.; Franciosini, E.; Smiljanic, R.; Randich, S.; Gilmore, G.; Bensby, T.; Bergemann, M.; Gonneau, A.; Guiglion, G.; Carraro, G.; Heiter, U.; Korn, A.; Magrini, L.; Morbidelli, L.; Zaggia, S. (2022).
  Astronomy & Astrophysics, Volume 658, A80
  https://doi.org/10.1051/0004-6361/202142234

- **The Gaia-ESO Survey: Preparing the ground for 4MOST and WEAVE galactic surveys. Chemical evolution of lithium with machine learning**
  Nepal, S.; Guiglion, G.; de Jong, R. S.; Valentini, M.; Chiappini, C.; Steinmetz, M.; **Ambrosch, M.**; Pancino, E.; Jeffries, R. D.; Bensby, T.; Romano, D.; Smiljanic, R.; Dantas, M. L. L.; Gilmore, G.; Randich, S.; Bayo, A.; Bergemann, M.; Franciosini, E.; Jiménez-Esteban, F.; Jofré, P.; Morbidelli, L.; Sacco, G. G.; Tautvaišienė, G.; Zaggia, S. (2023)
  Astronomy & Astrophysics, Volume 671, A61
  https://doi.org/10.1051/0004-6361/202244765

- **The Gaia-ESO Survey: Chemical evolution of Mg and Al in the Milky Way with machine learning**
  Ambrosch, M.; Guiglion, G. ; Mikolaitis, Š. ; Chiappini, C.; Tautvaišienė, G.; Nepal, S.; Gilmore, G.; Randich, S.; Bensby, T.; Bayo, A.; Bergemann, M. ; Morbidelli, L. ; Pancino, E.; Sacco, G. G. ; Smiljanic, R.; Zaggia, S.; Jofré, P. ; Jiménez-Esteban, F. M. (2023)
  Astronomy & Astrophysics, Volume 672, A46
  https://doi.org/10.1051/0004-6361/202244766

Additional publications

- **Chemical Composition of Bright Stars in the Northern Hemisphere: Star-Planet Connection**
  Tautvaišienė, G.; Mikolaitis, Š.; Drazdauskas, A.; Stonkutė, E.; Minkevičiūtė, R.; Pakštienė, E.; Kjeldsen, H.; Brogaard, K.; Chorniy, Y.; von Essen, C.; Grundahl, F.; **Ambrosch, M.**; Bagdonas, V.; Sharma, A.; Viscasillas Vázquez, C. (2022)
  The Astrophysical Journal Supplement Series, Volume 259, Number 2
  https://doi.org/10.3847/1538-4365/ac50b5

- **Ground-based contribution of average-sized telescopes to characterisation of PLATO targets**
  Mikolaitis, Š.; Tautvaišienė, G.; Drazdauskas, A.; Minkevičiūtė, R.; SStonkutė, E.; Pakštienė, Viscasillas Vázquez, C.; Janulis, R.; E.; **Ambrosch, M.**; Bagdonas,V.
  PLATO Stellar Science Conference 2023 (PSSC23), Milazzo, Italy.
  https://doi.org/10.5281/zenodo.8103989

Scientific talks

- **Investigating Gaia-ESO spectra with convolutional neural networks**
  Local seminar, September 22, 2021, Institute of Theoretical Physics and Astronomy, Vilnius, Lithuania

- **Parameterization of Gaia-ESO stellar spectra with convolutional neural networks**
  European Astronomical Society Annual Meeting, Contributed talk, June 27, 2022, Valencia, Spain

Posters

- **Spacial distribution of carbon, nitrogen, and oxygen in the Milky Way galaxy**
  e-Poster, European Astronomical Society Annual Meeting, July 10 to July 14, Kraków, Poland

- **Carbon and nitrogen as tracers of extra mixing processes in evolved stars**
  45th Lithuanian National Physics Conference, October 25-27, Vilnius, Lithuania

Contributions by the author

In the paper *The Gaia-ESO Survey: Chemical evolution of Mg and Al in the Milky Way with machine learning*, the quality control of the training set, the classification of the test set, the creation and testing of the adopted network architecture, the interpretation of the network gradients, and the analysis of the network predictions were performed by the author. The training spectra and their labels were provided by the Gaia-ESO survey, their preparation for the usage with the developed convolutional neural network was also done by the author. All *python* code for the neural network implementation, the analysis of the results, and the creation of the displayed figures was written by the author.

The data analysis, interpretation of results, and the creation of figures of the other two papers was the result of teamwork between the co-authors of the respective works.

1 Introduction

The use of machine–learning for the exploration of big data sets in astronomy was predicted over three decades ago (Rosenthal 1988). Yet, the high computational costs of this method have long delayed its advance. Some of the first applications of neural networks, a sub-field of machine–learning, include the automatic detection of sources in astronomical images (SExtractor, Bertin & Arnouts 1996), the morphological classification of galaxies (Lahav et al. 1996) and the classification of stellar spectra (Bailer-Jones 1997). In recent years, the increasing power of modern computer systems and the possibilities of cloud computing have led to a growing popularity of machine–learning methods. Powerful open-source libraries such as TensorFlow (Abadi et al. 2015) and PyTorch (Paszke et al. 2019) for Python programming offer easy-to-use frameworks for building and training several types of neural networks.

1.1 Analysis of stellar spectra with machine-learning methods

Convolutional Neural-Networks (CNNs) have recently been used to simultaneously infer multiple stellar labels (i.e., atmospheric parameters and chemical abundances) from stellar spectra. Every CNN contains convolutional layers which enable the network to find extended features in the input data. In stellar spectra, these features are absorption lines and continuum points; In 2-D images, such features could be eyes in a face or star clusters in a spiral galaxy (Bialopetravičius & Narbutis 2020). Neural network methods are purely data-driven and therefore require no input of any physical laws or models. Instead, during a training phase, the network learns to associate the strength of spectral features with the values of the stellar labels. This requires a training set of spectra with pre-determined labels, from which the network can learn. Training sets for spectral analysis typically contain several thousand stellar spectra with high quality labels. Current spectral surveys, which provide $\sim 10^5$ spectra with labels, are an ideal testing ground for the CNN approach to spectral parameterization.

Spectroscopic surveys provide insights into the evolution of individual stars, of large-scale structures such as globular clusters and of the Milky Way galaxy as a whole. Upcoming projects, for example the

William Herschel Telescope Enhanced Area Velocity Explorer (WEAVE, Dalton et al. 2018) and the 4-metre Multi-Object Spectroscopic Telescope (4MOST, de Jong et al. 2019) will observe millions of stars. Efficient automatic tools will be needed to analyze the large number of spectra that such surveys will deliver.

To find the atmospheric parameters and chemical composition of stars, classical spectroscopic methods either measure equivalent widths of absorption lines or compare observed spectra to synthetic spectra. These synthetic spectra can be generated on-the-fly or are part of a pre-computed spectral grid. Jofré et al. (2019) provide an overview of classical spectral analysis methods in the context of large spectroscopic surveys.

The main benefit of using machine-learning for spectra parameterization is computation speed. While classical methods typically take several minutes to determine parameters and abundances from a single spectrum, a trained CNN can parameterize several 10^4 spectra in the same amount of time. This speed is crucial to fully utilize the capabilities of the upcoming spectra surveys. For instance, 4MOST will observe $\sim$25 000 stars per night, with the goal of measuring up to 15 abundances per star. Machine-learning will be a way to manage such large amounts of data every day.

Examples of stellar parameterization using CNNs can be found in several recent studies. Fabbro et al. (2018) have developed StarNet, a CNN that can infer the stellar atmospheric parameters directly from observed spectra in the APO Galactic Evolution Experiment (APOGEE, Majewski et al. 2017). A grid of synthetic spectra was used to train and test StarNet. Purely observational data from APOGEE DR14 were used by Leung & Bovy (2019) to train their astroNN convolutional network. To mimic the methods of standard spectroscopic analysis, astroNN is designed to use the whole spectrum when predicting atmospheric parameters but is limited to individual spectral features for the prediction of chemical abundances. Guiglion et al. (2020) trained their CNN on medium-resolution stellar spectra from the RAdial Velocity Experiment (RAVE, Steinmetz et al. 2020) together with stellar labels that were derived from high-resolution APOGEE DR16 spectra. They also added absolute magnitudes and extinction corrections for their sample stars as

inputs for the network. This information allowed their CNN to put additional constraints on its predictions of the effective temperature and surface gravity.

This thesis tests a CNN approach in the context of the Gaia-ESO survey (GES, Gilmore et al. 2012; Randich et al. 2013). We use GIRAFFE spectra with labels from the sixth internal data release. The GES survey is designed to complement the astrometric data from the Gaia space observatory (Gaia Collaboration et al. 2016). The goal of the present project is to prepare the groundwork for machine–learning for the next generation of spectroscopic surveys, such as 4MOST and WEAVE.

1.2 The Gaia-ESO survey

The Gaia-ESO survey is a public spectroscopic survey. Its goal is to observe more than 10^5 stars to obtain a homogeneous overview of the chemical properties of various stellar populations in the Milky Way galaxy. When combined with the stellar kinematic properties from the *Gaia* space mission (Gaia Collaboration et al. 2016), the stellar chemical compositions provide insights into the formation and assembly histories of the Milky Way bulge, the thin and thick disks, the stellar halo, as well as globular and open clusters. Figure 1 shows the areas of the sky which have been observed by the Gaia-ESO survey.

Observations for the survey have been carried out with the FLAMES instrument of the ES0 Very Large Telescope (VLT) in Chile, between December 2011 and January 2018. FLAMES feeds two different spectrographs, GIRAFFE and UVES (Pasquini et al. 2002). Both of these are multi-object spectrographs. GIRAFFE can observe up to 130 objects at a time, with intermediate and high resolutions ($R \sim 5000$ to $R \sim 30000$), while UVES provides higher resolution ($R \sim 47000$) but can only observe 8 targets at once. The wavelength ranges of both spectrographs are divided into several predefined setups that together cover the whole visual range.

In the Gaia-ESO survey, multiple workgroups analyze the spectra from different spectrograph setups. The workgroups themselves are composed of various nodes, located in different countries across Europe (Gilmore et al. 2022). Each node used slightly different methods

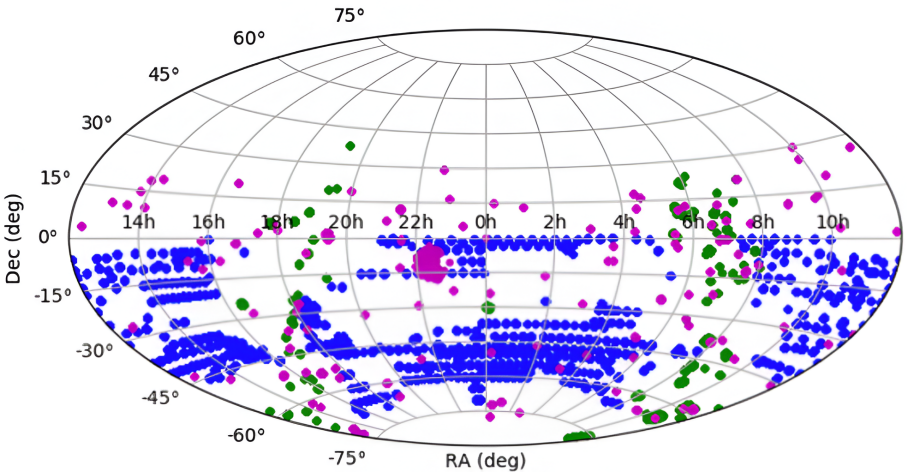


Figure 1: Fields in the sky as observed by Gaia-ESO survey. Different types of fields have different colors: Milky Way targets in blue, clusters in green, and calibration fields in magenta.

to determine the stellar parameters and abundances from the same spectra. The results from the individual nodes were then homogenized by using benchmark stars. In this work, we employ CNNs to investigate spectra from the GIRAFFE setups HR10, HR21, and HR15N. The *Vilnius* node analyzed spectra from the GIRAFFE HR10 and HR21 setups and provided the main stellar atmospheric parameters and abundances measurements for magnesium, aluminum, and iron, among other elements. Spectra from the GIRAFFE HR15N setup were investigated by the *Arcetri* node. This node measured the lithium abundance for the final data release of the survey (Franciosini et al. 2022). In the following section, we will summarize the methods used by the *Vilnius* and *Arcetri* nodes.

1.2.1 Classical methods for the determination of stellar parameters and chemical abundances

The term *classical* refers to methods that use knowledge about the physical processes within stellar atmospheres. This knowledge allows us to calculate how the radiation from the deeper layers of a star is transported through the stellar atmosphere, and how the atmospheric param-

eters and chemical elements shape the stellar spectra which we observe on earth. Given this knowledge, we can solve the inverse problem, and analyze the observed spectra to find the physical conditions in the stars. An in-depth description of the classical methods for the observation and analysis of stellar atmospheres is given in the book by Gray (2005).

The main atmospheric parameters of a star are the effective temperature $T_{\rm eff}$, surface gravity $\log(g)$, the metallicity [M/H][1], and microturbulence velocity $v_{\rm t}$. The *Vilnius* node determined these parameters by finding the model atmospheres, that satisfy the excitation and ionization balances in the GIRAFFE spectra. First, the equivalent widths of multiple Fe I and Fe II absorption lines were measured in every spectrum, across the whole spectral range. This was done automatically with the DAOSPEC tool (Stetson & Pancino 2008). The Fe lines were selected based on a line list that includes wavelengths, excitation potentials, and oscillator strengths of the absorption lines. Details about the selected lines and the used atomic data can be found in (Heiter et al. 2021a). The measured equivalent widths were then used as inputs for the spectral analysis code MOOG (Sneden 1973). The MOOG code calculates the iron abundance from a given iron line, based on the line's equivalent width and a given model atmosphere. Spectra from a grid of 1D MARCS model atmospheres (Gustafsson et al. 2008), assumed to be in local thermal equilibrium (LTE), were used as inputs for MOOG. These models are characterized by the parameters $T_{\rm eff}$, $\log(g)$, and $v_{\rm t}$. The model atmospheres were exchanged (by changing the values of $T_{\rm eff}$, $\log(g)$, and $v_{\rm t}$) until all iron abundances from the different lines were independent of the line excitation potential and equivalent width, and until Fe I and Fe II lines yielded the same average iron abundance (ionization balance). A Nelder-Mead simplex search (Nelder & Mead 1965) was used for the simultaneous optimization of $T_{\rm eff}$, $\log(g)$, and $v_{\rm t}$. Once the optimization is finished, the metallicity [M/H] is calculated by averaging the Fe abundances measured from all used Fe I and Fe II lines.

Elemental abundances of magnesium, aluminum, and iron ([Fe/H]) were found with a spectral synthesis method. Synthetic spectra were

[1]For clarity, we use *[M/H]* as the symbol for the metallicity as an atmospheric parameter. The iron abundance, determined by measuring selected iron absorption lines, is denoted as [Fe/H].

computed with the radiation transfer code TURBOSPECTRUM (Plez 2012), based on the same atomic line lists that are used for measuring the main atmospheric parameters. Turbospectrum takes MARCS model atmospheres with the previously measured atmospheric parameters as inputs. In addition to a model atmosphere, elemental abundances are passed as inputs to Turbospectrum. To measure the elemental abundance from an absorption line, multiple synthetic spectra with varying input abundances are generated. These synthetic spectra are then compared to the observed spectrum. The optimal input abundance (the result of the measurement) is found when the difference between the resulting synthetic spectrum and the observed spectrum is minimal. Examples of synthetic spectra fits to Mg and Al absorption lines in a GIRAFFE solar spectrum are shown in Fig. 2. If the input abundance is too low, the absorption line in the synthetic spectrum is too weak — its flux at the center of the line is much higher than the flux of the observed spectrum. When the input abundance is too high, then the flux at the line center is lower than the observed flux. As will be shown in Sect. 4.3, this physical connection between line depth and elemental abundance is learned by the CNN during the training phase.

The *Arcetri* node used a curve of growth method to determine the Li abundances from the GIRAFFE HR15N spectra. In this approach, the Li abundance is directly determined from the equivalent width of the Li I double feature at $\sim$ 6708 Å via an appropriate curve of growth. These curves show the relation between the abundance of an absorbing element and the equivalent width a specific absorption line of the element. Examples of curves of growths for the Li I double feature in HR15N are shown in Fig. 3. The shape of the curves depends on the stellar atmospheric parameters. To choose the appropriate curve of growth, the parameters of the given star must be known. For their Li measurements, the *Arcetri* node used atmospheric parameters from an earlier Gaia-ESO data release. These curves themselves were also computed by the *Arcetri* node. In total, they provided a grid of $\sim$120000 curves for various spectral types, ranging from $T_{\rm eff}$ = 3000 - 8000 K, log$g$ = 0.5 - 3.5, and [M/H] = $-$2.50 - 0.50. These curves have been interpolated to match the atmospheric parameters of the individual stars when determining the Li abundances. Additional details about the computation of

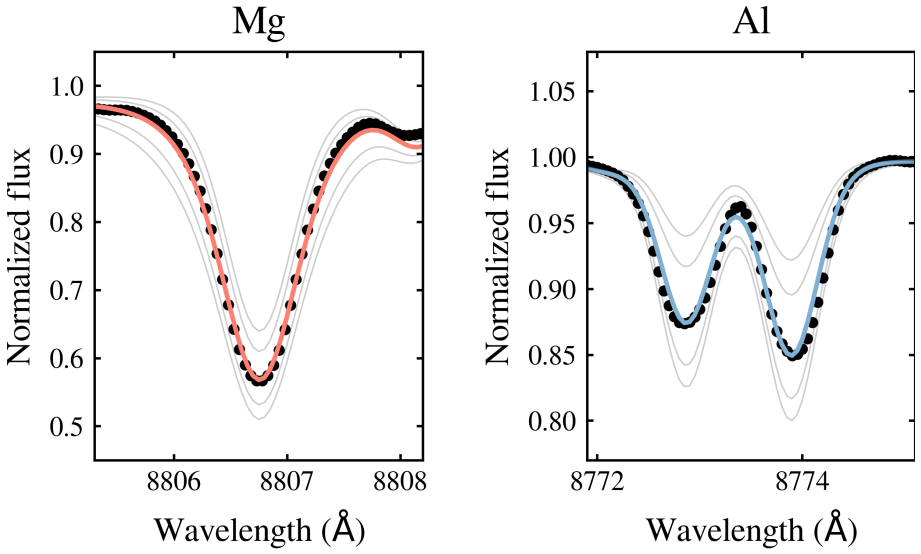


Figure 2: Examples of synthetic spectra fits to absorption lines in a GIRAFFE solar spectrum. Left panel: Fit to the Mg absorption line at 8806.75 Å. Right panel: Fit to the Al doublet around 8773 Å. In both panels, the observed GIRAFFE spectrum is shown by the black data points. The best fit synthetic spectra are indicated in color, while synthetic spectra with too low or too high input abundance are shown in gray.

the curves of growth and the Li abundance measurements can be found in Franciosini et al. (2022).

1.3 Structure of the thesis

We here outline the structure of the main body of this thesis: The following Section 2 presents an application of the classical spectral analysis methods. Precise abundances determinations from spectra of stars in the globular cluster NGC 1851 are used to separate the cluster stars into two chemically distinct subpopulations. The analysis shows that a precise investigation of multiple chemical elements is necessary to uncover this chemical separation.

Section 3 gives a theoretical introduction to convolutional neural networks. While this section does not include any scientific results, it presents our CNN architecture and explains how its different parts function. The details of the network training phase and the concept of network gradients are also given there.

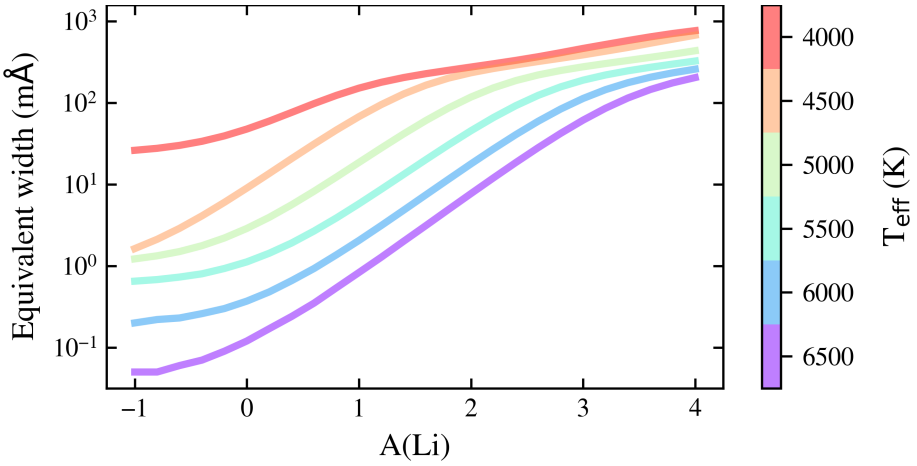


Figure 3: Curves of growth for FGK stars from Franciosini et al. (2022) for the Li I double feature at $\sim$ 6708 Å. The shown curves have been computed for $\log(g) = 4.5$ and [M/H] = 0.00 dex. The effective temperature for the curves increases from top to bottom.

In Section 4 it is shown that our convolutional neural network can predict multiple chemical abundances and atmospheric parameters from stellar spectra. In the same section, we present a method to identify spectra that are dissimilar to the training set spectra. The distinction between training-unlike and training-like spectra is crucial for real-life applications of neural networks to newly observed spectra. This sections also shows what exactly the network learns during the training phase.

In Section 5, we evaluate our network results by investigating some of the key properties of the Milky Way galaxy. Among others, we revisit the globular cluster NGC 1851, this time comparing the magnesium and aluminum abundances from Section 2 to the results from our network. In the same section, we also report the discovery of 31 previously unidentified lithium-rich giants in the Gaia-ESO iDR6 data set.

Finally, in Sections 6 and 7 we discuss some caveats of our neural network method and summarize the main points of this thesis.

2 The need for precise abundance measurements of multiple chemical elements: Two stellar populations in the globular cluster NGC 1851

Different elements are produced on different timescales, by various processes, in various environments. Studying the chemical composition of stars in a stellar population therefore gives insight into the evolution of the stars in the population and the formation history of the population as a whole. A successful application of this approach is the investigation of globular clusters in the Milky Way galaxy. Traditionally, these objects were considered to consist of stars that all have the same age and chemical composition (Bastian & Lardo 2018). It was therefore assumed that all stars in a globular cluster are of common origin. In the last two decades, this view has changed. The investigation of high-resolution spectra of many cluster member stars has revealed a rich chemical variety of stars in many globular clusters. These chemically peculiar clusters contain stellar populations with varying metallicities and differing abundances of slow neutron-capture process (s-process) elements (e.g., Carretta et al. 2010; Mucciarelli et al. 2015; Kovalev et al. 2019). This development was made possible by spectroscopic surveys, such as Gaia-ESO. These surveys provided high-resolution spectra of hundred thousand stars across the Milky Way galaxy in multiple wavelength ranges.

In this section, we present a detailed chemical study of the globular cluster NGC 1851. Abundances of up to 29 elements have been studied in 45 red giant stars in this cluster. The analysis reveals the existence of two chemically distinct sub-clusters. The figures and tables in the following sections are adapted from Tautvaišienė et al. (2022). All presented results are based on data that is published on the VizieR catalogue service[2].

Studies like this show the importance of obtaining precise abundances of multiple chemical elements.

[2]https://doi.org/10.26093/cds/vizier.36580080

2.1 Data and methods of analysis

The globular cluster NGC 1851 is massive ($M = 3.2 \cdot 10^5 M_\odot$) and relatively close with a distance of 15.1 kpc (Baumgardt & Hilker 2018). Its member stars are clearly separable from Milky Way field stars by the cluster's halo-like orbit and high radial velocity (320.5 km s^{-1}, Dinescu et al. 1999). For details on the target selection for globular clusters see Pancino et al. (2017a). Observations of the target stars have been carried out for the Gaia-ESO survey, using the high-resolution UVES spectrograph (see Sect. 1.2 for the survey details).

For our analysis, we use the stellar atmospheric parameters and chemical abundances (except those for C, N, and O) which have been published in the fourth internal data release of the Gaia-ESO survey. The presented abundances of Na and Ba have been corrected for non-local thermodynamic equilibrium (NLTE) effects. The Na corrections were taken from Lind et al. (2011), those for Ba from Korotin et al. (2014). We determined the abundances of carbon, nitrogen, and oxygen with a spectral synthesis method, as described in Sect. 1.2.1. Carbon abundances were measured from the two C_2 Swan bands at 5135.5 and 5635.2 Å. The oxygen abundance was determined by fitting the single forbidden [O I] line at 6300.3 Å. Due to the molecular equilibrium in stellar atmospheres, the C and O abundances are not independent of each other and have to be measured together. We did this by first measuring O and keeping its value fixed for the following C measurement. In the next step, the found C abundance was fixed for a new measurement of O. We repeated this cycle until both O and C abundances do not change anymore. The above-mentioned C and O features are well suited for measuring abundances because they are not sensitive to NLTE effects. In the case of the C_2 Swan bands, this is because they are formed by purely vibrational transitions. The [O I] line at 6300.3 Å also forms in LTE due to collisional coupling of the relevant energy levels (Asplund 2005). However, the region in the spectrum around our [O I] feature is contaminated by telluric lines. For our analysis, we used only those lines that were unaffected by telluric contamination. We also accounted for the two weak nickel lines that blend with the [O I] line. In our line list, we set the oscillator strengths of the two blending Ni lines to the values given by Johansson et al. (2003). Several $^{12}C^{14}N$

absorption features in the wavelength interval from 6470−6490 Å were investigated to determine nitrogen abundances. To make our abundance determination differential to the sun, we calibrated our analysis method to the solar spectrum from Kurucz (2005).

There are two categories of uncertainty sources in our abundance determinations. These are systematic errors, which affect all lines in a stellar spectrum simultaneously, and random errors in the analysis of individual lines. The systematic errors arise from uncertainties in the determined stellar atmospheric parameters. Random errors in the analysis of individual spectral lines are mainly due to errors in the normalization of the observed spectrum.

2.2 Chemical separation based on iron, nitrogen, and s-process elements

The investigation of the stars' composition reveals two chemically distinct sub-populations in the cluster. This distinction is most clear in the abundances of nitrogen and the s-processed elements (Y, Zr, Ba, La, Ce, Nd). The right panel of Fig. 1 shows the [s/Fe] ratios of the cluster stars relative to their nitrogen-to-iron ratio [N/Fe]. In this plot, two groups of data-points can be identified. Based on this, we separate our sample stars into the two sub-populations — the population with low average [s/Fe] and the population with high [s/Fe]. These two groups also differ in their average iron abundances (left panel of Fig. 1). We therefore refer to the two groups as the *metal-poor* and *metal-rich* populations. Their [Fe/H] values have a difference of 0.07 dex. The average [Fe/H] of the metal-rich sample is -0.98 ± 0.04 dex and the average of the metal-poor sample is -1.05 ± 0.05 dex. The average abundances of the s-processed elements and nitrogen, as well as the results for the other elements, are listed in Table 1.

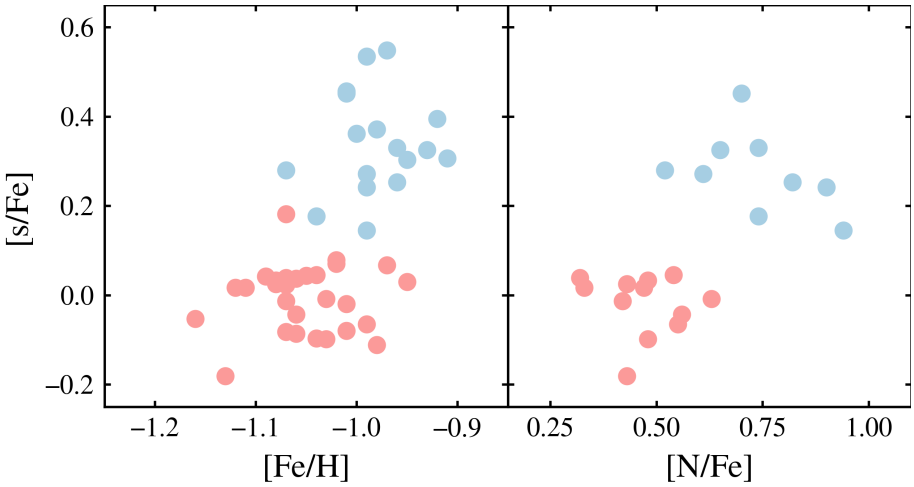


Figure 4: Chemical separation of cluster stars. Left panel: Average abundances of s-processed elements relative to [Fe/H]. Right panel: Average abundances of s-processed elements relative to their [N/Fe]. Our separation of the two subsamples is indicated by the colors of the data points (metal-rich sample in blue, metal-poor in red).

Table 1: Average abundances for the two populations and numbers of stars.

Parameter	Metal-poor			Metal-rich		
	Average	σ	N	Average	σ	N
[Fe/H]	−1.05	0.05	28	−0.98	0.04	17
A(Li I) $_{\rm 3D\,NLTE}$	0.14	0.66	27	0.46	0.59	17
[C/Fe] (C_2)	−0.31	0.08	14	−0.36	0.11	9
[N/Fe] (CN)	0.47	0.09	12	0.73	0.14	9
[O/Fe] ([O I])	0.27	0.18	28	−0.13	0.20	16
[Na I/Fe] $_{\rm NLTE}$	0.02	0.23	25	0.32	0.17	16
[Mg I/Fe]	0.26	0.08	28	0.25	0.10	17
[Al I/Fe]	0.17	0.19	28	0.38	0.15	17
[Si I/Fe]	0.07	0.05	28	0.09	0.05	17
[Ca I/Fe]	0.17	0.05	28	0.21	0.05	17
[Ca II/Fe]	0.24	0.12	25	0.29	0.15	15
[Ti I/Fe]	0.14	0.07	28	0.16	0.08	17
[Ti II/Fe]	0.17	0.05	28	0.18	0.04	17
[Sc I/Fe]	−0.01	0.12	28	−0.02	0.12	17
[Sc II/Fe]	−0.01	0.04	28	−0.02	0.05	17
[V I/Fe]	−0.11	0.07	28	−0.08	0.08	17
[Cr I/Fe]	−0.16	0.06	28	−0.12	0.07	17
[Cr II/Fe]	0.04	0.11	27	−0.05	0.08	17
[Mn I/Fe]	−0.37	0.05	28	−0.41	0.05	17
[Co I/Fe]	−0.09	0.04	28	−0.07	0.04	17
[Ni I/Fe]	−0.14	0.04	28	−0.15	0.06	17
[Cu I/Fe]	−0.33	0.17	28	−0.28	0.16	17
[Zn I/Fe]	−0.06	0.10	28	0.00	0.15	17
[Y II/Fe]	−0.24	0.06	28	−0.02	0.14	17
[Zr I/Fe]	0.15	0.18	27	0.37	0.15	17
[Mo I/Fe]	0.16	0.16	14	0.37	0.13	12
[Ba II/Fe]	0.10	0.13	28	0.55	0.19	17
[Ba II/Fe] $_{\rm NLTE}$	0.00	0.15	28	0.47	0.20	17
[La II/Fe]	−0.03	0.14	27	0.37	0.10	17
[Ce II/Fe]	−0.18	0.15	28	0.32	0.19	17
[Pr II/Fe]	0.46	0.05	15	0.63	0.12	12
[Nd II/Fe]	0.27	0.07	28	0.52	0.14	17
[Sm II/Fe]	0.37	0.13	14	0.55	0.15	12
[Eu II/Fe]	0.40	0.10	28	0.41	0.07	17
A(C+N+O)	7.97	0.11	12	7.94	0.08	9
C/N	0.66	0.13	12	0.35	0.13	9
[α/Fe]	0.16	0.04	28	0.18	0.05	17
[Iron peak/Fe]	-0.09	0.04	28	-0.07	0.04	17
[*s*/Fe]	−0.01	0.08	28	0.34	0.11	17

Figure 5 shows that the two populations also differ in the abundances of the individual s-processed elements. The average abundance ratios of Y, Zr, Ba, La, Ce, and Nd relative to iron are all higher in the metal-rich population than in the metal-poor one. By contrast, the ratio of the r-processed element Eu to iron is uniform across both cluster subsamples. There is also no difference between the metal-poor and metal-rich stars in the abundances of the eight iron-peak elements to iron which we investigate in this study (Fig. 6). As already seen for nitrogen, some other light elements also show differences between the two samples. The difference between the metal-rich and metal-poor sample is ≥ 0.3 dex for both [O/Fe], and [Na I/Fe]$_{\rm NLTE}$. To a lesser extent, the aluminum to iron ratios of the two samples are also different (by 0.21 dex). Carbon and the abundances of the α-elements (Mg, Si, Ca, Ti) relative to iron are almost the same in the two sub-populations. The distribution of the light elements is shown in Fig. 7.

2.3 Ages of the two cluster sub-populations from CNO abundances

In their high-precision photometric study, Milone et al. (2008) found that there exist two distinct subgiant branches in NGC 1851. They suggest that possible reasons for this are differences in age, overall [(C+N+O)/Fe] abundance, or both. This interpretation received support from Yong et al. (2009, 2015), who, in their analysis of 11 stars, observed a significant difference of 0.6 dex in [(C+N+O)/Fe] between the two branches. Simpson et al. (2017) obtained a narrower range in A(C+N+O) using medium-resolution spectra. However, Villanova et al. (2010) examined high-resolution spectra from 15 RGB stars in NGC 1851. They found no significant difference in the total A(C+N+O) content between the two populations.

The position of our two subpopulations in the color-magnitude diagram of NGC 1851 can be seen in Fig. 8. In this diagram, the metal-poor population is "on top" of the metal-rich one. This means that the metal-poor sub-branch is brighter than the metal-rich branch. We find a spread of $\sim$0.1 dex of A(C+N+O) in both populations, but no significant difference in their average C+N+O abundances (Table 1 and Fig. 9). Gratton et al. (2012) also observe a brighter, metal-poor sub-branch, and a fainter sub-branch which is more metal-rich in NGC 1851. More-

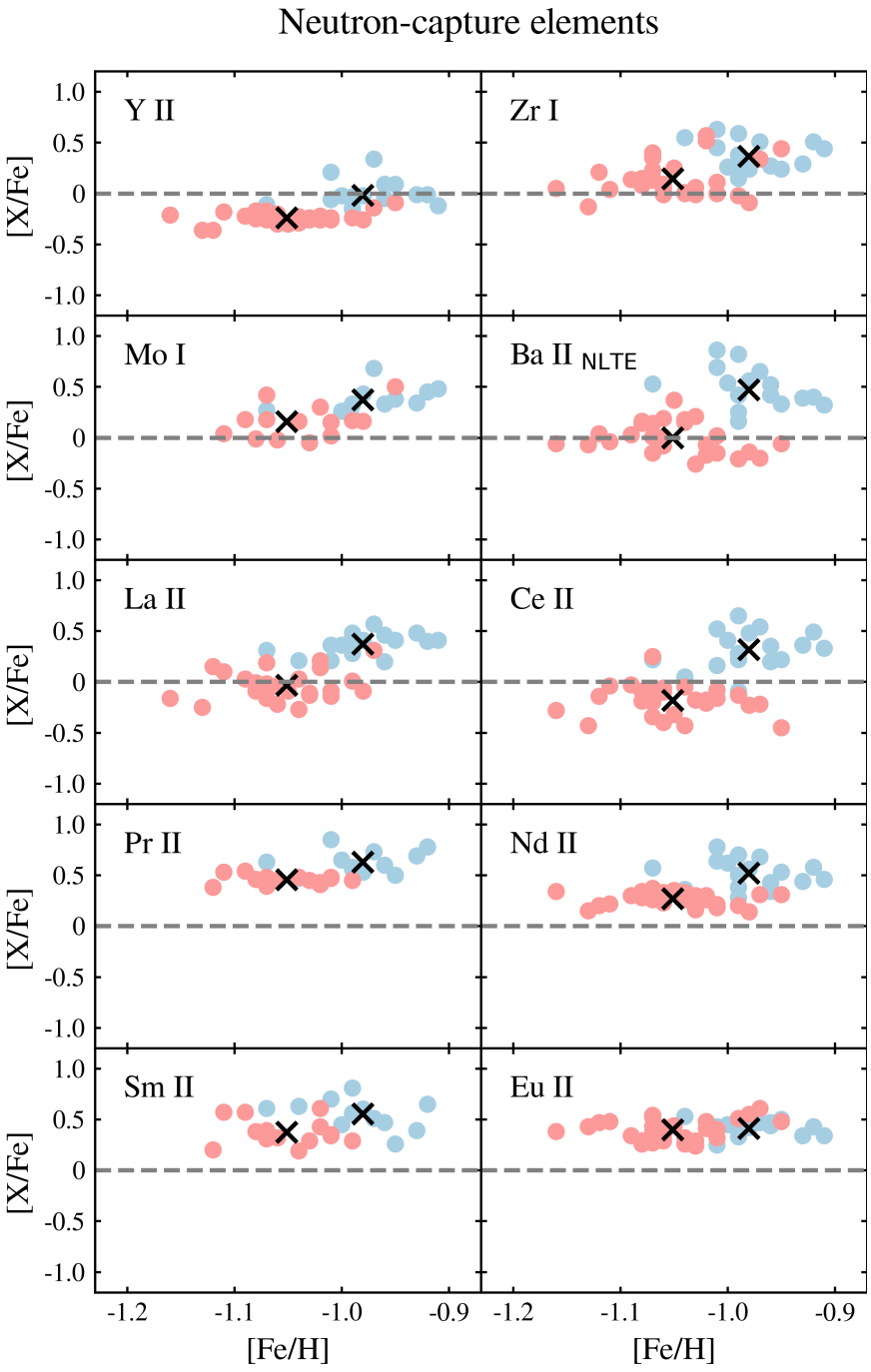


Figure 5: Element-to-iron ratios [X/Fe] of individual neutron-capture elements relative to iron. As in Fig. 1, the metal-rich population is shown in blue, the metal-poor one in red. The average abundances of the two sub-populations are marked by black crosses

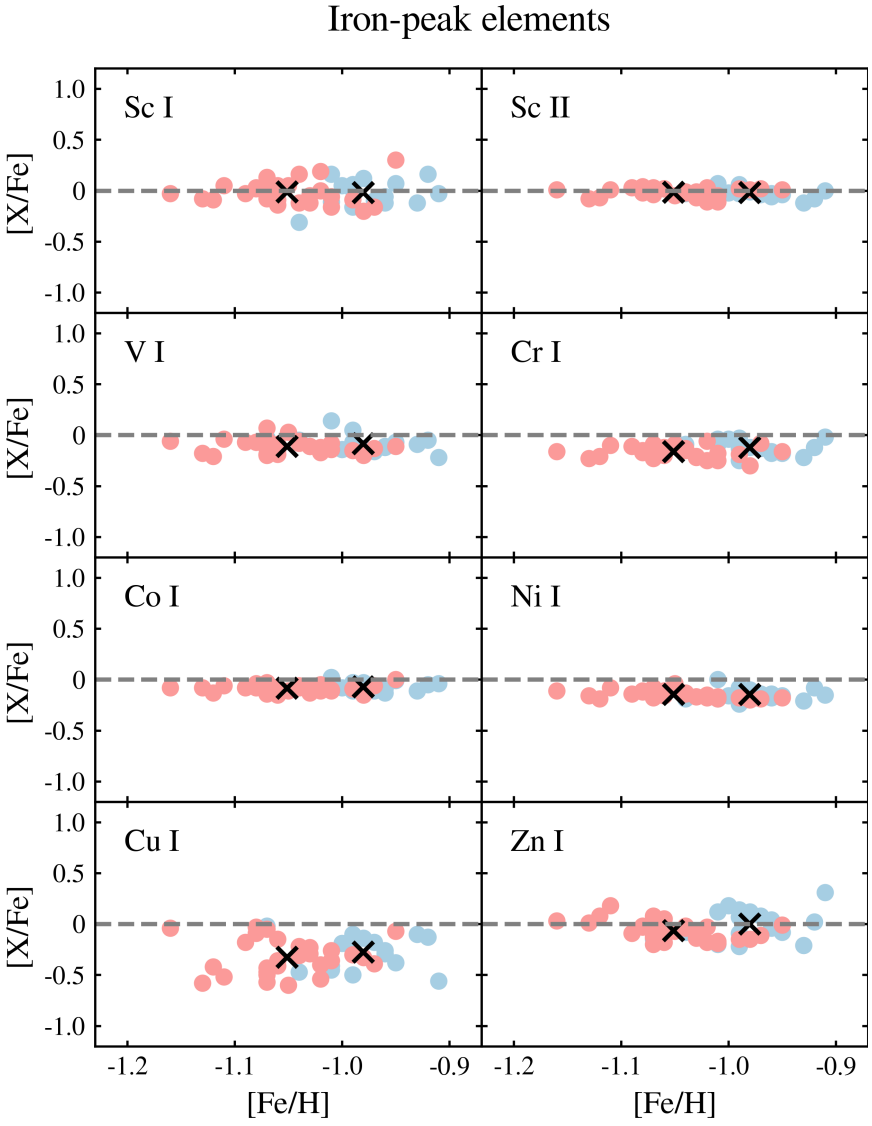


Figure 6: Same as Fig. 5, but for iron-peak elements.

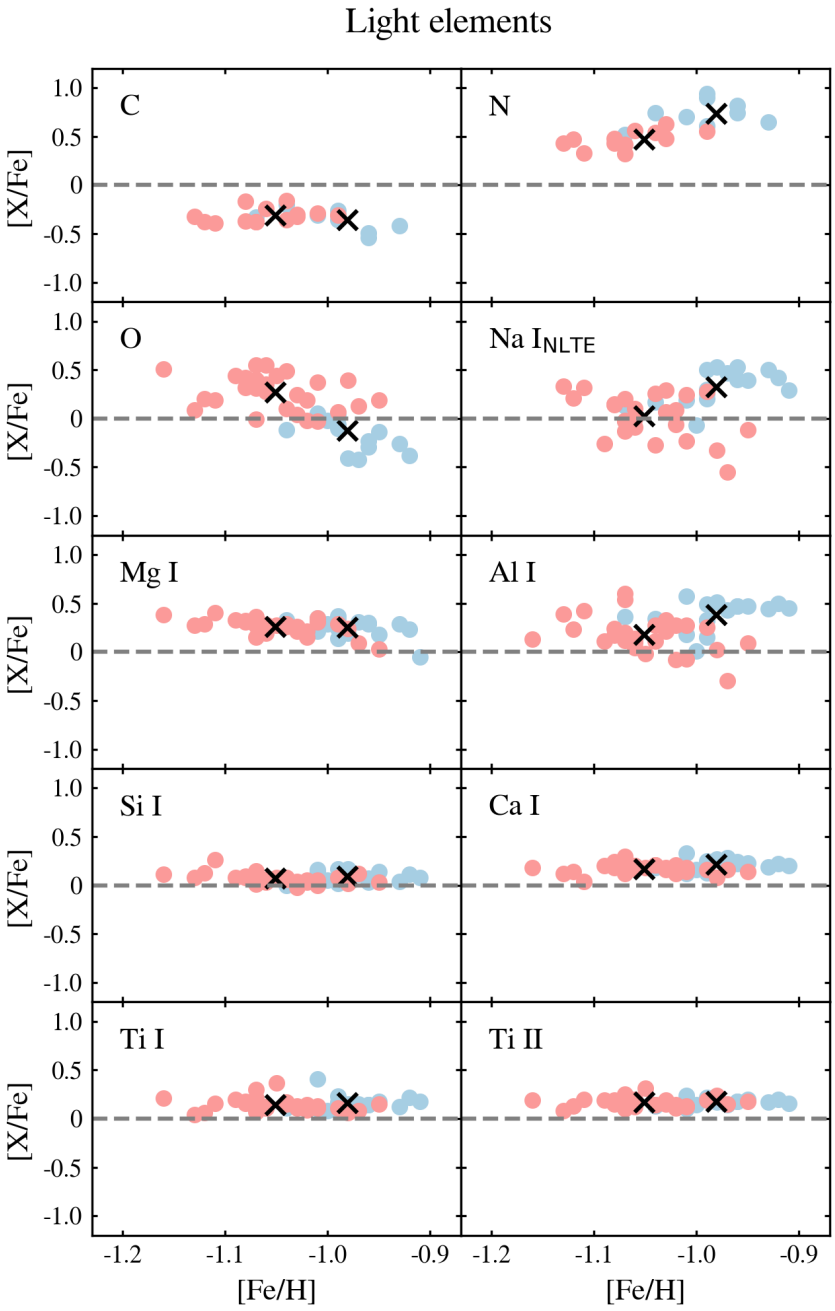


Figure 7: Same as Fig. 5, but for the light elements.

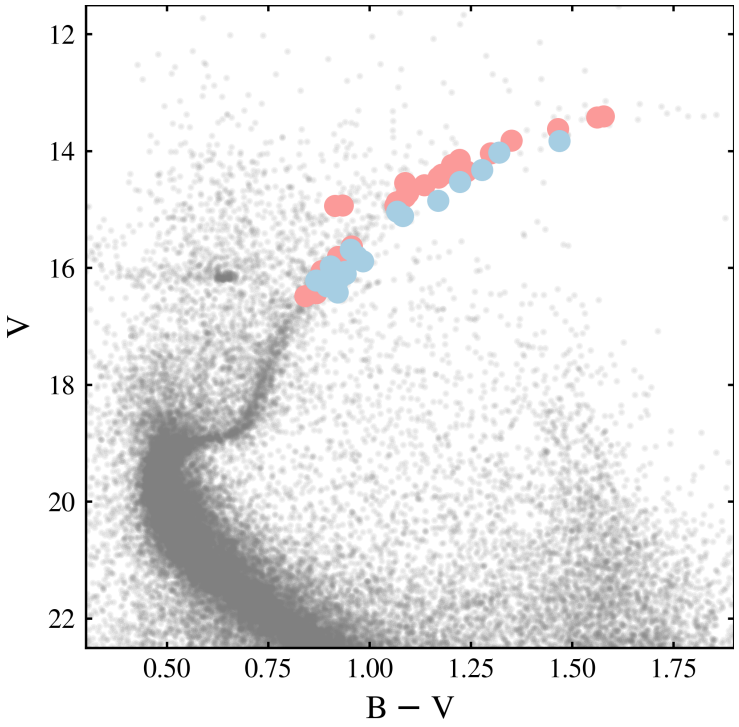


Figure 8: Positions of our metal-poor (red) and metal-rich (blue) subsamples in the color-magnitude diagram of NGC 1851. The V and B magnitudes of our 45 sample stars and of ∼60000 other cluster members were taken from the photometric catalog of Stetson et al. (2019).

over, they find that the metal-rich population would be ∼0.6 Gyr older than the metal-poor one, if the C+N+O abundance is the same for both. Based on this, we conclude conclude that our metal-rich subpopulation is by ∼0.6 Gyr older than the metal-poor one.

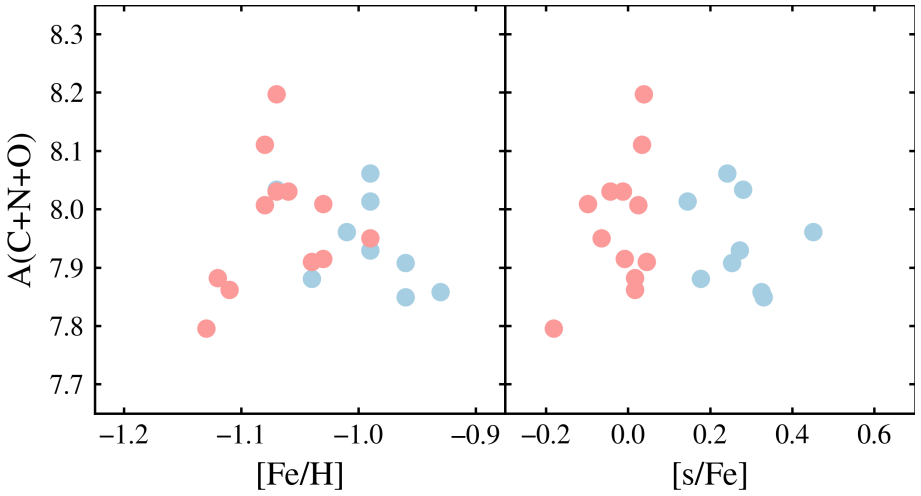


Figure 9: Total A(C+N+O) abundances of stars in our two subpopulations relative to [Fe/H] (left panel) and s-process elements [s/Fe] (right panel).

3 Convolutional neural networks

A convolutional neural network (CNN) is a type of machine learning method. Machine learning methods are complex, non-linear algorithms used to extract information from input data. These algorithms typically have many free parameters, which are not specified by the user. Instead, the algorithm "learns" the values of the parameters during a "training" phase. Subtypes of machine learning methods are defined by the distinct designs of their training phases:

- Supervised methods require labeled training data for the training phase. The labels of the input data, which can either be categorical or continuous, must be determined before the training. Training is the search for the parameter values that minimize the difference between the pre-determined labels and the algorithm's output label predictions. Supervised machine learning methods therefore learn by solving an optimization problem.

- Unsupervised methods work with unlabeled input data. These algorithms try to find relationships between the individual data points in the training data without any exterior information. Unsupervised methods are typically used to find clusters of similar

data points in large, high-dimensional data sets.

Convolutional neural networks belong to the class of supervised machine learning methods. The structure of CNN algorithms differs from other supervised machine learning methods. In the following, we describe the main building blocks of a CNN and their connections (the network "architecture"). We focus on our implementation of a CNN that can predict continuous labels from 1D stellar spectra.

3.1 Architecture of a CNN

The defining building blocks of a CNN are its convolution layers. These layers are designed to find features and patterns in the input data. In stellar spectra, these features could be single absorption lines, relations between groups of absorption lines, or the mean slope of the spectrum. The found spectral features are the basis for the calculation of the output label values. This calculation is done by the dense layers, which follow the convolutional layers in the CNN architecture. Every layer in a CNN can be followed by a regularization layer. In complex network architectures, regularization is needed to prevent the network from over-fitting to the training data. Table 2 shows the architecture of our CNN. We found this final architecture by experimentation, starting from a simple network with one convolution layer and one dense layer. Additional layers were added and the number of convolution kernels and dense layer neurons was adjusted until the network performance was satisfactory. Our choice of regularization layers was inspired by the network architecture of Guiglion et al. (2020). It should be mentioned that the architecture presented in Table 2 produces good results for our purposes. However, even better architectures for future projects might be found through further experimentation. Our final network contains 560 599 free parameters, most of them in the first dense layer. This first dense layer receives the flattened outputs from the convolution layers, which contains all the information about the found features in the training spectra. As will be detailed in Section 3.2.1, the free parameters will be optimized to accurately predict the labels of the training set spectra. We use 8807 training spectra with 8669 pixels each. In total, the $\sim 6 \cdot 10^5$ parameters will be fit to $8807 \cdot 8669 \approx 70 \cdot 10^6$ data points.

This low ratio of free parameters to data points ensures that the network training is not overdetermined by too many free parameters.

Below follow detailed descriptions of the CNN layer types.

3.1.1 Input layer

The input layer receives the raw input data and passes it on to the first convolution layer. Before the data is passed on, it may be prepared so that all samples are scaled and distributed evenly. Our input data are normalized stellar spectra, which have been corrected for redshift, and they all cover the same wavelength range. Therefore, no extra scaling or distribution has to be done in our input layer. An individual input spectrum is represented by a 1-dimensional array (or list) of flux values. The wavelength values of the individual points in a spectrum are not important for the prediction of the CNN. However, the order (index) of the spectrum's flux values must be preserved.

3.1.2 Convolution layers

In the first convolution layer, an input spectrum is convolved with one or multiple convolution kernels. The result of these operations are the feature maps, so named because they show the positions of various features in the input spectra. The type of feature, which is represented in a feature map, depends on the parameters of the used kernel. Figure 10 depicts the convolution of an input spectrum with the kernel $[-1, 1, -1]$.

The parameters of this kernel are chosen so that it detects absorption lines in the input spectrum. Therefore, the feature map shows the highest response at the positions of the absorption lines. If a spectrum without any absorption features is convolved with this kernel, the resulting feature map will be a horizontal line. The information about the lack of absorption is passed on to the other layers of the CNN by this flat feature map.

The kernel from Fig. 10 is hand-designed to detect absorption lines. However, in a typical CNN application, the values of the kernel parameters are unknown before the network training. During the training phase, the CNN finds the kernels that result in the best precision accuracy for the spectra in the training set. A convolution layer can have

Table 2: Architecture of our CNN. The input layer reads in the flux information of the stellar spectra. It is followed by three 1D convolution layers. An activation layer directly follows each convolution layer. The output from the third convolution block is then flattened to serve as input for the dense layers. Three dense layers (with a dropout layer after each) interpret the spectral features, found by the convolution layers, into output labels. The outputs from a last dense layer are the values of our five stellar labels (atmospheric parameters and elemental abundances). The order and number of the layers, as well as the hyperparameters, are set by the user. The values of the free parameters are initialized randomly and are optimized during the network training.

Layer	Hyperparameters	Free parameters
Input	Input shape: 8669	
1D convolution	filters: 8, kernel size: 20	168
Activation	LeakyReLU	
1D Max-pooling	pool size: 2	
1D convolution	filters: 6, kernel size: 20	966
Activation	LeakyReLU	
1D Max-pooling	pool size: 2	
1D convolution	filters: 4, kernel size: 20	484
Activation	LeakyReLU	
Flatten		
Dropout	dropout rate: 0.2	
Dense Layer	Neurons = 64	546 368
Dropout	dropout rate: 0.2	
Activaton	LeakyReLU	
Dense Layer	Neurons: 128	8320
Dropout	dropout rate: 0.2	
Activaton	LeakyReLU	
Dense Layer	Neurons: 32	4128
Dropout	dropout rate: 0.2	
Activaton	LeakyReLU	
Dense layer	Neurons: 5	165
Activation	Linear	
		Total: 560 599

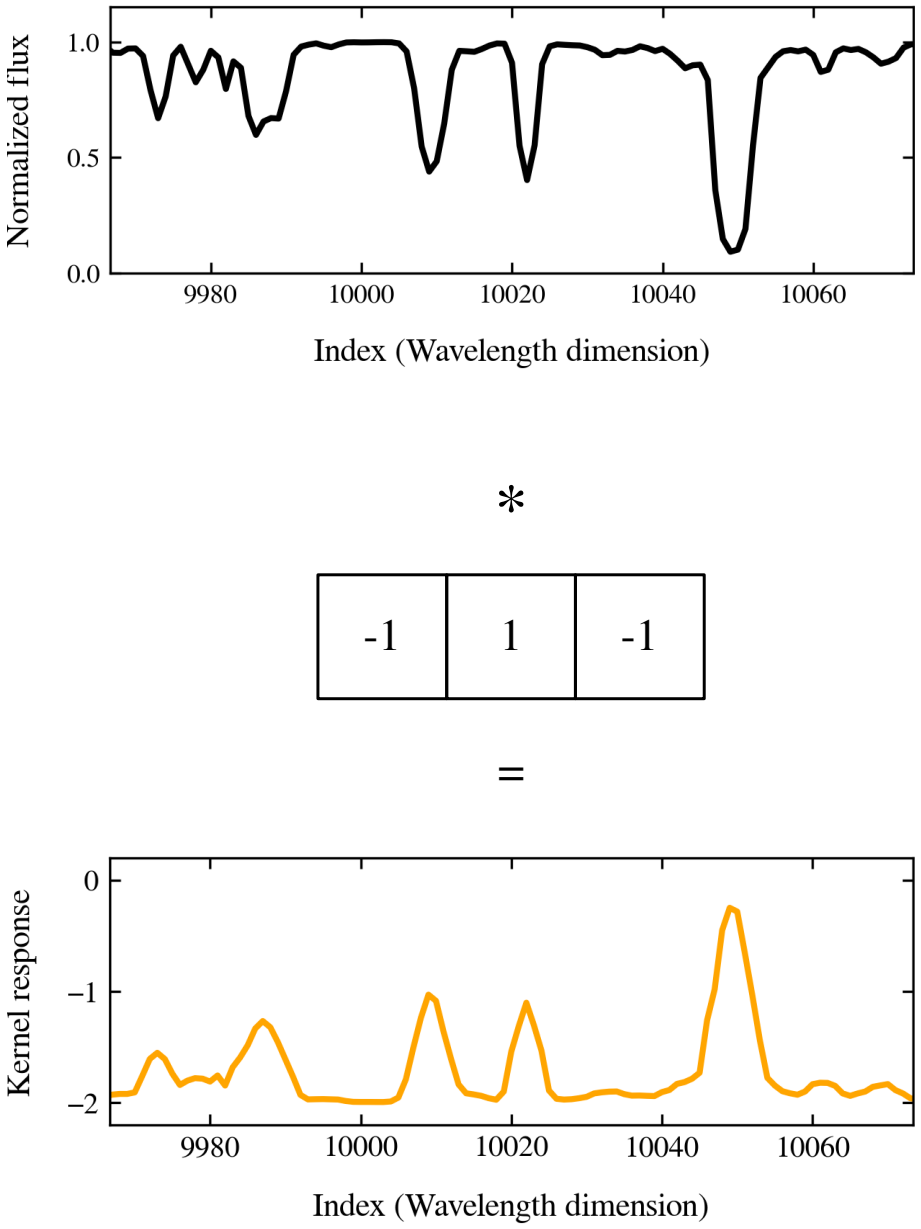


Figure 10: Example of the convolution of an input spectrum with a convolution kernel. The convolution operation is symbolized by an asterisk ($*$). Top panel: A part of the input spectrum. Bottom panel: Feature map as the result of the convolution of the input spectrum with the kernel $[-1, 1, -1]$.

multiple kernels, that independently produce feature maps of the layer input. Each of these different feature maps then represents a different aspect of the input spectrum. This is analogous to how the red, green, and blue color channels each represent a different aspect of a color image.

If a CNN has multiple convolution layers, the feature maps from the first layer will be convolved themselves in the following convolution layer. As the input data transforms as it passes from layer to layer, the resulting feature maps of feature maps become difficult to interpret. Figure 11 shows how a convolution layer with three inputs (feature maps from a previous layer) and four outputs works. Each of the three input maps is convolved with four individual kernels, making 12 different kernels in total. The 12 intermediate feature maps are then added together, so that each output is the sum of one intermediate feature map of each of the three inputs. In this way, each of the four outputs contains information from all three inputs.

A bias value (additional free parameter) is added to each of the four output channels. This effectively shifts the whole feature map up or down, as is demonstrated in Fig. 12.

The values of the pixels in the feature maps becomes important in the following activation operation. In the activation, the feature map is transformed according to an activation function. The value of every individual pixel in the map is changed in this operation. There are several types of activation functions, each adding non-linearity to the neural network. In recent machine–learning applications, the "Leaky ReLU" activation function (Fig. 13) is most often used. It leaves positive and zero output values unchanged, and multiplies negative outputs with a small positive value. Or, notated mathematically (Maas 2013):

$$LeakyReLU(x) = \begin{cases} a \cdot x & \text{if } x < 0 \\ x & \text{otherwise} \end{cases}$$

Where x is the value of an individual pixel in a feature map and $a \in (0,1)$. The effect of applying this activation function to a feature map is shown in Fig. 14.

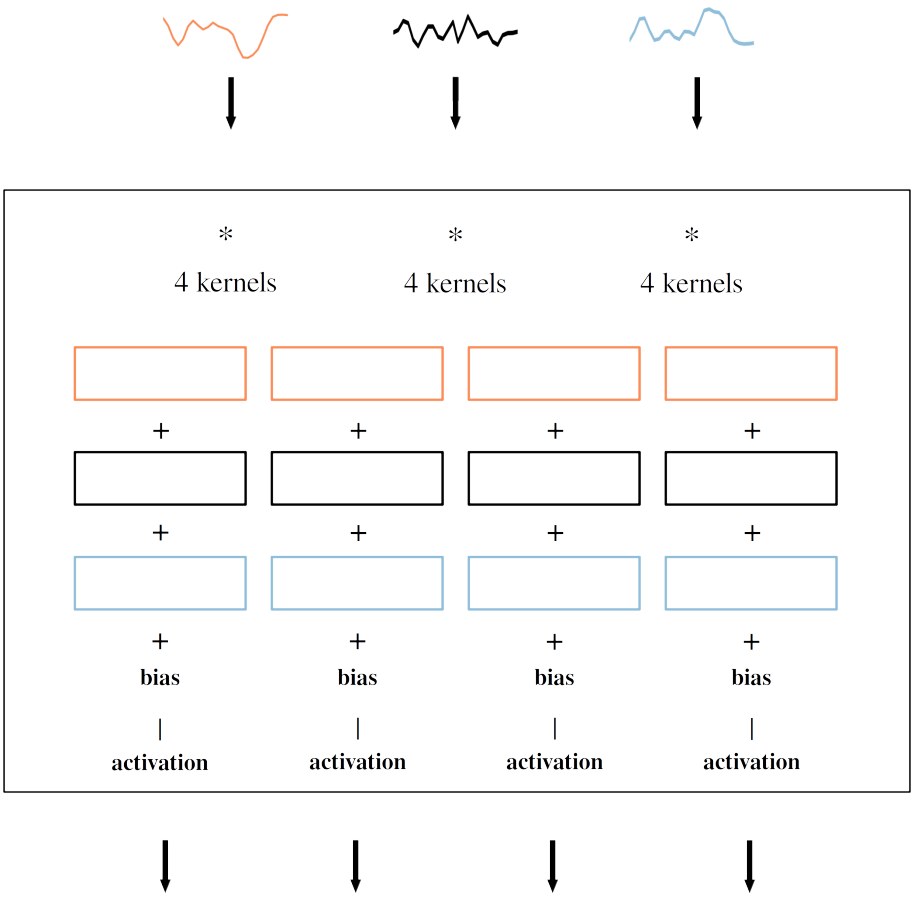


Figure 11: Operations inside a convolution layer with four filters. The layer receives three feature maps from a previous convolution layer, and forwards four feature maps to the next layer in the network architecture.

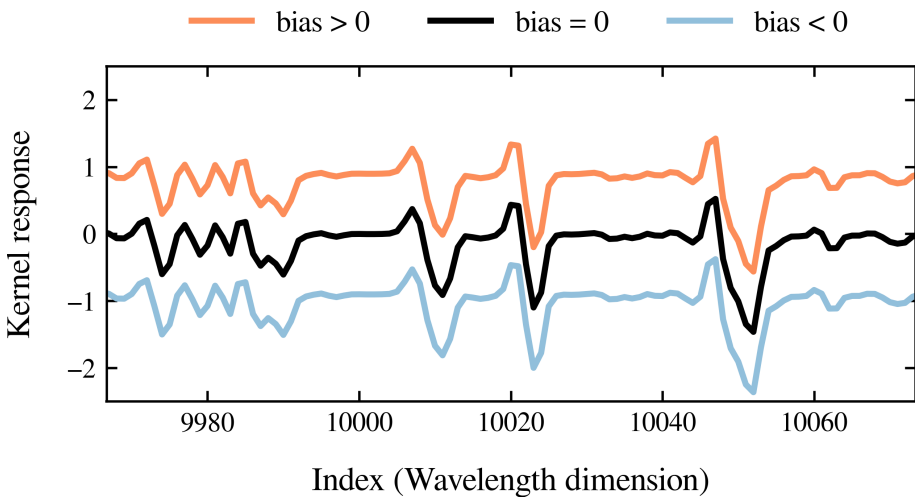


Figure 12: Effect of adding a bias value to a feature map. Positive bias shifts the feature map up (orange), negative bias shifts the map down (blue).

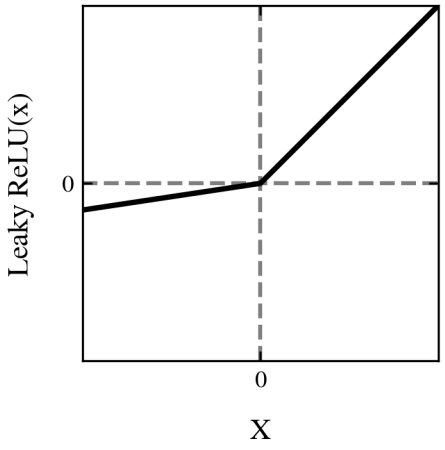


Figure 13: Leaky ReLU function. Input values x > 0 remain unchanged, while negative input values are multiplied by a value between 0 and 1.

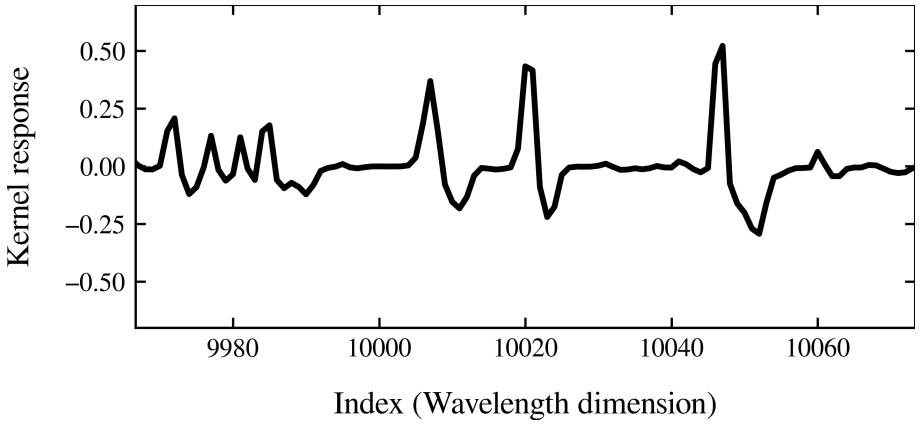


Figure 14: Convolution layer output after applying Leaky ReLU activation. This example shows the activated version of the black line from Fig. 12.

The output of the last convolution layer is a set of activated feature maps. These maps must be prepared in the following "flattening" layer before passing them to the first dense layer.

3.1.3 Flattening

The output from the last convolution layer consists of multiple feature maps (i.e., the output is multidimensional). Since the dense layers can only process one-dimensional input, the convolution output must be "flattened" before it can be used as input for the first dense layers. This is done by simply appending the individual feature maps to each other. For example, the flattening of four feature maps with 1000 pixels each will result in a one-dimensional output with 4000 pixels. Each of the neurons in the following dense layer will then take 4000 input values.

3.1.4 Dense layers

A dense layer consists of one or multiple neurons (also called units or nodes). Every neuron takes multiple inputs from a previous layer (either the flatten layer or a previous dense layer). Inside a single neuron, a linear combination of all input values is calculated, and a bias value is added. The result is transformed by an activation function and then passed as the neuron output to the next layer. As for the convolution layers, the activation function is typically "Leaky ReLU". A scheme of the operations inside a single neuron with four inputs is shown in Fig. 15. The four input values $[x_1, x_2, x_3, x_4]$ are multiplied elementwise with the four weights $[w_1, w_2, w_3, w_4]$ and added together. A bias value is then added to the result. The four weights and the bias term are free parameters of the CNN. Their optimal values are found in the network training. The result of this linear combination is a single value, that is passed to the activation function. The output of the activation function is then the final output of the neuron. This single output value is passed on to the neurons of the following dense layer. In this way, every layer neuron is connected to all neurons from the previous layer and passes its output value to all neurons of the next layer. Neurons within one layer are independent of each other.

3.1.5 Output layer

The output layer is a final dense layer, whose number of neurons is equal to the number of labels which should be predicted. The outputs of these neurons are directly proportional to the label predictions. If the

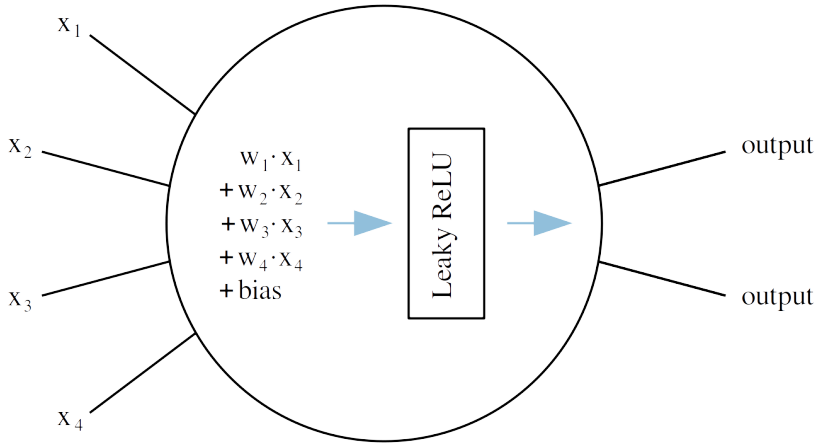


Figure 15: Scheme of the operations inside a single neuron. This neuron is part of an intermediate dense layer, between a layer with four neurons and a layer with two neurons.

network is used to solve a regression problem, no activation function is applied to the neuron outputs. If the network is built for classification tasks, sigmoid or softmax functions are often used (Singh et al. 2023).

3.1.6 Regularization layers

Regularization techniques can be used to prevent overfitting. Overfitting occurs when a network predicts accurate labels for the training set but performs poorly on other spectra. In this work, we add two types of regularization layers to the CNN architecture. These are the max-pooling and dropout layers.

Max-pooling is applied to the output feature maps of the convolution layers. It works by dividing a feature map into windows of equal widths. In each of these windows, only the maximum point is kept, and the other points are discarded. An example of the effect of max-pooling on a feature map is shown in Fig. 16.

The dropout layers are located between two dense layers in a network architecture. Dropout deactivates a fraction of neurons in the previous layer by setting their output values to zero (they are “dropped out”). This deactivation only happens during the network training. A

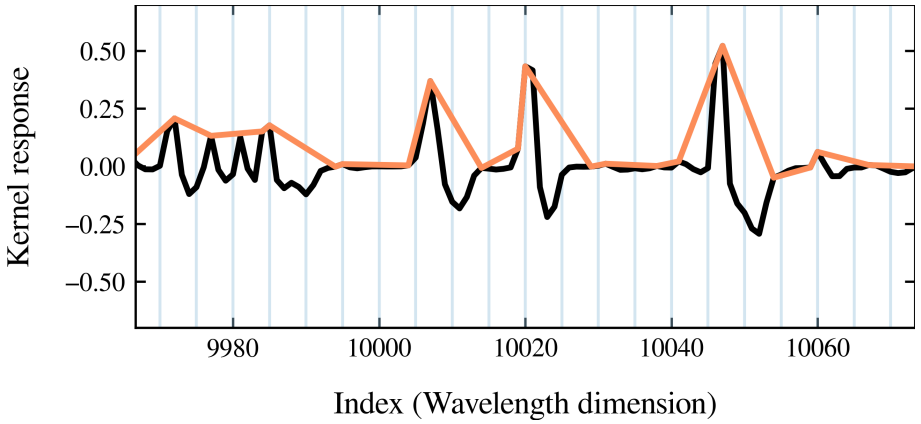


Figure 16: Max-pooling of a feature map. The original map is shown in black, the result of the max-pooling operation in red. Blue vertical lines show the borders of the max-pooling windows. Only the maximum flux value in a window is kept, the lower flux values are discarded. The windows have a width of 5 pixels, which corresponds to a wavelength range of 0.15 Å in this example.

new set of dropped neurons is chosen at random for every training spectrum. After the network training, dropout is not active anymore.

Other regularization methods include performing changes on the input data ("augmentation") and the transformation of the input labels. See Santos & Papa (2022) for a summary of regularization methods for convolutional neural networks.

3.2 Network training

When a network architecture is set up, the values of the convolution kernels, the weights in the dense layers, and the bias values are unknown. In the following text, we will call these values collectively the “free parameters” of the CNN. The search for the parameters that result in the best network output is called *network training*. We consider the network output optimal, when the variation between the pre-determined labels of a set of validation spectra and the CNN output for the same spectra is minimal.

3.2.1 Training and validation data sets

For the training of a neural network, two separate sets of spectra with associated labels are needed. One of these sets is the training set. In the training phase, the network repeatedly predicts the labels for the spectra in the training set. The values of the free parameters in the network are updated after every training iteration, with the goal of increasing the accuracy of the predictions in the next iteration.

The other data set is the validation set. It is used to detect over-fitting of the network, by comparing the network performance on the validation spectra to the performance on the training spectra. If the network performs poorly on the validation spectra but well on the training spectra, the network is over-fitting. This shows that the current network has specialized on the training set only and cannot generalize to other spectra. The user can then adjust the network architecture to prevent the over-fitting, by reducing the complexity of the network (fewer convolution and/or dense layers, fewer kernels/neurons in the individual layers). Another strategy to prevent over-fitting is to use regularization methods (see Sect. 3.1.6). The validation data is also used to monitor the training progress and to stop the training when the validation accuracy does not increase anymore.

3.2.2 Optimizing the free network parameters

The variation between input and output is also called the loss and is calculated with a loss-function. Network training is an optimization problem, where the loss is minimized by optimizing the free parameters of the network. The optimization process is carried out according to the following algorithm:

0. Random initialization of the free parameters
1. Network predicts labels for all spectra in the training set and validation set
2. Loss between pre-determined training labels and CNN output is calculated
3. Free parameters are updated

These steps are repeated, until the loss does not decrease anymore. One iteration of the steps 1-3, in which all training spectra pass once through the network, is called an epoch. The main task in the network training lies in step number 3, the updating of the parameter values. This must be done in a way that guarantees that the loss decreases with every epoch. A typical loss function for regression problems is the mean-squared error (MSE) between input labels and output labels. The total MSE loss is the average of the MSEs of the individual labels:

$$MSE_{total} = \frac{1}{N_{labels}} \sum_{label=1}^{N_{labels}} MSE_{label} \tag{1}$$

$$MSE_{label} = \frac{1}{N_{samples}} \sum_{i=1}^{N_{samples}} (input_i - output_i)^2 \tag{2}$$

Every individual output depends on the values of the free parameters in the CNN architecture. Equations 1 and 2 therefore connect the total loss to the free parameters.

Commonly used techniques for updating the network parameters are based on the gradient descent method. Gradient descent calculates the derivative of the loss function with respect to each individual free parameter in the network. Because of the connection between loss function and free parameters, it is possible to calculate the derivative of the total loss with respect to every individual parameter. These derivatives are then used to update the parameter values p for the next iteration in the optimization. If the value of the derivative is positive, then an increase of the parameter value also increases the total loss. Equation 3 shows the rule for updating a single parameter in the network:

$$p_{updated} = p_{old} - \frac{d(total\ loss)}{d(p_{old})} \cdot \alpha \tag{3}$$

where α, the so-called learning rate, is > 0. When this learning rate is too high, the updating steps are too large, and the gradient descent

method will "overstep" the optimal parameter value. Successive updating iterations will then cause the parameter value to oscillate around its optimum. If alpha is too small, the parameter will get stuck at a value where the loss function has a local minimum. The learning rate is therefore an important hyperparameter that must be set by the user to ensure a successful training of the CNN.

In the simplest case, the total loss is calculated after the network has predicted label values for all spectra in the training set. Gradient descent is then used to update all free parameters once per training epoch. A modification of this method is called stochastic gradient descent. Here, the parameters are updated after every individual training spectrum. An intermediate between these two approaches is the mini-batch gradient descent. This method updates the parameters multiple times per epoch, after a specified number of spectra has passed through the network.

For the training of our network we used the mini-batch gradient descent method, with a batch-size of 16 spectra and a learning rate of 0.0001. We also experimented with other optimization algorithms, including *Adam* and *RMSProp*. Both of theses are variants of the basic stochastic gradient descent method, and produced nearly identical results. We therefore chose to remain with the simple mini-batch gradient descent as our optimizer.

3.2.3 Testing of the network

After the successful training of the network, its performance on new spectra has to be tested. This requires a test set of labeled spectra. These spectra have not been used for training or validation during the training phase. Therefore, the performance on the test spectra shows what accuracy can be expected from applying the trained network to a large set of unlabeled spectra.

3.3 Explainability of the network predictions

As has been established above, a convolutional neural network is a complex non-linear function that takes an array of flux values (spectrum) as an input and calculates the label predictions as an output. As with other functions, we can calculate the derivative, or gradient, of this net-

work function with respect to the individual input variables (the fluxes). These gradients are a measure of the importance of individual flux values for the network output. If a network gradient at a certain pixel in the input spectrum is high, then this pixel is important for the label prediction. Positive gradients show that the higher the flux value at a pixel, the higher the value of the output prediction. Negative gradients indicate inverse correlation: Higher flux values then lead to lower values of the output label. The network gradients therefore supply important explainability for the predictions of neural networks, which are still often considered black box algorithms (Haar et al. 2023).

We show how network gradients can explain CNN label predictions for stellar spectra in Sect. 4.3.

4 Investigation of GES iDR6 spectra with CNNs

In the following sections, we show the application of the CNN method to spectra from the Gaia-ESO survey. The details of the sample selection, network training, testing, and physical validation of the network predictions are presented. We also demonstrate that CNNs can identify relevant spectral features and use them for the label predictions in a physically intuitive way.

The methods and results in these sections are based on the work that is presented in Ambrosch et al. (2023). This publication focuses on the training of a CNN for the simultaneous prediction of the labels $T_{\rm eff}$, log(g), [Mg/Fe], [Al/Fe], and [Fe/H]. These results are complimented by those published in Nepal et al. (2023), who trained a CNN to predict $T_{\rm eff}$, log(g), [Fe/H], and the lithium abundance A(Li). Both works use CNNs that were built and trained in a Python programming environment. Within Python, the open source deep-learning library KERAS (Chollet et al. 2015) was used together with the TENSORFLOW backend (Abadi et al. 2015).

4.1 Spectra and associated labels for the training and testing of the neural network

Our data set consists of spectra, associated stellar parameters, and abundances from the GES iDR6 data set. In the Gaia-ESO survey, atmospheric parameters and chemical abundances are determined by multiple workgroups that apply different codes and methodologies to the survey spectra. The spectra which we used in this study were taken with the GIRAFFE spectrograph that covers the visible wavelength range of 370–900 nm. Several setups divide the whole GIRAFFE spectral range into smaller parts. For this study we chose the HR10 (533.9–561.9 nm, R = 19800) and HR21 (848.4–900.1 nm, R = 16200) setups because they cover important Mg and Al absorption features.

4.1.1 Preparation of GES data for neural network training

For our analysis, we used normalized 1-D spectra from the GES archive. We removed bad pixels and cosmic ray spikes where necessary. To do so, we first calculated the median of all spectrum flux values. We then looked for cosmic ray spikes by finding all pixels with flux values that exceeded this median flux by five sigmas. The spikes were removed by setting their flux values to be equal to the spectrum median flux. Pixels with zero flux values were also set to the median flux. Afterward, we corrected the spectra for redshift based on the radial velocity provided by GES. To reduce the number of pixels per spectrum and therefore the computational cost of the further analysis, we re-binned the spectra to larger wavelength intervals per pixel. The HR10 spectra were resampled to 0.06 Å per pixel and the HR21 spectra to 0.1 Å per pixel. The original bin size for both setups is 0.05 Å per pixel. After re-binning, the spectra were truncated at the ends to ensure that all spectra from one setup share the exact same wavelength range. Eventually, we combined the HR10 and HR21 spectra to create one input spectrum per star for our network. The combined spectra are composed of 8669 pixels each and cover the wavelength ranges from 5350-5600 Å and 8480-8930 Å.

To build our training set, we performed several quality checks to ensure that our network will be trained on high-quality data. Spectra with signal-to-noise ratio (S/N) < 30 and large errors in atmospheric parameters and elemental abundances ($e T_{\rm eff} > 200$ K, $e\log(g) > 0.3$ dex, eA(element) > 0.2 dex) were discarded, as well as spectra that were marked with the TECH or PECULI flags or have rotation velocities > 20 km s^{-1}. Similar to Guiglion et al. (2020), we tested the inclusion of low S/N spectra into our training set. This increases the number of training spectra, but the training performance gets worse and the overall prediction quality of our network decreases. While too noisy spectra are worsening the performance of our network, a moderate degree of noise is beneficial because it plays a significant role in the regularization of the training process (Bishop et al. 1995, particularly Sect. 9.3 there). To exclude very noisy spectra, while still utilizing the regulatory effect of noise in the training data, we set the lower S/N limit for our training set to 30. We also removed spectra that showed a difference larger than 0.2 dex between the provided metallicity [M/H] (as a stellar atmospheric

parameter) and the iron elemental abundance [Fe/H].

We further examined the remaining spectra to find outliers and incorrect measurements. To investigate the similarity between all the spectra, a t-distributed stochastic neighbor embedding (t-SNE) analysis was employed. The t-SNE analysis is a popular unsupervised machine learning technique used to visualize the internal relationships and similarities in high dimensional data sets. This is done by giving each data point a location in a two- or three-dimensional similarity map (van der Maaten & Hinton 2008). In our case, the data points are the individual spectra, and the data set is n-dimensional, where n is the number of pixels in each spectrum. Figure 17 shows a two-dimensional similarity map for our combined spectra, obtained with the SKLEARN.MANIFOLD library for python programming (Pedregosa et al. 2011). Every point in the map corresponds to one spectrum, and the distance between the individual points is related to the similarity of the shapes of the individual spectra. There are two main branches in the map with several sub-structures. The two branches represent spectra from stars in two distinct populations: Main sequence stars with surface gravity $\log(g) \gtrsim 3.5$ and stars in the giant branch with lower $\log(g)$ values. The different physical properties in stellar atmospheres are reflected in the shapes of their spectra, which in turn determine their locations on the t-SNE map. The connection between physical parameters and spectral features is what our CNN learns during the training phase. It is worth to mention that t-SNE on its own has also been used to classify spectra: Traven et al. (2017), for example, used t-SNE as a tool to separate GALAH spectra into different, physically distinct classes; Matijevič et al. (2017) used t-SNE to search for metal-poor stars in the RAVE survey.

We see several outlier-spectra in our Fig. 17. Upon inspection, these spectra show signs of emission lines, have distorted absorption features, or have suffered from failed cosmic removal or wrong normalization. We excluded these outliers from the further analysis. For the analysis of future surveys such as WEAVE and 4MIDABLE-HR surveys, including emission line stars will be a necessity, as we expect many young stars to be observed. We note that the initialization of our t-SNE application includes an element of randomness, which results in slightly different shapes of the map after every run. The map will also look different for

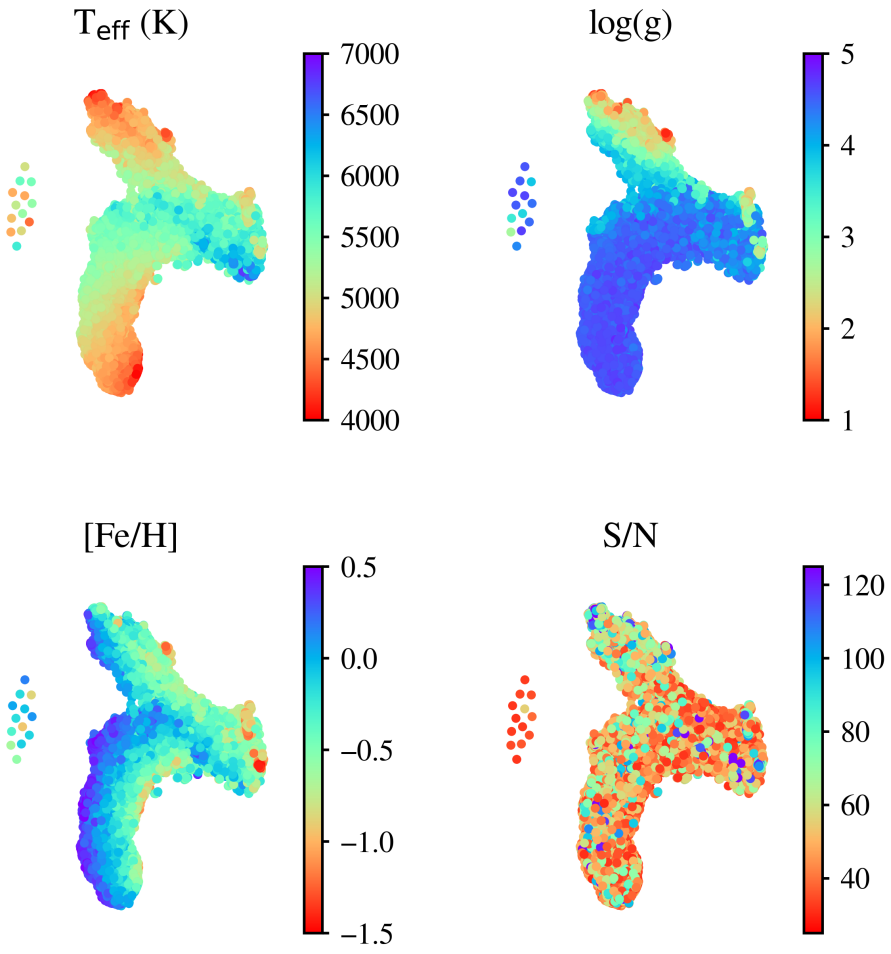


Figure 17: A t-SNE similarity map of our sample GIRAFFE spectra. The four panels show the same map, each color-coded with either a physical parameter or the spectrum S/N. While the relative distance of points in the map indicate the degree of similarity of the corresponding spectra, their X and Y coordinates themselves have no physical meaning. The map in this figure has been computed with perplexity 30. For our data, perplexity values between 20 and 50 produce qualitatively identical results.

different sets of spectra. However, in all our t-SNE runs, the outlier spectra were clearly identifiable.

Every training spectrum has a set of associated stellar labels. In our case these are the two atmospheric parameters $T_{\rm eff}$ and $\log(g)$ and the chemical abundances [Mg/Fe], [Al/Fe], and [Fe/H]. In the GES iDR6 data set the elemental abundances are given as absolute abundance values A(X), where X is any given element. We calculated [Fe/H] and [X/Fe] as follows: $[{\rm Fe/H}] = {\rm A(Fe)_{star}} - {\rm A(Fe)_{\odot}}$ and $[{\rm X/Fe}] = {\rm A(X)_{star}} - {\rm A(X)_{\odot}} - [{\rm Fe/H}]$. The absolute solar abundances were taken from Grevesse et al. (2007) to be consistent with the GES spectral analysis strategy. The decision to use these relative abundances instead of absolute abundances for the training of our network is justified in Sect. 4.3.

Magnesium and aluminum abundances are known to be sensitive to non-local thermodynamic equilibrium (NLTE) effects (Bergemann et al. 2017; Amarsi et al. 2020; Lind et al. 2022). These effects were not considered by GES during the parametrization of GIRAFFE spectra or during the homogenization (Hourihane et al., in prep). For dwarfs, NLTE corrections are well below 0.05 dex for both Al and Mg, while for giants they are in the range of $\sim 0.05 - 0.15$ (Amarsi et al. 2020). Strong NLTE effects may then have some effects on the training labels, but quantifying such an effect in out of the scope of the present thesis.

After applying all these constraints, we were left with 14 634 combined spectra with associated high-quality atmospheric parameters and elemental abundances. The training labels and the spectra themselves

Table 3: Values of the five training labels for the first few stars in our training set. The star CNAMES and indices are for identification purposes only and are not passed to the network for the training. The full training label input table has five columns and 14634 rows.

CNAME	Index	$T_{\rm eff}$ (K)	log(g)	[Mg/Fe]	[Al/Fe]	[Fe/H]
16153746-0822162	1	5780	4.35	0.05	0.06	0.17
12202074+0318445	2	4335	1.69	0.39	0.24	-0.32
12320544-4332104	3	4753	1.69	0.40	0.45	-0.41
20041963-3203228	4	4334	1.83	0.40	0.49	-0.51
21300375-1230399	5	4376	4.34	0.16	0.22	-0.16

Table 4: Spectrum fluxes of the five stars in the training set listed in Tab. 3. The indices are for identification only and are not used during the training. The full spectra input table for the training has 8669 columns and 14634 rows.

	Flux in wavelength bin								
Index	1	2	3	4	...	8666	8667	8668	8669
1	0.883	0.883	0.871	0.867	...	1.017	1.009	1.010	1.010
2	0.962	0.961	0.958	0.953	...	1.016	0.998	0.987	0.979
3	1.072	1.092	0.992	0.992	...	0.979	0.974	0.979	0.990
4	0.993	1.004	1.008	1.008	...	0.995	0.999	1.002	1.003
5	0.856	0.851	0.838	0.824	...	1.060	1.000	1.001	1.001

are passed as a data-tables to the python code (Table 3). As mentioned in Sect. 3.1.1, the absolute wavelength positions of the individual data points in the spectra are not important for the CNN training. Therefore, the ∼2900 Å gap between the GIRAFFE HR10 and HR21 setups does not influence the training process either. Table 4 shows the format of the training spectra input for the CNN. These 14 634 spectra were assigned to either the training set or the validation set for the training of our CNN. We found that using 40% of the total training data for the validation set yields the best training results for our application. That means that of our 14 696 spectra, 8817 spectra were assigned to the training set and 5879 to the validation set. Training and validation spectra were chosen at random before the training and we ensured that their labels cover the same parameter space.

4.1.2 Parameter space of training and validation set labels

To assess the parameter space of our training set input labels, we show the Kiel-diagram and abundance plots in Figs. 19 and 20. Effective temperatures range from $T_{\rm eff}$ = 4000 - 6987 K, the surface gravity $\log(g)$ is between 1.08 and 4.87 dex and [Fe/H] spans a range of ∼2 dex, from −1.53 to 0.72 dex. The color-coding in Fig. 19 reveals the metallicity sequence in the giant-branch of the Kiel-diagram.

Figure 20 shows density maps of the [Mg/Fe] and [Al/Fe] distribution of our training set. The [Mg/Fe] values range from −0.25 to 0.80 dex, [Al/Fe] values have a large spread of almost 2 dex, from −0.95 to 1.00 dex. The Mg distribution reveals two distinct regions of

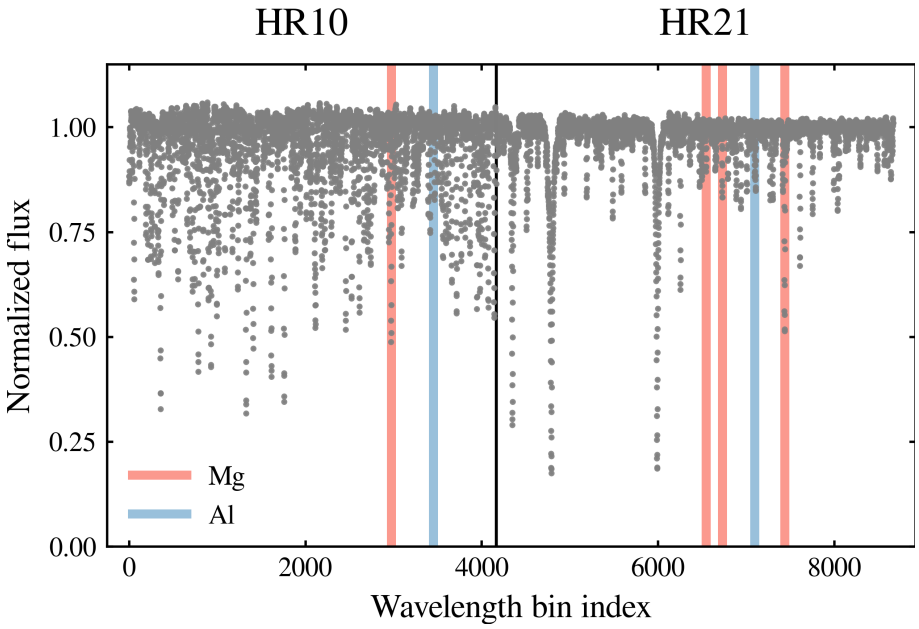


Figure 18: Input spectrum of the first star in our training set (CNAME 16153746-0822162). The spectrum is a combination of the individual GIRAFFE HR10 and HR21 spectra of this star. The locations of Mg and Al absorption features in the spectrum are marked by colored vertical lines.

enhanced density, separated by a narrow region of lower density. These two regions reflect the separation of Milky Way stars into a thin disk (low [Mg/Fe]) and a thick disk (enhanced [Mg/Fe]) population. Magnesium abundances are the best probe for this chemical separation between the thin and thick disk of our Galaxy (e.g. Fuhrmann (1998), Gratton et al. 2000). As expected, we do not see this separation in the [Al/Fe] plot. Our training set is dominated by nearby stars, due to the S/N cut and other quality criteria that we applied to the entire GES iDR6 data set. Therefore, our data does not cover some Milky Way properties that become apparent when one investigates a larger volume of our galaxy. Queiroz et al. (2020), for example, find two detached [Al/Fe] sequences for stars close to the galactic center ($R_{Gal} < 2$ kpc) in their sample of APOGEE stars. At low [Fe/H] several groups of stars can be observed in both the Mg and Al plots. The stars in these patches belong to different globular clusters. In the [Al/Fe] plot, the scatter of Al abundances in the globular clusters is higher than the scatter of Mg at equal metallicities. This large spread of Al abundances, especially in globular clusters at low metallicities, has already been observed in earlier GES releases (Fig. 4 in Pancino et al. 2017c) and indicates the existence of multiple stellar populations within the clusters.

4.1.3 "Inner" and "outer" test sets

In addition to the training and validation sets, we constructed a test set. This set is used to test the performance of our CNN on spectra that were not used in the training process. In this way, we can mimic the application of our CNN method to newly observed spectra, which have not yet been analyzed spectroscopically. The full test set therefore contains spectra without any applied quality constraints and spans wider S/N and label ranges than the training set. As we will show in Sect. 4.2.3, our network is not able to accurately label spectra that are outside the training set limits. Because of this, we have to find a way to identify spectra that are similar to our training spectra, as the labels of these spectra are likely to fall into the training set limits. We already demonstrated that t-SNE can show the similarity between spectra. We therefore employ t-SNE to identify those spectra in the test set whose labels are likely to be within the training set limits. A depiction of this method can be seen in

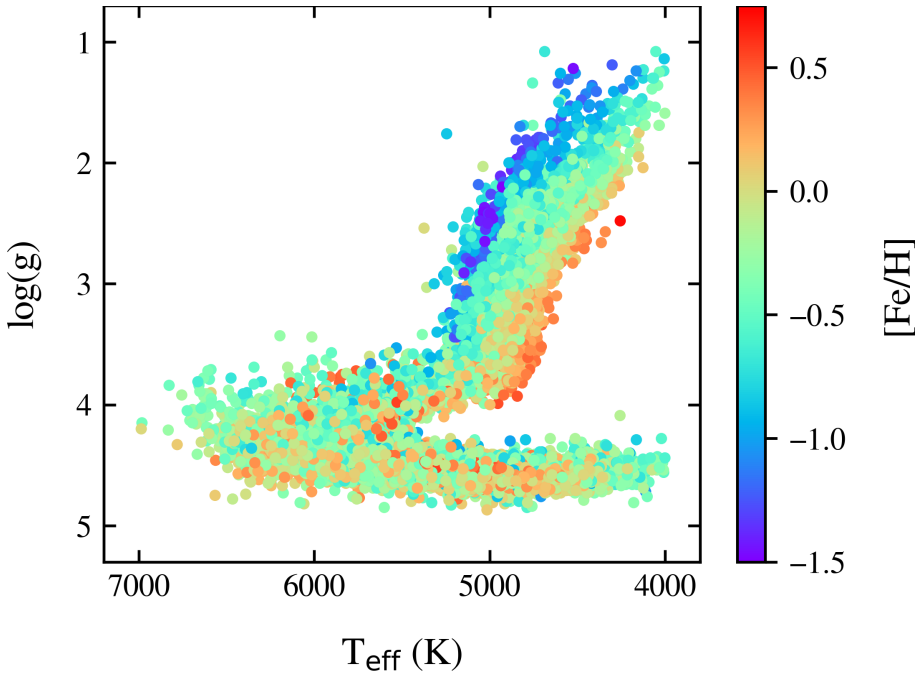


Figure 19: Kiel diagram containing the 14 634 stars that will be used to train and test our neural network. The color-coding indicates the metallicity gradient in the giant branch stars.

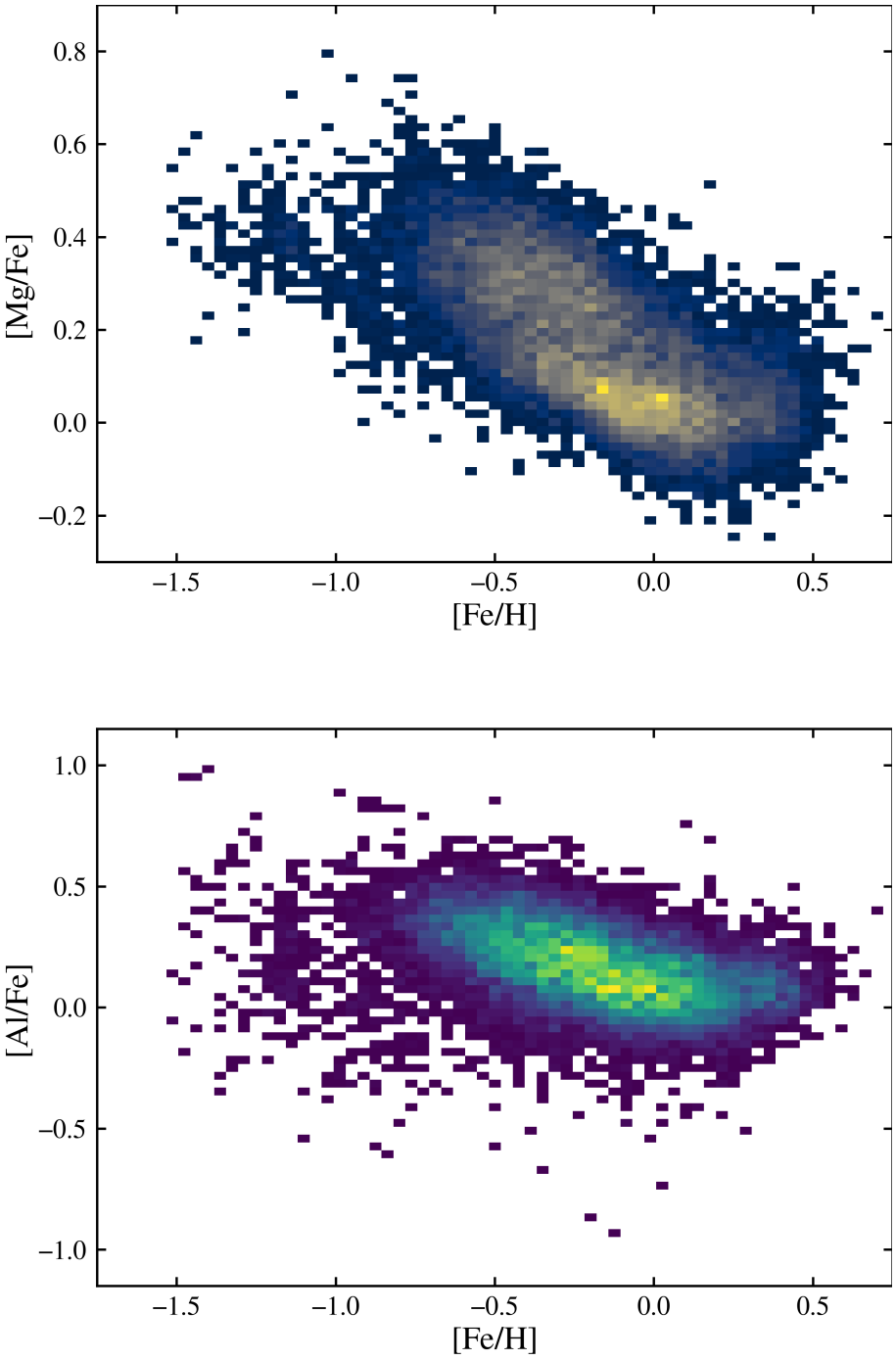


Figure 20: Density plots of [Mg/Fe] vs. [Fe/H] (top panel) and [Al/Fe] vs. [Fe/H] (bottom panel) for the 14 696 stars in the training and test sets. Brighter colors indicate a higher density of data points (linear color scale).

Fig. 21. In the top left panel, we show a t-SNE map that was calculated for all spectra in our data set. After constructing the t-SNE projection, we identified the training spectra in the map. The top right panel shows those spectra in the test set that cover the same area in the t-SNE map as the training spectra in the left panel. This was done by limiting the distance between training set data points and test set data points in the t-SNE map. The test spectra that are close to the training spectra in the map are similar to the training spectra. We call this set in the top right panel our "inner" test set. Finally, the bottom panel shows those spectra that are not similar to the training spectra, and their distance from the training spectra in the map exceeds our chosen maximum distance. This set of spectra is our "outer" test set, and we use it to test our network on spectra which are unlike the spectra in the training set. n Figs. 22 and 23 we show sample spectra from the training set together with spectra from the inner and outer test sets. All spectra from the areas *1*, *2*, and *3* of the t-SNE map are similar to each other. These spectra belong to the training and inner test sets. Based on this spectral similarity, we assume that the labels of these spectra are also similar to each other. The labels for the test set have already been determined by GES. Therefore, we can indeed confirm that the labels of all spectra in the top panels of Fig. 23 are within the training set limits. The spectra *a*, *b*, *c*, and *d* from the outer test are quite different from the spectra in the inner set. This difference is caused by the different physical conditions in the atmospheres of these stars. Spectrum *a* for example has a higher Fe abundance than any star in the training set, the surface gravity $\log(g)$ of spectrum *b* and the effective temperature of spectrum *c* exceed the training set limits, and the Fe abundance of spectrum *d* is by 0.2 dex lower than the lowest Fe value in the training set. With this t-SNE approach, we can therefore identify spectra that are likely to have labels that are outside the training set limits. These spectra can then be excluded from the further analysis, or their CNN label predictions can be flagged as uncertain. With the given GES labels of all our sample spectra, we can quantify the efficiency of this method: About 20% of all test spectra have GES labels outside the training set limits. The situation improves for our inner test set, with 12% of its spectra labels falling outside the training limits. When we additionally require the S/N of the inner set

to be ≥ 30 (to match the S/N range of the training set), less than 10% of these high S/N inner set spectra have labels exceeding the training limits. In a situation where the labels of an observed spectrum are not yet known, we therefore recommend to check if the spectrum is similar to the training set spectra.

The t-SNE method described above was developed in Ambrosch et al. (2023), and has since been successfully applied to a large set of *Gaia* spectra by Guiglion et al. (2024). In their study, the method was used to classify 841300 *Gaia* RVS spectra into *training-like* and *training-unlike* spectra. A table with CNN predictions for all *Gaia* RVS spectra has been published, with flags that indicate if a spectrum was similar to their training set or not.

Our full GIRAFFE test set contains 22 270 spectra, with a minimum S/N of 10 and including spectra with different shapes than those in the training set. The outer test set contains 3877 spectra. The inner test set then consists of the remaining 18 393 spectra which cover the same area in the t-SNE map as the training spectra. Of the inner test spectra, 4916 have S/N ≥ 30.

4.2 Training and testing of our CNN

4.2.1 Training of our network

At the start of the network training phase, the kernel values in the convolution layers, the weights for the dense layer neurons, and the bias values are initialized randomly. Also random is the assignment of training spectra to the batches for the mini-batch training. Due to this, the final parameter values of the trained network model, and therefore the network output, vary slightly every time the network is trained. To account for this variation, we performed ten training runs and recorded the results from each. We removed the two CNN models with the largest remaining validation losses at the end of their training phase. The remaining eight CNN models were then used to predict the labels for the spectra in the training, validation, and test sets. We report here the averages of the eight sets of labels as our results.

The training and validation sets remained unchanged for each of the ten training runs. We checked the label distributions for both the train-

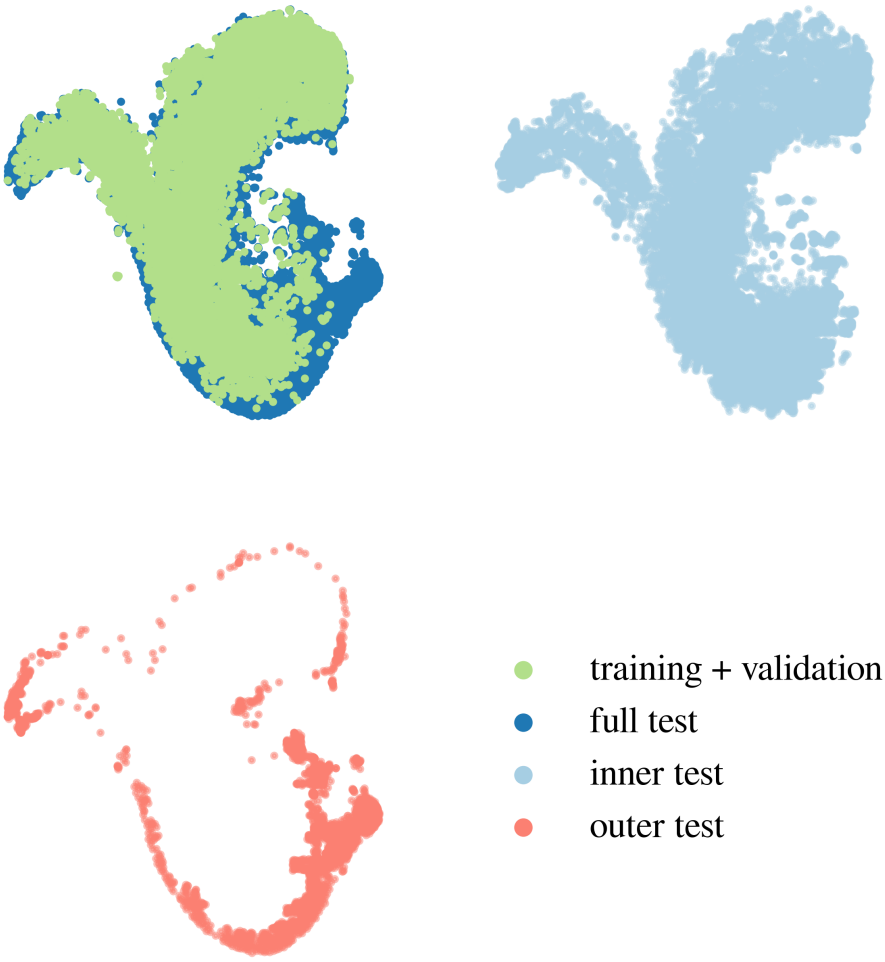


Figure 21: Top left panel: t-SNE map of all spectra in our GIRAFFE data set (dark blue), containing the training set spectra (green). Extreme outlier spectra have been removed. Top right panel: "Inner" test set, defined as the subset of test spectra that cover the same area in the t-SNE map as the training set spectra. Bottom left panel: Spectra in the test set that do not cover the same area in the map as the training set. This is our "outer" test set.

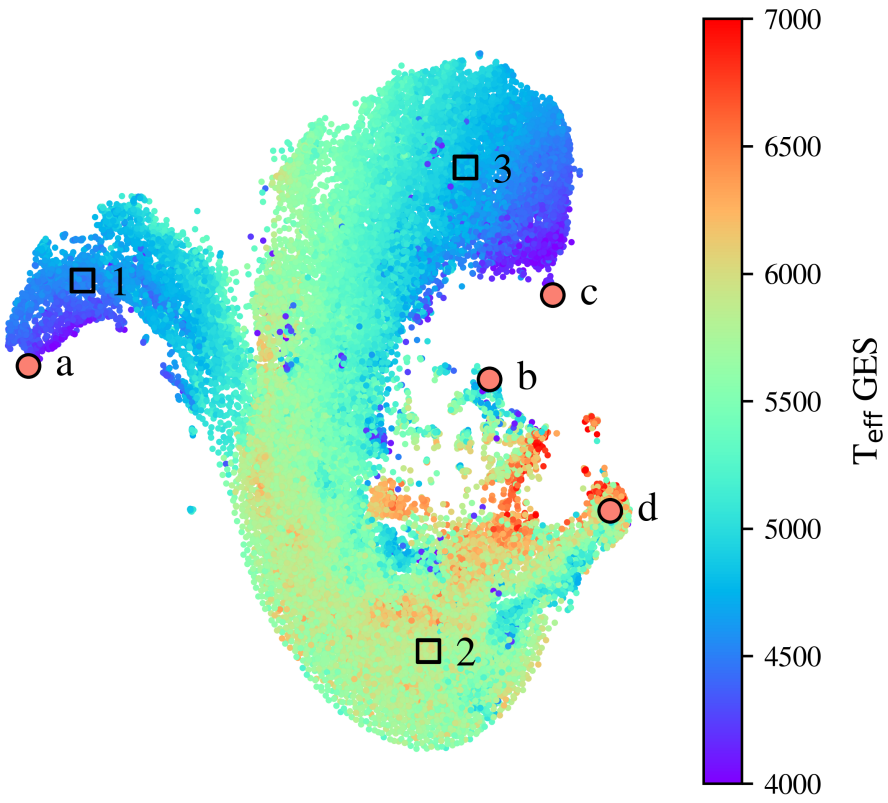


Figure 22: t-SNE map of all spectra in our data set. Colors indicate the GES label T_{eff} for each spectrum. The three black boxes mark areas in the map, that contain training, validation, and "inner" test spectra (see Fig. 21). The orange data points (*a*, *b*, *c*, *d*) mark spectra in the "outer" test set. We show these, together with spectra from the boxes *1*, *2*, and *3*, in Fig. 23.

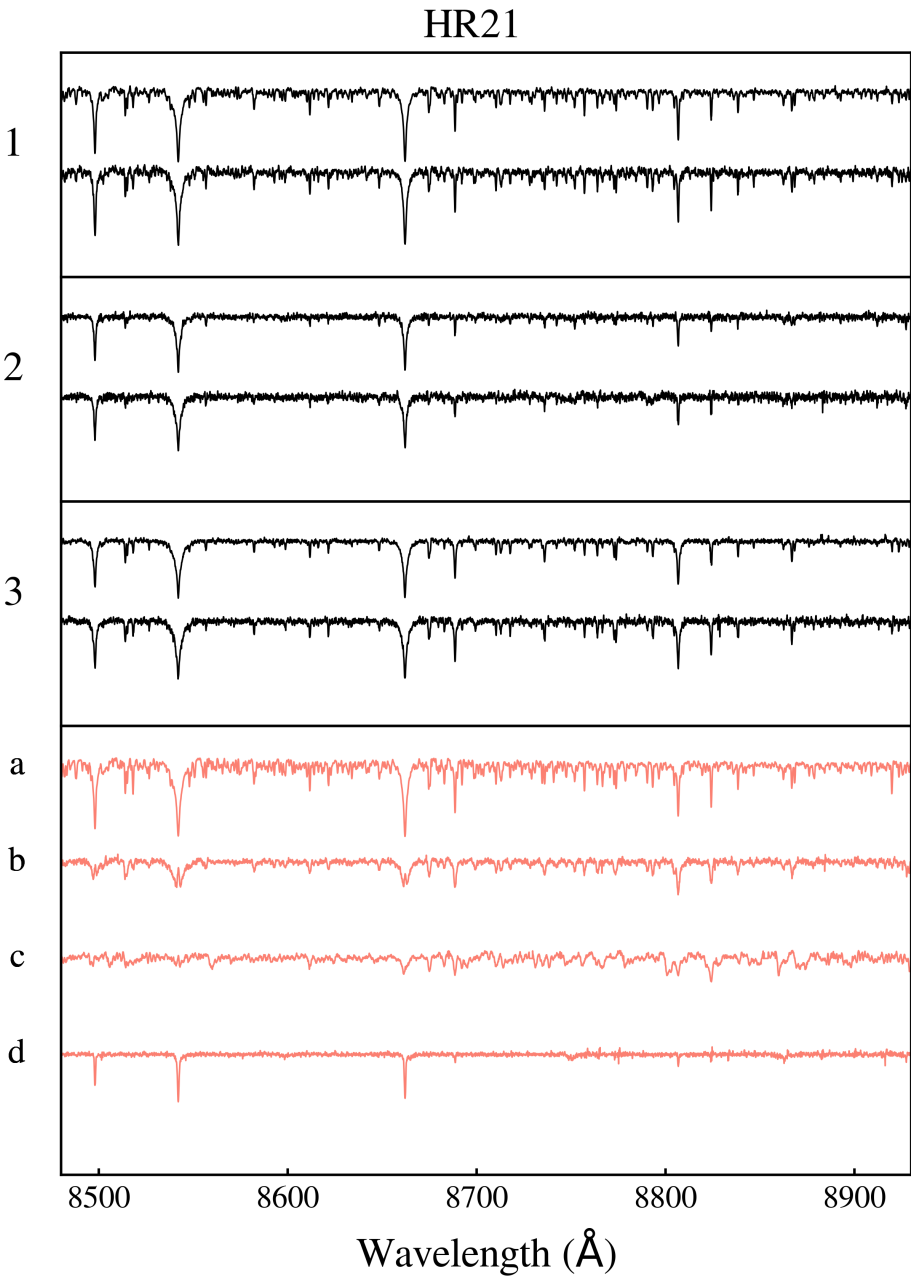


Figure 23: Sample spectra which are marked in the t-SNE map in Fig. 22. The three top panels contain spectra from the areas *1*, *2*, and *3*. In each of these panels, the top spectrum is a training spectrum, the bottom spectrum belongs to the "inner" test set. The bottom panel shows the spectra *a*, *b*, *c*, and *d* from the "outer" test set. For display, the spectra are shifted vertically, but the flux scale remains the same for all spectra.

ing and validation set and found that both span the same label ranges and are equally distributed. They also span the same area in the t-SNE map in Fig. 21. We therefore do not expect that keeping the training and validation sets constant will add any large uncertainties or training biases.

In every training epoch, the loss between the GES input labels and the CNN label predictions decreases (Fig. 24). In Fig. 25 we show the Kiel diagram and the [Mg/Fe] and [Al/Fe] distributions which our CNN predicts at different stages of the training phase. In early epochs, only the basic shapes of the parameter distributions are reproduced. As the training progresses, the finer details, which are visible in the GES input data, emerge in the CNN predictions. We stopped the training runs when the loss of the validation set did not decrease anymore in the last 30 training epochs.

On average, one training run lasted for 159 epochs and took $\sim$45 minutes to complete[3]. The label prediction was very fast: the parameterization of the $\sim$37 000 spectra in our data set took less than 20 seconds per CNN model.

4.2.2 Training results

In Fig. 26 we show the direct comparisons between the input GES measurements and the CNN predictions for the training and validation sets. There is a good agreement between the GES measurements and CNN predictions across all labels and for the two sets. Both the CNN predictions for the training set and validation set show the same offset (if any) and small dispersion around the 1:1 relation. This shows that the network performs well on spectra which it was not directly trained on and does not overfit. The dispersion around the 1:1 relation is uniform across most of the value ranges of all five labels. We use the dispersion of the training set as a measure for the training precision of our network: The training precision is 37 K for $T_{\rm eff}$, 0.06 dex for $\log(g)$, 0.05 dex for [Mg/Fe], 0.06 dex for [Al/Fe], and 0.04 dex for [Fe/H]. However, our CNN does not accurately reproduce the highest and lowest GES measurements. This is especially clear in the case of [Al/Fe],

[3]On a desktop PC, using only CPU (Intel Core i7-9700 CPU @ 3.00 GHz $\times$ 8)

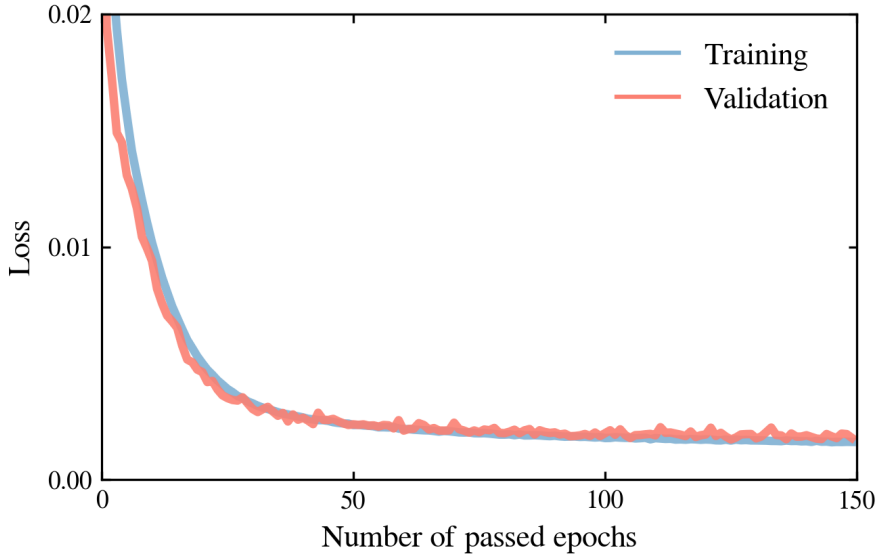


Figure 24: Evolution of the training and test losses during the network training phase. The loss of the test set is closely following the training loss. The small difference between training and test sets at the end of the training phase shows that the network is not over-fitting.

where the CNN predictions overestimate the lowest [Al/Fe] measurements by $\sim$0.5 dex, while the highest values are underestimated by the same amount. We explain this behavior by noting that only a few spectra with these extreme measurements were available for the network training. The CNN therefore predicts more moderate labels for these spectra.

4.2.3 Performance on the test sets

To evaluate the ability of our network to parameterize new spectra, that have not been involved in the training process at all, we compare the GES input labels to the CNN predictions for three different test sets. The definitions of our inner and outer test sets are given in Sect. 4.1.3. The top row of Fig. 3 shows the GES input to CNN output comparison for the inner test set spectra with S/N $\geq$ 30. In this subset, 90% of the GES labels lie within the training limits. Most of the remaining 10% of spectra are outside the [Mg/Fe] and [Al/Fe] limits. The reason for this

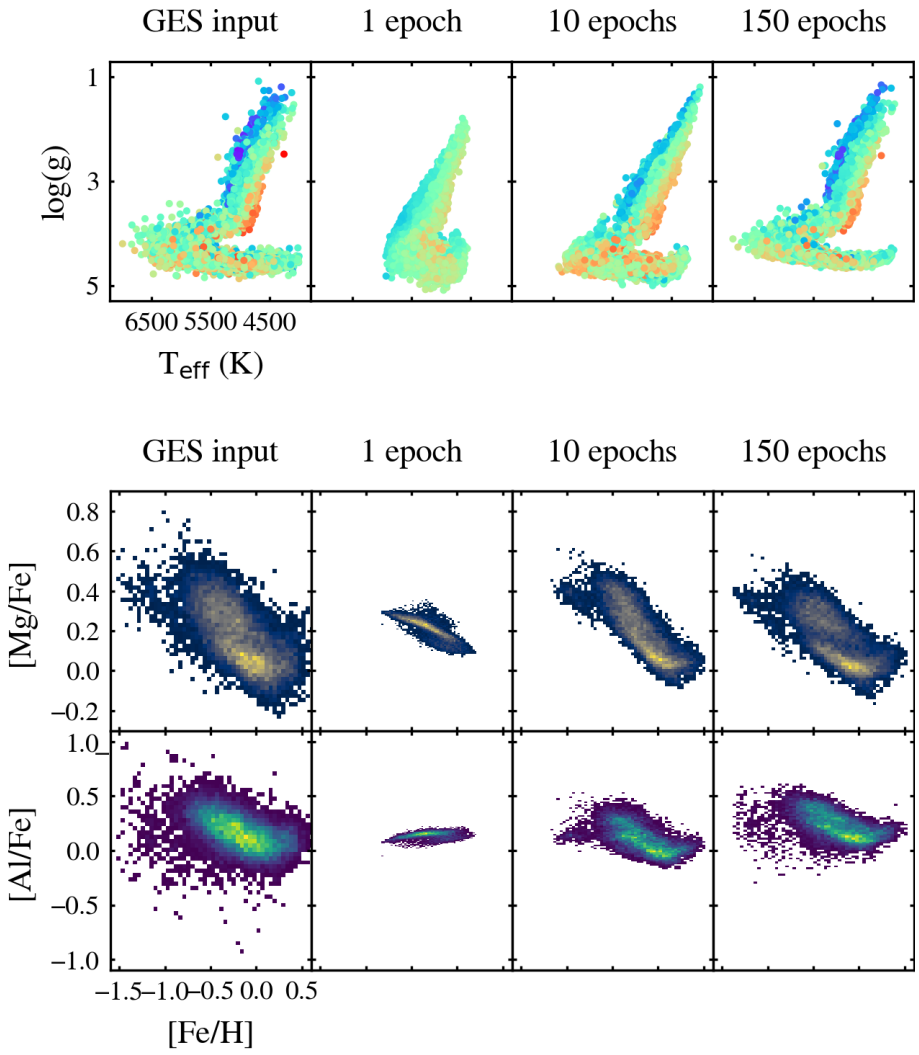


Figure 25: Progress of the network training. Left column: Kiel diagram, [Mg/Fe], and [Al/Fe] distributions, based on the GES input labels for our training and validation sets. Other columns: CNN predictions for the different label distributions after 1, 10, and 150 training epochs. The color-coding of the Kiel diagrams indicates the [Fe/H] values of each data point and is on the same scale as in Fig. 19. Brighter colors in the [Mg/Fe] and [Al/Fe] density plots indicate a higher density of data points (linear color scale).

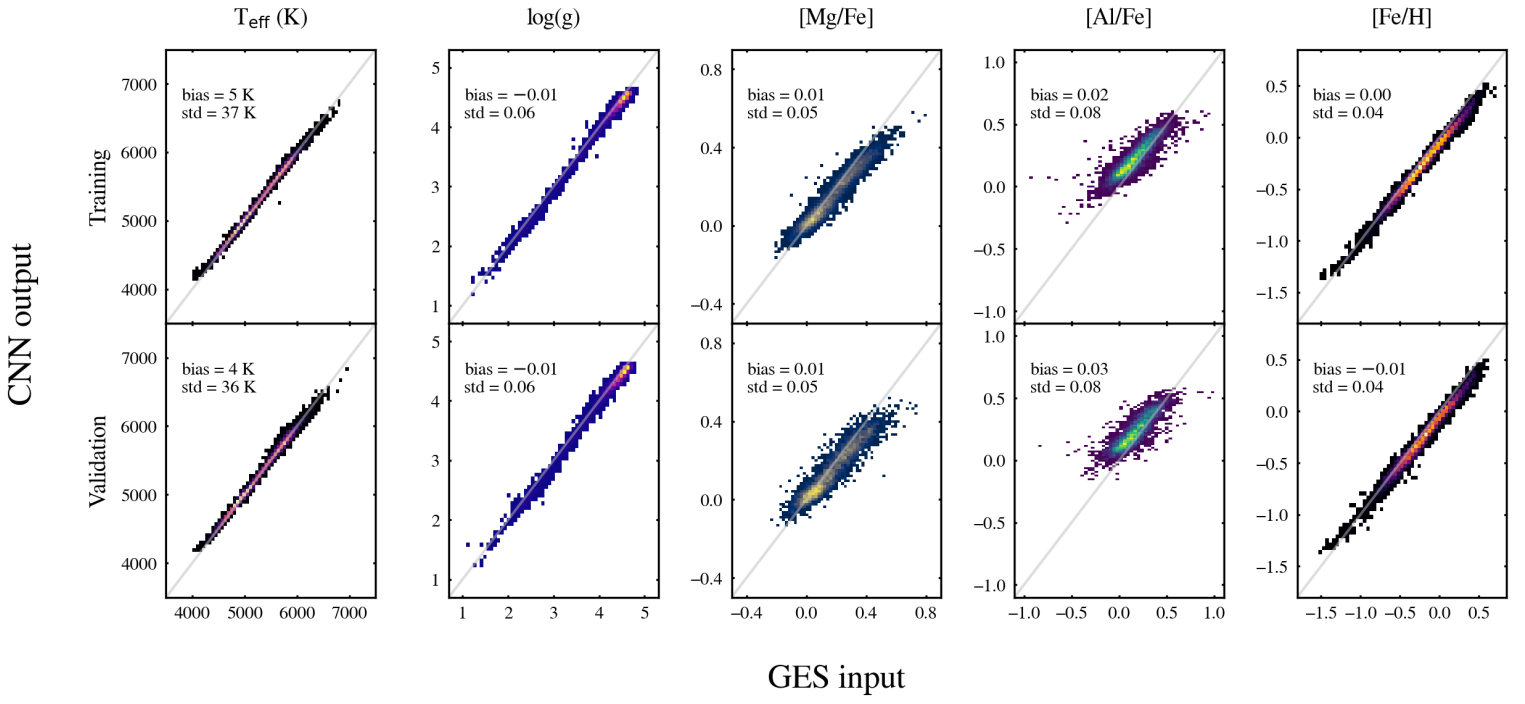


Figure 26: One-to-one comparison of labels from GES iDR6 and the values predicted by our CNN. The two rows show results for the training and validation sets. Each column contains the results for a different label. In every panel, the horizontal axis stands for the GES input labels, the vertical axis represents the labels predicted by our CNN. Brighter colors indicate a higher density of data points (linear color scale). The average bias and the standard deviation (scatter) of the results around the 1:1 relation is given in every panel. Solid diagonal lines show the 1:1 relation.

lies in the way we find our inner test set. This set contains only stars that occupy the same area in a t-SNE map as the training spectra (Fig. 21). The shape of a spectrum, and therefore its position in the t-SNE map, depends strongly on the labels $T_{\rm eff}$, $\log(g)$, and [Fe/H], while changes in [Al/Fe] only have a small effect on the overall spectrum. The same is true for [Mg/Fe], but to a lesser extent. This is because there are more Mg absorption lines than Al lines in our sample spectra (Heiter et al. 2021b). The accuracy of the network predictions starts to degrade with the low S/N inner test set (middle row of Fig. 3). This set contains spectra that are similar to the training spectra but have lower S/N (we remind that the minimum S/N of the training spectra is 30). The low S/N inner set contains more spectra that are outside the training limits. It is clear that our CNN cannot accurately parameterize spectra whose GES labels lie outside the training set range. The bottom row shows the results for the outer test set. Network predictions for this set are increasingly inaccurate, even for spectra inside the training set limits. The difference between GES input and CNN output for the outer test set is most prominent in [Al/Fe] and [Fe/H], where extremely low and high GES labels are not accurately predicted by our network.

The comparison of the CNN predictions for the different test sets highlight the importance of pre-selecting spectra that are likely to fall within the training set limits. Network predictions for spectra that are dissimilar to the training spectra or have lower S/N are likely to be inaccurate.

4.2.4 Estimation of internal precision

As described, the label predictions from our eight trained CNN models vary slightly. This variation can be used to estimate the internal precision of our methodology. We define the internal uncertainty of our results as the dispersion between the label predictions from the eight CNN models. In Fig. 28 we display the distribution of the internal uncertainty of our five labels relative to the predicted label values and to the spectra S/N. This analysis is done for both the inner test set with S/N$\geq$ 30, the inner test set with S/N < 30, and for the outer test set. The boxplots show the spread and median of the uncertainties in S/N bins of size 10, for the entire test set. Where S/N $\geq$ 30, the uncertainties are

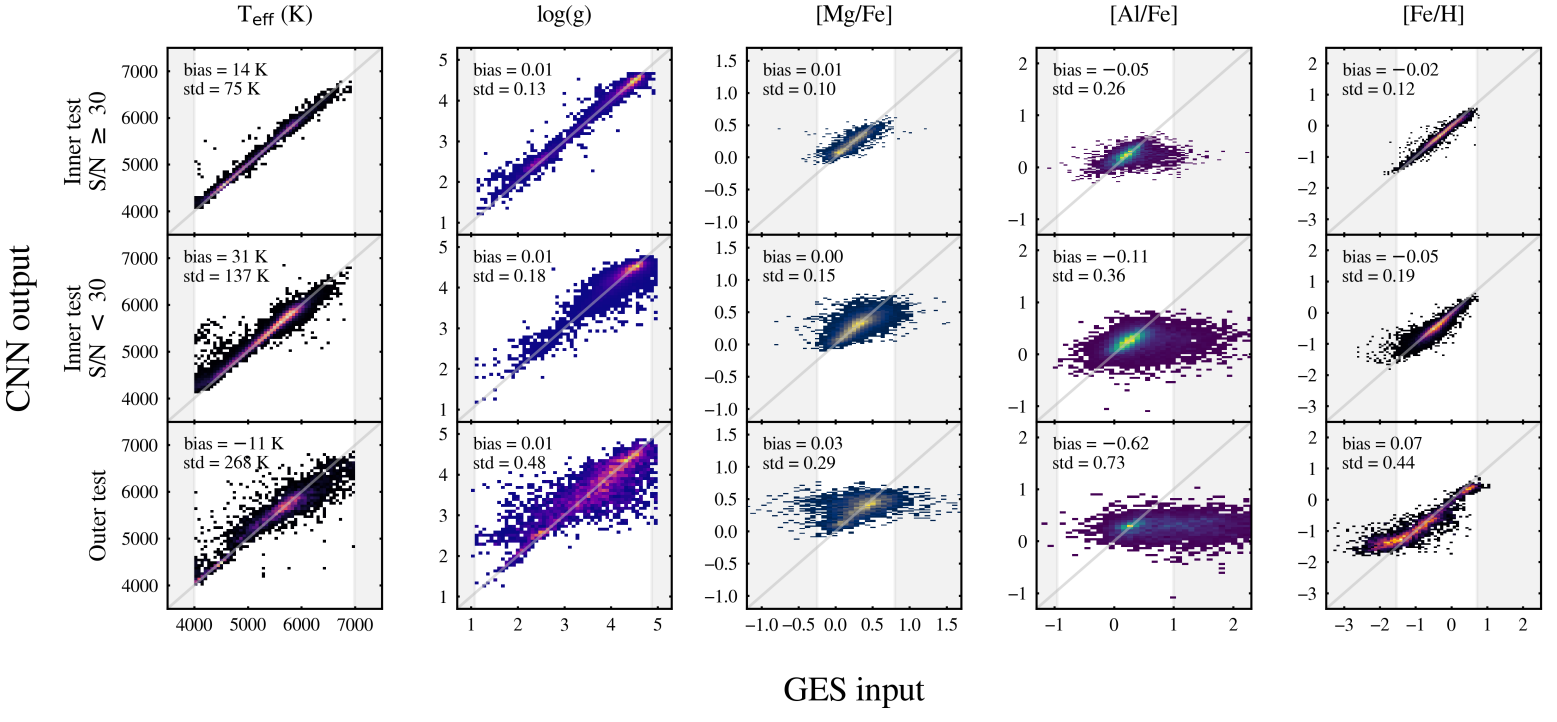


Figure 27: One-to-one comparison of labels from GES iDR6 and the values predicted by our CNN. The three rows show results for the inner test set in two different S/N ranges, and for the outer test set. Each column contains the results for a different label. In every panel, the horizontal axis stands for the GES input labels, the vertical axis represents the labels predicted by our CNN. Brighter colors indicate a higher density of data points (linear color scale). The average bias and the standard deviation (scatter) of the results around the 1:1 relation are given in every panel. Solid diagonal lines indicate the 1:1 relation. Shaded areas indicate regions that are outside the training label space.

small, with near constant median and spread across all bins. Towards lower S/N, both the median uncertainty and the spread in the bins increase. The mean uncertainties of the label predictions for the high S/N inner test set are small: 27 K for $T_{\rm eff}$, 0.04 for $\log(g)$, and 0.03 dex for [Mg/Fe], [Al/Fe], and [Fe/H] alike. The GES label errors for this set show little to no dependence on the absolute label value and S/N. Their mean values are 63 K for $T_{\rm eff}$, 0.15 for $\log(g)$, 0.22 dex for [Mg/Fe], 0.19 dex for [Al/Fe], and 0.18 dex for [Fe/H].

The CNN predictions with large uncertainties for one label also show large uncertainties for all other labels, while precise predictions are precise across all five labels. The internal precision of our $T_{\rm eff}$ and $\log(g)$ is highest where the training set density is highest. For $T_{\rm eff}$ this is the case between $\sim$ 4500 and 5775 K, for $\log(g)$ at $\sim$ 2.5 and 4.5 dex. Here, the uncertainties of the predictions for these two labels is lowest. Except for [Fe/H], the precision of the abundance predictions show no clear trends with the absolute label value. For [Fe/H], the uncertainty increases with lower [Fe/H] abundances. This is presumably due to the smaller number of stars in the metal-poor regime compared to the main bulk of the sample. Also, our CNN struggles to provide precise predictions due to the weak spectral features present in this [Fe/H] regime. We find that the uncertainties of the predictions for all five labels increase as the S/N of the spectra decreases.

The mean prediction uncertainties for the low S/N inner set and for the outer set are higher than for the high S/N test set. Precision for these sets also show strong trends with the absolute label value, especially for $T_{\rm eff}$ and [Fe/H].

We also tested how the uncertainty distributions change when we change the composition of the training and test sets for every training run. The resulting label uncertainties are similar to the uncertainties from our original approach. We leave the detailed investigation of the effect of varying train and test sets for future work.

4.2.5 Influence of radial velocity errors

As described in Sect. 4.1.1, we use the radial velocity of the stars for the redshift correction of the input spectra. Within GES, the radial velocities were measured by comparing the redshifted observed spectra

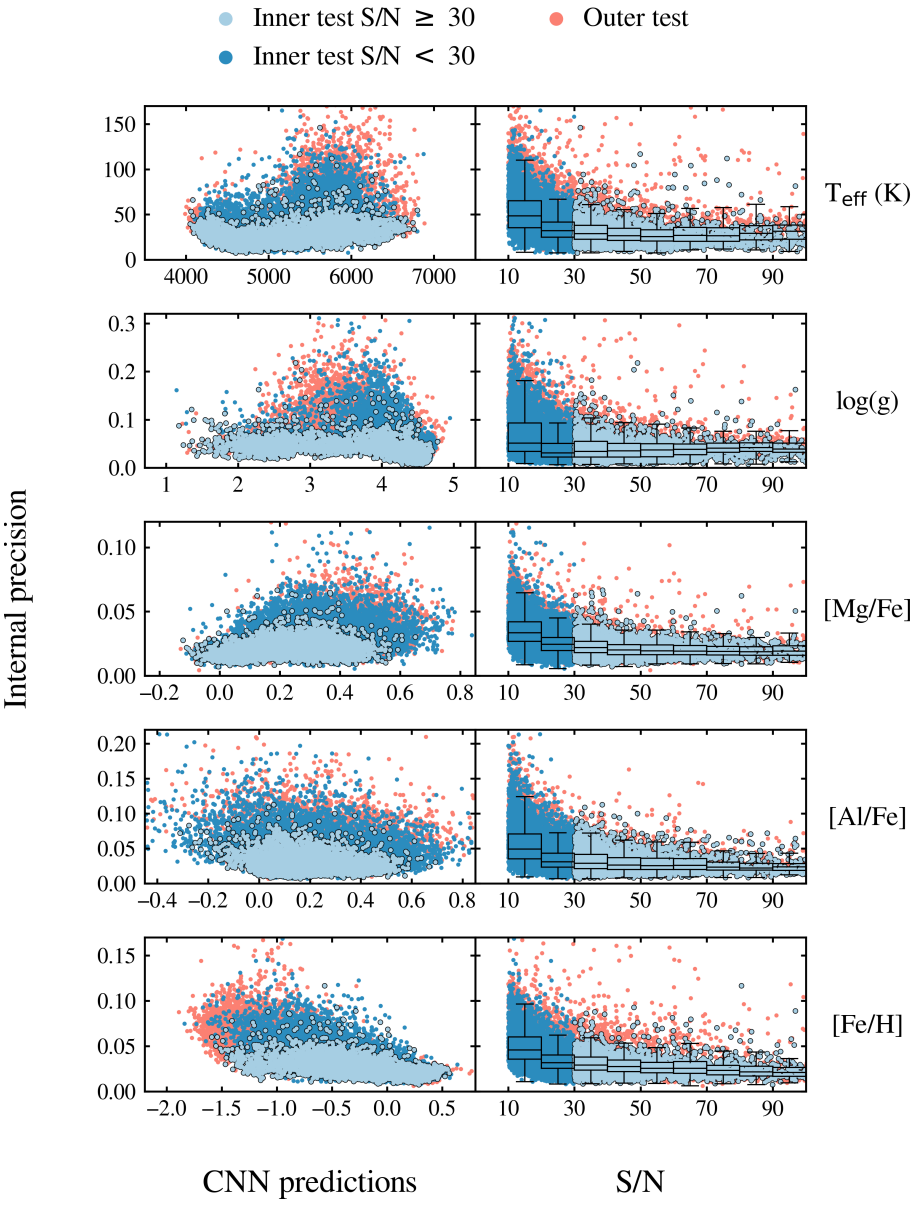


Figure 28: Internal precision of our CNN results for the inner test set with S/N $\geq$ 30 (light blue), with S/N < 30 (dark blue) and the outer test set (red). The left panels show the internal precision relative to the absolute CNN label values, the right panels show the precision relative to the spectra S/N. The box plots in the right panels show the median and spread of the uncertainties of the whole test set, in S/N bins of width 10, starting at S/N = 10.

to unshifted synthetic spectra (Gilmore et al. 2022). The velocities are measured from each GIRAFFE setup and then homogenized to provide one value for every star. We tested how uncertainties in the provided radial velocities for the HR15N spectra influence the accuracy of the CNN predictions of the labels $T_{\rm eff}$, log(g), [Fe/H], and A(Li). In Fig. 29, we show the difference Δ between the CNN label predictions for the HR15N test set and the GES input labels for these spectra. The data points in this figure are color-coded by the provided error of the radial velocity (E_Vrad) of these spectra. We see that the dispersion of Δ increases for all four labels, when the E_Vrad increases. The bias (average difference Δ) also increases with the uncertainty. This is most pronounced for the labels $T_{\rm eff}$ and [Fe/H]. Most of our stars with E_Vrad > 0.5 km s^{-1} are hotter than 5200 K. Measuring the radial velocity from GIRAFFE spectra becomes increasingly difficult for stars hotter than $\sim$5500 K, because of a lack of strong absorption lines in hot stars (Jackson et al. 2015).

We demonstrated that minor changes in input data wavelengths can significantly impact machine learning pipelines. For forthcoming surveys like 4MOST and WEAVE, accurate radial velocity estimation becomes crucial, given the extensive adoption of machine learning techniques due to increased observation volumes.

4.3 What does the network learn?

The purpose of the convolution layers in our CNN is to find spectral features. These spectral features are then interpreted into the labels by the dense layers. This approach is also used by classical spectral classification methods, where individual spectral features are investigated to derive the stellar parameters. However, since machine–learning is purely data-driven, the predictions of our CNN could merely be the result of our network learning correlations between labels in the training data. Inferring stellar parameters from correlations like these can lead to satisfying results for some spectra. Stars with exotic chemical compositions (for example stars with a non-solar mixture of elements, such as old thick disk stars) do not follow such trends and will not be parameterized well. This effect can be demonstrated by attempting to infer Li abundances from spectra that do not contain any Li absorption features.

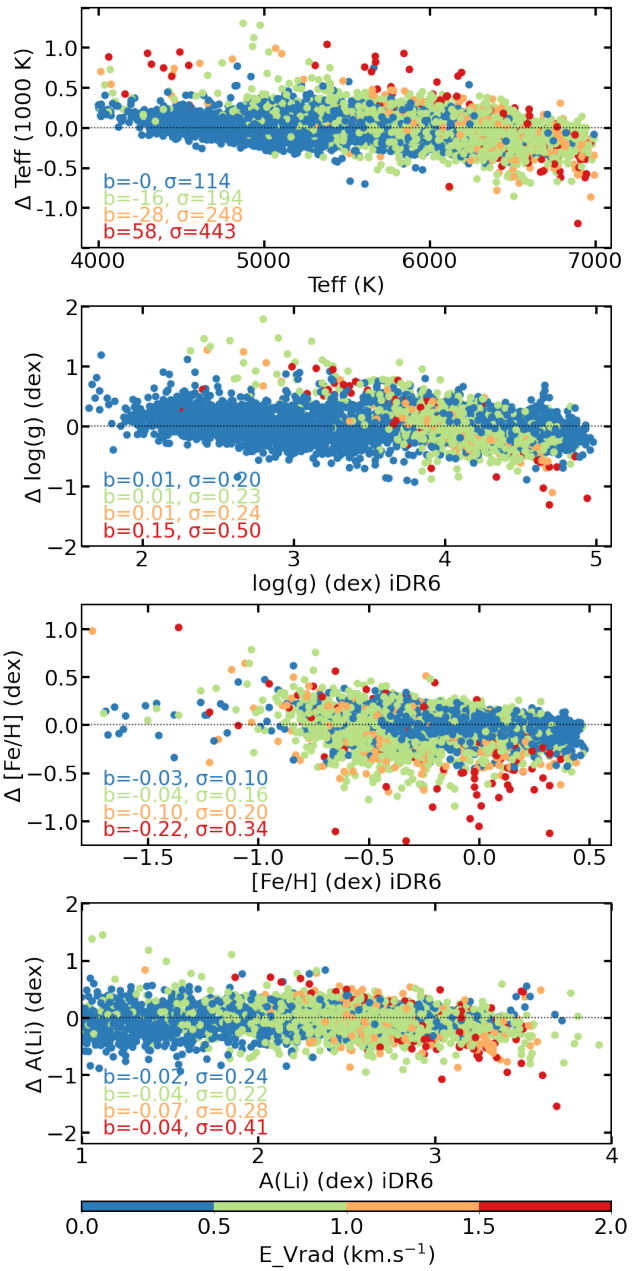


Figure 29: Difference Δ between CNN and GES labels as a function of GES labels for the HR12N test set. The stars are color-coded according to their radial velocity uncertainties. For each label, the bias = mean(CNN−GES) and σ = std(CNN−GES) in the four uncertainty bins are given. Figure from Nepal et al. (2023).

To this end, we trained a CNN on GIRAFFE HR15 spectra, where the Li line at 6707.8 Å has been masked out, with the training labels $T_{\rm eff}$, log(g), [Fe/H], and A(Li). Due to the lack of Li features in these modified training spectra, the network can only learn the astrophysical correlation between the input labels $T_{\rm eff}$ and A(Li). The A(Li) predictions of the so-trained CNN follow the general 1:1 relation when compared with the input A(Li) from GES iDR6 (Fig. 30). However, in addition to the large scatter of $\sim$0.5 dex, this version of the network fails to predict the high Li abundances of several stars. These stars are Li-rich giants, whose Li abundance does not follow the common Li-$T_{\rm eff}$ correlation. Without knowledge about the Li absorption line, the network cannot predict accurate Li abundances for these exotic objects. Individual elemental abundances can also be correlated. Magnesium and aluminum abundances for example are correlated with the iron abundance: Stars with low iron generally show low abundances of Mg and Li as well.

We therefore want to show that our CNN is indeed able to identify spectral lines and to associate them with the right labels.

To visualize where our CNN is active for a given label, we computed sensitivity maps using GradientTape from TensorFlow. In general, our CNN behaves like a function that maps input pixels in stellar spectra to the output label values. This mapping is differentiable, which means that the influence of every input pixel for the resulting label predictions can be quantified. Using automatic differentiation, one can compute the gradients, ∂ Label / ∂ λ, i.e., the sensitivity of the CNN to each pixel for every label. A large absolute gradient value at a wavelength bin then means that the network prediction for a given label is very sensitive to flux changes in that bin. In Fig. 4 we show the network gradients for our five labels across the whole wavelength range of the input spectra. The gradients are scattered randomly around zero for most of the wavelength range. Only at certain wavelength bins, the network is sensitive to flux changes. Here, the gradients show individual, narrow spikes. This is especially apparent in the gradients for [Mg/Fe] and [Al/Fe] in the HR21 part of our input spectra. The [Mg/Fe] gradients show two clear spikes at 8736.0 and 8806.8 Å. These are the locations of two Mg I absorption lines. The largest spike in the [Al/Fe] gradients mark the location of the Al I double feature at $\sim$8773 Å. These Mg and Al lines were used

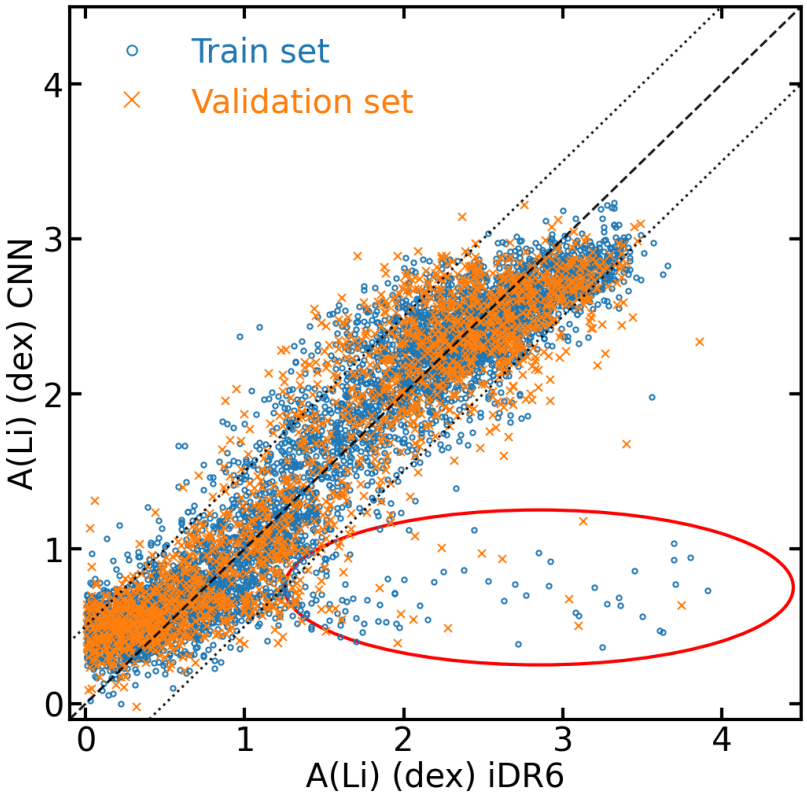


Figure 30: CNN vs. GES-iDR6 A(Li) for the CNN that has been trained on spectra with a masked 6707.8 Å Li line. Blue and orange represent train and validation sets, respectively. The dashed line shows the 1:1 relation, and two dotted lines are at ±0.5 dex. The red ellipse shows the Li-rich giants, whose high input Li abundance has been miss-predicted by the network. Figure from Nepal et al. (2023).

by GES for the determination of our input Mg and Al abundances. The gradients for Mg and Al also show spikes at the locations of the other absorption lines of these elements. Their precise wavelengths in the HR10 and HR21 setups are 5528.41, 8717.81, 8736.02 and 8806.75 Å for Mg and 5557.06, 8772.87 and 8773.90 Å for Al (Heiter et al. 2021b). We therefore see that our network can identify absorption lines in the input spectra. The negative gradient values at these wavelengths mean that if the flux at the absorption lines is low, the predicted abundance is high, and vice versa. This reflects the fact that stronger absorption features in spectra indicate higher elemental abundances in stellar atmospheres. Nepal et al. (2023) show analogous behavior of the CNN gradients of a network that has been trained to predict the Li abundance from GIRAFFE HR15N spectra. The Li gradient at the wavelength of the 6707.8 Å Li absorption line is strong and negative, while it is close to zero for all other wavelengths. The CNN label predictions are therefore directly based on the strength of the relevant absorption lines in the input spectra.

Further investigation of our gradient peaks gives interesting insights into the behavior of our CNN. Some spectral lines influence the network predictions for only one of the labels. An example in the HR10 setup is a Cr I line at ∼5410 Å, that corresponds to a peak in the gradient for $T_{\rm eff}$. Other lines affect multiple, uncorrelated labels. For deriving $T_{\rm eff}$ and $\log(g)$, our CNN is sensitive to the Ni I line in the red end of the HR10 setup. While this line coincides with the strongest peak in the $\log(g)$ gradient, only a minor peak is present in the $T_{\rm eff}$ gradient. A Fe I line at ∼8805 Å is also important for both the $T_{\rm eff}$ and $\log(g)$ predictions, but not for the [Fe/H] likely due to its blend with a Mg line.

The infrared calcium triplet (the three most prominent absorption lines in the HR21 setup) does not have a significant influence on the network predictions for any of the labels, but the Ca II line beyond 8900 Å causes a very strong response of the $T_{\rm eff}$ and [Fe/H] gradients. A deeper investigation of the CNN gradients could be done to search for complementary spectral features that could be used by standard spectroscopic pipelines, but this is out of the scope of this thesis.

We ran several tests to investigate the effect of correlations between the input labels in our training set. For this, we trained our CNN with

different combinations of input labels. We found that the gradients of a combination of $T_{\rm eff}$, $\log(g)$, and one or all abundances show no gradient correlations. This means that the CNN mainly learns from the spectral features. If the network is trained only with the highly correlated labels A(Mg), A(Al), and A(Fe) (absolute abundances), the gradients for the three labels are almost identical (Fig 32). In this case the CNN is still able to identify the locations of the Mg, Al and Fe absorption lines, but the network predictions for one element are also very sensitive to absorption lines of the other two elements. In addition, the quality of the CNN predictions starts to degrade, leading to larger differences between GES input labels and CNN predictions. This is because the network relies too much on the label correlations within the training set instead of the connection between spectral features and labels of individual spectra. For future surveys, we therefore recommend to carefully inspect the training data for strong correlations because they can influence the CNN performance.

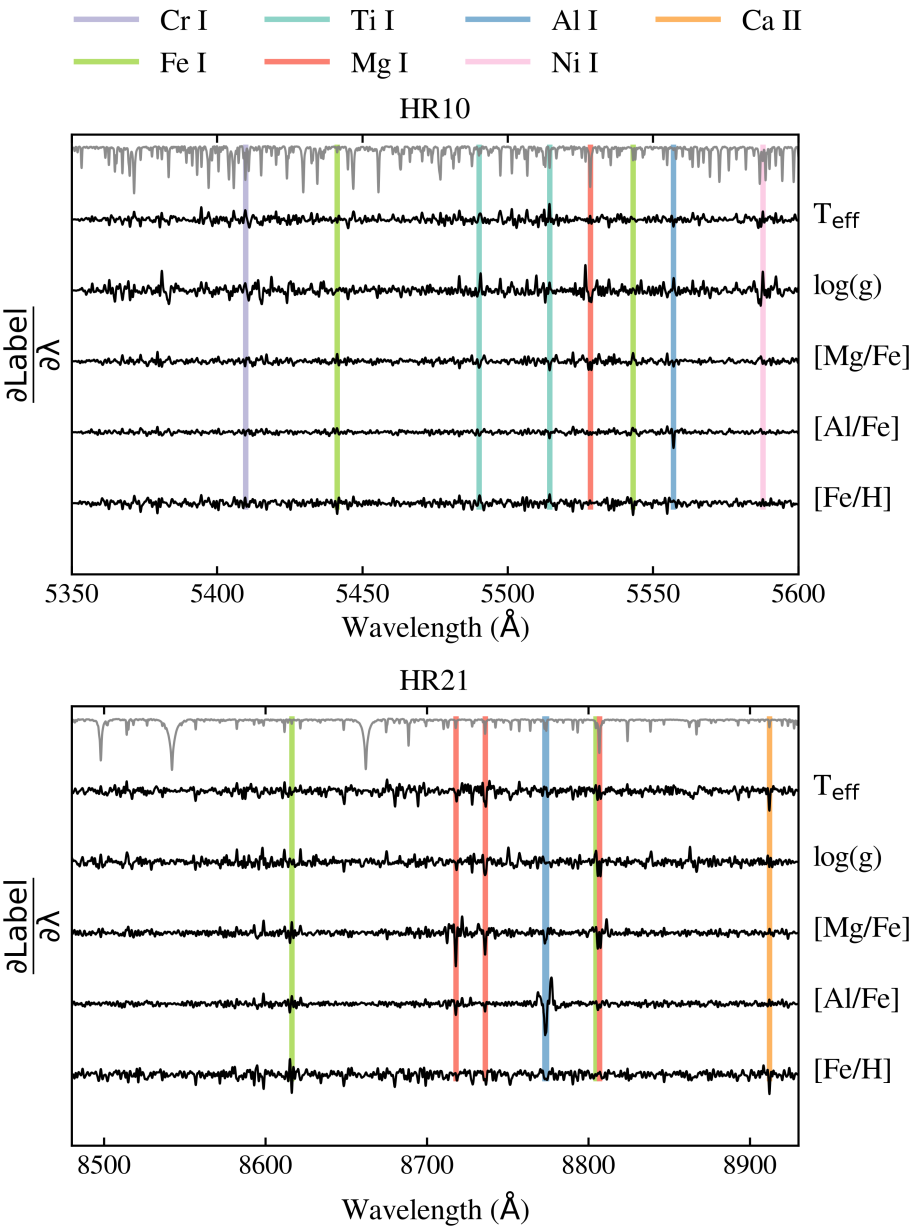


Figure 31: Network gradients for our five labels as a function of wavelength (black). The top panel shows the gradients across the GIRAFFE setup HR10, the bottom panel shows the same for the HR21 setup. An average input spectrum is shown in gray as the top line in both panels. The locations of selected absorption lines of different elements are marked with vertical colored lines.

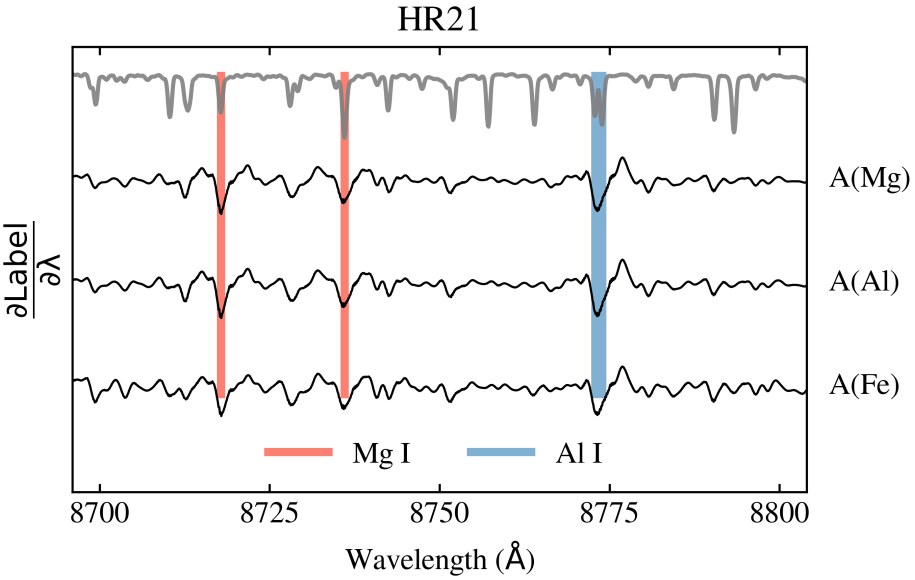


Figure 32: CNN sensitivity maps when trained only on the highly correlated labels A(Mg), A(Al), and A(Fe). For clarity, this figure focuses on a wavelength range in the GIRAFFE HR21 setup. It contains two Mg lines and the Al double absorption line.

5 Evaluation of the CNN results

5.1 Evaluation with benchmark stars

The GES iDR6 data set contains a number of benchmark stars with high quality spectra and precise stellar labels (Heiter et al. 2015). This benchmark set covers stars in different evolutionary stages with a wide range of stellar parameters and abundances, suited for the verification and calibration of large data sets (Pancino et al. 2017b). Our data set contains 25 benchmark stars, including the Sun. As for the rest of our data set, the labels for the benchmark stars were determined spectroscopically by GES. We note that none of the benchmark stars are present in the training or validation sample. Five of the benchmark stars are not part of the inner test set, meaning that their spectra are different from the training set spectra. Four of them have the lowest [Fe/H] of all benchmark stars, the fifth is the benchmark star with the highest [Fe/H] in our data set. The CNN predictions for these five stars do not match the GES input values well. The CNN predictions of the remaining 20

benchmark stars, which are part of the inner test set, agree well with the GES values across all five labels. The largest differences occur for stars on the edges of the parameter space, where the network predicts more moderate values compared to the extreme GES values. An example is HD 49933, the benchmark star with the highest $T_{\rm eff}$, for which our network predicts $\sim$350 K less than what is reported by GES. This star stays one of the hottest in our benchmark set, even with this reduction in $T_{\rm eff}$. Despite the large difference in one label, the CNN predictions for the other labels of HD 49933 agree well with the GES measurements. The label-specific bias and scatter between GES and CNN labels for the benchmark stars in the inner test set is comparable to the bias and scatter that we found for the training and validation sets in Fig. 26.

The CNN predicts similar label values for repeat spectra of our benchmark stars, oftentimes predicting identical labels for multiple repeats. The dispersions between repeated label predictions can be interpreted as the uncertainties of the CNN results. These CNN uncertainties are within the GES label uncertainties for the benchmark stars.

We conclude that our CNN can accurately predict multiple labels of individual stars, if their spectra are similar to the training set spectra. However, the most extreme CNN results should be used cautiously because they are likely to underestimate high values and overestimate low values.

5.2 Network predictions for globular clusters

Our data set includes stars that belong to several globular clusters. We identified member stars of five separate clusters based on their position in the sky and their scatter in [Fe/H] and radial velocities that are reported in GES iDR6. The position of the cluster members in the [Mg/Fe] and [Al/Fe] plots is displayed in Fig. 34. The CNN predictions reproduce the grouping of cluster members in the plots, with a small spread of [Fe/H] within each cluster. However, the CNN predictions show a smaller scatter in [Element/Fe] compared to the GES values, especially for Al. This reduced scatter reflects the results that we saw in Figs. 26 and 3, where the CNN predicts more moderate labels for spectra with extreme GES labels.

Our CNN results recover the Mg-Al anti-correlation, which is used

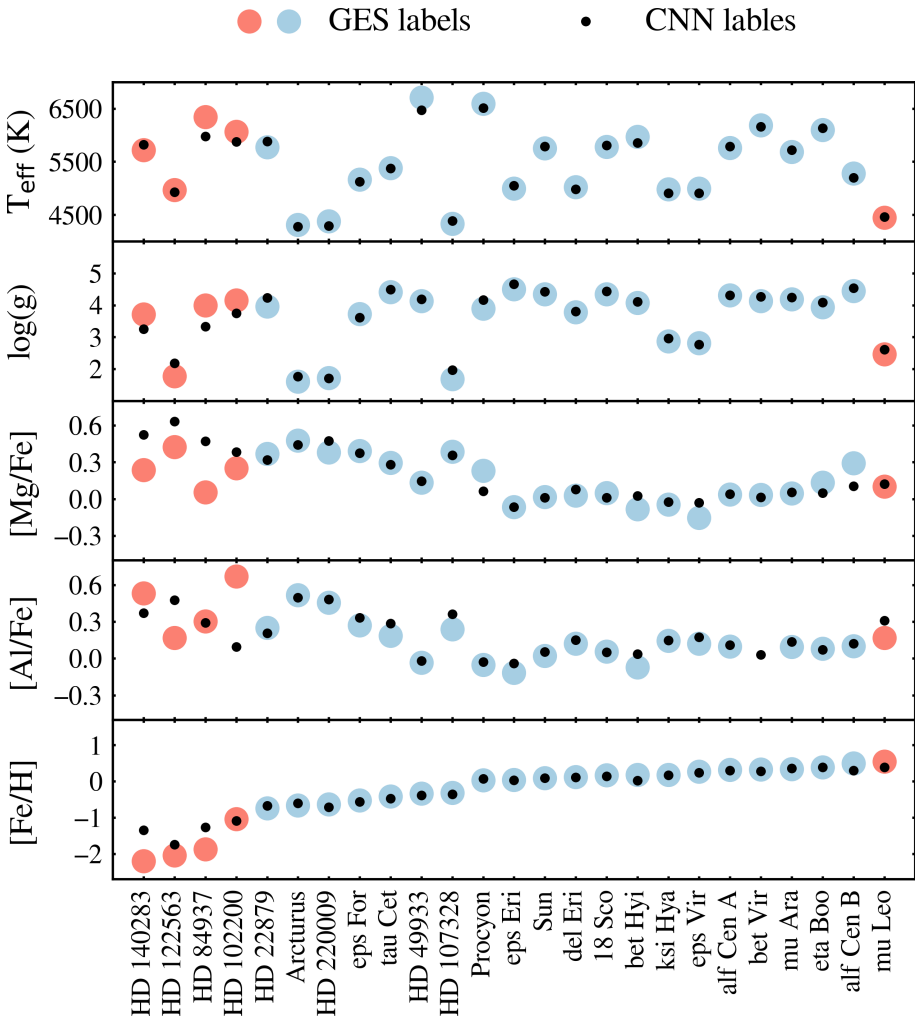


Figure 33: Comparison of GES input labels with CNN predictions for the benchmark stars. Red data points represent benchmark stars that are in the outer test set. Blue data points are benchmark stars in the inner test set and have S/N $\geq$ 30. Different data point sizes have no physical meaning and are for visualization purposes only.

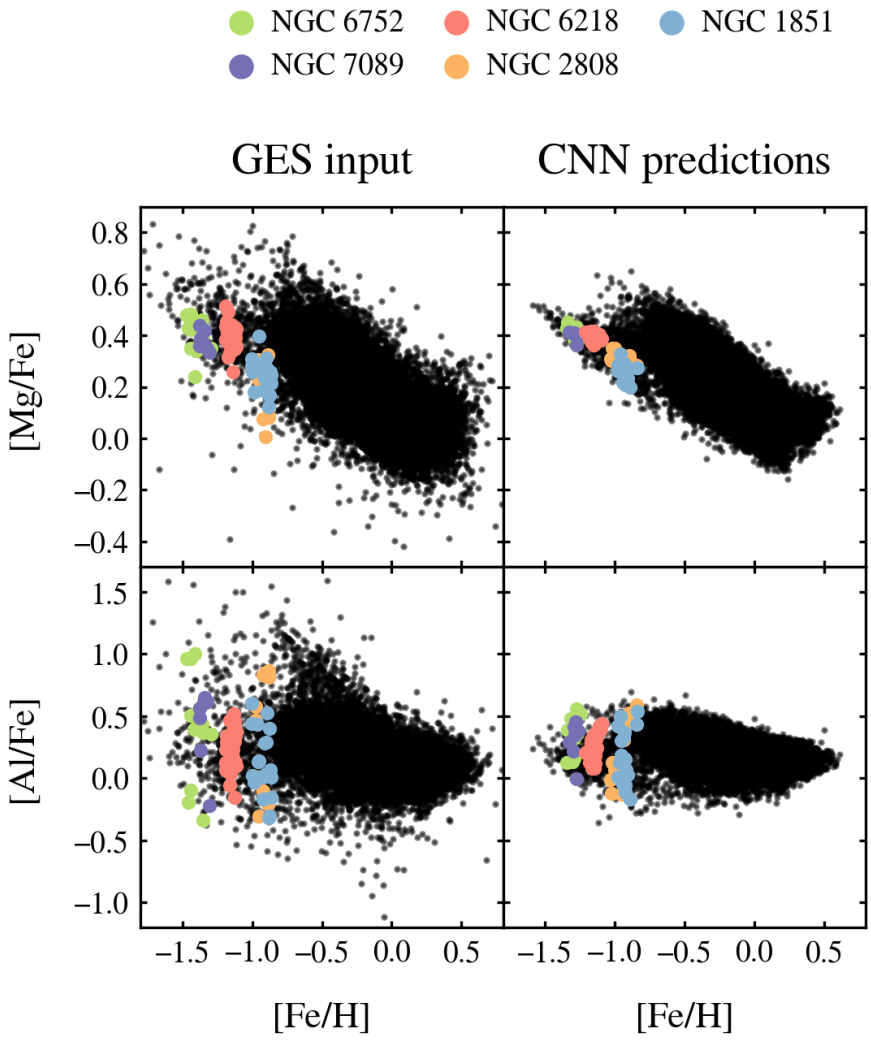


Figure 34: [Mg/Fe] and [Al/Fe] vs. [Fe/H] plots for stars in the training, validation, and test sets. The panels on the left show the distributions of the GES iDR6 values, the panels on the right are the predictions of the trained neural network. Cluster membership is indicated by differently colored datapoints.

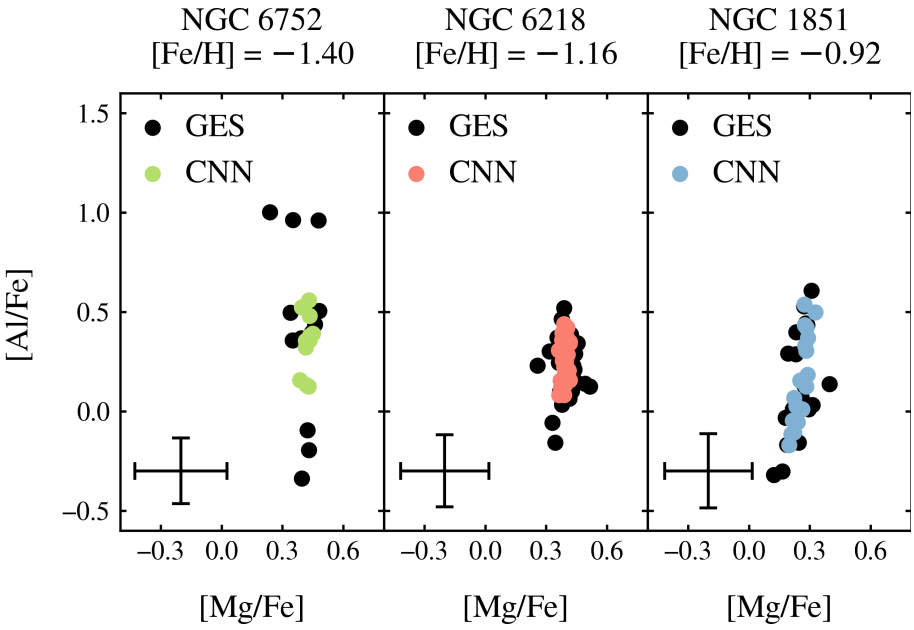


Figure 35: Mg-Al anti-correlation plots for three of our sample clusters with decreasing cluster metallicity. Colored data points show the labels predicted by the CNN; black points are the GES results. The color for the different clusters is the same as in Fig. 34. Average uncertainties of the GES results are shown in the lower left corner.

to investigate the chemical evolution of globular clusters (Pancino et al. 2017c). Figure 35 shows the Mg-Al anti-correlation in the clusters NGC 6752, NGC 6218, and NGC 1851. The average [Fe/H] values of these three clusters span a range of ∼0.5 dex. We see that the match between GES input and CNN output is improving with increasing cluster [Fe/H]. The cluster NGC 6752 contains stars with [Fe/H] values at the lower edge of the training set limit, where the density of training spectra is low. The density of the training set increases with [Fe/H], which leads to better CNN predictions for the cluster stars. Except for the two stars with the lowest [Al/Fe], all CNN predictions for the NGC 6218 agree with the GES results within their reported uncertainties. For NGC 1851, which has the highest average [Fe/H] value among the clusters in our sample, we observe a good match between the Mg-Al anti-correlation as measured by GES and our predicted anti-correlation.

5.2.1 NGC 1851 revisited

Here we investigate in more detail the CNN predictions for the globular cluster NGC 1851 and compare the results to what was found for this cluster in Sect. 2. The predicted [Fe/H] of the 20 cluster stars in our CNN sample spans a range of $\sim$0.2 dex. As already seen in Fig 34, and shown in more detail in Fig. 36, the [Al/Fe] of the stars in this cluster is quite varied.

There are two stars whose [Fe/H] is about 0.1 dex larger than the cluster mean of -0.92 dex. The predicted [Al/Fe] values of these two more metal-rich stars are 0.43 and 0.53 dex. Their average is 0.35 dex larger than the average predicted [Al/Fe] of the remaining, more metal poor stars (0.13 dex). The spread of [Mg/Fe] of the cluster shows little variance across the whole [Fe/H] range. This observation is comparable to what was found in Section 2 when analyzing NGC 1851 member stars with classical spectroscopic methods. There, the average [Al/Fe] value of the metal-rich cluster subpopulation is by 0.21 dex larger than the average of the metal-poor population. The average [Mg/Fe] values of both populations are within 0.01 dex (see the bottom panel of Fig. 36 and Table 1).

Our CNN sample has only one star in common with the cluster sample from Section **??** (CNAME 05134740-4004098). This star's CNN predictions for [Mg/Fe], [Al/Fe], and [Fe/H] are given in Table 5, together with the corresponding values from Section 2. The common star is part of our training set, and in Table 5 we also list its GES input values for the network training.

We remind here that the abundance values from Section 2 are from the 4th internal data release of the Gaia-ESO survey, obtained from

Table 5: CNN predicted values of the labels [Mg/Fe], [Al/Fe], and [Fe/H] for the common star 05134740-4004098, together with its corresponding values from Section **??** (GES iDR4) and the training input values (GES iDR6).

	[Mg/Fe]	[Al/Fe]	[Fe/H]
CNN	0.27	0.43	-0.84
GES iDR4 (UVES)	0.23	0.50	-0.92
GES iDR6 (GIRAFFE)	0.23	0.40	-0.88

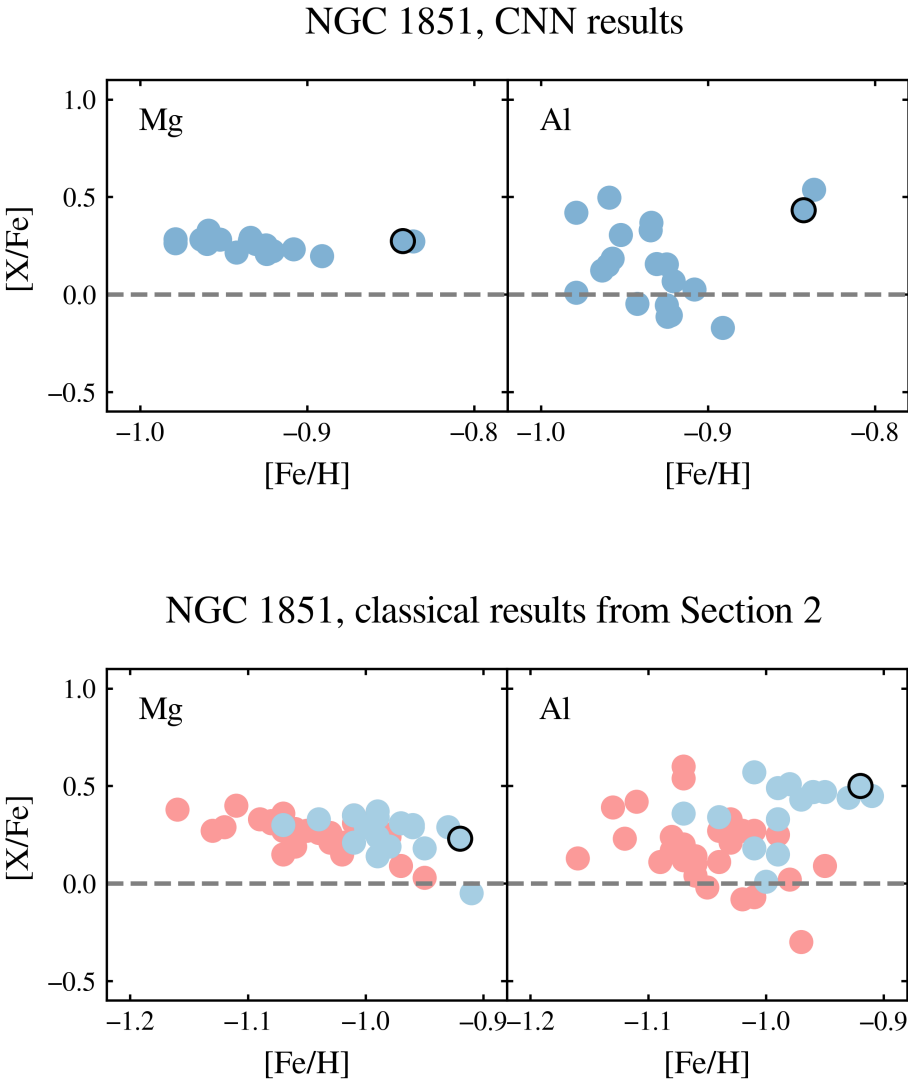


Figure 36: Top: Detailed view of the CNN predicted [Mg/Fe] and [Al/Fe] vs. [Fe/H] distributions of 20 stars in the cluster NGC 1851. Bottom: The same but for the sample of 45 cluster members from Section 2. The metal-rich population is shown in blue, the metal-poor one in red (compare to Fig. 7). In both subplots, results for the common star 05134740-4004098 are marked by a black circle.

spectra by the UVES spectrograph. Our training data is from the 6th internal Gaia-ESO data release and is based on spectra from the GIRAFFE spectrograph. As seen in Table 5, the CNN predictions agree quite well with the classically obtained results from both GES iDR4 and iDR6 data releases.

We conclude that our CNN is able to qualitatively reproduce the Mg and Al results that we found for the cluster NGC 1851 in Section 2. However, we only have CNN predictions for 20 cluster members. A detailed chemical study of the cluster would benefit from a larger sample. For a thorough separation of the cluster stars into chemically distinct subpopulations, abundance predictions for additional chemical elements (for example several s-processed elements and nitrogen) would be necessary.

5.3 Thin and thick disk populations

As discussed in Sect. 4.1.2, [Mg/Fe] values can be used to separate the Milky Way stars into a thin disk and a thick disk population. We performed this separation based on our CNN results for the inner observed set with S/N $\geq$ 30 in combination with the training and test sets. We also attempted to perform the separation for the combination of the low S/N inner observed set and outer observed set. The top panel of Fig. 37 shows the distribution of [Mg/Fe] vs. [Fe/H] for the CNN predictions for the low S/N set plus the outer observed set. We can see that the stars are not separated into the two distinct thin and thick disk populations. The CNN predictions for both [Mg/Fe] and [Fe/H] are strongly clustered around the label averages, and it is not possible to clearly separate the stars into a thin and thick disk population. The bottom panels show the same plot for the inner observed set with S/N $\geq$ 30 plus the test set. Here we can see the separation between the two disks: Thin disk stars with [Mg/Fe] lower than $\sim$0.2 dex and thick disk stars with enhanced [Mg/Fe]. To identify the thick and thin disk stars, we used the clustering algorithm HDBSCAN (Campello et al. 2013), which is implemented in the *hdbscan* library for Python programming. This algorithm assigns data points to different clusters, depending on the density of data points in a distribution. Two clusters are identified that correspond to the two stellar populations, as displayed in the two panels of Fig. 37. About

35% of the stars do not fall into any of the two clusters. Stars outside the two dense regions in the distribution are considered "noise" by the HDBSCAN algorithm and are not assigned to any cluster. In the literature, the chemical separation between thin and thick disk is often performed by splitting the distribution into several [Fe/H] bins and finding the [Mg/Fe] value in each bin where the density of stars is at a minimum (e.g. Adibekyan et al. 2011, Mikolaitis et al. 2014). Anders et al. (2018) use a sophisticated t-SNE approach to identify the different stellar populations. They include abundances measurements from 13 chemical elements to further dissect the thin and thick disk into additional subpopulations.

To investigate the age distributions of the two populations, we used the isochrone fitting code *A Unified tool to estimate Distances, Ages and Masses* (UniDAM). The UniDAM tool (Mints & Hekker 2017) follows a Bayesian approach of isochrone fitting. It compares stellar atmospheric parameters and absolute magnitudes from simulated PARSEC isochrones (Bressan et al. 2012) to the corresponding values in observed stars. All PARSEC isochrones also have stellar masses and ages associated with them. For the isochrone fitting we used the CNN predictions for the atmospheric parameters $T_{\rm eff}$ and $\log(g)$ in combination with [Fe/H]. Magnitudes of our sample stars in the J, H, and K bands were taken from the 2MASS catalog (Skrutskie et al. 2006). For UniDAM to calculate the absolute magnitudes, it is also necessary to provide the parallax value for each sample star. We used the parallaxes from Gaia EDR3 (Gaia Collaboration 2020). We removed stars with negative parallaxes as well as stars with relative parallax errors > 0.2. To get the most precise age estimates, we only considered turn-off stars in this analysis. Turn-off stars in our thin and thick disk samples were selected by their position in the Kiel-diagram. The resulting average age of the thin disk stars is 8.7 Gyr, the average thick disk age is 9.7 Gyr. This age difference between the two populations has been found in numerous studies and by using several age determination methods. Kilic et al. (2017) for example, find ages from 7.4–8.2 Gyr for the thin disk and 9.5–9.9 Gyr for the thick disk by analyzing luminosity functions of white dwarfs in the two disks. Using APOGEE spectra and precise age estimates based on asteroseismic constraints, Miglio et al. (2021)

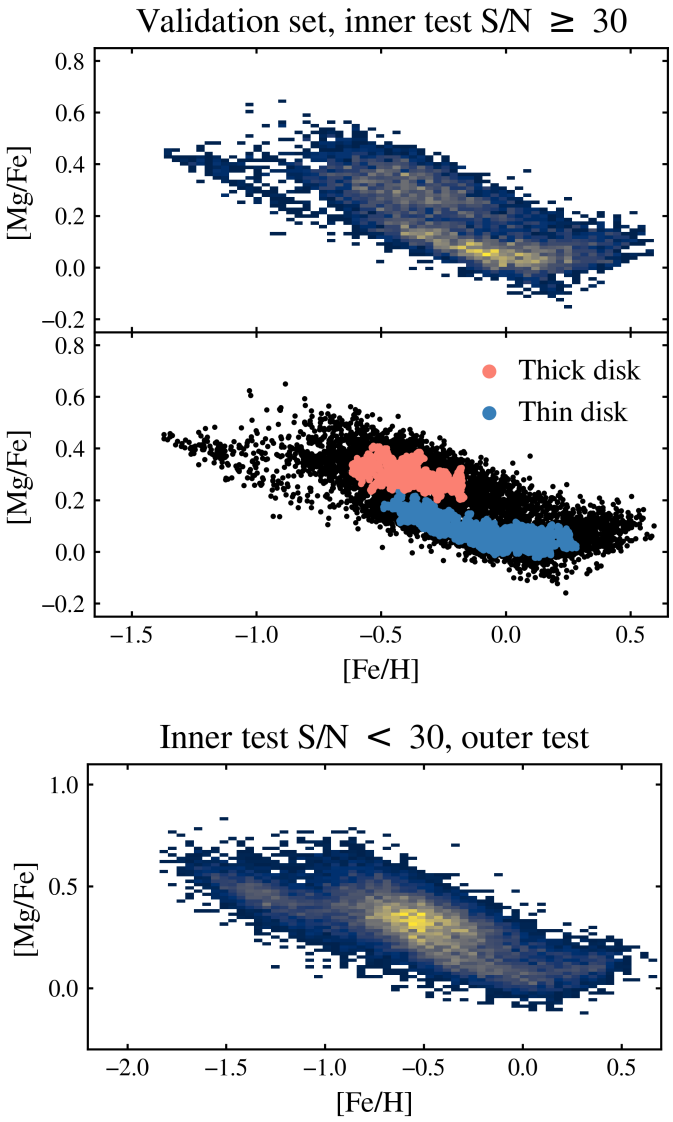


Figure 37: Top panels: Density map of the [Mg/Fe] vs. [Fe/H] distribution of our CNN results for the stars in the validation set and inner observed set with S/N $\geq$ 30. Thin and thick disk populations found by the HDBSCAN algorithm are shown at the bottom. The two populations correspond to the two dense regions in the panel above. Bottom panel: Same density map as above, but for stars in the low S/N inner + outer test sets. Brighter colors indicate a higher density of data points (linear color scale).

also show that the chemically selected thick disk stars are old, with a mean age of $\sim$11 Gyr. We note that the detailed age distribution of thin and thick disk members is sensitive to several selection criteria such as metallicity, kinematic properties, and the distance from the Milky Way center. A detailed investigation of the two stellar populations is out of the scope of this work.

We also investigated the kinematic properties of our thin and thick disk samples. Based on the current positions and velocities of the stars, we integrated their orbits for 5 Gyr in a theoretical Milky Way potential, using the Python-based galactic dynamics package *galpy* (Bovy 2015). For the integration we used the gravitational potential *MWPotential2014*, which combines bulge, disk, and halo potentials. Proper motions, sky coordinates and parallaxes of our sample were taken from the Gaia EDR3. In Fig. 38 we show the trends of the orbital eccentricities relative to [Fe/H] for our thick and thin disk stars. A linear regression model shows that the eccentricity e of thick disk orbits is decreasing with increasing [Fe/H]: $\Delta e/\Delta[\mathrm{Fe/H}] = -0.25$. The eccentricities of thin disk stars are on average lower than the thick disk eccentricities and show a slight positive trend ($\Delta e/\Delta[\mathrm{Fe/H}] = 0.02$). These results are consistent with the findings of Yan et al. (2019), who investigated the chemical and kinematic properties of thin and thick disk stars from the LAMOST data set (Zhao et al. 2012).

5.4 Discovery of 31 new lithium-rich giants

The training set of HR15N spectra with labels, which was used to train the CNN from Nepal et al. (2023) to predict Li abundances, contains 38 Li-rich giants ($\log(g < 3.5$ and $\mathrm{A(Li)} > 2$). When considering that the same training set contains $\sim$2500 giant stars in total, it becomes apparent that Li-rich giants are rare objects. Standard stellar evolution models (for example, from Lagarde et al. 2012) predict that an initial stellar A(Li) value of 3.3 dex decreases during the star's lifetime to below $\sim$1.50 dex. This reduction is predicted to happen mostly during and after the first dredge-up. In this mixing event, Li is transported from the stellar atmosphere into the center of the star, where it is destroyed by nuclear reactions. The existence of Li-rich giants implies the existence of one or several mechanisms that increase the stellar Li abundance

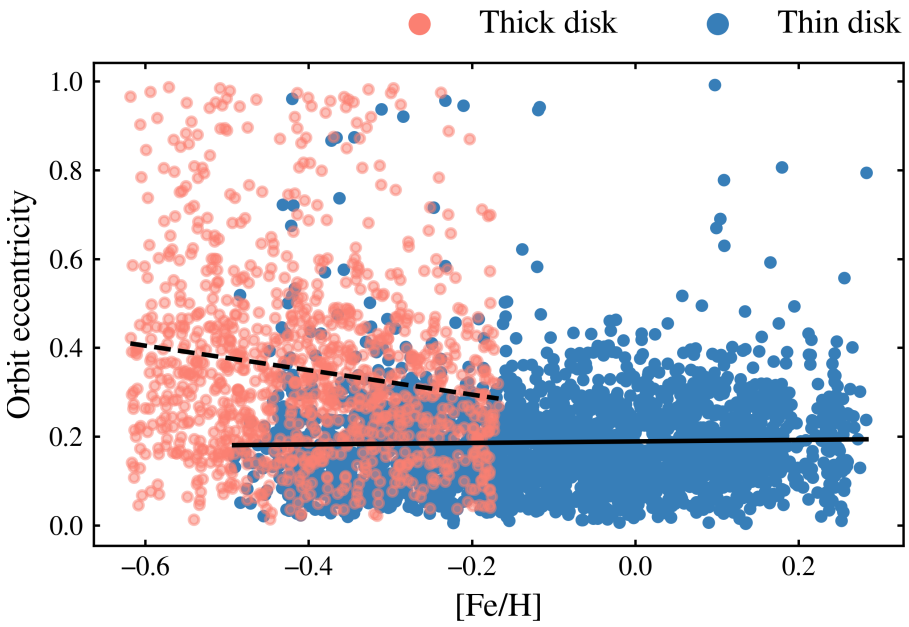


Figure 38: Eccentricities of stellar orbits as a function of [Fe/H] for our thick disk and thin disk samples. Black lines show linear fits to the thick disk (dashed) and thin disk data points (solid).

again after this depletion event. Cameron & Fowler (1971) propose a Li-enrichment mechanism in which, in a first step, Be^7 is produced in the stellar interior. Convection and non-standard mixing processes then rapidly transport the Be^7 to the stellar surface, where it decays to Li^7. External enrichment mechanisms have also been proposed, including planet engulfment and mass-transfer from a Li-rich companion star. The review by Casey et al. (2016) provides an overview over currently discussed Li-enrichment processes in giant stars.

We report here the discovery of 31 previously unidentified Li-rich giants as a result of our CNN predictions. These 31 new Li-rich giants are part of the test set from Nepal et al. (2023) and miss one or multiple labels ($T_{\rm eff}$, log(g), [M/H], or A(Li)) in the Gaia-ESO iDR6 data set. All of these giants have A(Li) > 2 dex, log(g) < 3.5, and $T_{\rm eff}$ < 5500 K. The S/N of their specra is > 25 and the internal CNN uncertainties of their label predictions are < 50 K for $T_{\rm eff}$, and < 0.1 dex for log(g), [M/H], and A(Li). None of these 31 stars have been reported as Li-rich giants in any previous Gaia-ESO papers. A search in the GALAH

survey catalog of Li-rich giants in the southern hemisphere Martell et al. (2021) also found no match with any of our newly discovered objects. We list our found Li-rich giants with their predicted labels in Table 6. Figure 5 shows the position of the found Li-rich giants together with label predictions for the stars from the above-mentioned training set. The giant branch of the Kiel diagram contains 67 Li-rich (A(Li) > 2 dex) in total, 38 of these are included in the training set.

Despite their rarity, Li-rich giants are believed to play an important role in the Li-enrichment of the interstellar medium (Romano et al. 2001). Studying these stars can therefore provide important insights into the formation and enrichment history of the Milky Way galaxy. Up to now, $\sim$11 000 Li-rich giants have been found (Hong-liang & Jianrong 2022). This number is expected to grow considerably with the upcoming large-scale spectroscopic surveys. Machine-learning methods like our CNN will be key in the search for the Li-rich giants and other rare objects.

Table 6: Gaia-ESO CNAMES of our 31 discovered Li-rich giants and their label predictions for T_{eff} (K), log(g) (dex), [M/H] (dex), and A(Li) (dex). The stars are sorted by descending A(Li).

CNAME	T_{eff}	log(g)	[M/H]	A(Li)
07434938−3841399	4841	2.84	−0.31	3.88
10495937−6345553	4805	2.72	−0.16	3.83
07464933−3750081	4948	2.9	−0.20	3.62
08064077−4736441	4797	2.66	−0.10	3.56
16271097−2455213	4920	2.82	−0.45	3.55
06410348+0905141	5071	3.13	−0.18	3.50
07493206−3759457	4799	2.69	−0.24	3.48
10430727−6456318	4619	2.47	−0.16	3.43
08110435−4853491	4831	2.68	−0.30	3.42
10400095−6419586	4525	2.34	0.01	3.29
08084532−4701292	4836	2.74	−0.19	3.26
07462219−3712141	4862	2.82	−0.20	3.22
06273069−0440141	4714	2.53	−0.68	3.21
08512566−4135067	4331	2.20	0.23	3.15
08102172−4845417	4514	2.36	−0.06	3.14
06255393−0457404	4981	2.93	−0.29	3.03
08083354−4711111	4441	2.31	0.10	3.00
07442999−3812166	4857	2.65	−0.24	2.98
10350175−6405092	4469	2.35	0.11	2.88
07475310−3733040	4853	2.80	−0.21	2.86
10483936−6327542	4383	2.21	0.08	2.74
10420066−6421333	4397	2.22	0.08	2.73
11130526−7617396	4815	2.67	−0.32	2.73
11123294−7727006	4315	2.31	0.16	2.61
10575316−7634459	4858	2.70	−0.21	2.42
07472841−3850499	5276	3.47	−0.12	2.35
10513847−6335341	4352	2.24	0.31	2.29
07472390−3856376	5049	2.93	−0.24	2.27
06272996−0518528	4522	2.44	−0.02	2.20
08075108−4744027	4719	2.57	−0.13	2.19
07483625−3724338	4939	3.01	−0.15	2.03

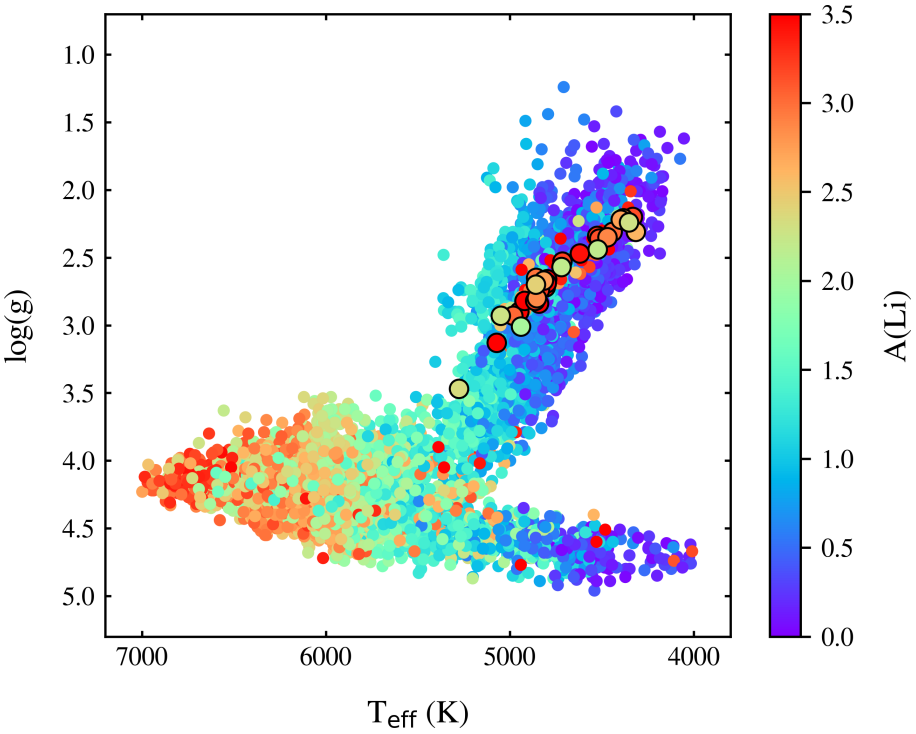


Figure 39: Kiel diagram of the training set stars from Nepal et al. (2023) and the 31 discovered Li-rich giants as larger data points with black edges. The data points are color-coded according to their lithium abundance. The label values are CNN predictions.

6 Caveats of the CNN method

During the network training, the GES input labels are considered to provide the true parameterization of the training spectra. The quality of the network predictions therefore depends entirely on the quality of the training data. We limited the uncertainties and errors in our training data by applying several quality constraints (Sect. 4.1.1), but there is still a possibility that the input labels may suffer from systematics. Inaccurate labels of a few input spectra will not have a noticeable effect on the training process. The cases with a large difference between GES input value and CNN prediction could therefore be the result of the network predicting accurate labels for spectra with inaccurate GES labels. Future work could investigate if and how CNNs can be used for the quality control of classically derived stellar parameters. Future surveys should also take care of including proper 3D and NLTE modeling when deriving atmospheric parameters and chemical abundances.

We estimate the internal uncertainties of our network predictions by training multiple CNN models on the same data. These uncertainties, however, do not consider the uncertainties of the training labels themselves. Bayesian deep learning frameworks account for both the training data uncertainties and model uncertainties (Kendall & Gal 2017). Future work could benefit from implementing this Bayesian approach into our CNN method.

The predictive power of our CNN is limited by the sparse training data that is available at the edges of the parameter space (Sect. 4.2.2). A more homogeneous coverage of the parameter space, achieved by increasing the number of training spectra with extreme parameter values, will increase the precision of the CNN predictions. In this way, the training sample is proactively built instead of relying on an existing set of labels.

During the training phase, our CNN not only learns the correlations between spectral features and stellar labels but is also sensitive to correlations within the training labels themselves. The effect of this is discussed in Sect. 4.3, where we see for example how the strength of Mg absorption lines also influences the network predictions for [Al/Fe]. These correlations can never be avoided when training the network to

predict multiple abundances at once. The alternative then is to train a separate network model for each abundance label. This strategy decreases the efficiency of the CNN approach, especially when the goal is to predict abundances of many chemical elements. Therefore, care should be taken to reduce the correlations in the training data without sacrificing the ability of the network to predict multiple labels at once.

7 Summary and conclusions

Here we summarize the main results of this thesis and the steps carried out to find these results.

- We investigated spectra of 45 red giant stars in the globular cluster NGC 1851 with classical spectroscopic methods. Stars in this cluster can be separated into two subsamples, with a difference of 0.07 dex in their mean metallicity. The subsamples also differ in their mean carbon-to-nitrogen ratios and abundances of the s-processed neutron-capture elements Y, Zr, Ba, La, Ce, and Nd. The average difference of these s-processed elements in the two subpopulations is 0.35 dex. We found no notable distinction in the average abundance-to-iron ratios of carbon, europium, α- and iron-peak elements. Analysis of the C, N, and O abundances in the cluster stars shows that the metal-rich subpopulation is $\sim$0.6 Gyr older than the metal-poor one.

- We built a training and a test set based on GES iDR6 spectra with S/N $\geq$ 30. Together, these sets consist of 14 634 stellar spectra with associated atmospheric parameters and chemical abundances. We applied several quality checks on these sets to ensure that our network is trained on high quality spectra and stellar labels. We use the parameters $T_{\rm eff}$ and $\log(g)$ and the abundances [Mg/Fe], [Al/Fe], and [Fe/H] as the input labels for our neural network. We also built an observed set of 22 270 spectra to test the performance of our CNN on spectra that were not involved in the training process.

- We used t-SNE to identify observed spectra that are similar in shape to the training set spectra. In this way we can identify spectra that are likely to have labels within the training label range, without relying on their GES labels. Less than 10% of the observed spectra that are similar to the training set spectra in shape and S/N range have GES labels outside the training set limits. This pre-selection step is important because neural networks are not able to accurately predict labels outside the training range.

- We then built a convolutional neural network with the Python-based library *Keras*. Our network architecture contains three convolutional layers, designed to detect features and absorption lines in input spectra. Three succeeding dense layers then convert the found spectral features into the values of the five output labels. We performed ten training runs, resulting in ten slightly different CNN models. We used the eight best CNN models to predict the labels of the training, test, and observed set spectra.

- On average, one training run took $\sim$45 minutes to complete on a desktop PC, using only CPU. Label prediction with our trained network is very fast: the parameterization of the $\sim$35 000 spectra in our data set took less than 20 seconds per CNN model.

- The CNN label predictions for the training and test sets agree well with the GES input labels. The bias (average offset) and scatter between CNN and GES labels are identical for the training and test sets, showing that our CNN is not over-fitting during the training. We use the scatter between GES input and CNN output for the training set as a measure for the training precision of our network: The training precision is 37 K for $T_{\rm eff}$, 0.06 dex for $\log(g)$, 0.05 dex for [Mg/Fe], 0.08 dex for [Al/Fe], and 0.04 dex for [Fe/H]. The results for the pre-selected observed set, with similar spectral shape and S/N range as the training set, are also in good agreement with the GES input values, albeit with a larger scatter between CNN and GES values. We find that the quality of the CNN results degrades for spectra with S/N < 30, especially for abundance predictions. Observed spectra that are different from the training set spectra are not parameterized accurately. We warn the community that machine–learning on low-S/N spectra may not be sufficient for deriving precise enough abundances. Surveys should therefore gather spectra with high-enough S/N (depending on their science goals).

- All sets of spectra have in common that the differences between CNN predictions and GES values increase at the edges of the parameter space. At the edges, the number of available training spectra is small. Increasing the number of training spectra in these

parameter regimes would increase the accuracy (mean bias) of the CNN predictions, as well as precision as the number of sample observations increases.

- The scatter between the predictions from the eight different CNN models can be used to assess the internal precision of our network. This scatter is small: On average 27 K for $T_{\rm eff}$, 0.04 for $\log(g)$, and 0.03 dex for [Mg/Fe], [Al/Fe], and [Fe/H] alike. However, the mean scatter may overestimate the precision of our network predictions. We find that the uncertainties increase at the edges of the parameter space. The uncertainties also increase as the spectra S/N decreases. Therefore, the spectra S/N and the position of the predicted labels in the parameter space should also be considered when estimating the label precision for individual spectra.

- We use network gradients to demonstrate the sensitivity of our network to different parts of the input spectra. The gradients show that the network can identify absorption lines in the input spectra and associates those lines to the relevant stellar labels. Caution should be applied when choosing input labels, because strongly correlated input labels lead to strongly correlated network gradients. The network then predicts labels based on unrelated spectral features (for example, absolute Al abundance from Mg absorption lines). Inferring stellar parameters from correlations like these can lead to satisfying results for some spectra. However, stars with exotic chemical compositions will not be parameterized well.

- The validation of our results with 25 GES benchmark stars shows that our CNN can precisely predict labels for individual stars over a large range of label values. Network predictions for repeat spectra of the benchmark stars show a small scatter per star. This scatter is within the GES uncertainties for the benchmark star labels.

- We investigated the Mg-Al anti-correlation in globular clusters, ranging from -0.92 to -1.40 in metallicity. In the most metal-poor regime, where our training set contains only a few stars, our CNN mainly recovers the Al spread in the clusters. The match between GES Mg-Al anti-correlation and CNN anti-correlation is

improving for clusters with higher [Fe/H], where the training data is denser.

- Our CNN is able to qualitatively reproduce the Mg and Al results that we found for the cluster NGC 1851 in Section 2. A detailed chemical study of the cluster would benefit from a larger sample. For a thorough separation of the cluster stars into chemically distinct subpopulations, abundance predictions for additional chemical elements (for example several s-processed elements and nitrogen) would be necessary.

- We found 31 previously unidentified lithium-rich giants in the Gaia-ESO iDR6 data set by training a convolutional neural network to predict Li abundances. The predicted A(Li) of all of these giants exceeds 2 dex, their $\log(g)$ is lower than 3.5 dex. The precision of the network label predictions for these 31 stars is high, with $eT_{\rm eff} < 50$ K and $e\log(g)$, e[M/H], and eA(Li) < 0.1 dex. None of these 31 Li-rich giants have previously been reported in any Gaia-ESO paper, and they do not appear in the GALAH survey catalog of Li-rich giants in the southern hemisphere. Lithium-rich giants are rare. Machine learning methods will play a crucial role in finding these objects in the observations of the upcoming large-scale spectroscopic surveys.

8 Acknowledgements

Here I would like to express my gratitude towards those who have made my PhD studies a joyful experience. First to my supervisor Šarūnas, whose guidance, patience, and open-mindedness gave me the confidence that is needed to finish a work like this. Then to Guillaume, who helped me to shape my first publication, which became the basis for this thesis. I want to thank my friends and colleagues, both here in Vilnius and outside: Edita, Arnas, Carlos, Vidas, Renata, Gražina, Kotryna, Barkha, Ashutosh, Samir, and Yuriy. Every day I look forward to meet them at the office. My gratitude is also towards the people of Lithuania, whose investment into education and science makes all of this possible. Many thanks to Miglė and most thanks to my parents − without them I wouldn't be here.

Dedicated to the memory of my mother

Santrauka (Summary in Lithuanian)

Darbo tikslai ir mokslinis naujumas

Pagrindinis šio darbo tikslas - sukurti ir apmokyti konvoliucinius neuroninius tinklus (convolutional neural networks - CNN)[4], kurie galėtų efektyviai ir greitai parametrizuoti didelės apimties žvaigždžių spektrų rinkinius. Tokie tinklai vienu metu nustato pagrindinius žvaigždžių atmosferos parametrus ir cheminių elementų gausas. Šiuo darbu siekiama parodyti, kad nauji mašininio mokymosi metodai yra patikimi įrankiai, kurie padės išnaudoti visą būsimų didelio masto spektroskopinių tyrimų potencialą. Taip pat bus parodyta, kad tinklo prognozių kokybė labai priklauso nuo mokomųjų spektrų ir susijusių etikečių kokybės. Tinklo apmokymui naudojamos etiketės nustatomos klasikiniais spektroskopiniais metodais. Siekiant įvertinti tinklo našumą, svarbu žinoti, kaip buvo sudarytas mokomasis rinkinys ir kokie yra jo apribojimai. Šiame darbe sukurti metodai tobulinti remiantis ankstesniais panašiais darbais: tinklo architektūros sudėtingumo, mokymo strategijų ir tinklo prognozių taiklumo bei tikslumo kontekste. Šioje studijoje pademonstruota, kad konvoliuciniai neuroniniai tinklai gali sąvarankiškai atrinkti tyrimui naudojamas spektro linijas. Tinklas prognozuoja šių linijų etiketes tokiu būdu, kuris yra intuityvus ir atitinka tokias procedūras, kokiomis klasikiniais metodais analizuojami žvaigždžių spektrai. Norint nustatyti chemiškai ypatingas žvaigždes, labai svarbu remtis esamais spektriniais požymiais, o ne vien tik numatyti astrofizikinėmis koreliacijomis pagrįstas etiketes. Šis metodas taip pat padidina tinklo prognozių supratimą.

Naujausioje literatūroje aprašoma, kad šiame darbe sukurti metodai buvo išplėtoti taip, kad būtų galima ištirti daugiau kaip 800 000 *Gaia*-RVS spektrų. Su šia disertacija susijusios publikacijos yra pasirengimas būsimoms WEAVE ir 4MOST spektroskopinėms apžvalgoms. Šiose

[4] angliškų terminų vertimui į lietuvių kalbą naudotas mašininio ir giliojo mokymosi sąvokų žodynas, inicijuotas dr. Lino Petkevičiaus iš VU Matematikos ir Informatikos fakulteto: https://github.com/linas-p/ML-AI-2-LT

apžvalgose bus stebima milijonai žvaigždžių, esančių visose Paukščių Tako dalyse.

Ginamieji teiginiai

1. Žvaigždes, esančias kamuoliniame spiečiuje NGC 1851, galima suskirstyti į dvi grupes, kurių vidutinis metalingumas skiriasi per 0,07 dex. Šios grupės taip pat skiriasi vidutiniu anglies ir azoto gausų santykiu bei letojo neutronų pagavimo branduolinių reakcijų (s proceso) elementų Y, Zr, Ba, La, Ce ir Nd elementų gausomis. Vidutinis šių *s*-proceso elementų gausų skirtumas abiejose sub-populiacijose yra 0,35 dex. Nepastebėta jokių reikšmingų skirtumų tarp vidutinių anglies, europio ir geležies piko ir α elementų gausų.

2. Naudodami neprižiūrimo mašininio mokymosi algoritmą t-SNE galime nustatyti, ar stebėto spektro forma yra panaši į mokomosios aibės spektrų rinkinį. Spektrai, kurių forma ir S/N santykis yra panašūs į mokomojo rinkinio, CNN nustatyti rezultatai gerai dera su GES vertėmis. Taip pat nustatyta, kad CNN rezultatų kokybė pablogėja, kai spektrų S/N < 30, ypač prognozuojant elementų gausas. Spektrai, kurių forma skiriasi nuo mokomosios aibės spektrų, parametrizuojami netiksliai.

3. Po apmokymo neuroniniai tinklai per kelias minutes gali vienu metu įvertinti žvaigždžių atmosferos parametrus $T_{\rm eff}$ ir $\log(g)$ bei cheminių elementų gausas [Mg/Fe], [Al/Fe] ir [Fe/H] iš $\sim$ 35 000 žvaigždžių spektrų. $T_{\rm eff}$ nustatymo tikslumas yra 37 K, $\log(g)$ - 0,06 dex, cheminių gausų - < 0,08 dex.

   Šiame darbe tyrėme Mg-Al antikoreliaciją kamuoliniuose spiečiuose, kurių metalingumas svyruoja nuo 0,92 iki 1,40 dex. Nemetalingiausiuose speičiuose, kurių mokymo rinkinyje yra tik kelios žvaigždės, CNN daugiausia atkuria aliuminio gausų sklaidą. GES Mg-Al antikoreliacijos ir CNN antikoreliacijos geriau atkuriamos metalingesniuose spiečiuose, kuriuose mokymo duomenų tankis yra didesnis.

   Remdamiesi CNN [Mg/Fe] ir [Fe/H] įverčiais savo imties žvaigždes suskirstėme į plono ir storo disko populiacijas. Nustatėme, kad storojo disko žvaigždės yra vidutiniškai 1 mlrd. m. senesnės už plonojo disko žvaigždes. Storo disko žvaigždžių orbitos eks-

centricitetas turi neigiamą tendenciją nuo metalingmo ($\Delta e/\Delta$[Fe/H] $= -0{,}26$). Šie CNN pagrįsti rezultatai atitinka panašius literatūroje pateiktus tyrimus.

4. Apmokę konvoliucinį neuroninį tinklą nustatyti ličio gausą žvaigždžių atmosferose, "Gaia-ESO iDR6" duomenų rinkinyje radome 31-ą anksčiau neidentifikuotą ličiu praturtintą raudonąją milžinę ($\log(g) < 3{,}5$ dex). Šioms žvaigždėms nustatyta A(Li) gausa viršija 2.0 dex. Tinklo etikečių prognozių tikslumas šioms žvaigždėms yra didelis: e$T_{\rm eff} < 50$ K ir elog(g), e[M/H] bei eA(Li) $< 0{,}1$ dex. Nė viena iš šių 31 Li turtingų raudonųjų milžinių anksčiau nebuvo aprašyta jokiame "Gaia-ESO" straipsnyje. Jų nėra ir GALAH apžvalgos kataloge, kuriame aprašomos pietų pusrutulio ličiu praturtintos raudonosios milžinės. Turint omenyje, kad ličiu praturtintos raudonosios milžinės yra retos, galima sakyti, kad mašininio mokymosi metodai bus labai svarbūs ieškant šių objektų būsimose didelio masto spektroskopinėse apžvalgose.

5. CNN gradientus naudojame tam, kad pademonstruotume mūsų tinklo jautrumą įvairiems įvesties spektro segmentams. Gradientai rodo, kad tinklas sugeba panaudoti absorbcijos linijas įvesties spektruose. Mūsų CNN susieja rastų sugerties linijų stiprumą su atitinkamomis žvaigždžių etiketėmis. Renkantis įvesties etiketes reikia atidumo, nes stipriai koreliuotos įvesties etiketės lemia stipriai koreliuotus tinklo gradientus. Tada tinklas prognozuoja etiketes pagal nesusijusias spektrines savybes (pavyzdžiui, absoliučią Al gausą pagal Mg sugerties linijas). Žvaigždžių parametrus nustatant pagal tokias koreliacijas, kai kurių spektrų atveju galima gauti patenkinamus rezultatus. Tačiau žvaigždžių, kurių cheminė sudėtis yra ne tipinė, parametrai nebus gerai nustatyti.

6. Mūsų mokymo, tikrinimo ir testavimo domenų rinkiniams bendra tai, kad skirtumai tarp CNN prognozių ir GES verčių didėja parametrų erdvės pakraščiuose. Kraštuose esančių tinklo mokymui skirtų spektrų skaičius yra mažas. Didinant tokių spektrų skaičių šiose parametrų erdvėse būtų galima padidinti CNN prognozių tikslumą (vidutinį nuokrypį), taip pat preciziškumą didėjant stebėjimų imčiai.

Kelių skirtingų CNN modelių prognozių sklaida gali būti naudojama mūsų tinklo vidiniam tikslumui įvertinti. Ši sklaida yra nedidelė: vidutiniškai 27 K, jei tai T_{eff}, 0,04 $\log(g)$ ir 0,03 [Mg/Fe], [Al/Fe] bei [Fe/H]. Nustatėme, kad neapibrėžtys didėja parametrų erdvės pakraščiuose. Neapibrėžtys taip pat didėja mažėjant spektro signalo-triukšmo santykiui. Todėl vertinant atskirų spektrų etikečių tikslumą, taip pat reikėtų atsižvelgti ir į spektrų S/N ir prognozuojamų etikečių padėtį parametrų erdvėje.

1 Įvadas

1.1 Žvaigždžių spektrų analizė taikant mašininio mokymosi metodus

Jau daugiau nei prieš tris dešimtmečius buvo prognozojama, kad mašininis mokymasis bus naudajimas didelių duomenų rinkinių tyrimui astronomijoje (Rosenthal 1988). Tačiau didelės šio metodo skaičiavimo sąnaudos ilgai stabdė šio metodo plėtrą. Vieni pirmųjų mašininio mokymosi neuroninių tinklų taikymų buvo automatinė objektų paieška astronominiuose vaizduose (SExtractor, Bertin & Arnouts 1996), galaktikų morfologinis klasifikavimas (Lahav et al. 1996) ir žvaigždžių spektrų klasifikavimas (Bailer-Jones 1997). Pastaraisiais metais dėl tobulėjančių šiuolaikinių kompiuterinių sistemų ir debesų kompiuterijos galimybių vis labiau populiarėja mašininio mokymosi metodai. Galingos Python atvirojo kodo bibliotekos, tokios kaip TensorFlow (Abadi et al. 2015) ir PyTorch (Paszke et al. 2019) siūlo lengvai naudojamas kelių tipų neuroninių tinklų kūrimo ir mokymo sistemas.

Didelės spektroskopinės apžvalgos leidžia tirti didelių žvaigždžių struktūrų, tokių kaip kamuoliniai spiečiai, evoliuciją, taip pat ir viso Paukščių Tako evoliuciją. Būsimi projektai, pavyzdžiui, Viljamo Heršelio teleskopo „William Herschel Telescope Enhanced Area Velocity Explorer“ (WEAVE, Dalton et al. 2018) ir 4 metrų multiobjektinis spektroskopinis teleskopas (4MOST, de Jong et al. 2019), leis stebėti milijonus žvaigždžių. Reikės veiksmingų automatinių priemonių didelės apimties rinkinių spektrinei analizei, kuriuos pateiks tokios apžvalgos. Šiomis aplinkybėmis dabartinė „Gaia-ESO“ spektroskopinė apžvalga pateikia idealų duomenų rinkinį mašininio mokymosi metodams kurti ir testuoti.

Pastaruoju metu konvoliuciniai neuronų tinklai (CNN) naudojami vienu metu iš žvaigždžių spektrų nustatyti kelias žvaigždžių etiketes (t. y. atmosferos parametrus ir chemines gausas). Kiekvieną CNN sudaro konvoliuciniai sluoksniai, kurie leidžia tinklui įvesties duomenyse rasti išplėstinius požymius. Žvaigždžių spektruose tokie požymiai yra absorbcijos linijos ir kontinuumo taškai; dvimačiuose vaizduose tokie požymiai gali būti akys veide arba žvaigždžių telkiniai spiralinėje galaktikoje (Bialopetravičius & Narbutis 2020). Neuroninių tinklų metodai pagrįsti tik duomenimis, todėl jiems nereikia įvesti jokių fizikinių

dėsnių ar modelių. Vietoj to per, mokymo etapą tinklas išmoksta susieti spektrinių požymių stiprumą su žvaigždžių etikečių reikšmėmis. Tam reikia mokomojo spektrų rinkinio su iš anksto nustatytomis etiketėmis, iš kurių tinklas gali mokytis. Spektrinės analizės mokomuosius rinkinius paprastai sudaro keli tūkstančiai žvaigždžių spektrų su kokybiškai nustatytomis etiketėmis. Dabartinės spektrinės apžvalgos, kuriose pateikiami $\sim 10^5$ spektrai su etiketėmis, yra ideali terpė CNN metodų kūrimui ir tobulinimui.

Pagrindinis mašininio mokymosi naudojimo spektrų parametrams nustatyti privalumas yra skaičiavimo greitis. Klasikiniais metodais automatinis atmosferos parametrų ir gausų nustatymas iš vieno spektro paprastai trunka kelias minutes, o apmokytas CNN per tą patį laiką gali parametrizuoti 10^4 spektrų. Šis greitis yra labai svarbus, norint visiškai išnaudoti būsimų spektrų tyrimų galimybes. Pavyzdžiui, 4MOST stebės $\approx$ 25 000 žvaigždžių kiekvieną naktį, o kiekvienai žvaigždei bus galima išmatuoti iki 15-os elementų gausų. Mašininis mokymasis bus labai tinkamas metodas analizuoti tokius didelius duomenų kiekius.

Žvaigždžių parametrizavimo naudojant CNN pavyzdžių galima rasti naujausiuose tyrimuose. Fabbro et al. (2018) sukūrė StarNet - CNN, kuris gali tiesiogiai išvesti žvaigždžių atmosferos parametrus iš stebimų spektrų APO Galaktikos evoliucijos eksperimente (APOGEE, Majewski et al. 2017). StarNet mokymui ir testavimui buvo naudojamas sintetinių spektrų tinklas.

Leung & Bovy (2019) savo astroNN konvoliuciniam tinklui apmokyti naudojo tikrus stebėjimo duomenis iš APOGEE DR14. Siekiant imituoti standartinės spektroskopinės analizės metodus, astroNN sukurtas taip, kad prognozuojant atmosferos parametrus būtų naudojamas visas spektras, tačiau prognozuojant chemines gausas apsiribojama atskirais spektro segmentais. Guiglion et al. (2020) savo darbe CNN apmokė su vidutinės skiriamosios gebos žvaigždžių spektrais iš RAdial Velocity Experiment (RAVE, Steinmetz et al. 2020) kartu su žvaigždžių etiketėmis, kurios buvo gautos iš didelės skiriamosios gebos APOGEE DR16 spektrų. Jie taip pat pridėjo žvaigždžių absoliutinius ryškius ir ekstinkcijos pataisas, kaip tinklo įvesties duomenis. Ši informacija leido jų CNN nustatyti ir naudoti papildomus apribojimus efektinės temperatūros ir laisvo kritimo pagreičio įvertinimui.

Šioje disertacijoje CNN metodas naudojamas „Gaia-ESO“ spektrinės apžvalgos (GES, Gilmore et al. 2012; Randich et al. 2013) kontekste. Naudojome GIRAFFE spektrus su etiketėmis iš iDR6 rinkinio. Bendrai, GES yra skirtas papildyti kosminės observatorijos „Gaia“ astrometrinius duomenis (Gaia Collaboration et al. 2016). Šio darbo tikslas - paruošti pagrindus mašininiam mokymuisi naujos kartos spektroskopinėms apžvalgoms, tokioms kaip 4MOST ir WEAVE.

1.2 Gaia-ESO spektrinė apžvalga

„Gaia-ESO“ spektrinės apžvalgos tikslas - stebėti daugiau kaip 10^5 žvaigždžių, kad būtų gauta vienalytė įvairių Paukščių Tako žvaigždžių populiacijų cheminių savybių apžvalga. Kartu su žvaigždžių kinematinėmis savybėmis, gautomis iš *Gaia* kosminės misijos (Gaia Collaboration et al. 2016), žvaigždžių cheminė sudėtis leidžia analizuoti mūsų Galaktikos bei sudėtinių jos struktūrų, tokių kaip centrinio telkinio, plono ir storo disko, formavimąsi ir evoliuciją.

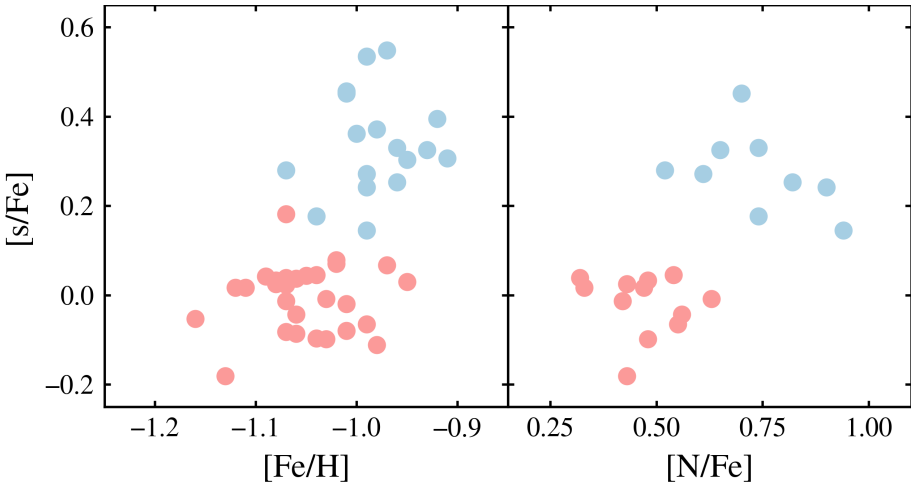
Tyrimo stebėjimai buvo atliekami naudojant Čilėje esančio Europos pietinės observatorijos (ESO) VLT teleskopo FLAMES instrumentą. Nuo 2011 m. gruodžio FLAMES dirba su dviem skirtingais spektrografais: GIRAFFE ir UVES (Pasquini et al. 2002). Abu jie yra daugiaobjektiniai spektrografai. GIRAFFE vienu metu gali stebėti iki 130 objektų, skiriamoji geba nuo R $\sim$ 5000 iki R $\sim$ 30000, o UVES užtikrina didesnę skiriamąją gebą (R $\sim$ 47000), tačiau vienu metu gali stebėti tik 8 objektus. Abiejų spektrografų bangų ilgių segmentai apima visą regimąjį diapazoną.

Šiame darbe naudojame GIRAFFE HR10, HR21 ir HR15N spektrinius segmentus.

2 Poreikis atlikti tikslius kelių cheminių elementų gausų matavimus: dvi žvaigždžių populiacijos kamuoliniame spiečiuje NGC 1851

Čia pateikiame išsamų kamuolinio žvaigždžių spiečiaus NGC 1851 cheminių gausų analizę kurioje 45 raudonųjų milžinių žvaigždėse buvo ištirta iki 29 elementų gausos.

Žvaigždžių atmosferų cheminės sudėties tyrimas atskleidė dvi cheminiu požiūriu skirtingas žvaigždžių subpopuliacijas. Šis skirtumas ryškiausias azoto ir s-proceso elementų (Y, Zr, Ba, La, Ce, Nd) gausose. 1 pav. kairėje pavaizduoti speičiaus žvaigždžių [s/Fe] gausų santykiai, lyginant su [Fe/H] ir dešinėje – lyginant su [N/Fe]. Šiame grafike galima išskirti dvi grupes. Tuo remiantis, mūsų atrinktas žvaigždes skirstome į dvi subpopuliacijas: populiaciją su mažu vidutiniu [s/Fe] santykiu ir populiaciją su dideliu [s/Fe]. Šios dvi grupės taip pat skiriasi vidutine geležies gausa (1 pav. kairėje). Todėl šias dvi grupes skirstome į populiacijas žvaigdžių su mažesne metalų gausa (metal-poor) ir, atitinkamai, didesne metalų gausa (metal-rich). Jų [Fe/H] vertės skiriasi per 0,07 dex. Metalingesnės populiacijos vidutinė [Fe/H] yra $-0{,}98 \pm 0{,}04$ dex, o nemetalingesnės - $-1{,}05 \pm 0{,}05$ dex. Vidutinės s-proceso elementų ir azoto gausos, taip pat kitų elementų rezultatai pateikti 1 lentelėje.



1 pav.: Cheminis spiečiaus žvaigždžių pasiskirstymas. Kairėje: vidutinės s-proceso elementų gausos lyginant su [Fe/H]. Dešinėje: vidutinės s-proceso elementų gausos lyginant su [N/Fe]. Taškų spalvomis pažymėtas dviejų populiacijų atskyrimas (melsvi taškai – metalingesnė populiacija, rausvi taškai – nemetalinga populiacija).

1 lentelė: Vidutinės abiejų populiacijų cheminių elementų gausos ir žvaigždžių skaičius.

Parametras	Nemetalinga			Metalinga		
	Vidurkis	σ	N	Vidurkis	σ	N
[Fe/H]	−1.05	0.05	28	−0.98	0.04	17
A(Li I) $_{\rm 3D\,NLTE}$	0.14	0.66	27	0.46	0.59	17
[C/Fe] (C_2)	−0.31	0.08	14	−0.36	0.11	9
[N/Fe] (CN)	0.47	0.09	12	0.73	0.14	9
[O/Fe] ([O I])	0.27	0.18	28	−0.13	0.20	16
[Na I/Fe] $_{\rm NLTE}$	0.02	0.23	25	0.32	0.17	16
[Mg I/Fe]	0.26	0.08	28	0.25	0.10	17
[Al I/Fe]	0.17	0.19	28	0.38	0.15	17
[Si I/Fe]	0.07	0.05	28	0.09	0.05	17
[Ca I/Fe]	0.17	0.05	28	0.21	0.05	17
[Ca II/Fe]	0.24	0.12	25	0.29	0.15	15
[Ti I/Fe]	0.14	0.07	28	0.16	0.08	17
[Ti II/Fe]	0.17	0.05	28	0.18	0.04	17
[Sc I/Fe]	−0.01	0.12	28	−0.02	0.12	17
[Sc II/Fe]	−0.01	0.04	28	−0.02	0.05	17
[V I/Fe]	−0.11	0.07	28	−0.08	0.08	17
[Cr I/Fe]	−0.16	0.06	28	−0.12	0.07	17
[Cr II/Fe]	0.04	0.11	27	−0.05	0.08	17
[Mn I/Fe]	−0.37	0.05	28	−0.41	0.05	17
[Co I/Fe]	−0.09	0.04	28	−0.07	0.04	17
[Ni I/Fe]	−0.14	0.04	28	−0.15	0.06	17
[Cu I/Fe]	−0.33	0.17	28	−0.28	0.16	17
[Zn I/Fe]	−0.06	0.10	28	0.00	0.15	17
[Y II/Fe]	−0.24	0.06	28	−0.02	0.14	17
[Zr I/Fe]	0.15	0.18	27	0.37	0.15	17
[Mo I/Fe]	0.16	0.16	14	0.37	0.13	12
[Ba II/Fe]	0.10	0.13	28	0.55	0.19	17
[Ba II/Fe] $_{\rm NLTE}$	0.00	0.15	28	0.47	0.20	17
[La II/Fe]	−0.03	0.14	27	0.37	0.10	17
[Ce II/Fe]	−0.18	0.15	28	0.32	0.19	17
[Pr II/Fe]	0.46	0.05	15	0.63	0.12	12
[Nd II/Fe]	0.27	0.07	28	0.52	0.14	17
[Sm II/Fe]	0.37	0.13	14	0.55	0.15	12
[Eu II/Fe]	0.40	0.10	28	0.41	0.07	17
A(C+N+O)	7.97	0.11	12	7.94	0.08	9
C/N	0.66	0.13	12	0.35	0.13	9
[α/Fe]	0.16	0.04	28	0.18	0.05	17
[Iron peak/Fe]	-0.09	0.04	28	-0.07	0.04	17
[*s*/Fe]	−0.01	0.08	28	0.34	0.11	17

3 Konvoliuciniai neuroniniai tinklai

Konvoliucinis neuroninis tinklas (CNN) yra mašininio mokymosi metodas. Mašininio mokymosi metodai - tai sudėtingi, netiesiniai algoritmai, naudojami informacijai iš įvesties duomenų išgauti. Šie algoritmai paprastai turi daug laisvų parametrų, kurių naudotojas nenurodo. Vietoj to algoritmas "išmoksta" parametrų reikšmes per "mokymo" etapą.

3.1 CNN architektūra

Pagrindinės CNN sudedamosios dalys yra jo konvoliucijos sluoksniai. Šie sluoksniai skirti įvesties duomenyse rasti tam tikras ypatybes ir dėsningumus. Rasti spektriniai ypatumai yra pagrindas išvesties etikečių reikšmėms apskaičiuoti. Šį skaičiavimą atlieka tankūs sluoksniai, kurie CNN architektūroje eina po konvoliuciniais sluoksniais. Sudėtingose tinklo architektūrose reikalingas reguliavimas, kad tinklas pernelyg nepritaptų prie mokymo duomenų (over-fitting). 2 lentelėje pateikta mūsų CNN architektūra.

Įvesties sluoksnis gauna neapdorotus įvesties duomenis ir perduoda juos pirmajam konvoliucijos sluoksniui. Prieš pateikiant duomenis, juos reikia paruošti taip, kad jų struktūra būtų vienoda.

Pirmajame konvoliuciniame sluoksnyje įvestas spektras konvoliuojamas su vienu ar keliais konvoliuciniais filtrais. Šių operacijų rezultatas yra parametrų žemėlapiai. Juose parodoma įvairių parametrų padėtis įvesties spektruose. Parametrų žemėlapyje vaizduojamų požymių tipas priklauso nuo naudojamo filtro parametrų. Mokymo etape CNN randa filtrus, kurie užtikrina geriausią mokymo aibės spektrų tikslumą. Konvoliucijos sluoksnis gali turėti kelis filtrus, kurie nepriklausomai sukuria sluoksnio įvesties požymių žemėlapius. Jei CNN turi kelis konvoliucinius sluoksnius, pirmojo sluoksnio požymių žemėlapiai patys bus konvoliuojami kitame konvoliuciniame sluoksnyje.

Paskutinio konvoliucijos sluoksnio išvesties formatą sudaro keli požymių žemėlapiai (t. y. išvestis yra daugiamatė). Tankieji sluoksniai gali apdoroti tik vienmatį įvesties formatą, todėl reikia paruošti paduodant duomenis pirmiesiems tankiesiems sluoksniams vienmačiu pavidalu.

Tankųjį sluoksnį sudaro vienas ar keli neuronai (dar vadinami vie-

2 lentelė: Mūsų CNN architektūra. Paskutinio tankaus sluoksnio išvestys yra mūsų penkių žvaigždžių etikečių (atmosferos parametrų ir elementų gausų) vertės. Sluoksnių eiliškumą ir skaičių, taip pat ir hiperparametrus nustato vartotojas.

Layer	Hyperparameters	Free parameters
Input	Input shape: 8669	
1D convolution	filters: 8, kernel size: 20	168
Activation	LeakyReLU	
1D Max-pooling	pool size: 2	
1D convolution	filters: 6, kernel size: 20	966
Activation	LeakyReLU	
1D Max-pooling	pool size: 2	
1D convolution	filters: 4, kernel size: 20	484
Activation	LeakyReLU	
Flatten		
Dropout	dropout rate: 0.2	
Dense Layer	Neurons = 64	546 368
Dropout	dropout rate: 0.2	
Activaton	LeakyReLU	
Dense Layer	Neurons: 128	8320
Dropout	dropout rate: 0.2	
Activaton	LeakyReLU	
Dense Layer	Neurons: 32	4128
Dropout	dropout rate: 0.2	
Activaton	LeakyReLU	
Dense layer	Neurons: 5	165
Activation	Linear	
		Total: 560 599

netais arba mazgais). Kiekvienas neuronas gauna kelis įvesties kanalus iš ankstesnio sluoksnio (plokščiojo sluoksnio arba ankstesnio tankaus sluoksnio). Vieno neurono viduje apskaičiuojama visų įvesties verčių tiesinė kombinacija ir pridedama nuokrypio vertė. Rezultatas transformuojamas aktyvacijos funkcija ir perduodamas kaip neurono išvestis į kitą sluoksnį.

Norint išvengti tinklo permokymo, galima naudoti reguliarizavimo metodus. Permokymas pasireiškia tada, kai tinklas numato tikslias mokomosios aibės etiketes, bet prastai veikia dirbant su kitais spektrais. Šiame darbe CNN architektūrą papildome dviejų tipų reguliarizavimo sluoksniais. Tai maksimalaus kaupimo (max-pooling) ir išmetimo (dropout) sluoksniai. Max-pooling taikomas konvoliucijos sluoksnių išvesties požymių žemėlapiams. Jis veikia padalijant požymių žemėlapį į vienodo pločio langus. Kiekviename iš šių langų išsaugomas tik didžiausios vertės, o kitos vertės atmetamos. Tinklo architektūroje išmetimo sluoksniai yra tarp dviejų tankių sluoksnių. Išmetimo sluoksnis deaktyvuoja dalį ankstesnio sluoksnio neuronų, nustatydamas jų išvesties vertes lygiomis nuliui (jie "išmetami").

Išvesties sluoksnis yra galutinis tankus sluoksnis, kurio neuronų skaičius yra lygus prognozuojamų etikečių skaičiui. Šių neuronų išvestys yra tiesiogiai proporcingos nustatytoms etiketėms.

3.2 Tinklo mokymas

Sudarant tinklo architektūrą, nežinomos konvoliucijos filtro branduolių, tankiųjų sluoksnių svorių ir šališkumo reikšmės (bendrai vadinamos "laisvaisiais parametrais"). Parametrų, kurių dėka gaunamas geriausias tinklo rezultatas, paieška vadinama *tinklo mokymu*. Tinklo išvestį laikome optimalia, kai skirtumas tarp iš anksto nustatytų validavimo spektrų rinkinio etikečių ir CNN išvesties tiems patiems spektrams yra minimalus. Sėkmingai apmokius tinklą, reikia patikrinti jo veikimą su naujais spektrais. Tam reikia sukurti testinį pažymėtų spektrų, kurie nebuvo naudojami tinklo mokymui, rinkinį.

4 GES iDR6 spektrų parametrų nustatymas naudojant CNN

4.1 Mokymo, validavimo ir testavimo rinkiniai

Mūsų duomenų rinkinį sudaro GES iDR6 spektrai, atitinkamų žvaigždžių parametrai ir cheminių elementų gausos. Spektrai, kuriuos naudojome šiame tyrime, buvo gauti naudojant GIRAFFE spektrografą, apimantį regimąjį 370-900 nm bangų ilgių diapazoną. Visgi GIRAFFE spektrografu visas 370-900 nm nestebimas, o parenkami tyrimui reikalingi specifiniai mažesni regionai. Šiame tyrime naudoti HR10 (533,9-561,9 nm, R = 19800) ir HR21 (848,4-900,1 nm, R = 16200) regionai, nes jie apima svarbias Mg ir Al sugerties linijas.

Siekdami sukurti mokymo rinkinį, atlikome keletą kokybės patikrinimų, kad užtikrintume, jog mūsų tinklas bus apmokytas naudojant aukštos kokybės duomenis. Spektrai, kurių signalo ir triukšmo santykis S/N < 30 ir atmosferos parametrų bei elementų gausų paklaidos ($e T_{\rm eff}$ > 200 K, $e\log(g)$ > 0.3 dex, eA(elementas) > 0,2 dex) buvo atmesti, taip pat atmesti spektrai, kurie buvo pažymėti TECH arba PECULI žymomis, o taip pat tokios žvaigždėws, kurių sukimosi greičiai > 20 km s^{-1}.

Siekiant ištirti visų spektrų panašumą, taikyta (t-SNE - distributed stochastic neighbor embedding) analizė. t-SNE analizė yra populiarus nekontroliuojamas mašininio mokymosi metodas, naudojamas vidiniams ryšiams ir panašumams didelės dimensijos duomenų rinkiniuose atvaizduoti. Tai atliekama kiekvienam duomenų taškui suteikiant vietą dvimačiame arba trimačiame panašumo žemėlapyje (van der Maaten & Hinton 2008). Skirtingos žvaigždžių atmosferų fizikinės savybės atsispindi jų spektrų formose, kurios savo ruožtu lemia jų vietą t-SNE žemėlapyje. Ryšį tarp fizikinių parametrų ir spektrinių savybių mūsų CNN išmoksta mokymo etape.

Kiekvienas apmokymo spektras turi susijusių žvaigždžių etikečių rinkinį. Mūsų atveju tai yra du atmosferos parametrai $T_{\rm eff}$ ir $\log(g)$, bei cheminės gausos [Mg/Fe], [Al/Fe] ir [Fe/H]. GES iDR6 duomenų rinkinyje elementų gausos pateikiamos kaip absoliučiosios gausos vertės A(Elementas). Apskaičiavome [Fe/H] ir [Elementas/Fe] taip: $[\mathrm{Fe/H}] =$

$A(Fe)_{star} - A(Fe)_{\odot}$ ir $[Element/Fe] = A(Element)_{star} - A(Element)_{\odot} - [Fe/H]$.

Pritaikius visus šiuos apribojimus, liko 14 634 spektrai, kurie turėjo aukštos kokybės atmosferos parametrų ir elementų gausų etiketes. Šie spektrai atsitiktine tvarka priskiriami mokymo arba tikrinimo rinkiniui. Šių dviejų rinkinių spektrų etikečių intervalas yra toks: efektinė temperatūra kinta nuo $T_{\rm eff}$ = 4000 - 6987 K, laisvo kritimo pagreitis $\log(g)$ yra nuo 1,08 iki 4,87 dex, o [Fe/H] kinta nuo $-1,53$ iki 0,72 dex. [Mg/Fe] vertės svyruoja nuo $-0,25$ iki 0,80 dex, o [Al/Fe] vertės labai išsibarščiusios, beveik per 2 dex: nuo $-0,95$ iki 1,00 dex.

Be mokymo ir validavimo rinkinių, sudarėme ir testavimo rinkinį. Šis rinkinys naudojamas mūsų CNN veikimui patikrinti naudojant spektrus, kurie nebuvo naudojami mokymo procese. Taip galime imituoti mūsų CNN metodo taikymą naujai stebėtiems spektrams, kurie dar nebuvo analizuoti spektroskopiškai. Todėl visą testavimo rinkinį sudaro spektrai be jokių taikomų kokybės apribojimų ir jis apima platesnius S/N ir etikečių diapazonus nei mokymo rinkinys. Mūsų tinklas negali patikimai įvertinti spektrų, kurie nepatenka į mokomojo rinkinio ribas. Todėl turime rasti būdą, kaip identifikuoti spektrus, kurie yra panašūs į mūsų mokomuosius spektrus, nes tikėtina, kad šių spektrų etiketės pateks į mokomosios aibės ribas. Jau įrodėme, kad t-SNE gali parodyti spektrų panašumą. Todėl naudojame t-SNE, kad testavimo rinkinyje nustatytume tuos spektrus, kurių etiketės greičiausiai patenka į mokymo rinkinio ribas. Šio metodo atvaizdavimas pateiktas 2 pav. Viršutiniame kairiajame paveikslėlyje parodytas t-SNE žemėlapis, kuris buvo apskaičiuotas visiems mūsų duomenų rinkinio spektrams. Sudarę t-SNE projekciją, žemėlapyje nustatėme mokymo rinkinio spektrus. Viršutiniame dešiniajame paveikslėlyje parodyti tie testinio rinkinio spektrai, kurie t-SNE žemėlapyje užima tą patį plotą kaip ir kairiajame paveikslėlyje pateikti mokomieji spektrai. Bandomieji spektrai, kurie žemėlapyje yra artimi mokomiesiems spektrams, yra panašūs į mokomuosius spektrus. Šį viršutiniame dešiniajame paveikslėlyje esantį rinkinį vadiname "vidiniu" testavimo rinkiniu. Galiausiai apatiniame paveikslėlyje rodomi tie spektrai, kurie nėra panašūs į mokomuosius spektrus. Šis spektrų rinkinys yra mūsų "išorinis" testavimo rinkinys, kurį naudojame savo tinklui išbandyti spektrams, kurie nepanašūs į mo-

kymo rinkinio spektrus.

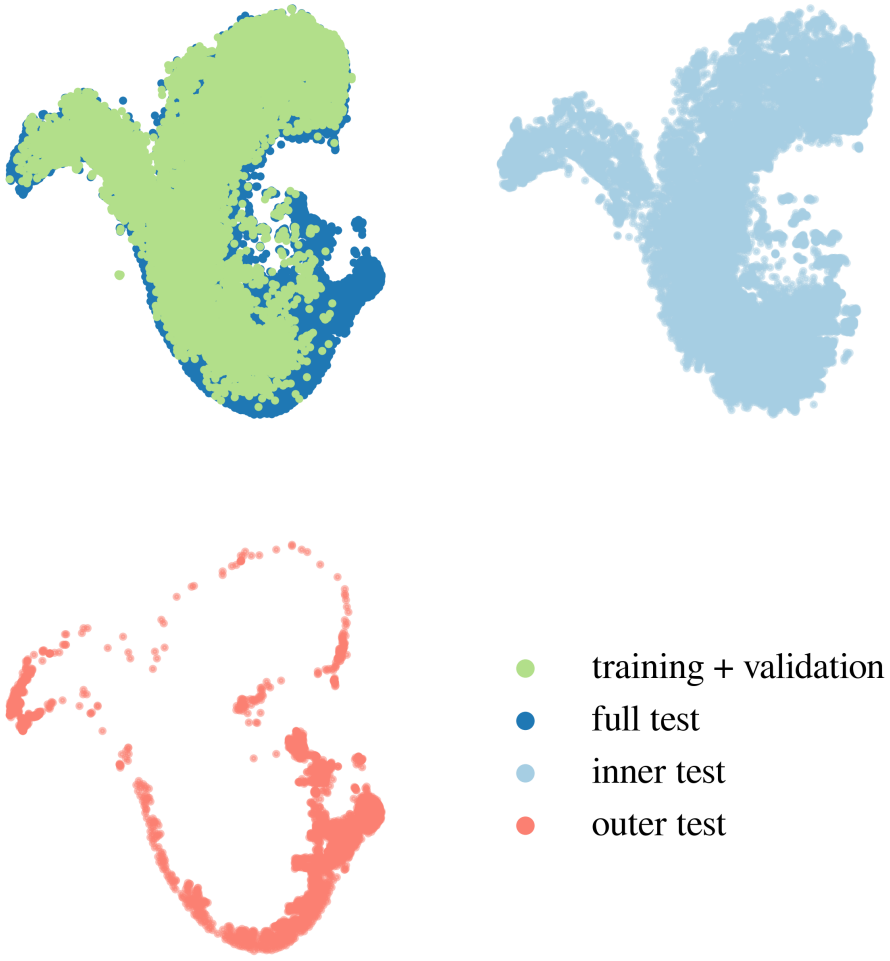
4.2 Mokymo etapas ir etikečių prognozės testavimo rinkiniui

Vidutiniškai vienas mokymo ciklas truko 159 epochas, o jo trukmė buvo $\sim$45 minutės[5]. Galutinės apmokyto tinklo modelio parametrų vertės, taigi ir tinklo išvestis, šiek tiek skiriasi kiekvieną kartą, kai tinklas yra mokomas. Norėdami atsižvelgti į šį svyravimą, atlikome dešimt mokymų ir užfiksavome kiekvieno iš jų rezultatus. Mokymo etapo pabaigoje pašalinome du CNN modelius su didžiausiais likusiais validavimo nuostoliais. Likę aštuoni CNN modeliai buvo naudojami spektrų, esančių mokymo, validavimo ir testavimo rinkiniuose, etiketėms prognozuoti. Čia pateikiame aštuonių etikečių rinkinių vidurkius kaip mūsų rezultatus.

Tiek mokymo, tiek validavimo rinkinyje GES matavimai ir CNN prognozės gerai dera su visomis etiketėmis. Mokymo ir validavimo rinkinių prognozės rodo tokį patį nuokrypį (jei toks yra) ir nedidelę dispersiją apie santykį 1:1. Tai rodo, kad tinklas gerai veikia su spektrais, su kuriais jis nebuvo tiesiogiai apmokytas, ir nėra permokytas. Dispersija aplink 1:1 santykį yra vientisa daugumoje intervalų tarp visų penkių etikečių reikšmių. Mokymo rinkinio dispersiją naudojame kaip mūsų tinklo mokymo tikslumo matą: $T_{\rm eff}$ mokymo tikslumas yra 37 K, $\log(g)$ - 0,06 dex, [Mg/Fe] - 0,05 dex, [Al/Fe] - 0,06 dex, [Fe/H] - 0,04 dex. Tiesa, mūsų CNN netiksliai atkuria didžiausias ir mažiausias GES verte.

Siekdami įvertinti mūsų tinklo gebėjimą parametrizuoti naujus spektrus, kurie visiškai nedalyvavo mokymo procese, lyginame GES įvesties etiketes su CNN prognozėmis trims skirtingiems testavmo rinkiniams. Viršutinėje 3 paveikslo eilutėje parodytas GES įvesties ir CNN išvesties palyginimas vidiniam testavimo rinkinio spektrui, kurio S/N $\geq$ 30. Šiame porinkinyje 90% GES etikečių yra mokymo ribose. Dauguma likusių 10% spektrų yra už [Mg/Fe] ir [Al/Fe] ribų. To priežastis yra mūsų vidinio testavimo rinkinio paieškos būdas. Šį rinkinį sudaro tik tos žvaigždės, kurios t-SNE žemėlapyje užima tą pačią sritį kaip ir mokomieji spektrai (2 pav.). Spektro forma, taigi ir jo padėtis t-SNE žemėlapyje, labai priklauso nuo etikečių $T_{\rm eff}$, $\log(g)$ ir [Fe/H], tuo tarpu

[5]Staliniame kompiuteryje, naudojant tik CPU (Intel Core i7-9700 CPU @ 3.00 GHz $\times$ 8)

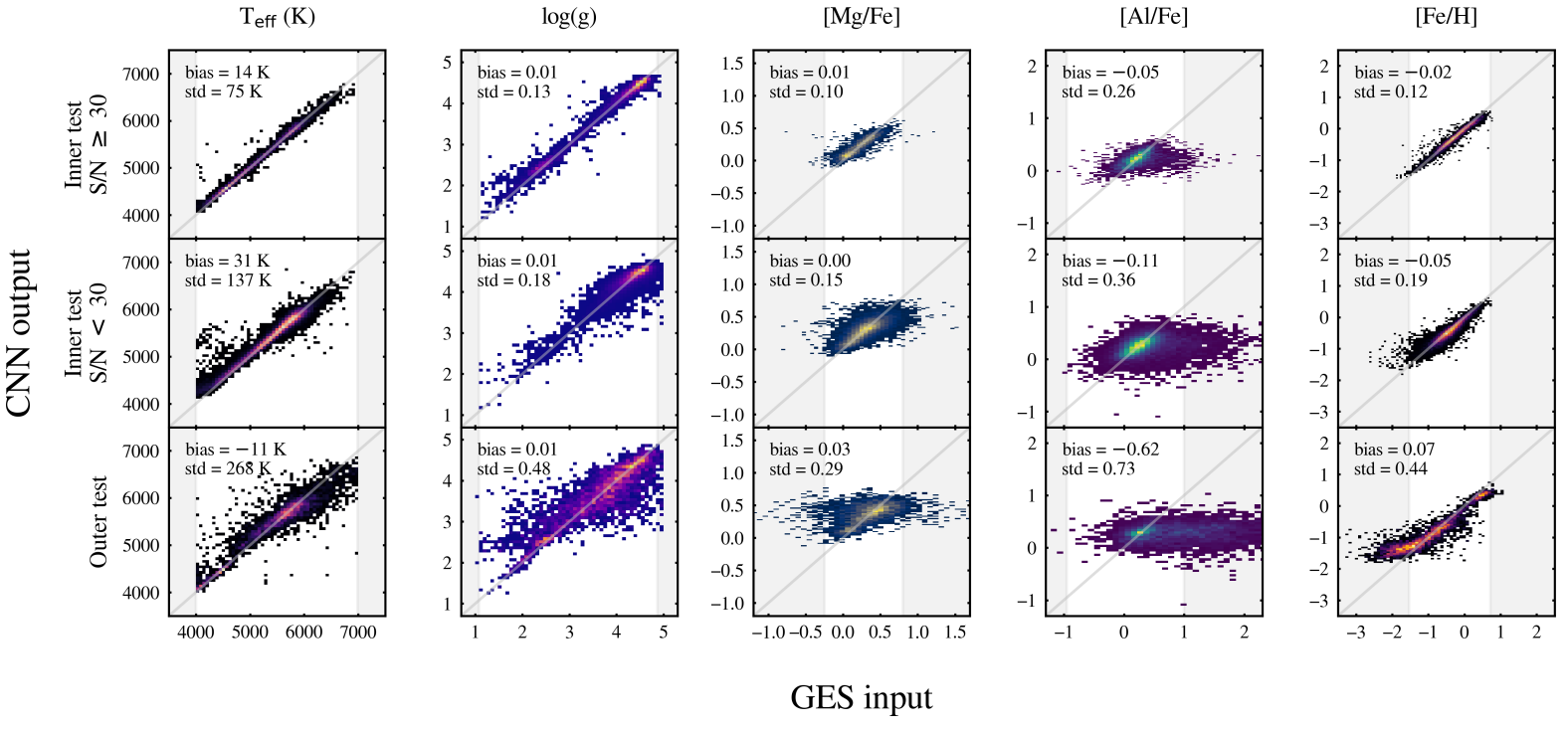


2 pav.: Viršutiniame kairiajame paveiksle: visų GIRAFFE duomenų rinkinio spektrų (tamsiai mėlyna spalva) t-SNE žemėlapis su mokymo rinkinio spektrais (žalia spalva). Itin nutolę taškai buvo pašalinti. Viršutinis dešinysis paveikslas: "vidinis" testavimo rinkinys, apibrėžiamas kaip testavimo spektrų, kurie t-SNE žemėlapyje apima tą pačią sritį kaip ir mokymo rinkinio spektrai, poaibis. Apatinis kairysis paveikslas: testavimo rinkinio spektrai, kurie neužima tos pačios žemėlapio srities kaip mokymo rinkinys. Tai mūsų "išorinis" testavimo rinkinys.

[Al/Fe] pokyčiai bendram spektrui turi tik nedidelę įtaką. Tas pats pasakytina ir apie [Mg/Fe], tačiau mažesniu mastu. Taip yra todėl, kad mūsų imties spektruose (Heiter et al. 2021b) yra daugiau Mg sugerties linijų nei Al linijų. Tinklo prognozių tikslumas pradeda mažėti, kai vidinis bandymų rinkinys turi mažą S/N (3 pav. vidurinė eilutė). Šiame rinkinyje yra spektrų, kurie yra panašūs į mokymams skirtus spektrus, tačiau jų S/N yra mažesnis (primename, kad mažiausias mokymo spektrų S/N yra 30). Mažo S/N vidiniame rinkinyje yra daugiau spektrų, kurie nepatenka į mokymo ribas. Akivaizdu, kad mūsų CNN negali tiksliai parametrizuoti spektrų, kurių GES etiketės yra už mokomosios aibės ribų. 3 pav. apatinėje eilutėje pateikti išorinio testavimo rinkinio rezultatai. Tinklo prognozės šiam rinkiniui yra vis mažiau tikslios, net ir spektrams, esantiems mokymo rinkinio ribose. Skirtumas tarp GES įvesties ir CNN išvesties išorinėje testavimo aibėje labiausiai pastebimas [Al/Fe] ir [Fe/H] srityse, kur mūsų tinklas netiksliai prognozuoja itin mažas ir dideles GES etiketes.

Mūsų rezultatų vidinę neapibrėžtį apibūdiname kaip aštuonių CNN modelių etikečių prognozių dispersiją. Vidutinės etikečių prognozių neapibrėžtys vidiniame bandymų rinkinyje su aukštu S/N yra mažos: 27 K - $T_{\rm eff}$, 0,04 - log(g) ir 0,03 dex - [Mg/Fe], [Al/Fe] ir [Fe/H]. Šio rinkinio GES etikečių paklaidos beveik nepriklauso nuo absoliučios etikečių vertės ir S/N. Jų vidutinės vertės yra 63 K ($T_{\rm eff}$), 0,15 (log(g), 0,22 [Mg/Fe], 0,19 [Al/Fe] ir 0,18 [Fe/H].

CNN prognozės su didele neapibrėžtimi vienai etiketei taip pat pasižymi didele neapibrėžtimi visoms kitoms etiketėms, o tikslios prognozės yra tikslios visoms penkioms etiketėms. Vidinis mūsų $T_{\rm eff}$ ir log(g) tikslumas yra didžiausias ten, kur mokymo rinkinio tankis yra didžiausias. $T_{\rm eff}$ atveju taip yra tarp $\sim$ 4500 ir 5775 K, o log(g) - tarp $\sim$ 2,5 ir 4,5 dex. Čia šių dviejų etikečių prognozių neapibrėžtys yra mažiausios. Išskyrus [Fe/H], cheminių elementų gausos prognozių tikslumas nerodo aiškių tendencijų, susijusių su absoliutine etiketės verte. Kalbant apie [Fe/H], neapibrėžtis didėja mažesnėms [Fe/H] gausos. Tai, tikriausiai, susiję su tuo, kad nemetalingų žvaigždžių santykinai yra gerokai mažiau lyginant su visa imtimi. Nustatėme, kad mažėjant spektrų S/N mažėja visų penkių etikečių prognozių neapibrėžtys.



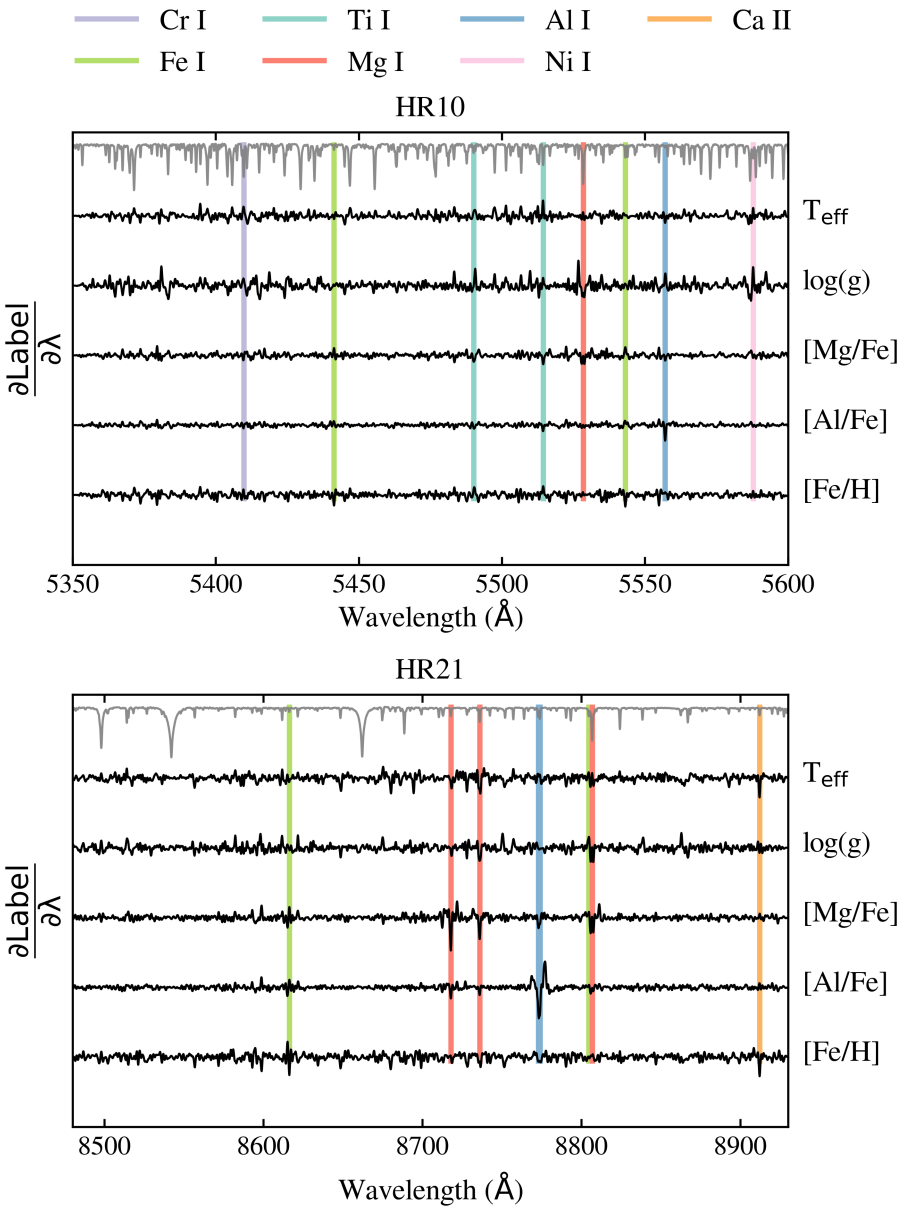
3 pav.: GES iDR6 etikečių ir mūsų CNN prognozuojamų reikšmių palyginimas. Trijose eilutėse pateikiami vidinio testavimo rinkinio rezultatai dviems skirtingiems S/N, ir išorinio testavimo rinkinio rezultatai. Kiekviename stulpelyje pateikiami skirtingų etikečių rezultatai. Kiekviename stulpelyje horizontalioji ašis reiškia GES įvesties etiketes, vertikalioji - mūsų CNN prognozuojamas etiketes. Šviesesnės spalvos rodo didesnį duomenų taškų tankį (tiesinė spalvų skalė). Kiekviename grafike pateikiamas rezultatų vidutinis nuokrypis ir standartinis nuokrypis (sklaida) aplink 1:1 santykį. Ištisinės linijos rodo 1:1 santykį. Šešėliu pažymėtos sritys rodo regionus, kurie yra už mokymo etikečių srities ribų.

4.3 Ko mokosi tinklas?

Norėdami vizualizuoti spektro sritis, kur mūsų CNN yra aktyvus tam tikros etiketės atžvilgiu, sudarėme jautrumo žemėlapius naudodami "GradientTape" iš "TensorFlow". Apskritai, mūsų CNN elgiasi kaip funkcija, kuri įvesties pikselius žvaigždžių spektruose atvaizduoja į išvesties etikečių reikšmes. Šis atvaizdavimas yra diferencijuojamas, o tai reiškia, kad galima kiekybiškai įvertinti kiekvieno įvesties pikselio įtaką gautoms etikečių prognozėms. Naudojant automatinį diferencijavimą, galima apskaičiuoti gradientus, ∂ Label / ∂ λ, t. y. CNN jautrumą kiekvienam pikseliui kiekvienai etiketei. Didelė absoliutinė gradiento vertė bangų ilgių diapozone reiškia, kad tinklo prognozė tam tikrai etiketei yra labai jautri srauto pokyčiams būtent tame diapozone. 4 pav. pavaizduoti tinklo gradientai mūsų penkioms etiketėms visame įvesties spektrų bangų ilgių diapazone. Gradientai yra atsitiktinai išsibarstę apie nulį beveik visame spektre. Tik tam tikruose bangos ilgių intervaluose tinklas yra jautrus srauto pokyčiams. Čia gradientai rodo atskirus siaurus šuolius. Tai ypač akivaizdu [Mg/Fe] ir [Al/Fe] gradientuose mūsų spektrų HR21 segmente. [Mg/Fe] gradientai rodo du aiškius šuolius ties 8736,0 ir 8806,8 Å. Tai yra dviejų Mg I sugerties linijų pozicijos. [Al/Fe] gradientų pikas žymi Al I dubleto poziciją ties $\sim$8773 Å. Taigi matome, kad mūsų tinklas gali nustatyti absorbcijos linijas įvesties spektruose. Neigiamos gradiento reikšmės šiuose bangos ilgiuose reiškia, kad jei absorbcijos linijų srautas yra mažas, prognozuojama gausa yra didelė, ir atvirkščiai. Tai intuityviai parodo, kad gilesnės absorbcijos linijos spektruose rodo didesnes elementų gausas žvaigždžių atmosferose.

4.4 Paukščių Tako charakteristikų analizė pagal CNN prognozes

Į GES iDR6 duomenų rinkinį įtrauktos žvaigždės, priklausančios keliems kamuoliniams spiečiams. Mūsų CNN rezultatai leidžia atkurti Mg-Al antikoreliaciją NGC 6752, NGC 6218 ir NGC 1851 spiečiuose. Tokia antikoreliacija naudojama kamuolinių spiečių cheminei evoliucijai tirti (Pancino et al. 2017c). Spiečių vidutinės [Fe/H] vertės apima $\sim$0,5 dex intervalą. Matome, kad GES įvesties ir CNN išvesties sutapimas gerėja didėjant spiečiaus [Fe/H]. Spiečiuje NGC 6752 yra



4 pav.: Penkių etikečių tinklo gradientai pagal bangos ilgį (juoda spalva). Viršuje pavaizduoti GIRAFFE HR10 gradientai, o apačioje - HR21 gradientai. Viršutinė pilka linija vaizduoja vidutinį įvesties spektrą. Pasirinktų skirtingų elementų sugerties linijų vietos pažymėtos vertikaliomis spalvotomis linijomis. Pažymėtos Mg ir Al linijos buvo naudojamos GES nustatant įvesties Mg ir Al gausas. Jų bangos ilgiai yra 5528,41, 8717,81, 8736,02 ir 8806,75 Å Mg ir 5557,06, 8772,87 ir 8773,90 Å Al (Heiter et al. 2021b).

žvaigždžių, kurių [Fe/H] vertės yra ties apatine mokymo rinkinio riba, kur mokymo spektrų tankis yra mažas. Mokymo rinkinio tankis didėja didėjant [Fe/H], todėl CNN geriau prognozuoja metalingesnių spiečių žvaigždžių savybes. Išskyrus dvi žvaigždes su mažiausiu [Al/Fe], visos CNN prognozės NGC 6218 spiečiuje sutampa su GES rezultatais, neviršijant nurodytų paklaidų. NGC 1851, kurio vidutinė [Fe/H] vertė yra didžiausia iš visų mūsų spiečių, Mg-Al antikoreliacija stebima tiek GES tiek ir mūsų metodo prognozėse.

[Mg/Fe] vertės gali būti naudojamos skirstant Paukščių Tako žvaigždes į plono ir storo disko populiacijas. Šį skirstymą atlikome remdamiesi CNN rezultatais, gautais vidiniam stebėjimų rinkiniui su S/N ≥ 30, kartu su mokymo ir testavimo rinkiniais. Storojo ir plonojo disko žvaigždėms nustatyti naudojome klasterizavimo algoritmą HDBSCAN (Campello et al. 2013), kuris įgyvendintas "Python" programavimo bibliotekoje *hdbscan*. Šis algoritmas priskiria duomenų taškus skirtingoms grupėms, priklausomai nuo duomenų taškų tankio pasiskirstymo. Išskirtos dvi grupės, atitinkančios dvi žvaigždžių populiacijas.

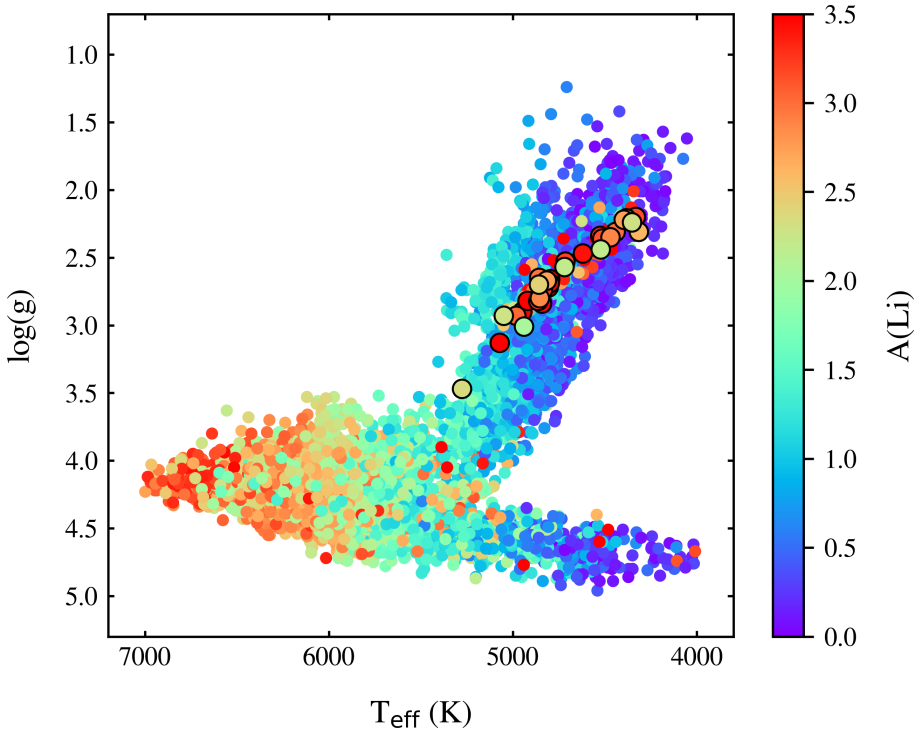
Norėdami ištirti abiejų populiacijų amžiaus pasiskirstymą, naudojome programą *A Unified tool to estimate Distances, Ages and Masses* (UniDAM), kuri gali įvertinti amžių naudojant izochronas. Gautas vidutinis plonojo disko žvaigždžių amžius yra 8,7 mlrd. m., o storojo disko - 9,7 mlrd. m. Toks amžiaus skirtumas tarp abiejų populiacijų nustatytas atlikus ne vieną tyrimą ir naudojant keletą amžiaus nustatymo metodų.

Taip pat tyrėme plono ir storo disko populiacijų kinematines savybes. Remdamiesi dabartinėmis žvaigždžių padėtimis ir greičiais, integravome jų orbitas 5 mlrd. m. laikotarpiui naudojant teorinį Paukščių Tako potencialą, pasitelkę Python pagrindu sukurtą galaktikos dinamikos paketą *galpy* (Bovy 2015). Integravimui naudojome gravitacinį potencialą *MWPotencial2014*, kuris apjungia centrinio telkinio, disko ir halo potencialus. Žvaigždžių savieji judėjimai, dangaus koordinatės ir paralaksai paimti iš Gaia EDR3. Tiesinės regresijos modelis rodo, kad storojo disko orbitų ekscentricitetas e mažėja didėjant [Fe/H]: $\Delta e/\Delta[\mathrm{Fe/H}] = -0{,}25$. Plono disko žvaigždžių orbitų ekscentricitetas yra vidutiniškai mažesnis už storo disko ir turi nedidelę teigiamą tendenciją ($\Delta e/\Delta[\mathrm{Fe/H}] = 0{,}02$). Šie rezultatai sutampa su Yan et al. (2019) išvadomis. Jie tyrė plono ir storo disko žvaigždžių chemines ir kinematines savybes

remiantis LAMOST duomenų rinkiniu (Zhao et al. 2012).

4.5 31-a nauja ličiu praturtinta raudonoji milžinė

Šiame darbe rasta 31-a anksčiau neidentifikuota Li praturtinta raudonoji milžinė, pasinaudojant mūsų CNN prognozių rezultatais. Šios 31 naujos ličiu praturtintos žvaigždės yra dalis Nepal et al. (2023) testavimo rinkinio. Joms trūksta vienos ar kelių etikečių ($T_{\rm eff}$, log(g), [M/H] arba A(Li)) Gaia-ESO iDR6 duomenų rinkinyje. Visos šios milžinės turi A(Li) > 2 dex, log(g) < 3,5, o $T_{\rm eff}$ < 5500 K (5 pav.). Šių spektrų signalo-triukšmo santykis yra > 25, o jų etikečių prognozių vidinės CNN neapibrėžtys yra < 50 K $T_{\rm eff}$ ir < 0,1 dex log(g), [M/H] ir A(Li). Nė viena iš šių žvaigždžių nebuvo identifikuota kaip Li praturtinta milžinė Gaia-ESO darbuose. GALAH tyrimo kataloge Martell et al. (2021) taip pat neidentifikavo nė vieno iš mūsų naujai aptiktų objektų kaip Li praturtintų milžinių. Kylio ($T_{\rm eff}$-log(g)) diagramos milžinių šakoje yra 67 Li praturtintos žvaigždždės (A(Li) > 2 dex), 38 iš jų įtrauktos į mokymo rinkinį.



5 pav.: Kylio diagrama (T_{eff}-log(g)), sudaryta iš Nepal et al. (2023). 31 aptiktos Li praturtintos milžinės pažymėtos didesniais apskritimais. Taškų spalvos žymi ličio gausas.

References

Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, software available from tensorflow.org

Adibekyan, V. Z., Santos, N. C., Sousa, S. G., & Israelian, G. 2011, A new α-enhanced super-solar metallicity population, Astron. Astrophys., 535, L11

Amarsi, A. M., Lind, K., Osorio, Y., et al. 2020, The GALAH Survey: non-LTE departure coefficients for large spectroscopic surveys, Astron. Astrophys., 642, A62

Ambrosch, M., Guiglion, G., Mikolaitis, Š., et al. 2023, The Gaia-ESO Survey: Chemical evolution of Mg and Al in the Milky Way with machine learning, Astron. Astrophys., 672, A46

Anders, F., Chiappini, C., Santiago, B. X., et al. 2018, Dissecting stellar chemical abundance space with t-SNE, Astron. Astrophys., 619, A125

Asplund, M. 2005, New Light on Stellar Abundance Analyses: Departures from LTE and Homogeneity, Annual Review of Astronomy and Astrophysics, 43, 481

Bailer-Jones, C. A. L. 1997, Neural network classification of stellar spectra, The Observatory, 117, 250

Bastian, N. & Lardo, C. 2018, Multiple Stellar Populations in Globular Clusters, Annual Review of Astronomy and Astrophysics, 56, 83

Baumgardt, H. & Hilker, M. 2018, A catalogue of masses, structural parameters, and velocity dispersion profiles of 112 Milky Way globular clusters, Mon. Not. R. Astron. Soc., 478, 1520

Bergemann, M., Collet, R., Amarsi, A. M., et al. 2017, Non-local Thermodynamic Equilibrium Stellar Spectroscopy with 1D and <3D> Models. I. Methods and Application to Magnesium Abundances in Standard Stars, Astrophys. J., 847, 15

Bertin, E. & Arnouts, S. 1996, SExtractor: Software for source extraction., Astron. Astrophys. Suppl., 117, 393

Bialopetravičius, J. & Narbutis, D. 2020, Study of Star Clusters in the M83 Galaxy with a Convolutional Neural Network, Astron. J., 160, 264

Bishop, C. M. et al. 1995, Neural networks for pattern recognition (Oxford university press)

Bovy, J. 2015, galpy: A python Library for Galactic Dynamics, Astrophys. J. Suppl. Ser., 216, 29

Bressan, A., Marigo, P., Girardi, L., et al. 2012, PARSEC: stellar tracks and isochrones with the PAdova and TRieste Stellar Evolution Code, Mon. Not. R. Astron. Soc., 427, 127

Cameron, A. G. W. & Fowler, W. A. 1971, Lithium and the s-PROCESS in Red-Giant Stars, Astrophys. J., 164, 111

Campello, R. J. G. B., Moulavi, D., & Sander, J. 2013, in Advances in Knowledge Discovery and Data Mining, ed. J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Berlin, Heidelberg: Springer Berlin Heidelberg), 160–172

Carretta, E., Gratton, R. G., Lucatello, S., et al. 2010, Abundances for a Large Sample of Red Giants in NGC 1851: Hints for a Merger of Two Clusters?, Astrophys. J. Lett., 722, L1

Casey, A. R., Ruchti, G., Masseron, T., et al. 2016, The Gaia-ESO Survey: revisiting the Li-rich giant problem, Mon. Not. R. Astron. Soc., 461, 3336

Chollet, F. et al. 2015, Keras, https://github.com/fchollet/keras

Dalton, G., Trager, S., Abrams, D. C., et al. 2018, in Ground-based and Airborne Instrumentation for Astronomy VII, ed. C. J. Evans, L. Simard, & H. Takami, Vol. 10702, International Society for Optics and Photonics (SPIE), 388 – 397

de Jong, R. S., Agertz, O., Berbel, A. A., et al. 2019, 4MOST: Project overview and information for the First Call for Proposals, The Messenger, 175, 3

Dinescu, D. I., Girard, T. M., & van Altena, W. F. 1999, Space Velocities of Globular Clusters. III. Cluster Orbits and Halo Substructure, Astron. J., 117, 1792

Fabbro, S., Venn, K. A., O'Briain, T., et al. 2018, An application of deep learning in the analysis of stellar spectra, Mon. Not. R. Astron. Soc., 475, 2978

Franciosini, E., Randich, S., de Laverny, P., et al. 2022, The Gaia-ESO Survey: Lithium measurements and new curves of growth, Astron. Astrophys., 668, A49

Fuhrmann, K. 1998, Nearby stars of the Galactic disk and halo, Astron. Astrophys., 338, 161

Gaia Collaboration. 2020, VizieR Online Data Catalog: Gaia EDR3 (Gaia Collaboration, 2020), VizieR Online Data Catalog, I/350

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, The Gaia mission, Astron. Astrophys., 595, A1

Gilmore, G., Randich, S., Asplund, M., et al. 2012, The Gaia-ESO Public Spectroscopic Survey, The Messenger, 147, 25

Gilmore, G., Randich, S., Worley, C. C., et al. 2022, The Gaia-ESO Public Spectroscopic Survey: Motivation, implementation, GIRAFFE data processing, analysis, and final data products, Astron. Astrophys., 666, A120

Gratton, R. G., Carretta, E., Matteucci, F., & Sneden, C. 2000, Abundances of light elements in metal-poor stars. IV. [Fe/O] and [Fe/Mg] ratios and the history of star formation in the solar neighborhood, Astron. Astrophys., 358, 671

Gratton, R. G., Villanova, S., Lucatello, S., et al. 2012, Spectroscopic analysis of the two subgiant branches of the globular cluster NGC 1851, Astron. Astrophys., 544, A12

Gray, D. F. 2005, The Observation and Analysis of Stellar Photospheres, 3rd edn. (Cambridge University Press)

Grevesse, N., Asplund, M., & Sauval, A. J. 2007, The Solar Chemical Composition, Space Sci. Rev., 130, 105

Guiglion, G., Matijevič, G., Queiroz, A. B. A., et al. 2020, The RAdial Velocity Experiment (RAVE): Parameterisation of RAVE spectra based on convolutional neural networks, Astron. Astrophys., 644, A168

Guiglion, G., Nepal, S., Chiappini, C., et al. 2024, 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., 682, A9

Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, A grid of MARCS model atmospheres for late-type stars. I. Methods and general properties, Astron. Astrophys., 486, 951

Haar, L. V., Elvira, T., & Ochoa, O. 2023, An analysis of explainability methods for convolutional neural networks, Engineering Applications of Artificial Intelligence, 117, 105606

Heiter, U., Jofré, P., Gustafsson, B., et al. 2015, Gaia FGK benchmark stars: Effective temperatures and surface gravities, Astron. Astrophys., 582, A49

Heiter, U., Lind, K., Bergemann, M., et al. 2021a, Atomic data for the Gaia-ESO Survey, Astron. Astrophys., 645, A106

Heiter, U., Lind, K., Bergemann, M., et al. 2021b, Atomic data for the Gaia-ESO Survey, Astron. Astrophys., 645, A106

Hong-liang, Y. & Jian-rong, S. 2022, A Review to the Studies of Lithium-Rich Giants, Chinese Astronomy and Astrophysics, 46, 1

Jackson, R. J., Jeffries, R. D., Lewis, J., et al. 2015, The Gaia-ESO Survey: Empirical determination of the precision of stellar radial velocities and projected rotation velocities, Astron. Astrophys., 580, A75

Jofré, P., Heiter, U., & Soubiran, C. 2019, Accuracy and Precision of Industrial Stellar Abundances, Annual Review of Astronomy and Astrophysics, 57, 571

Johansson, S., Litzén, U., Lundberg, H., & Zhang, Z. 2003, Experimental f-Value and Isotopic Structure for the Ni I Line Blended with [O I] at 6300 Å, Astrophys. J. Lett., 584, L107

Kendall, A. & Gal, Y. 2017, in Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17 (Red Hook, NY, USA: Curran Associates Inc.), 5580–5590

Kilic, M., Munn, J. A., Harris, H. C., et al. 2017, The Ages of the Thin Disk, Thick Disk, and the Halo from Nearby White Dwarfs, The Astrophysical Journal, 837, 162

Korotin, S. A., Andrievsky, S. M., Luck, R. E., et al. 2014, Oxygen abundance distribution in the Galactic disc, Mon. Not. R. Astron. Soc., 444, 3301

Kovalev, M., Bergemann, M., Ting, Y.-S., & Rix, H.-W. 2019, Non-LTE chemical abundances in Galactic open and globular clusters, Astron. Astrophys., 628, A54

Kurucz, R. L. 2005, New atlases for solar flux, irradiance, central intensity, and limb intensity, Memorie della Societa Astronomica Italiana Supplementi, 8, 189

Lagarde, N., Decressin, T., Charbonnel, C., et al. 2012, Thermohaline instability and rotation-induced mixing. III. Grid of stellar models and asymptotic asteroseismic quantities from the pre-main sequence up to the AGB for low- and intermediate-mass stars of various metallicities, Astron. Astrophys., 543, A108

Lahav, O., Naim, A., Sodré, L., J., & Storrie-Lombardi, M. C. 1996, Neural computation as a tool for galaxy classification: methods and examples, Mon. Not. R. Astron. Soc., 283, 207

Leung, H. W. & Bovy, J. 2019, Deep learning of multi-element abundances from high-resolution spectroscopic data, Mon. Not. R. Astron. Soc., 483, 3255

Lind, K., Asplund, M., Barklem, P. S., & Belyaev, A. K. 2011, Non-LTE calculations for neutral Na in late-type stars using improved atomic data, Astron. Astrophys., 528, A103

Lind, K., Nordlander, T., Wehrhahn, A., et al. 2022, Non-LTE abundance corrections for late-type stars from 2000Å to 3{}m: I. Na, Mg, and Al, arXiv e-prints, arXiv:2206.11070

Maas, A. L. 2013, in Rectifier Nonlinearities Improve Neural Network Acoustic Models

Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, The Apache Point Observatory Galactic Evolution Experiment (APOGEE), Astron. J., 154, 94

Martell, S. L., Simpson, J. D., Balasubramaniam, A. G., et al. 2021, The GALAH survey: a census of lithium-rich giant stars, Mon. Not. R. Astron. Soc., 505, 5340

Matijevič, G., Chiappini, C., Grebel, E. K., et al. 2017, Very metal-poor stars observed by the RAVE survey, Astron. Astrophys., 603, A19

Miglio, A., Chiappini, C., Mackereth, J. T., et al. 2021, Age dissection of the Milky Way discs: Red giants in the Kepler field, Astron. Astrophys., 645, A85

Mikolaitis, Š., Hill, V., Recio-Blanco, A., et al. 2014, The Gaia-ESO Survey: the chemical structure of the Galactic discs from the first internal data release, Astron. Astrophys., 572, A33

Milone, A. P., Bedin, L. R., Piotto, G., et al. 2008, The ACS Survey of Galactic Globular Clusters. III. The Double Subgiant Branch of NGC 1851, Astrophys. J., 673, 241

Mints, A. & Hekker, S. 2017, A Unified tool to estimate Distances, Ages, and Masses (UniDAM) from spectrophotometric data, Astron. Astrophys., 604, A108

Mucciarelli, A., Lapenna, E., Massari, D., et al. 2015, A Chemical Trompe-L'oeil: No Iron Spread in the Globular Cluster M22, Astrophys. J., 809, 128

Nelder, J. A. & Mead, R. 1965, A Simplex Method for Function Minimization, The Computer Journal, 7, 308

Nepal, S., Guiglion, G., de Jong, R. S., et al. 2023, The Gaia-ESO Survey: Preparing the ground for 4MOST and WEAVE galactic surveys. Chemical evolution of lithium with machine learning, Astron. Astrophys., 671, A61

Pancino, E., Lardo, C., Altavilla, G., et al. 2017a, The Gaia-ESO Survey: Calibration strategy, Astron. Astrophys., 598, A5

Pancino, E., Lardo, C., Altavilla, G., et al. 2017b, The Gaia-ESO Survey: Calibration strategy, Astron. Astrophys., 598, A5

Pancino, E., Romano, D., Tang, B., et al. 2017c, The Gaia-ESO Survey. Mg-Al anti-correlation in iDR4 globular clusters, Astron. Astrophys., 601, A112

Pasquini, L., Avila, G., Blecha, A., et al. 2002, Installation and commissioning of FLAMES, the VLT Multifibre Facility, The Messenger, 110, 1

Paszke, A., Gross, S., Massa, F., et al. 2019, in Advances in Neural Information Processing Systems 32, ed. H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Curran Associates, Inc.), 8024–8035

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825

Plez, B. 2012, Turbospectrum: Code for spectral synthesis, Astrophysics Source Code Library, record ascl:1205.004

Queiroz, A. B. A., Anders, F., Chiappini, C., et al. 2020, From the bulge to the outer disc: StarHorse stellar parameters, distances, and extinctions for stars in APOGEE DR16 and other spectroscopic surveys, Astron. Astrophys., 638, A76

Randich, S., Gilmore, G., & Gaia-ESO Consortium. 2013, The Gaia-ESO Large Public Spectroscopic Survey, The Messenger, 154, 47

Romano, D., Matteucci, F., Ventura, P., & D'Antona, F. 2001, The stellar origin of ^{7}Li. Do AGB stars contribute a substantial fraction of the local Galactic lithium abundance?, Astron. Astrophys., 374, 646

Rosenthal, D. A. 1988, in European Southern Observatory Conference and Workshop Proceedings, Vol. 28, European Southern Observatory Conference and Workshop Proceedings, 245–259

Santos, C. F. G. D. & Papa, J. a. P. 2022, Avoiding Overfitting: A Survey on Regularization Methods for Convolutional Neural Networks, ACM Comput. Surv., 54

Simpson, J. D., Martell, S. L., & Navin, C. A. 2017, A broad perspective on multiple abundance populations in the globular cluster NGC 1851, Mon. Not. R. Astron. Soc., 465, 1123

Singh, Y., Saini, M., & Savita. 2023, in 2023 IEEE International Conference on Contemporary Computing and Communications (InC4), Vol. 1, 1–7

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, The Two Micron All Sky Survey (2MASS), Astron. J., 131, 1163

Sneden, C. A. 1973, PhD thesis, University of Texas, Austin

Steinmetz, M., Matijevič, G., Enke, H., et al. 2020, The Sixth Data Release of the Radial Velocity Experiment (RAVE). I. Survey Description, Spectra, and Radial Velocities, The Astronomical Journal, 160, 82

Stetson, P. B. & Pancino, E. 2008, DAOSPEC: An Automatic Code for Measuring Equivalent Widths in High-Resolution Stellar Spectra, Publ. Astron. Soc. Pac., 120, 1332

Stetson, P. B., Pancino, E., Zocchi, A., Sanna, N., & Monelli, M. 2019, Homogeneous photometry - VII. Globular clusters in the Gaia era, Mon. Not. R. Astron. Soc., 485, 3042

Tautvaišienė, G., Drazdauskas, A., Bragaglia, A., et al. 2022, Gaia-ESO Survey: Detailed elemental abundances in red giants of the peculiar globular cluster NGC 1851, Astron. Astrophys., 658, A80

Traven, G., Matijevič, G., Zwitter, T., et al. 2017, The Galah Survey: Classification and Diagnostics with t-SNE Reduction of Spectral Information, Astrophys. J. Suppl. Ser., 228, 24

van der Maaten, L. & Hinton, G. 2008, Visualizing Data using t-SNE, Journal of Machine Learning Research, 9, 2579

Villanova, S., Geisler, D., & Piotto, G. 2010, Detailed Abundances of Red Giants in the Globular Cluster NGC 1851: C+N+O and the Origin of Multiple Populations, Astrophys. J. Lett., 722, L18

Yan, Y., Du, C., Liu, S., et al. 2019, Chemical and Kinematic Properties of the Galactic Disk from the LAMOST and iGaia/i Sample Stars, The Astrophysical Journal, 880, 36

Yong, D., Grundahl, F., D'Antona, F., et al. 2009, A Large C+N+O Abundance Spread in Giant Stars of the Globular Cluster NGC 1851, Astrophys. J. Lett., 695, L62

Yong, D., Grundahl, F., & Norris, J. E. 2015, CNO abundances in the globular clusters NGC 1851 and NGC 6752, Mon. Not. R. Astron. Soc., 446, 3319

Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, LAMOST spectral survey — An overview, Research in Astronomy and Astrophysics, 12, 723

Vilnius University Press
9 Saulėtekio Ave., Building III, LT-10222 Vilnius
Email: info@leidykla.vu.lt, www.leidykla.vu.lt
bookshop.vu.lt, journals.vu.lt
Print run 15