

LETTER TO THE EDITOR

A neural network approach to determining photometric metallicities of M-type dwarf stars

C. Duque-Arribas^{1,*}, H. M. Tabernero^{2,3}, D. Montes¹, J. A. Caballero⁴, and E. Galceran¹

¹ Departamento de Física de la Tierra y Astrofísica & IPARCOS-UCM (Instituto de Física de Partículas y del Cosmos de la UCM), Facultad de Ciencias Físicas, Universidad Complutense de Madrid, 28040 Madrid, Spain

² Institut de Ciències de l'Espai (ICE, CSIC), Campus UAB, c/ de Can Magrans s/n, 08193 Cerdanyola del Vallès, Barcelona, Spain

³ Institut d'Estudis Espacials de Catalunya (IEEC), C/Esteve Terradas, 1, Edifici RDIT, Campus PMT-UPC, 08860 Castelldefels, (Barcelona), Spain

⁴ Centro de Astrobiología (CSIC-INTA), ESAC Campus, Camino bajo del castillo s/n, 28692 Villanueva de la Cañada, Madrid, Spain

Received 24 March 2025 / Accepted 18 May 2025

ABSTRACT

Context. M dwarfs are the most abundant stars in the Galaxy and serve as key targets for stellar and exoplanetary studies. It is particularly challenging to determine their metallicities because their spectra are complex. For this reason, several authors have focused on photometric estimates of the M-dwarf metallicity. Although artificial neural networks have been used in the framework of modern astrophysics, their application to a photometric metallicity estimate for M dwarfs remains unexplored.

Aims. We develop an accurate method for estimating the photometric metallicities of M dwarfs using artificial neural networks to address the limitations of traditional empirical approaches.

Methods. We trained a neural network on a dataset of M dwarfs with spectroscopically derived metallicities. We used eight absolute magnitudes in the visible and infrared from *Gaia*, 2MASS, and WISE as input features. Batch normalization and dropout regularization stabilized the training and prevented overfitting. We applied the Monte Carlo dropout technique to obtain more robust predictions.

Results. The neural network demonstrated a strong performance in estimating photometric metallicities for M dwarfs in the range of $-0.45 \leq [\text{Fe}/\text{H}] \leq +0.45$ dex and for spectral types as late as M5.0 V. On the test sample, the predictions showed uncertainties down to 0.08 dex. This surpasses the accuracy of previous methods. We further validated our results using an additional sample of 46 M dwarfs in wide binary systems with FGK-type primary stars with well-defined metallicities and achieved an excellent predictive performance that surpassed the 0.1 dex error threshold.

Conclusions. This study introduces a machine-learning-based framework for estimating the photometric metallicities of M dwarfs and provides a scalable data-driven solution for analyzing large photometric surveys. The results outline the potential of artificial neural networks to enhance the determination of stellar parameters, and they offer promising prospects for future applications.

Key words. stars: abundances – stars: fundamental parameters – Hertzsprung–Russell and C–M diagrams – stars: late-type – stars: low-mass

1. Introduction

Knowledge of stellar parameters and in particular, the chemical composition of stars in the Milky Way, is crucial for several areas of astrophysics. M dwarfs are the most common type of star in our Galaxy by far. They constitute over 75% of all stars (Henry et al. 2006; Winters et al. 2015; Reylé et al. 2021). Studies of the metallicity of these cool stars have been made regarding the chemical and dynamical evolution of our Galaxy (Bahcall & Soneira 1980; Reid et al. 1997; Chabrier 2003; Ferguson et al. 2017) and exoplanet discoveries and characterization (e.g. Nutzman & Charbonneau 2008; Shields et al. 2016; Reiners et al. 2018; Trifonov et al. 2018; Luque et al. 2019; Caballero et al. 2022).

Nevertheless, determining the metallicity of M dwarfs still remains a challenging task because their spectra are complex. They are dominated by strong molecular features that erode the stellar continuum (Allard et al. 1997; Van Eck et al. 2017; Passegger et al. 2018; Marfil et al. 2021). For this reason, sev-

eral studies relied on photometric data to estimate the metallicity of these stars, using techniques such as frequentist or Bayesian statistics, k -nearest neighbors, or Gaussian-process regressors to provide photometric calibrations for the M-dwarf metallicities (e.g. Bonfils et al. 2005; Johnson & Apps 2009; Neves et al. 2012; Mann et al. 2013; Davenport & Dorn-Wallenstein 2019; Rains et al. 2021; Duque-Arribas et al. 2023).

Our study proposes a novel method for estimating photometric metallicities of M dwarfs using artificial neural networks (ANNs). The literature knows several ANNs that were applied to astrophysical studies, including the determination of stellar parameters from the Sloan Digital Sky Survey–III APOGEE¹ spectra (Fabbro et al. 2018) or from CARMENES² spectra (Passegger et al. 2020; Bello-García et al. 2023; Mas-Buitrago et al. 2024), evolutionary states of red

¹ Apache Point Observatory Galactic Evolution Experiment.

² Calar Alto high-Resolution search for M dwarfs with Exoearths with Near-infrared and optical Echelle Spectrographs; <https://carmenes.caha.es/index.html>

* Corresponding author: chrduque@ucm.es

Table 1. Data-filtering criteria applied to astrophotometric data.

Survey	Filter
<i>Gaia</i> EDR3	parallax_over_error > 10 ruwe < 1.4 photo_g_mean_flux_over_error > 50 photo_bp_mean_flux_over_error > 20 photo_rp_mean_flux_over_error > 20
2MASS	Qf1 = AAA
CatWISE2020	qph = AA**

Notes. Qf1 and qph are the quality flags in the 2MASS *JHK_s* bands and in the WISE *W1W2W3W4* bands, respectively.

giants from asteroseismology (Hon et al. 2017), star-galaxy classification (Kim & Brunner 2017), and a stellar spectral classification (Kheirdastan & Bazarghan 2016). Relatively few studies have explored the use of ANNs for deriving photometric metallicities, however. Whitten et al. (2019) estimated effective temperatures and metallicities for stars hotter than $T_{\text{eff}} > 4500$ K using J-PLUS³ photometry (Cenarro et al. 2019). Fallows & Sanders (2022) estimated metallicities of red giant stars from *Gaia* EDR3 (Gaia Collaboration 2016, 2021) and the 2MASS⁴ (Skrutskie et al. 2006) and WISE⁵ (Wright et al. 2010) surveys, and Molina-Jorquera et al. (2024) relied on S-PLUS⁶ photometry (Mendes Oliveira et al. 2019) to classify giant and dwarf stars and to determine the metallicity of red giants. Finally, Ferreira Lopes et al. (2025) estimated the stellar parameters for about five million stars from S-PLUS multiband photometry, although the authors indicated that the results for stars with $T_{\text{eff}} < 4000$ K should be interpreted with caution. Despite these advancements, no study to date has specifically focused on applying ANNs to estimate photometric metallicities for M dwarfs as a standalone population. We present an ANN that we used to estimate the metallicity of M-type dwarf stars based on their visible and infrared photometry.

2. Method

2.1. Data and preprocessing

The proposed approach was carried out on the stellar sample described by Duque-Arribas et al. (2023). They made use of the sample of 5875 early- and intermediate-type M dwarfs presented by Birky et al. (2020), with stellar parameters derived from APOGEE spectra (Majewski et al. 2017; Abolfathi et al. 2018) using The Cannon (Ness et al. 2015), and they cross-matched these stars with the *Gaia* DR3 (Gaia Collaboration 2023), 2MASS, and CatWISE2020 (Marocco et al. 2021) catalogs. Then, they filtered the resulting sample based on the photometric and astrometric quality criteria compiled in Table 1 while removing young and evolved objects. They obtained a final sample of 4919 M dwarfs.

This sample was then randomly divided into three subsets: 60% for training, 20% for validation, and 20% for testing, that is, 2951, 984 and 984 stars, respectively. The first subset was used to train the ANN, the second subset to estimate its accuracy and tune the hyperparameters, and the third subset to test the ANN.

Following the approach of Duque-Arribas et al. (2023), we evaluated the predictive performance of the ANN using a separate dataset of FGK+M binary systems from Montes et al. (2018). The primary stars have well-determined stellar atmospheric parameters derived from high-resolution spectroscopy with the equivalent width method (Tabernero et al. 2019), and it is reasonable to assume that the M-dwarf companions share the same chemical composition (Desidera et al. 2006; Andrews et al. 2018).

To prepare the dataset, we cross-matched the M-dwarf companions with the *Gaia* DR3, 2MASS, and CatWISE2020 catalogs and identified 115 stars in common. We then applied the same data-filtering criteria as described earlier and constrained stars with parameters similar to the Birky et al. (2020) sample. Specifically, we limited the sample to stars with $-0.45 \leq [\text{Fe}/\text{H}] \leq 0.45$ dex, $1.85 \text{ mag} \leq G - J \leq 3.10$ mag, and $-0.10 \text{ mag} \leq W1 - W2 \leq 0.24$ mag, which correspond to early- and intermediate-M dwarfs (spectral types M0.0 V to M5.0 V approximately; Cifuentes et al. 2020). After applying these constraints, we obtained a final sample of 46 M-dwarf companions for testing the ANN on a completely independent dataset.

2.2. ANN architecture

We employed the Keras⁷ API, which operates on the TensorFlow⁸ framework, to build our model. Hyperparameter optimization, in our case, for the number of hidden layers, the number of neurons per layer, and the learning rate, was performed using the RandomizedSearchCV class from the scikit-learn⁹ library. This approach efficiently explores the hyperparameter space by sampling it randomly, allowing for the identification of near-optimal configurations in less time compared to exhaustive methods, such as the grid search. The optimization process used a *k*-fold cross-validation, with four folds in our implementation, and it evaluated the model based on a modified mean squared error (MSE) as the loss metric.

The modification of the MSE was included to mitigate the overabundance of solar-metallicity stars in the sample, as noted by Fallows & Sanders (2022). Otherwise, the ANN might inadvertently interpret this imbalance as a trend in the data, which would lead to biased predictions. To counteract this, we included a weighting factor $W = 4 \times |[\text{Fe}/\text{H}]| + 0.5$ as a linear multiplier on the MSE. This weighting reduced the influence of the stars with $[\text{Fe}/\text{H}] \approx 0$ while it linearly increased the effect for objects with higher or lower metallicities. This ensured a more balanced contribution throughout the entire metallicity range. We tested alternative weighting strategies, including quadratic and exponential functions, but found that they did not improve the results. The linear form was chosen for its simplicity and effectiveness. The specific coefficients were determined empirically: The slope ensured a gradual yet sufficient increase in weight as the metallicity deviated from solar, while the intercept of 0.5 prevented complete suppression of the most abundant metallicities. This balance was found to stabilize training and improve model generalization, in particular, for stars at the extremes of our metallicity range.

The architecture of the neural network was chosen based on the empirical performance during the hyperparameter tuning. We explored a range of configurations that are summarized in Table 2. Our final architecture, with two hidden layers of

³ Javalambre-Photometric Local Universe Survey.

⁴ Two Micron All-Sky Survey.

⁵ Wide-field Infrared Survey Explorer.

⁶ Southern Photometric Local Universe Survey.

⁷ <https://keras.io/about/>

⁸ <https://www.tensorflow.org/>

⁹ <https://scikit-learn.org/stable/>

Table 2. Summary of the tested ANN hyperparameters.

Hyperparameter	Explored values
Number of hidden layers	1, 2, 3, 4, 5
Neurons per layer	8, 16, 32, 64, 128, 256
Learning rate	reciprocal($1 \cdot 10^{-4}$, $1 \cdot 10^{-2}$)
Dropout rate	0.1, 0.2, 0.3, 0.4, 0.5

64 neurons each, consistently achieved the lowest validation loss without signs of overfitting. Simpler architectures showed higher validation errors, while more complex architectures offered no additional performance gain and occasionally led to overfitting.

The input layer incorporated eight features, each corresponding to an absolute magnitude using the *Gaia* parallaxes: $M_{G_{BP}}$, M_G , and $M_{G_{RP}}$ from *Gaia*, M_J , M_H , and M_{K_S} from 2MASS, and M_{W1} , and M_{W2} from CatWISE2020. The hidden layers, fully connected, contained 64 neurons per layer. The output layer consisted of a single neuron for regression. The network was trained using the Adam optimizer (a stochastic gradient descent method; Kingma & Ba 2017) with a learning rate of $2 \cdot 10^{-4}$. For the learning rate, we sampled values from a reciprocal (log-uniform) distribution in the range $[1 \cdot 10^{-4}, 1 \cdot 10^{-2}]$, as this approach gives greater weight to lower values, which are more likely to yield stable convergence during training. A schematic representation of the neural network is displayed in Fig. 1.

The neurons in the input and hidden layers incorporated linear activation functions, whereas the output layer adopted a hyperbolic tangent (tanh) activation function. This choice constrained the predictions to the range $(-1, +1)$ dex. Because the training sample spans a narrower metallicity range (-0.45 dex to $+0.45$ dex), this activation ensured that predictions remained bounded within a safe interval during training. It also implies a limitation, however: The model is not suited for extrapolating beyond the training range. For stars with true metallicities outside this domain, the output will saturate toward ± 1 , reflecting the asymptotic behaviour of the tanh function. These predictions should not be interpreted as accurate estimates, but rather as an indication that the input is outside the reliable operating range of the model.

We also incorporated batch normalization after the input and hidden layers. This technique normalizes activations during training, stabilizing and accelerating convergence by mitigating internal covariate shifts (Ioffe & Szegedy 2015). Additionally, dropout was implemented after the hidden layers, with the dropout rate included as a tunable hyperparameter. This technique is discussed in greater detail in the following subsection.

2.3. Monte Carlo dropout

Dropout is one of the most popular regularization techniques for neural networks that aims at mitigating overfitting during the training stage by randomly dropping a fraction of the neurons (Hinton et al. 2012; Srivastava et al. 2014). During every training step, each neuron has a certain probability that is included as a tuneable hyperparameter of being ignored but being active during the next step. By applying this technique, the ANN learns as a cohesive unit and avoids over-reliance on any particular neuron. We applied dropout to the layers of our neural network and set the dropout rate at 20%.

In the classical dropout, neurons are no longer dropped after the training stage. Gal & Ghahramani (2016) introduced

a technique called Monte Carlo (MC) dropout, however, which enhances the performance of a trained dropout model and provides a more accurate measure of the model uncertainty. This is achieved by ensuring that the dropout layers remain active after the training, which produces a distribution of predictions instead of a single estimate. Averaging over these multiple predictions with dropout provides a Monte Carlo estimate that is more reliable than a single prediction. Gal & Ghahramani (2016) also established a deep connection between dropout networks and the approximate Bayesian inference.

3. Results and discussion

After training the ANN and optimizing its hyperparameters using the training and validation datasets, we evaluated its performance on the test sample. As outlined in the previous section, we applied MC dropout to generate 100 stochastic forward passes per star. This provided more robust estimates and an uncertainty measure for these predictions. We also generated 500 samples, but the predictions did not change.

In Fig. 2 we compare the spectroscopic metallicity values reported by Birky et al. (2020) and the photometric metallicity predictions generated by our ANN for the test subsample, color-coded by $G_{BP} - G_{RP}$, which is a proxy for the stellar effective temperature Cifuentes et al. (2020). Most of the estimated metallicities closely align with the one-to-one relation. This indicates the reliability of the model. Furthermore, as expected, no significant correlation is observed between the predicted metallicities and effective temperature. The prediction residuals exhibit a standard deviation of 0.08 dex, which is compatible with the values reported by Duque-Arribas et al. (2023) using a Bayesian approach.

In Fig. 3 we present the color-magnitude diagram for the test subsample, color-coded by metallicity estimates from our ANN. The diagram shows the clear recovery of the expected metallicity gradient along the main sequence. This alignment underscores the ability of the ANN to capture and reproduce well-known trends in stellar populations. For a detailed discussion of metallicity gradients in the color-color and color-magnitude diagrams, we refer to Duque-Arribas et al. (2023).

Finally, as an additional test, we applied our ANN to the M-dwarf secondaries in wide binary systems reported by Montes et al. (2018). In Fig. 4 we compare the spectroscopic metallicities of the primary stars reported by Montes et al. (2018) and the photometrically estimated values for the M-dwarf companions using our ANN. The statistics of the residuals between the spectroscopic and photometric metallicities for the 46 stars with good-quality data, obtained with previous photometric studies found in the literature and with our ANN, are reported in Table 3. We achieved a strong correlation between photometric and spectroscopic metallicities of the M dwarfs compared with these previous results, as illustrated by Fig. A.1.

Among the most recent studies, both Rains et al. (2021) and Duque-Arribas et al. (2023) employed polynomial fits to the main sequence to estimate the photometric metallicities of M dwarfs. The model by Rains et al. (2021) relied on *Gaia* and 2MASS photometry alone and omitted the W1–W2 color index, which is a known metallicity-sensitive feature (Schmidt et al. 2016). This omission likely contributes to its slightly larger dispersion compared to the model by Duque-Arribas et al. (2023) and the present work. Our results are similar to those of Duque-Arribas et al. (2023), as expected because both studies used the same underlying dataset. The modest improvement

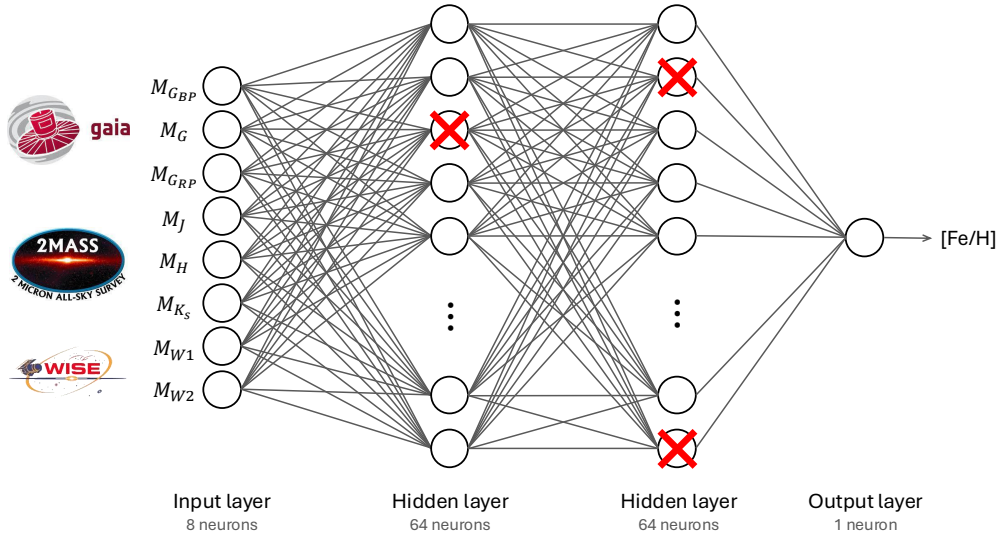


Fig. 1. Schematic representation of our ANN architecture. The red crosses indicate dropout neurons.

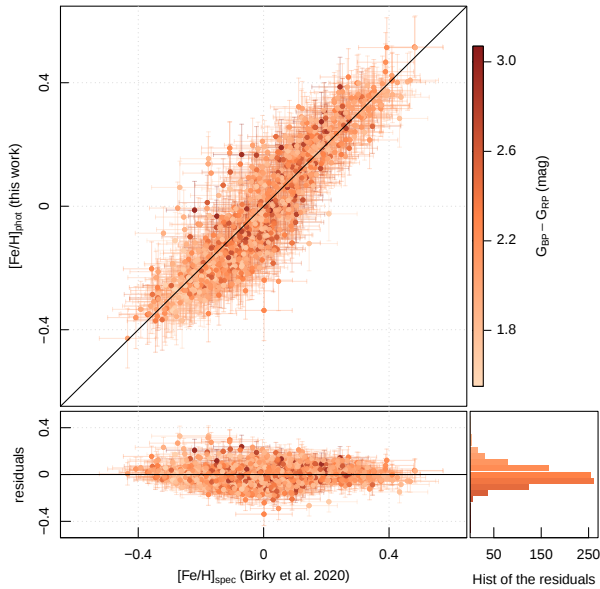


Fig. 2. Comparison between spectroscopic metallicities reported by Birky et al. (2020) and photometric estimates by our ANN. The 984 stars from the test sample and their corresponding residuals are color-coded by $G_{BP} - G_{RP}$ color (darker symbols show cooler stars, and lighter symbols show warmer stars). The solid lines denote the 1:1 relation, and the residuals are zero. In the bottom right corner we represent the histogram of the residuals.

achieved here can probably be attributed to the greater flexibility of the ANN relative to the polynomial fitting approach.

4. Conclusions

We presented a machine-learning framework for estimating the photometric metallicities of M dwarfs, for which we exploited the capabilities of ANNs combined with the accuracy and homogeneity of the visible and infrared photometry from the *Gaia* DR3, 2MASS, and CatWISE surveys. By employing robust techniques such as weighting adjustments in the loss function to mitigate the effects of sample imbalances, MC dropout for

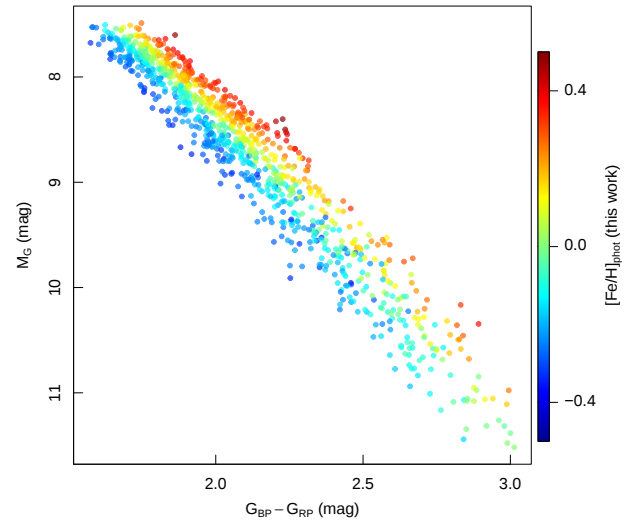


Fig. 3. Color-magnitude diagram of the 984 stars from the test sample, color-coded by the metallicity estimates provided by our ANN.

the uncertainty estimation, and hyperparameter optimization, we demonstrated that our ANN with two hidden layers can achieve highly accurate and reliable predictions.

We trained the model with the sample presented by Birky et al. (2020). Its performance on the test sample showed a mean residual of 0.00 ± 0.08 dex. Additionally, the results confirmed that the predicted metallicities closely match spectroscopic values and exhibit no correlation with the effective temperature. This validated the ability of the model to generalize without introducing biases. The validation on an independent dataset of 46 M dwarfs in FGK+M binary systems from Montes et al. (2018) further reinforced the predictive accuracy of the ANN and underscored its versatility when applied to new data. These results agree well with those presented by Duque-Arribas et al. (2023) using a Bayesian linear regression approach.

This ANN-based approach provides a scalable and efficient alternative for processing large photometric datasets. Its ability to infer stellar metallicities using only broadband photometry

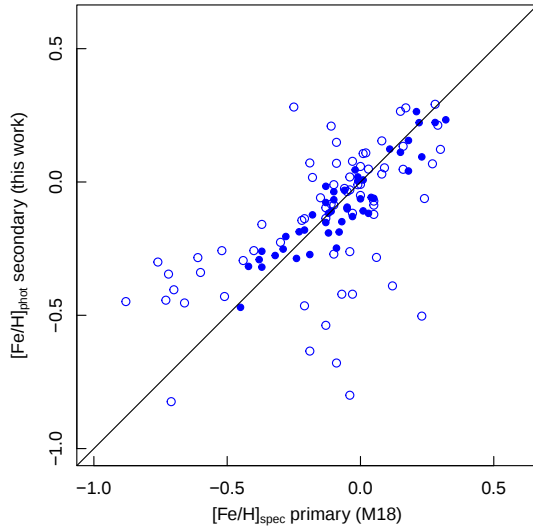


Fig. 4. Comparison between spectroscopic metallicity of the FGK-type primary stars measured by Montes et al. (2018) and the photometric metallicity estimated for the M-dwarf companions by this work using an ANN. The filled circles represent the 46 good-quality subsample.

Table 3. Statistics of the differences between the primary $[\text{Fe}/\text{H}]_{\text{spec}}$ and the secondary $[\text{Fe}/\text{H}]_{\text{phot}}$.

	$\bar{x}$	$\tilde{x}$	σ_x	$\text{MAD}(x)$	r	r_s
B05	+0.144	+0.111	0.147	0.105	0.71	0.67
JA09	-0.064	-0.086	0.206	0.144	0.69	0.66
N12	+0.071	+0.039	0.149	0.115	0.69	0.66
M13	+0.043	+0.063	0.144	0.156	0.65	0.64
DD19	+0.047	+0.056	0.176	0.167	0.53	0.52
R21	-0.007	+0.001	0.105	0.089	0.84	0.78
DA23	-0.012	-0.004	0.087	0.081	0.90	0.83
This work	+0.015	+0.001	0.082	0.075	0.92	0.88

Notes. Mean ($\bar{x}$), median ($\tilde{x}$), standard deviation (σ_x), median absolute deviation ($\text{MAD}(x)$), and Pearson's (r) and Spearman's (r_s) correlation coefficients, where $x = [\text{Fe}/\text{H}]_{\text{spec}} - [\text{Fe}/\text{H}]_{\text{phot}}$ in dex.

significantly reduces the reliance on time-intensive spectroscopic observations. This framework can be extended to incorporate additional features or input data to enhance the accuracy and applicability of the predictions. Furthermore, the model could be adapted to other stellar types or parameter determinations, which would broaden its applicability in stellar astrophysics. The Python code is publicly available at GitHub¹⁰.

In summary, this work highlights the potential of machine learning, particularly ANNs, in advancing our ability to determine the stellar metallicity efficiently and accurately. It paves the way for transformative applications in the era of large-scale astronomical surveys.

Acknowledgements. We thank the anonymous referee for the instructive and detailed report, which clearly improved our manuscript. We acknowledge financial support from the Universidad Complutense de Madrid and the Agencia Estatal de Investigación (AEI/10.13039/501100011033) of the Ministerio de Ciencia e Innovación and the ERDF “A way of making Europe” through projects PID2019-107427GB-C31, PID2022-137241NB-C4 Allard et al. (1997), Bahcall & Soneira (1980), and PID2022-138855NB-C31, and the European

Research Council (ERC) under the European Union’s Horizon Europe programme (ERC Advanced Grant SPOTLESS; no. 101140786).

References

- Abolfathi, B., Aguado, D. S., Aguilar, G., et al. 2018, *ApJS*, **235**, 42
Allard, F., Hauschildt, P. H., Alexander, D. R., & Starrfield, S. 1997, *ARA&A*, **35**, 137
Andrews, J. J., Chanamé, J., & Agüeros, M. A. 2018, *MNRAS*, **473**, 5393
Bahcall, J. N., & Soneira, R. M. 1980, *ApJS*, **44**, 73
Bello-García, A., Passegger, V. M., Ordieres-Meré, J., et al. 2023, *A&A*, **673**, A105
Birky, J., Hogg, D. W., Mann, A. W., & Burgasser, A. 2020, *ApJ*, **892**, 31
Bonfils, X., Delfosse, X., Udry, S., et al. 2005, *A&A*, **442**, 635
Caballero, J. A., González-Álvarez, E., Brady, M., et al. 2022, *A&A*, **665**, A120
Cenarro, A. J., Moles, M., Cristóbal-Hornillos, D., et al. 2019, *A&A*, **622**, A176
Chabrier, G. 2003, *PASP*, **115**, 763
Cifuentes, C., Caballero, J. A., Cortés-Contreras, M., et al. 2020, *A&A*, **642**, A115
Davenport, J. R. A., & Dorn-Wallenstein, T. Z. 2019, *Res. Notes Am. Astron. Soc.*, **3**, 54
Desidera, S., Gratton, R. G., Lucatello, S., & Claudi, R. U. 2006, *A&A*, **454**, 581
Duque-Arribas, C., Montes, D., Tabernero, H. M., et al. 2023, *ApJ*, **944**, 106
Fabbro, S., Venn, K. A., O’Brien, T., et al. 2018, *MNRAS*, **475**, 2978
Fallows, C. P., & Sanders, J. L. 2022, *MNRAS*, **516**, 5521
Ferguson, D., Gardner, S., & Yanny, B. 2017, *ApJ*, **843**, 141
Ferreira Lopes, C. E., Gutiérrez-Soto, L. A., Ferreira Alberice, V. S., et al. 2025, *A&A*, **693**, A306
Gaia Collaboration (Prusti, T., et al.) 2016, *A&A*, **595**, A1
Gaia Collaboration (Brown, A. G. A., et al.) 2021, *A&A*, **649**, A1
Gaia Collaboration (Vallenari, A., et al.) 2023, *A&A*, **674**, A1
Gal, Y., & Ghahramani, Z. 2016, in Proceedings of The 33rd International Conference on Machine Learning, eds. M. F. Balcan, & K. Q. Weinberger (New York: PMLR), *Proceedings of Machine Learning Research*, **48**, 1050
Henry, T. J., Jao, W.-C., Subasavage, J. P., et al. 2006, *AJ*, **132**, 2360
Hinton, G. E., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. 2012, ArXiv e-prints [arXiv:1207.0580]
Hon, M., Stello, D., & Yu, J. 2017, *MNRAS*, **469**, 4578
Ioffe, S., & Szegedy, C. 2015, ArXiv e-prints [arXiv:1502.03167]
Johnson, J. A., & Apps, K. 2009, *ApJ*, **699**, 933
Kheirastan, S., & Bazarghan, M. 2016, *Ap&SS*, **361**, 304
Kim, E. J., & Brunner, R. J. 2017, *MNRAS*, **464**, 4463
Kingma, D. P., & Ba, J. 2017, ArXiv e-prints [arXiv:1412.6980]
Luque, R., Pallé, E., Kossakowski, D., et al. 2019, *A&A*, **628**, A39
Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, *AJ*, **154**, 94
Mann, A. W., Brewer, J. M., Gaidos, E., Lépine, S., & Hilton, E. J. 2013, *AJ*, **145**, 52
Marfil, E., Tabernero, H. M., Montes, D., et al. 2021, *A&A*, **656**, A162
Marocco, F., Eisenhardt, P. R. M., Fowler, J. W., et al. 2021, *ApJS*, **253**, 8
Mas-Buitrago, P., González-Marcos, A., Solano, E., et al. 2024, *A&A*, **687**, A205
Mendes Oliveira, C., Ribeiro, T., Schoenell, W., et al. 2019, *MNRAS*, **489**, 241
Molina-Jorquera, F., Damke, G., Fernández-Olivares, D., et al. 2024, *A&A*, **691**, A144
Montes, D., González-Peinado, R., Tabernero, H. M., et al. 2018, *MNRAS*, **479**, 1332
Ness, M., Hogg, D. W., Rix, H. W., Ho, A. Y. Q., & Zasowski, G. 2015, *ApJ*, **808**, 16
Neves, V., Bonfils, X., Santos, N. C., et al. 2012, *A&A*, **538**, A25
Nutzman, P., & Charbonneau, D. 2008, *PASP*, **120**, 317
Passegger, V. M., Reiners, A., Jeffers, S. V., et al. 2018, *A&A*, **615**, A6
Passegger, V. M., Bello-García, A., Ordieres-Meré, J., et al. 2020, *A&A*, **642**, A22
Rains, A. D., Žerjal, M., Ireland, M. J., et al. 2021, *MNRAS*, **504**, 5788
Reid, I. N., Gizis, J. E., Cohen, J. G., et al. 1997, *PASP*, **109**, 559
Reiners, A., Ribas, I., Zechmeister, M., et al. 2018, *A&A*, **609**, L5
Reylé, C., Jardine, K., Fouqué, P., et al. 2021, *A&A*, **650**, A201
Schmidt, S. J., Wagoner, E. L., Johnson, J. A., et al. 2016, *MNRAS*, **460**, 2611
Shields, A. L., Ballard, S., & Johnson, J. A. 2016, *Phys. Rep.*, **663**, 1
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, **131**, 1163
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014, *J. Mach. Learn. Res.*, **15**, 1929
Tabernero, H. M., Marfil, E., Montes, D., & González Hernández, J. I. 2019, *A&A*, **628**, A131
Trifonov, T., Kürster, M., Zechmeister, M., et al. 2018, *A&A*, **609**, A117
Van Eck, S., Neykens, P., Jorissen, A., et al. 2017, *A&A*, **601**, A10
Whitten, D. D., Placco, V. M., Beers, T. C., et al. 2019, *A&A*, **622**, A182
Winters, J. G., Henry, T. J., Lurie, J. C., et al. 2015, *AJ*, **149**, 5
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, **140**, 1868

¹⁰ <https://github.com/chrduque/metamorphosis-NN.git>

Appendix A: Additional figure

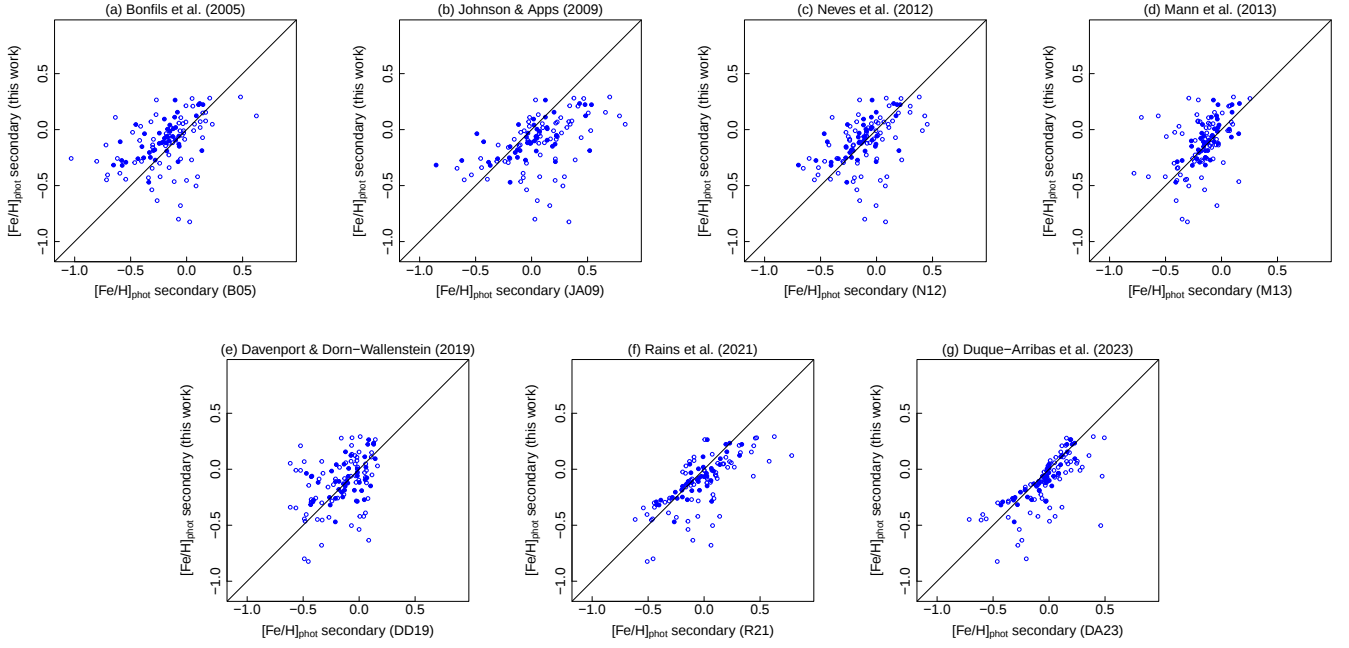


Fig. A.1. Same as Fig. 4, but for the photometric metallicities estimated by [Bonfils et al. \(2005\)](#), [Johnson & Apps \(2009\)](#), [Neves et al. \(2012\)](#), [Mann et al. \(2013\)](#), [Davenport & Dorn-Wallenstein \(2019\)](#), [Rains et al. \(2021\)](#), and [Duque-Arribas et al. \(2023\)](#).