ASSESSING THE RELIABILITY OF GAIA DR3 EFFECTIVE TEMPERATURES *

Avdeeva A., Kovaleva D., Malkov O.
Institute of Astronomy RAS
Moscow, Russia
avdeeva@inasan.ru

ABSTRACT

Effective temperatures play a vital role in numerous astronomical applications, including stellar classification, interstellar extinction estimation, and understanding of stellar atmospheres. Accurate effective temperatures are crucial for interpreting observational data and advancing our understanding of the Universe. This study focuses on assessing the reliability of effective temperatures derived from Gaia Data Release 3 (DR3) for interstellar reddening calculations. To extract the temperatures, suitable for the individual reddening evaluation, Machine Learning methods such as XGBoost and Support Vector Machines (SVM) are employed. The reliability of Gaia DR3 temperatures, selected by the model, is evaluated through comparative analyses with high-resolution spectroscopy surveys. The findings shed light on the accuracy of Gaia DR3 effective temperatures and the effectiveness of machine learning methods in extracting temperatures for interstellar reddening calculations.

***K*eywords** surveys · effective temperatures · Gaia

1 Introduction

Gaia Data Release 3 (DR3) [1, 2] represents a remarkable milestone in our understanding of the Milky Way and beyond, providing an unprecedented wealth of precise astrometric, photometric, and stellar parameter data that will revolutionize various fields of astronomy. It includes information on astrophysical characteristics for hundreds of millions of stars obtained via several independent pipelines having different sets of input data from Gaia observables [3]. One of the main modules, General Stellar Parametriser from photometry (GSP-Phot,) estimates effective temperature Teff, logarithm of surface gravity log g, metallicity, absolute magnitude MG, radius R, distance r, line-of-sight extinctions A0, AG, ABP, and ARP, and the reddening E(BP-RP) by forward-modelling the low-resolution BP/RP spectra, apparent G magnitude, and parallax using a Markov Chain Monte Carlo (MCMC) method. To this end, GSP-Phot employs stellar evolutionary models in forward-model interpolation in order to obtain self-consistent temperatures, surface gravities, metallicities, radii, and absolute magnitudes [4]. Gaia DR3 provides stellar parameters obtained with a GSP-Phot module for 471 million sources. These parameters enable detailed studies of stellar properties, stellar evolution, and the composition of different stellar populations across the galaxy. However, self-consistent determination of the parameters could lead to additional systematic errors, particularly, in $T_{\rm eff}$ estimations.

On the other hand, accurate determination of stellar temperatures is of paramount importance in various fields of astronomy and astrophysics. The knowledge of precise and reliable temperatures of celestial objects allows us to unravel their intrinsic properties, understand their evolutionary stages, and gain insights into fundamental astrophysical processes.

At the moment there are a number of spectroscopic surveys with a decent resolution that could be considered a reliable source of fundamental stellar parameters, such as APOGEE [5], GALAH [6], Gaia-ESO [7], etc. APOGEE survey DR16 ($R \sim 22500$) covers a decent amount of areas in the Nothern sky with some parts in the Southern sky. APOGEE primarily aims to study evolved stars across the Galactic disk, the Galactic Center, and the outer halo. A baseline magnitude limit of H = 12.2 was adopted for "normal" APOGEE fields. GALAH survey ($R \sim 28000$) explores the

* *Citation*: **Authors. Title. Pages.... DOI:000000/11111.**

stars with the following magnitude selection: $12 < V < 14$, and the galactic latitude $|b| > 10$ deg, so it probes mainly the thin and the thick disc of the Galaxy. Gaia-ESO survey ($R \sim 16000 - 25000$) is a public spectroscopic survey that aims to obtain high-quality spectroscopy of 100,000 Milky Way stars, systematically covering all major components of the Galaxy. Gaia-ESO survey operates in the Southern sky and targets primarily faint stars $V > 16$. Although these surveys have sufficient resolution and signal-to-noise ratio, they are limited in sky coverage and brightness of the stars.

The objective of this study is to evaluate the dependability of effective temperatures derived from Gaia DR3 in order to explore if we can use these temperatures to obtain the individual reddening of the stars. By utilizing a comprehensive dataset comprising spectroscopic parameters, photometric measurements, and intrinsic colour-temperature relationships, the Machine Learning models are trained to decide if the temperature from Gaia DR3 is suitable for assessing the individual reddening. We compare the temperatures, approved by the models, from Gaia DR3 with temperatures from APOGEE and GALAH surveys to evaluate their accuracy.

This article is organised as follows. In Sec.2 we describe the method of calculation interstellar reddening from Gaia DR3 data and compare the result with Gaia DR3 reddening. In Sec.3 we explore the dependence of the calculated reddening with the effective temperature of Gaia DR3 and apparent magnitude G. The machine learning methods are described and applied in Sec.4. In Sec.5 we discuss the results and make a conclusion.

2 Calculation of the interstellar reddening and comparison with Gaia calculated reddening

We have selected 20 test areas, uniformly distributed on the sky (see Fig.1). The radius of the area depends on the galactic latitude b: $r = 0.1 + {}^{|b|}/_{100}$ so that the number of objects at lower latitudes is compatible with the number of objects at higher latitudes.

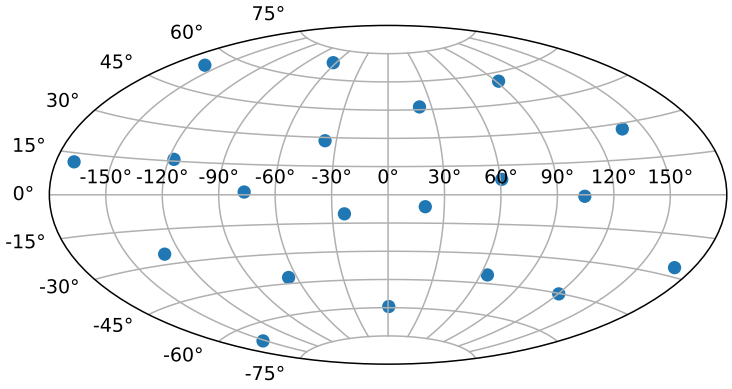

Figure 1: The distribution of areas in galactic coordinates.

For individual stars in each area, we have calculated the reddening $E(BP - RP) = (BP - RP) - (BP - RP)_0$. The intrinsic colours $(BP - RP)_0$ were calculated according to [8] [2], where the intrinsic colours provided for different effective temperatures.

We then compare the calculated reddening with the one provided by Gaia DR3 (Fig. 2). The green dots are calculated in this work using [8], red dots are from Gaia DR3. The remarkable features are the area below zero and near zero at large distances (green dots) and the reddening peak closer than 2 kpc (both green and red dots). The main difference between these results is the no negative reddening in the data provided by Gaia. This is most likely due to the algorithm, used in Gaia, which assumes the impossibility of negative reddening and extinction.

Through the similarity of our reddening and the Gaia ones, we conclude that this result is not caused by the specific color-$T_{\rm eff}$ relationship we use, but with mistaken effective temperatures from Gaia DR3.

3 The dependence of the reddening on the Gaia G magnitude and T$_{eff}$

We found two main causes for the features, described in Sec. 2. Those stars that show negative reddening are most likely faint stars with $G_{mag} > 17^m.5$ for which the Gaia atmospheric algorithm does not perform well (See Fig.3).

[2] `http://www.pas.rochester.edu/~emamajek/EEM_dwarf_UBVIJHK_colors_Teff.txt`

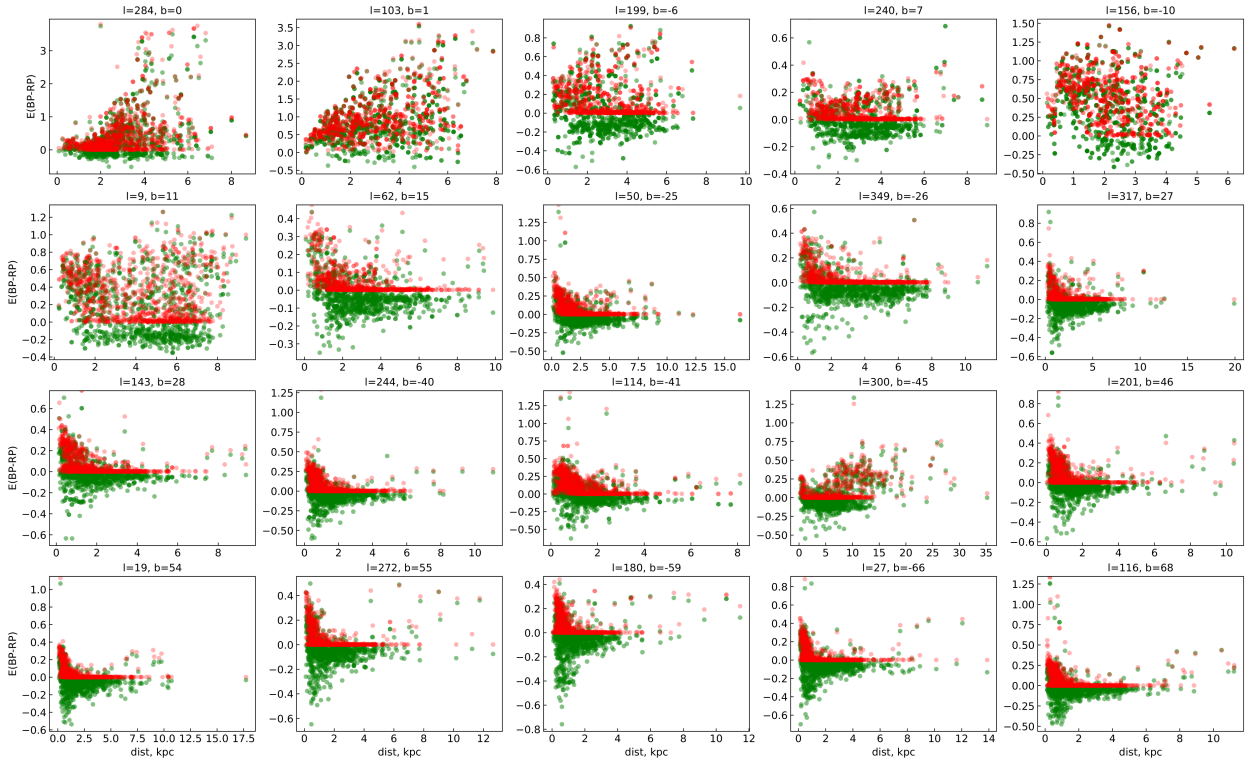

Figure 2: Reddening vs distance dependence based on Gaia DR3 data. The green dots are calculated in this work using Gaia $teff\_gspphot$ temperatures, the red ones are reddening provided by Gaia DR3. The distance scale and reddening scale are different for each plot for better visibility.

Possible systematics at the faint end was mentioned in Chapter 11 of Gaia Documentation [9], stating that discrepancies may occur at $G_{mag} > 16^m$ for $logg$. We use a more tolerant threshold for magnitude G_{mag} to preserve as much good-enough data as possible. The quality of the remaining temperatures will be later evaluated using high-resolution spectroscopic surveys.

Another ambiguous feature of the obtained plots is the high reddening peak within the nearest 2 kpc, decreasing with the distance. Fig. 4 shows the same areas as in Fig. 3 after applying $G_{mag} < 17^m.5$ coloured according to the effective temperature of the object. The peak is strongly correlated with the lower temperatures $T_{\text{eff}} < 5000$. This is especially noticeable at higher latitudes ($|b| > 50$), where the reddening is usually small. Thus, we set another threshold $T_{\text{eff}} \geq 5000$ to get rid of inadequately large reddenings for nearby stars.

Fig. 5 shows the result of applying two thresholds: $G_{mag} < 17.5$ and $T_{\text{eff}} \geq 5000$, to the Gaia samples. Although most non-conforming results are cut off by these filters, there are still areas where the reddening seems to be misidentified. Namely, a dense horizontal line near zero that extends to greater distances. The most prominent example of such behaviour is an area with coordinates $l = 240$, $b = 7$, where this line goes along with the increasing with distance reddening. There are also some stars with significantly larger $E(BP - RP)$ within the 2 kpc. In the next section, we use Machine Learning methods to achieve better results.

4 Application of ML to the extraction of reliable temperatures for the reddening calculation

4.1 Dataset and features

We compiled the training set from the very same data we used above, in total 53241 objects from 20 areas on the sky. Then, we marked objects with $T_{\text{eff}} < 5000$ or $G_{mag} > 17.5$ (as we already infer to be the indicators of bad astrophysical parameters solution) as the objects of the negative class (label= 0). Also, we said that the objects with $E(BP - RP) < 0$ (these are objects that mostly make up horizontal lines) belong to the negative class. Moreover, to pay additional attention to the objects with extremely high reddening at close distances, we said, that all objects with $E(BP - RP) > 0.2$ within 1 kpc to be of negative class as well. Thus objects that do not have any of the listed

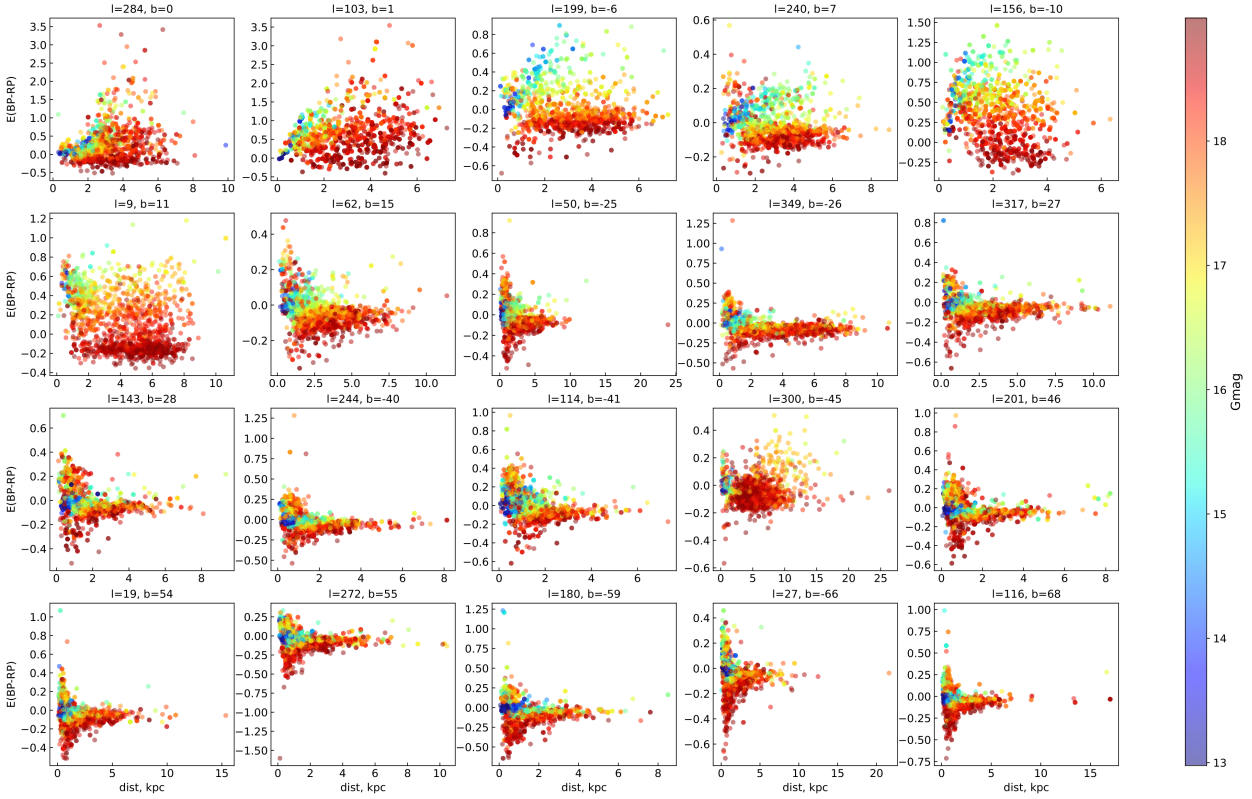

Figure 3: Dependence of reddening to distance with magnitude indicated by colour. As can be seen, the negative reddening are associated with faint stars.

disadvantages are the objects of the positive class (label$= 1$). Out of the whole dataset, 6950 objects are of positive class and 46291 are of negative class. The models then need to distinguish the positive and negative class objects according to the features.

The features we used are presented in Fig. 6 correlation matrix. We use astrometric parameters (coordinates, parallaxes and their errors as well as astrometric quality flags), photometric parameters (magnitudes, fluxes and their quality flags from Gaia DR3, with C* introduced in [10]) and atmospheric parameters from Gaia DR3 (effective temperatures, logg and metallicity as well as their upper and lower estimations). According to the correlation matrix, there are some correlations (or anti-correlations) between label and parallax error, magnitudes and fluxes, effective temperature and $logg$, so these features most likely influence the result. It should be noted that although models know the effective temperatures and G_{mag}, we expose no information on reddening or extinction, either provided by Gaia DR3 or calculated in Sec. 2.

4.2 Machine learning models

In this work, we use two approaches: boosting algorithm XGBoost [11] and Support Vector Machine Classifier [12]. Boosting is a popular Machine Learning technique that uses ensemble learning to train models sequentially, with each new model correcting the errors of previous ones. The Support Vector Machine is another widely used method that finds a line, surface, or hypersurface to separate classes in the feature space by maximizing the distance from each point to the interface. We use the Sckit-learn library [13] of Python to conduct this research.

The data is split into training, validation, and test sets with a 60/20/20 split in a stratified fashion across all of the areas. Fig. 7 shows the distribution of train/test objects by latitude and longitude.

Features are scaled using StandardScaler and model hyperparameters are selected using optuna [14] on the validation set. The final model performance is evaluated on the test set. However, as our labelling is not physically driven, but rather empirical, we compare the temperatures of objects, selected by the model, with temperatures from high-resolution spectroscopic surveys, namely, APOGEE and GALAH.

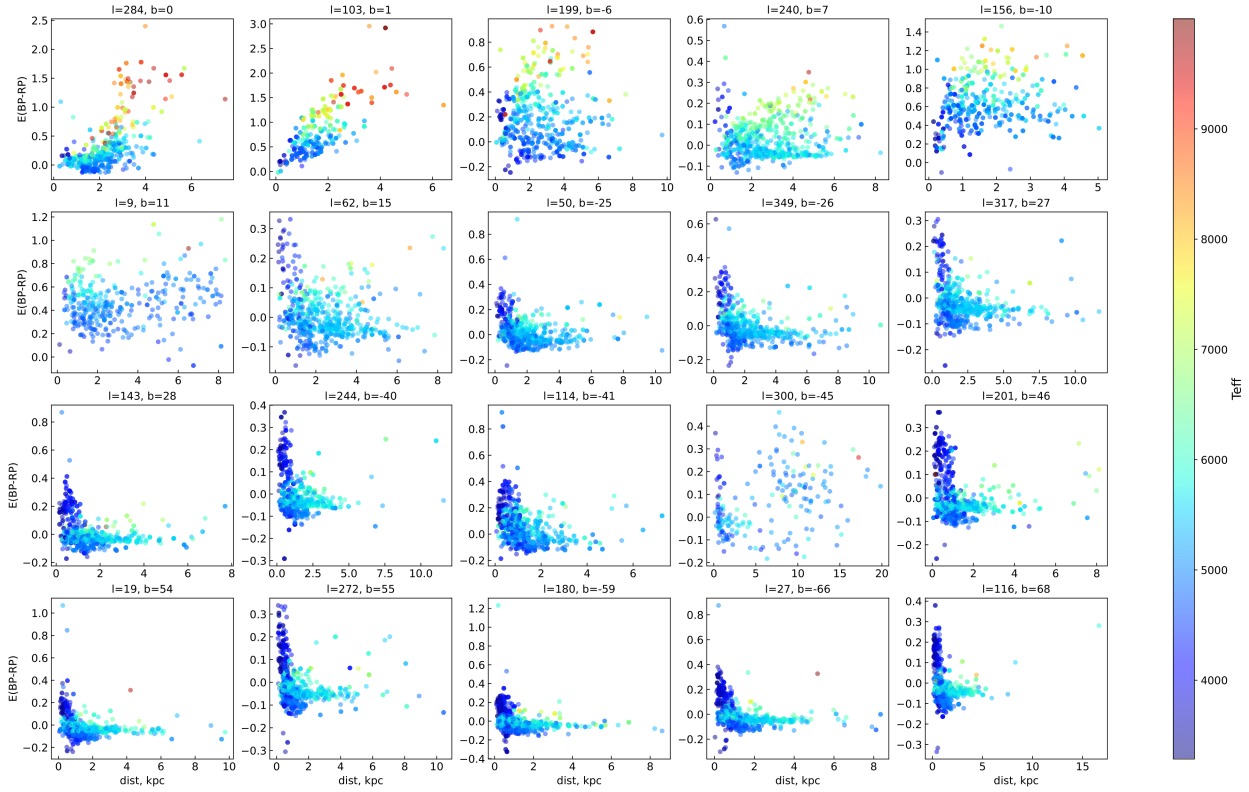

Figure 4: Dependence of reddening to distance with the effective temperature of the object indicated by colour. Abnormally large reddening at close distances is associated with low temperatures.

It should be noted that since SVM models do not allow missing values, models are only applied to data without missing values.

The hyperparameters received with optuna for XGBoost are:

```
max_depth=10,
learning_rate=0.034775804985354486,
subsample=0.9916425167426951,
gamma=0.5183702022766637,
reg_alpha=0.42379209845380494
```

with the overall precision and recall scores on the test dataset, respectively, 0.958 and 0.977.

The hyperparameters for SVM Classifier are:

```
C=0.7285456142566975
kernel=linear
desicion_function_shape=ovr
gamma=0.00012025962588239793
```

The precision and recall scores on the test part of the dataset of 0.680 and 0.859, respectively.

We have also tried the voting strategy. Voting Classifier is not a model itself but a way to take into account the peculiarities of each algorithm. We used the soft voting mode, meaning that the label of the object is predicted taking into account the probability of being an object of this class, calculated by each model. We obtain the precision score of Voting Classifier of 0.940 and a recall score of 0.953 on the test part of the dataset.

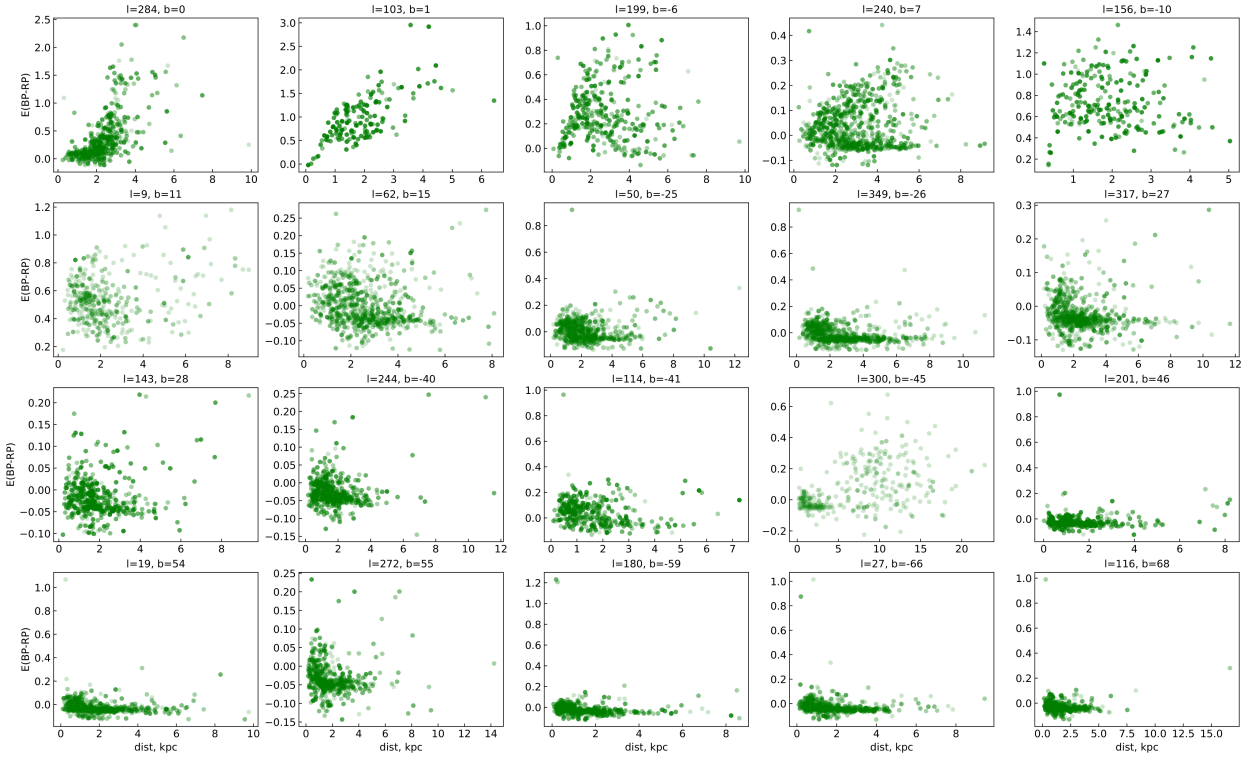

Figure 5: Reddening dependence on distance after applying two filters: $G_{mag} < 17.5$ and $T_{\rm eff} \geq 5000$. Outliers persist with excessively high reddening at short distances and nearly zero reddening at long distances

4.3 Verification of the results with high-resolution surveys

We assume that the quality of reddening vs distance plots relies mainly on temperature quality, so we expect the objects selected by the models to be the ones with higher $T_{\rm eff}$ quality. We test this hypothesis by applying the model to the intersection of Gaia DR3 and APOGEE DR16 or GALAH DR2 in different parts of the sky. The cross-matching of Gaia DR2 with both APOGEE DR16 and GALAH DR2 was done by [15] while Gaia DR2 and Gaia DR3 are connected through the official cross-matching catalogue presented by Gaia.

We have applied XGBoost, SVM and Voting Classifier to the intersection of Gaia DR3 and APOGEE DR16/GALAH DR2. Fig 8 and Fig. 9 show the result of the application of the models to the APOGEE-Gaia and GALAH-Gaia datasets, respectively. The left panels are the comparison of Gaia DR3 temperatures with temperatures of the same objects from the high-resolution spectroscopic survey. The right panels are density plots of the difference between temperatures.

It is seen from the results that the best coherence between Gaia temperatures and high-quality temperatures is obtained via XGBoost model, while both SVM and Voting Classifier allow more divergent results. This result also has flaws, especially in the $T_{\rm eff} > 8000$ region in the case of APOGEE. It does not, however, reproduce in the case of GALAH data, meaning that the difference may occur through some peculiarities of the APOGEE survey. The design of labelling also does not allow for the lowest temperatures.

On the other hand, SVM, being the model with the lowest precision and recall scores, also shows the poorest coherence with high-resolution surveys. Despite the fact that Voting Classifier improves the performance of the SVM model, it is not without its shortcomings. Both APOGEE and GALAH datasets indicate that SVM and Voting Classifier lead to the selection of the objects, whose temperatures to a large extent systematically differ from the temperatures of high-resolution surveys.

5 Discussion and conclusion

In this work, we have investigated the reliability of the temperatures derived from low-resolution BP/RP spectra of Gaia DR3. We have assessed the temperatures both in an indirect way, using color-$T_{\rm eff}$ relations to produce reddening vs

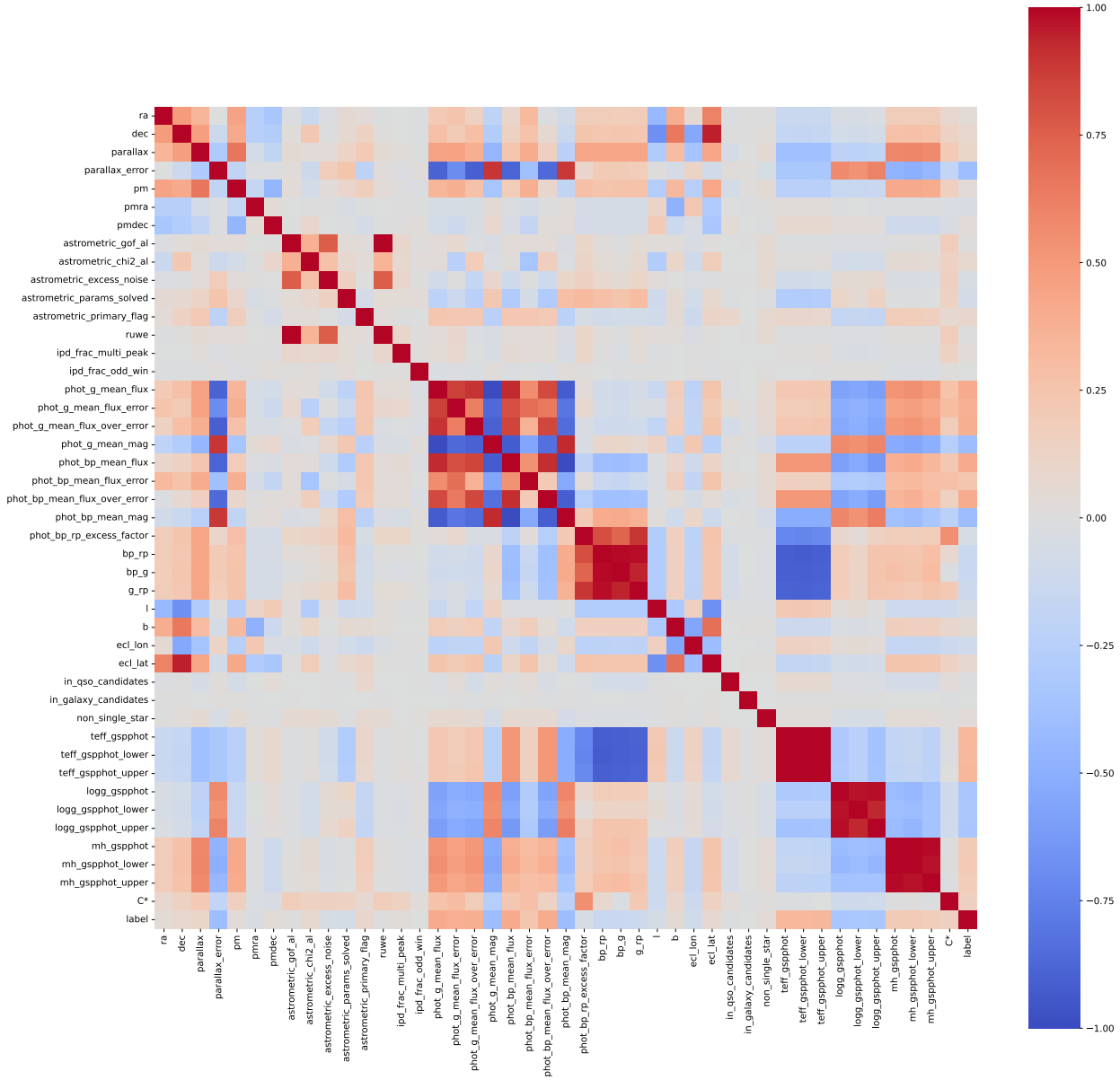

Figure 6: Correlation matrix displaying the relationships between selected columns in the catalogue. Each cell in the matrix represents the correlation coefficient between two variables, with colour intensity indicating the strength and direction of the correlation.

distance plots, and direct way by comparing the temperatures from Gaia DR3 with the temperatures from high-resolution surveys, such as APOGEE and GALAH.

By studying reddening vs distance plots, we found that the effective temperatures below 5000 degrees, as well as temperatures of objects with magnitudes $G > 17.5$, produce odd reddening results. Thus, we conclude that these objects have outrageous temperatures. There were, however, other objects with abnormal reddening which we assume are also associated with an incorrectly determined temperature.

We investigated if we can access reliable temperatures using Machine Learning techniques. We have labelled all the objects with an outstanding deviation in reddening as objects of a negative class and trained the XGBoost and SVM models to classify objects into two classes. The XGBoost model has shown decent precision and recall scores, while the precision score of SVM model was poor.

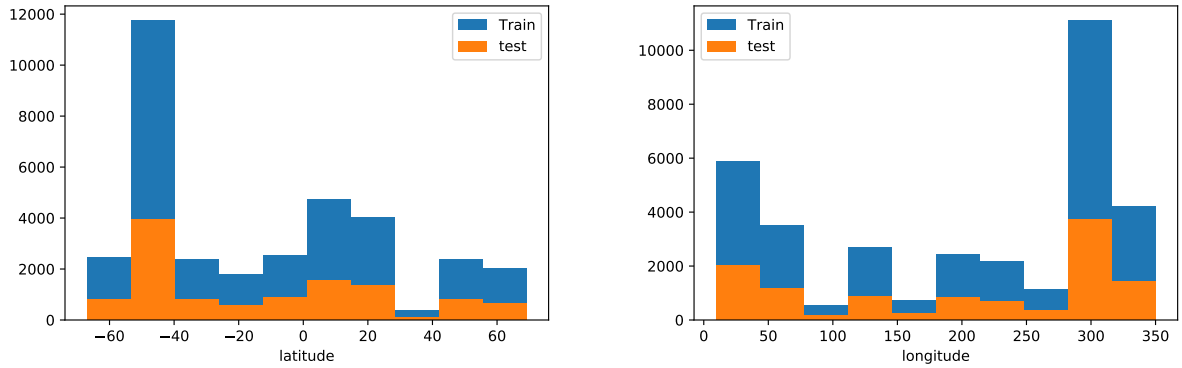

Figure 7: Distribution of train and test objects by position on the sky. Histograms show that train and test objects are proportionally distributed in all areas.

As the labelling for the training dataset was rather empirical, we have evaluated the results by comparison of Gaia DR3 temperatures and high-quality temperatures from APOGEE/GALAH. From the rather good agreement between high-quality temperatures and the temperatures of objects selected by XGBoost model, we can conclude that the criteria we used for the labelling process are eligible. They are not only empirical but also indicate the real reasons associated with the temperature discrepancy.

Although the XGBoost model can extract features with better-than-average temperatures from the entire dataset, it still has drawbacks. One of them is the absence of lower temperatures due to the labelling process. Another unambiguous result is the large discrepancy between APOGEE temperatures and the Gaia DR3 ones at $T_{\rm eff} > 8000$. Such discrepancies do not occur while applying the same model to the GALAH data, in this case, there is a clear cutoff around $T_{\rm eff} = 8000$. This may indicate some features of the APOGEE survey compared to GALAH.

On the contrary, SVM model and Voting Classifier have shown no efficiency in extracting reliable temperatures. SVM model has shown both poor precision results and failed the test on Gaia - APOGEE/GALAH intersection. The SVM model's poor performance in temperature classification, compared to XGBoost, can be attributed to two key factors: sensitivity to non-linear relationships and challenges posed by the imbalanced class distribution of the temperature dataset. SVM struggles to capture complex non-linear patterns in the temperature data, while XGBoost excels at modelling non-linear relationships. Moreover, the imbalanced class distribution, with a majority of objects having deviant temperatures and a smaller portion exhibiting good temperatures, can hinder SVM's ability to effectively learn from the minority class, resulting in biased results.

Despite the fact that the temperatures of the objects selected by the models, primarily by the XGBboost, still have a systematic difference from the APOGEE/GALAH temperatures and are only a small part of the complete Gaia dataset, the results of the work indicate that it is possible in principle to extract a subset of sufficiently accurate temperatures from the Gaia. These temperatures then could be used in various astronomical applications, such as individual reddening estimation.

Acknowledgments

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (`https://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, `https://www.cosmos.esa.int/web/gaia/dpac/consortium`).

6 List of reviewers remarks

———————— REVIEW 1 ————————

SUBMISSION: 52

TITLE: Assessing the Reliability of Gaia DR3 Effective Temperatures

AUTHORS: Aleksandra Avdeeva, Dana Kovaleva and Oleg Malkov

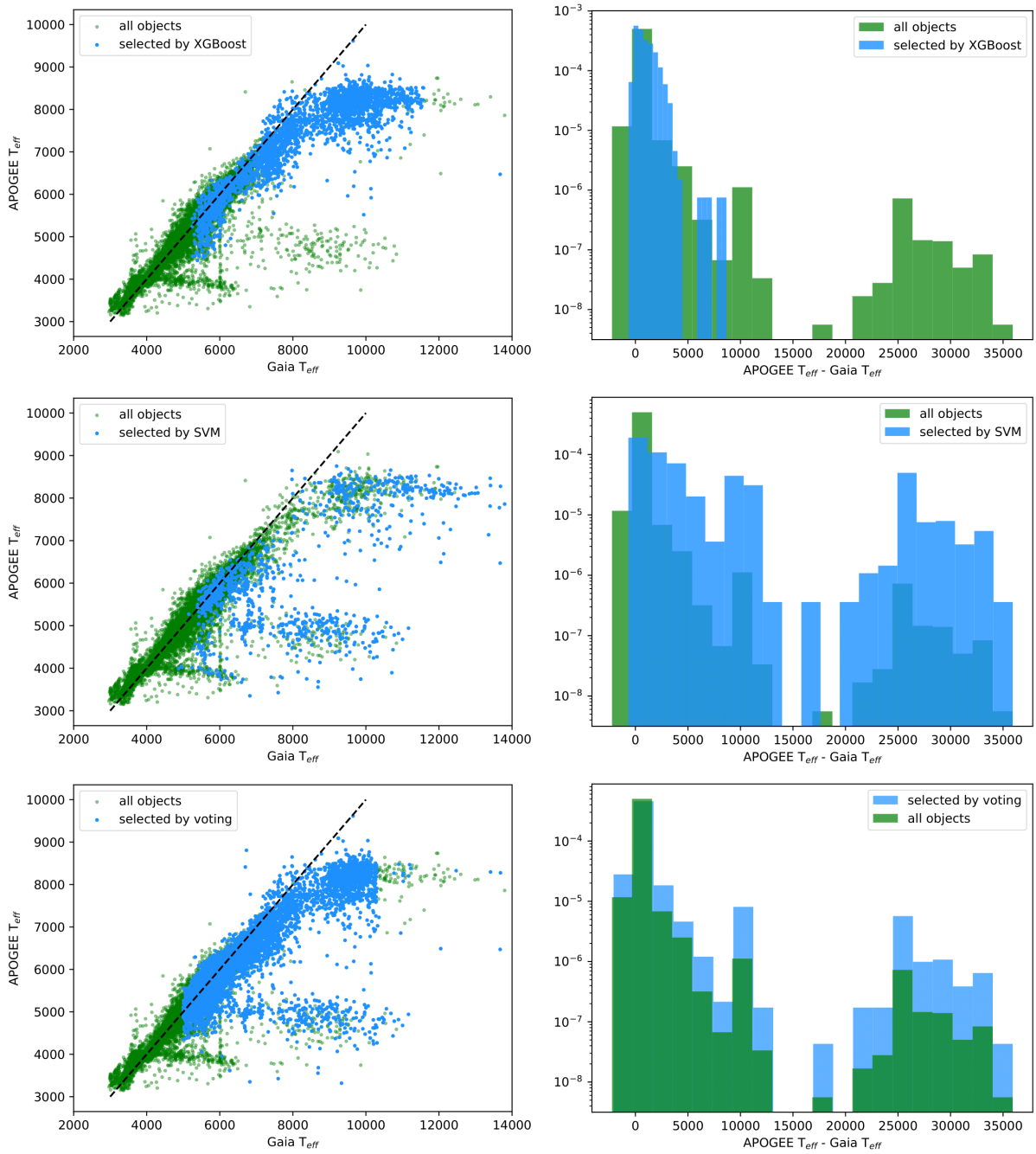

Figure 8: Comparison of Gaia DR3 temperatures with high-resolution spectroscopic survey temperatures for the APOGEE-Gaia dataset. Left panels show the temperature comparison, while right panels depict density plots of temperature differences.

——— Overall evaluation ———

SCORE: 3 (strong accept)

—— TEXT: The article is devoted to the study of the reliability of effective temperatures derived from Gaia DR3 for interstellar reddening calculations using machine learning methods. The task is relevant and no significant remarks found. The work was done at a high level and is recommended for publication.

——————— REVIEW 2 ———————

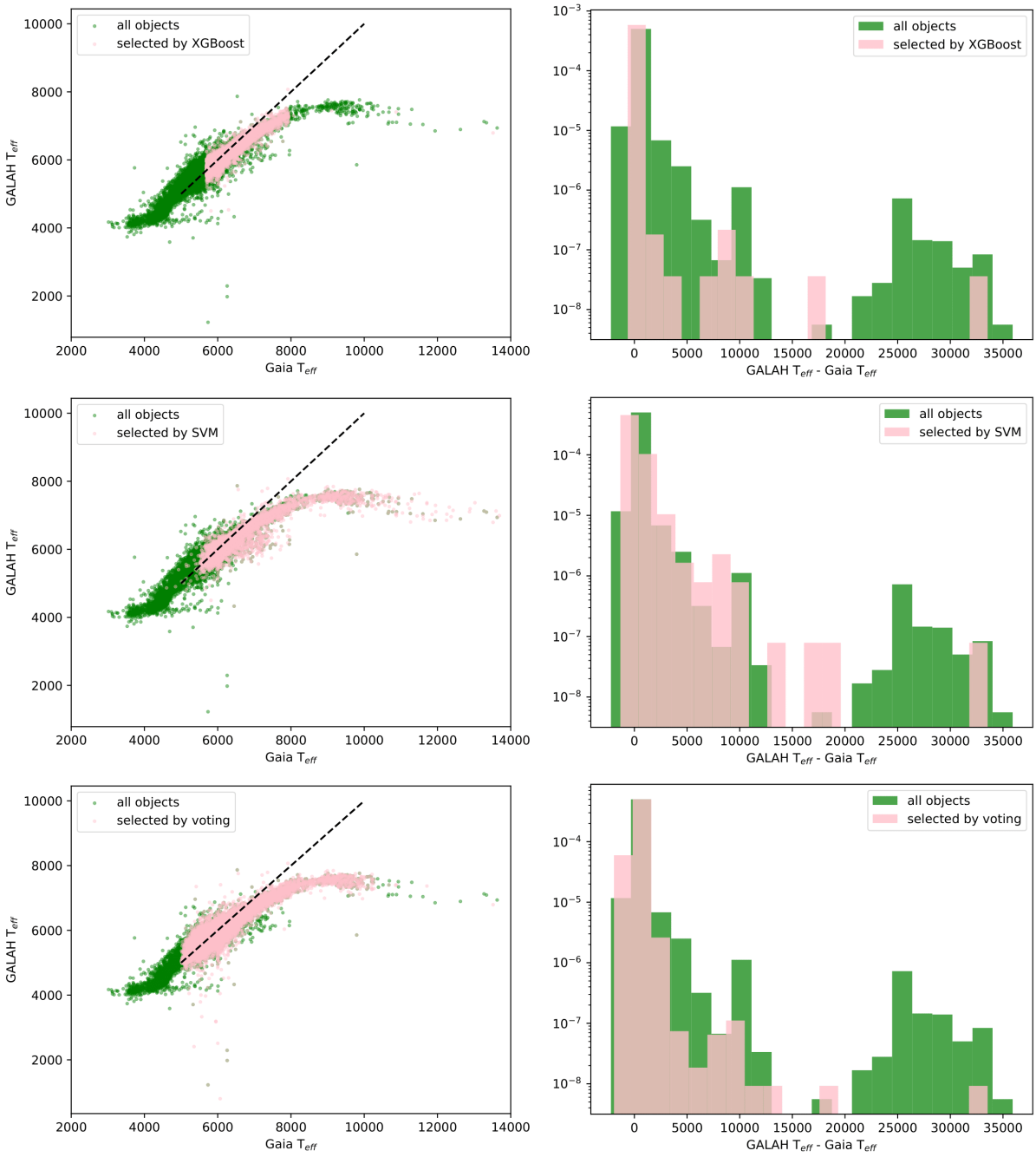

Figure 9: Evaluation of Gaia DR3 temperatures compared to high-resolution spectroscopic survey temperatures for the GALAH-Gaia dataset. The left panels illustrate the temperature comparison, while the the right panel presents density plots depicting the distribution of temperature differences.

SUBMISSION: 52

TITLE: Assessing the Reliability of Gaia DR3 Effective Temperatures

AUTHORS: Aleksandra Avdeeva, Dana Kovaleva and Oleg Malkov

——— Overall evaluation ———

SCORE: 3 (strong accept)

—— TEXT: (translated by authors, original in Russian) Article is on the topic of the conference. There are no serious comments. It is necessary to fix the typos. Small comments are marked in the text of the original article.

———————— REVIEW 3 ————————

SUBMISSION: 52

TITLE: Assessing the Reliability of Gaia DR3 Effective Temperatures

AUTHORS: Aleksandra Avdeeva, Dana Kovaleva and Oleg Malkov

———— Overall evaluation ————

SCORE: 1 (weak accept)

—— TEXT: In this work authors focus on evaluating the dependability of effective temperatures derived from Gaia DR3 to individual reddening of the stars. The data set has around 55000 samples from 20 areas of the sky.

It is not stated if the 60/20/20 split is random or done by the areas of the sky. To have more robust metrics, the split should be done by sky areas with some cross-validation added.

Authors use gradient boosting model and SVM to build binary classifiers. There are two labels – the objects that are "bad" and "positive". The labeling is done using heuristic rule. It is not stated, how imbalanced are the classes. Only precision and recall metrics are given. Depending on how imbalanced the classes are, I suggest adding ROC-AUC and PR-AUC, as well as F1-score. Also, as the positive class seems to be of greater interest, so the number of false positives is of importance and should be tracked and reduced.

Verification procedure is qualitative, not quantitative, and is based on visual comparison of high-resolution spectroscopic survey temperatures. It would be interesting to add some quantitative score here, like error or correlation coefficient.

Some more advanced classifiers such as DNN can be added to the model list.

In several paces (e. g. last sentence of section 1) there are unfinished phrases [This article is organised as follows. . .]. Results are provided in heavy or corrupted (?) pictures which do not allow article to be viewed in Adobe Acrobat or browser in convenient way.

7 Our answers

We greatly thank the reviewers for careful reading and high evaluation of the work. As for the remarks, we have fixed the images and most of the typos and unfinished sentences. We have also specified how imbalanced the dataset is and specified the train/validation/test split in more details (see Fig.7). Other metrics and quantitative comparison will be added closer to the publication.

References

[1] T. Prusti, J. H. J. de Bruijne, and Gaia Collaboration. The gaia mission. , 595:A1, November 2016. doi: 10.1051/0004-6361/201629272. URL https://ui.adsabs.harvard.edu/abs/2016A&A...595A...1G.

[2] A. Vallenari, A. G. A. Brown, and Gaia Collaboration. Gaia data release 3: Summary of the content and survey properties. *arXiv e-prints*, art. arXiv:2208.00211, July 2022. doi: 10.48550/arXiv.2208.00211. URL https://ui.adsabs.harvard.edu/abs/2022arXiv220800211G.

[3] O. Creevey, R. Sordo, and F. Pailler and. Gaia data release 3. astrophysical parameters inference system (apsis). i. methods and content overview. *Astronomy &amp Astrophysics*, jul 2022. doi: 10.1051/0004-6361/202243688.

[4] R. Andrae, M. Fouesneau, R. Sordo, C.A.L. Bailer-Jones, T. Dharmawardena, J. Rybizki, F. De Angeli, H.E.P. Lindstrom, D. Marshall, R. Drimmel, A.J. Korn, C. Soubiran, N. Brouillet, and L. Casamiquela and. Gaia data release 3: Analysis of the gaia BP/RP spectra using the general stellar parameterizer from photometry. *Astronomy &amp Astrophysics*, jun 2022. doi: 10.1051/0004-6361/202243462.

[5] Henrik Jönsson, Jon A. Holtzman, and et. al. Apogee data and spectral analysis from sdss data release 16: Seven years of observations including first results from apogee-south. , 160(3):120, September 2020. doi: 10.3847/1538-3881/aba592. URL https://ui.adsabs.harvard.edu/abs/2020AJ....160..120J.

[6] Luca Casagrande, Jane Lin, Adam D. Rains, Fan Liu, Sven Buder, Jonathan Horner, Martin Asplund, Geraint F. Lewis, Sarah L. Martell, Thomas Nordlander, Dennis Stello, Yuan-Sen Ting, Robert A. Wittenmyer, Joss Bland-Hawthorn, Andrew R. Casey, Gayandhi M. De Silva, Valentina D'Orazi, Ken C. Freeman, Michael R. Hayden, Janez Kos, Karin Lind, Katharine J. Schlesinger, Sanjib Sharma, Jeffrey D. Simpson, Daniel B. Zucker, and Tomaž

Zwitter. The galah survey: effective temperature calibration from the infrared flux method in the gaia system. , 507(2):2684–2696, October 2021. doi: 10.1093/mnras/stab2304. URL https://ui.adsabs.harvard.edu/abs/2021MNRAS.507.2684C.

[7] G. Gilmore, S. Randich, and et. al. The gaia-eso public spectroscopic survey. *The Messenger*, 147:25–31, March 2012. URL https://ui.adsabs.harvard.edu/abs/2012Msngr.147...25G.

[8] Mark J. Pecaut and Eric E. Mamajek. Intrinsic colors, temperatures, and bolometric corrections of pre-main-sequence stars. , 208(1):9, September 2013. doi: 10.1088/0067-0049/208/1/9. URL https://ui.adsabs.harvard.edu/abs/2013ApJS..208....9P.

[9] A. Ulla, O. L. Creevey, M. A. Álvarez, R. Andrae, C. A. L. Bailer-Jones, I. Bellas-Velidis, E. Brugaletta, R. Carballo, C. Dafonte, L. Delchambre, T. Dharmawardena, R. Drimmel, M. Fouesneau, Y. Frémat, D. Garabato, D. Hatzidimitriou, U. Heiter, G. Kordopatis, A. J. Korn, A. Lanzafame, A. Lobel, M. Manteiga, D. J. Marshall, F. Pailler, L. Pallas-Quintela, A. Recio-Blanco, J. Rybizki, L. M. Sarro Baro, M. Schultheis, R. Sordo, C. Soubiran, F. Thévenin, and A. Vallenari. Gaia dr3 documentation chapter 11: Astrophysical parameters. Gaia DR3 documentation, European Space Agency; Gaia Data Processing and Analysis Consortium. Online at <A href="https://gea.esac.esa.int/archive/documentation/GDR3/index.html">https://gea.esac.esa.int/archive/documentation/GDR3/index id. 11, June 2022. URL https://ui.adsabs.harvard.edu/abs/2022gdr3.reptE..11U.

[10] M. Riello, F. De Angeli, D. W. Evans, P. Montegriffo, J. M. Carrasco, G. Busso, L. Palaversa, P. W. Burgess, C. Diener, M. Davidson, N. Rowell, C. Fabricius, C. Jordi, M. Bellazzini, E. Pancino, D. L. Harrison, C. Cacciari, F. van Leeuwen, N. C. Hambly, S. T. Hodgkin, P. J. Osborne, G. Altavilla, M. A. Barstow, A. G. A. Brown, M. Castellani, S. Cowell, F. De Luise, G. Gilmore, G. Giuffrida, S. Hidalgo, G. Holland, S. Marinoni, C. Pagani, A. M. Piersimoni, L. Pulone, S. Ragaini, M. Rainer, P. J. Richards, N. Sanna, N. A. Walton, M. Weiler, and A. Yoldas. Gaia early data release 3. photometric content and validation. , 649:A3, May 2021. doi: 10.1051/0004-6361/202039587. URL https://ui.adsabs.harvard.edu/abs/2021A&A...649A...3R.

[11] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. doi: 10.1145/2939672.2939785.

[12] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[13] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Andreas Müller, Joel Nothman, Gilles Louppe, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, October 2011. doi: 10.48550/arXiv.1201.0490. URL https://ui.adsabs.harvard.edu/abs/2011JMLR...12.2825P.

[14] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework.

[15] M. Tsantaki, E. Pancino, P. Marrese, S. Marinoni, M. Rainer, N. Sanna, A. Turchi, S. Randich, C. Gallart, G. Battaglia, and T. Masseron. Survey of surveys. i. the largest compilation of radial velocities for the galaxy. , 659:A95, March 2022. doi: 10.1051/0004-6361/202141702. URL https://ui.adsabs.harvard.edu/abs/2022A&A...659A..95T.