UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

PREDICTING STELLAR AGES FROM CHEMICAL ABUNDANCES WITH XGBOOST: A MACHINE LEARNING APPROACH TO GALACTIC ARCHAEOLOGY

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By TUCKER CAPPS

Norman, Oklahoma
2025

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/341606

PREDICTING STELLAR AGES FROM CHEMICAL ABUNDANCES WITH XGBOOST:
A MACHINE LEARNING APPROACH TO GALACTIC ARCHAEOLOGY

A THESIS APPROVED FOR THE

GALLOGLY COLLEGE OF ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Charles Nicholson, Chair

Dr. Andres Gonzalez Huertas

Dr. Michael Hayden

©Copyright by TUCKER CAPPS 2025
All Rights Reserved.

Contents

List of Tables **v**

List of Figures **vi**

Abstract **viii**

1 Introduction **1**
1.1 Motivation 1
1.2 Objectives 2
1.3 Scope and Significance 3

2 Background **4**
2.1 Stellar Ages and Isochrone Fitting 4
2.2 Chemical Abundances as Age Indicators 8
2.3 Machine Learning for Stellar Age Determination 10

3 Data **12**
3.1 GALAH Data Release 4 12
3.2 Data Selection 13
3.3 Training+Test Set 18

4 Methodology **19**
4.1 Bayesian Isochrone Fitting 19
4.2 eXtreme Gradient Boosting (XGBoost) 22

5 Results **28**
5.1 Cross-Validation and Final Hyperparamters 28
5.2 Model Performance on Training Set 29
5.3 Model Performance on Test Set 30
5.4 Residual Analysis 33
5.4.1 Training Set Residuals 33
5.4.2 Test Set Residuals 38
5.4.3 Test Set Residuals vs Chemical Abundances 41
5.5 Feature Importance 43
5.6 Age-Abundance Relations 45

6 Discussion **51**

7 Conclusion **56**

References **58**

List of Tables

1 Quality cuts used to select data from GALAH DR4. 17
2 Additional cuts used to select MSTO data used for model training. . 19
3 Hyperparameter grid used to tune the XGBoost model, consisting of 103,680 possible hyperparameter combinations. 28
4 Optimal configuration of hyperparameters that minimized the average RMSE across all folds. 29

List of Figures

1 "In the Hertzsprung-Russell diagram the temperatures of stars are plotted against their luminosities. The position of a star in the diagram provides information about its present stage and its mass. Stars that burn hydrogen into helium lie on the diagonal Branch, the so-called Main Sequence." **Credit:** ESO 6

2 A model isochrone with solar metallicity and solar age. The near horizontal line of points near the bottom of the plot denotes the Main Sequence. The region in the red box is the Main Sequence Turn-Off. The diagonal line that extends up and to the right is the Giant Branch. Each data point on this plot represents a star with a unique initial mass, with initial mass increasing as you move along the Main Sequence and up to the Giant Branch. 7

3 Periodic table of elements displaying the relative abundance from each nucleosynthetic production site. **Credit:** Kobayashi et al. (2020) 8

4 A set of theoretical isochrones having the same metallicity. Notice how the Main Sequence and Giant Branch are stacked on top of each other while the data points in the Main Sequence Turn-Off are discernible. 11

5 A Kiel Diagram of 663,075 stars having flag_sp = 0 from the GALAH DR4 dataset. Notice the Main Sequence, Main Sequence Turn-Off, and Giant Branch. 14

6 A Kiel Diagram of 421,934 stars having good measurements for the desired elemental abundances. 15

7 A Kiel Diagram of 412,405 quality-ensured stars having $-1.0 \leq [Fe/H] \leq 0.5$ (disk stars). 16

8 The ages predicted by the XGBoost model versus the ground-truth ages from isochrone fitting for the training set. The dashed red line denotes a 1:1 relationship. 31

9 The ages predicted by the XGBoost model versus the ground-truth ages from isochrone fitting for the test set. The dashed red line denotes a 1:1 relationship. 32

10 Histogram of the residuals (true minus predicted values) for the training set, displaying an approximately normal shape centered at zero. 34

11 Residuals versus the predicted age for the training set. Evidently, the model is much more confident at the extremes and less confident at intermediate ages. 35

12 Histogram of the isochrone-derived stellar ages in the training set, showing the skew toward younger ages and under-representation of old stars. 37

13 Histogram of the residuals (true minus predicted values) for the test set, displaying an approximately normal shape centered at zero. . . 39

14 Residuals versus predicted age for the test set, showing a clear increase in variance at the extremes and lower variance at intermediate ages. 40
15 Panel displaying the residuals versus the features for all 15 features utilized in this study. 42
16 Feature importance based on gain for the XGBoost model. [Mg/Fe] exhibits the highest gain, indicating it contributes the most to reducing the model's loss function during training. 46
17 Feature importance based on weight for the XGBoost model. [Fe/H] has the highest weight, suggesting it played the largest role in the model's decision making. 47
18 Pearson correlation matrix for the elemental abundance ratios used as model features. Strong correlations among α-elements and between certain s-process elements are evident, consistent with their astrophysical origins. 48
19 Elemental abundances versus XGBoost predicted age for all stars in the test set. [Fe/H] shows a monotonic decrease with age, which is physically sound. Additionally, α-element abundances show an increase with age, which is also physically sound. 50
20 Elemental abundances versus XGBoost predicted age for all disk stars. The monotonic decrease with age of [Fe/H] is evident, just as before. The increase with age of the α-element abundance is also present. These patterns indicate the model is generalizing well to unseen stellar populations. 52

Abstract

Estimating stellar ages remains a central challenge in astrophysics, particularly for stars outside the narrow evolutionary phases where traditional methods like isochrone fitting are reliable. This study explores the use of supervised machine learning, specifically, the Extreme Gradient Boosting (XGBoost) algorithm, to infer stellar ages from chemical abundance ratios derived from the GALAH DR4 spectroscopic survey. A training set of 20,303 Main Sequence Turn-Off stars was constructed using strict quality cuts and age labels obtained via Bayesian isochrone fitting. The XGBoost model was trained on 15 elemental abundance ratios and tuned through exhaustive hyperparameter optimization with five-fold cross-validation.

The final model achieves a root mean squared error (RMSE) of 1.38 Gyr on a held-out test set, generalizes well across stellar populations not seen during training, and reproduces known chemical evolution trends such as the anti-correlation between [Fe/H] and age and the positive correlation between [α/Fe] and age. Feature importance analysis reveals that [Mg/Fe], [Ca/Fe], and [Y/Fe] are the most informative predictors of age, consistent with theoretical nucleosynthetic expectations. Residual analyses expose a tendency for the model to regress predictions toward the mean age in underrepresented regions, particularly for very young and very old stars.

These results demonstrate that machine learning can extract robust age information from chemical abundance patterns and extend age-dating to stellar populations for which traditional techniques break down. This work highlights the power of data-driven approaches in galactic archaeology and paves the way for scalable age estimation in next-generation spectroscopic surveys.

1 Introduction

1.1 Motivation

In recent years, machine learning (ML) has become an indispensable tool across many Branches of science, enabling researchers to extract meaningful patterns from complex, high-dimensional datasets (Jordan and Mitchell, 2015; Carleo et al., 2019). In fields such as genomics, materials science, and climate modeling, machine learning techniques have unlocked new ways of analyzing large and noisy datasets that defy traditional analytic approaches. Physics has been no exception, with machine learning methods now applied in condensed matter, quantum many-body systems, and cosmology (Mehta et al., 2019).

Astronomy, in particular, stands out as a data-rich science where machine learning is rapidly gaining ground (Baron, 2019; Fluke and Jacobs, 2019). The proliferation of large-scale surveys such as Gaia, SDSS, APOGEE, and GALAH has provided unprecedented volumes of stellar data, challenging traditional analysis pipelines and motivating the adoption of data-driven approaches. In this context, machine learning has been applied to problems ranging from galaxy morphology classification to photometric redshift estimation, and increasingly, to stellar parameter inference.

One domain where machine learning shows exceptional promise is in determining stellar ages—a longstanding challenge in astrophysics. Stellar age is not a directly observable quantity, yet it is a fundamental parameter for understanding galactic evolution. Together with mass, age is also one of the two principal quantities that govern stellar evolution. Several age-dating methods that do not involve machine learning exist, but every method has advantages and shortcomings that depend on stellar type, data quality, and data availability. These traditional methods and their specifics are discussed in Section 2.1.

Recent work has focused on using machine learning to model the complex, non-linear relationships between stellar properties and chemical abundances derived from spectra. For instance, neural networks and data-driven models like The Cannon have been used to estimate stellar parameters from spectral data (Leung and Bovy, 2018; Ness et al., 2015; Fabbro et al., 2017). Other studies have applied supervised learning directly to abundance patterns to infer stellar ages, particularly for stars in surveys such as GALAH (Hayden et al., 2022). Such approaches suggest that chemical abundances, particularly ratios involving α-elements and neutron-capture elements, can serve as empirical proxies, or *chemical clocks*, for age prediction (Nissen, 2015; Feltzing et al., 2016).

1.2 Objectives

The primary objective of this project is to evaluate the effectiveness of supervised machine learning, specifically the Extreme Gradient Boosting (XGBoost) algorithm (Chen and Guestrin, 2016), for predicting stellar ages from chemical abundance ratios in the fourth data release (DR4) of the GALAH survey (Buder et al., 2024). Building on recent work that has shown an empirical relation between certain abundance patterns and stellar age (Hayden et al., 2022), this study aims to develop a model that generalizes beyond stars for which a reliable age-dating method exists and provides reliable age estimates for a broad range of stellar populations.

To accomplish this, the model will be trained on a subset of stars with well-determined ages, using their abundance ratios as input features. The predictive performance of the model will be assessed using cross-validation and its generalization will be evaluated by applying it to unseen stars of the same population and those of other populations (stars in different evolutionary phases). The resulting age distribution will be compared to known chemical-age trends in the Milky Way as a qualitative validation.

In addition to predictive and generalization performance, this project also aims to interpret the model's behavior through feature importance and residual analysis to identify which abundance ratios contribute most strongly to age prediction and find where the method begins to fail. The study will identify current limitations and propose future improvements for extending abundance-based machine learning models to the next generation of spectroscopic surveys (surveys in which the primary data product is the atomic spectra of the object observed).

1.3 Scope and Significance

This project focuses on the application and evaluation of the XGBoost algorithm for predicting stellar ages from chemical abundance ratios, using high-quality spectroscopic data from the GALAH DR4 survey. The training set is restricted to high signal-to-noise stars with both reliable ages and abundance measurements, while the broader goal is to generalize these predictions across a wider range of stellar evolutionary phases.

The significance of this work lies in both its scientific and methodological contributions. By improving age estimation for stars that fall outside the regime for which current age estimation methods exist, this study offers a complementary approach for reconstructing galactic evolutionary history. In particular, chemically derived ages may provide a path forward in extending current age-dating methods to stellar populations in which they currently do not work.

More broadly, this project contributes to the growing integration of machine learning in astrophysical data analysis. As future spectroscopic surveys expand in scale and complexity, robust, interpretable, and scalable machine learning methods will be essential. The tools and insights developed here are intended to serve as a foundation for that future work.

2 Background

2.1 Stellar Ages and Isochrone Fitting

Determining the ages of stars remains one of the most persistent challenges in astrophysics. Unlike parameters such as mass, temperature, or chemical composition, which can be inferred directly from observations, stellar age must be estimated indirectly using models of stellar evolution. Yet age is a critical quantity, as it underpins our understanding of stellar life-cycles, the formation history of the Milky Way, and the broader process of chemical enrichment in galaxies (Soderblom, 2010). Because of the importance of stellar age, multiple age-dating methods have arisen.

Asteroseismology is a powerful age-dating method that probes a star's internal structure through the study of oscillation modes. These oscillations, detectable as periodic brightness variations, are sensitive to fundamental stellar properties including mass, radius, and age. In particular, observations from space missions like *Kepler* and *CoRoT* have enabled precise measurements of solar-like oscillations in thousands of stars, greatly improving age determinations, especially for Red Giants and Sub-Giants (Chaplin and Miglio, 2013; Hekker and Christensen-Dalsgaard, 2017). Asteroseismic ages can reach precisions of around 10–20% in favorable cases, providing a significant improvement over isochrone-based estimates alone. However, the requirement for high-quality, long-duration photometric time series limits this method to relatively bright and nearby stars.

Gyrochronology is another technique that estimates stellar ages by leveraging the relationship between a star's rotation period and its age. Main sequence stars gradually lose angular momentum over time via magnetized stellar winds, leading to predictable spin-down behavior that can be empirically calibrated against cluster stars with known ages (Barnes, 2007; Meibom et al., 2015). In practice, gyrochronology is most reliable for cool stars (spectral types later than F) and for ages

up to a few gigayears, beyond which magnetic braking becomes less efficient and calibrations become uncertain. Like asteroseismology, gyrochronology provides an independent and physically motivated approach that, when combined with other age-dating methods, helps break degeneracies and enhance our understanding of stellar populations.

One of the most widely used methods—and the method that will be used in this study—for estimating stellar ages is isochrone fitting, which involves comparing a star's observed properties to grids of theoretical models that describe how those properties change over time. These models, known as isochrones, trace the predicted position of stars of a given age and metallicity across a range of masses on the Hertzsprung-Russell (HR) diagram (Bressan et al., 2012; Dotter et al., 2008). Isochrone fitting is particularly effective for Turn-Off stars where age-sensitive features are more pronounced and the spacing between isochrones in parameter space is larger than the errors in the observations.

Figure 1 shows an HR diagram, where stars are plotted by luminosity and effective temperature. The location of a star on this diagram reflects its current evolutionary stage, with the Main Sequence, Sub-Giant Branch, and Red Giant Branch each corresponding to the different phases in stellar evolution. Isochrones model these phases, allowing astronomers to estimate stellar ages by finding the best-fitting theoretical track for a given star or population.

Isochrones are generated using stellar evolution codes that simulate internal processes such as nuclear burning, energy transport, convection, and mass loss over time. Modern tools like PARSEC (Bressan et al., 2012), MIST (Choi et al., 2016), and BaSTI (Hidalgo et al., 2018) provide extensive model grids covering a wide range of stellar masses, metallicities, and evolutionary timescales. From these grids, theoretical luminosities and temperatures are computer for synthetic stars, and their positions on the HR diagram are used to define isochrone curves. An

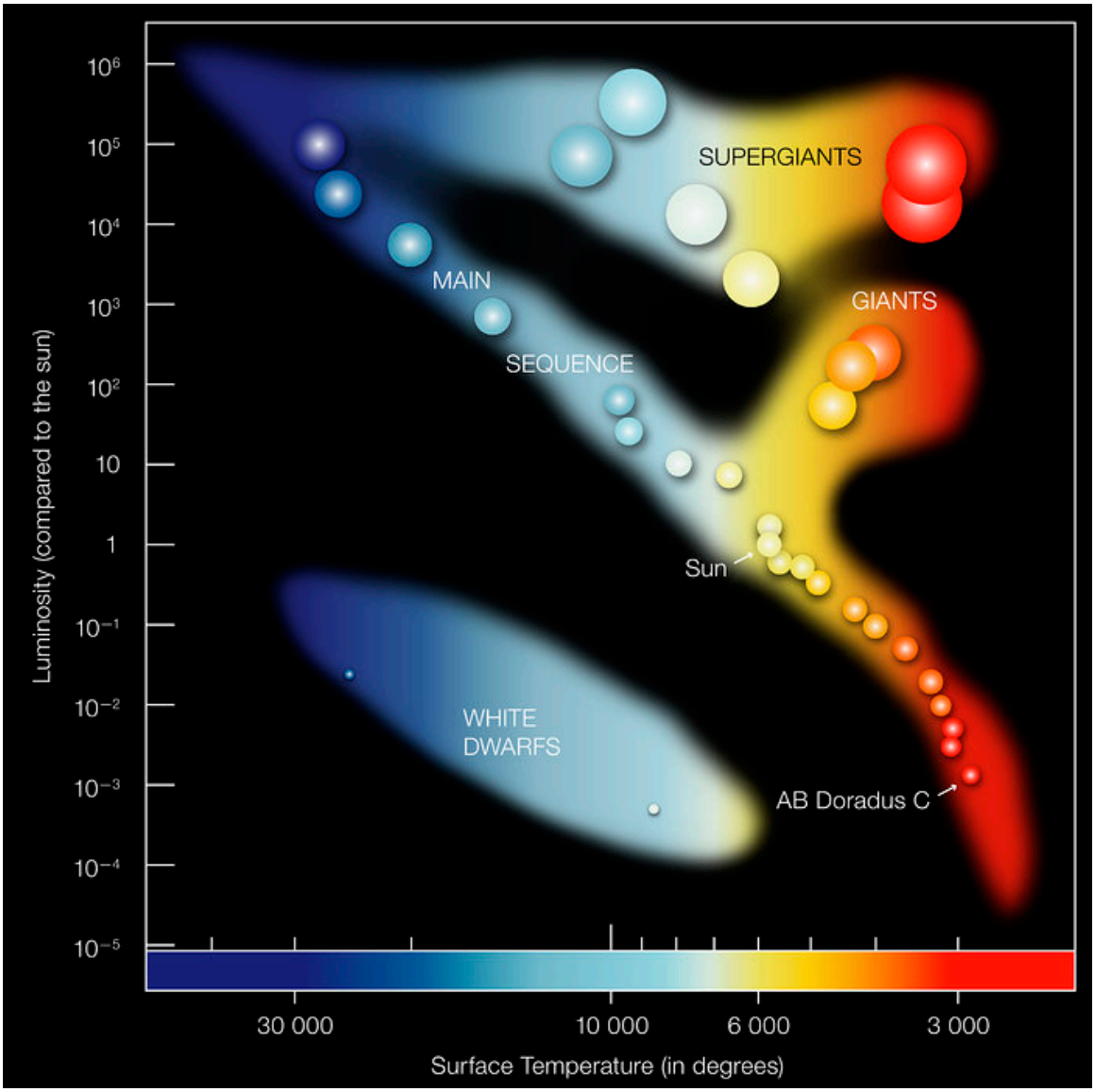


Figure 1: "In the Hertzsprung-Russell diagram the temperatures of stars are plotted against their luminosities. The position of a star in the diagram provides information about its present stage and its mass. Stars that burn hydrogen into helium lie on the diagonal Branch, the so-called Main Sequence." **Credit:** ESO

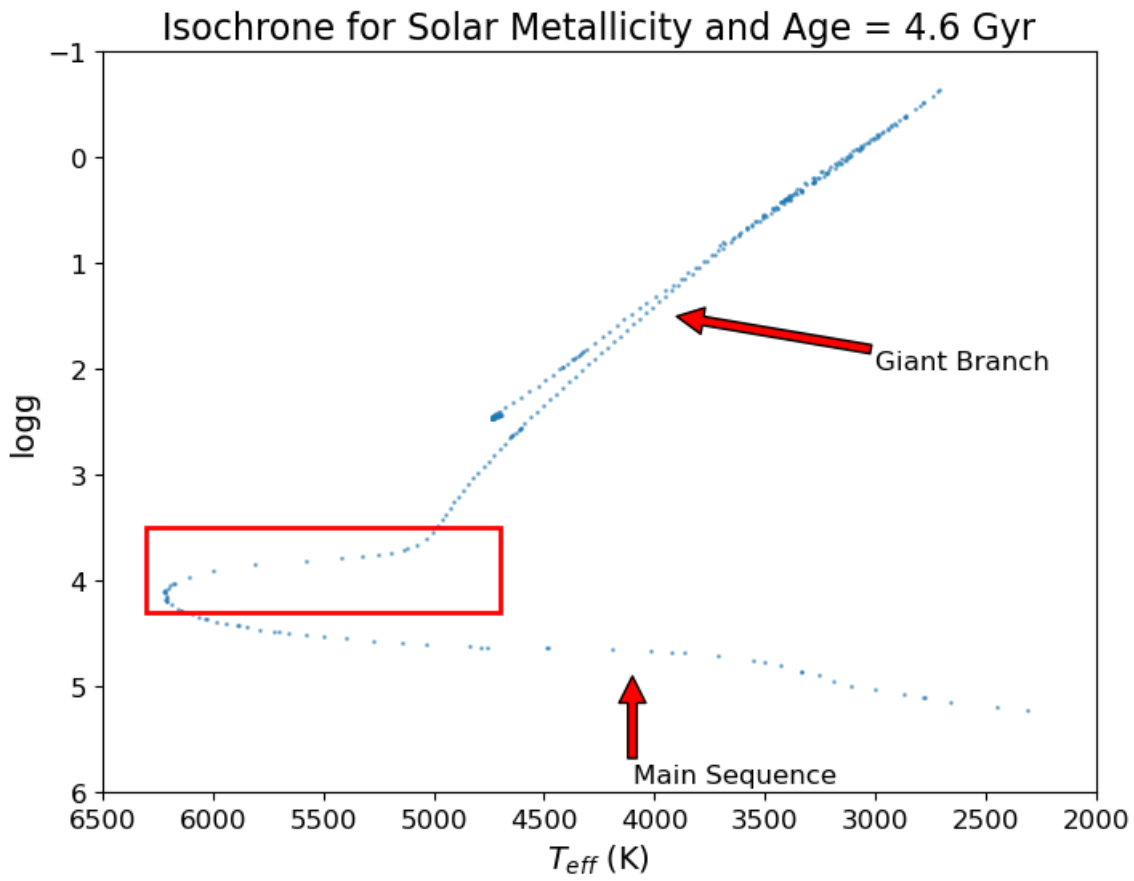


Figure 2: A model isochrone with solar metallicity and solar age. The near horizontal line of points near the bottom of the plot denotes the Main Sequence. The region in the red box is the Main Sequence Turn-Off. The diagonal line that extends up and to the right is the Giant Branch. Each data point on this plot represents a star with a unique initial mass, with initial mass increasing as you move along the Main Sequence and up to the Giant Branch.

example solar metallicity isochrone is shown in Figure 2.

For individual field stars, Bayesian isochrone fitting techniques have become increasingly common. These methods combine observed stellar parameters such as effective temperature, surface gravity, and metallicity with prior distributions on stellar mass and age to derive posterior age estimates (Jørgensen and Lindegren, 2005; Serenelli et al., 2013). Turn-off stars, which are in a rapid phase of evolution and exhibit significant changes in observable parameters over short timescales, are particularly useful for these analyses because they allow for more precise age de-

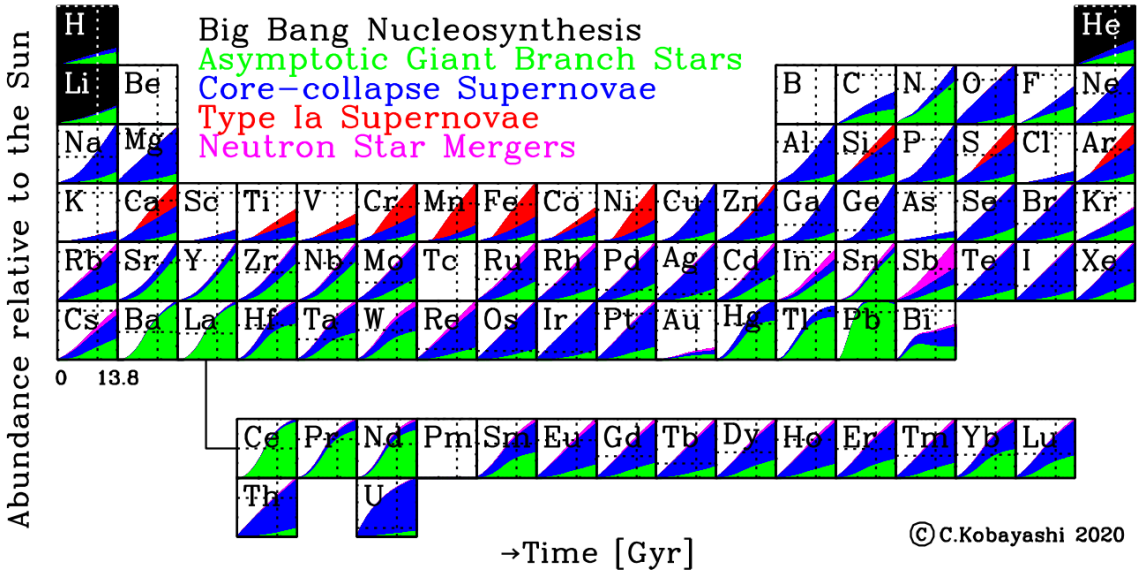


Figure 3: Periodic table of elements displaying the relative abundance from each nucleosynthetic production site. **Credit:** Kobayashi et al. (2020)

terminations.

Despite their widespread use, isochrone-based ages are subject to significant uncertainties in many regions of the HR diagram. On the Main Sequence and Red Giant Branch, isochrones of different ages are tightly clustered, which makes is difficult to distinguish between stars of different ages based on their position alone (Soderblom, 2010). These degeneracies, along with measurement uncertainties, often limit the precision of isochrone-based age estimates to several gigayears for individual stars. This limitation has motivated the exploration of alternative age-dating methods that incorporate different information, such as asteroseismic data or chemical abundances.

2.2 Chemical Abundances as Age Indicators

The chemical composition of the earliest gas clouds in the Universe was set during Big Bang nucleosynthesis, which produced a mixture dominated by hydrogen and helium, with trace amounts of lithium and beryllium (Fields et al., 2020). Heavier

elements, referred to as "metals" in astronomy, were absent. The first generation of stars, Population III stars, formed from this pristine material and initiated the chemical enrichment of the Universe.

Because stellar lifetime is inversely proportional to mass, the most massive stars evolve and die quickly. Stars that are massive enough to end as core-collapse supernovae (Type II, or SNeII) have lifetimes less than 30 million years. These early supernovae contributed significantly to the enrichment of α-elements such as oxygen, magnesium, silicon, and calcium. In contrast, lower-mass stars evolve more slowly and, in some cases, end their lives as Type Ia supernovae (SNe Ia). These result from interactions of a white dwarf and its binary companion and occur on longer timescales, peaking at 1 billion years after star formation. These events are the primary sources of iron-peak elements, including iron, nickel, and chromium (Kobayashi et al., 2020).

This difference in timescales between the fast production of α-elements and the delayed enrichment of iron-peak elements by SNe Ia forms the basis of the chemical age diagnostic. As proposed by Tinsley (1979), early generations of stars exhibit elevated [α/Fe] ratios because they formed before iron-rich material released by SNe Ia began enriching the interstellar medium. As the Galaxy evolves and more iron from delayed SNe Ia accumulates, newly formed stars show progressively lower [α/Fe] ratios and higher metallicity.

This time-dependent abundance pattern means that chemical compositions, particularly ratios like [α/Fe], carry an imprint of a star's formation epoch. These ratios thus serve as powerful empirical proxies for age. Theoretical models, such as those from Kobayashi et al. (2020), provide detailed predictions for how different nucleosynthetic sources contribute to the chemical enrichment of the Galaxy over time.

Figure 3 summarizes the dominant production sites for each element and shows

the cumulative enrichment of the periodic table from the Big Bang to the present. The left edge of each cell corresponds to $t = 0$, while the right edge marks the present day. Elements produced by core-collapse supernovae appear early in cosmic history, while contributions from delayed sources like SNe Ia and asymptotic giant Branch stars begin later, supporting their utility in tracing stellar ages.

By precisely measuring the chemical abundances of stars today, particularly those involving α- and iron-peak elements, astronomers can infer the relative timing of their formation. These abundance patterns thus serve as reliable indicators of stellar age, allowing us to understand the chronological chemical evolution of the Galaxy.

2.3 Machine Learning for Stellar Age Determination

Traditional methods of determining stellar ages, such as isochrone fitting, have been the standard in stellar astrophysics. As stated above, isochrone fitting compares observed stellar properties, such as temperature, luminosity, and surface gravity, to theoretical models of stellar evolution, allowing for age estimation. However, this technique is most reliable for certain evolutionary phases, particularly for Turn-Off stars, where stellar properties change significantly with age. In large spectroscopic surveys like GALAH, where we have access to high-resolution chemical abundance data, an alternative approach is to use machine learning to infer stellar ages based on chemical abundances. This is useful for multiple reasons.

First, isochrone ages are highly uncertain for Main Sequence and giant stars because the measurement errors on the stellar parameters are larger than the distance between isochrones in parameter space. Put simply, the isochrones stack on top of each other on the Main Sequence and Giant Branch but not on the Main Sequence Turn-Off (Figure 4). This makes direct isochrone fitting unreliable outside of the Turn-Off region. Machine learning models can be trained on a subset of stars with

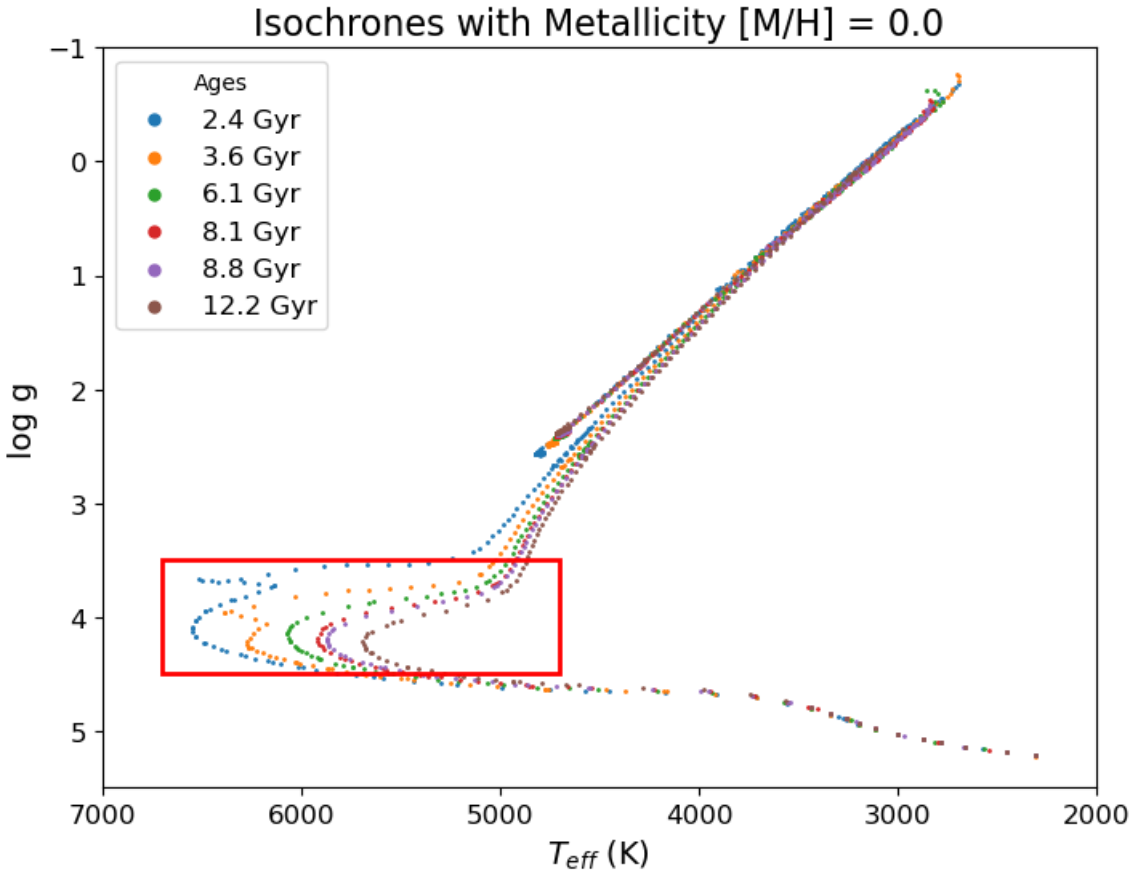


Figure 4: A set of theoretical isochrones having the same metallicity. Notice how the Main Sequence and Giant Branch are stacked on top of each other while the data points in the Main Sequence Turn-Off are discernible.

well-determined isochrone ages (Turn-Off stars) and generalize these age estimates to the rest of the population using chemical signatures.

Second, the chemical composition of a star carries fossil records of its formation epoch. Abundance ratios such as [Fe/H] and [α/Fe] evolve over time due to the different timescales of Type Ia and Type II supernovae. Machine learning can recognize complex, nonlinear patterns in multi-element abundance space, allowing for improved age inference beyond traditional methods.

Lastly, isochrone fitting requires detailed spectroscopic and photometric data, which can be computationally expensive to apply consistently across hundreds of thousands of stars. Machine learning, once trained, can rapidly predict ages for vast datasets, making it highly scalable for surveys like GALAH.

While isochrone fitting remains the best method for precise stellar age determination in Turn-Off stars, it becomes unreliable for other evolutionary phases. Machine learning provides a powerful alternative by leveraging chemical abundance patterns to estimate stellar ages across a much larger dataset, making it an essential tool for large-scale studies of stellar populations and galactic archaeology.

3 Data

3.1 GALAH Data Release 4

The dataset used in this project is derived from the GALAH (Galactic Archaeology with HERMES) Data Release 4 (DR4) Buder et al. (2024), which provides detailed spectroscopic information for nearly one million stars in the Milky Way. GALAH is a large-scale stellar survey conducted using the High Efficiency and Resolution Multi-Element Spectrograph (HERMES) spectrograph on the 3.9m Anglo-Australian Telescope, and it aims to explore the chemical and dynamical evolution of our galaxy by measuring high-resolution spectra of stars across a wide range of positions and

stellar types. The survey provides precise measurements of stellar parameters such as effective temperature (T_{eff}), surface gravity ($\log g$), and elemental abundances for up to 32 elements.

3.2 Data Selection

For this project, the full GALAH DR4 catalog was initially downloaded, containing 917,588 stars in total. However, not all of these stars have high-quality measurements for the parameters required in this analysis. While the overall goal of this project is predicting stellar ages from surface abundances, actually getting the ages for the training set necessitates accurate estimates of $\log g$ and $\log T_{eff}$. GALAH includes quality flags for these parameters (flag_sp), allowing users to filter for reliable data. By requiring that both $\log g$ and $\log T_{eff}$ be flagged as reliable, the dataset was reduced to 663,075 stars. Figure 5 shows a Kiel Diagram ($\log g$ on the y-axis instead of luminosity) of the sample after the flag_sp cut.

Because the elemental abundances will serve as the features for the machine learning model used in this project, it is important that all of the elemental abundances used are reliable. GALAH provides reliability flags for each elemental abundance in the dataset (flag_X_fe being the quality flag for [X/Fe]). However, there are several elements in the dataset that have few reliable measurements. Because an unreliable abundance measurement for a single elemental would remove the row that contains it from the dataset, it is not wise to include every elemental abundance. To ensure a balance between investigating as many elements as possible and having a sufficient amount of data, 14 elemental abundances (with respect to iron) were selected to serve as features for the model: [Mg/Fe], [Ca/Fe], [Ti I/Fe], [Si/Fe], [Mn/Fe], [Na/Fe], [K/Fe], [Y/Fe], [Ba/Fe], [Cr/Fe], [Zn/Fe], [Al/Fe], [Sc/Fe], and [O/Fe]. Additionally, the iron-to-hydrogen abundance ratio ([Fe/H]) was also required to be reliable (flag_fe_h). Requiring that the abundance measurements

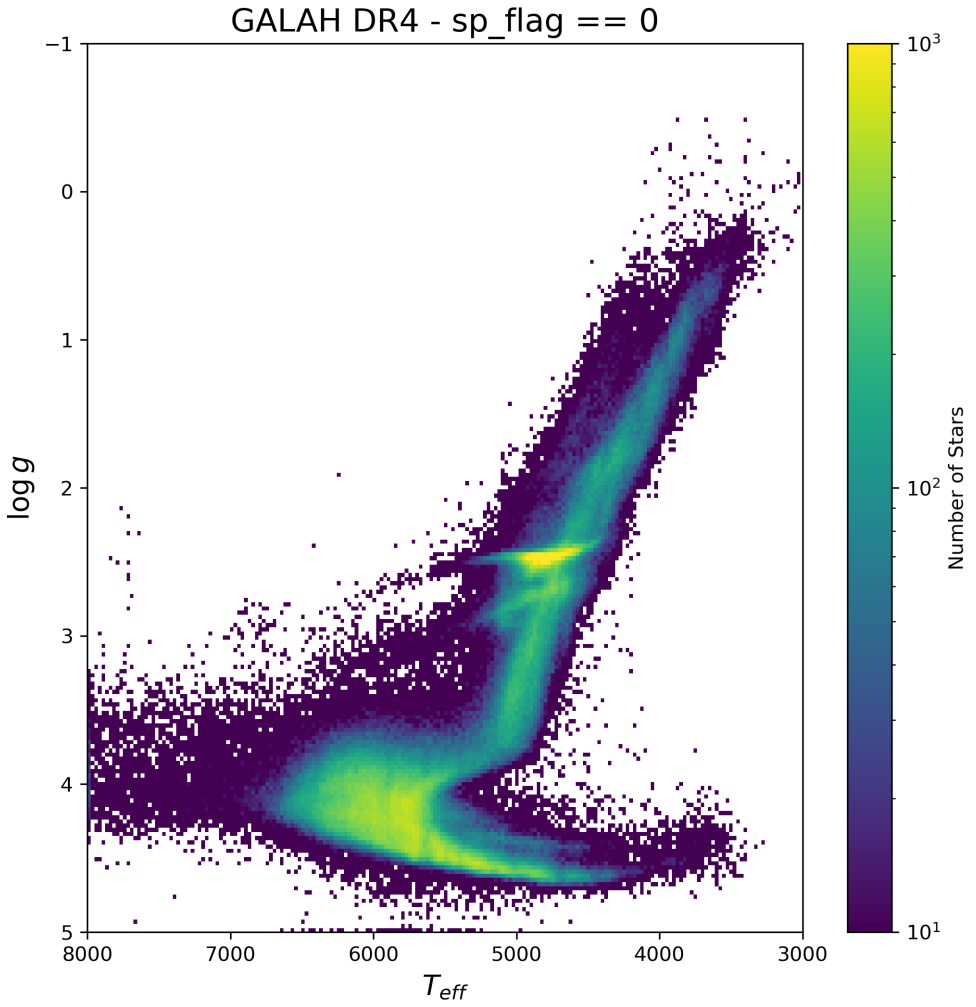


Figure 5: A Kiel Diagram of 663,075 stars having flag_sp = 0 from the GALAH DR4 dataset. Notice the Main Sequence, Main Sequence Turn-Off, and Giant Branch.

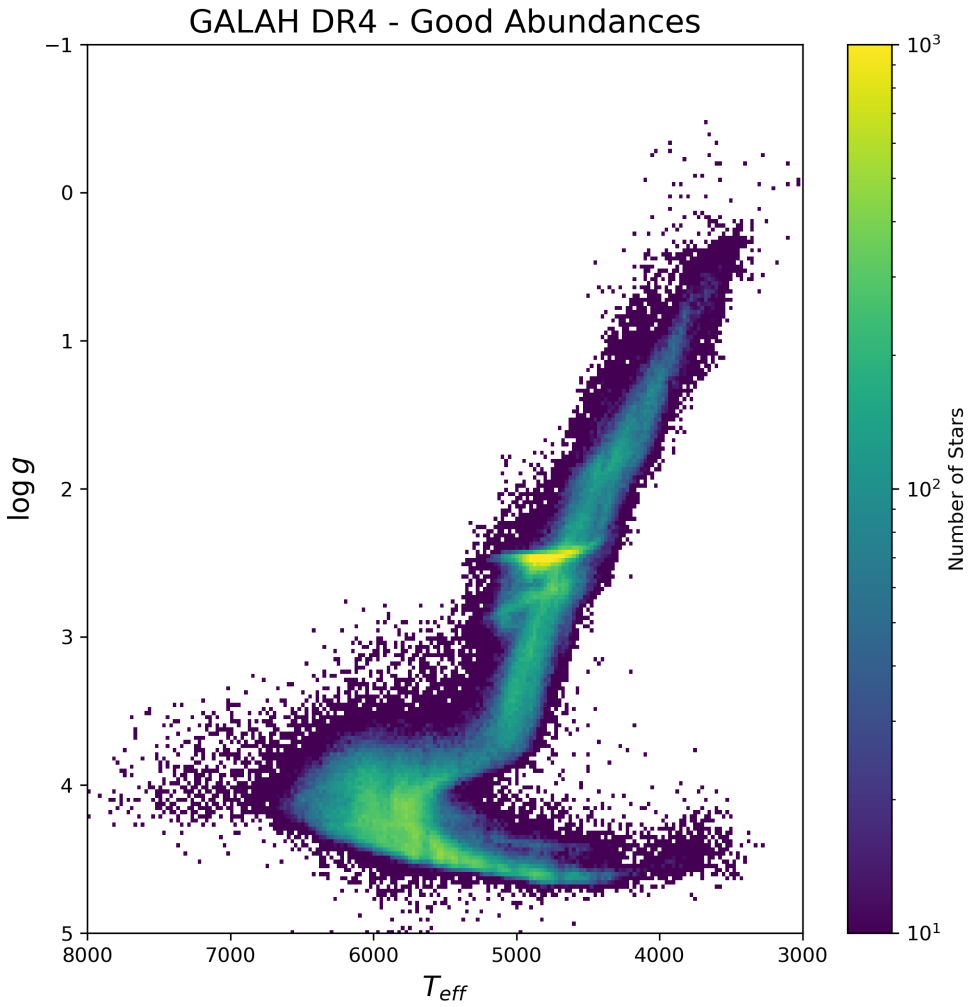


Figure 6: A Kiel Diagram of 421,934 stars having good measurements for the desired elemental abundances.

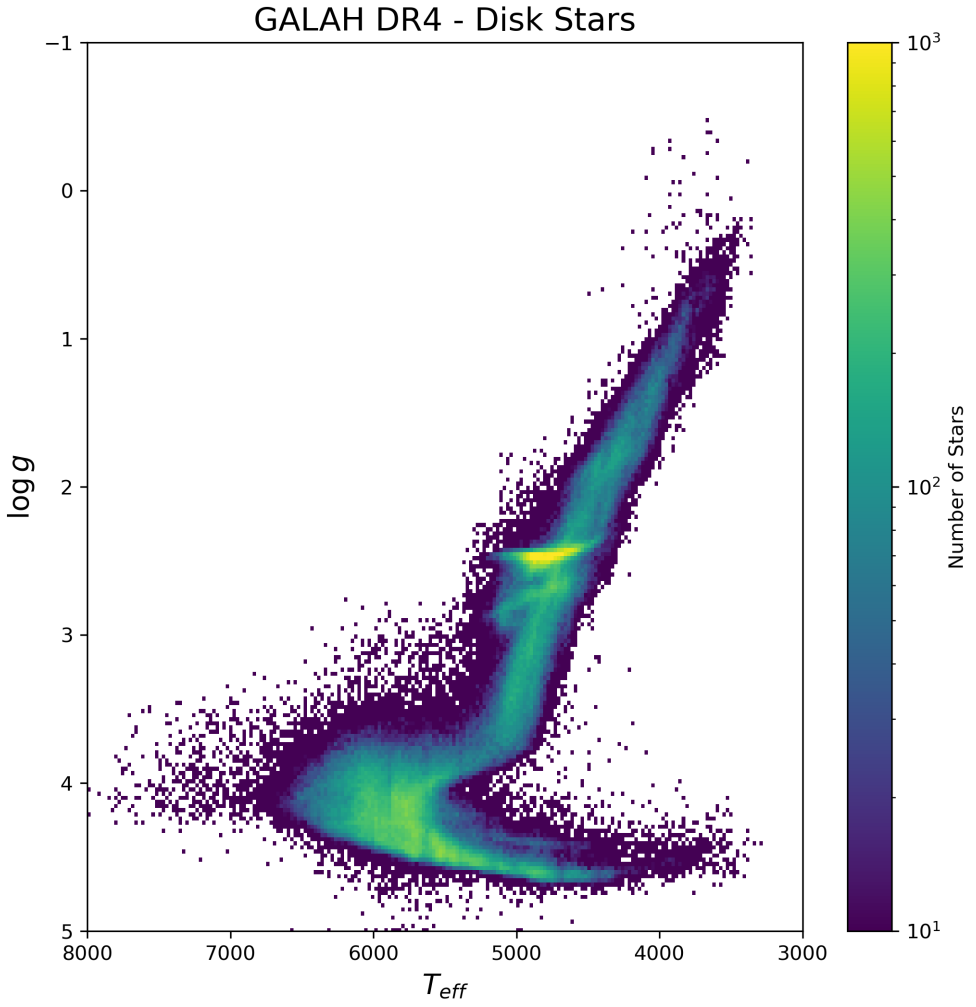


Figure 7: A Kiel Diagram of 412,405 quality-ensured stars having $-1.0 \leq [Fe/H] \leq 0.5$ (disk stars).

Parameter	Value
flag_sp	0
flag_fe_h	0
flag_X_fe [X/Fe]	0
χ^2	≤ 4
SNR	≥ 30
$[Fe/H]$	$-1.0 \leq [Fe/H] \leq 0.5$

Table 1: Quality cuts used to select data from GALAH DR4.

be reliable further restricted the dataset to 421,934 stars. Figure 6 displays this smaller sample in a Kiel Diagram. Evidently, abundance measurements are less reliable for the hottest stars.

While the quality flags indicate that nothing went catastrophically wrong during the data reduction pipeline, they do not directly reflect the quality of the spectral fit or the precision of the derived measurements. To ensure a good fit between the observed and synthetic spectra, a chi-square threshold of less than 4 was required ($\chi^2 < 4$). This helps ensure that the model accurately reproduces the observed spectrum, leading to more reliable estimates of stellar atmospheric parameters and chemical abundances. In addition, a minimum signal-to-noise ratio of 30 (SNR > 30) was required. A higher SNR improves the precision of measurements by reducing the impact of random noise, effectively shrinking the error bars on flux values and leading to more tightly constrained parameter estimates. These cuts reduced the dataset to 415,343 stars.

A final cut was made that ensured only disk stars be present in the dataset, $-1 \leq [Fe/H] \leq 0.5$. Stars outside of the disk, e.g. halo stars, have unique chemistry and kinematics as they have often formed elsewhere, such as other galaxies, and were gained through mergers with these satellite systems. Including objects that do not share formation pathways would lead to unnecessary spread. This metallicity cut reduces the dataset to 412,405 stars. Figure 7 shows a Kiel Diagram of the quality-ensured disk stars and Table 2 outlines the quality cuts made.

3.3 Training+Test Set

The training+test set for this study consists of stars selected to be near the Main Sequence Turn-Off, where stellar ages can be more reliably constrained using isochrone fitting. A surface gravity cut of $3.5 \leq \log g \leq 4.1$ was applied to isolate stars evolving off the Main Sequence but not yet on the Giant Branch. This region of the Hertzsprung-Russell diagram is known to be sensitive to age, allowing for more precise isochrone fitting compared to giants or dwarfs (this sensitivity is discussed in Section 2.3). This brings the number of stars in training set to 83,420.

To ensure the highest quality, an additional signal-to-noise cut was made, (SNR $>$ 70). The goal with this quality cut is to ensure that the errors on the measured parameters are as small as possible, allowing the isochrone fitting process (described in Section 4.1) to find as accurate an age as possible. This signal-to-noise requirement reduces the training+test set 25,550 stars. The age for each of these stars was found using the methodology described in Section 4.1, where [α/Fe] was also found as an intermediate step.

Because hot, young stars have very few emission lines it is difficult to get accurate chemical abundances for the youngest stars. Therefore, a cut was made to exclude stars younger than 1.8 Gyr from the training+test set, which also serves as a temperature cut. Additionally, some thick disk stars ([α/Fe]-enhanced) have been discovered to be uncharacteristically young. The physical origin of these young, α-rich stars is a matter of great debate (Chiappini et al., 2015; Martig et al., 2015). Because of this, stars with [α/Fe] > 0.15 are required to be greater than 8 Gyr old. Making these final cuts to the training+test set leaves 25,379 stars, approximately 6.2% of the sample. Table 3.3 outlines the cuts made to create the training+test set. For model training and evaluation purposes, the training+test set was divided using an 80/20 train-test split. This split puts the training set at 20,303 stars and the test set at 5,076 stars.

Parameter	Value
$\log g$	$3.5 \leq \log g \leq 4.1$
SNR	≥ 70
τ	≥ 1.8 Gyr

Table 2: Additional cuts used to select MSTO data used for model training.

4 Methodology

4.1 Bayesian Isochrone Fitting

The ages for stars in the training set described in Section 3.3 were calculated using a Bayesian framework to fit the observed data points to theoretical model isochrones. These model isochrones were processed using the PARSEC isochrone web application, where different metallicity and age values were supplied. The age of the isochrones range from 500 million years to 13.5 billion years, with a step size of 35 million years, yielding 375 distinct ages. The metallicity values range from -1.0 to 0.5 with a step size of 0.1, yielding 15 distinct metallicity values. Because each isochrone has a single age and a single metallicity value, each star is fit to a total 5,625 model isochrones.

Observations for $\log g$ and T_{eff} are available for each star. However, while abundances are available for multiple elements, GALAH does not directly supply a metallicity value. This metallicity value will incorporate the abundance of every element heavier than helium. Because this is not feasible, many studies use [Fe/H] as a proxy for metallicity. However, this project utilized a more accurate approach, Equation 1, described in Buder et al. (2021). The [α/Fe] value is a measure of the abundance of the α-elements. This was calculated as an intermediate step by taking the average of the Mg, Ti I, and Si abundances.

$$[M/H] = [Fe/H] + \log(0.694 \times 10^{[\alpha/Fe]} + 0.306) \quad (1)$$

The errors on both [α/Fe] and [M/H] were propagated using the formalism described in Taylor (1997). Because [α/Fe] is an average, its error is simply the individual abundance errors added in quadrature. However, for more complex, multi-variate measurements, the error is found using Equation 2. Thus, the error on [M/H] is found using Equation 3.

$$\delta q = \sqrt{\left(\frac{\partial q}{\partial x}\delta x\right)^2 + \ \ldots \ + \left(\frac{\partial q}{\partial z}\delta z\right)^2} \tag{2}$$

$$\sigma_{[M/H]} = \sqrt{\sigma^2_{[Fe/H]} + \left(\left(\frac{0.694 \times 10^{[\alpha/Fe]} \times \ln(10)}{0.694 \times 10^{[\alpha/Fe]} + 0.306}\right)\sigma_{[\alpha/Fe]}\right)^2} \tag{3}$$

To derive stellar ages from the observed parameters, a Bayesian isochrone fitting procedure was applied. The goal is to estimate the posterior probability distribution of a star's age given its observed parameters. Bayes' theorem provides the foundation:

$$P(\tau \mid \mathbf{D}) = \frac{P(\mathbf{D} \mid \tau)\ P(\tau)}{P(\mathbf{D})} \tag{4}$$

where $P(\tau \mid \mathbf{D})$ is the posterior probability of the age τ given the data $\mathbf{D}$ (observed parameters), $P(\mathbf{D} \mid \tau)$ is the likelihood of observing the data for a given age, $P(\tau)$ is the prior on age, and $P(\mathbf{D})$ is the marginal likelihood or normalization constant.

In this analysis, the prior $P(\tau)$ was taken to be uniform, implying no preferred age range before observing the data. The posterior is therefore proportional to the likelihood:

$$P(\tau \mid \mathbf{D}) \propto P(\mathbf{D} \mid \tau) \tag{5}$$

The likelihood $P(\mathbf{D} \mid \tau)$ was computed by comparing each star's observed pa-

rameters to each of the model isochrones. These parameters include T_{eff}, $\log g$, and [M/H], all of which are matched against corresponding values in the model isochrones for a given age and metallicity.

Assuming independent Gaussian errors on each parameter, the likelihood for a given isochrone point is defined as:

$$\mathcal{L} \propto \prod_j \exp\left(-\frac{(X_{j,obs} - X_{j,iso})^2}{2\sigma_{X_j}^2}\right) \tag{6}$$

where $X_j = \{T_{eff}, \log g, [M/H]\}$.

To ensure that the likelihood was associated with a particular age, an additional Gaussian weight centered on the specific age bin was applied. This weight acts like a smoothing kernel over the discrete age grid and can be interpreted as a kernel density prior over age with width equal to the age bin step:

$$\exp\left(-\frac{(\tau - \tau_i)^2}{2\Delta\tau^2}\right)$$

For each star, the likelihoods from all matching isochrone points at a given age bin were summed to obtain an unnormalized posterior probability for that age. After evaluating this across all age bins, the full posterior probability distribution was normalized to integrate to one:

$$P(\tau_i \mid \mathbf{D}) = \frac{\Sigma_k \mathcal{L}_{ik}}{\Sigma_l \Sigma_k \mathcal{L}_{lk}}$$

where $\mathcal{L}_{ik}$ is the likelihood contribution from the k-th isochrone point at age bin i.

The result is a discrete posterior probability distribution for each star, sampled over the 375 model age bins. These posterior distributions were used to assign a most probable age (e.g., the maximum a posteriori value).

4.2 eXtreme Gradient Boosting (XGBoost)

This project utilizes the machine learning algorithm XGBoost (eXtreme Gradient Boosting) (Chen and Guestrin, 2016) to predict stellar ages from chemical abundances. XGBoost is a scalable and regularized implementation of gradient boosted decision trees that has been widely adopted across a variety of domains due to its strong predictive performance and computational efficiency (Dhaliwal et al., 2018; Li et al., 2019; Ogunleye and Wang, 2020; Ma et al., 2021). Given the tabular structure of the input data and the tendency of chemical features to exhibit nonlinear relationships with stellar age, XGBoost is particularly well-suited to this regression task. Its built-in regularization mechanisms further reduce the risk of overfitting, making it a robust choice for models that must generalize beyond the training set.

XGBoost belongs to a family of ensemble learning algorithms based on gradient boosting. Ensemble methods improve performance by combining multiple weaker models (in this case, shallow decision trees) into a more powerful predictor. Unlike bagging methods such as random forests, where models are trained independently in parallel, boosting takes a sequential approach. Each model in the sequence is trained to correct the errors made by the ensemble up to that point, refining the prediction at every step.

This iterative refinement begins with a naive prediction, typically the mean of the target variable across all training examples. The model then calculates the residuals, which are the differences between the ground truth and predicted values. A new tree is trained to predict these residuals, effectively learning how to correct the mistakes of the current model. The output of this tree is scaled by a learning rate and added to the existing prediction. This process is repeated over many iterations, with each tree learning how to fix the errors made by the ensemble so far. Mathematically, this is equivalent to taking a step in the direction of the negative gradient of the loss function with respect to the current model output, hence the

term gradient boosting. Over time, this process reduces the overall loss and allows the model to capture complex, nonlinear dependencies between the features and the target.

Each individual decision tree in the ensemble learns by a process called splitting. At each node in the tree, the algorithm evaluates all possible thresholds for every feature (e.g., [Fe/H], [Mg/Fe]) and selects the one that most effectively reduces the residual error. For example, the tree might learn that stars with [Fe/H] below a certain threshold tend to be systematically older, and split the data accordingly. This recursive splitting process continues until a stopping criterion is met, such as a maximum tree depth or minimum child weight, resulting in a hierarchical structure that partitions the data into regions of relatively uniform stellar age. The final prediction for a star is determined by the average age of the training stars that land in the same terminal node (leaf).

By combining many such trees, each specialized in correcting the missteps of its predecessors, XGBoost constructs a highly flexible and accurate model. It further enhances this process through regularization techniques that penalize model complexity. These include constraints on tree depth and leaf weights, as well as penalties for unnecessary splits. Together, these features allow XGBoost to balance flexibility with generalizability, making it a powerful tool for modeling the chemical evolution of stars.

To determine the optimal configuration for the XGBoost model, an exhaustive grid search was performed over a defined hyperparameter space. Model performance was evaluated using five-fold cross-validation, wherein the training data was partitioned into five equally sized subsets (James et al., 2014). In each fold, four subsets were used to train the model while the remaining subset served as the validation set. This process was repeated five times, ensuring that each subset was used once for validation. The performance of each hyperparameter configuration

was assessed by averaging the root mean squared error (RMSE) across all five folds. This strategy provides a more reliable estimate of the model's generalization ability than a single train-test split and helps mitigate overfitting to any one subset of the data.

The RMSE was selected as the evaluation metric for hyperparameter tuning because it penalizes large errors more severely than small ones, making it particularly suitable for a scientific regression task where gross misestimations of stellar age are especially undesirable. As RMSE is reported in the same units as the target variable (Gyr), it also provides an interpretable measure of predictive accuracy.

The learning objective of the XGBoost model was set to `reg:squarederror`, instructing the model to minimize the squared difference between predicted and true stellar ages. This is the standard loss function for continuous regression tasks and aligns naturally with the use of RMSE as the evaluation metric.

The `learning_rate` hyperparameter, commonly referred to as the learning rate, determines the scale of the contribution made by each individual tree to the overall prediction. After a tree is trained to model the residual its outputs are scaled by the learning rate before being added to the cumulative prediction. A smaller learning rate reduces the impact of each tree, requiring more trees to fully minimize the loss but making the training process more stable. For instance, with eta = 0.01, each tree only nudges the model output slightly toward the correct value, promoting gradual learning. Conversely, a high learning rate (e.g., 0.1) may cause the model to overcorrect early and fit noise. The values tested ([0.01, 0.03, 0.05, 0.1]) allowed for exploration of both conservative and moderately aggressive update schemes.

The `n_estimators` hyperparameter, which defines the number of boosting rounds (i.e., trees), was varied to determine the appropriate model capacity. A higher number of estimators allows the model to capture more complexity, particularly when paired with a small learning rate. For instance, if the learning rate is set to 0.01,

the model may require 700–1000 trees to fully minimize the loss. However, with a higher learning rate such as 0.1, fewer trees may suffice, potentially around 200–300. Selecting an appropriate value ensured the ensemble was expressive enough to learn meaningful age-abundance relationships without unnecessarily increasing model size or training time.

The max`max_depth` hyperparameter sets the maximum number of hierarchical splits a tree can make from its root to its leaf nodes. This directly limits the number of interactions the model can learn between features in a single tree. Shallow trees (e.g., `max_depth` = 4) result in broader, more generalized splits that capture large-scale patterns, such as separating old and young stars using only a couple of chemical features. In contrast, deeper trees (e.g., `max_depth` = 10) allow the model to build intricate rules, such as segmenting stars into very narrow chemical abundance ranges, which increases risk of overfitting to noise or outliers. By controlling depth, the model was restricted from learning overly specific rules unless the complexity was warranted by the data.

The `min_child_weight` parameter controls the minimum sum of instance weights required in a child node for a split to be allowed. In regression tasks, instance weights default to 1, so this effectively sets the minimum number of samples in a leaf. During tree construction, a potential split is only accepted if both resulting Branches contain a sufficient number of training samples. This prevents the model from focusing too heavily on small, potentially noisy subsets. For example, with `min_child_weight` = 7, a split that would produce a leaf with only 3 stars would be discarded even if it slightly reduces training error. Larger values force the model to generalize more broadly and avoid overfitting to rare abundance combinations.

The `subsample` hyperparameter determines the fraction of the training data randomly sampled (without replacement) to train each tree. This means that each tree is trained on a different subset of the full data, introducing stochasticity into the

ensemble. Sub-sampling prevents the model from becoming overly reliant on particular training examples, thereby reducing variance and improving generalization. For example, with `subsample` = 0.7, each tree is trained on only 70% of the data, making the ensemble more robust to outliers and local anomalies. This is especially useful when the training set is large or when certain data points dominate the residuals due to measurement error or rare chemical abundance patterns.

The `colsample_bytree` hyperparameter specifies the fraction of available input features (in this case, chemical abundance ratios) that are randomly selected to be used when constructing each tree. This feature-level sub-sampling introduces diversity among trees and reduces the model's tendency to overfit to a small subset of strong predictors. For instance, with `colsample_bytree` = 0.7, only 70% of the features are considered when finding splits in any given tree. This forces different trees in the ensemble to explore different combinations of features, resulting in a more generalized and less biased model. This is particularly useful when input features are correlated, such as multiple alpha elements or iron-peak elements.

The `gamma` hyperparameter sets the minimum required reduction in the loss function (e.g., mean squared error) for a split to be made. During tree construction, a potential split is evaluated not only for how well it separates the data but also for whether the improvement in loss justifies the added complexity. If the improvement is less than `gamma`, the split is discarded. For example, if `gamma` = 5, then a split must reduce the overall loss by at least 5 units to be considered. This constraint discourages shallow gains from marginal splits, like slightly separating a few mid-age stars. This promotes simpler, more meaningful trees that only grow when there is a substantial benefit to doing so.

To discourage large leaf output values and improve the stability of predictions, the `lambda` hyperparameter was tuned. This hyperparameter controls L2 regularization on leaf weights. Larger values of `lambda` constrain the model's predictions

to remain moderate, especially in regions with limited data. For example, if a group of stars with low [Mg/Fe] is associated with high age variance (1 - 8 Gyr in the thin disk), a high `lambda` setting can prevent the model from assigning an extreme prediction to such a group, thus avoiding overfitting to rare cases.

The `alpha` parameter applies L1 regularization to the leaf weights, encouraging sparsity in the model. Tuning this hyperparameter allows the model to zero out uninformative or weakly contributing leaf outputs. For instance, if a specific combination of abundances appears in the training data but does not meaningfully improve age prediction, a nonzero `alpha` value may eliminate those contributions by driving their leaf outputs toward zero. This can be particularly useful in high-dimensional datasets, where sparsity aids both interpretability and generalization.

Together, the tuning of these hyperparameters allows the model to strike a balance between flexibility and generalization, enabling it to capture complex, nonlinear relationships between chemical abundances and stellar age while avoiding overfitting to noise in the training data. Parameters governing model complexity, such as `max_depth`, `n_estimators`, and `learning_rate`, carefully coordinate to ensure that the ensemble had sufficient capacity without becoming unstable. Regularization terms (`gamma`, `min_child_weight`, `lambda`, and `alpha`) act to constrain unnecessary complexity and promoted sparsity in the learned structure. Sub-sampling parameters (`subsample` and `colsample_bytree`) introduce randomness that further enhanced the robustness of the ensemble. By exploring these settings through a grid search (Table 3 displays the full grid) validated with five-fold cross-validation, a set of hyperparameters was identified that consistently minimized prediction error across all folds, resulting in a model that was both accurate and resilient to overfitting. This deliberate tuning process was critical for achieving the RMSE observed and for ensuring that the final model would generalize effectively to new, unseen stellar data.

Hyperparameter	Values
learning_rate	[0.01, 0.03, 0.05, 0.1]
n_estimators	[100, 300, 500, 700]
max_depth	[4, 6, 8, 10]
min_child_weight	[1, 3, 7]
subsample	[0.7, 0.85, 1.0]
colsample_bytree	[0.7, 0.85, 1.0]
gamma	[0, 0.1, 0.5, 1, 5]
lambda	[1, 5, 10]
alpha	[0, 0.1, 1, 5]

Table 3: Hyperparameter grid used to tune the XGBoost model, consisting of 103,680 possible hyperparameter combinations.

5 Results

5.1 Cross-Validation and Final Hyperparamters

As stated in Section 4, a comprehensive grid search was conducted using five-fold cross-validation on the training dataset. This approach systematically evaluated 103,680 combinations of hyperparameters by training the model on four folds and validating it on the fifth, rotating through all partitions. This led to 518,400 total XGBoost models fit. The performance metric used for optimization was the root mean squared error (RMSE), which quantifies the average deviation between the predicted and true stellar ages. The best-performing hyperparameter configuration, identified as the one that minimized the average RMSE across all folds, are presented in Table 4.

This configuration achieved an average RMSE of 1.42 Gyr across the five folds, reflecting the model's expected generalization error on unseen data. These hyperparameters strike a balance between model complexity and regularization: deeper trees and a large ensemble (`max_depth` = 10, `n_estimators` = 700) provide sufficient capacity to capture non-linear relationships in the data, while regularization terms (`alpha`, `lambda`, and `gamma`) act to prevent overfitting. Sub-sampling parameters

Hyperparameter	Values
learning_rate	0.03
n_estimators	700
max_depth	10
min_child_weight	7
subsample	0.7
colsample_bytree	0.85
gamma	1
lambda	10
alpha	5

Table 4: Optimal configuration of hyperparameters that minimized the average RMSE across all folds.

(subsample and colsample_bytree) introduce additional randomness, improving robustness. This cross-validated model configuration was subsequently used to train the final model on the full training set.

5.2 Model Performance on Training Set

To evaluate how well the model captured the relationship between stellar parameters and age, the final XGBoost model was trained on the full training sample of stars with isochrone-derived ages. On the training set, it achieved a root mean squared error (RMSE) of 0.72 Gyr, a mean absolute error (MAE) of 0.54 Gyr, and an R^2 score of 0.93, indicating strong internal consistency and a high capacity to explain variance within the training data.

This level of performance is notable given the inherent complexity of the age–abundance relation. That the model achieved such low error suggests that the abundance patterns in this dataset do encode meaningful age information, and that the model was able to uncover these relationships. The relatively high R^2 further reinforces this point by indicating that the majority of age variance across the sample can be explained using chemical abundances alone.

Still, strong performance on the training set does not guarantee the model will

generalize. A model that memorizes training labels can appear highly accurate but still fail to predict unseen data correctly. This is especially true for tree-based models that tend to overfit the training data. Although the residual structure (investigated further in Section 5.4) and low error provide some reassurance that this model has learned meaningful patterns rather than simply memorizing, generalization can only be confirmed through evaluation on a separate test set.

Figure 8 shows predicted versus true stellar ages, with most predictions falling tightly along the one-to-one line. Small deviations are more apparent at the youngest and oldest ends of the age range. At these extremes, the model tends to slightly overestimate the ages of young stars and underestimate those of old stars.

5.3 Model Performance on Test Set

To assess the model's ability to generalize to unseen data, the final XGBoost model was evaluated on a held-out test set comprising stars not used in either training or cross-validation. On this test set, the model achieved a root mean squared error (RMSE) of 1.38 Gyr, a mean absolute error (MAE) of 1.01 Gyr, and an R^2 score of 0.73.

These results confirm a meaningful degree of generalization, with performance degradation relative to the training set (RMSE = 0.72 Gyr) that is expected in the presence of a non-trivial predictive task and chemically diverse stellar populations. The test RMSE closely matches the average RMSE from the 5-fold cross-validation. This indicates that the cross-validation was performed properly and the model performs consistently on unseen data, even improving slightly when allowed to train on all 5 folds. The lower R^2 on the test set reflects both intrinsic uncertainties in the age–abundance relation and limitations imposed by the finite feature set.

Figure 9 shows the predicted versus isochrone-derived ages for the test set. The scatter increases relative to the training set, but the overall structure remains intact.

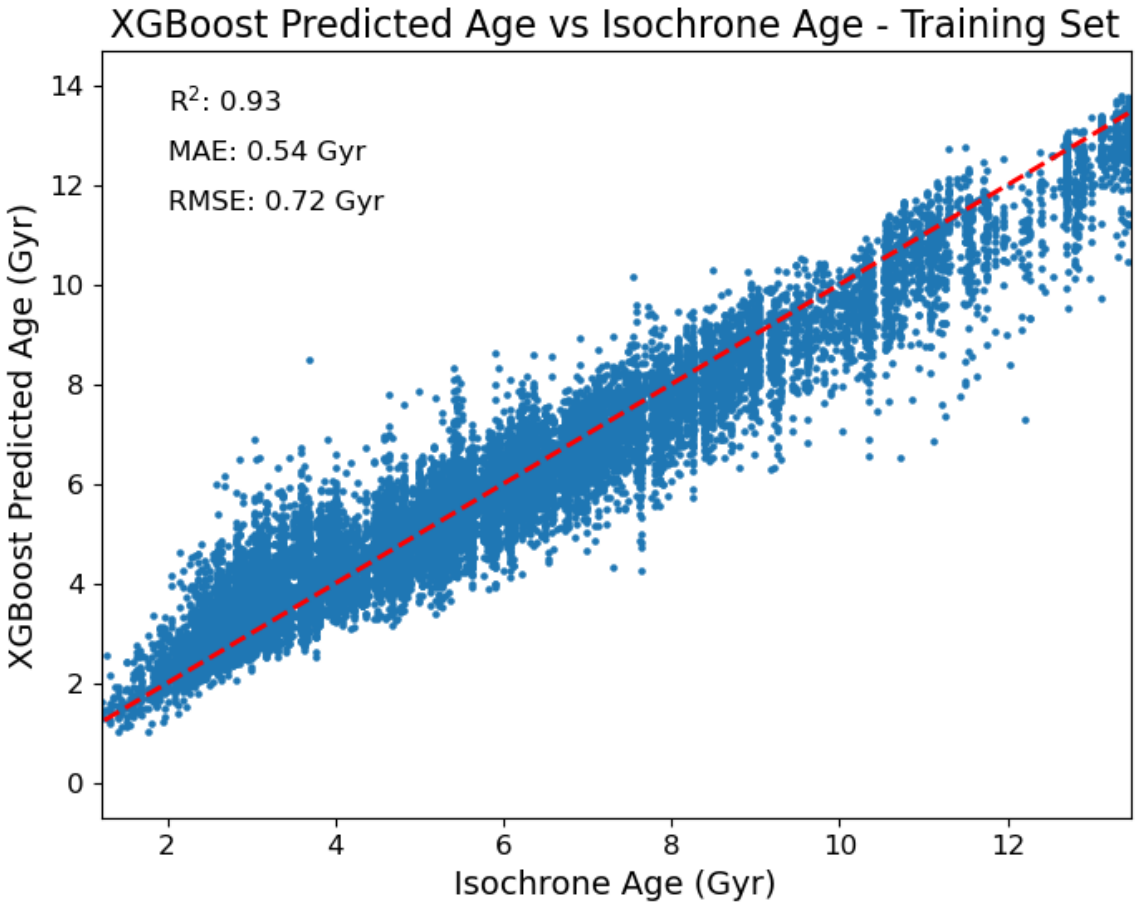


Figure 8: The ages predicted by the XGBoost model versus the ground-truth ages from isochrone fitting for the training set. The dashed red line denotes a 1:1 relationship.

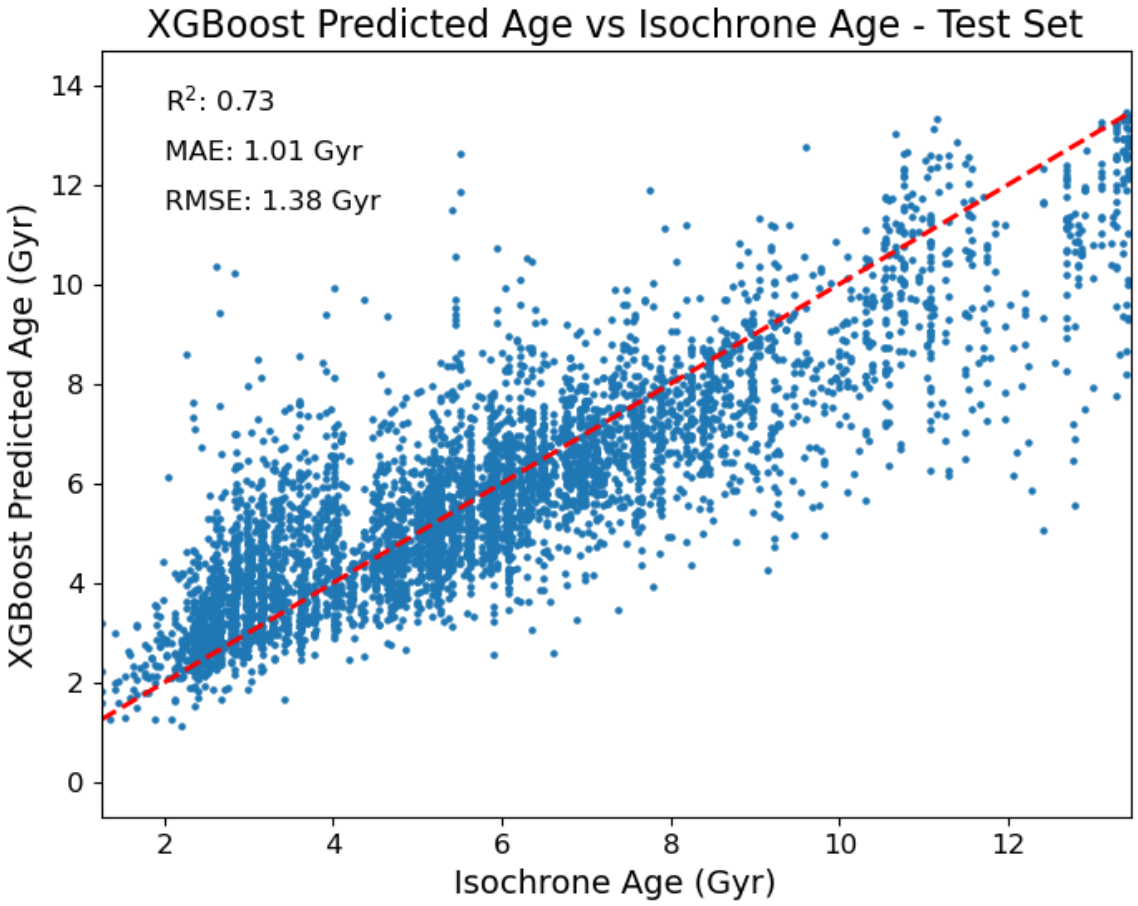


Figure 9: The ages predicted by the XGBoost model versus the ground-truth ages from isochrone fitting for the test set. The dashed red line denotes a 1:1 relationship.

Predictions generally follow the one-to-one line, with modest bias emerging at the low- and high-age extremes. Specifically, the model tends to overestimate the ages of young stars and significantly underestimate the ages of old stars. While a similar trend was also observed in the training set, it is much more pronounced on test set, particularly for old stars.

This pattern of bias at the edges of the age distribution is a common feature in regression models trained on imbalanced datasets. The sparsity of very young and very old stars in the training set leads to increased uncertainty and a tendency to regress predictions toward the mean. This highlights a key limitation: while the model captures broad trends in the data and performs well in the mid-age regime, it struggles to extrapolate confidently in underrepresented regions.

Overall, the model demonstrates solid predictive capability, recovering stellar ages with ~ 1 Gyr precision across a chemically diverse test sample. The drop in performance from training to test is within acceptable bounds for a problem of this complexity, particularly given the degeneracies and uncertainties inherent in stellar age estimation. These results validate the use of XGBoost for age prediction from chemical abundances and motivate further analysis into the structure of model errors and the contributions of individual features.

5.4 Residual Analysis

5.4.1 Training Set Residuals

Figure 10 shows the distribution of residuals, defined as the difference between the true and predicted values, which are approximately symmetric and centered near zero. This usually suggests that the model does not systematically over- or under-predict across the majority of the parameter space. While this is the case for the intermediate ages, the model does over-predict for young stars and under-predict

for old stars.

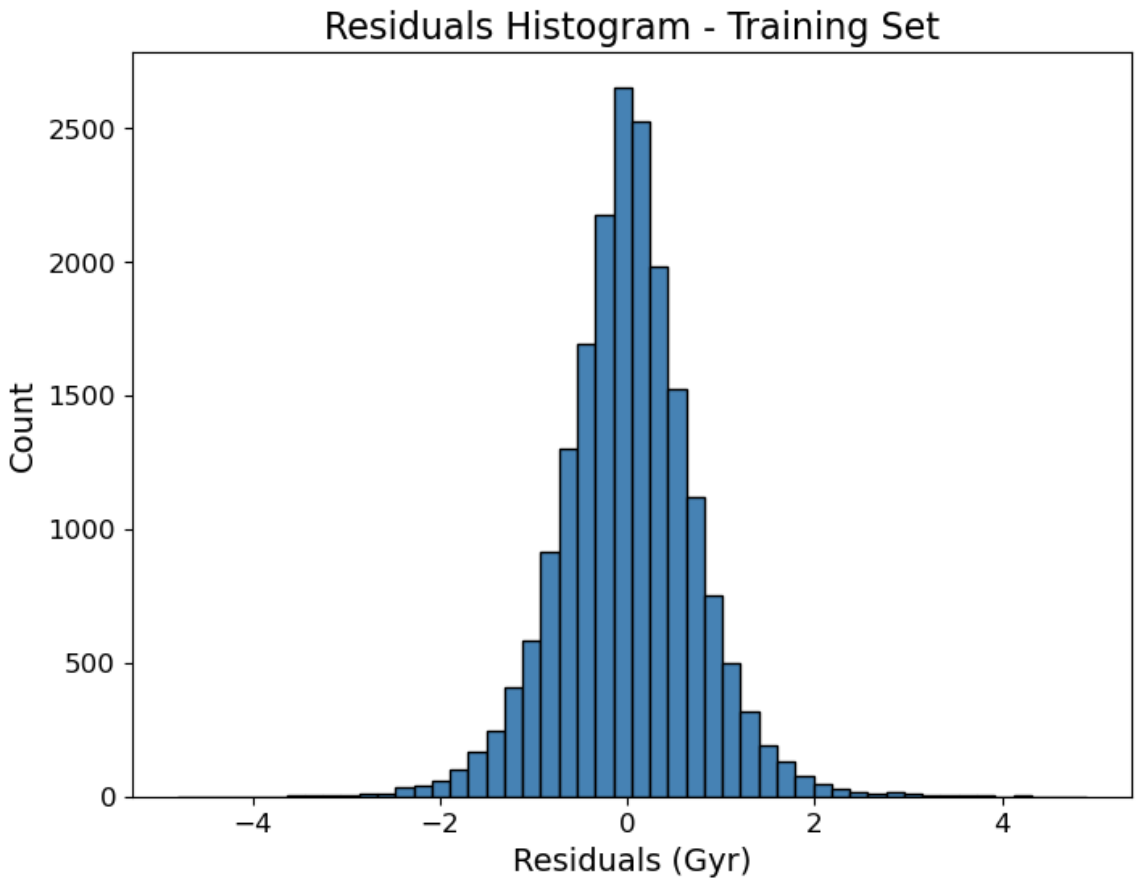


Figure 10: Histogram of the residuals (true minus predicted values) for the training set, displaying an approximately normal shape centered at zero.

Figure 11 shows the residuals as a function of the predicted age on the training set. A clear pattern emerges: the residual spread is smallest at the youngest and oldest predicted ages, and becomes widest in the middle of the predicted age range, particularly between 4 and 10 Gyr. This indicates that the model is most confident in its predictions at the extremes and has greater difficulty resolving stellar ages in the mid-range.

The narrow spread of residuals at low and high predicted ages suggests that very young and very old stars are chemically distinct enough for the model to identify them reliably. In these regions, stars tend to have abundance patterns that stand

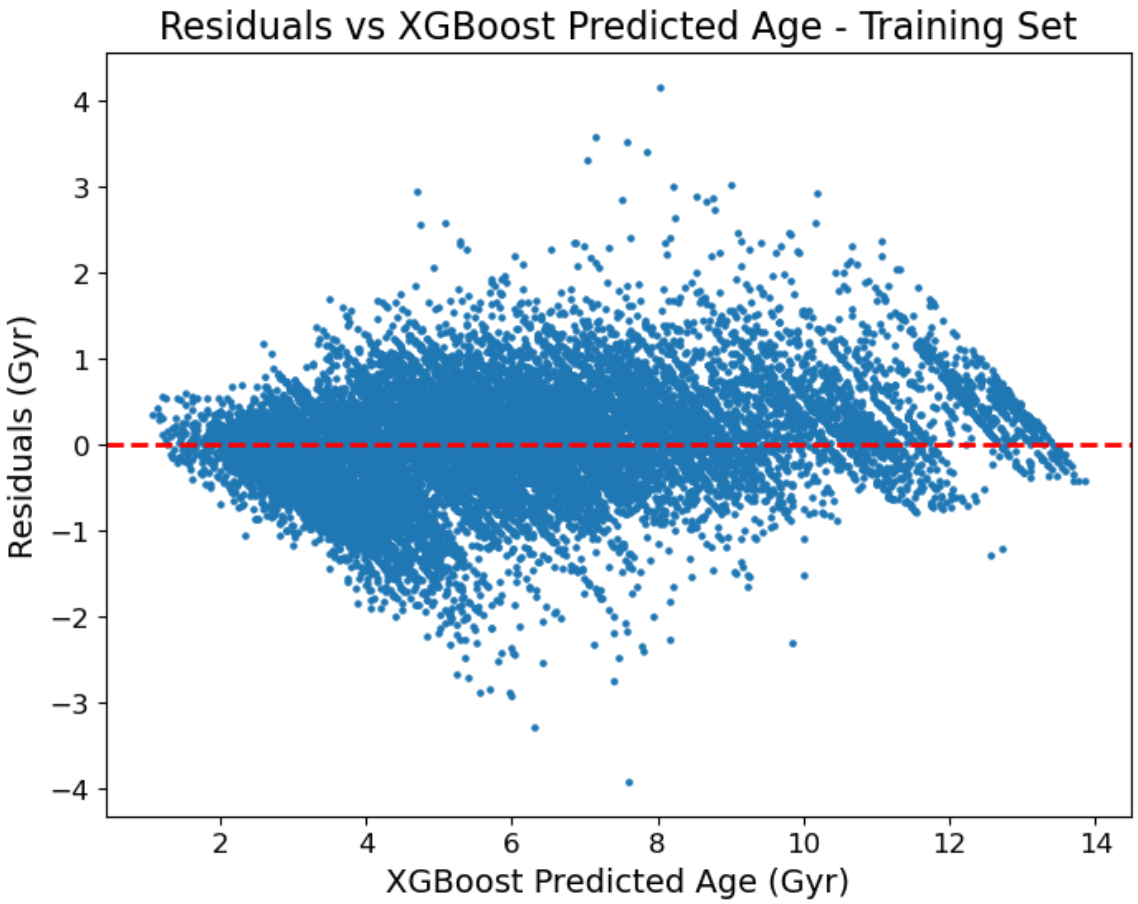


Figure 11: Residuals versus the predicted age for the training set. Evidently, the model is much more confident at the extremes and less confident at intermediate ages.

out more clearly; young stars typically have high metallicity and low α-enhancement, while old stars often exhibit low metallicity and elevated α-element ratios. These distinctive signatures likely help the model anchor its age predictions with greater certainty.

In contrast, the residuals become noticeably more dispersed for stars with predicted ages near the middle of the range. This increased variance reflects a higher degree of ambiguity in the model's estimates, which may be due to greater chemical overlap between stellar populations in this regime. Mid-aged stars in the Galactic disk are known to be chemically degenerate, with different formation histories leading to similar abundance patterns. As a result, the model may struggle to dis-

entangle age differences using abundances alone, leading to noisier predictions.

The residuals show a clear directional trend across the predicted age range. At low predicted ages, residuals skew negative, indicating that the model overestimates the ages of young stars. At high predicted ages, residuals skew positive, indicating that the model underestimates the ages of old stars. This structure suggests a local bias that pulls predictions toward the center of the age distribution, a sign of regression to the mean. While the overall residual distribution is centered around zero, this pattern highlights systematic error at the edges of the prediction space. This error is likely caused by data sparsity, increased degeneracy in the age-abundance relation, or both.

The distribution of isochrone-derived ages in the training set, shown in Figure 12, is highly non-uniform and skewed toward younger stars. The majority of stars fall between 2 and 6 Gyr, with a sharp decline in representation at older ages. Stars older than 9 Gyr are relatively rare in the sample. This imbalance introduces a natural bias into the learning process: the model receives far more training examples in the intermediate-age range and fewer examples at the extremes, which affects both predictive accuracy and uncertainty.

The regime with the highest residual spread corresponds to the most densely populated portion of the training distribution. This may be a consequence of degeneracy in the abundance–age relation within this region, where multiple stellar populations converge chemically despite different formation histories. In contrast, the improved performance at the low and high ends of the age range may stem not from high representation, but from distinctive abundance signatures that allow the model to anchor predictions more confidently, despite smaller sample sizes.

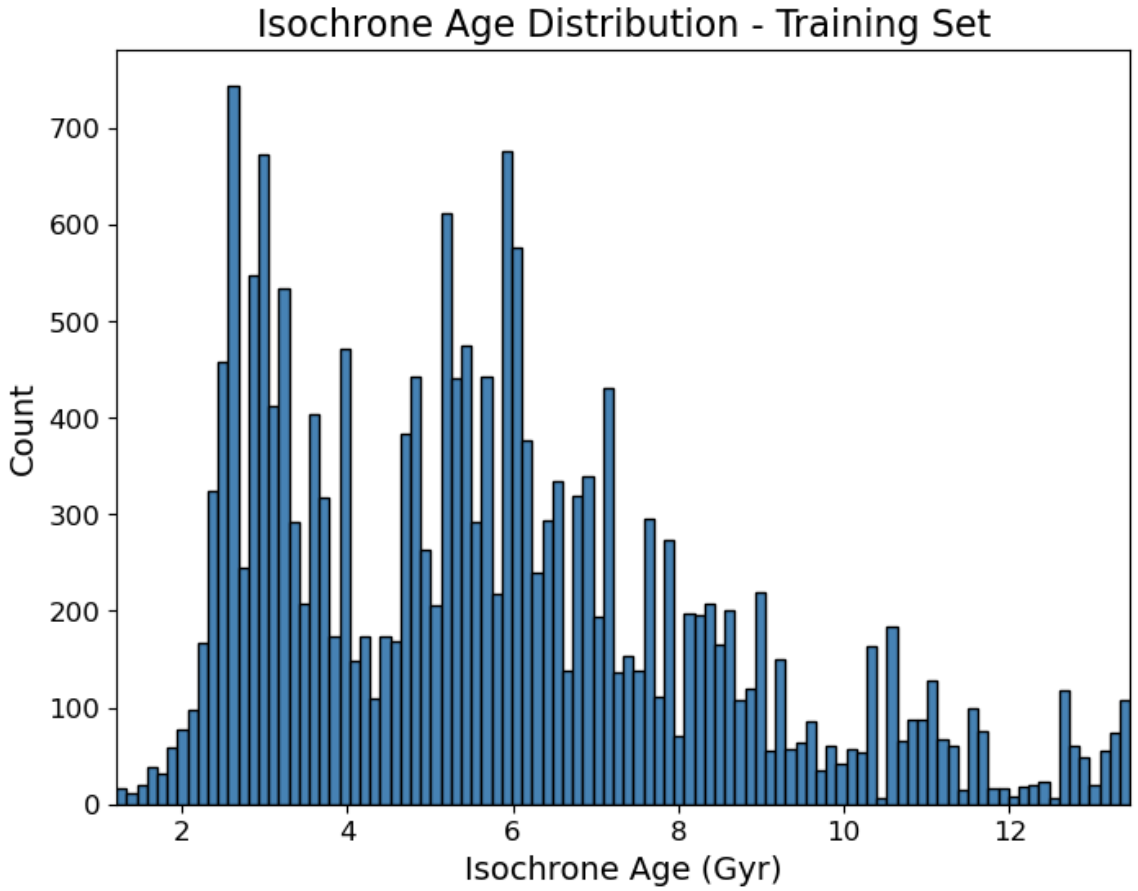


Figure 12: Histogram of the isochrone-derived stellar ages in the training set, showing the skew toward younger ages and under-representation of old stars.

5.4.2 Test Set Residuals

Figure 13 shows a histogram of the test residuals. The residuals are centered approximately at zero and display a normal shape, with a larger standard deviation than that of the training set. As discussed previously, this shape usually indicates no global bias. However, the residual structure reveals clear local biases at both ends of the age range, motivating a more detailed residual analysis as a function of the model input features.

Figure 14 displays the residuals as a function of the predicted age for the test set. Unlike the training set, where residuals were most tightly clustered at the extremes, the test set exhibits greater residual spread at both the low and high ends of the predicted age range. This increased uncertainty likely reflects the model's reduced confidence in extrapolating beyond the densest regions of the training distribution, where very young and very old stars were underrepresented.

As in the case of the training set, the test set residuals show a directional trend with negative values at low predicted ages and positive values at high predicted ages. This regression to the mean behavior indicates that the model systematically over-predicts the ages of young stars and under-predicts those of old stars. However, in contrast to the training set, the residual spread increases at the extremes, with the most pronounced variance observed for the youngest and oldest stars. This inversion of the pattern seen in training suggests that the model's apparent confidence at the edges during training was due in part to familiarity with specific examples, rather than a true ability to generalize in those regions.

The broader residual spread in the test set highlights the model's reduced certainty when encountering unseen data with extreme age values. This is consistent with the non-uniform age distribution of the training set, where stars at the low and high ends were relatively rare. In such sparsely sampled regions, the model lacks sufficient exposure to confidently learn and reproduce the mapping from abun-

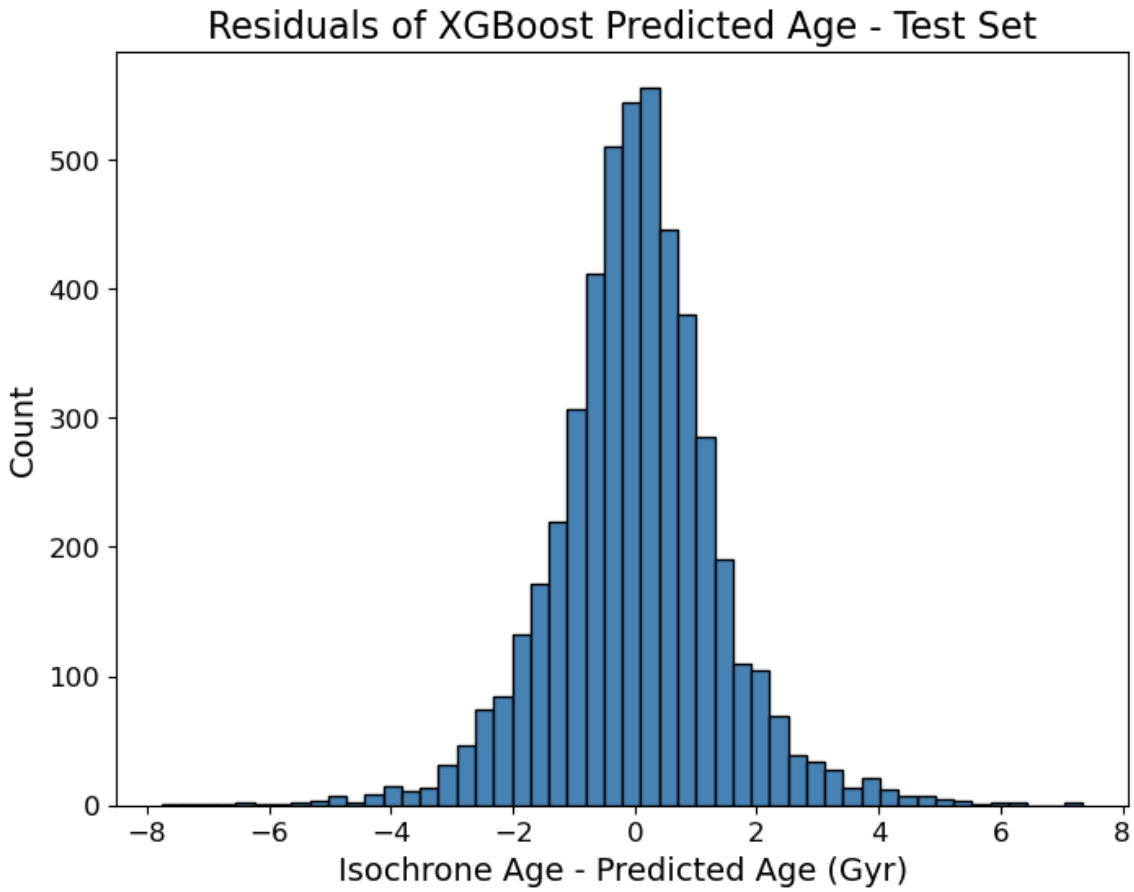


Figure 13: Histogram of the residuals (true minus predicted values) for the test set, displaying an approximately normal shape centered at zero.

dance space to stellar age. As a result, predictions at the boundaries become more variable, even as the average trend continues to follow a regression-to-the-mean trajectory.

Taken together, these results illustrate a key limitation of the model: while it performs well in the core of the age distribution, its predictive power diminishes at the tails. The systematic residual structure observed here provides strong motivation to examine how these residuals behave as a function of the model's input features.

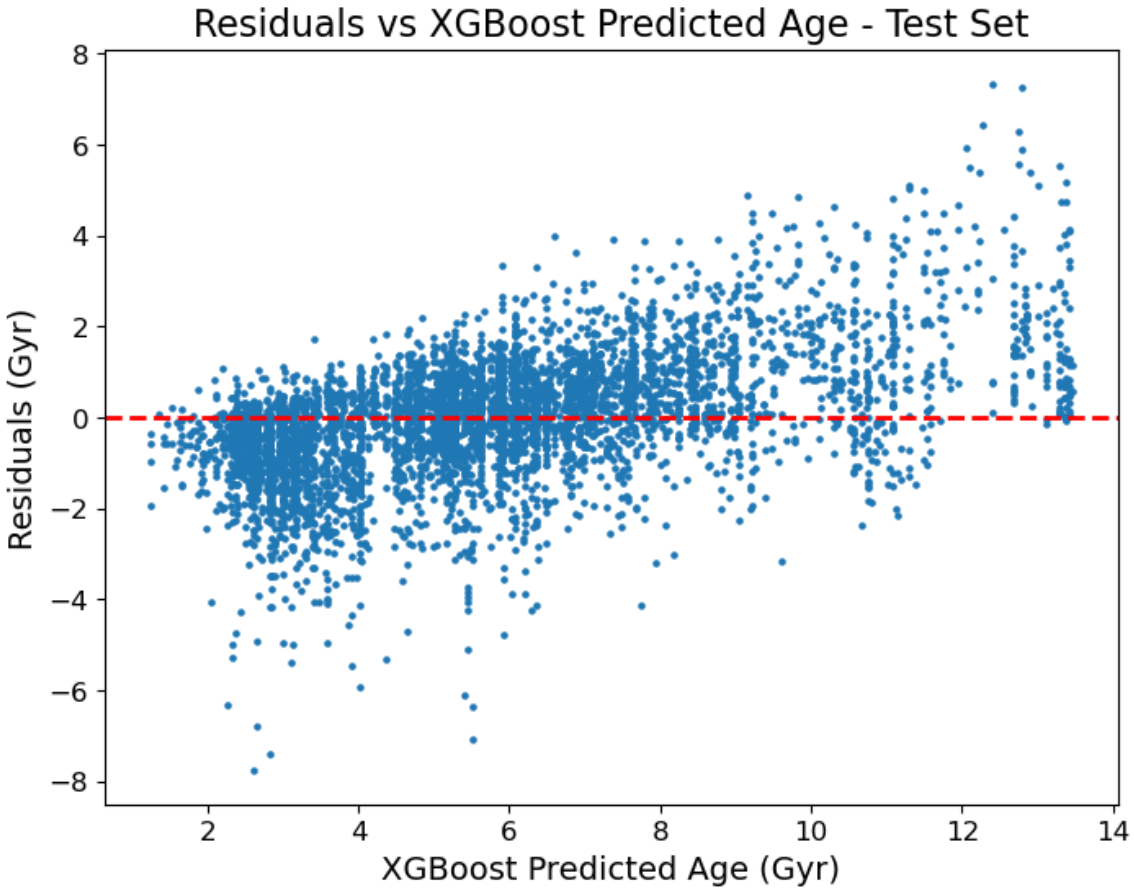


Figure 14: Residuals versus predicted age for the test set, showing a clear increase in variance at the extremes and lower variance at intermediate ages.

5.4.3 Test Set Residuals vs Chemical Abundances

To examine whether prediction errors are systematically related to specific elemental abundance ratios, residuals were plotted as a function of each individual feature used in the model. The resulting trends are displayed in Figure 15, where each panel shows residuals as a function of a single abundance ratio, with a dashed red line indicating zero residual. There are only two abundances that show strong residual structure: [Cr/Fe] and [Y/Fe].

Chromium is an iron-peak element and, as seen in Figure 3, follows a similar production history to iron—being predominantly synthesized in Type Ia supernovae (SNe Ia) later in the Galaxy's chemical evolution. As a result, the [Cr/Fe] ratio tends to remain close to solar for much of the stellar population, reflecting their common origin. The model residuals show increased scatter around [Cr/Fe] $\approx$ 0 but reduced variance at super-solar [Cr/Fe] values. This suggests that for stars with near-solar chromium abundances, chromium offers little age information, making it harder for the model to disentangle subtle trends. In contrast, stars with elevated [Cr/Fe] may represent a more chemically distinct subset, perhaps associated with localized enrichment, which the model is able to fit more accurately.

Yttrium, on the other hand, is a neutron-capture element primarily produced by the slow neutron capture process (s-process) in asymptotic giant Branch (AGB) stars. Its abundance increases more gradually over time, and it has been proposed as part of an empirical chemical clock when used in ratios such as [Y/Mg] or [Y/Al] (Nissen, 2015). The residuals for [Y/Fe] exhibit a similar pattern to chromium, with greater scatter near sub-solar [Y/Fe] values and reduced variance at super-solar [Y/Fe] values. This trend is consistent with the fact that stars exhibiting elevated [Y/Fe] tend to be younger, occupying regions of chemical space where age estimates are better constrained. In contrast, stars with lower [Y/Fe] are generally older and span a broader range of possible ages, contributing to greater prediction error. This

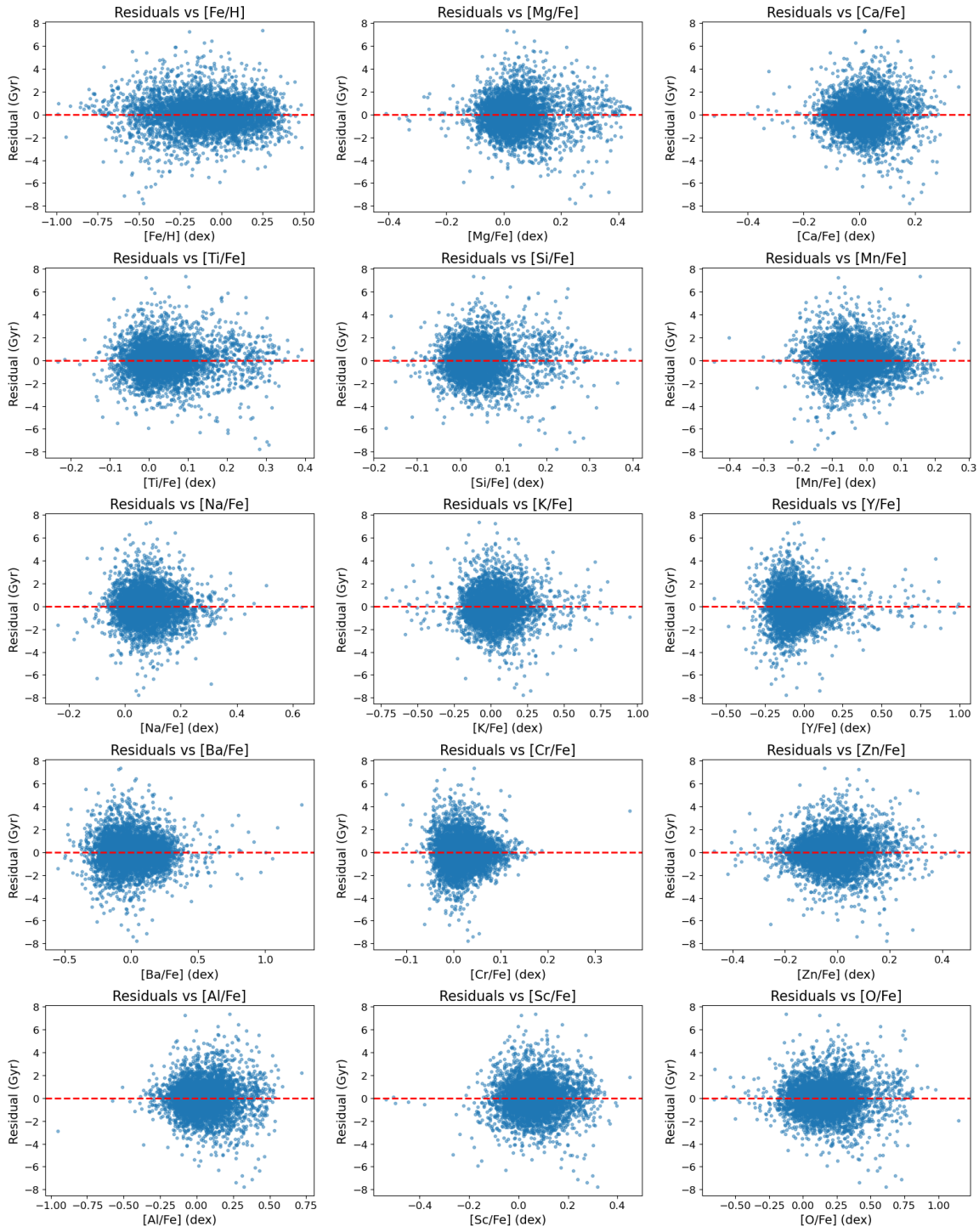


Figure 15: Panel displaying the residuals versus the features for all 15 features utilized in this study.

residual structure likely reflects a combination of true astrophysical variance and limitations in age inference for chemically older populations.

5.5 Feature Importance

To interpret the decision-making process of the XGBoost model and to evaluate which elemental abundance ratios most effectively contribute to the prediction of stellar ages, feature importance was quantified using two metrics provided by the model: weight and gain.

The weight metric represents the total number of times a particular feature is used to split data across all decision trees within the ensemble. It serves as a proxy for how often a feature is involved in partitioning the input space. As depicted in Figure 17, the most frequently used features included [Fe/H], [Cr/Fe], [Al/Fe], and [Ba/Fe]. These features appeared in a high number of decision nodes, suggesting that they contribute to the overall model structure by frequently aiding in differentiating between age regimes. However, weight alone does not reflect the effectiveness of these splits in reducing prediction error.

To capture the actual predictive utility of each feature, the gain metric was also examined. Gain is defined as the average improvement in the model's objective function (i.e., reduction in loss) brought about by splits on a given feature. As such, it measures how much a feature improves model accuracy when it is used. Figure 16 illustrates that [Mg/Fe] overwhelmingly dominated in terms of gain, contributing significantly more than any other feature. [Ca/Fe] and [Y/Fe] were the next most important features by this metric, whereas many of the high-weight features such as [Fe/H] and [Cr/Fe] exhibited relatively low gain.

This divergence between weight and gain underscores a crucial insight: a feature that is frequently used does not necessarily contribute substantial predictive power. For instance, [Fe/H] may be involved in many splits, but the correspond-

ing improvements in the model's predictions are modest on average. In contrast, [Mg/Fe], although selected less often, is highly valuable when it is used, leading to major reductions in the model's loss function. This finding highlights the importance of considering both the frequency and quality of splits when interpreting tree-based model behavior.

The scientific relevance of these results is notable. Both [Mg/Fe] and [Ca/Fe] are α-elements primarily produced in Type II (core-collapse) supernovae and are widely regarded as reliable proxies for stellar population age. The abundance of these elements relative to iron traces the chemical enrichment timescale of the interstellar medium from which stars form. Higher [α/Fe] values typically indicate older stellar populations that formed rapidly before substantial iron enrichment from Type Ia supernovae occurred. Therefore, the prominence of [Mg/Fe] and [Ca/Fe] in the gain-based ranking aligns well with established astrophysical interpretations.

The elevated gain of [Y/Fe] is also consistent with recent findings that s-process elements can be effective age indicators, especially in combination with α-element abundances. Yttrium is synthesized predominantly in asymptotic giant Branch (AGB) stars, and its abundance relative to α-elements can reflect the relative contribution of slower, delayed enrichment channels.

Meanwhile, the relatively low gain of [Fe/H], despite its frequent use, suggests that while metallicity is helpful in separating subgroups of stars, it lacks the specificity required for precise age determination in isolation. This observation reflects the well-known degeneracy in the age–metallicity relation, particularly in the thin disk, where stars of similar metallicity can span a wide range of ages.

Taken together, the feature importance analysis reveals that while numerous chemical abundance ratios contribute to the construction of the model, only a subset, including [Mg/Fe], [Ca/Fe], and [Y/Fe], provide high predictive value. These

features are not only useful statistically, but they are also physically motivated, reinforcing confidence in the model's internal logic and interpretability. This analysis also informs future feature selection efforts, offering a principled way to reduce dimensionality and computational cost while retaining performance. The alignment of model-driven importance with known nucleosynthetic pathways lends support to the broader use of machine learning in galactic archaeology and stellar population studies.

Additionally, to explore potential relationships among the elemental abundance ratios used as features, a Pearson correlation matrix was computed and is shown in Figure 18. This plot reveals notable correlations among α-elements, such as [Mg/Fe], [Ca/Fe], and [Ti/Fe], which are consistent with their shared nucleosynthetic origins in Type II supernovae. Conversely, certain s-process elements (e.g., [Y/Fe] and [Ba/Fe]) show weaker correlations with α-elements but exhibit moderate mutual correlation, reflecting their common production in AGB stars. The correlation structure highlights where redundancy or co-variation might exist, providing important context for interpreting feature importance metrics. In particular, highly correlated features can share predictive power, which may cause importance to be distributed across them in the model even if only one is truly informative in a physical sense. Thus, the correlation analysis offers an additional layer of interpretability, complementing the gain and weight assessments.

5.6 Age-Abundance Relations

To investigate the connections between stellar chemical composition and inferred ages, the relationships between predicted ages and individual elemental abundance ratios were examined. This analysis was performed to assess whether known empirical trends, such as the anti-correlation between [Fe/H] and age or the positive correlation between [α/Fe] and age, were preserved in the machine learning

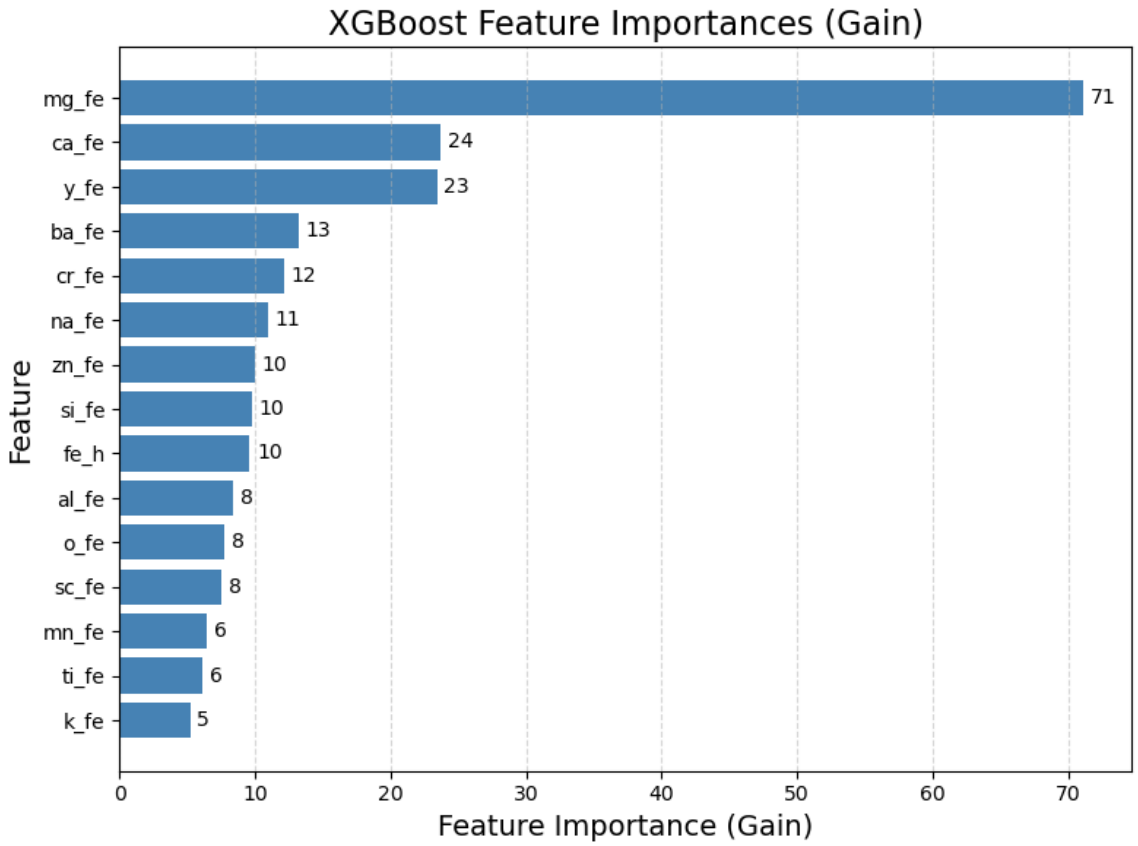


Figure 16: Feature importance based on gain for the XGBoost model. [Mg/Fe] exhibits the highest gain, indicating it contributes the most to reducing the model's loss function during training.

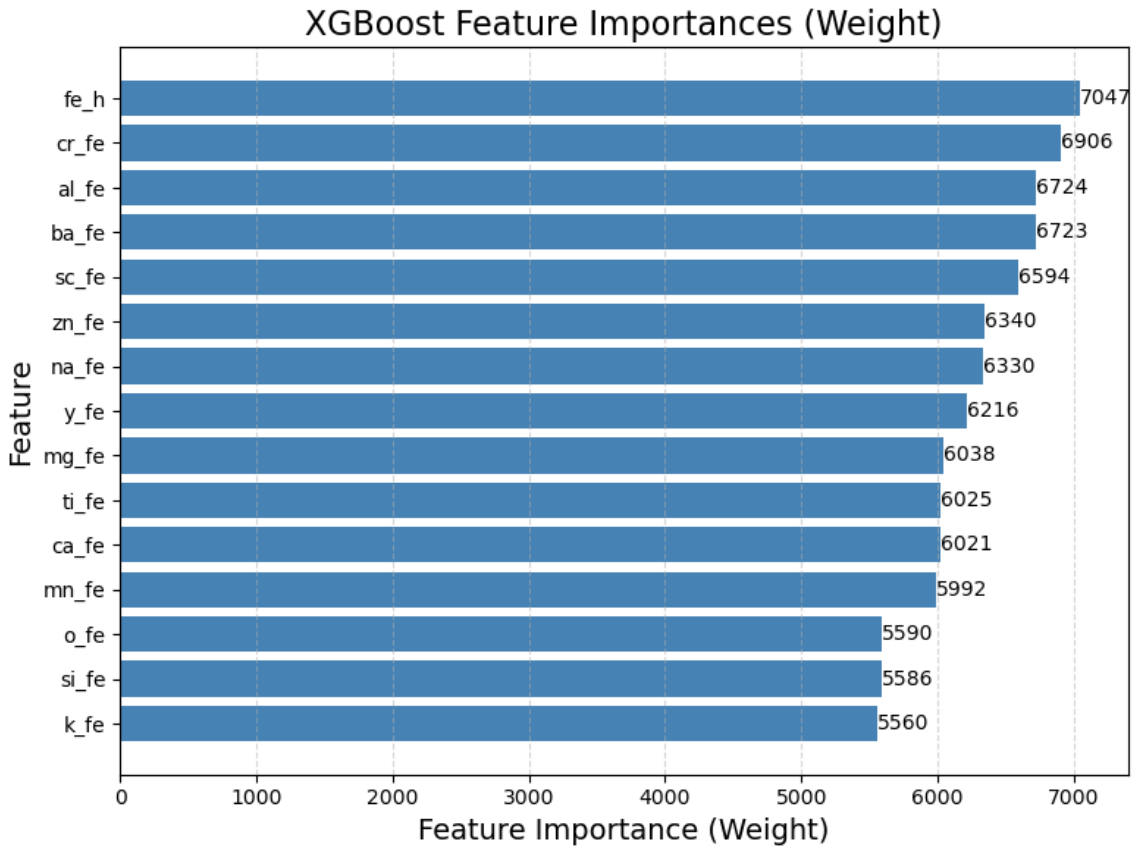


Figure 17: Feature importance based on weight for the XGBoost model. [Fe/H] has the highest weight, suggesting it played the largest role in the model's decision making.

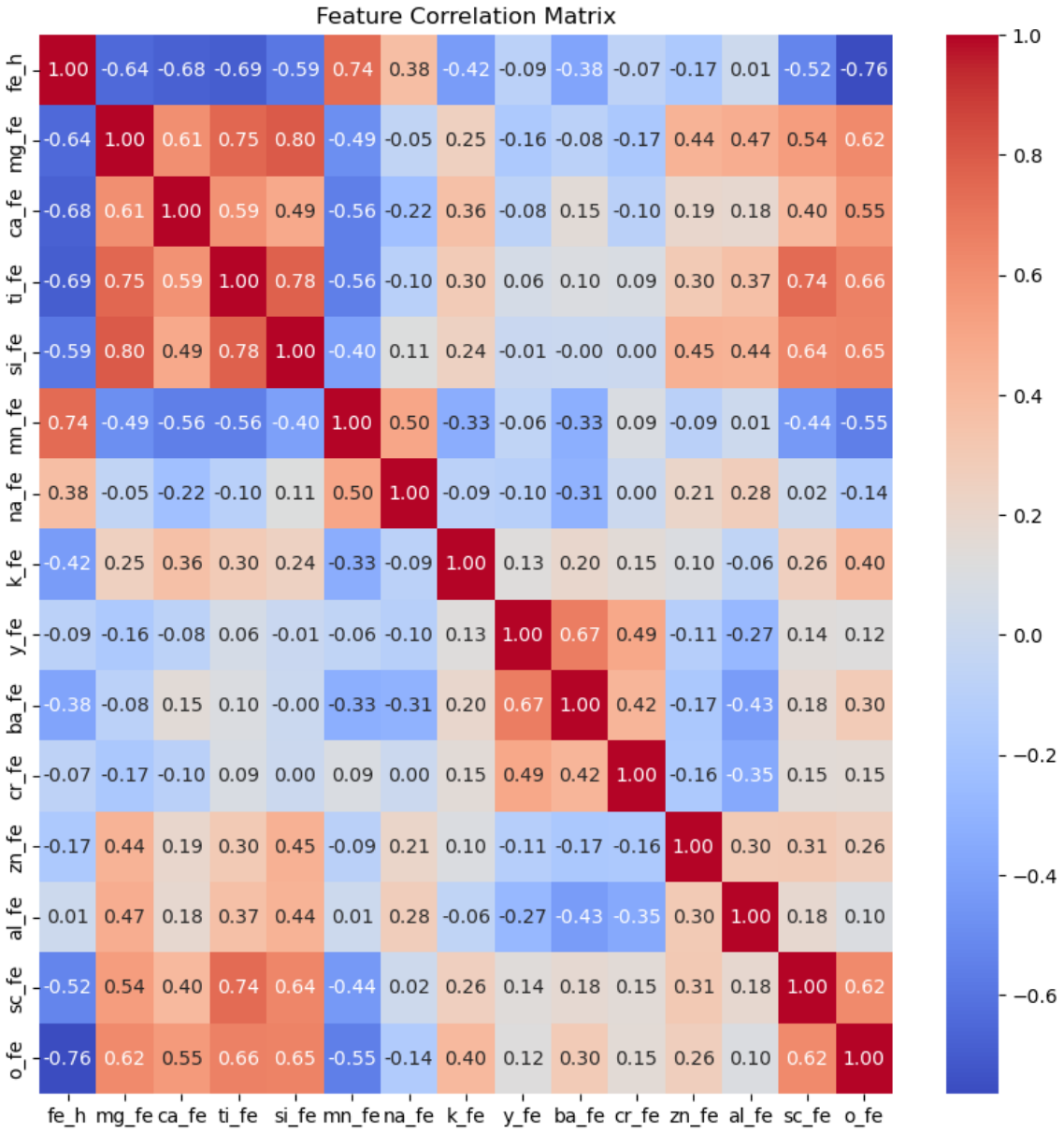


Figure 18: Pearson correlation matrix for the elemental abundance ratios used as model features. Strong correlations among α-elements and between certain s-process elements are evident, consistent with their astrophysical origins.

model's predictions.

For the test set, which consisted of turnoff stars with Bayesian isochrone-derived ages, the model reproduced well-known empirical age-abundance trends (Figure 19). A clear anti-correlation was observed between [Fe/H] and age, wherein older stars were generally more metal-poor. Additionally, strong positive correlations were recovered for α-elements such as [Mg/Fe], [Si/Fe], [Ti/Fe] and [Ca/Fe], consistent with early enrichment from core-collapse supernovae. These trends served as validation that the predicted ages retained physical meaning and were reflective of known evolutionary sequences.

When gathering data that a machine learning model will be trained on, an assumption is made that the mechanism by which the training set is created will continue to create instances by the same means. To investigate the extrapolative power of the model and whether the age-abundance mechanism persists throughout the H-R diagram, ages were also predicted for the 387,826 stars outside the training set. Specifically, this includes the disk stars that that were not on the MSTO (328,985 stars) and the stars that were on the MSTO but had SNR < 70 (58,841 stars). Despite the lack of direct training on these evolutionary stages, the age-abundance trends persisted in the extended sample (Figure 20).

The α-elements relations with age remained largely monotonic and consistent in slope with those observed in the test set, while [Fe/H] continued to exhibit a broad inverse trend, albeit with increased scatter. This suggests that the abundance–age relations learned by the model are robust across a broader region of parameter space than that originally used for training.

These results indicate that elemental abundance patterns contain sufficient information to predict stellar ages across a wide range of evolutionary stages. Moreover, the ability of the model to generalize beyond the training set underscores the potential of data-driven approaches to extend age-dating methods to populations

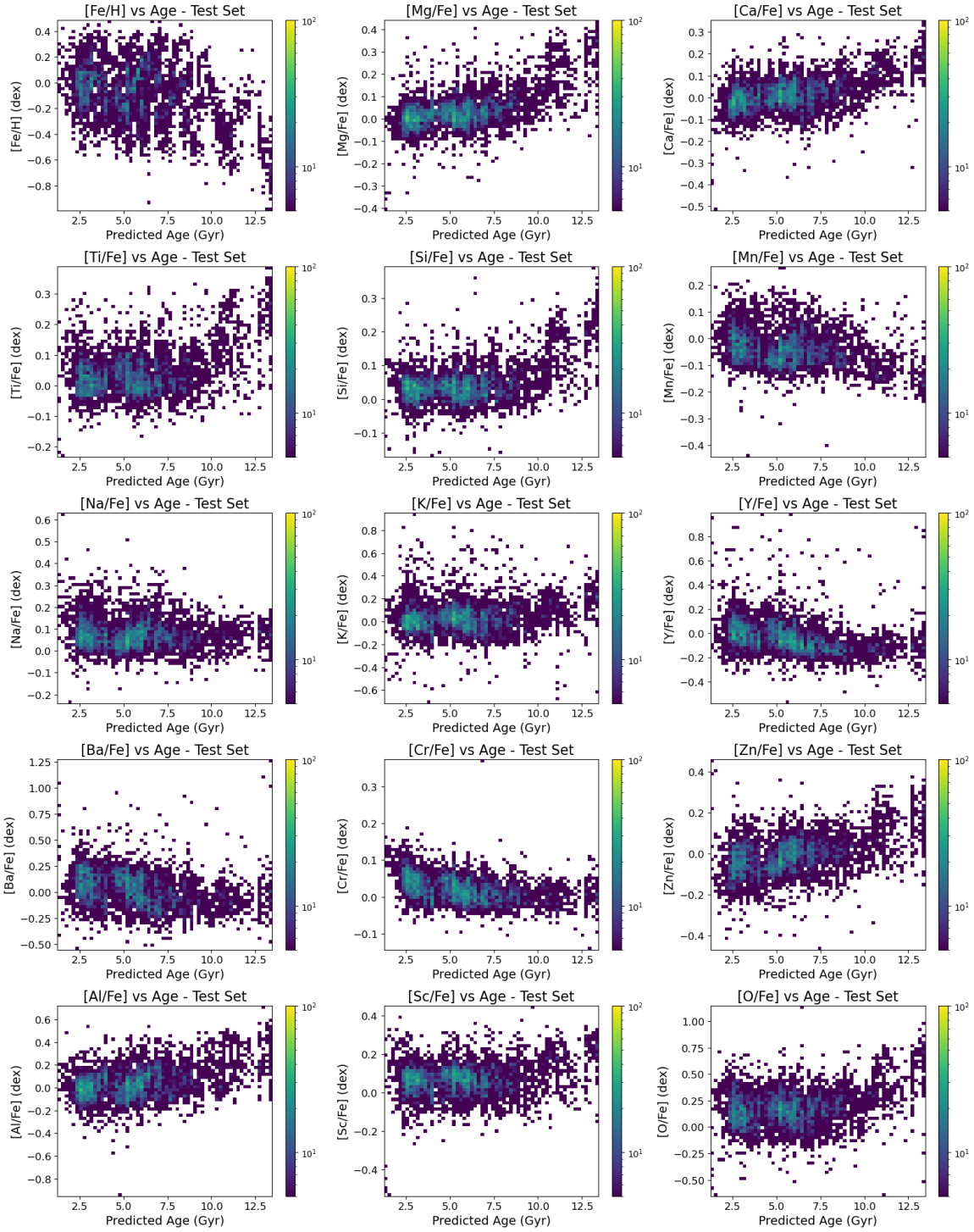


Figure 19: Elemental abundances versus XGBoost predicted age for all stars in the test set. [Fe/H] shows a monotonic decrease with age, which is physically sound. Additionally, α-element abundances show an increase with age, which is also physically sound.

for which isochrone fitting is ineffective or ambiguous.

6 Discussion

This study demonstrates that stellar chemical abundance patterns encode sufficient information to recover relative stellar ages with ~ 1 Gyr precision, even in the absence of precise isochrone-derived age labels for most evolutionary phases. By training an XGBoost regression model on a carefully curated set of Main Sequence Turn-Off stars with Bayesian isochrone ages, and then applying it to the broader GALAH DR4 disk population, it was shown that data-driven age estimation generalizes effectively across a large region of parameter space.

The model's performance on the test set, an RMSE of 1.38 Gyr and R^2 of 0.73, indicates that the chemical–age relation learned from the training sample captures meaningful astrophysical structure rather than simply overfitting to the training labels. Residual analysis revealed systematic biases near the extremes of the age distribution, consistent with regression to the mean behavior due to under-representation of very young and very old stars in the training set. Nonetheless, the model maintained high fidelity to the one-to-one age prediction line across most of the age range, with no catastrophic failure modes.

A central result of this study is the preservation of empirical age-abundance trends in the model outputs. The model reproduced known correlations between age and [Fe/H], [Mg/Fe], [Si/Fe], and other α-elements in the test set, validating the scientific utility of the predictions. When extended to non-turnoff stars, specifically, the Main Sequence and Red Giant Branch stars excluded from training, the predicted age-abundance patterns persisted with only modest increases in scatter. This extrapolative success supports the claim that the abundance–age mapping is sufficiently robust to extend beyond the evolutionary phases for which isochrone

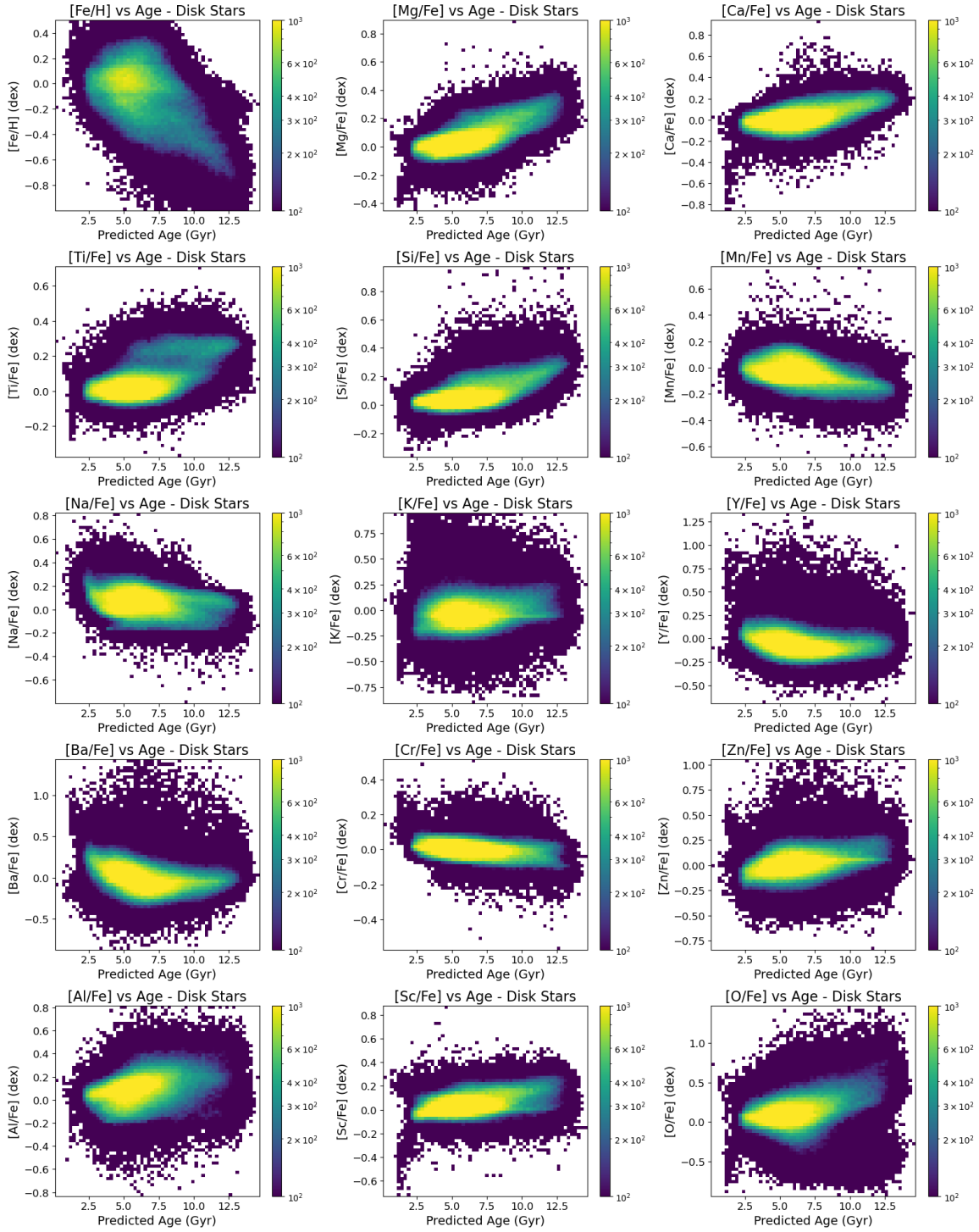


Figure 20: Elemental abundances versus XGBoost predicted age for all disk stars. The monotonic decrease with age of [Fe/H] is evident, just as before. The increase with age of the α-element abundance is also present. These patterns indicate the model is generalizing well to unseen stellar populations.

fitting is reliable.

The feature importance analysis further supports the model's interpretability and alignment with astrophysical expectations. [Mg/Fe], [Ca/Fe], and [Y/Fe] emerged as the most predictive features by gain, in agreement with theoretical and observational work showing that α-enhancement and s-process abundances track chemical enrichment timescales and thus stellar ages. In contrast, [Fe/H], despite being the most frequently used feature by weight, contributed relatively little predictive power in terms of gain, reflecting the degeneracy of the age–metallicity relation in the Galactic disk. These findings reinforce the idea that abundance-based age estimation is most effective when leveraging meaningful (in a nucleosynthetic sense) features.

An important consideration when interpreting the model's age predictions is the propagation of observational uncertainties in the stellar parameters and abundances. Spectroscopic measurements of $T_{\rm eff}$, $\log g$, [M/H], and individual element abundances all carry non-negligible errors, typically ranging from 30–100 K in effective temperature, 0.1–0.2 dex in surface gravity, and 0.05–0.1 dex in abundances depending on the element and signal-to-noise ratio. Because the XGBoost model learns complex nonlinear relations between these features and stellar age, even modest uncertainties can propagate in nontrivial ways, potentially amplifying the scatter in predicted ages. In particular, errors in α-element abundances like [Mg/Fe] and [Si/Fe], which are among the most important age predictors by gain, can directly bias the inferred enrichment timescales and lead to systematic age shifts. Moreover, correlated uncertainties (e.g., between $T_{\rm eff}$ and $\log g$ or between different abundance ratios) can further complicate error propagation. While the model's robustness is supported by the recovery of age-abundance trends, local predictions for individual stars should be interpreted with caution, especially in low signal-to-noise regimes. Future work could incorporate explicit uncertainty mod-

eling, such as perturbing input parameters within their error distributions (Monte Carlo sampling) or adopting Bayesian frameworks to propagate these uncertainties through to the final age predictions. Addressing these measurement-driven uncertainties is essential to fully realizing the potential of chemical chronometry in galactic archaeology, where the precise dating of stars underpins efforts to reconstruct birth clusters, merger events, and the temporal evolution of the Galactic disk.

Despite the promising results, several limitations remain. First, the age predictions rely on isochrone fitting for Turn-Off stars, which, while more precise than for other phases, still carries inherent model dependencies and uncertainties. Any systematics in the stellar models (e.g., opacities, mixing lengths, helium abundances, diffusion) will propagate into the machine learning predictions. Second, the training set was constructed with stringent quality cuts, particularly in signal-to-noise and age range, which likely biases the learned relations toward a cleaner but narrower region of chemical space. This conservative approach prioritizes label accuracy over sample diversity, which improves reliability but may limit generalizability to more chemically complex or noisier populations.

A critical factor to consider when interpreting the age predictions is the mismatch between the intrinsic age distribution of the Galactic disk and that of the training sample. The global star formation rate history (SFRH) of the Milky Way indicates that the majority of stars formed early in its evolution followed by a declining star formation rate to the present day. As a result, the overall stellar population is dominated by older stars. However, due to our location in the solar neighborhood and observational selection effects, the training sample used here is biased toward intermediate-age and younger stars. This local sampling bias arises because the solar vicinity is dominated by thin disk stars formed in more recent epochs, while older thick disk stars are relatively underrepresented both in the observed sample

and in the training set. Consequently, the learned abundance–age relation is anchored in a population that does not fully reflect the true global age distribution of the Milky Way. This bias could contribute to the systematic underestimation of ages for the oldest stars and overestimation for the youngest stars observed in the residual analysis.

A key challenge moving forward is the model's reduced confidence in sparsely populated age regimes. The systematic underestimation of old star ages and overestimation of young ones points to an imbalance in the training distribution. The stellar population in the Solar neighborhood is approximately 85% thin disk and 15% thick disk. As discussed in the Introduction, the stars of the thick disk are old. So, while the training sample is of the same distribution of the physical population, this may not be the best approach to take. A possible remedy to this is the construction of a more uniform training distribution. Alternatively, ensemble modeling or Bayesian machine learning approaches could be explored to quantify prediction uncertainty more rigorously in these regimes.

Finally, the methodology developed here has implications beyond age estimation. Because chemical patterns encode information about star formation histories, stellar migration, and enrichment processes, this approach offers a scalable framework for chemically tagging stars and reconstructing Galactic structure. With future surveys like 4MOST poised to produce even larger and higher-quality spectroscopic datasets, the techniques demonstrated here can be extended and refined for broader galactic archaeology applications.

In summary, this work validates the use of machine learning, specifically XGBoost regression, as a powerful tool for stellar age estimation in large-scale spectroscopic surveys. By exploiting the latent chronology encoded in chemical abundances, the model recovers physical trends, extrapolates effectively, and yields interpretable outputs that align with our current understanding of stellar and Galactic

evolution. This paves the way for more comprehensive and automated approaches to mapping the Milky Way's history.

7 Conclusion

This study set out to address the longstanding challenge of stellar age estimation by leveraging machine learning techniques applied to high-resolution spectroscopic data. By training an XGBoost regression model on a carefully selected set of Main Sequence Turn-Off stars with Bayesian isochrone-derived ages, this work demonstrated that stellar chemical abundances alone can predict relative ages with a root mean squared error of approximately 1.4 Gyr across a diverse population of disk stars.

The model not only performed well on a held-out test set, but also retained physical interpretability—reproducing well-established empirical age-abundance trends and identifying astrophysically meaningful features such as [Mg/Fe], [Ca/Fe], and [Y/Fe] as the most informative predictors. When applied to the broader GALAH DR4 sample, including Main Sequence and Red Giant Branch stars excluded from training, the model continued to recover known chemical evolution signatures, suggesting robust generalization beyond the training regime.

While limitations remain—particularly with respect to data imbalance, label uncertainties, and prediction confidence at age extremes—this work establishes a scalable and interpretable framework for chemical age estimation. The use of physically motivated features and a well-regularized learning algorithm ensured that the model learned meaningful patterns rather than spurious correlations. Importantly, this approach circumvents the limitations of isochrone fitting in evolutionary phases where traditional methods fail, expanding the utility of stellar age estimates across the Hertzsprung–Russell diagram.

The results presented here not only validate the use of ensemble machine learning for stellar chronometry but also pave the way for its application to upcoming large-scale spectroscopic surveys. As the volume and precision of stellar abundance data continue to grow, data-driven techniques such as this will play an increasingly central role in unraveling the formation and chemical enrichment history of the Milky Way. In particular, accurate stellar age estimates derived from chemical abundances open the door to powerful galactic archaeology applications. By combining precise ages with detailed chemical fingerprints, it becomes possible to chemically tag stars back to their original birth clusters, even if they have since dispersed throughout the Galaxy. This enables researchers to reconstruct the hierarchical assembly and merger history of the Milky Way, identify accreted stellar populations, and trace radial migration processes in the disk. Moreover, age information, when coupled with elemental abundance patterns, provides critical constraints on the sequence of star formation events and the timescales of chemical enrichment from different nucleosynthetic sources. Thus, the methods developed in this work not only improve our ability to date individual stars but also serve as a cornerstone for dissecting the complex evolutionary history of our Galaxy and our place within it.

References

S. A. Barnes. Ages for Illustrative Field Stars Using Gyrochronology: Viability, Limitations, and Errors. , 669(2):1167–1189, Nov. 2007. doi: 10.1086/519295.

D. Baron. Machine learning in astronomy: a practical overview, 2019. URL https://arxiv.org/abs/1904.07248.

A. Bressan, P. Marigo, L. Girardi, B. Salasnich, C. Dal Cero, S. Rubele, and A. Nanni. <scp>parsec</scp>: stellar tracks and isochrones with the padova and trieste stellar evolution code: <scp>parsec</scp>:tracks and isochrones. *Monthly Notices of the Royal Astronomical Society*, 427(1):127–145, Oct. 2012. ISSN 1365-2966. doi: 10.1111/j.1365-2966.2012.21948.x. URL http://dx.doi.org/10.1111/j.1365-2966.2012.21948.x.

S. Buder, S. Sharma, J. Kos, A. M. Amarsi, T. Nordlander, K. Lind, S. L. Martell, M. Asplund, J. Bland-Hawthorn, A. R. Casey, G. M. de Silva, V. D'Orazi, K. C. Freeman, M. R. Hayden, G. F. Lewis, J. Lin, K. J. Schlesinger, J. D. Simpson, D. Stello, D. B. Zucker, T. Zwitter, K. L. Beeson, T. Buck, L. Casagrande, J. T. Clark, K. Čotar, G. S. da Costa, R. de Grijs, D. Feuillet, J. Horner, P. R. Kafle, S. Khanna, C. Kobayashi, F. Liu, B. T. Montet, G. Nandakumar, D. M. Nataf, M. K. Ness, L. Spina, T. Tepper-García, Y.-S. Ting, G. Traven, R. Vogrinčič, R. A. Wittenmyer, R. F. G. Wyse, M. Žerjal, and Galah Collaboration. The GALAH+ survey: Third data release. , 506(1):150–201, Sept. 2021. doi: 10.1093/mnras/stab1242.

S. Buder, J. Kos, E. X. Wang, M. McKenzie, M. Howell, S. L. Martell, M. R. Hayden, D. B. Zucker, T. Nordlander, B. T. Montet, G. Traven, J. Bland-Hawthorn, G. M. De Silva, K. C. Freeman, G. F. Lewis, K. Lind, S. Sharma, J. D. Simpson, D. Stello, T. Zwitter, A. M. Amarsi, J. J. Armstrong, K. Banks, M. A. Beavis, K. Beeson,

B. Chen, I. Ciucă, G. S. Da Costa, R. de Grijs, B. Martin, D. M. Nataf, M. K. Ness, A. D. Rains, T. Scarr, R. Vogrinčič, Z. Wang, R. A. Wittenmyer, Y. Xie, and The GALAH Collaboration. The GALAH Survey: Data Release 4. *arXiv e-prints*, art. arXiv:2409.19858, Sept. 2024. doi: 10.48550/arXiv.2409.19858.

G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová. Machine learning and the physical sciences. *Reviews of Modern Physics*, 91(4), Dec. 2019. ISSN 1539-0756. doi: 10.1103/revmodphys.91.045002. URL http://dx.doi.org/10.1103/RevModPhys.91.045002.

W. J. Chaplin and A. Miglio. Asteroseismology of solar-type and red-giant stars. *Annual Review of Astronomy and Astrophysics*, 51(1):353–392, Aug. 2013. ISSN 1545-4282. doi: 10.1146/annurev-astro-082812-140938. URL http://dx.doi.org/10.1146/annurev-astro-082812-140938.

T. Chen and C. Guestrin. XGBoost: A Scalable Tree Boosting System. *arXiv e-prints*, art. arXiv:1603.02754, Mar. 2016. doi: 10.48550/arXiv.1603.02754.

C. Chiappini, F. Anders, T. S. Rodrigues, A. Miglio, J. Montalbán, B. Mosser, L. Girardi, M. Valentini, A. Noels, T. Morel, I. Minchev, M. Steinmetz, B. X. Santiago, M. Schultheis, M. Martig, L. N. da Costa, M. A. G. Maia, C. Allende Prieto, R. de Assis Peralta, S. Hekker, N. Themeßl, T. Kallinger, R. A. García, S. Mathur, F. Baudin, T. C. Beers, K. Cunha, P. Harding, J. Holtzman, S. Majewski, S. Mészáros, D. Nidever, K. Pan, R. P. Schiavon, M. D. Shetrone, D. P. Schneider, and K. Stassun. Young [α/fe]-enhanced stars discovered by corot and apogee: What is their origin? *Astronomy amp; Astrophysics*, 576: L12, Apr. 2015. ISSN 1432-0746. doi: 10.1051/0004-6361/201525865. URL http://dx.doi.org/10.1051/0004-6361/201525865.

J. Choi, A. Dotter, C. Conroy, M. Cantiello, B. Paxton, and B. D. Johnson. Mesa

isochrones and stellar tracks (mist). i. solar-scaled models. *The Astrophysical Journal*, 823(2):102, May 2016. ISSN 1538-4357. doi: 10.3847/0004-637x/823/2/102. URL http://dx.doi.org/10.3847/0004-637X/823/2/102.

S. S. Dhaliwal, A.-A. Nahid, and R. Abbas. Effective intrusion detection system using xgboost. *Information*, 9(7), 2018. ISSN 2078-2489. doi: 10.3390/info9070149. URL https://www.mdpi.com/2078-2489/9/7/149.

A. Dotter, B. Chaboyer, D. Jevremović, V. Kostov, E. Baron, and J. W. Ferguson. The dartmouth stellar evolution database. *The Astrophysical Journal Supplement Series*, 178(1):89–101, Sept. 2008. ISSN 1538-4365. doi: 10.1086/589654. URL http://dx.doi.org/10.1086/589654.

S. Fabbro, K. A. Venn, T. O'Briain, S. Bialek, C. L. Kielty, F. Jahandar, and S. Monty. An application of deep learning in the analysis of stellar spectra. *Monthly Notices of the Royal Astronomical Society*, 475(3):2978–2993, Dec. 2017. ISSN 1365-2966. doi: 10.1093/mnras/stx3298. URL http://dx.doi.org/10.1093/mnras/stx3298.

S. Feltzing, L. M. Howes, P. J. McMillan, and E. Stonkutė. On the metallicity dependence of the [y/mg]–age relation for solar-type stars. *Monthly Notices of the Royal Astronomical Society: Letters*, 465(1):L109–L113, 10 2016. doi: 10.1093/mnrasl/slw209. URL https://doi.org/10.1093/mnrasl/slw209.

B. D. Fields, K. A. Olive, T.-H. Yeh, and C. Young. Big-bang nucleosynthesis after planck. *Journal of Cosmology and Astroparticle Physics*, 2020(03):010–010, Mar. 2020. ISSN 1475-7516. doi: 10.1088/1475-7516/2020/03/010. URL http://dx.doi.org/10.1088/1475-7516/2020/03/010.

C. J. Fluke and C. Jacobs. Surveying the reach and maturity of machine learning and artificial intelligence in astronomy. *WIREs Data Mining and Knowledge*

Discovery, 10(2), Dec. 2019. ISSN 1942-4795. doi: 10.1002/widm.1349. URL http://dx.doi.org/10.1002/widm.1349.

M. R. Hayden, S. Sharma, J. Bland-Hawthorn, L. Spina, S. Buder, I. Ciucă, M. Asplund, A. R. Casey, G. M. De Silva, V. D'Orazi, K. C. Freeman, J. Kos, G. F. Lewis, J. Lin, K. Lind, S. L. Martell, K. J. Schlesinger, J. D. Simpson, D. B. Zucker, T. Zwitter, B. Chen, K. Čotar, D. Feuillet, J. Horner, M. Joyce, T. Nordlander, D. Stello, T. Tepper-Garcia, Y.-s. Ting, P. Wang, R. Wittenmyer, and R. Wyse. The galah survey: chemical clocks. *Monthly Notices of the Royal Astronomical Society*, 517(4):5325–5339, Oct. 2022. ISSN 1365-2966. doi: 10.1093/mnras/stac2787. URL http://dx.doi.org/10.1093/mnras/stac2787.

S. Hekker and J. Christensen-Dalsgaard. Giant star seismology. *The Astronomy and Astrophysics Review*, 25(1), June 2017. ISSN 1432-0754. doi: 10.1007/s00159-017-0101-x. URL http://dx.doi.org/10.1007/s00159-017-0101-x.

S. L. Hidalgo, A. Pietrinferni, S. Cassisi, M. Salaris, A. Mucciarelli, A. Savino, A. Aparicio, V. S. Aguirre, and K. Verma. The updated basti stellar evolution models and isochrones. i. solar-scaled calculations. *The Astrophysical Journal*, 856(2):125, Mar. 2018. ISSN 1538-4357. doi: 10.3847/1538-4357/aab158. URL http://dx.doi.org/10.3847/1538-4357/aab158.

G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning: with Applications in R*. Springer Publishing Company, Incorporated, 2014. ISBN 1461471370.

M. I. Jordan and T. M. Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015.

B. R. Jørgensen and L. Lindegren. Determination of stellar ages from isochrones:

Bayesian estimation versus isochrone fitting. , 436(1):127–143, June 2005. doi: 10.1051/0004-6361:20042185.

C. Kobayashi, A. I. Karakas, and M. Lugaro. The Origin of Elements from Carbon to Uranium. , 900(2):179, Sept. 2020. doi: 10.3847/1538-4357/abae65.

H. W. Leung and J. Bovy. Deep learning of multi-element abundances from high-resolution spectroscopic data. *Monthly Notices of the Royal Astronomical Society*, Nov. 2018. ISSN 1365-2966. doi: 10.1093/mnras/sty3217. URL http://dx.doi.org/10.1093/mnras/sty3217.

W. Li, Y. Yin, X. Quan, and H. Zhang. Gene expression value prediction based on xgboost algorithm. *Frontiers in genetics*, 10:1077, 2019.

M. Ma, G. Zhao, B. He, Q. Li, H. Dong, S. Wang, and Z. Wang. Xgboost-based method for flash flood risk assessment. *Journal of Hydrology*, 598:126382, 2021. ISSN 0022-1694. doi: https://doi.org/10.1016/j.jhydrol.2021.126382. URL https://www.sciencedirect.com/science/article/pii/S0022169421004297.

M. Martig, H.-W. Rix, V. Silva Aguirre, S. Hekker, B. Mosser, Y. Elsworth, J. Bovy, D. Stello, F. Anders, R. A. García, J. Tayar, T. S. Rodrigues, S. Basu, R. Carrera, T. Ceillier, W. J. Chaplin, C. Chiappini, P. M. Frinchaboy, D. A. García-Hernández, F. R. Hearty, J. Holtzman, J. A. Johnson, S. R. Majewski, S. Mathur, S. Mészáros, A. Miglio, D. Nidever, K. Pan, M. Pinsonneault, R. P. Schiavon, D. P. Schneider, A. Serenelli, M. Shetrone, and O. Zamora. Young α-enriched giant stars in the solar neighbourhood. , 451(2):2230–2243, Aug. 2015. doi: 10.1093/mnras/stv1071.

P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab. A high-bias, low-variance introduction to

machine learning for physicists. *Physics Reports*, 810:1–124, May 2019. ISSN 0370-1573. doi: 10.1016/j.physrep.2019.03.001. URL http://dx.doi.org/10.1016/j.physrep.2019.03.001.

S. Meibom, S. A. Barnes, I. Platais, R. L. Gilliland, D. W. Latham, and R. D. Mathieu. A spin-down clock for cool stars from observations of a 2.5-billion-year-old cluster. *Nature*, 517(7536):589–591, Jan. 2015. ISSN 1476-4687. doi: 10.1038/nature14118. URL http://dx.doi.org/10.1038/nature14118.

M. Ness, D. W. Hogg, H.-W. Rix, A. Y. Q. Ho, and G. Zasowski. The cannon: A data-driven approach to stellar label determination. *The Astrophysical Journal*, 808 (1):16, July 2015. ISSN 1538-4357. doi: 10.1088/0004-637x/808/1/16. URL http://dx.doi.org/10.1088/0004-637X/808/1/16.

P. E. Nissen. High-precision abundances of elements in solar twin stars: Trends with stellar age and elemental condensation temperature□□□. *Astronomy amp; Astrophysics*, 579:A52, June 2015. ISSN 1432-0746. doi: 10.1051/0004-6361/201526269. URL http://dx.doi.org/10.1051/0004-6361/201526269.

A. Ogunleye and Q.-G. Wang. Xgboost model for chronic kidney disease diagnosis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 17(6): 2131–2140, 2020. doi: 10.1109/TCBB.2019.2911071.

A. M. Serenelli, M. Bergemann, G. Ruchti, and L. Casagrande. Bayesian analysis of ages, masses and distances to cool stars with non-LTE spectroscopic parameters. , 429(4):3645–3657, Mar. 2013. doi: 10.1093/mnras/sts648.

D. R. Soderblom. The ages of stars. *Annual Review of Astronomy and Astrophysics*, 48(1):581–629, Aug. 2010. ISSN 1545-4282. doi: 10.1146/annurev-astro-081309-130806. URL http://dx.doi.org/10.1146/annurev-astro-081309-130806.

J. R. Taylor. *Introduction to Error Analysis, 2nd Ed. (cloth)*. University Science Books, 1997.

B. M. Tinsley. Stellar lifetimes and abundance ratios in chemical evolution. , 229: 1046–1056, May 1979. doi: 10.1086/157039.