Draft version June 6, 2023
Typeset using LaTeX **twocolumn** style in AASTeX631

Asteroseismology applied to constrain structure parameters of δ Scuti stars with machine learning

Subrata K. Panda,[1] Siddharth Dhanpal,[1] Simon J. Murphy,[2] Shravan Hanasoge,[1] and Timothy R. Bedding[3]

[1] *Tata Institute of Fundamental Research, Colaba, Mumbai 400005, Maharashtra, India*
[2] *Centre for Astrophysics, University of Southern Queensland, Toowoomba, QLD 4350, Australia*
[3] *Sydney Institute for Astronomy (SIfA), School of Physics, University of Sydney, NSW 2006, Australia*

ABSTRACT

Asteroseismology is a powerful tool to probe the structure of stars. Space-borne instruments like CoRoT, Kepler and TESS have observed the oscillations of numerous stars, among which δ Scutis are particularly interesting owing to their fast rotation rates and complex pulsation mechanisms. In this work, using machine learning as well as other algorithms, we inferred *model-dependent* masses, metallicities and ages of 60 δ Scuti stars from their photometric, spectroscopic and asteroseismic observations. We found most of these stars to be younger than 30 Myr, with masses around 1.6 $M_\odot$ and metallicities lower than $Z = 0.010$. This work emphasizes the importance of the large-frequency separation ($\Delta\nu$) in studies of δ Scuti stars. Our studies have further revealed that constraining dipole modes can help in significantly improving age estimation and that radial modes succinctly encode information pertaining to stellar luminosity and temperature. We also gained qualitative insights into the importance of stellar observables in estimating mass, metalicity, and age. The effective surface temperature $T_{\rm eff}$ strongly affects the inference of all structure parameters and the asteroseismic offset parameter ϵ plays an essential role in the inference of age.

1. INTRODUCTION

Asteroseismic observations from TESS (Ricker et al. 2014) and Kepler (Borucki et al. 2010) have shed light on the dynamics and interiors of thousands of pulsating stars (Paparó 2019; Bowman 2020; Aerts 2021; Kurtz 2022). A sizeable fraction of the pulsating class from these missions are δ Scuti stars (Bowman et al. 2018; Guzik 2021), which lie at the junction of the instability strip and the main sequence (MS) band in the Hertzsprung-Russell (HR) diagram (Uytterhoeven, K. et al. 2011), providing direct views of both these classes. These are low-to-intermediate mass ($1.5-2.5M_\odot$) main-sequence variables with spectral types ranging from A-F (Bowman et al. 2018; Murphy et al. 2021b; Kurtz 2022). These stars are used as standard candles (McNamara et al. 2007), and in assessing the metallicities and ages of stellar clusters (Murphy et al. 2022), in turn enabling Galactic archaeology.

Asteroseismology is a powerful tool which can be used to estimate various structure parameters such as mass, composition, and age on different pulsating classes, e.g., stochastic oscillators (Chaplin et al. 2013; Hekker & Christensen-Dalsgaard 2017; Vrard, M. et al. 2016; Hon et al. 2017, 2018; Dhanpal et al. 2022), γ-Doradus stars (Li et al. 2019; Mombarg, J. S. G. et al. 2021) and high-mass coherent oscillators (Hendriks et al. 2019), among others. Some efforts have been applied to similar studies of δ Sct stars (Suárez, J. C. et al. 2014a; García Hernández et al. 2017; Barceló Forteza, S. et al. 2020; Pamos Ortega et al. 2022). In this paper, we develop a methodology to measure structure parameters - mass, metallicity, and ages of δ Sct stars. These measurements can in principle constrain the metallicities and ages of the host clusters to which these stars might belong.

δ Sct stars oscillate predominantly in low radial order pressure modes ($\nu > 5$ d^{-1}), which are mainly driven by the κ mechanism (Chevalier 1971) in the Helium ionization zone. The pressure modes propagate primarily in the stellar envelope and probe the near-surface regions.

Although we have precise estimates of the luminosity (L), effective temperature ($T_{\rm eff}$) and pulsation frequencies, the seismology and the parameter inference of δ Sct stars is challenging because of the following reasons: (a) they have lower-overtone p-modes, where asymptotic theory fails, (b) many stars exhibit fast rotation, leading to ellipsoidal deformation (Reese, D. et al. 2006), (c) complex mode-selection mechanisms (Dziembowski et al. 1990) influencing the observed spectra, and (d) appearance of island modes (Reese, D. et al. 2006), chaotic modes (Barceló Forteza, S. et al. 2017) etc. making it very difficult to characterize δ Sct spectra.

Finding regular patterns in δ Sct spectra is very helpful for carrying out asteroseismology. Although some δ Sct stars had been previously reported to show regular frequency spacings (Zwintz, K. et al. 2011; Paparó, M. et al. 2013; Suárez, J. C. et al. 2014b), a larger ensemble of such stars was found by Bedding et al. (2020), who identified 60 δ Sct stars from *TESS* and *Kepler* exhibiting regular pulsation patterns. They were able to identify some modes, label their radial orders and deduce the large-frequency separation ($\Delta\nu$) and p-mode offset (ϵ). Many more examples of high-frequency patterns in δ Sct stars have subsequently been identified (e.g., Murphy et al. 2020; Murphy et al. 2021b; Hasanzadeh et al. 2021; Le Dizès et al. 2021; Kahraman Aliçavuş et al. 2022; Murphy et al. 2022). Here, we use seismic parameters to infer stellar structure through the application of neural networks and other techniques. As the relationships between observables such as luminosity, temperature, pulsation frequencies and structure parameters can be highly non-linear, neural networks can potentially be useful in building a model connecting them all. In addition, a well-trained neural network infers these structure parameters substantially faster than conventional methods such as MCMC, making it a capable method for ensemble studies.

In this paper, we developed three deep-learning methods and a least-squares fitting technique to deduce stellar structure parameters. All of them require synthetic stellar structure models. Section 2 contains descriptions of this synthetic dataset. We provide a brief background on machine learning methods in Section 3. In section 4, we describe how we trained a simple neural network and inferred stellar parameters from 43 observed pulsation spectra. In sections 5 and 6, we design two superior methods involving the manipulation of eigenfrequencies that aided in more accurate parameter inference. In these three methods, we assessed the feature importance of different stellar observables qualitatively. However, none of the above methods worked uniformly well over all the 60 stars of our sample because of their various limitations. This led us to develop a grid-search based least-squares minimization technique, described in Section 7, with which we were able to confidently infer (M, Z, τ) parameters for all the 60 δ Scutis in our sample.

2. GRID OF STELLAR MODELS

We built a grid of synthetic δ Sct models following the procedure described in Murphy et al. (2022). First, we constructed a large library of evolutionary models, comprising 371 evolutionary tracks, each of which has an initial mass (M) and metallicity (Z) drawn from a uniformly homogeneous distribution over the ranges $[1.3-2.2]$ $M_\odot$ and $[0.002-0.026]$, with $0.1M_\odot$ (M) and 0.002 (Z) spacing respectively. These models were then evolved using the stellar evolution code MESA (Paxton et al. 2010, 2013, 2015, 2018). The evolution of each stellar track was initiated from a pre-main sequence (pre-MS) phase with a central temperature of 9×10^5 K to ensure computational convergence (Soderblom 2010). We stopped the evolution at approximately one third of the terminal-age main-sequence age, before the avoided crossings of mixed modes disrupt the neat pulsation patterns. Since we did not evaluate mode amplitudes or pulsation driving/damping effects, we approximated the δ Sct instability strip with the effective temperature limits $T_{\rm eff}\in[6800{\rm K}, 10000{\rm K}]$. We considered models within this range to be potential δ Sct stars – around 49,000 of the $\sim$ 96,470 models were included in this range.

We implemented convection using mixing-length theory (Canuto 1990), with the Henyey formalism (Henyey et al. 1965), in which we set the mixing-length parameter $\alpha_{\rm MLT}$ to 1.9. We did not consider $\alpha_{\rm MLT}$ as a variable parameter because chemically homogeneous stellar models do not change significantly with varying values of $\alpha_{\rm MLT}$ (Murphy et al. 2021b). In our simulations, we also included overshooting (Herwig 2005) past the core (with $f_0 = 0.002,\ f = 0.022$) and at the surface (with $f_0 = 0.001,\ f = 0.006$). To model the atmosphere, we adopted the Eddington $T-\tau$ formalism (Eddington 1926). Our models were non-rotating and did not include element diffusion. From these MESA models we obtained $L,\ T_{\rm eff},\ \log g,\ \rho,\ R$, and other quantities as a function of evolutionary age.

We used the stellar oscillation code GYRE (Townsend et al. 2013, 2017; Goldstein et al. 2020) to calculate pulsation frequencies at each time step of the evolution. To obtain realistic resonant modes, GYRE performed a search in the range $5-95$ d^{-1} for $m=0$ p modes having degrees ℓ =0 and 1. We then fitted a linear polynomial to the radial ($\ell=0$) p-mode resonant frequencies at radial orders $n\in[5-9]$ to determine the large-frequency separation ($\Delta\nu$) and offset parameter (ϵ).

We used the results to construct a large grid that contains structure parameters (M, Z, τ, ρ, etc.), observables ($L, T_{\rm eff}$), and asteroseismic quantities ($\Delta\nu, \epsilon$). This grid was used to train the neural networks to infer structure parameters from observables and asteroseismic quantities.

The HR diagram (Figure 1) shows the stellar models computed using MESA, including the potential δ Sct models ($3.83\le\log T_{\rm eff}\le4$) on this HR diagram. For

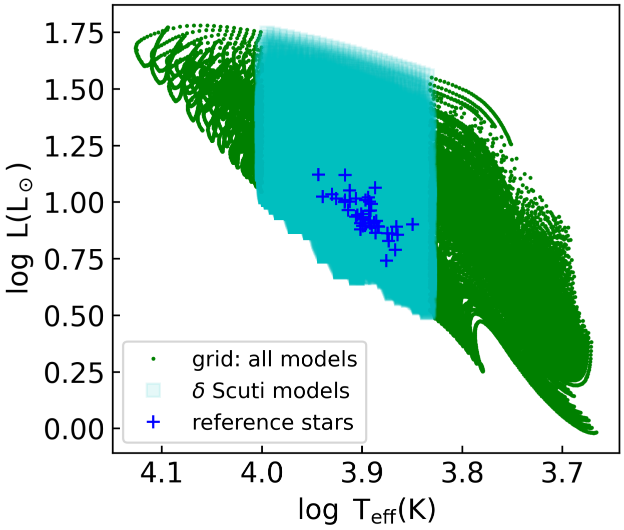

Figure 1: HR diagram showing the positions of all synthetic stellar models, with δ Sct stars present in the solid patch. The δ Sct targets of this paper are denoted using '+' symbols.

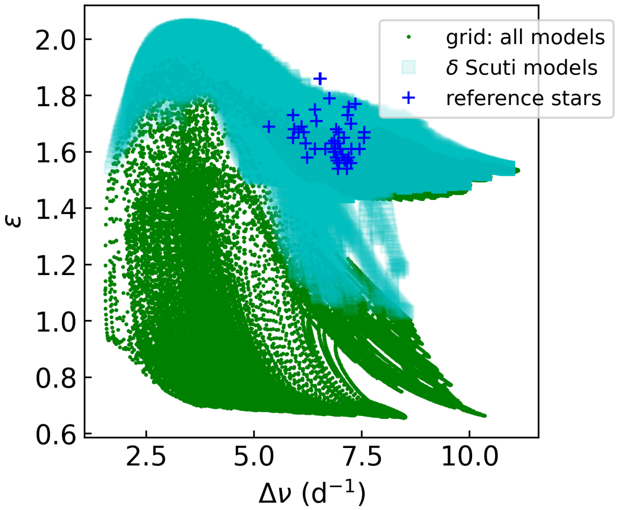

Figure 2: $\Delta\nu - \epsilon$ diagram denoting the positions of all synthetic stellar models, with δ Sct stars present in the solid patch. The δ Sct targets of this paper are denoted using '+' symbols.

reference, we mark the target stars of this work using '+' symbols, which are taken from the observed δ Sct stars in Bedding et al. (2020). Figure 2 contains the same models as Figure 1, but showing the asteroseismic $\Delta\nu$–ϵ diagram. Both figures demonstrate that the phase space of our stellar grid is broad enough to represent the wide range of observed δ Sct stars.

3. MACHINE LEARNING

The application of machine learning (ML) algorithms to the fitting of stellar models has shown great promise (Verma et al. 2016; Hendriks et al. 2019; Dhanpal et al. 2022; Scutt et al. 2023). Neural networks (Glorot et al. 2011) constitute a specialized machine-learning algorithm, which are intended to accomplish regression, classification, encoding-decoding, sample generation, and time-series prediction tasks. They are often used to discover unknown results from big data and fit models to observations much more quickly than MCMC-like methods.

For the work presented in this article, we have used regression networks to infer continuous values of the parameters of interest. Our networks were developed using the Python libraries TENSORFLOW (Abadi et al. 2015) and KERAS (Chollet et al. 2015).

We configured multi-layer perceptron networks to infer M, Z and age (τ). The accuracy of the inference appeared to increase when the network predicted these parameters separately, instead of a single monolithic network predicting all outputs simultaneously. This limits the estimation of the uncertainty covariance matrix to on-diagonal elements, and a more general methodology to simultaneously infer all parameters would be ideal, which we will attempt in a future effort. We did not develop a neural network to infer radius (R), since it may be directly obtained from $\{L, T_{\rm eff}\}$ using the black-body radiation equation.

In each set of ML experiments, we randomly split the δ Sct model grid into 'training' and 'test' (validation) data at a 95 : 5 ratio. The network used the training data to learn different features and compute the output. Unseen test data were used to quantify the accuracy. If the accuracy was deemed acceptable on the validation data, we concluded that the network was successfully trained.

To train this neural network, we optimized the mean-squared error (MSE) between the actual and predicted values, as shown in equation 1.

$$\mathrm{MSE} = \sum_{i=1}^{N} \frac{(y^i_{\rm predicted} - y^i_{\rm true})^2}{N} \tag{1}$$

where N is the number of samples. We used the ADAM optimizer (Kingma et al. 2014) to minimize the loss function.

The neural networks predicted a single number for each parameter of interest. For the uncertainty associated with each parameter, we used the cross-validation method (Mosteller et al. 1968), which involved dividing the dataset into k equal subsets, training the model on $k-1$ parts, and keeping the $k^{\rm th}$ dataset for validation. This resulted in k networks, each with different weights.

Setting $k = 40$, we trained 40 networks in parallel to yield 40 instances of predictions of M, $\log Z$, and $\log \tau$. We considered the median of the 40 samples as the most probable result. We found that the 40 instances of inferences sometimes followed highly skewed distributions. Hence we set the 16^{th} and 84^{th} percentiles of the inferred distributions as the lower and higher uncertainty limits, respectively.

Normalization is an important aspect of ML workflows in order to deal with inputs that fluctuate over different scales (or ranges). Hence, we first normalized the output values of our dataset to the range [-1, 1]. Therefore, for the inferences of observations, trained networks also predicted values between [-1, 1]. To obtain the appropriate inference, we processed the network outputs by un-normalizing them. The output layer of each network was equipped with an L2 regularizer with regularizing parameter $\lambda = 10^{-6}$, chosen as such to prevent overfitting.

4. METHOD 1: ML USING SEISMIC INDICES

In this method we deployed three regression networks to infer M, Z and τ from four input parameters: $\{L, T_{\text{eff}}, \Delta\nu, \epsilon\}$. The following subsections describe the network architecture and the success of the method in fitting structure parameters to physical observables. Although the overall accuracy is encouraging, we find it difficult to deduce the age using this method – the reasons for which we also discuss. Finally, we report the result of parameter fits to 43 δ Sct stars from Bedding et al. 2020. As an add-on to ML methods, we show how we can gain qualitative insights into feature importances of different stellar observables, where analytical understanding is not well established.

4.1. *Model architecture*

Through a series of trials, we found that the network achieves optimal accuracy when it has access to certain observables (inputs) to infer desired structure parameters (output). This has been summarized in Table 1. Although radius (R) is dependent on L and T_{eff}, using it as an additional input helps the networks train quickly and achieve higher accuracy. For the network to infer age (τ), it typically requires prior knowledge of M and Z: this is because each physical parameter (θ) is a functions of M, Z and τ, which is why age inversion requires (M, Z) to be treated as inputs.

$$\theta = f(M, Z, \tau) \implies \tau = \tilde{f}^{-1}(M, Z, \theta)$$

Since we developed a few alternative methods, we label this procedure as *Method-1* for easy reference.

Output	Inputs
M	$L,\ T_{\text{eff}},\ R,\ \Delta\nu,\ \epsilon$
Z	$L,\ T_{\text{eff}},\ R,\ \Delta\nu,\ \epsilon$
τ	$M,\ Z,\ L,\ T_{\text{eff}},\ R,\ \Delta\nu,\ \epsilon$

Table 1: Stellar structure parameters (output) may be inferred from input observables.

As the outputs from each of these networks are physical quantities with continuous values, we used regression models to make predictions. These regression networks have sequential architectures, with each network comprising 1 input layer, 10 intermediate layers, and 1 output layer. The intermediate layers have 400 neurons, the output layers have 1 neuron and the input layers have 5 or 7 neurons depending on the output parameter (Table 1). With the exception of the output layer, which has a *tanh* activation function, all other neurons were activated using the rectified linear unit (reLU) function. Since the output layer has *tanh* activation, its outputs are constrained to within [-1, 1]. This was chosen so as to limit the output to within the boundary of the model grid. We show the model architecture in Figure 3.

4.2. *Results on synthetics*

We trained the networks and compared the network-predicted values for the validation data with the corresponding true values (Figure 4). The network was able to achieve high accuracy in predictions of M, and Z. Figure 5 shows the distributions of errors in our predictions on the validation data. It indicates that, for 99% of samples, the network predicted M to within 1% of the actual values. Similarly, these networks were able to recover Z to within 3.5% and $\log \tau$ to within 13% for 90% of the stars.

4.3. *Challenges in inferring age*

From Figure 4 (c), it is evident that age inferences are accurate in younger ($\log \tau < 1$) and older ($\log \tau > 2.5$) stars. The large scatter in Figure 4(c) around $\log \tau :\sim 1.3 - 2.5$ shows that the network systematically over-predicts the ages of stars near Zero Age Main Sequence (ZAMS).

Possible reason for this discrepancy is that no observable evolves as a monotonic function of age. All structure and seismic parameters cross their pre-main-sequence values sometimes within the main sequence, as shown in Figure 6. Hence, none of our inputs may be used to uniquely distinguish between pre-main sequence ($\log \tau \sim 1$) and main-sequence ($\log \tau \sim 2.5$) phases. This fact was also discussed in Murphy et al. 2021a.

4.4. *Feature Importance*

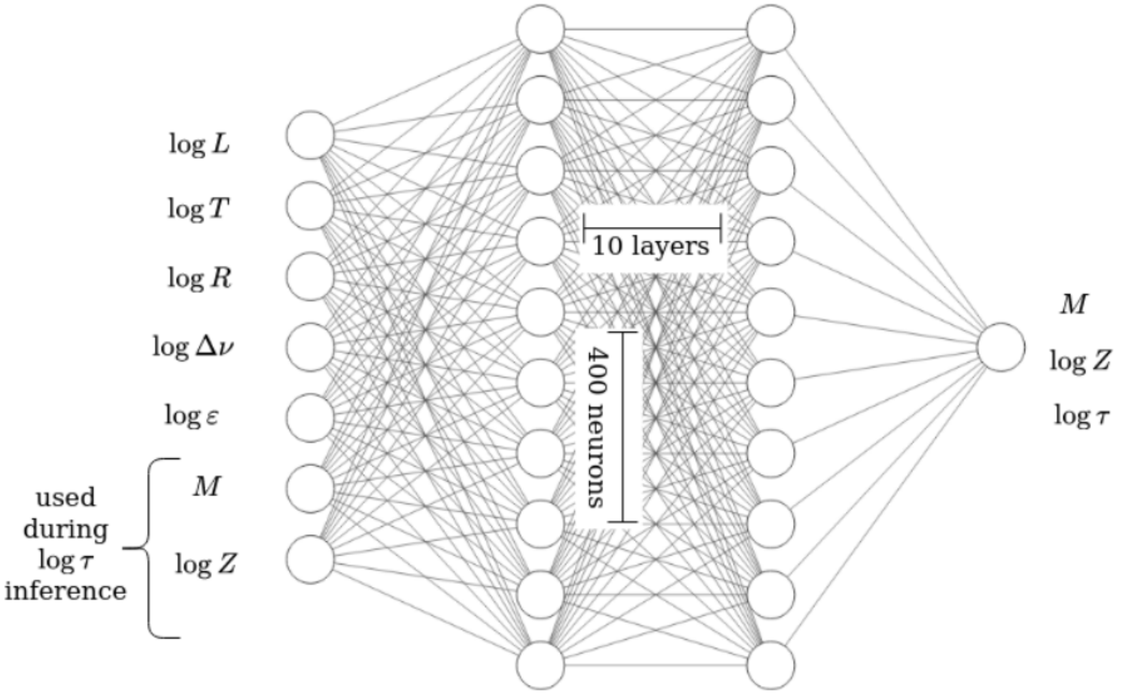

Figure 3: Architecture of the neural-network models.

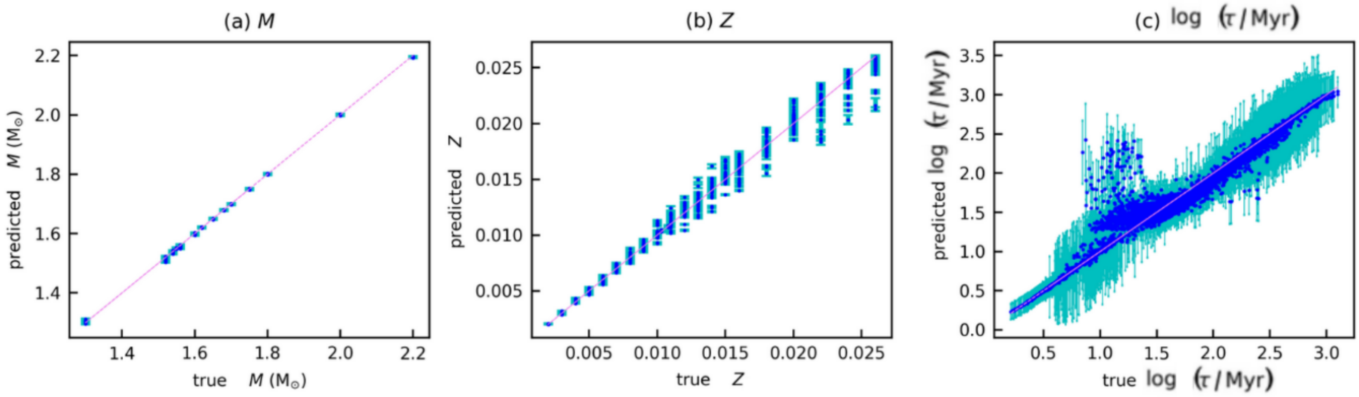

Figure 4: Plots showing network predictions for validation samples against their corresponding true values. Shown are (a) M, (b) Z, and (c) $\log \tau$. Uncertainties associated with predicted values are also shown.

No known analytic relations connect the observables with stellar-structure parameters. However, there have been efforts (Bellinger et al. 2016) to determine the importance of different observables connected to stellar parameters. Similarly, we determined the qualitative importance of various inputs to determine how strongly they influence parameter estimation. In other words, this reveals how errors in the input quantities propagate through the network and perturb the output.

The inner workings of neural networks are not easily interpreted (Montavon et al. 2018); however, the network finds correlations between different input quantities and the output. To measure independent contributions arising from the inputs, we perturbed each input quantity by 0.5% (without changing other inputs) to measure the relative differences in outputs. We expected that the contribution of the input is proportional to the relative difference in output. Hence, this output error may be considered a measure of the strength of the corresponding input quantity. We carried out this process for all input quantities to determine their qualitative importance. We show the average relative differences (a proxy for feature importance) in Figure 7.

Figure 7 indicates that $\log T_{\rm eff}$ contributes most strongly towards the inference of all parameters. M and Z are necessary inputs for the inference of age. Al-

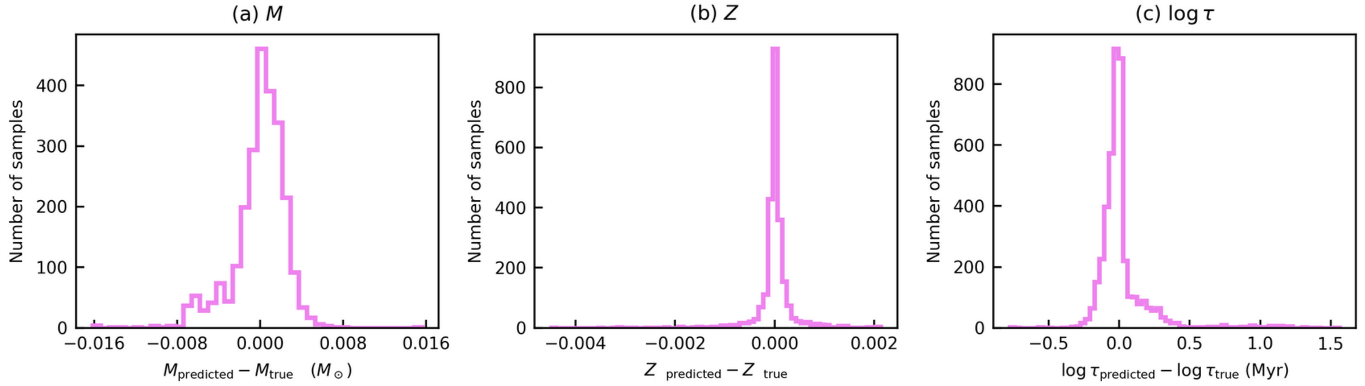

Figure 5: Errors between network predictions for validation samples and their corresponding true values: shown are (a) M, (b) Z, and (c) $\log \tau$.

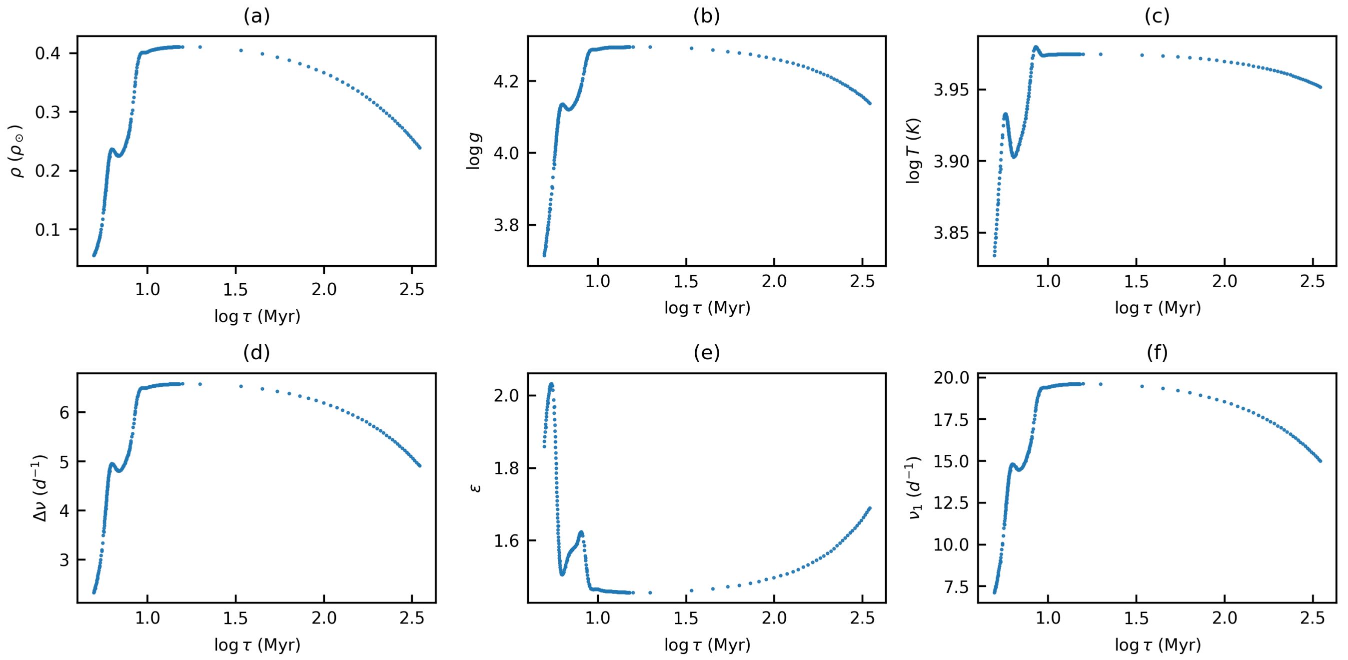

Figure 6: Evolution with time of stellar properties for a model with mass 2.2 $M_\odot$ and $Z = 0.026$. Most structural and seismic quantities cross their pre-MS values during MS. Hence, values of these parameters are degenerate at two different ages. Crossing effects are shown for (a) ρ, (b) $\log g$, (c) $\log T_{\rm eff}$, (d) $\Delta\nu$, (e) ϵ and (f) ν_1 (frequency corresponding to $n = 1, \ell = 0$).

though ϵ is an offset parameter that controls the shift in the $\ell = 0$ frequencies, it plays an important role in the inference of τ. This was also emphasized in Bedding et al. 2020, where the $\Delta\nu - \epsilon$ diagram was shown to encode information about age (τ). $\Delta\nu$ plays an important role in the inference of Z_i. It is well known that $\Delta\nu$ highly correlates with square root of mean stellar density – similarly, it also tightly related to metallicity (Z) because the higher the metallicity, the lower the stellar density.

To clarify the method implemented so far, we have used seismic indices $\Delta\nu$ and ϵ (and not individual mode frequencies) as inputs along with L and $T_{\rm eff}$. This formalism has been used in the next subsection, where we obtain results for 43 δ Sct stars.

4.5. *Results: Inference for 43 stars*

We implemented our trained networks on 43 *TESS* δ Sct stars to infer their M, Z and age (τ). These stars are a part of the 60 δ Sct stars available from Bedding et al. (2020) and have simultaneously available measurements of $T_{\rm eff}, L, \Delta\nu, \epsilon$. For the remaining 17 stars, we could not apply our method because we could not find all of

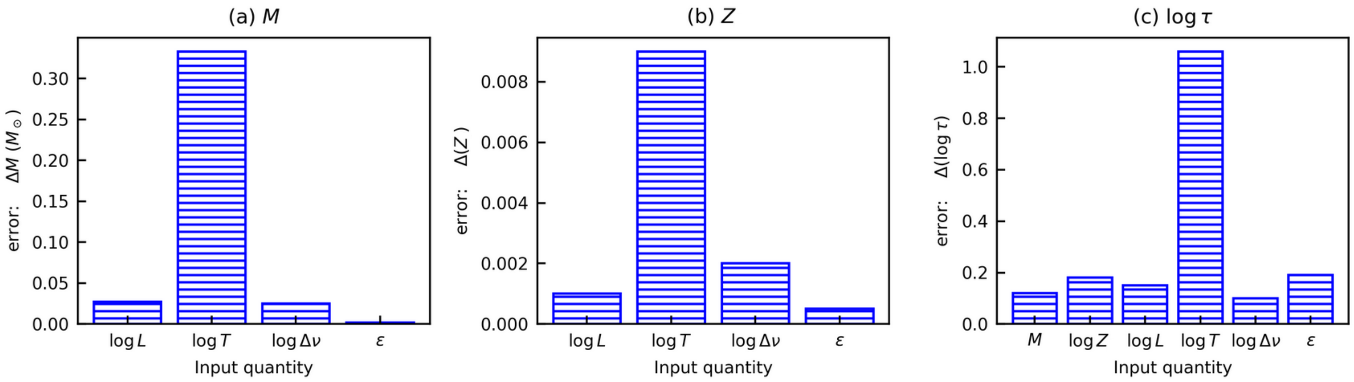

Figure 7: Different inputs contribute unequally towards overall inferences of (a) M, (b) Z and (c) $\log\tau$. This figure qualitatively depicts the contribution strengths corresponding to the input parameters when perturbed by 0.5%.

these 4 parameters, although $\Delta\nu$ was measured for all of them.

No inconsistencies were found in our investigation of the statistics of inferred M and τ of observed stars. However, the inferred Z distribution posed a challenge – most of the inferences were 0.026, which is the upper boundary of our model grid. We guess the input observables of these stars probably fall in sparse regions of the training hyperspace – which leads our networks to do unreliable interpolation. Moreover, since the networks interpolate through the model grid, it is not possible to make direct assessments of the pulsation frequencies corresponding to the fitted parameters and then compare with the observed modes. This is possible only if one can evolve the star (corresponding to fitted parameters) and perform the pulsation calculation anew. One way to visualize frequency matching is to find a suitable model from the dense grid that closely fits all the observations. This is indeed the method we present in Section 7.

5. METHOD 2: ML USING RADIAL MODES

Thus far, we have only used L, $T_{\rm eff}$, and seismic indices - $\Delta\nu$ and ϵ in Method-1 (Section 4) to infer structure parameters. We have not used individual mode frequencies as inputs, with which we will train the neural net in this section.

Although we used L as an input quantity in Method-1, the reliability of its measurements depends on the accuracy of stellar distance, as L is calculated using magnitude and distance. Since structure parameters are very sensitive to $\log T_{\rm eff}$ (evident from section 4.4), errors in measuring the latter propagate into inferences of the former.

In this section, we show that if we use eigenfrequencies directly instead of these observables $(L, T_{\rm eff})$, we still achieve similar results while overcoming existing issues.

In order to keep this method simple, we used radial modes as inputs (no dipole modes). Figures from Bedding et al. (2020) convey that radial modes are far more easily identified than dipole modes. Conventional techniques such as period ratios (Petersen & Christensen-Dalsgaard (1996)) are also helpful for identifying radial modes.

5.1. *Method*

Since most of our synthetic samples have nearly seven overtones of radial-mode oscillations, we decided to use seven frequencies as inputs to train the neural network. This is also supported by the observed spectra of several δ Sct stars (Bedding et al. 2020; Murphy et al. 2022).

5.2. *Results on synthetics*

In this formalism, neural networks could learn all the parameters, even without needing $\{L, T_{\rm eff}, \Delta\nu, \epsilon\}$, as seen in Figure 8, which shows that the network inferences are very close to the true values. Since $\Delta\nu, \epsilon$ are computed from radial modes, we did not use those as inputs. The networks were able to recover M to within 1%, Z to within 5% and $\log\tau$ to within 7% for 90% of the stars. In Table 2, we compare the average learning accuracy (= 100% - average% error) associated with each parameter inference using the current method and Method-1 of Section 4. This method is as accurate as Method-1 in inferring M and Z. However, it infers age at a significantly higher accuracy (Table 2). We designate this procedure as *Method-2* in order to distinguish it from Method-1.

This method highlights the importance of the $\ell = 0$ eigenfrequencies of radial orders $n = 1 - 7$. It shows that the non-asymptotic modes succinctly preserve the representation of L and $T_{\rm eff}$, although how this is so is unclear.

Despite its robustness, we were unable to apply this method to the observed stars since identifying continuous radial overtones ($\nu_1 - \nu_7$) is difficult. Additionally, it was harder to automate this method for multiple stars

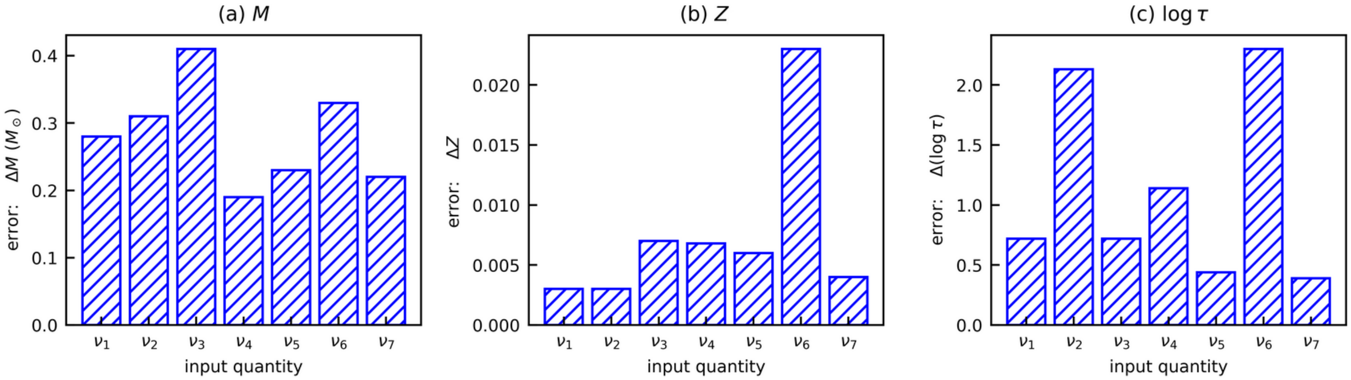

Figure 8: Network predictions for validation samples for (a) M, (b) Z_i and (c) $\log\tau$, on taking eigenfrequencies $\{\nu_1, \ldots, \nu_7\}$ as inputs (without L, $T_{\rm eff}$, $\Delta\nu$, ϵ). This implies that these non-asymptotic modes are helpful for parametric inference, particularly useful when L or $T_{\rm eff}$ are not available or reliable.

	Method-1	Method-2
	(Section 4)	(Current Method)
Inputs:	$(L, T_{\rm eff}, \Delta\nu, \epsilon)$	$\{\nu_1 - \nu_7\}$
M	99.9%	99.5%
Z	99.7%	99.5%
$\log\tau$	93%	96%

Table 2: Comparing performances as quantified using average learning accuracy of the robust method discussed in this section and Method-1 of Section 4. While both perform similarly, Method-2 works without $\{L, T_{\rm eff}, \Delta\nu, \epsilon\}$.

because identifying radial orders requires significant human intervention. Identifying radial modes crucially depends on choosing the correct $\Delta\nu$, without which échelle diagrams cannot be constructed. Even if this method worked efficiently, we were unable to resolve the age discrepancy (similar as Method-1) over the synthetics, as evident from Figure 8 (c). In Section 6, we propose another method that improves over this issue.

We investigated the strengths of contributions arising from each mode which participated in the inference process. We measured their importance in a manner similar to that in Section 4.4. The attendant measures are shown in Figure C5. It may be understood from the figure that all frequencies are not equally important. ν_6 has the highest contribution, followed by ν_2 and ν_3. Hence, both non-asymptotic and asymptotic modes seem to be important.

6. METHOD 3: ML USING RADIAL AND DIPOLE MODES

In this section, we present the final follow-up experiment to study the importance of dipole modes. Every experiment carried out so far has dealt with radial ($\ell = 0$) modes, or quantities such as $\{\Delta\nu, \epsilon\}$ that depend on radial modes. However, since dipole modes are often seen in the échelle diagrams (Bedding et al. 2020) of δ Sct stars, it is useful to conduct these experiments in order to study their contribution.

6.1. *Method*

In the échelle diagrams of observed δ Sct stars (Bedding et al. 2020), higher-order overtones ($\sim n > 4$) of radial and dipole modes are vertically aligned. Frequency separations tend to increase with the lower-order modes, seen as a curvature in the mode ridges. Sometimes, the low radial-order modes either do not get excited or we cannot reliably identify them, e.g., because they may be surrounded by neighbouring modes. In this method, we built a model that took observables $\{L, T_{\rm eff}\}$ and eigenfrequencies with $n \in [4-7]$ from each of the $\ell = 0$ and 1 ridges, and produced as output the structure parameters. For age inference, we supplemented pre-inferred M and Z to the frequencies and $\{L, T_{\rm eff}\}$. We refer to this procedure as *Method-3*.

6.2. *Results on synthetics*

Using the current method (Method-3), the networks were able to learn M,Z, and $\log\tau$. Figure 9 shows the results of the network on synthetics. The network could recover M to within 1% for 96% of stars. Similarly, it was able to recover Z to within 5% and $\log\tau$ to within 4% for 90% of the stars. Figure 9 and Table 3 indicate that the accuracy of age inference has improved. Since the current method adequately constrains age, this emphasizes the importance of dipole modes.

At present, we were unable to apply this method to measure structure parameters in observed stars as it is challenging to accurately label the dipole modes, because of unreliable mode identification.

To understand the importance of the dipole and radial modes, we carried out a feature-importance experi-

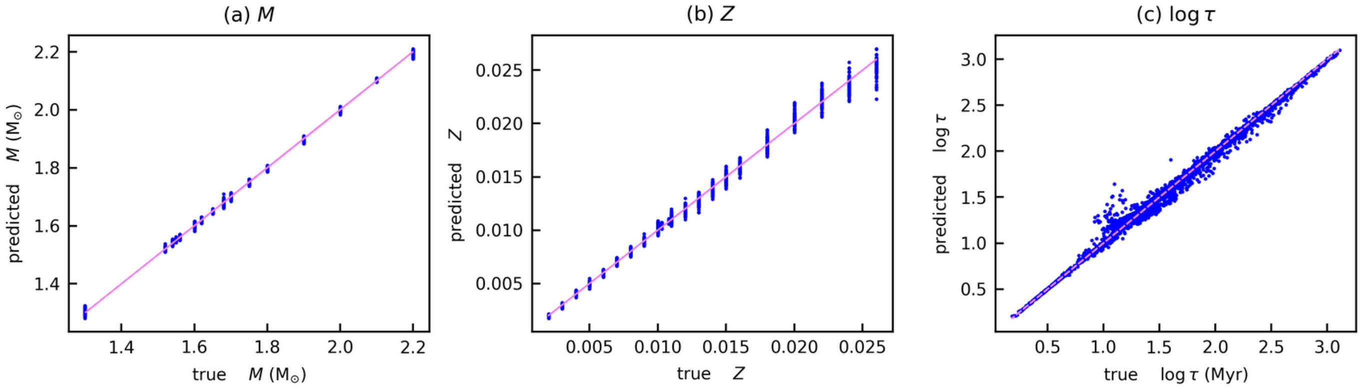

Figure 9: Improved network prediction after inputting radial-dipole ($\ell = 0, 1$) modes of overtone $n \in [4-7]$ along with $\{L, T_{\rm eff}\}$. The age inference is much more accurate than the approaches described earlier.

	Method-1 (Section 4)	Method-2 (Section 5)	Method-3 (Current Method)
Inputs:	$\{L, T_{\rm eff}, \Delta\nu, \epsilon\}$	$\{\nu_1 - \nu_7\}$	$\{L, T_{\rm eff}$, 8 modes$\}$
M	99.9%	99.5%	99.8%
Z	99.7%	99.5%	99.6%
$\log\tau$	93%	96%	98.1%

Table 3: Comparing performances (average learning accuracy) of Methods-1 and 2 (Section 4, 5) with Method-3 (this section), where we use $T_{\rm eff}$, L, 4 radial and 4 zonal dipole modes as inputs.

ment for these parameters and show the plots in Figure D6. Both radial and dipole modes have higher contributions than $\{L, T_{\rm eff}\}$ in age (τ) inference. Dipole modes add important information supplementary to the radial modes, enabling higher-accuracy inferences, e.g., in determining age.

With more number of inputs to neural network, it is supposed to gain higher accuracy. We also measured χ^2/N for all the three experiments to assess which method is more efficient. Here χ mean average squared differences between true and machine predicted values and N represents the number of inputs fed to the networks. Table 4 summarizes these facts.

	Method-1 (Section 4)	Method-2 (Section 5)	Method-3 (Current Method)
Inputs:	$\{L, T_{\rm eff}, \Delta\nu, \epsilon\}$	$\{\nu_1 - \nu_7\}$	$\{L, T_{\rm eff}$, 8 modes$\}$
M	7×10^{-5}	10^{-5}	2×10^{-6}
Z	10^{-7}	5×10^{-8}	9×10^{-9}
$\log\tau$	3×10^{-3}	6.5×10^{-4}	10^{-3}

Table 4: Comparing χ^2/N for all the methods.

7. METHOD 4: LEAST SQUARE FITTING OF STELLAR PARAMETERS

Each method described above has its own set of advantages and disadvantages. Results from ML methods are susceptible to extrapolation if their inputs are not covered within the training data or if they fall within a sparse region of the data. This can lead to higher uncertainties associated with inferences drawn from the results.

In this section, we employed a grid-search algorithm to minimize the least-square loss functions in order to identify the model best matching to the observation. This method requires interactive inspection to obtain accurate fits, provided that there exists a close fit model to the observation. A similar approach was used by Steindl, T. et al. 2022 to fit stellar parameters to a small number of δ Sct stars.

One of the advantages of this method is it can be applied to the 60 stars from Bedding et al. 2020 to achieve good results. Even for stars with missing values of L or $T_{\rm eff}$, this method can obtain a good fit to observed spectra successfully.

7.1. *Method*

The method of least squares is a common technique to fit photometric (L), spectroscopic ($T_{\rm eff}$) and seismic quantities (eigen-frequencies) between observations and models. The underlying principle involves assigning a χ^2 value (as shown in equation 2) to each sample in the model, which represents a weighted combination of the squared differences between the observed and model values. The objective is to find the model sample that results in the minimum value of χ^2. The values of σ_L, $\sigma_{T_{\rm eff}}$, and σ_{ν_i} are critical parameters in this approach, as they represent the magnitude of the errors that can be tolerated. This methodology can be highly effective if the appropriate σ values are used.

$$\chi^2 = \left\{ \frac{(L_{\rm obs} - L_{\rm model})^2}{\sigma_L^2} + \frac{(T_{\rm eff;obs} - T_{\rm eff;model})^2}{\sigma_{T_{\rm eff}}^2} + \sum_{i=1}^{N} \frac{(\nu_{i,\rm obs} - \nu_{i,\rm model})^2}{\sigma_{\nu_i}^2} \right\} \tag{2}$$

The last part of equation 2 indicates the requirement of determining both the observed ($\nu_{i,\rm obs}$) and model frequencies ($\nu_{i,\rm model}$) of the same radial order and angular degree. However, accurately identifying and labeling the observed modes in δ Sct stars is challenging due to the high density contamination, missing modes as well as presence of modes with unknown origin. Identification of genuine modes requires iterative pre-whitening followed by elimination of frequency combinations – which adds to the complexity of the fitting routine.

In our approach, we have made slight modifications to a similar method. Our primary objective is to search for a model that can reprodcue as many observed modes as possible without having additional spurious peaks. We have applied constraints to the models using the observed values of $\Delta\nu$ (which are obtained from Bedding et al. 2020) to prevent selection of erroneous fits - since a model with much lower $\Delta\nu$ can accurately match a relatively large number of significant observed modes. Finally, we introduce a χ^2 term to compare few observed frequencies to their nearest model frequencies ($\nu_{\rm model}^{\rm closest}$). Our modified χ^2 looks like equation 3.

$$\chi^2 = \left\{ \frac{(L_{\rm obs} - L_{\rm model})^2}{\sigma_L^2} + \frac{(T_{\rm eff;\ obs} - T_{\rm eff;\ model})^2}{\sigma_{T_{\rm eff}}^2} + \frac{(\Delta\nu_{\rm obs} - \Delta\nu_{\rm model})^2}{\sigma_{\Delta\nu}^2} + \sum_{\nu_1}^{\nu_N} \frac{(\nu_{\rm obs} - \nu_{\rm model}^{\rm closest})^2}{\sigma_\nu^2} \right\} \tag{3}$$

The χ^2 formula therefore has similarity to that of Steindl, T. et al. (2022) and the likelihood function of Scutt et al. (2023).

Values of σ_L have been taken from Bedding et al. (2020). Following the same, we have set the uncertainty in $T_{\rm eff}$ to be 2% and that of $\Delta\nu$ to be 0.02 $\rm d^{-1}$. Rayleigh frequency resolution criterion has been taken as the uncertainty in mode frequencies, similar to Steindl, T. et al. (2022). Since we fitted three parameters simultaneously (M, Z, τ), 1σ uncertainties (or 68% confidence interval) associated with the best-fit parameters correspond to 3.5 increase from $\chi^2_{\rm min}$ (Table 1 of Avni (1976)). Hence we assembled all the model parameters corresponding to $\chi^2 \in [\chi^2_{\rm min}, \chi^2_{\rm min} + 3.5]$ and used their spread to calculate the uncertainties.

Regarding the frequency component of the χ^2, we neither calculated eigen-frequenies from iterative pre-whitening nor assigned any possible identifications to the peaks. Instead, we visually inspected the spectrum and selected usually four modes from the higher frequency end, even if their amplitudes were relatively faint. We selected these modes through trial and error. After begining with these arbitrary set of modes we searched for their closest eigen-frequencies ($\nu_{\rm model}^{\rm closest}$) across all the model samples disregarding their n or ℓ. Treating these frequenies as inputs to equation 3, we calculated χ^2 across all the models and inspected for $\chi^2_{\rm min}$. We kept on selecting different sets of input frequencies as well as kept on reducing the number of fitted modes (from N = 4 to 1) and repeated the entire fitting process afresh until we achieve a minimum possible $\chi^2_{\rm min}$.

The reason we favoured selecting very few modes is as follows. It is not always possible to obtain exact fits to all the observed modes simultaneously. In such scenarios, fitting larger numbers of modes would demand a highly sophisticated mode selection process since we are picking up the modellable modes through trial and error. Inappropriate mode selection usually leads to a solution where none of the modes are comparable to the observed modes. However, for each star, we were successful to identify a smaller number of genuine modes using which we could fit most of the observed modes.

This fitting routine takes us $\sim$ 10 seconds given that we vectorized this operation across 112 cpu cores using NumPy. Otherwise, this would correspond to $\sim$ 20 cpu minutes (per single star) without any core-level parallelization. However the mode selection process (being manual and interactive) is the most difficult one and it took us around 10 minutes per a single star.

In Figure 10, we present an example of our successfully fitted results using this methodology. It was challenging for us to find a close-fit model where lower-order model modes would precisely align with the observed modes. Additionally, dipole modes of the fitted models were sometimes seen to be present in the vicinity of the observed ones but not precisely. This discrepancy could possibly be due to the unequal splitting of dipole modes and the rotation-induced shift of $m = 0$ components, even in radial modes.

7.2. *Results*

In this section, we present the results of our method on 60 stars from Bedding et al. 2020. For 3 of these stars, either one or both of $(L, T_{\rm eff})$ inputs were missing. For

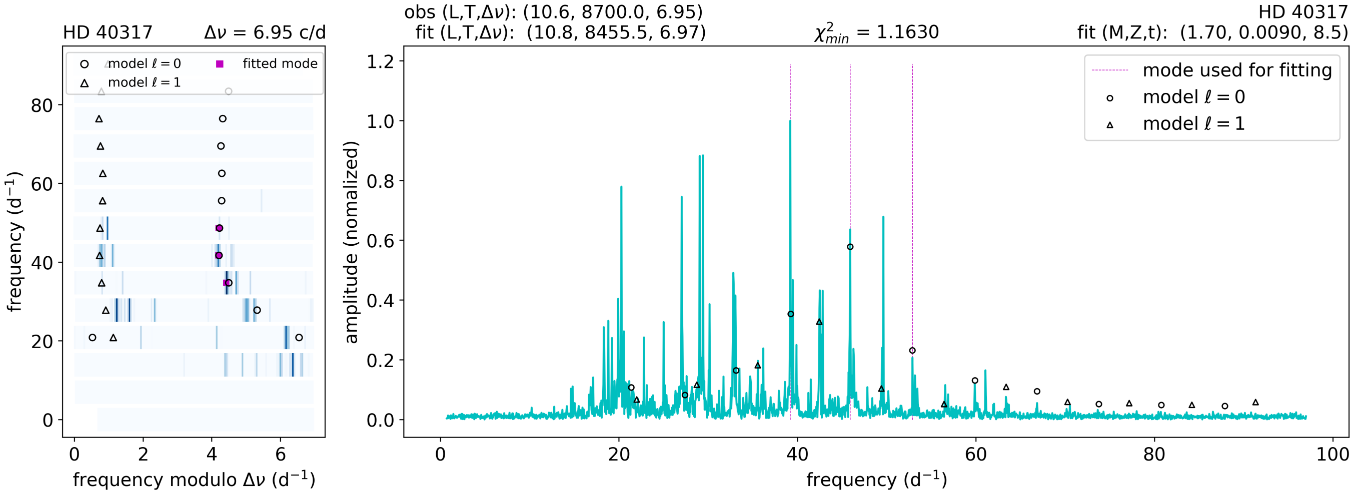

Figure 10: Obtaining a best-fit model to observation of HD 40317. The dotted lines in the right side spectrum shows the 3 modes we selected for the fitting routine. These 3 modes are also shown in the échelle diagram (on left) as filled rectangle symbols. After determining the best-fit model, we show its radial and dipole modes as open circles and triangles over both the échelle diagram and the oscillation spectrum. Some of these open symbols do not exactly fall on the observed modes because the fitted frequencies fall within the spectral windows in the vicinity of the sharp peaks. We also report the observed and best-fitted values of $L, T_{\rm eff}$ and $\Delta\nu$ in the title of the spectrum plot. The (M, Z, τ) values of the best fitted model are presented in the same title.

such stars, we ignored the corresponding contributions of χ^2. As the best fit model has an inherent value of missing parameters, one can have a crude estimate of the parameters for these stars.

In Table 5 and Figure A1 - A3 of appendix A, we have presented the fitted structure parameters of the individual stars. Additionally, Figure 11 summarizes the statistics, indicating that most of the fitted masses are approximately ~ 1.6 M$_\odot$, while the dominant metallicity values are distributed around $Z = 0.010$, corresponding to [Fe/H] ~ -0.23. We also observed a bi-modal age distribution among the stars, where most of them are young (around $\sim$ 10 Myr) and a few are very old (over 100 Myr). However, we emphasize that age inferences are not highly precise due to the degeneracy effect, which means that at two different ages, the star can have similar physical and seismic structure.

Table 5. Neural network inferred values of mass, composition, and age of 60 δ Sct stars taken from Bedding et al. (2020).

Id	mass $(M_\odot)$	Z_i	age (Myr)
HD 2280	$1.35^{+0.1}_{-0.05}$	$0.008^{+0.001}_{-0.002}$	$977.7^{+202.9}_{-1.23}$
HD 3622	$1.7^{+0.05}_{-0.15}$	$0.022^{+0.002}_{-0.014}$	$50.6^{+75.3}_{-42.04}$
HD 10779	$1.6^{+0.1}_{-0.02}$	$0.009^{+0.013}_{-0.001}$	$9.0^{+156.43}_{-0.25}$
HD 17341	$1.65^{+0.1}_{-0.01}$	$0.012^{+0.006}_{-0.001}$	$9.0^{+2.67}_{-0.01}$
HD 17693	$1.7^{+0.05}_{-0.05}$	$0.012^{+0.001}_{-0.003}$	$9.7^{+0.01}_{-1.59}$
HD 20203	$1.5^{+0.04}_{-0.05}$	$0.008^{+0.001}_{-0.001}$	$692.5^{+0.06}_{-132.78}$
HD 20232	$1.68^{+0.02}_{-0.13}$	$0.018^{+0.002}_{-0.011}$	$229.8^{+104.08}_{-221.47}$
HD 24572	$1.52^{+0.06}_{-0.02}$	$0.013^{+0.003}_{-0.004}$	$344.0^{+132.62}_{-331.75}$
HD 24975	$1.7^{+0.05}_{-0.06}$	$0.014^{+0.001}_{-0.003}$	$10.4^{+0.6}_{-1.55}$
HD 25674	$1.54^{+0.01}_{-0.04}$	$0.007^{+0.001}_{-0.001}$	$829.4^{+77.82}_{-0.08}$
HD 28548	$1.7^{+0.05}_{-0.06}$	$0.012^{+0.001}_{-0.006}$	$163.0^{+158.0}_{-155.63}$
HD 30422	$1.5^{+0.02}_{-0.05}$	$0.006^{+0.001}_{-0.001}$	$8.3^{+0.01}_{-0.01}$
HD 31322	$1.75^{+0.05}_{-0.05}$	$0.011^{+0.001}_{-0.002}$	$7.7^{+0.01}_{-0.75}$
HD 31640	$1.55^{+0.09}_{-0.03}$	$0.009^{+0.004}_{-0.001}$	$9.3^{+2.44}_{-0.23}$
HD 31901	$1.58^{+0.06}_{-0.02}$	$0.009^{+0.001}_{-0.001}$	$9.4^{+0.01}_{-0.58}$
HD 32433	$1.62^{+0.03}_{-0.1}$	$0.018^{+0.002}_{-0.01}$	$201.1^{+193.23}_{-191.66}$
HD 38597	$1.58^{+0.06}_{-0.02}$	$0.008^{+0.002}_{-0.001}$	$694.8^{+0.02}_{-185.18}$
HD 38629	$1.45^{+0.05}_{-0.1}$	$0.004^{+0.001}_{-0.001}$	$972.4^{+0.21}_{-0.86}$
HD 40317	$1.7^{+0.05}_{-0.08}$	$0.009^{+0.001}_{-0.002}$	$8.5^{+0.16}_{-1.17}$
HD 42005	$1.6^{+0.02}_{-0.02}$	$0.008^{+0.001}_{-0.001}$	$8.6^{+0.01}_{-0.01}$
HD 42608	$1.75^{+0.05}_{-0.1}$	$0.018^{+0.002}_{-0.01}$	$13.8^{+96.09}_{-5.7}$
HD 44726	$1.55^{+0.01}_{-0.03}$	$0.011^{+0.001}_{-0.003}$	$396.6^{+253.75}_{-383.59}$
HD 44930	$1.64^{+0.01}_{-0.1}$	$0.018^{+0.008}_{-0.005}$	$13.7^{+453.78}_{-3.78}$

Table 5 *continued*

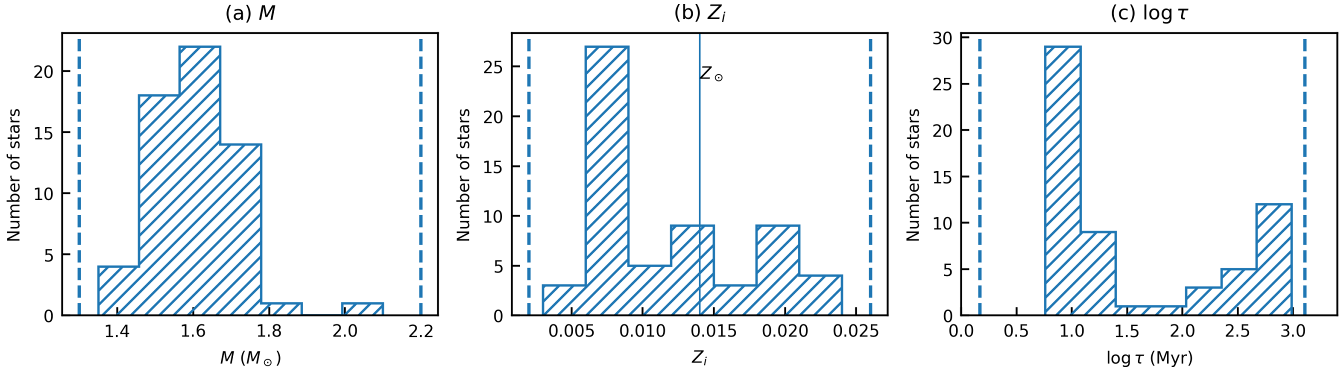

Figure 11: Statistics of the fundamental parameters of 60 δ Sct stars, inferred using the least-squares minimization method. The dotted vertical lines represent the boundary of the model grid.

Table 5 *(continued)*

Id	mass ($M_\odot$)	Z_i	age (Myr)
HD 44958	$1.6^{+0.1}_{-0.02}$	$0.009^{+0.013}_{-0.001}$	$9.1^{+11.22}_{-0.05}$
HD 45424	$1.55^{+0.01}_{-0.03}$	$0.008^{+0.001}_{-0.001}$	$823.0^{+54.73}_{-0.06}$
HD 46722	$1.52^{+0.01}_{-0.02}$	$0.008^{+0.001}_{-0.001}$	$9.1^{+0.27}_{-0.01}$
HD 48985	$1.64^{+0.04}_{-0.02}$	$0.008^{+0.001}_{-0.001}$	$8.5^{+0.01}_{-0.67}$
HD 50153	$1.7^{+0.05}_{-0.02}$	$0.02^{+0.002}_{-0.002}$	$15.9^{+0.01}_{-0.01}$
HD 54711	$1.62^{+0.06}_{-0.12}$	$0.009^{+0.006}_{-0.003}$	$9.8^{+378.22}_{-1.95}$
HD 55863	$1.64^{+0.01}_{-0.06}$	$0.009^{+0.001}_{-0.001}$	$8.6^{+0.69}_{-0.09}$
HD 59104	$1.56^{+0.02}_{-0.02}$	$0.018^{+0.002}_{-0.002}$	$19.1^{+297.27}_{-1.96}$
HD 59594	$1.6^{+0.05}_{-0.1}$	$0.015^{+0.005}_{-0.007}$	$13.9^{+476.0}_{-4.77}$
HD 78198	$1.62^{+0.06}_{-0.02}$	$0.014^{+0.006}_{-0.001}$	$9.9^{+4.74}_{-0.37}$
HD 99506	$1.65^{+0.03}_{-0.1}$	$0.018^{+0.002}_{-0.01}$	$16.2^{+246.96}_{-7.1}$
HD 223011	$1.62^{+0.18}_{-0.02}$	$0.01^{+0.014}_{-0.001}$	$8.7^{+438.49}_{-0.32}$
HD 290799	$1.55^{+0.01}_{-0.01}$	$0.004^{+0.001}_{-0.001}$	$7.1^{+0.01}_{-0.93}$
TIC 349645354	$1.75^{+0.05}_{-0.05}$	$0.024^{+0.002}_{-0.004}$	$12.3^{+0.01}_{-2.14}$
TIC 431695696	$1.54^{+0.08}_{-0.04}$	$0.006^{+0.002}_{-0.001}$	$8.1^{+1.14}_{-0.01}$
TIC 124381332	$1.68^{+0.07}_{-0.06}$	$0.014^{+0.001}_{-0.007}$	$13.3^{+145.36}_{-5.31}$
TIC 340358522	$1.62^{+0.02}_{-0.02}$	$0.008^{+0.001}_{-0.001}$	$8.4^{+0.01}_{-0.01}$
HD 187547	$1.58^{+0.02}_{-0.04}$	$0.018^{+0.002}_{-0.002}$	$177.9^{+0.64}_{-160.44}$
KIC 8415752	$1.75^{+0.05}_{-0.05}$	$0.024^{+0.002}_{-0.002}$	$344.6^{+0.04}_{-0.02}$
KIC 9450940	$1.62^{+0.03}_{-0.02}$	$0.015^{+0.001}_{-0.001}$	$702.5^{+53.63}_{-51.24}$
HD 37286	$1.52^{+0.06}_{-0.02}$	$0.006^{+0.005}_{-0.001}$	$8.5^{+4.51}_{-0.01}$
HD 39060	$1.45^{+0.05}_{-0.1}$	$0.008^{+0.001}_{-0.001}$	$861.4^{+0.45}_{-0.53}$
HD 42915	$1.55^{+0.09}_{-0.01}$	$0.006^{+0.001}_{-0.001}$	$7.6^{+0.01}_{-0.87}$
HD 290750	$2.1^{+0.1}_{-0.1}$	$0.013^{+0.013}_{-0.001}$	$5.8^{+185.38}_{-0.07}$
TIC 143381070	$1.5^{+0.02}_{-0.05}$	$0.007^{+0.001}_{-0.001}$	$594.4^{+0.03}_{-117.6}$
TIC 260161111	$1.54^{+0.06}_{-0.04}$	$0.008^{+0.002}_{-0.003}$	$599.3^{+51.98}_{-591.2}$
HD 10961	$1.35^{+0.1}_{-0.05}$	$0.003^{+0.001}_{-0.001}$	$8.4^{+0.01}_{-0.01}$
HD 25248	$1.5^{+0.02}_{-0.05}$	$0.009^{+0.005}_{-0.001}$	$10.9^{+4.14}_{-0.01}$
HD 67688	$1.68^{+0.02}_{-0.03}$	$0.008^{+0.001}_{-0.001}$	$7.6^{+0.01}_{-0.01}$
HD 70510	$1.65^{+0.03}_{-0.01}$	$0.018^{+0.002}_{-0.002}$	$90.0^{+0.12}_{-74.06}$
HD 75040	$1.6^{+0.02}_{-0.05}$	$0.011^{+0.001}_{-0.001}$	$10.0^{+0.58}_{-0.09}$
HD 222496	$1.8^{+0.05}_{-0.15}$	$0.022^{+0.002}_{-0.009}$	$12.1^{+485.56}_{-3.64}$
HD 34282	$1.6^{+0.02}_{-0.1}$	$0.009^{+0.001}_{-0.004}$	$421.4^{+390.4}_{-0.02}$
HD 29783	$1.68^{+0.02}_{-0.04}$	$0.009^{+0.001}_{-0.001}$	$8.5^{+0.3}_{-0.61}$
HD 220811	$1.54^{+0.04}_{-0.01}$	$0.015^{+0.003}_{-0.001}$	$849.5^{+0.05}_{-154.81}$
HD 25369	$1.58^{+0.06}_{-0.04}$	$0.012^{+0.006}_{-0.001}$	$9.9^{+5.47}_{-0.33}$
HD 89263	$1.52^{+0.23}_{-0.07}$	$0.011^{+0.013}_{-0.006}$	$561.8^{+194.6}_{-555.8}$

Table 5 *continued*

8. CONCLUSION

We trained three neural networks as well as a least-squares minimization technique to predict M, Z ([Fe/H]) and age (τ) for 60 δ Sct stars that show regular p-mode pulsation patterns, which allow $\Delta\nu$ to be measured. We provided the first inferences of these quantities for these stars.

We found the masses of most of these stars to be distributed around $1.6M_\odot$, with the exception of two stars of lower mass (HD 2280, HD 10961: $\sim 1.4M_\odot$) and one star with the highest mass (HD 290750: $2.1M_\odot$). Metallicity of a significant fraction of the stars are found to hover around $Z = 0.010$ (or [Fe/H] = -0.23). In this sample, we also found a few stars having very low as well as very high values of metallicity. TIC 349645354 and KIC 8415752 have the highest metallicity, at $Z = 0.024$. Finally, most stars are younger than 30 Myr with very few stars older than $\sim$ 100s of Myr.

On the synthetic data, our neural networks were able to infer M and Z much more accurately using $\{L, T_{\rm eff}, \Delta\nu, \epsilon\}$ as inputs. However, age inferences were in general not as accurate as M or Z. This is because, over the course of their MS evolution, stars cross their pre-MS trajectory, which make *all* their structure parameters degenerate in these two stages of evolution. A similar behaviour was demonstrated in Murphy et al. 2021a. Therefore, age inference is not expected to be confident.

We determined the relative importance of different input quantities using machine leaning approaches. $T_{\rm eff}$ plays a critical role in the inference of all parameters. ϵ was found to have significant importance when constraining age.

We observed that a longer pattern of radial modes (starting from $n = 1$) contains essential information about δ Sct structures. This set of frequencies may be treated as an essential substitute to $\{L, T_{\rm eff}\}$ as we found that they can constrain stellar parameters even without L and $T_{\rm eff}$. Doing the feature importance experiment, we found that ν_6, ν_2 and ν_3 are the most significant radial modes.

Finally, we developed a formalism that was able to more precisely determine δ Sct age. We used L, $T_{\rm eff}$ and frequencies of both radial ($\ell = 0$) and dipole ($\ell = 1$) modes with radial orders $n \in [4-7]$ as inputs to infer different parameters. The degeneracy problem in age was reduced while using these inputs. We therefore conclude that dipole modes act as independent quantities with which to constrain stellar parameters and they add supplementary information to radial modes and $\{\Delta\nu, \epsilon\}$. In the future, we aim to use this method to precisely measure the ages of multiple δ Sct stars observed in missions such as *TESS* and *Kepler*.

ACKNOWLEDGMENTS

S.D. acknowledges SERB, DST, Government of India, CII and Intel Technology India Pvt. Ltd. for the Prime minister's fellowship and facilitating research. We have performed all computations in the Intel Lab Academic Compute Environment using Intel® Xeon® Platinum 8280 CPU. MESA (Paxton et al. 2010) and GYRE (Townsend et al. 2013) codes have been very useful in simulating the data, without which this project would not have been possible. We have used SciPy (Oliphant 2007), Matplotlib (Hunter 2007), NumPy (van der Walt et al. 2011), Keras (Chollet et al. 2015), and Tensorflow (Abadi et al. 2015) like Python packages. This research made use of Lightkurve (Lightkurve Collaboration et al. 2018), a Python package for Kepler and TESS data analysis. GNU parallel (Tange 2011) has been used to do parallel computation. SJM was supported by the Australian Research Council (ARC) through Future

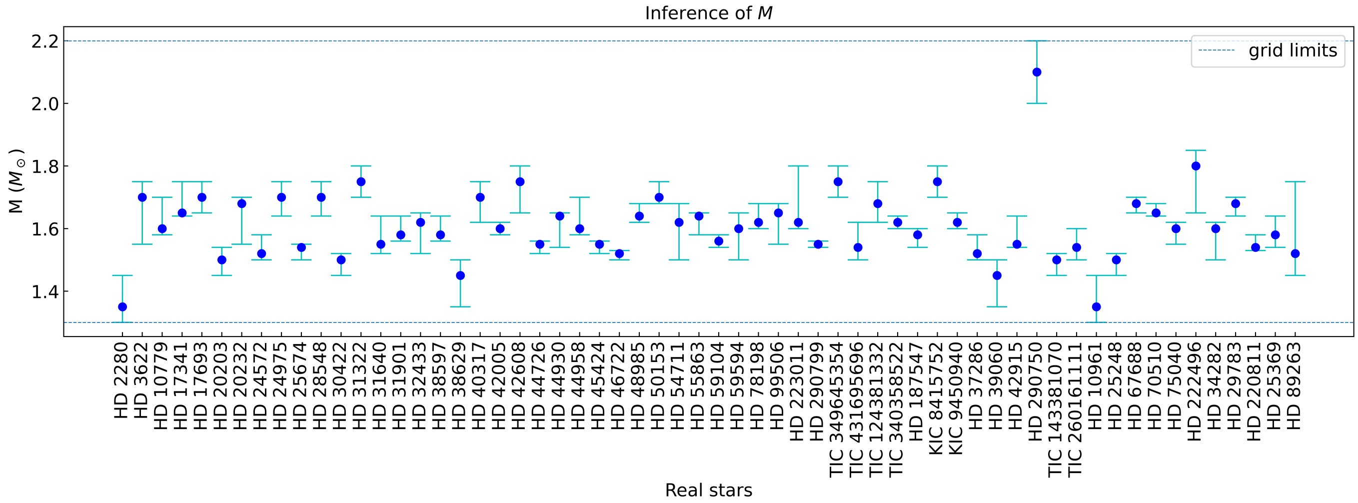

Figure A1: M inference for 60 δ Scuti stars.

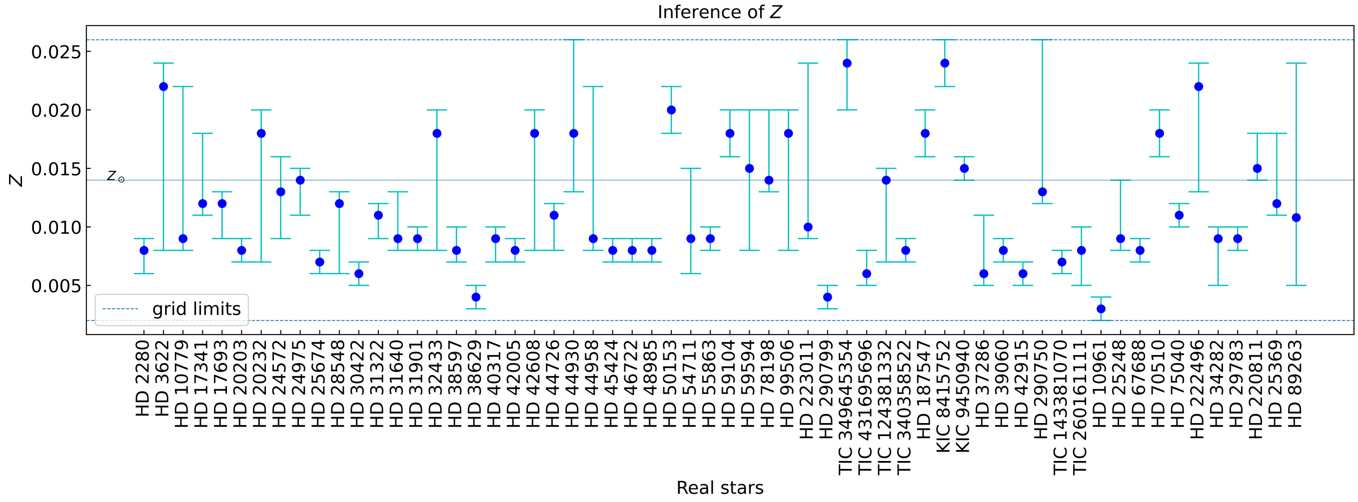

Figure A2: Z inference for 60 δ Scuti stars.

Fellowship FT210100485. TRB was also supported by the ARC, through DP210103119 and FL220100117, and by the Danish National Research Foundation (Grant DNRF106) through its funding for the Stellar Astrophysics Centre (SAC).

APPENDIX

A. INFERENCE FOR EACH STAR

We show below in Figures A1 - A3, M, Z, and τ inferences from method-4 for 60 δ Sct stars taken from Bedding et al. 2020. The x-axis contains the IDs of the stars and the y-axis displays the inferred quantities.

B. GROUP-WISE INFERENCE ACCURACY

For most of the stars, measured values of L and $T_{\rm eff}$ are available, although these quantities are not sufficient to constrain different stellar parameters. Additional independent quantities are expected to assist in constraining these structure parameters.

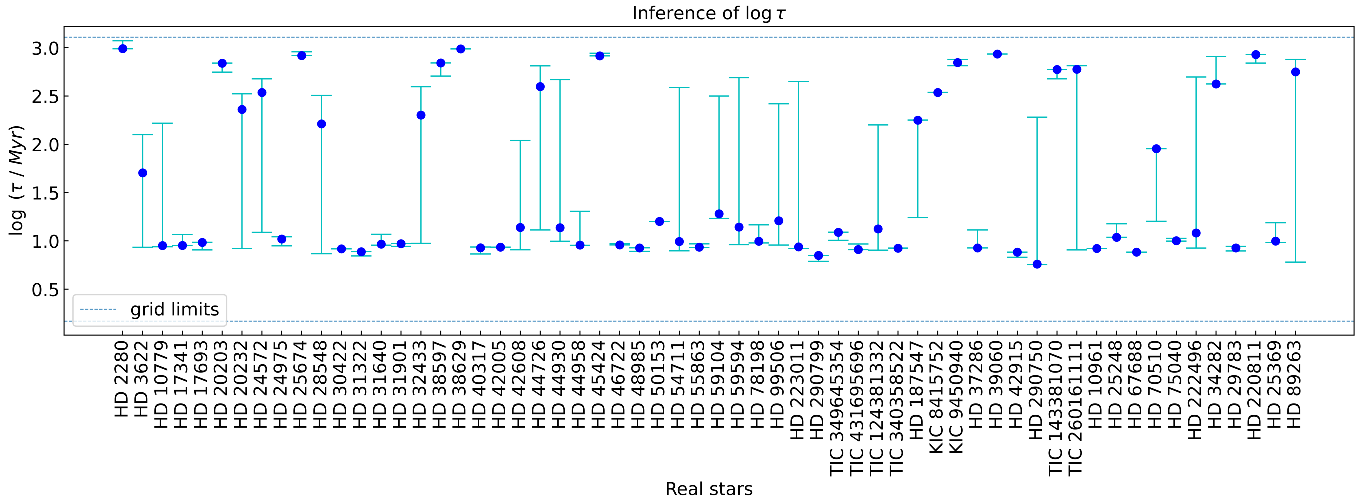

Figure A3: $\log_{10}(\tau/\mathrm{Myr})$ inference for 60 δ Scuti stars.

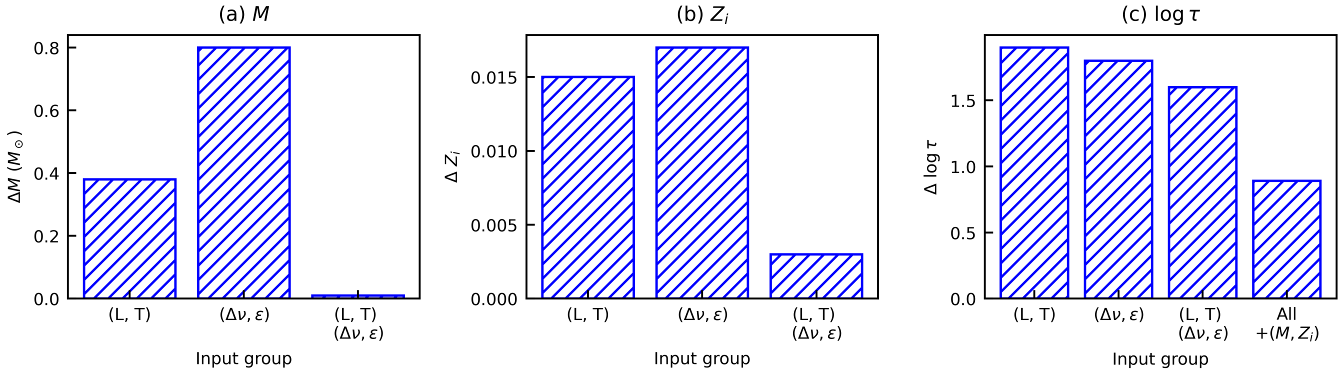

Figure B4: Maximum absolute errors between network prediction and true values, for a different group of input quantities shown in the x-axis. Figures for prediction of (a) M in $M_\odot$, (b) Z and (c) $\log\tau$, where τ is in Myr.

Asteroseismic quantities $\Delta\nu$ and ϵ are suitable for this purpose. But neither (L, T_{eff}) nor $(\Delta\nu, \epsilon)$ alone can accurately infer the stellar parameters. However, when combined, these parameters can efficiently constrain those stellar parameters.

We considered $\{L, T_{\mathrm{eff}}, R\}$ as inputs, trained the networks with these, and measured the maximum absolute errors between true values and network predictions (over the validation samples). Similarly, we took asteroseismic quantities $\{\Delta\nu, \epsilon\}$ and repeated this validation process. Finally, we combined all of them and again measured the maximum absolute error associated with the predictions. During age inference, we supplemented M, Z like quantities and also assembled the values of validation errors. We visualize all of these errors in Figure B4, which emphasizes that validation error decreases drastically when observables and asteroseismic quantities are simultaneously considered as inputs.

In Table B1, we show the Pearson-R correlation coefficients (Rodgers et al. 1988) between network predictions (over validation samples) and corresponding true values. It is a measure of inference accuracy, i.e., correlation between true and inferred values. The Pearson-R coefficient can have values between -1 and +1. The higher the R coefficient, the higher the correlation between true values and network prediction. This Pearson-R correlation coefficient between the two sets of measurements {p} and {q} is calculated by equation B1.

$$R = \frac{\sum_i (p_i - \overline{p})(q_i - \overline{q})}{\sqrt{\sum_i (p_i - \overline{p})^2 \sum_i (q_i - \overline{q})^2}} \tag{B1}$$

where, $\overline{p}, \overline{q}$ are the means of the measurements.

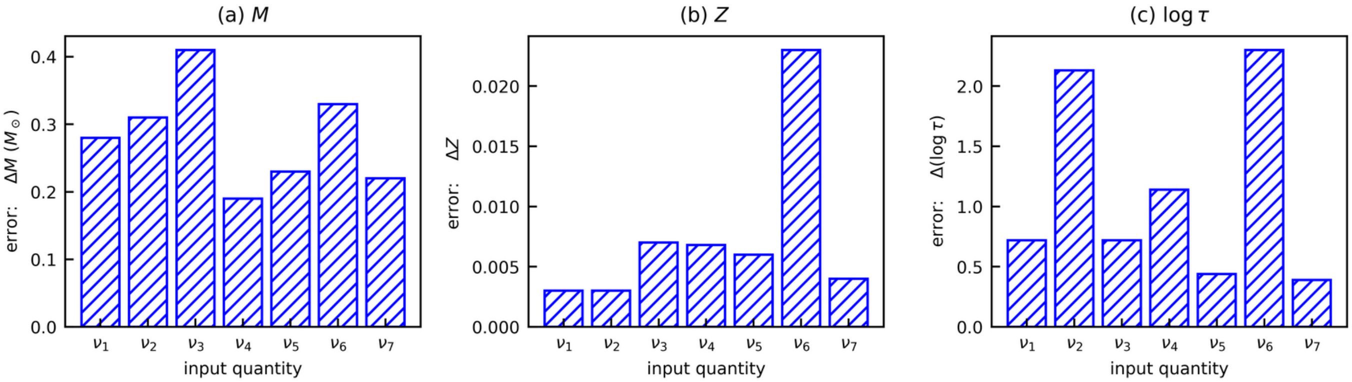

Figure C5: Importance of first 7 eigenfrequencies in the inference of (a) M, (b) Z and (c) $\log\tau$.

Adding (M, Z) as additional inputs increases the Pearson-R coefficient by only 0.01. However, from Figure B4 (c), it is evident that (M, Z) inputs actually assist in reducing the absolute error associated with $\log\tau$ prediction.

	$\{L, T_{\rm eff}\}$	$\{\Delta\nu, \epsilon\}$	$\{L, T_{\rm eff}, \Delta\nu, \epsilon\}$	$\{L, T_{\rm eff}, \Delta\nu, \epsilon, M, Z\}$
M	0.93	0.86	0.99	
$\log Z$	0.77	0.85	0.99	
$\log\tau$	0.67	0.78	0.98	0.99

Table B1: Pearson-R correlation coefficients between network predictions and true values, for different groups of input quantities.

C. FEATURE IMPORTANCE OF RADIAL MODES

In Section 5, we devised a robust method in which we considered eigenmodes ($\nu_1 - \nu_7$) as inputs and showed that almost every parameter may be learned from these eigenfrequencies. In continuation of our feature importance studies of Section 4.4, we also here intend to study the contribution strengths coming from each frequency.

We follow the same procedure as Section 4.4 to measure the feature importance. Using trained networks over the validation samples, we perturbed one of the input frequencies by 0.5% (keeping others constant) and measured the error propagating to the output layer (inference). We repeated this procedure for each parameter, assembled the inference errors and finally normalized them to get an overall idea of feature importance. We considered these errors as qualitative proxies of feature contributions.

We show the feature importance results in Figure C5. It turns out that all 7 frequencies have non-negligible importance. When inferring M, some of the frequencies have relatively low contributions. However, there is no regular trend of importance variation as we move from ν_1 to ν_7.

D. FEATURE IMPORTANCE OF DIPOLE MODES

In section 6, we showed that every physical parameter may be learnt from $\{L, T_{\rm eff}, \ell = 0, 1 \text{ modes}\}$. Hence, it immediately requires us to learn their importance, similar to what we did for earlier methods (Section 4, 5). We follow the same procedures as in the earlier Sections (4.4, 5) with 0.5% perturbation of each input quantity followed by measurements of average errors associated with the inferences. These errors are considered as qualitative proxies for the feature importance.

We show the feature contribution plot in Figure D6. The dipole and radial modes have non-negligible importance for $\log Z$ and $\log\tau$ predictions. However, this is not true for the M inference. For age inference, dipole modes have the greatest importance. Also, the contributions of radial modes cannot be ruled out. However, the radial modes are themselves not completely able to lift the age degeneracy in earlier experiments. Hence, dipole modes play an

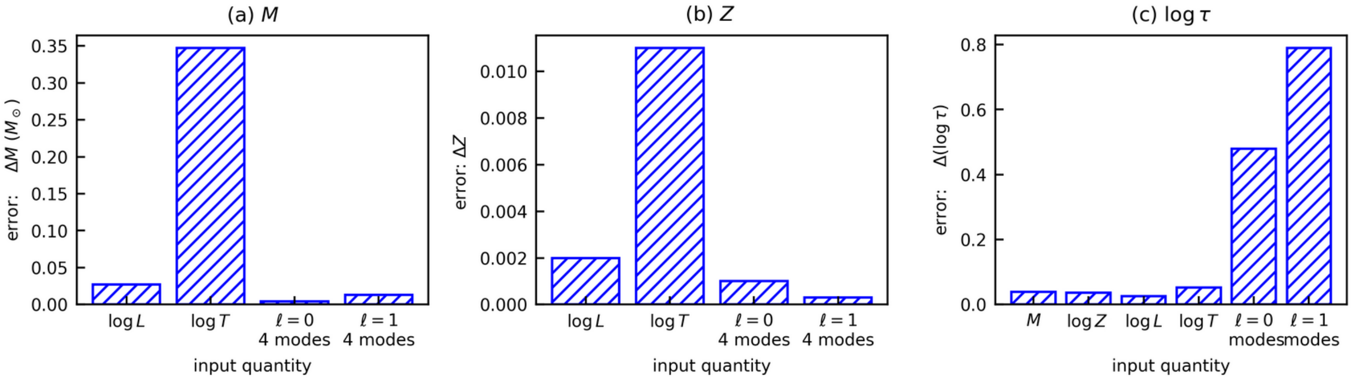

Figure D6: Qualitative contributions arising from input quantities for inferences of (a) M, (b) Z and (c) $\log \tau$.

important supplementary role in determining the stellar parameters precisely. This again emphasizes the usefulness of asteroseismology.

REFERENCES

Abadi, M., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems

Aerts, C. 2021, Reviews of Modern Physics, 93, 015001, doi: 10.1103/RevModPhys.93.015001

Avni, Y. 1976, ApJ, 210, 642, doi: 10.1086/154870

Barceló Forteza, S., Moya, A., Barrado, D., et al. 2020, A&A, 638, A59, doi: 10.1051/0004-6361/201937262

Barceló Forteza, S., Roca Cortés, T., García Hernández, A., & García, R. A. 2017, A&A, 601, A57, doi: 10.1051/0004-6361/201628675

Bedding, T. R., et al. 2020, Nature, 581, 147, doi: 10.1038/s41586-020-2226-8

Bellinger, E. P., et al. 2016, The Astrophysical Journal, 830, 31, doi: 10.3847/0004-637x/830/1/31

Borucki, W. J., et al. 2010, Science, 327, 977, doi: 10.1126/science.1185402

Bowman, D. M. 2020, Frontiers in Astronomy and Space Sciences, 7, doi: 10.3389/fspas.2020.578584

Bowman, D. M., et al. 2018, MNRAS, 476, 3169, doi: 10.1093/mnras/sty449

Canuto, V. M. 1990, A&A, 227, 282

Chaplin, W. J., et al. 2013, ARA&A, 51, 353, doi: 10.1146/annurev-astro-082812-140938

Chevalier, C. 1971, A&A, 14, 24

Chollet, F., et al. 2015, Keras, https://keras.io

Dhanpal, S., et al. 2022, The Astrophysical Journal, 928, 188, doi: 10.3847/1538-4357/ac5247

Dziembowski, W., et al. 1990, AcA, 40, 19

Eddington, A. S. 1926, The Internal Constitution of the Stars

García Hernández, A., Suárez, J. C., Moya, A., et al. 2017, Monthly Notices of the Royal Astronomical Society: Letters, 471, L140, doi: 10.1093/mnrasl/slx117

Glorot, X., et al. 2011, in Proceedings of Machine Learning Research, Vol. 15, Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (Fort Lauderdale, FL, USA: PMLR), 315–323

Goldstein, J., et al. 2020, The Astrophysical Journal, 899, 116, doi: 10.3847/1538-4357/aba748

Guzik, J. A. 2021, Frontiers in Astronomy and Space Sciences, 8, 55, doi: 10.3389/fspas.2021.653558

Hasanzadeh, A., Safari, H., & Ghasemi, H. 2021, MNRAS, 505, 1476, doi: 10.1093/mnras/stab1411

Hekker, S., & Christensen-Dalsgaard, J. 2017, A&A Rv, 25, 1, doi: 10.1007/s00159-017-0101-x

Hendriks, L., et al. 2019, Publications of the Astronomical Society of the Pacific, 131, 108001, doi: 10.1088/1538-3873/aaeeec

Henyey, L., Vardya, M. S., & Bodenheimer, P. 1965, ApJ, 142, 841, doi: 10.1086/148357

Herwig, F. 2005, Annual Review of Astronomy and Astrophysics, 43, 435, doi: 10.1146/annurev.astro.43.072103.150600

Hon, M., et al. 2017, Monthly Notices of the Royal Astronomical Society, 469, 4578, doi: 10.1093/mnras/stx1174

—. 2018, Monthly Notices of the Royal Astronomical Society, 476, 3233, doi: 10.1093/mnras/sty483

Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90, doi: 10.1109/MCSE.2007.55

Kahraman Aliçavuş, F., Handler, G., Aliçavuş, F., et al. 2022, MNRAS, 510, 1413, doi: 10.1093/mnras/stab3515

Kingma, D. P., et al. 2014, Adam: A Method for Stochastic Optimization, arXiv, doi: 10.48550/ARXIV.1412.6980

Kurtz, D. W. 2022, Annual Review of Astronomy and Astrophysics, 60, 31, doi: 10.1146/annurev-astro-052920-094232

Le Dizès, C., Rieutord, M., & Charpinet, S. 2021, A&A, 653, A26, doi: 10.1051/0004-6361/202141291

Li, G., et al. 2019, Monthly Notices of the Royal Astronomical Society, 491, 3586, doi: 10.1093/mnras/stz2906

Lightkurve Collaboration, Cardoso, J. V. d. M., Hedges, C., et al. 2018, Lightkurve: Kepler and TESS time series analysis in Python, Astrophysics Source Code Library. http://ascl.net/1812.013

McNamara, D. H., et al. 2007, The Astronomical Journal, 133, 2752, doi: 10.1086/513717

Mombarg, J. S. G., et al. 2021, A&A, 650, A58, doi: 10.1051/0004-6361/202039543

Montavon, G., Samek, W., & Müller, K.-R. 2018, Digital Signal Processing, 73, 1, doi: https://doi.org/10.1016/j.dsp.2017.10.011

Mosteller, F., et al. 1968, in Handbook of Social Psychology, Vol. 2 (Addison-Wesley)

Murphy, S. J., Joyce, M., Bedding, T. R., White, T. R., & Kama, M. 2021a, Monthly Notices of the Royal Astronomical Society, 502, 1633, doi: 10.1093/mnras/stab144

Murphy, S. J., Paunzen, E., Bedding, T. R., Walczak, P., & Huber, D. 2020, MNRAS, 495, 1888, doi: 10.1093/mnras/staa1271

Murphy, S. J., et al. 2021b, Monthly Notices of the Royal Astronomical Society, 502, 1633, doi: 10.1093/mnras/stab144

—. 2022, Monthly Notices of the Royal Astronomical Society, 511, 5718, doi: 10.1093/mnras/stac240

Oliphant, T. E. 2007, Computing in Science & Engineering, 9, 10, doi: 10.1109/MCSE.2007.58

Pamos Ortega, D., García Hernández, A., Suárez, J. C., et al. 2022, Monthly Notices of the Royal Astronomical Society, 513, 374, doi: 10.1093/mnras/stac864

Paparó, M., et al. 2013, A&A, 557, A27, doi: 10.1051/0004-6361/201321792

Paparó, M. 2019, Frontiers in Astronomy and Space Sciences, 6, doi: 10.3389/fspas.2019.00026

Paxton, B., et al. 2010, The Astrophysical Journal Supplement Series, 192, 3, doi: 10.1088/0067-0049/192/1/3

—. 2013, The Astrophysical Journal Supplement Series, 208, 4, doi: 10.1088/0067-0049/208/1/4

—. 2015, The Astrophysical Journal Supplement Series, 220, 15, doi: 10.1088/0067-0049/220/1/15

—. 2018, The Astrophysical Journal Supplement Series, 234, 34, doi: 10.3847/1538-4365/aaa5a8

Petersen, J. O., & Christensen-Dalsgaard, J. 1996, A&A, 312, 463

Reese, D., Lignières, F., & Rieutord, M. 2006, A&A, 455, 621, doi: 10.1051/0004-6361:20065269

Ricker, G. R., et al. 2014, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003, doi: 10.1117/1.JATIS.1.1.014003

Rodgers, J. L., et al. 1988, The American Statistician, 42, 59, doi: 10.1080/00031305.1988.10475524

Scutt, O. J., Murphy, S. J., Nielsen, M. B., et al. 2023. https://arxiv.org/abs/2302.11025

Soderblom, D. R. 2010, Annual Review of Astronomy and Astrophysics, 48, 581, doi: 10.1146/annurev-astro-081309-130806

Steindl, T., Zwintz, K., & Müllner, M. 2022, A&A, 664, A32, doi: 10.1051/0004-6361/202243242

Suárez, J. C., García Hernández, A., Moya, A., et al. 2014a, A&A, 563, A7, doi: 10.1051/0004-6361/201322270

Suárez, J. C., et al. 2014b, A&A, 563, A7, doi: 10.1051/0004-6361/201322270

Tange, O. 2011, ;login: The USENIX Magazine, 36, 42, doi: 10.5281/zenodo.16303

Townsend, R. H. D., et al. 2013, Monthly Notices of the Royal Astronomical Society, 435, 3406, doi: 10.1093/mnras/stt1533

—. 2017, Monthly Notices of the Royal Astronomical Society, 475, 879, doi: 10.1093/mnras/stx3142

Uytterhoeven, K., et al. 2011, A&A, 534, A125, doi: 10.1051/0004-6361/201117368

van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, Computing in Science & Engineering, 13, 22, doi: 10.1109/MCSE.2011.37

Verma, K., et al. 2016, Monthly Notices of the Royal Astronomical Society, 461, 4206, doi: 10.1093/mnras/stw1621

Vrard, M., et al. 2016, A&A, 588, A87, doi: 10.1051/0004-6361/201527259

Zwintz, K., et al. 2011, A&A, 533, A133, doi: 10.1051/0004-6361/201117272