The Astronomical Journal, 171:313 (9pp), 2026 May https://doi.org/10.3847/1538-3881/ae4bd8

Ensemble Machine Learning Approach to Estimate the Asteroseismic Indices for δ Scuti Stars Observed by TESS

Akram Bolouki[1], Amir Hasanzadeh[2], and Hossein Safari[1,3]
[1] Department of Physics, Faculty of Science, University of Zanjan, University Blvd., Zanjan, 45371-38791, Zanjan, Iran; safari@znu.ac.ir
[2] Centre for Fusion, Space and Astrophysics, Department of Physics, University of Warwick, Coventry CV4 7AL, UK
[3] Observatory, Faculty of Science, University of Zanjan, University Blvd., Zanjan, 45371-38791, Zanjan, Iran
Received 2025 July 4; revised 2026 February 23; accepted 2026 February 23; published 2026 April 24

Abstract

We develop an ensemble machine learning framework to estimate key asteroseismic parameters—namely the frequency of the highest peak ($\nu(A_{\max})$), the frequency of maximum oscillation power ($\nu_{\max}$), and the large frequency separation ($\Delta\nu$)—directly from TESS light curves of δ Scuti stars (δ Sct stars). For each light curve, we extract a set of features comprising statistical moments, Principal Component Analysis, autocorrelation function, spectral features from the fast Fourier transform, and multiscale features from the discrete wavelet transform. These features are used to train a stacked regressor model composed of Random Forest, Gradient Boosting, and Ridge Regressors. We train and evaluate the model using 583 δ Sct stars randomly selected from a total of 643 samples, repeatedly applying random 80/20 splits across 100 iterations. The model achieves high predictive accuracy with R^2 scores exceeding 0.77 for all asteroseismic quantities. We validate generalizability by applying the trained model to the remaining 60 δ Sct stars, not seen during training. The predicted values show strong agreement with traditional asteroseismic measurements, confirming the effectiveness of this framework for large-scale, automated asteroseismic analysis. Furthermore, the proposed framework allows for the estimation of asteroseismic indices across new 251 δ Sct stars.

Unified Astronomy Thesaurus concepts: Asteroseismology (73); Stellar oscillations (1617); Delta Scuti variable stars (370)

Materials only available in the online version of record: data behind figures

1. Introduction

Asteroseismology provides a powerful diagnostic tool for probing the internal structures of stars by analyzing their natural oscillation modes (C. Aerts et al. 2010; C. Aerts 2021; D. W. Kurtz 2022). Among various classes of pulsating variables, δ Sct stars are of particular interest due to their hybrid pulsation modes and dense frequency spectra (L. A. Balona 2014; P. A. Bradley et al. 2015; D. M. Bowman 2017; J. P. Sánchez Arias et al. 2017; J. A. Guzik 2021; S. Barceló Forteza et al. 2024). These intermediate-mass stars, typically spanning 1.5–2.5 solar masses (L. A. Balona & W. A. Dziembowski 2011; V. Antoci et al. 2019; F. Vasigh et al. 2024), exhibit both radial and nonradial oscillations (R. Garrido & E. Rodriguez 1996; M. A. Dupret et al. 2005; C. Lv et al. 2023). The driving mechanism behind these pulsations is the κ-mechanism, which operates within the partial ionization zones of helium (M. A. Dupret et al. 2004a, 2004b; D. R. Xiong et al. 2016). The internal structure of δ Sct stars is further complicated by the presence of convective cores, which introduce processes such as convective overshooting (J. Daszyńska-Daszkiewicz et al. 2024), chemical mixing (S. Turcotte et al. 2000), and angular momentum redistribution (D. W. Kurtz et al. 2014). These factors substantially influence the observed oscillation spectra (C. C. Lovekin & J. A. Guzik 2017; V. Dornan & C. C. Lovekin 2022). Moreover, stellar rotation-often moderate to rapid in δ Sct stars adds to the complexity by introducing mode coupling and frequency splitting, thereby complicating mode identification (M. J. Goupil et al. 2000; M. Breger et al. 2009; S. Barceló Forteza et al. 2017; G. M. Mirouh et al. 2019; T. R. Bedding et al. 2020). Asteroseismic indices such as the frequency corresponding to the highest peak in the power spectrum, the frequency of maximum oscillation power, and the large frequency separation serve as essential diagnostics for determining stellar properties and evolutionary states (H. Kjeldsen & T. R. Bedding 1995; D. M. Bowman & D. W. Kurtz 2018). Among these, $\Delta\nu$ —the spacing between consecutive overtone modes of the same spherical degree—is particularly sensitive to the stellar mean density and has become a cornerstone in ensemble asteroseismology (B. Mosser et al. 2012; A. García Hernández et al. 2017). However, the determination of $\Delta\nu$ in δ Sct stars poses significant challenges due to their rapid rotation, presence of mixed modes (M. Templeton et al. 2001), and mode interactions (S. Barceló Forteza et al. 2015; A. Ramón-Ballesta et al. 2021a), all of which disrupt the regular frequency spacing (B. Mosser et al. 2013). Recent studies (A. Hasanzadeh et al. 2021; K. Gootkin et al. 2024; C. Lv et al. 2024; A.-Y. Zhou 2025) have significantly expanded the database of asteroseismic parameters for δ Sct stars, particularly those observed by the TESS mission (G. R. Ricker et al. 2015), thereby facilitating population-level analyses. To extract asteroseismic indices from such extensive observational datasets, various methodologies have been proposed, each offering distinct advantages and limitations (B. Mosser & T. Appourchaux 2009; L. S. Viani et al. 2019; S. J. Murphy et al. 2020). Nevertheless, the accurate determination of these indices remains complex due to observational

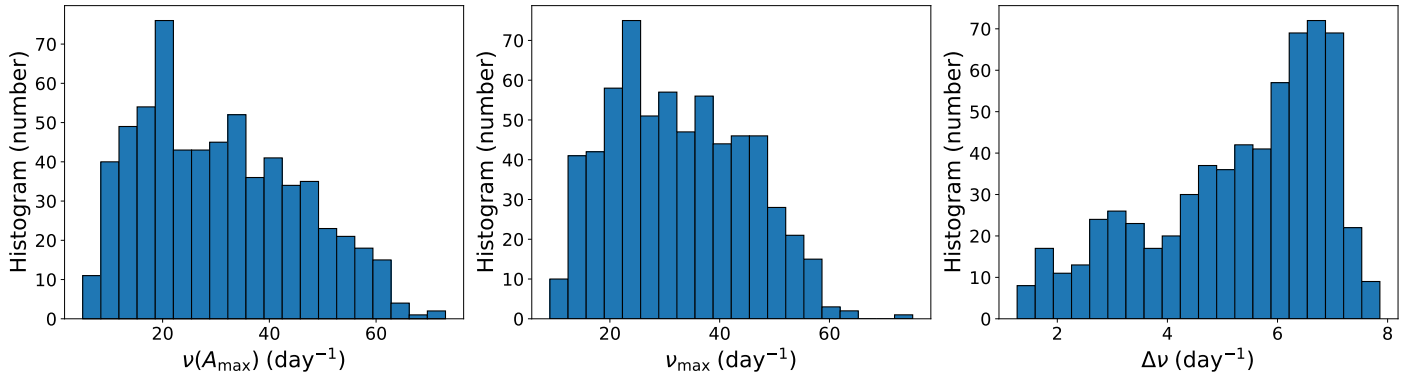

Figure 1. Histogram (number) for $\nu(A_{max})$ (left panel), ν_{max} (middle panel), and $\Delta\nu$ (right panel) of 643 δ Sct stars (see supplementary electronic table). (The data used to create this figure are available in the online article.)

limitations and the intrinsic physical intricacies of δ Sct stars (S. K. Panda et al. 2024b). The processing and analysis of large volumes of light curve data also require substantial computational resources and expertize (S. J. Murphy et al. 2023; M. B. Nielsen et al. 2023; D. Hey & C. Aerts 2024). With the exponential growth of time-series data, machine learning (ML) techniques have emerged as powerful tools for automating and enhancing the extraction of asteroseismic and stellar parameters (H. Rauer et al. 2014; S. K. Panda et al. 2024a; S. Abbasvand et al. 2025; N. Kumar et al. 2025). In particular, ensemble learning methods offer improved accuracy and robustness by combining the predictive strengths of multiple base models. Previous studies have shown the utility of data-driven approaches in astronomy using techniques such as Random Forests (L. Breiman 2001), Gradient Boosting (J. H. Friedman 2001), and regularized linear models, including Ridge Regression (A. E. Hoerl & R. W. Kennard 1970). In the present study, we utilize a stacked ensemble ML framework to predict asteroseismic indices for 643 δ Sct stars based on features (including statistical moments, Principal Component Analysis, autocorrelation function, fast Fourier transform, and discrete wavelet transform) extracted from their TESS light curves. First, we construct the ensemble regression model using a subset of 583 δ Sct stars, randomly selected from a total of 643 targets. In each iteration, we split this subset into 80% for training and 20% for testing to evaluate the model's predictive performance. Second, we assess the generalization capability of the trained model using the remaining 60 δ Sct stars, which were entirely excluded from the training process. We predict the asteroseismic indices for these 60 stars and compare the results with values obtained from traditional asteroseismic methods. We repeat this two-step procedure (model training and generalization) more than 10 times, each with a different random selection of the 583 training/testing stars, to verify the stability and reliability of both the learning and generalization performance of the ensemble regressor. Third, we apply the method to determine the asteroseismic indices for 251 δ Sct stars. Section 2 describes the data utilized in this study. Section 3 details the feature extraction algorithm and the ensemble ML methodology. Sections 4 and 5 present the results and their implications. Finally, Section 6 concludes with a summary of our findings.

2. Data

We compile light curves and asteroseismic indices—including ($\nu(A_{max})$), (ν_{max}), and ($\Delta\nu$)—for a sample of 643 δ Sct stars analyzed by C. Lv et al. (2024), A. Hasanzadeh et al. (2021), and S. J. Murphy et al. (2024). While A. Hasanzadeh et al. (2021) reported asteroseismic parameters for 438 δ Sct stars, C. Lv et al. (2024) provided updated indices for 145 stars, and S. J. Murphy et al. (2024) calculated asteroseismic parameters for 60 stars. From an initial pool of 677 δ Sct stars, we select those with TESS short-cadence (2 minutes) light curves available from the MAST archive, containing at least 7000 data points per sector. The light curves are calibrated using the Pre-search Data Conditioning Simple Aperture Photometry method. We further constrain the sample to stars for which all three asteroseismic indices are available. Following the application of these selection criteria, the curated dataset includes 643 δ Sct stars. The supplementary electronic file accompanying Figure 1, provides complete information on these stars. We assemble a dataset of 251 δ Sct stars C. Lv et al. (2024) and S. J. Murphy et al. (2024) with available short-cadence TESS light curves, for which a set of asteroseismic indices have not yet been published. An additional requirement for predicting the true asteroseismic indices of these stars is that each light curve must contain at least 7000 data points across a minimum of three sectors. The electronic table accompanying Figure 7 contains the TESS identifiers for 251 δ Sct stars.

3. Methods

3.1. Features Extraction

To analyze variability and oscillatory patterns in stellar flux, we pick up the TESS light curves for δ Sct stars. From each light curve, we extract a set of statistical and frequency-domain features that describe both the distribution and dynamic behavior of stellar brightness over time. We compute statistical descriptors such as the mean, standard deviation, median, minimum, maximum, first and third quartiles (25th and 75th percentiles), and the interquartile range. To describe the shape and symmetry of the flux distribution, we also calculate skewness and kurtosis. To mitigate redundancy among highly correlated flux values of stellar light curve, we apply the Principal Component Analysis (PCA) to extract the first ten principal components are retained as additional features. PCA projects the original time-series data onto an orthogonal basis ordered by explained variance, allowing the dominant modes of variability to be preserved while suppressing noise (J. Chang et al. 2018).

To evaluate temporal correlation in the stellar flux, we calculate the first twelve values of the autocorrelation function

(ACF). The ACF features, particularly probe time-lagged correlations at delays that are closer to the dominant quasi-periodic pulsation timescales. We use the fast Fourier transform (FFT) to extract frequency-domain characteristics and retain 100 highest peaks Fourier coefficients to represent dominant periodic components in each light curve. We also identify the index of the dominant frequency component, corresponding to the largest Fourier amplitude (excluding the zero-frequency term), to capture the main oscillatory mode of each light curve. Additionally, we apply the discrete wavelet transform (DWT) using the Daubechies-4 mother wavelet up to level three (G. R. Lee et al. 2019). We discard any samples with missing values in either the input features or the target asteroseismic indices to preserve data integrity and ensure robust model training. We normalize the extracted features of δ Sct stars light curves based on a reference δ Sct star (TIC 28679106, Sector 14, which is the first target in our database) to make them suitable for ML algorithms.

3.2. Resampling of Minority Features

Some δ Sct stars observed by TESS have light curves from many sectors, while others have only one or a few in MAST. Additionally, certain ranges of asteroseismic indices are underrepresented compared to other δ Sct stars in the dataset (Section 4). This scarcity of light curves in some sectors or low-population ranges of asteroseismic indices creates an imbalance in the features, complicating the estimation of these indices. To mitigate this imbalance and ensure that extracted features contribute more evenly to the ML process, we implement a resampling algorithm on the fluxes of δ Sct stars light curves, creating synthetic flux values by sampling from a normal distribution derived from the original light curve data.

3.3. Stacking-based Ensemble Regression Framework

To estimate the asteroseismic indices from the extracted features, we use a stacked ensemble learning framework with three base regressors of Random Forest, Gradient Boosting, and Ridge Regression. The Random Forest Regressor combines multiple decision trees and averages their outputs to reduce variance and overfitting (L. Breiman 2001). The Gradient Boosting Regressor builds models sequentially, with each one correcting the errors of the previous model to improve accuracy (J. H. Friedman 2001). The Ridge Regressor is a regularized linear model that handles multicollinearity and reduces overfitting by applying an L2 penalty to the coefficients (A. E. Hoerl & R. W. Kennard 1970). We integrate these base models using a Stacking Regressor, which uses their predictions as inputs to a final meta regressor. This meta model combines the strengths of the base regressors to improve generalization and reduce prediction error. To predict each asteroseismic parameter index, a hyperparameter tuning is carried out using a randomized search strategy applied to two ensemble models of the Random Forest and Gradient Boosting. For each of the 100 training iterations, the algorithm explores a set of candidate hyperparameter values—such as the number of trees, maximum depth, learning rate, and subsampling fraction—by sampling them randomly from predefined ranges. Each candidate is evaluated through cross-validation, and the combination that yields the best predictive performance is selected as the optimal configuration for that iteration. These tuned models are then combined in a stacking framework, ensuring that the predictions are based on models that have been systematically optimized rather than relying on arbitrary default settings. Our ensemble strategy is consistent with current best practices in modern ML (Z.-H. Zhou 2025).

4. Result

We employ an ensemble ML approach to estimate key asteroseismic parameters of 643 δ Sct stars from TESS observations. We extract a compact set of features from each light curve with at least 7000 data points. This transformation enables the model to capture the distributional characteristics of each light curve without relying on the raw time-series data. These features provide a detailed representation of each light curve's statistical structure, periodicity, and temporal variability for all 643 δ Sct stars. Figure 1 shows the histogram for asteroseismic indices of δ Sct stars. As shown in the figure, $\nu(A_{\max})$, $\nu_{\max}$, and $\Delta\nu$ are ranging 5.02 to 73.01, 9.07 to 75.17, and 1.27 to 7.86, respectively. Our analysis of the 643 targets in the training and testing dataset reveals that some stars exhibit relatively lower surface gravity, suggesting a more advanced evolutionary stage among these δ Sct stars. Consequently, they are expected to show slightly reduced $\Delta\nu$ values (D. M. Bowman & D. W. Kurtz 2018; A. Hasanzadeh et al. 2021). By contrast, S. J. Murphy et al. (2023) reported $\Delta\nu$ values greater than 4 for young δ Sct stars based on MESA evolutionary models. As illustrated in the figure, certain ranges of asteroseismic indices, such as low values of $\Delta\nu$, occur less frequently than others. To address this imbalance in the distribution of features for ML, we perform resampling of the light curves.

Total of 643 targets, we randomly select 583 δ Sct stars for training and testing (Section 4.1). The remaining 60 δ Sct stars (Section 4.2) are used in the generalization dataset. In addition, we estimate the asteroseismic indices for the 251 targets described in Section 4.4.

4.1. Prediction Results

We train and evaluate the ensemble regressor using a sample of 583 δ Sct stars. In each iteration, we randomly divide the dataset into training (80%) and testing (20%) subsets. To ensure robustness and account for variability in the sampling process, we repeat this random division and model evaluation 100 times. This iterative approach allows us to assess the stability and generalizability of the ensemble regressor across different data splits. Our analysis shows that each δ Sct star is included in the testing subset at least three times throughout the iterations.

Figure 2 presents the comparison between the measured and predicted values of the three principal asteroseismic indices—$\nu(A_{\max})$, $\nu_{\max}$, and $\Delta\nu$—for a sample of 583 δ Sct stars. Each column shows the median predicted values with their associated upper and lower uncertainties, alongside the corresponding fractional errors plotted against the measured quantities. For each unique TIC, we aggregate all corresponding prediction samples, computed the median predicted values, and derived the 16th–84th percentile range to quantify the asymmetric upper and lower uncertainties relative to the median. This asymmetry between the upper and lower uncertainties likely arises from the non-normal distribution of the asteroseismic indices (Figure 1), a behavior that is expected for stellar and asteroseismic parameters in oscillating

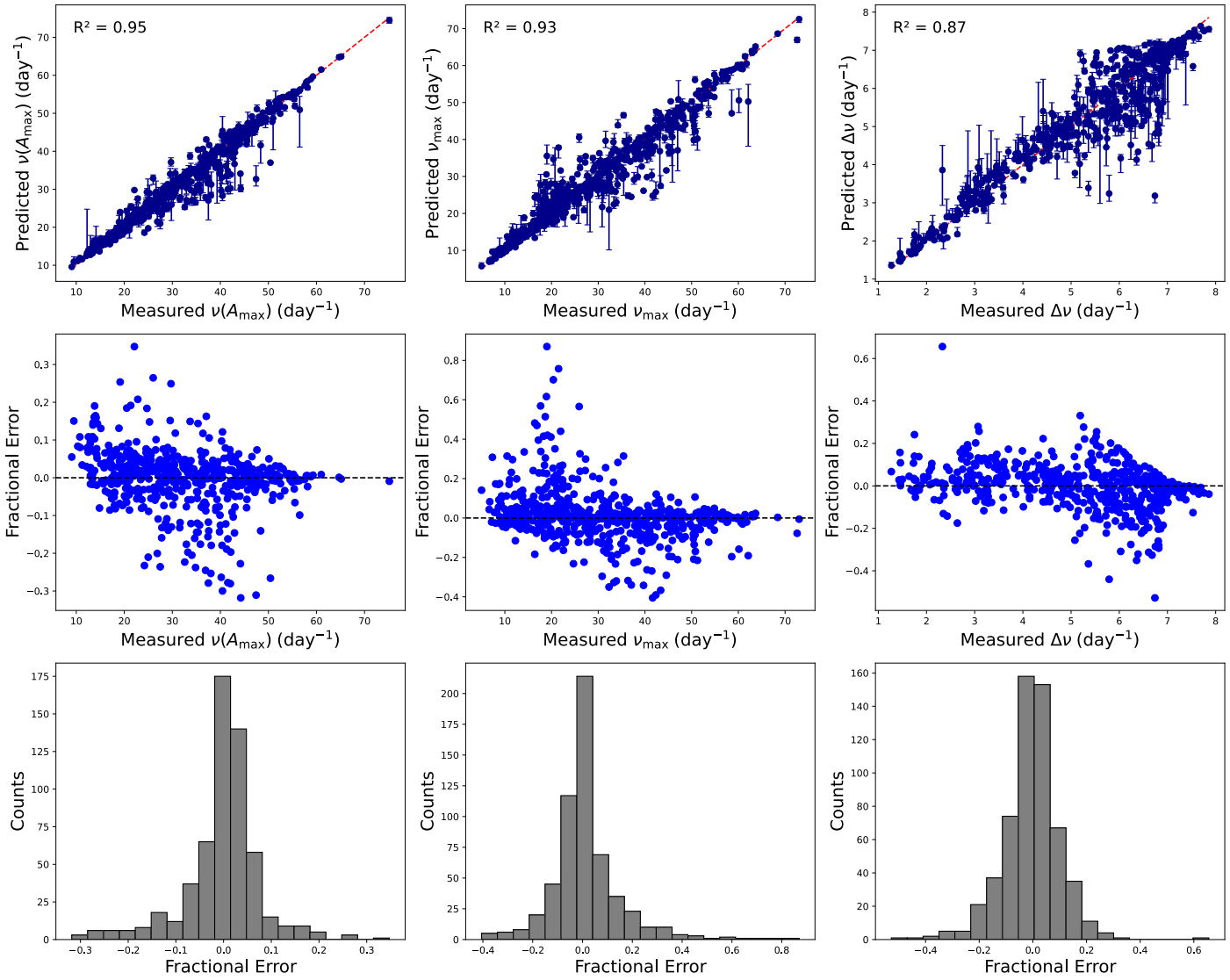

Figure 2. The predicted median with associated upper and lower uncertainties vs. the measured values for 583 δ Sct stars (top row), fractional errors (middle row), and distribution of fractional error (bottom row) obtained by ensemble ML for $\nu(A_{\max})$ (left panel), $\nu_{\max}$ (middle panel), and $\Delta\nu$ (right panel; see supplementary electronic table hosted by Figure 1). The red dashed line ($y = x$) shows the ideal prediction line in the top row, whereas the black dashed line marks the zero fractional error reference in the bottom row.

stars (S. Abbasvand et al. 2025). For $\nu(A_{\max})$, the ensemble model achieves an R^2 of 0.95, demonstrating excellent agreement with the observed values. Similarly, for $\nu_{\max}$, the performance remains strong with $R^2 = 0.93$. In the case of $\Delta\nu$, the model attains $R^2 = 0.87$, indicating a slightly lower but still reliable predictive accuracy. We find that most fractional errors are below 0.2, indicating relatively small errors in estimating asteroseismic indices. Larger errors, while present, occur less frequently.

To evaluate the relative contribution of different input features in predicting the asteroseismic parameters, we employ a ML approach based on the Random Forest and Gradient Boosting Regressors. After training the model on the scaled feature set, the feature importance for each target is derived from the Random Forest, which quantifies how much each feature reduces the prediction error across the ensemble of decision trees. In practice, features that lead to larger reductions in variance when used for splitting are assigned higher importance values. By ranking these importance scores, we identified the most influential physical and statistical descriptors of the stellar time series for each asteroseismic parameter. Figure 3 highlights the top 20 contributing features for the three asteroseismic indices obtained by Random Forest Regressor. The results show that the ACF features consistently rank among the most important predictors, reflecting their strong diagnostic power for oscillatory signals. FFT and DWT-based features occupy the next level of importance, while statistical descriptors of the time series also contribute significantly within the top 20 features. The statistical moments, like skewness and kurtosis, also contribute to the top 20 features. A comparable feature importance analysis was performed for the Gradient Boosting Regressor, but the results are not shown here. Please note that Ridge Regression, serving as the meta model in our stacking approach, does not yield feature importance scores, since it is a linear model that operates on the outputs of the base models rather than on the original input features directly.

Figure 4 presents a sample of (one of 100 iterations) the learning curve for ensemble model used to derive the value of $\nu(A_{\max})$, $\nu_{\max}$, and $\Delta\nu$ for δ Sct stars. Both the training and

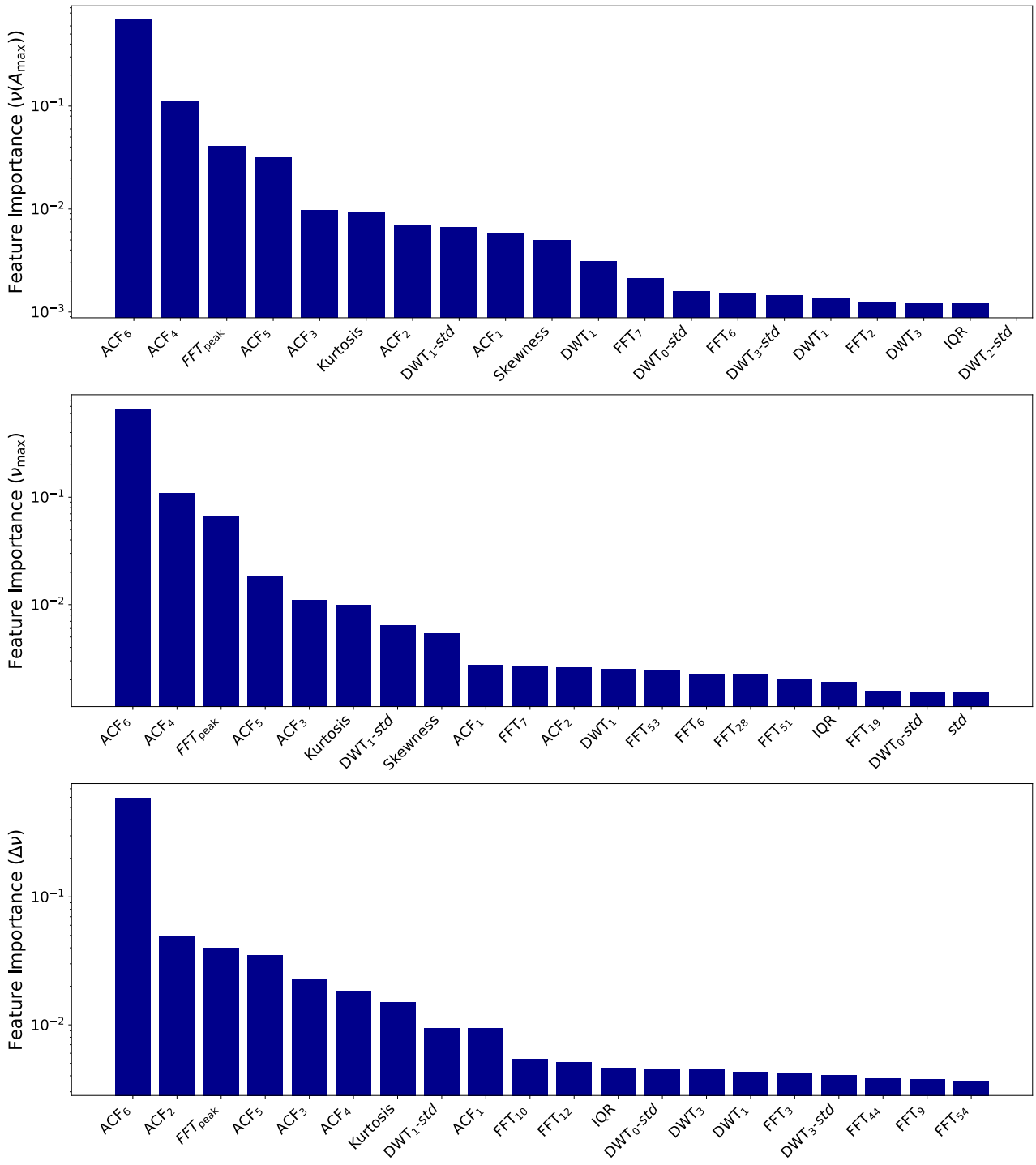

Figure 3. The 20 top-ranked features, including ACF, Wavelet (DWT), FFT, and statistical moment-based parameters, for $\nu(A_{max})$ (top panel), ν_{max} (middle panel), and $\Delta\nu$ (bottom panel) obtained during 100 training/testing iterations. Feature subscripts indicate their position within each feature set. For example, ACF$_6$ is the sixth of twelve ACF features and shows the highest mean importance over 100 Random Forest training and testing runs.

validation sets exhibit initially substantial R^2 values. As the training dataset size increases from 10% to 100% in increments of 10%, the R^2 values for validation gradually increasing to the higher values for predicting asteroseismic indices. This trend for learning curve demonstrates improved model accuracy and stability, indicating that the ensemble model effectively converges in estimating the value of asteroseismic indices. This pattern in the training and validation performance demonstrates the effectiveness of the hyperparameter tuning throughout the learning process. We likewise observe consistent trends in the learning curves across all training/testing split iterations.

4.2. Generalization to Unseen Data

We evaluate the generalizability of our ensemble model by applying it to an independent set of 60 δ Sct stars that are not used during training (Section 4.1). These 60 δ Sct stars are

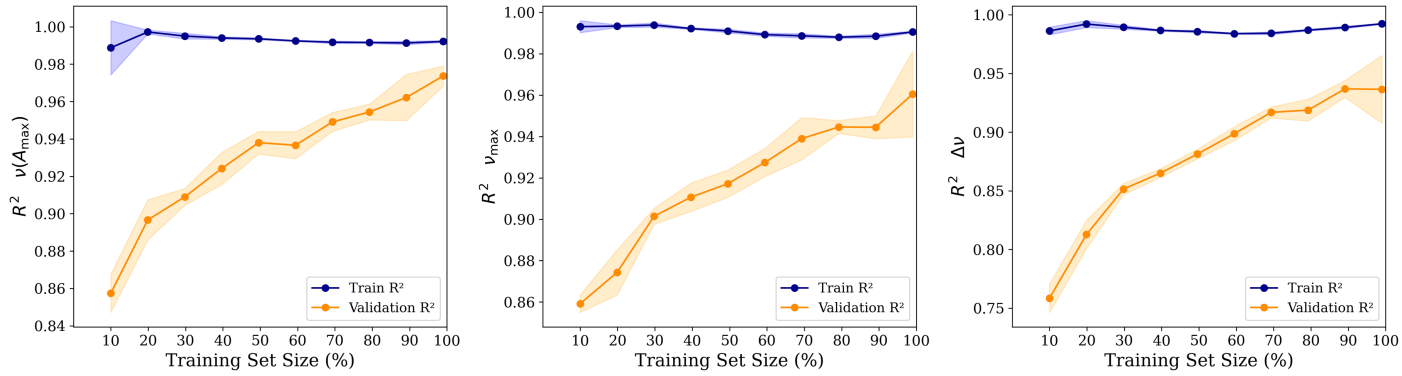

Figure 4. A sample of learning curve including validation R^2 and train R^2 for $\nu(A_{max})$ (left panel), ν_{max} (middle panel), and $\Delta\nu$ (right panel) estimation in δ Sct stars using the ensemble model for a training/testing process. Increasing R^2 indicates convergence and effective generalization.

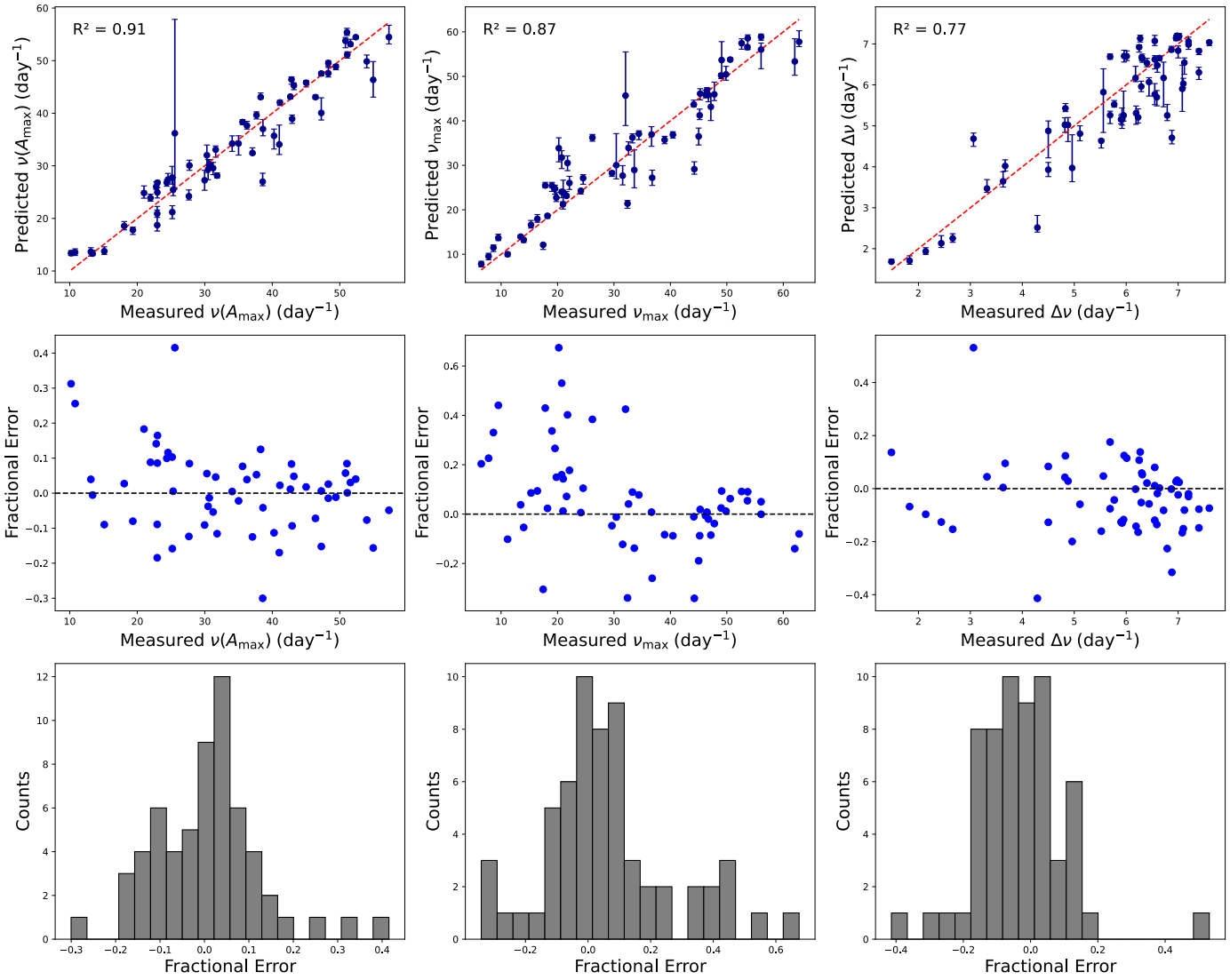

Figure 5. The predicted median with associated upper and lower uncertainties vs. the measured values for 60 δ Sct stars (top row), fractional errors (middle row), and distribution of fractional error (bottom row) obtained by ensemble ML for $\nu(A_{max})$ (left panel), ν_{max} (middle panel), and $\Delta\nu$ (right panel; see supplementary electronic table hosted by Figure 1). The red dashed line ($y = x$) shows the ideal prediction line in the top row, whereas the black dashed line marks the zero fractional error reference in the bottom row.

marked with an asterisk ("*") in the supplementary electronic table (see Figure 1). The measured and predicted values, along with their associated uncertainties, for the 583 stars (Figures 1 and 2)in the training-testing dataset and the 60 stars (Figure 5) in the generalization dataset are also listed in the supplementary electronic table accompanying Figure 1. For each of these stars, the asteroseismic parameters are estimated over 100 iterations of the ensemble regressor.

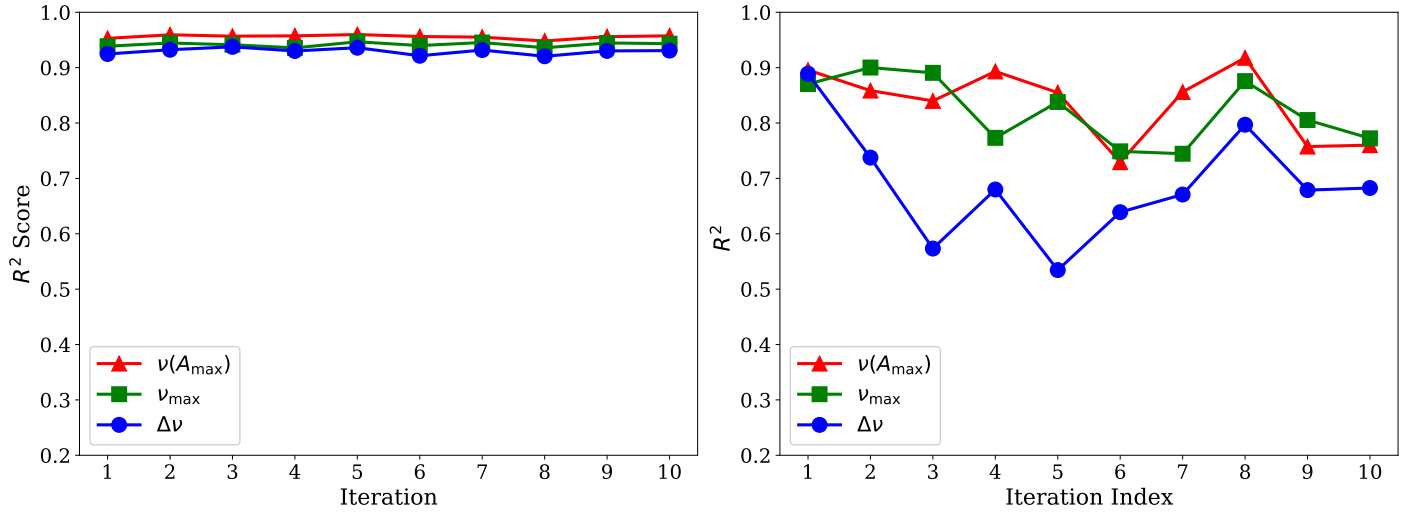

Figure 6. The R^2 score for predicted $\nu(A_{max})$ (red triangle), ν_{max} (green square), and $\Delta\nu$ (blue circle) for test (left panel) and generalization (right panel) samples across 10 iterations of random selection. The results are based on 583 and 60 stars from a total of 643 δ Sct stars.

Figure 5 presents the predicted values (median with upper and lower uncertainties) of $\nu(A_{max})$, ν_{max}, and $\Delta\nu$ alongside their observed values and corresponding fractional errors for 60 δ Sct stars. As illustrated, the predictions show strong agreement with the measured quantities. The resulting R^2 scores for $\nu(A_{max})$, ν_{max}, and $\Delta\nu$ are 0.91, 0.87, and 0.77, respectively, indicating robust model performance. Moreover, the fractional errors for these 60 δ Sct stars display similar patterns to those observed in the training/testing phase (remaining symmetrically distributed around zero with comparable magnitudes). This similarity between the training/testing and generalization fractional error trends reinforces the model's stability.

4.3. Random Training and Generalization

To evaluate the performance and generalization capability of the ensemble regressor under random sampling, we repeat the training and prediction process over 10 iterations. In each iteration, we randomly select 583 out of 643 δ Sct stars for training and testing, applying an 80/20 split. The remaining 60 δ Sct stars, which the model does not see during training, are used exclusively for testing the generalization performance. For every iteration, the ensemble model is trained on the 80% subset and evaluated on the 20% test set. Simultaneously, it predicts the asteroseismic parameters—$\nu(A_{max})$, ν_{max}, and $\Delta\nu$—for the 60 unseen stars. Figure 6 presents the R^2 scores for the test and generalizations samples across all 10 iterations. As shown in the figure, the R^2 scores for both the test stars and the unseen targets remain consistently high with minimal variation across iterations. This stability confirms the robustness and reliability of the ensemble regressor in predicting asteroseismic indices.

4.4. Asteroseismic Indices for 251 δ Sct stars

We apply the stacked ML framework to estimate the asteroseismic indices for 251 δ Sct stars. The method requires that each target have observations in at least three TESS sectors, with a minimum of 7000 data points per sector, to ensure that the asteroseismic indices and their associated uncertainties can be reliably computed using the ML approach. To further guarantee the validity of the analysis, we examine the power spectrum of each target to confirm the presence of multiple pulsation modes, which are necessary for accurately estimating parameters such as ν_{max} and $\Delta\nu$. Processing a light curve through the algorithm yields a set of asteroseismic indices that characterize the star's oscillation properties. However, for some δ Sct stars, particularly those with high-amplitude pulsations, these results may be less reliable, as the algorithm may not fully capture the extreme oscillation behavior. Therefore, a manual or automated procedure to verify the presence of multiple frequencies in the power spectrum should be performed prior to determining the asteroseismic indices with this ML method. Figure 7 shows the distributions of the predicted asteroseismic indices for these stars. The results demonstrate that the estimated values of $\nu(A_{max})$, ν_{max}, and $\Delta\nu$ generally lie within the expected ranges for δ Sct stars. The supplementary electronic file (see Figure 7) provides the predicted asteroseismic indices for all 251 stars, together with their corresponding upper and lower uncertainties (Figure 7).

5. Discussion

δ Sct stars, which have intermediate masses and rotate rapidly, display complex pulsation patterns that make traditional asteroseismic analysis particularly challenging (D. M. Bowman 2017; M. Lares-Martiz 2022; J. Daszyńska-Daszkiewicz et al. 2024). One of the most important parameters for characterizing these stars is the large frequency separation ($\Delta\nu$), which provides direct insight into their mean stellar density (J. C. Suárez et al. 2014; T. R. Bedding et al. 2020; A. Hasanzadeh et al. 2021; C. Lv et al. 2024). However, conventional methods, such as autocorrelation of the power spectrum, often fail to robustly identify $\Delta\nu$ due to the effects of rotational splitting, amplitude modulation, mixed modes, and mode crowding (J. C. Suárez et al. 2006; D. M. Bowman et al. 2016; X. Chen & Y. Li 2017; A. Ramón-Ballesta et al. 2021b). Similarly, determining other key asteroseismic parameters, such as the frequency of maximum oscillation power (ν_{max}) and the frequency of maximum amplitude ($\nu(A_{max})$), is complicated by the densely packed and often overlapping frequency modes in δ Sct stars (T. R. Bedding et al. 2023; S. J. Murphy et al. 2024; S. K. Panda et al. 2024b). These complexities result in significant uncertainty when

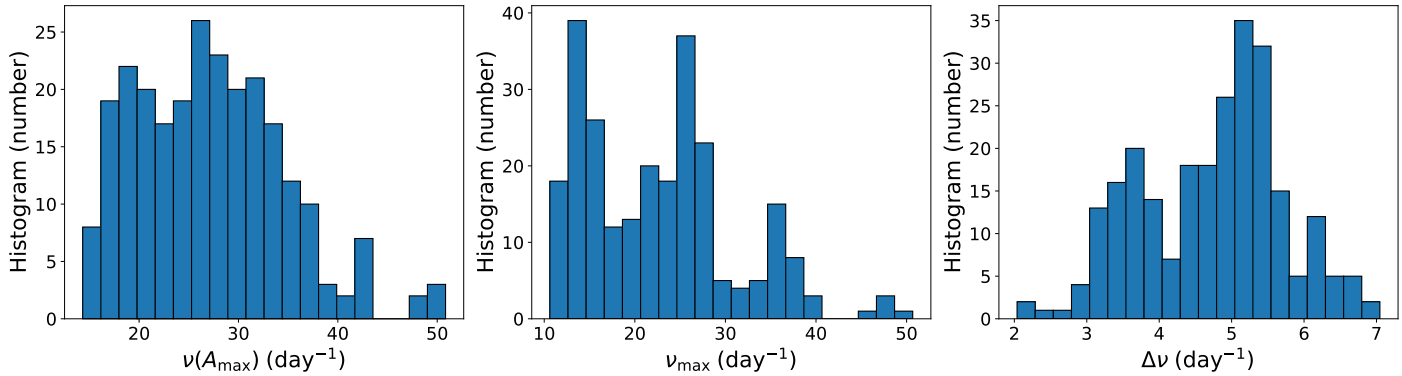

Figure 7. Histogram for $\nu(A_{max})$ (left panel), ν_{max} (middle panel), and $\Delta\nu$ (right panel) of 251 δ Sct stars predicted by ensemble ML (see supplementary electronic table).

(The data used to create this figure are available in the online article.)

applying traditional, threshold-based, or peak-picking techniques, which often require extensive manual intervention and are highly sensitive to noise. But the ML ensemble regression algorithm used to estimate asteroseismic parameters with a considerably short light curve i.e., 7000 data points is a solution compared with a traditional approach in determining the key parameters. Furthermore, scaling these traditional methods to the vast data streams from current and future space missions such as TESS and PLATO presents an additional computational burden (M. Hon et al. 2018a; J. A. Guzik & M. Roth 2021; M. Hon et al. 2021; S. J. Murphy et al. 2023). To address these limitations, ML algorithms have emerged as powerful alternatives. ML models, especially those designed for sequence and pattern recognition, can learn complex nonlinear relationships directly from the frequency spectra, making them highly effective at extracting asteroseismic parameters in a data-driven fashion (E. P. Bellinger et al. 2016; M. Hon et al. 2017, 2018b, 2020; J. S. G. Mombarg et al. 2021). In particular, ensemble approaches—where multiple base learners are combined—have demonstrated increased robustness and accuracy in identifying $\nu(A_{max})$, ν_{max}, and $\Delta\nu$ across diverse pulsating behaviors and observational conditions (S. K. Panda et al. 2024a). These ML models surpass traditional methods in accuracy and provide automated processing, allowing efficient analysis of the large datasets-often termed "big data" from contemporary asteroseismic surveys. With space missions such as TESS and the upcoming PLATO (H. Rauer et al. 2014) generating vast amounts of high-precision photometric time series, ML-based pipelines are becoming essential for deriving the physical properties of thousands of variable stars with minimal manual effort. Moreover, ML techniques directly learn to predict asteroseismic parameters from light curves, bypassing the need for intermediate frequency extraction steps. This approach proves especially useful for stars where traditional preprocessing fails or becomes unreliable due to low signal-to-noise ratios or instrumental artifacts. ML models also handle noisy, incomplete, or irregularly sampled data effectively, which further increases their applicability to real-world observations (C. Reyes et al. 2022; O. J. Scutt et al. 2023; J.-S. Pan et al. 2024). Finally, the creation of a comprehensive, curated database of δ Sct stars asteroseismic parameters—including $\Delta\nu$, ν_{max}, and $\nu(A_{max})$—across a wide range of stellar types would significantly improve model generalization and performance. Such a dataset would enable the development of next-generation ML algorithms capable of inferring not only seismic parameters but also broader physical properties (e.g., mass, age, composition) directly from light curves. In summary, ML-particularly ensemble-based and scalable implementations-offers a transformative path forward for asteroseismic analysis of δ Sct stars. It addresses the limitations of traditional techniques, facilitates the processing of large and complex datasets, and opens the door to precise, automated parameter inference in the era of big-data-driven stellar astrophysics.

6. Conclusion

This study presents a ML framework designed to estimate key asteroseismic parameters directly from stellar light curves. The framework is applied to a sample of 643 δ Sct stars observed by the TESS mission. Each light curve needs at least 7000 data points to extract 153 features. These include statistical descriptors, PCA, ACF, FFT, and DWT. Empirically, we found that using these 153 features provides optimal performance in predicting asteroseismic indices while keeping computational costs low in the ML approach. Using fewer features led to underfitting, reducing predictive accuracy, whereas using more features caused overfitting without substantial gain, thus confirming that 153 features offer a balanced and empirically optimal representation. Also, we find that for light curves less than 7000 data points, the extracted features may insufficient to achieve the high performance ML base regressor. We implement a supervised stacking ensemble regressor composed of Random Forest, Gradient Boosting, and Ridge Regression. The framework is evaluated using a subset of 583 δ Sct stars for training and testing, and an independent set of 60 δ Sct stars for generalization. To ensure robustness, we repeat the random 80/20 split of the training/testing set over 100 iterations. The ensemble model achieves high predictive accuracy ($R^2 > 0.77$) to predict asteroseismic indices.

The trained model generalizes well to the 60 unseen δ Sct stars, for which predicted asteroseismic parameters show good agreement with those derived from traditional asteroseismology methods. We additionally derived the asteroseismic indices for 251 δ Sct stars, producing a new catalog. In this work, for the features selected from the light curves of δ Sct stars using the stacked ML approach, we achieve significant estimates of asteroseismic parameters. Although modifying the number of features and employing more advanced ML models, which are rapidly developing, could potentially further

improve the performance of the present method, there is also room for improvement by computing features that provide a better interpretation of the light curves. One important limitation of the current method is the requirement of at least three light curves per star, each containing a minimum of 7000 data points, in order to estimate the mean values and uncertainties of asteroseismic quantities. Moreover, the ML algorithm predicts asteroseismic parameters without considering the nature of the input light curve, which means that the light curve may not correspond to the type of stars used in training or may belong to stars that inherently lack measurable asteroseismic quantities such as $\nu_{\max}$, and $\Delta\nu$. Therefore, it is necessary to examine the periodogram of each star before feeding it to the algorithm to ensure the presence of multiple pulsation frequencies and the feasibility of estimating asteroseismic parameters. However, in this study, the periodogram is not used directly to calculate these quantities. In future work, this approach can be extended to estimate additional asteroseismic indices such as small frequency separations, parameters of other types of variable stars, and to derive stellar parameters and asteroseismic indices using long-cadence data from various observations.

Acknowledgments

This paper includes data collected by the TESS mission, which are publicly available from the Mikulski Archive for Space Telescopes (MAST).

This work makes use of the Lightkurve package for accessing and processing the TESS light curves. The analysis also utilizes external software tools, including Scikit-learn and PyWavelets (PyWT; https://scikit-learn.org). We sincerely thank the reviewer for their constructive comments and valuable suggestions. This work is based upon research funded by the Iran National Science Foundation (INSF) under projects No. 4038404 and 40402831.

ORCID iDs

Akram Bolouki https://orcid.org/0009-0000-1511-6311
Amir Hasanzadeh https://orcid.org/0000-0002-7286-1438
Hossein Safari https://orcid.org/0000-0003-2326-3201

References

Abbasvand, S., Hasanzadeh, A., Alipour, N., & Safari, H. 2025, MNRAS, 544, L1
Aerts, C. 2021, RvMP, 93, 015001
Aerts, C., Christensen-Dalsgaard, J., & Kurtz, D. W. 2010, Asteroseismology (Springer), 1
Antoci, V., Cunha, M. S., Bowman, D. M., et al. 2019, MNRAS, 490, 4040
Balona, L. A. 2014, MNRAS, 437, 1476
Balona, L. A., & Dziembowski, W. A. 2011, MNRAS, 417, 591
Barceló Forteza, S., Michel, E., Roca Cortés, T., & García, R. A. 2015, A&A, 579, A133
Barceló Forteza, S., Pascual-Granado, J., Suárez, J. C., et al. 2024, MNRAS, 535, 2189
Barceló Forteza, S., Roca Cortés, T., García Hernández, A., & García, R. A. 2017, A&A, 601, A57
Bedding, T. R., Murphy, S. J., Hey, D. R., et al. 2020, Natur, 581, 147
Bedding, T. R., Murphy, S. J., Crawford, C., et al. 2023, ApJL, 946, L10
Bellinger, E. P., Angelou, G. C., Hekker, S., et al. 2016, ApJ, 830, 31
Bowman, D. M. 2017, Amplitude Modulation of Pulsation Modes in Delta Scuti Stars (Springer)
Bowman, D. M., & Kurtz, D. W. 2018, MNRAS, 476, 3169
Bowman, D. M., Kurtz, D. W., Breger, M., Murphy, S. J., & Holdsworth, D. L. 2016, MNRAS, 460, 1970
Bradley, P. A., Guzik, J. A., Miles, L. F., et al. 2015, AJ, 149, 68
Breger, M., Lenz, P., & Pamyatnykh, A. A. 2009, MNRAS, 396, 291
Breiman, L. 2001, Mach. Learn., 45, 5
Chang, J., Guo, B., & Yao, Q. 2018, AnSta, 46, 2094
Chen, X., & Li, Y. 2017, ApJ, 838, 31
Daszyńska-Daszkiewicz, J., Szewczuk, W., & Walczak, P. 2024, MNRAS, 532, 1140
Dornan, V., & Lovekin, C. C. 2022, ApJ, 924, 130
Dupret, M. A., Grigahcène, A., Garrido, R., Gabriel, M., & Scuflaire, R. 2004a, A&A, 414, L17
Dupret, M. A., Grigahcène, A., Garrido, R., Gabriel, M., & Scuflaire, R. 2005, A&A, 435, 927
Dupret, M. A., Thoul, A., Scuflaire, R., et al. 2004b, A&A, 415, 251
Friedman, J. H. 2001, AnSta, 29, 1189
García Hernández, A., Suárez, J. C., Moya, A., et al. 2017, MNRAS, 471, L140
Garrido, R., & Rodriguez, E. 1996, MNRAS, 281, 696
Gootkin, K., Hon, M., Huber, D., et al. 2024, ApJ, 972, 137
Goupil, M. J., Dziembowski, W. A., Pamyatnykh, A. A., & Talon, S. 2000, ASPC, 210, 267
Guzik, J. A. 2021, FrASS, 8, 55
Guzik, J. A., & Roth, M. 2021, FrASS, 8, 97
Hasanzadeh, A., Safari, H., & Ghasemi, H. 2021, MNRAS, 505, 1476
Hey, D., & Aerts, C. 2024, A&A, 688, A93
Hoerl, A. E., & Kennard, R. W. 1970, Technometrics, 12, 55
Hon, M., Bellinger, E. P., Hekker, S., Stello, D., & Kuszlewicz, J. S. 2020, MNRAS, 499, 2445
Hon, M., Stello, D., & Yu, J. 2017, MNRAS, 469, 4578
Hon, M., Stello, D., & Yu, J. 2018a, MNRAS, 476, 3233
Hon, M., Stello, D., & Zinn, J. C. 2018b, ApJ, 859, 64
Hon, M., Huber, D., Kuszlewicz, J. S., et al. 2021, ApJ, 919, 131
Kjeldsen, H., & Bedding, T. R. 1995, A&A, 293, 87
Kumar, N., Singh, H. P., Malkov, O., et al. 2025, Univ, 11, 207
Kurtz, D. W. 2022, ARA&A, 60, 31
Kurtz, D. W., Saio, H., Takata, M., et al. 2014, MNRAS, 444, 102
Lares-Martiz, M. 2022, FrASS, 9, 301
Lee, G. R., Gommers, R., Waselewski, F., Wohlfahrt, K., & O'Leary, A. 2019, JOSS, 4, 1237
Lovekin, C. C., & Guzik, J. A. 2017, ApJ, 849, 38
Lv, C., Esamdin, A., Hasanzadeh, A., et al. 2024, A&A, 686, A174
Lv, C., Esamdin, A., Hasanzadeh, A., et al. 2023, ApJ, 959, 33
Mirouh, G. M., Angelou, G. C., Reese, D. R., & Costa, G. 2019, MNRAS, 483, L28
Mombarg, J. S. G., Van Reeth, T., & Aerts, C. 2021, A&A, 650, A58
Mosser, B., & Appourchaux, T. 2009, A&A, 508, 877
Mosser, B., Elsworth, Y., Hekker, S., et al. 2012, A&A, 537, A30
Mosser, B., Michel, E., Belkacem, K., et al. 2013, A&A, 550, A126
Murphy, S. J., Bedding, T. R., Gautam, A., & Joyce, M. 2023, MNRAS, 526, 3779
Murphy, S. J., Bedding, T. R., Gautam, A., Kerr, R. P., & Mani, P. 2024, MNRAS, 534, 3022
Murphy, S. J., Paunzen, E., Bedding, T. R., Walczak, P., & Huber, D. 2020, MNRAS, 495, 1888
Nielsen, M. B., Davies, G. R., Chaplin, W. J., et al. 2023, A&A, 676, A117
Pan, J.-S., Ting, Y.-S., & Yu, J. 2024, MNRAS, 528, 5890
Panda, S. K., Dhanpal, S., Murphy, S. J., Hanasoge, S., & Bedding, T. R. 2024a, ApJ, 960, 94
Panda, S. K., Hanasoge, S., Dhanpal, S., & DC, V. 2024b, ApJL, 975, L12
Ramón-Ballesta, A., García Hernández, A., Suárez, J. C., et al. 2021a, MNRAS, 505, 6217
Ramón-Ballesta, A., García Hernández, A., Suárez, J. C., et al. 2021b, Plato Mission Conference 2021, Zenodo, doi:10.5281/zenodo.5562837
Rauer, H., Catala, C., Aerts, C., et al. 2014, ExA, 38, 249
Reyes, C., Stello, D., Hon, M., & Zinn, J. C. 2022, MNRAS, 511, 5578
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2015, JATIS, 1, 014003
Sánchez Arias, J. P., Córsico, A. H., & Althaus, L. G. 2017, A&A, 597, A29
Scutt, O. J., Murphy, S. J., Nielsen, M. B., et al. 2023, MNRAS, 525, 5235
Suárez, J. C., García Hernández, A., Moya, A., et al. 2014, A&A, 563, A7
Suárez, J. C., Garrido, R., & Goupil, M. J. 2006, A&A, 447, 649
Templeton, M., Basu, S., & Demarque, P. 2001, ApJ, 563, 999
Turcotte, S., Richer, J., Michaud, G., & Christensen-Dalsgaard, J. 2000, A&A, 360, 603
Vasigh, F., Ziaali, E., & Safari, H. 2024, ApJ, 969, 19
Viani, L. S., Basu, S., Corsaro, E., Ball, W. H., & Chaplin, W. J. 2019, ApJ, 879, 33
Xiong, D. R., Deng, L., Zhang, C., & Wang, K. 2016, MNRAS, 457, 3163
Zhou, A.-Y. 2025, RNAAS, 9, 56
Zhou, Z.-H. 2025, Ensemble Methods: Foundations and Algorithms (2nd ed.; Chapman & Hall/CRC), 364