Original Research Article

Multiple Input Neural Network with Fourier Series to Classify Variable Stars

Jayveer Kochhar

Dhirubhai Ambani International School,Taraangan, 24 presidency society, NS road no. 7, JVPD Scheme, Mumbai, Maharashtra, 400098, India

ABSTRACT

Measurement of astronomical distances, understanding of stellar evolution, and improvement of our knowledge of galactic structure depend on accurate classification of variable stars. Conventional classification techniques find it difficult to handle the rising volume of data produced by massive sky surveys such as Large Synoptic Survey Telescope (LSST), All Sky Automated Survey for Supernovae (ASAS-SN), and Optical Gravitational Lensing Experiment (OGLE). This work offers a hybrid deep learning method using Multiple-Input Neural Networks (MINN) to improve classification accuracy by means of image-based light curve analysis combined with astrophysical parameters derived via Fourier decomposition and skewness analysis. Problems including class imbalance, phase misalignment, and subtype distinction are addressed in the suggested approach. Minima Phase Standardization aligns phase-folded light curves for uniformity; the Fourier Best Fit model extracts important coefficients reflecting light curve shape. A Variable Star Light Curve Simulator creates synthetic data for underrepresented classes, especially ACeps and Type II Cepheids, to reduce dataset imbalance, therefore guaranteeing a more balanced training dataset. In ten epochs, the hybrid model attained an overall classification accuracy of 89.8%; considerable gains for rare classes were obtained. For common classes, the convolutional neural network (CNN) alone achieved 98.1% accuracy. This work emphasizes, especially for rare and underrepresented variable star types, the need of integrating deep learning with astrophysical insights to increase classification accuracy. By laying the groundwork for automated, large-scale classification of variable stars, the proposed framework accelerates the study of huge astronomical datasets and improves our knowledge of the structure and development of the universe.

Keywords: Multiple-Input Neural Networks; Fourier decomposition; Phase-Folded Light Curves; Variable Stars; Data Augmentation

INTRODUCTION

One of the toughest problems in the science of astrophysics is trying to compare the vast distances of the universe. Knowledge of the structure, dynamics, and evolution of the cosmos depends on exact measurement of astronomical distances. Accurate measurement of these distances enables us to map galaxies, probe cosmic expansion, and determine fundamental properties of

Corresponding author: Jayveer Kochhar, E-mail: jsk130908@gmail.com.

Accepted November 7, 2025
https://doi.org/10.70251/HYJR2348.36432448

celestial objects. There are multiple different ways one can measure distance at these scales, each appropriate for a different magnitude of measurement. The parallax method uses the apparent change in the location of surrounding stars resulting from Earth's orbit. Although accurate, it fails at larger scales due to limitations in currently available technology (1). On the other hand, one can use cosmological redshift (2, 3): the stretching of light from far-off galaxies arising from the universe's expansion, therefore providing a measure of distance at cosmic scales but requiring presumptions about the structure of the universe. These techniques have several shortcomings even if they are efficient. Beyond a few thousand light-years, parallax loses efficacy since the little angular shift brought about by Earth's orbit is too small to precisely detect.

Conversely, redshift depends on broad cosmological models and assumptions about the nature of the cosmos. Nevertheless, standard candles are a vital instrument for exploring the distant reaches of the universe since they may be utilized over considerably more distances (3). Astronomers fill in such spaces with a particular group of celestial objects known as standard candles. Standard candles are astronomical bodies whose luminosity, that is, the whole amount of naturally emitted light, is either known or can be exactly computed. By applying the inverse square law of light, which maintains that the apparent brightness (how bright a star appears to us) of an object is directly proportional to the square of its distance from the observer, astronomers can ascertain the distance of an object. By matching the observed brightness and how bright it looks from Earth with the inherent brightness and how much light it naturally emits, astronomers can exactly ascertain the distance.

Many standard candles, especially variable stars like Cepheids and RR Lyrae, have a defining quality in their period-luminosity relationship, a predictable pattern in which the time it takes for the star's brightness to vary is exactly linked with its intrinsic brightness. Originally found by Henrietta Leavitt and detailed in (4) Physics Department *et al*., this link lets astronomers find the absolute luminosity of a star just by timing its brightness fluctuations.

Standard candles are not limited by distance unlike techniques like parallax and redshift. Reliable distance measurements depend on precise classification of variable stars. Misclassifications can cause major mistakes, therefore influencing the conclusions derived from the measurements and ultimately, our understanding of the cosmos. Traditional categorization techniques have failed to keep up as astronomical datasets are expanding fast with large-scale sky surveys such the Optical Gravitational Lensing Experiment (OGLE) and the Large Synoptic Survey Telescope (LSST). Emerging as a potent option is machine learning (ML), which automates the classification of variable stars with remarkable speed and accuracy and allows the examination of vast, complicated datasets. Astronomers can better classify variable stars, increase accuracy of distance estimations, and further our knowledge of the structure and development of the cosmos by using ML. This combination of sophisticated computer methods and astrophysical knowledge heralds a new phase in our cosmic inquiry.

LITERATURE REVIEW

Variable stars are a diverse group of stars whose brightness changes over time, either due to intrinsic processes within the star or external interactions. Intrinsic variability often arises from pulsations in the star's outer layers, driven by changes in pressure and temperature. Standard reference works on variable star characteristics can be found in introductory texts (5).

Types of variable stars

Cepheids are standard candles and pulse periodically because of the interaction between ionized helium layers and the star's gravitational forces. With ionized helium serving as a "valve," these pulsations arise as the outer layers of the star expand and shrink cyclically. The ionized helium becomes opaquer during compression, therefore trapping heat and raising pressure. This results in outward growth. The helium gets less ionized and more transparent as the star grows, allowing heat to escape and driving contraction of the outer layers once more. Determining distances to galaxies depends critically on the known relationship between the pulsation period of Cepheid and luminosity, the period-luminosity connection.

RR Lyrae stars are another prominent type of variable star. Though their pulsation-driven variability is similar to that of Cepheids they are older, metal-poor stars and are frequently observed in globular clusters. Though they are smaller stars, their shorter pulsation periods, usually less than a day, are powered by comparable helium ionization processes such as Cepheids. RR Lyrae stars, unlike Cepheids, have almost homogeneous intrinsic luminosity, which makes them very helpful for investigating galaxy structures.

Eclipsing binaries are a class of variable stars in which the interaction between two stars causes periodic variation in apparent brightness. Two stars orbit one another in these systems, and as seen from Earth, the measured light regularly dims when one star passes in front of the other (Figure 1). An eclipsing binary's light curve offers useful information on the orbital dynamics, mass, and size of the stars. Eclipsing binaries also prove to be useful in revealing the existence of extra companions or exoplanets through thorough study of these systems.

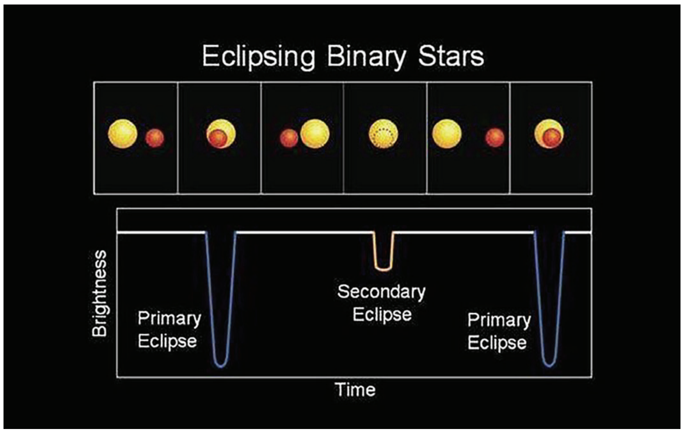

Figure 1. Visual representation of an eclipsing binary through 1 complete period.

Less common but absolutely vital subclasses are Type II Cepheids (T2Cephs) and ACeps. Considered as metal-poor stars in binary systems, anomalous Cepheids are sometimes formed by mass transfer in eclipsing binary systems influencing their variability. Though their luminosities are smaller and their periods are shorter than those of traditional Cepheids, they nevertheless adhere to a slightly more complicated period luminosity relation. Usually older and less mass, T2Cephs are split into subgroups according to their pulsation periods: RV Tauri stars, W Virginis stars, and BL Her stars. Every subgroup has unique physical properties, usually including interactions between star pulsations and variations in atmospheric opacity resulting from ionization.

There are many more types of variable stars such as heartbeat stars and other types of standard candles such as supernovae but don't require classification. This is because they possess highly distinct features and only a handful of these classes exist within the observable universe.

The Currently Available Data

The Optical Gravitational Lensing Experiment's fourth iteration (OGLE-IV) survey offers the most robust and varied dataset having the most star types and highest overall number of stars. Covering many star fields such as the Galactic bulge, disc, and Magellanic Clouds, OGLE-IV catalogs' millions of stars with broad photometric observations in the I-band and V-band. The dataset includes time series data of the star's magnitude, the position of the star in the night sky, period, Fourier coefficients, reference times, and more important data of the star. Because of its coverage, accuracy, and well-documented metadata, which give a strong basis for variable star classification, OGLE-IV's dataset is a pillar for astrophysical study. The OGLE-IV dataset has been described in detail by Soszyński (6).

Challenges in Variable Star Classification

Manual techniques and human knowledge have always been the foundation of accurate classification of variable stars. But the volume of data has increased exponentially, surpassing the capabilities of conventional techniques with the arrival of massive surveys like OGLE-IV and LSST, which calls for the introduction of automated methods. Machine learning has proved to be effective at handling these massive datasets and raising classification accuracy.

Previous classifiers captured temporal patterns in light curve data using recurrent neural networks (RNNs). RNNs were good in spotting general categories of variable stars, but their accuracy suffered when predicting under-represented classes like ACeps and T2Cephs due to imbalanced datasets. Moreover, these models lacked generalizability over datasets, especially those with noise or lacking data.

Recent advances have seen the application of phase-folding the light curve and inputting into a convolutional neural network (CNN), which leverages the periodicity of variable stars by folding their light curves over one or two periods. This approach enhances feature extraction for periodic signals, leading to better classification outcomes. Studies such as (7) Szklenár *et al.* and (8) Jayasinghe *et al.* highlight these challenges and emphasize the need for improved representation of rare classes. However, even these advanced models face limitations. For instance, phase-folded CNNs often suffer from biases due to class imbalances in training data, resulting in poor performance for rare subclasses of variable stars.

Gaps in Existing Research

Although variable star categorization has been automated with considerable progress, important gaps still exist. Current models can miss completely using astrophysical domain knowledge, including evolutionary properties and period-luminosity correlations. Furthermore, limiting the dependability of current classifiers is the under-representation of some classes together with difficulties resulting from observational gaps and noisy data. Dealing with these gaps calls for an integrated approach combining astrophysical insights with machine learning methods to guarantee strong categorization throughout several datasets and star types.

Research question

Can we develop machine learning models for variable star classification which can be made more transferable across different astronomical surveys by integrating astrophysical knowledge, leveraging explainability techniques, and incorporating uncertainty quantification to handle observational gaps and noisy data?

This study directly addresses these limitations by developing a hybrid deep learning framework that integrates astrophysical knowledge with data-driven methods for more robust and generalizable variable star classification. While previous approaches have primarily relied on image-based light curve analysis, this work incorporates Fourier-derived physical parameters, skewness coefficients, and period information alongside phase-folded light curve images within a Multiple-Input Neural Network (MINN) architecture. This combination enables the model to capture both the morphological and physical characteristics of stellar variability, improving discrimination among visually similar subtypes such as Type II and Anomalous Cepheids. Furthermore, a Variable Star Light Curve Simulator is introduced to generate realistic synthetic data for underrepresented classes, mitigating the persistent issue of class imbalance observed in large-scale surveys. By uniting astrophysical domain insights with machine learning, this research not only enhances classification accuracy for rare variable stars but also establishes a transferable methodology applicable across multiple surveys. Consequently, this study contributes a significant methodological advancement toward scalable, interpretable, and physically grounded automated classification in modern astrophysics.

METHODS AND MATERIALS

Downloading and Structuring Data

The datasets from large astronomical studies such as the OGLE are often distributed in .dat or .txt formats, sometimes archived as.tar.gz files containing multiple light curves. Given the scale of these surveys, an automated script was developed to download and structure the data efficiently. The folders that are processed are names as per star fields Galactic bulge (blg), Small Magellanic Cloud (smc), Large Magellanic Cloud (lmc), Galactic Disk (gd) and anomalous Cepheids in the entire galaxy (gal).

After extraction, the raw files were organized into a directory, and subsequently, a comprehensive metadata file, referred to as the "mega data file", was created by merging all individual files corresponding to different star subtypes. This consolidated file contained detailed information for all stars in the dataset. The key columns in the metadata file included: Star ID, a unique identifier formatted as *OGLE4–(star field)–(star type)–(star number)*; Type, representing the broad classification of the variable star (such as Cepheid, RR Lyrae, or Eclipsing Binary); Subtype, denoting a more specific classification that captures distinct pulsation modes (fundamental, first overtone, or multimode); Period (P), defined as the time in days between successive maxima of magnitude; and Epoch (t_0), the reference time corresponding to the maxima for pulsating stars or minima for eclipsing binaries.

This structured format ensures that every light curve is linked to its corresponding metadata, allowing for streamlined preprocessing and classification. By implementing this automated data retrieval system, the study efficiently handles large volumes of astronomical data while maintaining accuracy in variable star identification.

Magnitude to Flux Conversion

In astronomy, magnitude is a logarithmic measurement of object brightness. Flux, on the other hand, is a direct physical indicator of a star's brightness and shows the real quantity of light energy obtained per unit area. More suited for computer models, flux is linearly proportional to luminosity unlike magnitude.

To transform magnitude into flux, the following standard relation is applied:

$$F = 10^{-0.4(m_0 - m)} \tag{1}$$

Where: F is the flux in arbitrary units, m is the observed magnitude from the dataset, m_0 is a reference magnitude, often chosen to normalize flux values.

For consistency across all stars, a standard reference magnitude of 25 has been used. Using a fixed reference magnitude also standardizes the light curves generated, allowing for better model generalization.

Normalization of Flux

After converting magnitudes to flux, due to errors in the data provided or in the processing algorithm certain data points sometimes exceed 1. To resolve this the flux values are normalized by dividing by the maximum flux observed in each light curve:

$$F' = \frac{F}{F_{max}} \tag{2}$$

Where: F' is the normalized flux, F is the original flux, F_{max} is the maximum flux value for that particular star.

This simple normalization method ensures that all values fall within the range [0,1] preserving the relative variations in brightness while preventing extreme numerical values. It also standardizes the representation of light curves across different types of variable stars, ensuring that the model learns variability patterns rather than being influenced by differences in absolute luminosity. However, completely excluding absolute luminosity can obscure key physical distinctions between classes, particularly for types like Cepheids, RR Lyrae, and Type II Cepheids, where intrinsic brightness is astrophysically meaningful. To strike an optimal balance, we preserve relative brightness information while avoiding overfitting to survey-specific magnitude scales. As discussed in Section 2.3.2, the average absolute magnitude is included as a normalized numerical input, allowing the model to incorporate luminosity in a consistent and physically grounded way without letting it override temporal variability features.

Phase Folding the Light Curves

Variable stars show periodic brightness fluctuations, so their light curves repeat over a given length. Still, the raw light curve comprises magnitude measurements dispersed over several time stamps when observations are taken at different times. Expressing all data points in terms of their phase (how much of the period has been completed) allows phase folding to convert these observations from a time-based representation into a single, recurring cycle. Each observation is thus allocated a phase value between 0 and 1 instead of absolute time, essentially "stacking" several cycles of the brightness changes of the star onto a single period. This lets astronomers see and examine a whole cycle of fluctuation independent of the timing of individual observations.

Phase folding is performed using the equation:

$$\phi = \frac{(t-t_0)}{P} \ (mod\ 1) \tag{3}$$

Where: ϕ is the phase (a value between 0 and 1), t is the observation time, t_0 is the reference epoch, P is the period of the variable star.

This step is essential when using CNNs for classification because it allows the conversion of light curves from a text-based format (time-series data) into images which allows the model to actually understand the shape of the curve and classify accordingly (Figure 2).

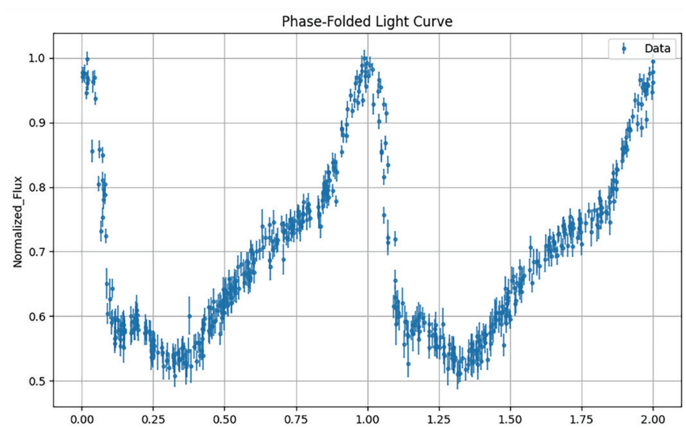

Figure 2. Phase-folded light curves show the periodic variability of stellar brightness after aligning observations by phase.

Conversion of Phase-Folded Light Curves to Images

After phase folding, the variable star light curves remain in a text-based format, consisting of phase values and corresponding normalized flux values. However, since we aim to classify stars based on the shape of their light curves using CNN, we must convert these numerical data points into images. The light curves were converted into 128 x 128-pixel grayscale images to standardize the classification procedure therefore guaranteeing a consistent input format for model training. With a width split into 128 discrete bins, the phase values, which span 0 to 2, are mapped onto the x-axis of the image and a similar method is used for the y axis with flux.

The conversion process is carried out using the Pillow library in Python. First, a blank image of size 128 × 128 pixels is created to act as the canvas for plotting the phase-folded data. The phase values, which range from 0 to 2, are mapped onto the x-axis, while the normalized flux values are projected onto the y-axis, with each data point being placed at a corresponding pixel coordinate based on its phase and brightness. To encode point density, the pixel intensity is adjusted by counting how many data points fall within each pixel and taking the natural logarithm of this count. The pixel with the highest log value is scaled to an intensity of 255, and all other pixels are scaled proportionally relative to this maximum. Finally, the resulting image is saved as a grayscale PNG file, which serves as the input for the CNN.

To further enhance learning, the images include two consecutive phase cycles (0 to 2) instead of a single cycle (0 to 1). This change enables the model to grasp the periodic character of variable stars, therefore avoiding treatment of phase borders as unconnected edges. It also lessens edge effects, therefore guaranteeing that transitions between the start and end of the light curve seem natural rather like manufactured cut-off. Additionally, some variable star classes exhibit subtle differences that become clearer over multiple cycles. Particularly for rare categories including Anomalous Cepheids and Type II Cepheids, the model increases its capacity to discriminate between comparable subtypes by presenting two full phases (Figure 3).

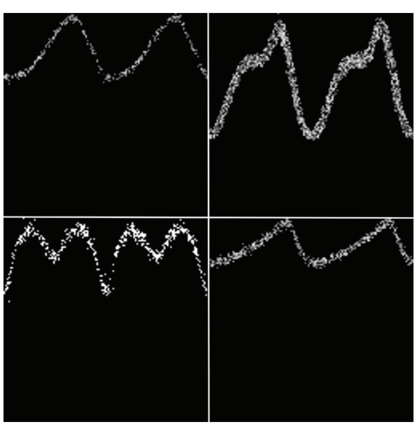

Figure 3. Phase-folded light curves converted into 128×128 grayscale image representations, enabling the convolutional neural network to process temporal features spatially.

Initial CNN Model and Observed Challenges

A basic CNN was implemented to classify the phase-folded light curve images into main variability types such as Cepheids (cep), RR Lyrae (rrlyr), and Eclipsing Binaries (ecl). Although CNN achieved an overall accuracy of 98.1%, this metric largely reflects the dominant classes in the dataset and does not represent balanced performance across all types. As shown in the confusion matrix, certain rare classes such as Anomalous Cepheids were not correctly classified at all, while others like T2Cephswere frequently misclassified. This limitation arises from the significant class imbalance—where common classes such as ECL and RR Lyrae dominate—which inflates the overall accuracy. Therefore, overall accuracy alone is not an adequate measure of the model's performance, particularly for under-represented classes. These shortcomings motivated the development of additional strategies, such as incorporating astrophysical parameters and synthetic data augmentation, to achieve fairer classification across all classes. However, several key challenges emerged during experimentation. First, the limitations of using only image-based input became apparent. While the CNN was able to identify broad shape patterns effectively, image data alone proved insufficient for precise subtype classification. Critical numerical features—such as period, amplitude, and asymmetry metrics—are indispensable for distinguishing closely related subclasses, and their absence significantly weakened the model's ability to separate categories like classical Cepheids and Type II Cepheids. Second, inconsistent phase alignment across different variable star types introduced unintended biases. For pulsating stars, the phase reference starts at maximum brightness, whereas for eclipsing binaries, it begins at minimum brightness. As a result, the model began to use this phase offset as a shortcut for classification instead of learning the intrinsic structure of the light curve. Finally, class imbalance issues caused poor performance on rare categories such as Acep and T2Cephs, which were often misclassified as more common types like Cepheids or RR Lyrae due to the dominance of these classes in the training set.

Although CNN attained good overall accuracy, these constraints revealed shortcomings in its capacity to handle rare classes, inconsistencies in data representation, and the lack of fundamental characteristics needed for fine-grained categorization. Three significant corrections were done to address class imbalance, phase misalignment, and absence of significant astrophysical properties. Important characteristics like harmonic distortions, phase shifts (Φ), and amplitudes (A) were first derived using Fourier Best Fit, therefore providing a systematic representation of variability patterns. These properties should enable better categorization of stars having similar light curve shapes. Second, Minima Phase standardization helped all pulsing variable stars to be at least in brightness oriented to a reference epoch. This modification provides consistency in phase-folding by eliminating phase misalignment between different types of stars, hence removing biases (Figure 4). Third, the Variable Star Light Curve Simulator was designed to generate synthetic light curves for underrepresented groups like Type II Cepheids (T2Cep) and ACep. The simulator preserves astrophysical realism by using Fourier-based statistical distributions, Gaussian noise, amplitude fluctuations, irregular sampling, and missing data points to reproduce genuine observations. Together, boosting classification accuracy, balancing dataset representation, and incorporating astrophysical domain knowledge ensure a more consistent and robust classification framework.

Fourier Best Fit with Skewness

A mathematical representation called the Fourier decomposition model divides periodic functions, like the brightness fluctuations of variable stars, into a sum of sinusoidal components. Fourier series offer a good method to faithfully depict the forms of variable star light curves since they are naturally periodic. Offering a thorough numerical depiction of a star's variability, the obtained Fourier coefficients capture important properties including amplitude fluctuations, phase shifts, and higher-order harmonic structures. Later on, these coefficients are used in the MINN, where they enhance discrimination between visually similar but physically different variable stars by adding astrophysical background, hence complementing the image-based classification. A more detailed discussion of how these Fourier coefficients are incorporated into the classification pipeline is provided in Section 2.3.2, where the MINN architecture is explained.

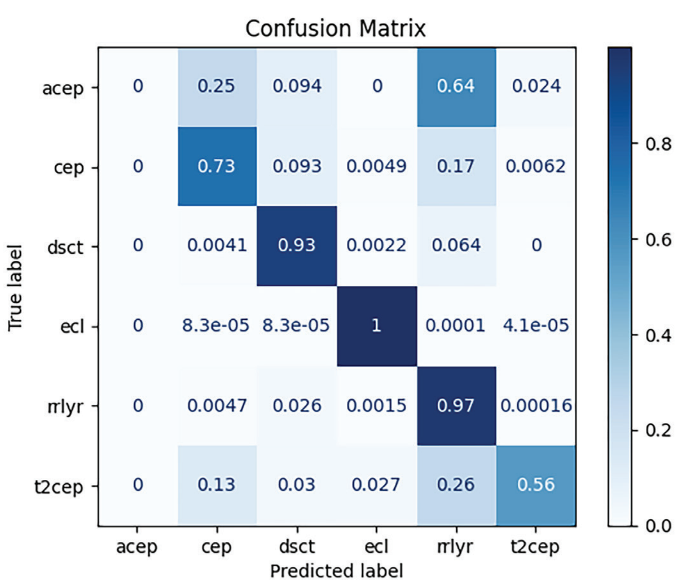

Figure 4. Confusion matrix of the basic type of classifier, indicating prediction accuracy and misclassification rates across major variable star categories.

The Fourier series used to fit the light curve is:

$$F(\phi) = A_0 + \sum_{i=1}^{H} A_i sin(2\pi i\phi + \phi_i) \quad (4)$$

Where: $F(\phi)$ represents the modeled flux at phase, H represents the number of harmonics, A_0 represents the mean flux, A_i represents the amplitude of the i-th harmonic and ϕ_i represents the phase shift of the i-th harmonic.

The periodic variability of many variable stars, such as pulsating stars (e.g., Cepheids, RR Lyrae), often exhibits asymmetry. This asymmetry is characterized by a steep rise to maximum brightness followed by a gradual decline to minimum brightness. A standard Fourier model with linearly increasing phase struggles to capture this asymmetry accurately, as it assumes that the changes in brightness occur symmetrically over the cycle.

To overcome this limitation, phase distortion is introduced in the Fourier model. This approach modifies the phase dynamically using sine (α) and cosine (β) terms.

$$\phi_{dis} = \phi + \alpha sin(2\pi\phi) + \beta sin(2\pi\phi) \quad (5)$$

The phase distortion essentially stretches or compresses specific parts of the phase cycle, allowing the model to adapt to asymmetrical patterns in the light curve.

Sine Distortion (α): This parameter is used to enhance the asymmetry of the light curve by modifying the steepness of the brightness variations. By adjusting α, the rise or fall in flux can be made either sharper or smoother, allowing the model to better capture the characteristic skewness observed in many variable star profiles.

Cosine Distortion (β): This parameter fine-tunes the alignment of key structural features such as peaks and troughs. By introducing a cosine-based adjustment, β ensures that the modeled flux aligns more accurately with the observed light curve shape, improving the representation of phase-dependent brightness variations.

While the Fourier model provides a flexible representation, high-order harmonics can sometimes introduce overfitting, especially in stars with complex light curves such as Cepheids (CEP) and Delta Scuti (DSCT) stars (Figure 5). To mitigate this, a regularization term (λ) was introduced to penalize excessive reliance on higher harmonics, thereby preventing unnatural oscillations in the model. A regularized cost function balances fitting the data well and keeping the model simple by adding a penalty for complexity. This prevents overfitting and ensures smoother, more reliable models.

A

B

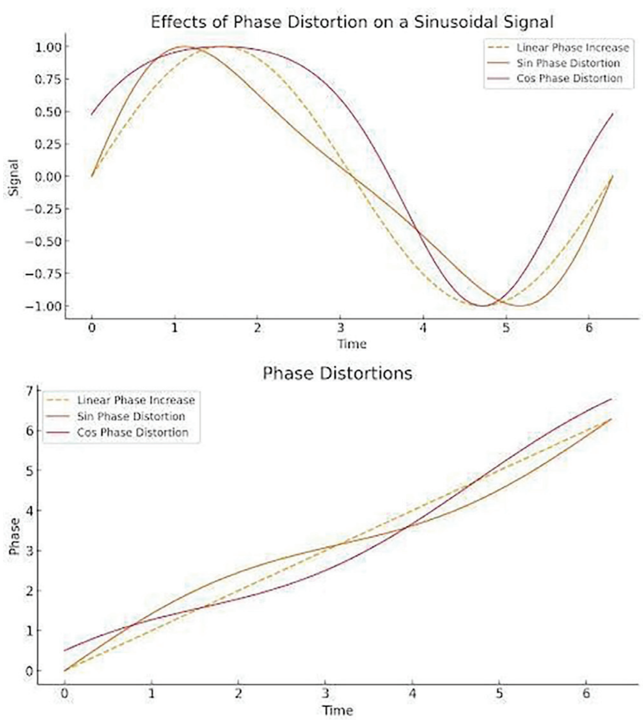

Figure 5. Effects of different phase distortions on a standard periodic signal, showing how linear, sine, and cosine phase perturbations modify both waveform shape and phase progression.

The regularized cost function is:

$$MSE_{reg} = \frac{1}{N}\sum_{j=1}^{N}\left(F_{obs}(\phi_j) - F_{model}(\phi_j)\right)^2 + \lambda \sum_{i=2}^{H}(i+1)^2 (A_i)^2 \quad (6)$$

Where λ is the regularization factor which controls how much each higher harmonic is penalized. This encourages the model to prioritize low-order harmonics, leading to smoother and more physically meaningful fits.

For CEP and DSCT stars, where overfitting was particularly problematic, an optimal λ value of 0.1 was chosen empirically to balance model accuracy and smoothness. Indeed, different λ values were tested for various classes by fitting light curves with several regularization strengths and comparing the resulting MSE and smoothness of the fits. Although λ was tuned per class, this adjustment affects only how flexibly the Fourier model fits the data, not the overall network complexity. It improves the physical realism of fits without adding significant computational cost or extra model parameters. Fourier decomposition has been widely applied in astrophysics (9).

The L-BFGS-B algorithm was initially used to optimize the Fourier coefficients by minimizing the error. However, since L-BFGS-B is a local optimization algorithm, it sometimes converged to suboptimal solutions, particularly in cases where the second and third harmonic amplitudes) exceeded 1, which is physically implausible and the model failed to converge properly, leading to higher-than-expected errors (Error>0.001).

To correct these anomalies, Differential Evolution (DE), a global optimization algorithm, was used in such cases. DE explores the entire parameter space more thoroughly, helping escape poor local minima. Once DE found a better initial guess using a maxiter of 100 and a population of 15, L-BFGS-B was used again to refine the final parameters. This hybrid approach of global (DE) + local (L-BFGS-B) optimization resulted in more reliable fits, particularly for Cepheids and Delta Scuti stars, where the light curves exhibit complex non-linear behavior.

Determining the Minima Phase for pulsators

Once the best-fit model was obtained, the minima phase was calculated by finding the phase at which the modeled flux is at its lowest value using the minimize_scalar function from scipy. This is critical for standardizing the light curve representations across variable star types. For pulsating stars, where the original epoch typically corresponds to maximum brightness, the new epoch is recalculated to correspond to minimum brightness using the formula:

$$t_{min} = t_0 + \phi_{min} \times P \quad (7)$$

Where: t_0 is the original reference time, P is the period of the star, ϕ_{min} is the phase of minimum flux.

This adjustment eliminates the model bias caused by inconsistent t_0 definitions, ensuring that all phase cycles begin at minimum brightness, regardless of the star type (Figure 6).

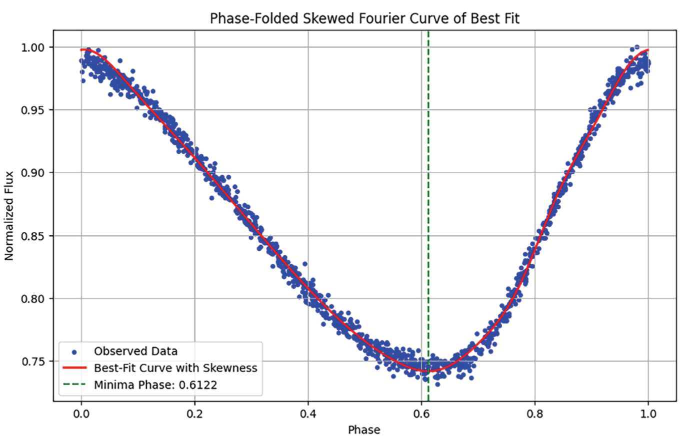

Figure 6. Phase-folded light curve with the best-fit skewed Fourier model overlaid, highlighting the phase of minimum flux and the degree of asymmetry in the pulsation cycle.

To address the data imbalance problem and improve the classification of underrepresented variable star classes, a Variable Star Light Curve Simulator was developed. This simulator generates synthetic light curves that mimic real observational data, ensuring that rare classes, such as ACeps and Type II Cepheids (T2Cep), have sufficient training samples for machine learning models. The simulator is based on the Fourier phase-distorted model, which accurately represents the periodic nature of variable stars. This model captures the underlying periodic structure while incorporating distortions to account for real-world asymmetries in brightness changes. The generated light curves are further refined by introducing observational effects that simulate real astrophysical data.

Random Gaussian noise used to replicate instrumental uncertainty is a fundamental component of the simulator. Instrumental limits cause sensor noise in real-world astronomical observations; hence, including this into the simulated data guarantees that the synthetic light curves precisely like real observational conditions. Apart from noise, the simulator generates astrophysical variability by periodically changing the light curve's amplitude. Many variable stars vary their brightness for long periods outside their periodic pulsations. A set amount is changed in the amplitude every specified number of periods to replicate these effects. This guarantees that the models of classification developed on this dataset can extend to actual stars showing modest evolutionary variations in their brightness.

The simulator also randomizes the time interval between data points, another crucial element. Real-world astronomical surveys find that weather circumstances and telescope scheduling limits cause observations to be not regularly spaced. The simulator chooses the time intervals between subsequent data points at random to duplicate this, therefore producing a dataset that reflects actual irregular sampling patterns. A small portion of the data produced are also randomly deleted to replicate missing observations brought on by telescope downtime, atmospheric circumstances, or other observational restrictions. The best fit data from Fourier Best Fit with Skewness is crucial at this stage to make sure the simulator generates accurate light curves. To make sure the light curves we generate are concordant with those already available we take means and standard deviations of all Fourier coefficients and period which are then used to generate accurate Fourier coefficients for the augmented data.

A problem that comes up when using this approach however is that amplitudes in higher harmonics vary greatly based on the initial amplitude and the same is true for phi values. This leads to high standard deviations in these values and in turn the simulator generates random data. To mitigate this, the amplitudes are expressed as a ratio to the first amplitude, and the phi values are represented as the difference from the first phase. This reduces the standard deviation and allows for accurate data augmentation. Analyzing the best fit data helps the simulator to replicate the flux variations observed in real data, hence maintaining an astrophysically accurate synthetic dataset free from artificial bias into the classification model. Synthetic data augmentation, as demonstrated by Becker *et al.* (10), helps improve classifier performance for rare classes.

The Machine Learning Model

Based on their phase-folded light curves, the CNN is meant to distinguish different stars (Figure 7). Its four-block design guarantees efficient separation between several variable star kinds by progressively extracting important aspects. CNNs have been widely used for image-based classification in astronomy (11). The CNN is built in a four-block deep architecture whereby every block gradually extracts better features from the light curve images. Larger kernel sizes in the first layers enable the network to capture wide periodicity and changes. Later layers with smaller kernel sizes concentrate on complex brightness fluctuations. To find the best architecture, several options for dropout rates, max pooling sizes, and stride lengths were investigated throughout model development.

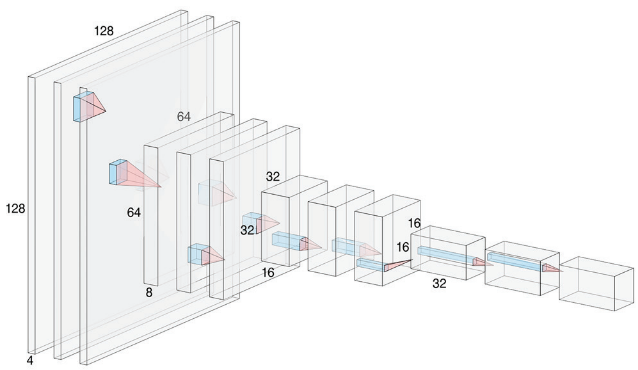

Figure 7. Three-dimensional schematic of the convolutional neural network architecture, illustrating progressive feature extraction across convolutional and pooling layers.

Multiple convolutional layers, each using ReLU activation functions to capture abrupt shifts in brightness unique of various variable stars, make up the last CNN design (Table 1). Two-by-two max pooling layers are used after every convolutional block to lower computational complexity while nevertheless preserving important spatial characteristics. Dropout layers are incorporated to improve generalization and prevent overfitting. The last completely linked layers improve the obtained features, thereby ensuring the model distinguishes between different sorts of stars. The softmax activation function in the output layer turns CNN's predictions into probability distributions spanning all conceivable star classes, therefore allowing the network to assign the most likely classification to any input image.

The Multiple Input Neural Network

While the CNN branch successfully classifies stars based on light curve morphology, many variable star types share highly similar visual patterns. This is particularly true for stars that have overlapping period-luminosity relationships, such as Classical Cepheids (Cep), Type II Cepheids (T2Cep), and RR Lyrae (RRLyr) stars. Additionally, rare variable stars such as ACep are significantly underrepresented in most training datasets, making them difficult to classify with CNN-based models alone. An extra branch was added in the network to incorporate numerical astrophysical characteristics acquired from Fourier decomposition and skewness analysis of the light curves, hence improving classification accuracy and lowering misclassification among similar-looking classes. This hybrid method lets the model combine numerical and image-based components, hence enhancing its capacity to distinguish between closely similar variable star types (Figure 8).

The CNN architecture of the final model is rather standard following the trend of having deeper and smaller filters as the model progresses to extract more

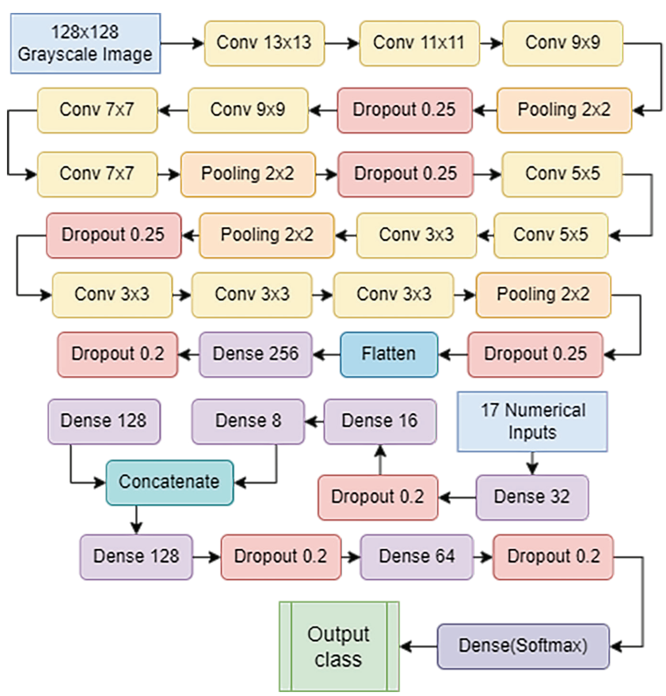

Figure 8. Flowchart of the multiple-input neural network architecture combining image-based and numerical features, followed by concatenation and dense layers for classification.

high-level features followed by 2 dense layers which are also standard. For the numerical input, it was found that having 3 layers before the concatenate allowed for the model to accurately capture all the trends from the numerical data and reducing the fully connected layers here would lead to insufficient use of the numerical data. Finally, an additional 2 fully connected layers were added before the final softmax layer which is necessary for a multiple class classification. All layers had a ReLu activation function to prevent vanishing gradients while not overcomplicating the model.

The numerical branch processes a set of astrophysical parameters taken from the light curve of every star. These attributes help the model to differentiate across classes with comparable visual traits by offering extra understanding of the fundamental physical traits of the stars.

The selected attributes encompass both Fourier-based features and general astrophysical properties, ensuring a comprehensive characterization of stellar variability. The Fourier-based features include the fundamental Fourier amplitude coefficient (A_1), which represents the dominant oscillation mode of the star, and the fundamental Fourier phase coefficient (Φ_1), which captures the phase shift within the pulsation cycle. To account for higher-order oscillations, normalized amplitude ratios (A_2–A_6 / A_1) were included to quantify the relative contribution of successive harmonics, while normalized phase differences (Φ_2–Φ_6 − Φ_1) describe the evolution of phase shifts across these harmonics. In addition, two skewness parameters were introduced: α (alpha), representing asymmetry between the rising and declining brightness segments, and β (beta), which captures secondary asymmetry effects that further refine the light curve morphology.

The general astrophysical properties complement these Fourier descriptors by providing broader physical context. These include the period, which defines the duration (in days) of one complete pulsation cycle; the mean magnitude, representing the star's average brightness over time; and the Fourier fit error (MSE), which quantifies the residual between the observed light curve and its Fourier model, serving as an indicator of the complexity or irregularity in the variability pattern. Together, these attributes provide a balanced and interpretable feature set for accurate classification of variable stars.

The integration of numerical and image-based features has shown promise in recent studies (see, e.g., (12) Eyer *et al*.). The numerical features are processed through a fully connected dense network consisting of multiple layers that refine the information and capture complex interactions between different astrophysical parameters. This branch operates in parallel to CNN and generates a numerical representation of each star.

Following each branch's respective feature extraction, they are concatenated to let the model use the morphological characteristics of the CNN and the astrophysical insights from the numerical branch.

Extra fully connected layers pass this mixed representation, where the network learns to balance both kinds of features before applying a softmax activation function for the final classification.

RESULTS

The MINN and CNN models were trained to classify variable stars into their respective types and subtypes. The results indicate that the combination of phase-folded light curve images and astrophysical parameters such as Fourier coefficients, skewness metrics, and period significantly enhanced classification accuracy for rare classes of stars. The model achieved an overall classification accuracy of approximately 90% after 10 epochs, demonstrating both efficiency and robustness. This overall figure reflects aggregate performance and is influenced by class imbalance; rare classes show lower accuracy despite targeted improvements, as evident in the confusion matrix.

Training Performance

The training and validation loss curves indicate that the model converges rapidly, with a significant drop in loss within the first 3 epochs before stabilizing. The validation loss remains slightly lower than the training loss throughout, suggesting that the model does not suffer from overfitting.

The accuracy curves (Figure 9A) show that validation accuracy surpasses 88% within just 3 epochs and continues to improve, reaching 89.8% by epoch 10. The training accuracy follows a similar trend, demonstrating that the model effectively learns from the dataset. The model used a 75-25 train test split.

Confusion Matrix Analysis

Figure 10 presents the confusion matrix, providing a detailed breakdown of classification performance across all variable star types and subtypes. This visualization highlights how well the model differentiates between classes and identifies areas of

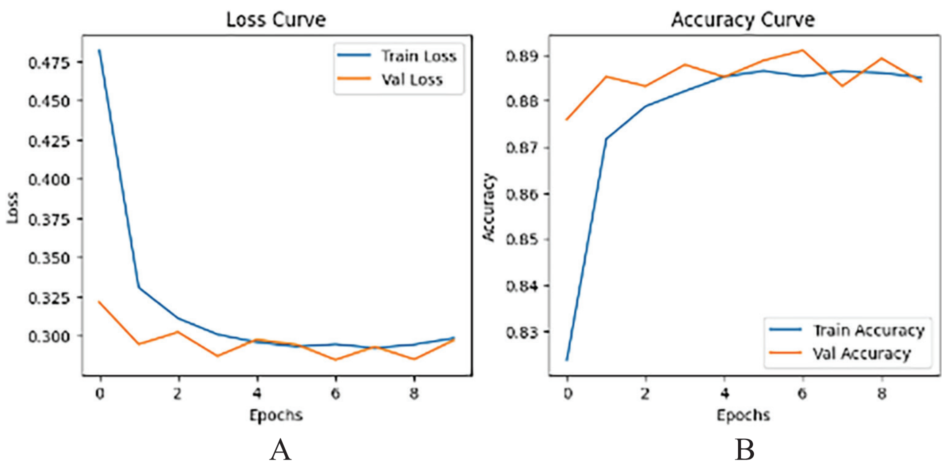

Figure 9. Training and validation loss and accuracy curves across epochs, demonstrating model convergence and generalization performance.

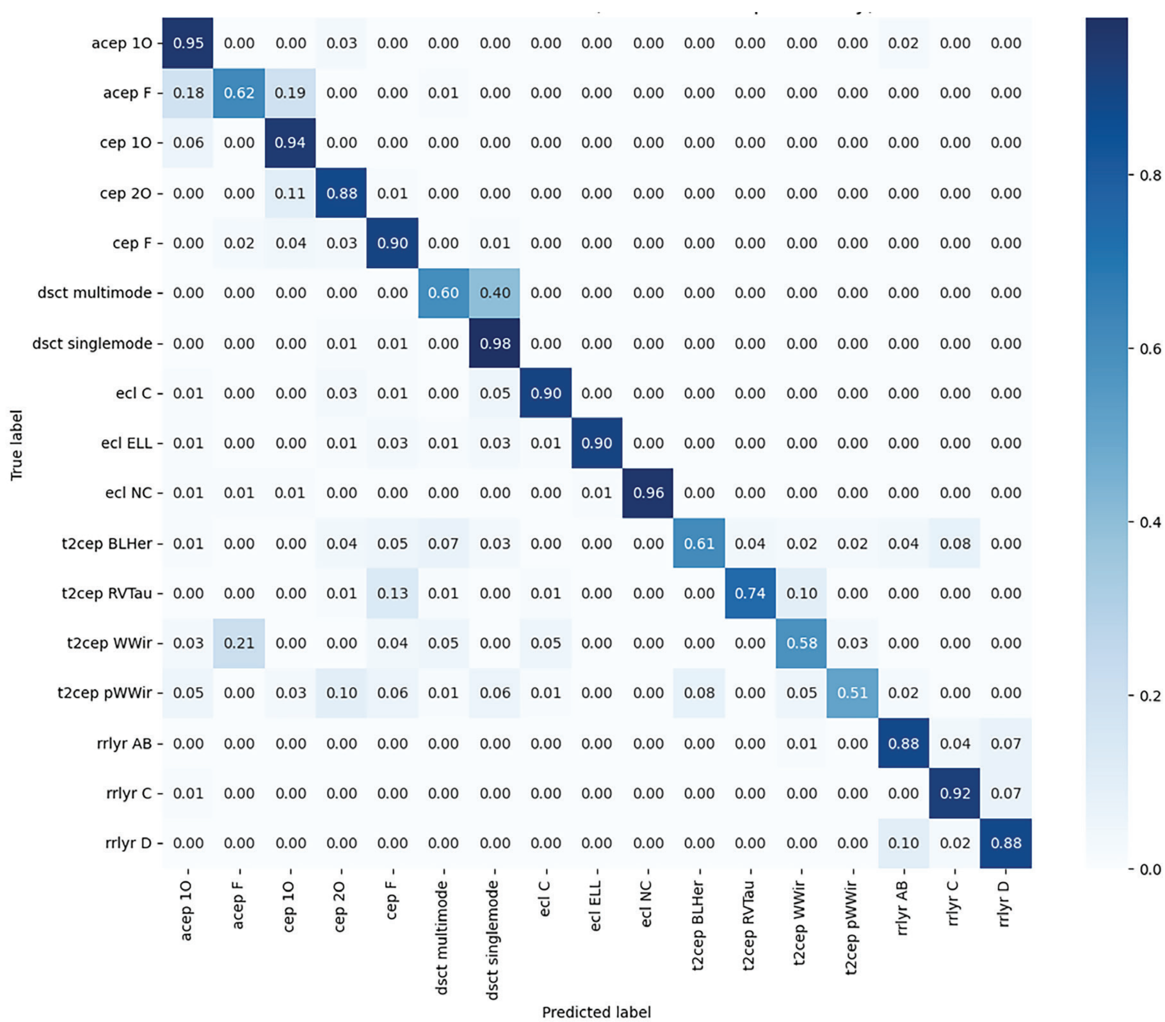

Figure 10. Final confusion matrix of the optimized model, summarizing classification performance across all variable star subclasses after hyperparameter tuning.

misclassification.

Among the best-classified stars, Classical Cepheids and RR Lyrae stars continue to achieve high accuracy levels, with over 89% correct classification. Their distinct period-luminosity relationships and well-documented variability make them relatively easier for the model to identify. Eclipsing binaries and single-mode δ Scuti stars are also well-classified, with minimal confusion, demonstrating the model's ability to differentiate between pulsating and non-pulsating stars.

However, misclassifications remain prominent in specific categories. Type II Cepheids, particularly W Virginis (WVir) and BL Herculis (BLHer) stars, continue to be misidentified as Classical Cepheids or into each other's subtypes. While accuracy has improved compared to the initial confusion matrix, the overlap between Type II Cepheid subtypes suggests that period-based classification alone is insufficient for differentiation. BLHer stars are frequently confused with Classical Cepheids, likely due to their overlapping periods, while WVir stars are misclassified as pWVir (peculiar W Virginis). This trend indicates that the boundaries between these subtypes are less well-defined than initially assumed.

A particularly striking issue is the misclassification of multimode δ Scuti stars, where nearly 40% of them are confused with other δ Scuti subtypes. While the model recognizes that these stars belong to the δ Scuti category, it struggles to determine whether they are single-mode or multimode. This suggests that additional astrophysical parameters, such as metallicity or secondary oscillation modes, could improve classification performance for this group.

Examining trends across different classification accuracies, three broad groupings emerge. High-accuracy groups include Classical Cepheids, RR Lyrae stars, and eclipsing binaries, all of which are well-represented in training datasets and have distinct variability features. A moderate-accuracy group consists of multimode δ Scuti stars and some Type II Cepheid subtypes, particularly BLHer and RVTau, which exhibit confusion among themselves but remain somewhat distinct from Classical Cepheids. Finally, lower-accuracy categories include peculiar W Virginis stars and some Anomalous Cepheids, which show frequent misclassification into Cepheids or RR Lyrae stars. These trends suggest that while well-represented stars are accurately classified, subtypes with overlapping variability properties remain a challenge.

Unexpected misclassifications further highlight areas for refinement. Anomalous Cepheids, particularly ACep F, are frequently misclassified as Cepheids, which is surprising given that they originate from different evolutionary pathways. Their light curves bear some resemblance to Cepheids, but their physical properties differ, suggesting that incorporating additional features such as metallicity or distance estimates could improve classification. Another anomaly is the misclassification of WVir stars into Classical Cepheids at a higher-than-expected rate. Although their light curves share some morphological similarities, their underlying pulsation mechanisms differ. This suggests that expanding the model to incorporate more complex period-luminosity relationships might improve differentiation.

These findings confirm that the addition of astrophysical parameters significantly enhances classification performance for rare variable star types. However, challenges remain in distinguishing subtypes with overlapping characteristics. The continued misclassification of Type II Cepheids and multimode δ Scuti stars suggests that further improvements could be achieved through enhanced feature engineering, refined synthetic data augmentation, and the inclusion of additional astrophysical priors. Future research should explore these possibilities to further improve the robustness and generalizability of machine learning models in variable star classification.

Overall accuracy works as an aggregate measure but lack subtleties across the various classes of the dataset. Model performance can be better examined if metrics like precision, recall, and F1-score are calculated for each class of the variable-star population. These metrics shed light on the model's ability to determine both the common and the rare types.

The macro-averaged F1-score somewhet elevated to 0.87, which shows that there is still a lack of balance between recall and precision, but the weighted-average F1-score, which is adjusted to class imbalance, reached 0.91. In fact, for the dominant classes—Classical Cepheid and RR Lyrae—the F1-scores were above 0.93 while the less common classes of Anomalous and Type II Cepheids F1-scores were 0.59 and 0.53, respectively. Such metrics continue to support the trends that were identified in the confusion matrix, which show that the model is strong on types that are well-represented in the dataset, but weak on the types that are underrepresented.

In order to better understand the effect of augmentation of the light-curve dataset on the performance of the model, statistical tests were conducted which compared augmentation to no augmentation. A paired t-test on per-class F1-scores across five random seeds produced p=0.018, which allows us to determine that the gain for rare classes (average delta F1 of around 0.09) is statistically significant at the 95% level. This improvement is also substantiated by bootstrapped macro-F1 delta confidence interval [0.035, 0.148].

These supplementary analyses ensure that the improvements in performance are not the result of chance fluctuation or random chance optimization. Instead, they result from an enhanced classifier that utilizes astrophysical constraints along with synthetic data to distinguish temporal and structural features of uniquely rare variable star classes.

DISCUSSION

The results of this study demonstrate the efficacy of a hybrid machine learning approach that combines CNNs for image-based classification with MINN with astrophysical parameters. By combining domain-specific knowledge with deep learning, the model greatly enhances accuracy, robustness, and adaptability in classifying variable stars, enabling the classifier to more effectively distinguish visually similar light curves using numerical metrics like Fourier amplitudes, period, and skewness for a more thorough analysis. After just 10 training epochs, the model achieved an overall classification accuracy of approximately 90%, outperforming traditional CNN-based methods that classify subclasses that rely solely on phase-folded light curve images.

Key Insights from the Results

One of the major contributions of this work is the significant improvement of identification of uncommon classes of variable stars, i.e., ACeps and Type II Cepheids (T2Cep). Past methods using machine learning tended to misclassify these classes due to their underrepresentation in training samples, and therefore lower precision in identifying these stars in real astronomical surveys.

Another intriguing finding is the high classification accuracy of Classical Cepheids (Cep) and RR Lyrae (RR Lyr) stars, which were more than 92% in terms of F1-score. These two types are two of the most extensively researched and well-represented variable star types and are hence ideal test subjects for model performance evaluation. The high classification accuracy of these stars is an indication of the efficacy of the phase-folded CNN approach, as these stars possess well-defined periodic pulsations with relatively low observational noise.

The confusion matrix analysis indicates that the most serious misclassification is between Classical Cepheids and Type II Cepheids, as evidenced by significant off-diagonal values. For instance, in our matrix, Cep F is misclassified as T2Cep WVir with a proportion of 0.13, while conversely, T2Cep WVir is misclassified as Cep F at 0.20, highlighting their overlapping light curve characteristics. This reciprocal confusion underscores the difficulty in separating these classes based solely on morphological features and emphasizes the need for additional astrophysical parameters to improve discrimination. Nevertheless, the addition of astrophysical parameters, including Fourier amplitude ratios, skewness coefficients (α, β), and period, served to mitigate this uncertainty by introducing further discriminative factors beyond the mere morphology of the light curve.

Comparison with Prior Studies

The classification accuracy of 90% achieved within 10 epochs surpasses previous deep learning-based models that relied solely on image-based classification which were trained with the same number of epochs; most models such as those in (7) Szklenár *et al.* and (8) Jayasinghe *et al.* that surpass this accuracy often train for upwards of 100 or 200 epochs. Past studies using Recurrent Neural Networks (RNNs) and time-series classifiers faced challenges such as missing data, irregular sampling, and class imbalances.

When compared directly to Szklenár *et al.* (2022), our MINN model demonstrates clear improvements for some rare classes but also highlights areas where further work is needed. For instance, Szklenár *et al.* reported only about 45% accuracy for first-overtone ACep, whereas our model reaches 0.59, and similarly improves the classification of T2Cep WVir from Szklenár's ~50% to 0.53. However, Szklenár's model still outperforms ours for some well-represented classes: their accuracy for Classical Cepheids fundamental mode (Cep F) ranges between 95–98%, while our model achieves about 0.88, and they report over 96% for single-mode δ Scuti stars compared to our 0.86. Moreover, for multimode δ Scuti stars, both models struggle, with Szklenár achieving ≤60% and our model around 0.40, suggesting this remains a challenging class regardless of the approach. Despite these differences, our hybrid model requires only 10 training epochs, significantly fewer than the 100–200 epochs used by Szklenár *et al*. and still manages comparable or better performance for several rare and underrepresented classes, emphasizing the benefits of integrating astrophysical parameters and synthetic augmentation.

In this paper, the new hybrid deep learning method solves these problems by including astrophysical information in its architecture. By making use of Fourier-based descriptions for light curves, enriching them with abundant astrophysical information, and benefiting from synthetic generation of data, not only does it excel in performance in terms of accuracy but in handling varied datasets as well, making it a promising tool for classifying variable stars in extensive surveys.

Impact of Synthetic Data Augmentation

This study's major accomplishment is the development of the Variable Star Light Curve Simulator. This tool is important because it creates realistic synthetic light curves for variable star types that lack sufficient data. A key challenge in classifying variable stars is the imbalance in training datasets. Some types, such as Classical Cepheids and RR Lyrae, have a lot of data, while others, like Anomalous Cepheids and Type II Cepheids, have much less. This imbalance skews the classifier to favor the well-represented types, resulting in poor performance when identifying those with limited data.

By means of Fourier coefficient distributions derived from real-world observational data, the synthetic light curves stay astrophysically consistent even if the sample size of rare classes is expanded. While adding synthetic data just 1.5% improved the general classifier accuracy, the classification accuracy for rare classes like ACeps and T2Ceph climbed noticeably. Since these stars are fundamental components of galactic structure modeling and cosmology, this focused enhancement has important consequences for studies on stellar evolution as well as cosmic distance measurements. The accuracy of ACeps and T2Cephs was much lower without synthetic data augmentation, staying between 0.15 and 0 but in the confusion matrix above one can see how most of these under-represented categories have more accuracy than 0.5. Simulated light curves allowed the algorithm to learn the unique variability traits of these stars, hence lowering misclassification errors and increasing classification confidence.

While the simulator strives to replicate real observational conditions closely, even subtle differences between synthetic and real light curves, such as unmodeled noise patterns, rare variability features, or survey-specific artifacts, can affect the model's performance and generalization. These discrepancies may lead to overconfidence in rare class predictions if the synthetic data fails to capture the true diversity of stellar behaviors. Future work should further validate synthetic light curves against unseen survey data and refine simulation methods to minimize these gaps.

It is not feasible to simulate the entire star dataset because generating high-fidelity synthetic light curves for millions of stars would demand substantial computational resources and time, especially given the complexity of Fourier fitting and noise modeling. Moreover, excessive reliance on synthetic data risks introducing biases or artifacts that could cause the model to learn features absent in real observations, potentially degrading performance on actual survey data. Therefore, synthetic data is used selectively to balance rare classes while preserving the integrity of the real dataset.

Future work could explore the integration of complementary simulation approaches, such as combining our Fourier-based light curve generator with Gaussian Process (GP)-based methods like those used by Szklenár *et al*. This would require careful harmonization of priors, variability assumptions, and class definitions, but could potentially yield hybrid training datasets that enhance classification accuracy across all star types. Additionally, ensemble learning or domain adaptation techniques may help bridge differences in simulator outputs while preserving physical interpretability.

Limitations, Challenges and Scope

While the model demonstrates high accuracy and robustness, certain challenges remain, particularly in classifying variable star types that exhibit strong overlap in their observational properties. Some of the primary limitations include the significant challenge from the overlap in period–luminosity characteristics between T2Cephs and Classical Cepheids, as both exhibit similar pulsation periods and brightness levels. This similarity makes it difficult to distinguish between them using light curve data alone. Incorporating additional features, such as metallicity indicators or multi-band photometric measurements, could provide the necessary separation in feature space to improve model performance for these closely related classes. Future research should focus on training models that generalize well across multiple surveys (e.g., OGLE, LSST, ASAS-SN) to ensure consistent classification accuracy despite variations in observation strategies.

Fine-Tuning for Subtype Classification: While the current model achieves high accuracy for broad variable star categories, classification at the subtype level—especially for multimode pulsators and hybrid stars—remains challenging. Enhancing feature engineering, exploring clustering methods in period-amplitude space, and integrating ensemble-based architectures could significantly improve subtype discrimination and lead to more nuanced astrophysical insights.

Implications for Future Research

This work has a solid foundation for large-scale automatic variable star classification and has important consequences for upcoming astronomical sky surveys such as the Legacy Survey of Space and Time, the All-Sky Automated Survey for Supernovae, and the Optical Gravitational Lensing Experiment.

Future research could focus on several promising directions. One avenue involves incorporating additional astrophysical data, such as multi-band photometry, the Weinstein Index, and other derived quantities from light curves, to enrich the feature space and improve classification precision. Another important direction is expanding and refining the synthetic data augmentation approach by enhancing the underlying mathematical models used to simulate light curves or leveraging greater computational power to optimize simulation accuracy. Applying transfer learning techniques across surveys such as ASAS-Zn and LSST, in conjunction with the OGLE4 dataset, would also help ensure stronger cross-survey adaptability. Furthermore, conducting a comprehensive sensitivity analysis could provide deeper insights into how various parameters influence classification performance. Finally, implementing targeted retraining strategies with adaptive feedback loops may help further reduce residual misclassifications while maintaining robust generalization across variable star types.

Using machine learning, insights from astrophysics, and artificial data creation can significantly improve how we classify stars, impacting our understanding of galaxy structures, star evolution, and the universe. This study lays the groundwork for more accurate and quicker categorization of stars that change brightness, known as variable stars. Future astronomy projects will be able to automatically and accurately sort millions of these stars, helping us explore and understand the universe more efficiently.

CONCLUSION

This study presents a hybrid deep-learning approach that combines CNN-based image classification with astrophysical parameters, resulting in an overall classification accuracy of 89.8% after just 10 epochs. By incorporating period, Fourier coefficients, and synthetic data augmentation, the model significantly improved classification performance, particularly for rare variable star types.

The key findings demonstrate high classification accuracy and strong transferability for major variable star types, including Classical Cepheids, RR Lyrae, and eclipsing binaries. Notably, the approach achieves significant improvements in the classification of rare classes, particularly Type II Cepheids and Anomalous Cepheids. Additionally, the study effectively mitigates class imbalance by employing synthetic light curve generation, enhancing the robustness and generalization of the model across diverse stellar populations.

This research demonstrates that machine learning models leveraging astrophysical domain knowledge can effectively classify variable stars at high accuracy levels. Future work should focus on further refining classification accuracy for overlapping subclasses, integrating additional astrophysical priors, and ensuring adaptability across multiple astronomical surveys.

This study paves the way for improving automated classification methods in large-scale surveys such as LSST, ASAS-SN, and Gaia, ultimately enhancing our ability to map and understand the variable star populations of the universe.

ACKNOWLEDGEMENTS

Thank you to Eleftheria Fassman from Cornell University for the guidance in the development of this research paper

CODE AVAILABILITY

The full code used is available at https://github.com/3mb3rr/VarStarClassifier. This model was trained on a NVIDIA RTX 3080 ti GPU with 16GB of RAM.

CONFLICT OF INTEREST

The author(s) declare that there are no conflicts of interest regarding the publication of this article.

REFERENCES

1. Leeuwen FV. Hipparcos, the New Reduction of the Raw Data. Astrophysics and Space Science Library, 2007, ISSN: 978-90-481-7612-0. https://doi.org/10.1007/978-1-4020-6342-8
2. Hogg DW. Distance measures in cosmology. Astrophysics. 1999; p. 1-16
3. Freedman WL, Madore BF. The Hubble Constant. *Annual Review of Astronomy and Astrophysics*. 2010; 48: 673-710. https://doi.org/10.1146/annurev-astro-082708-101829
4. Physics Department, University of California, Santa Barbara, CA, Period-luminosity relation of the Cepheid variable stars and its importance in measurement of cosmic distance Available from: https://www.deepspace.ucsb.edu/wp-content/uploads/2023/04/Period-luminosity-Relation-of-the-Cepheid-Variable-Stars-and-Its-Importance-in-Measurement-of-Cosmic-Distance.pdf (accessed on 2025-07-08)
5. Variable Star Astronomy. Available from: https://www.aavso.org/education/vsa (accessed on 2025-07-08)
6. Soszyński I, Udalski A, Szymanowski MK, Wyrzykowski Ł, *et al*. The OGLE Collection of Variable Stars. Classical, Type II, and Anomalous Cepheids toward the Galactic Center. Acta Astronomica. 2017; 67: 297-316
7. Szklenár T, Bódi A, Tarczay-Nehéz D, Vida K, *et al*. Variable Star Classification with a Multiple-input Neural Network. *The Astrophysical Journal*. 2022; 938 (1): 37. https://doi.org/10.3847/1538-4357/ac8df3
8. Jayasinghe T, Stanek K, Kochanek C, Shappee B, *et al*. The ASAS-SN Catalog of Variable Stars II: Uniform Classification of 412,000 known variables. 2018. https://www.semanticscholar.org/paper/The-ASAS-SN-Catalog-of-Variable-Stars-II%3A-Uniform-Jayasinghe-Stanek/d61cdccb8a412c97d7e06d81c22eb101a67dfa1c
9. Liu A, Xu JC. Machine learning-based classification of variable stars. *Research Notes of the AAS*. 2024; 8 (3): 83. https://doi.org/10.3847/2515-5172/ad34b2
10. Becker I, Pichara K, Catelan M, *et al*. Scalable end-to-end recurrent neural network for variable star classification. *Monthly Notices of the Royal Astronomical Society*. 2020; 493 (2): 2981-2995. https://doi.org/10.1093/mnras/staa350
11. Kang Z, Zhang Y, Zhang J, Li C, *et al*. Periodic Variable Star Classification with Deep Learning: Handling Data Imbalance in an Ensemble Augmentation Way. *Publications of the Astronomical Society of the Pacific*. 2023; 135 (1051): 1-10. https://doi.org/10.1088/1538-3873/acf15e
12. Eyer L, Audard M, Holl B, Rimoldini L, *et al*. GaiaData Release 3. *Astronomy and Astrophysics*. 2023; 674 (13): 1-19. https://doi.org/10.1051/0004-6361/202243626