Comparing Neural, Margin-Based, and Tree-Based Models for Stellar Classification

Tathagat Agrawal

Abstract

This study investigates the use of three supervised learning algorithms—Multilayer Perceptron (MLP), Support Vector Classifier (SVC), and Decision Tree (DT)—to classify six types of stars based on stellar parameters (temperature, luminosity, radius, absolute magnitude, colour, and spectral class). Using a standard train/test split on the publicly available "Stars" dataset, features were preprocessed via label encoding and standard scaling, and each model was trained under comparable conditions. Performance was evaluated through loss curves, confusion matrices, and standard metrics (accuracy, precision, recall, and F1-score).

1. Introduction

Morphological and spectral classification of stars has guided astrophysical research since the early 20th century, following the development of the Hertzsprung-Russell diagram (or the H-R diagram) which revealed relationships between temperature and luminosity (Hertzsprung, 1905; Russell, 1919). The H-R diagram, which is often plotted with effective temperature along the horizontal axis and luminosity along the vertical axis (Figure 1), demonstrates that different stellar classes occupy distinct loci within this two-dimensional space. Main-sequence stars form a well-defined diagonal band; white dwarfs cluster in a lower-left region of high temperature but low luminosity; supergiants and hypergiants reside in the upper-right.

In practice, astronomers classify a star by determining its position on the H-R diagram and assigning the corresponding class. However, as contemporary surveys such as Gaia and SDSS deliver photometric and spectroscopic measurements for millions of stars, manual plotting and visual classification have become prohibitively time-consuming. Since the underlying task reduces to mapping a vector of astrophysical features (temperature, luminosity, radius, absolute magnitude, colour, and spectral class) to a finite set of class labels, it naturally lends itself to supervised learning. Machine learning algorithms can learn this mapping directly from data, automating classification at scale while preserving or even improving upon the accuracy of manual methods.

This study evaluates three representative algorithms—a feed-forward neural network (MLP), a margin-based classifier (SVC), and a tree-based model (DT)—on a standardized six-class stellar dataset, with the aims of quantifying classification accuracy, examining overfitting dynamics, and offering guidance for hyperparameter selection in astronomical contexts.

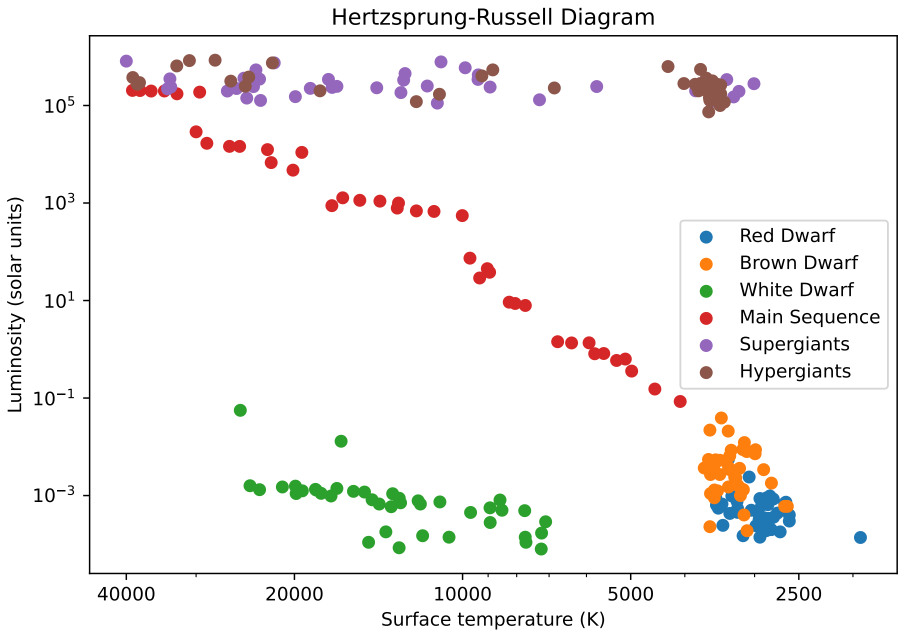

Figure 1: H-R diagram for the "Stars" dataset

2. Methodology

2.1. Dataset and preprocessing

This study utilized the publicly available "Stars" dataset (Baidya, 2019), comprising 240 samples distributed across six stellar classes: Brown Dwarf, Red Dwarf, White Dwarf, Main Sequence, Supergiants, and Hypergiants. Each instance is characterized by six features: four continuous variables—Temperature, Luminosity, Radius, and Absolute Magnitude—and two categorical variables—Colour and Spectral Class.

Prior to model training, preprocessing was conducted in three key stages. First, categorical features were numerically encoded via label encoding, mapping discrete classes to integers while preserving their nominal status. Second, continuous features were standardized using z-score normalization, defined as:

$$z = \frac{x - \mu}{\sigma}$$

where x is the feature value, μ the sample mean, and σ the sample standard deviation. This transformation ensures that each feature has zero mean and unit variance, promoting numerical stability in optimization. Finally, the dataset was partitioned into training and test sets using an 80/20 stratified split to maintain class proportions across splits, reducing sampling bias.

2.2. Models

Three supervised learning models were implemented: a Multilayer Perceptron (MLP), a Support Vector Classifier (SVC), and a Decision Tree (DT). Each model was trained under comparable experimental conditions.

The MLP architecture consisted of two hidden layers, each containing 64 neurons activated by the Rectified Linear Unit (ReLU) function, with a softmax output layer for multi-class classification. Training minimized the cross-entropy loss using the Adam optimizer. Hyperparameters were set as follows: learning rate $\eta = 0.001$, L_2 regularization coefficient $\alpha = 0.0001$, batch size of 32, and a maximum of 50 iterations.

The SVC utilized a Radial Basis Function (RBF) kernel defined by:

$$K(x, x') = \exp\Big(-\gamma \|x - x'\|^2\Big)$$

with the kernel coefficient defined as

$$\gamma = \frac{1}{n \cdot \mathrm{Var}(X)}$$

for $n = 6$ features. The penalty parameter $C = 1.0$ controlled the trade-off between margin maximization and classification error.

The Decision Tree classifier was trained using Gini impurity as the splitting criterion. No pruning was applied, allowing the tree to expand until pure leaves or exhaustion of splits.

2.3. Evaluation metrics

For a set of predictions $\{\hat{y}_i\}_{i=1}^m$ on the test set ($m = 48$) and ground truths $\{y_i\}_{i=1}^m$, model performance was assessed via accuracy, precision, recall, and F1-score:

$$\begin{aligned}
\text{accuracy} &= \frac{1}{m} \sum_i \mathbb{I}[\hat{y}_i = y_i] \\
\text{precision} &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \\
\text{recall} &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \\
\text{F1-score} &= 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}.
\end{aligned}$$

3. Results and Discussion

3.1. MLP results

The MLP model achieved 100% test accuracy, as shown in Table 1 and the confusion matrix (Figure 2). Every sample was correctly classified, with per-class precision, recall, and F1-scores all equal to unity. The loss curve (Figure 3) exhibits smooth exponential decay, characteristic of well-conditioned optimization under Adam, indicating stable convergence.

Table 1: Performance metrics for the MLP

	precision	recall	f1-score	support
Brown Dwarf	1.00	1.00	1.00	8
Hypergiants	1.00	1.00	1.00	8
Main Sequence	1.00	1.00	1.00	8
Red Dwarf	1.00	1.00	1.00	8
Supergiants	1.00	1.00	1.00	8
White Dwarf	1.00	1.00	1.00	8
accuracy			1.00	48
macro avg	1.00	1.00	1.00	48
weighted avg	1.00	1.00	1.00	48

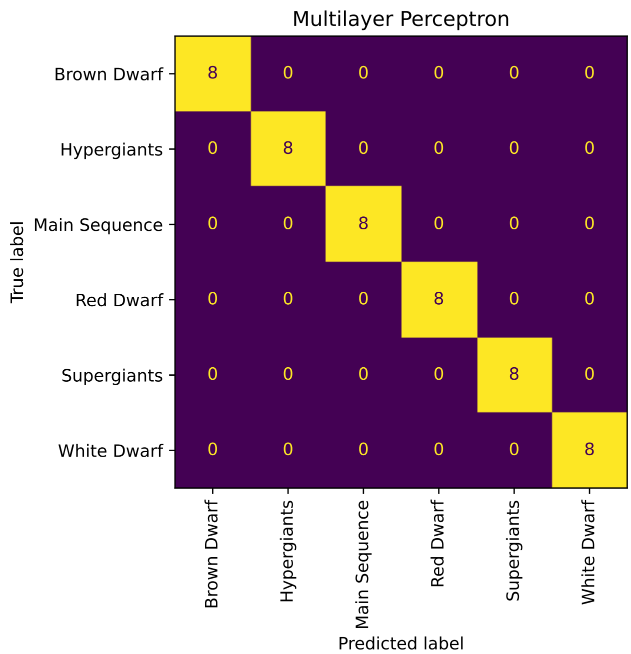

Figure 2: Confusion matrix for the MLP

3.2. SVC results

The SVC classifier achieved a slightly lower test accuracy of 95.83%. As shown in Table 2 and the confusion matrix (Figure 4), most classes were predicted accurately. However, minor confusion was observed between Supergiants and Main Sequence stars, possibly due to overlapping feature distributions in the transformed space. The precision and recall metrics remained robust across classes, though the recall and F1-score for Supergiants notably dipped to 0.75 and 0.86, respectively.

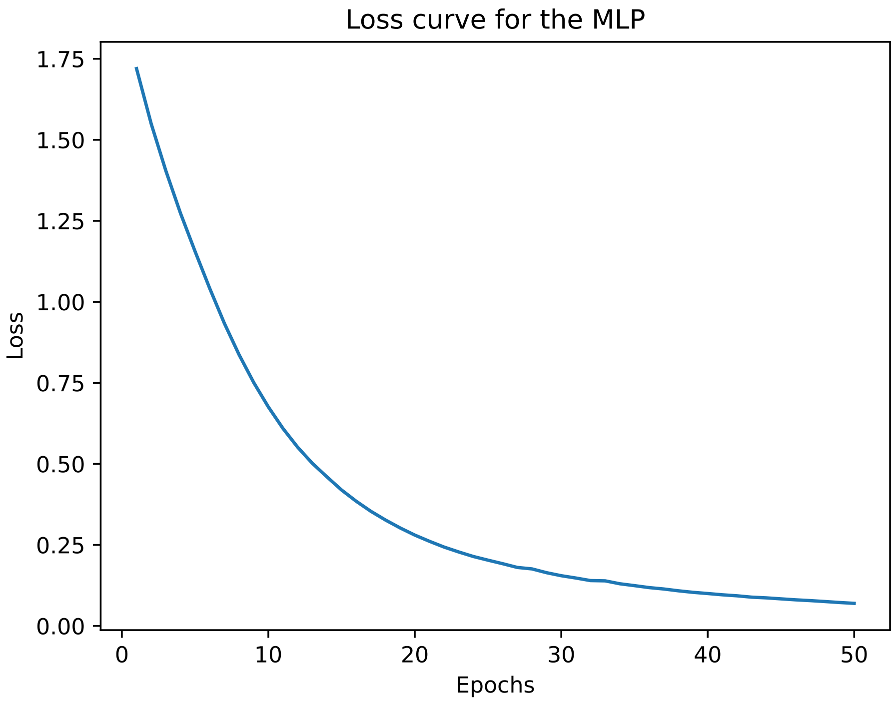

Figure 3: MLP loss curve

Table 2: Performance metrics for the SVC

	precision	recall	f1-score	support
Brown Dwarf	1.00	1.00	1.00	8
Hypergiants	1.00	1.00	1.00	8
Main Sequence	0.80	1.00	0.89	8
Red Dwarf	1.00	1.00	1.00	8
Supergiants	1.00	0.75	0.86	8
White Dwarf	1.00	1.00	1.00	8
accuracy			0.96	48
macro avg	0.97	0.96	0.96	48
weighted avg	0.97	0.96	0.96	48

3.3. DT results

The DT model achieved 100% test accuracy, as shown in Table 3 and the confusion matrix (Figure 5). Every sample was correctly classified, with per-class precision, recall, and F1-scores all equal to unity. It is quite interesting to note that the decision tree only considered the absolute magnitude and temperature as the distinguishing features (Figure 6).

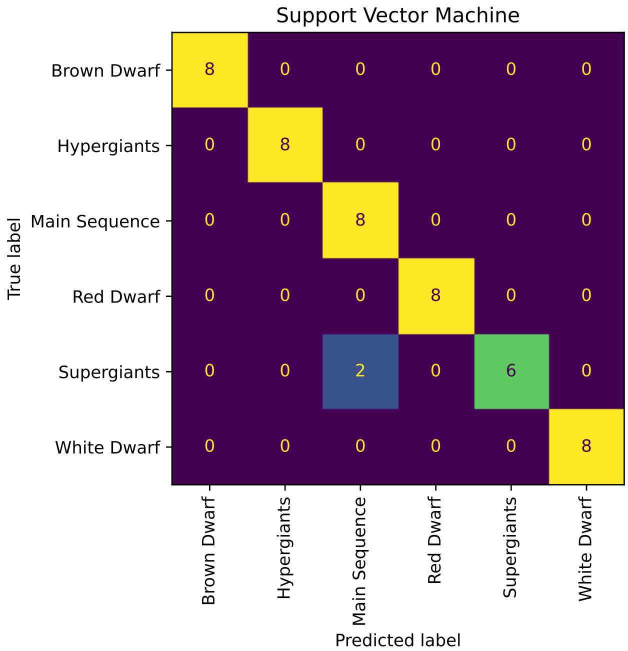

Figure 4: Confusion matrix for the SVC

Table 3: Performance metrics for the DT

	precision	recall	f1-score	support
Brown Dwarf	1.00	1.00	1.00	8
Hypergiants	1.00	1.00	1.00	8
Main Sequence	1.00	1.00	1.00	8
Red Dwarf	1.00	1.00	1.00	8
Supergiants	1.00	1.00	1.00	8
White Dwarf	1.00	1.00	1.00	8
accuracy			1.00	48
macro avg	1.00	1.00	1.00	48
weighted avg	1.00	1.00	1.00	48

3.4. Comparative discussion

The key performance metrics of the three models are summarized in Table 4. It is clear that the MLP and the DT achieved better performance than the SVC. However, due to the small dataset size, it is possible that these models have memorized the training distribution.

4. Conclusions and Future Work

This study evaluated three standard supervised learning models—MLP, SVC, and DT—on a six-class stellar classification task using real astrophysical features.

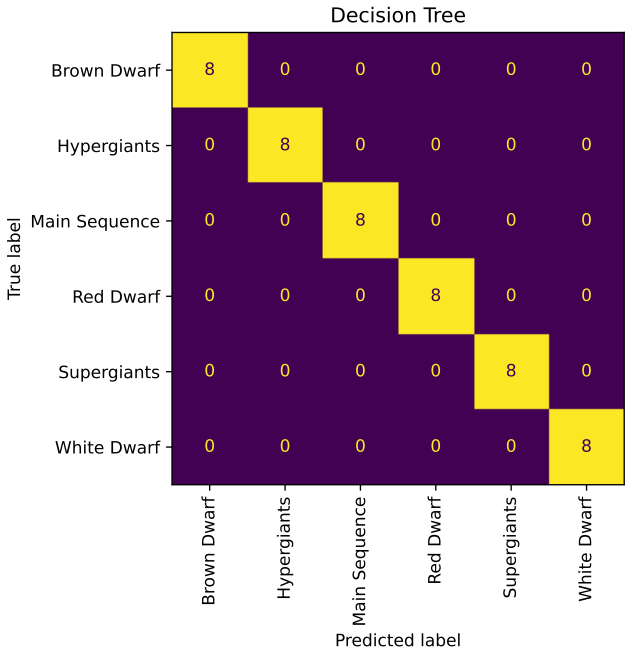

Figure 5: Confusion matrix for the DT

Table 4: Model performance comparison on the test set

Model	Accuracy	Precision	Recall	F1-score
MLP	1.0000	1.0000	1.0000	1.0000
SVC	0.9583	0.9666	0.9583	0.9576
DT	1.0000	1.0000	1.0000	1.0000

While the decision tree achieved perfect test accuracy, its unpruned structure suggested severe overfitting. The MLP and SVC, by contrast, demonstrated robust generalization, with the MLP slightly outperforming the SVC in terms of accuracy and balanced class-wise performance.

Future work should explore the use of ensemble methods, such as Random Forests or Gradient Boosting, to mitigate overfitting while leveraging the interpretability of tree-based models. Moreover, hyperparameter optimization through grid search or Bayesian optimization could systematically improve model selection. Further experiments incorporating additional astrophysical features, synthetic augmentation, and more challenging datasets could also enhance model robustness and applicability.

A promising direction would involve the use of convolutional neural networks (CNNs) directly on stellar spectra or photometric time-series data, enabling the models to learn hierarchical feature representations beyond manually engineered variables.

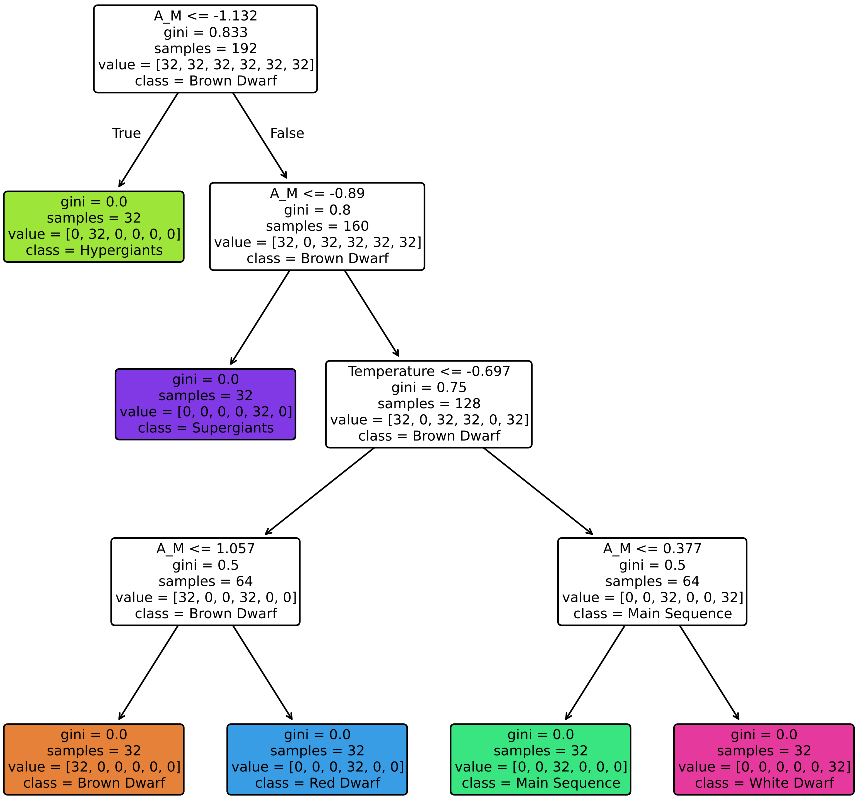

Figure 6: Decision tree

5. Related Work

The application of machine learning techniques to stellar classification tasks has a rich history, reflecting the need for scalable and accurate alternatives to manual spectral analysis. Early efforts, such as those by Bailer-Jones (1997), demonstrated the feasibility of using artificial neural networks (ANNs) for stellar spectral type prediction, establishing that data-driven approaches could rival traditional classification accuracy.

Subsequent studies expanded the use of support vector machines (SVMs) in astrophysical settings. Zhang and Zhao (2004) applied SVMs to classify stars and galaxies in large astronomical surveys, leveraging the algorithm's ability to model complex, non-linear decision boundaries. Their work highlighted the strength of margin-based classifiers in handling noisy and high-dimensional astrophysical data.

Decision trees and ensemble methods have also found application in this domain. Ball et al. (2006) used decision trees and random forests for automated classification tasks in the Sloan Digital Sky Survey (SDSS), emphasizing the interpretability of tree-based models, which is often a critical requirement in scientific analyses. However, these methods were also found to be susceptible to overfitting if not properly regularized.

More recent work has incorporated deep learning approaches. For instance, Fabbro et al. (2018) utilized convolutional neural networks (CNNs) directly on raw spectroscopic data for star-galaxy classification, demonstrating that hierarchical feature learning can outperform manually engineered features. Such methods, however, typically require much larger datasets than conventional models like SVCs or decision trees.

In comparison to these prior works, the present study systematically evaluates three distinct classes of supervised learning models—MLP (neural), SVC (margin-based), and DT (tree-based)—on a relatively small, well-defined stellar dataset. Unlike many previous studies that focused solely on accuracy, this work also critically examines overfitting behaviour and generalization dynamics, offering deeper insights into the trade-offs inherent in model selection for astronomical classification tasks.

References

Baidya, Deepraj (2019). *Star dataset to predict star types.* `https://www.kaggle.com/datasets/deepu1109/star-dataset`. Kaggle.

Bailer-Jones, Coryn AL (1997). "Neural network classification of stellar spectra". In: *Publications of the Astronomical Society of the Pacific* 109.738, p. 932.

Ball, Nicholas M et al. (2006). "Robust machine learning applied to astronomical data sets. I. star-galaxy classification of the Sloan Digital Sky Survey DR3 using decision trees". In: *The Astrophysical Journal* 650.1, p. 497.

Fabbro, S et al. (2018). "An application of deep learning in the analysis of stellar spectra". In: *Monthly Notices of the Royal Astronomical Society* 475.3, pp. 2978–2993.

Hertzsprung, Ejnar (1905). "Zur Strahlung Der Strerne". In: *Zeitschrift fur wissenschaftliche Photographie* 3, pp. 442–449.

Russell, Henry Norris (1919). "Some problems of sidereal astronomy". In: *Proceedings of the National Academy of Sciences* 5.10, pp. 391–416.

Zhang, Yanxia and Yongheng Zhao (2004). "Automated clustering algorithms for classification of astronomical objects". In: *Astronomy & Astrophysics* 422.3, pp. 1113–1121.