Grenze International Journal of Engineering and Technology, January Issue

Deep Learning for Spectral Typing of Stars: A CNN Framework on Sloan Digital Sky Survey Observations

Sowmya R B[1], Sanjana M Nagaraj[2] and Sowjanya M Nagaraj[3]
[1]Department of Information Science and Engineering, Siddaganga Institute of Technology, Tumkur, India
Email: sowmyabnaik19@gmail.com
[2]Department of Computer Science and Business Systems, Dayananda Sagar College of Engineering, Bangalore, India
Email: sanjanamnagaraj18@gmail.com
[3]Department of Management Studies, Dayananda Sagar College of Arts, Science and Commerce, Bangalore, India
Email: mnagarajsowjanya@gmail.com

***Abstract*—We present a Convolutional Neural Network (CNN) model for automated stellar classification using data from the Sloan Digital Sky Survey (SDSS) DR17. The proposed model achieved a precision of 0.94, recall of 0.91, and F1-score of 0.92, surpassing traditional approaches such as Random Forest (F1-score 0.83). The CNN demonstrated robustness under noisy conditions (F1-score 0.89) and achieved superior accuracy for M-type stars (F1-score 0.96), while minor confusion between G- and K-types was observed due to color similarities. Feature attribution using SHAP values revealed that g−r and r−i indices were the most influential, reflecting their physical correlation with stellar temperature. The model's generalization was confirmed through 5-fold cross-validation (mean F1-score 0.91 ± 0.02) and high AUC scores (0.90–0.98). These results highlight the potential of CNN-based methods for scalable, explainable, and reliable star classification, paving the way for integration with Gaia, WISE, and LSST datasets for advanced astronomical analyses.**

***Index Terms*— Convolutional Neural Network, Stellar Classification, SDSS DR17, Explainable AI, Feature Importance, Astronomical Surveys.**

I. INTRODUCTION

Imagine standing under a night sky, each star a unique book in a cosmic library, waiting to be sorted by its story its temperature, size, and chemical makeup. Astronomers have long classified stars into spectral types (O, B, A, F, G, K, M), from scorching O-types to cool M-types, using spectra that act like a star's fingerprint [1]. In the early 20th century, pioneers like Annie Jump Cannon manually sorted thousands of stars by eye, a heroic but slow task. Today, surveys like the Sloan Digital Sky Survey (SDSS) catalog millions, making manual methods as outdated as a horse-drawn carriage [2]. Photometric data brightness in colors like ultraviolet (u), green (g), and red (r) offer a faster way, like judging a book by its cover, but they're trickier due to noise and overlapping traits. This is where artificial intelligence (AI) steps in, like a super-smart librarian. Deep learning, especially convolutional neural networks (CNNs), can spot patterns in data faster than any human, making sense of messy photometric measurements without needing a PhD in astrophysics [3]. Picture a CNN as a librarian who doesn't just skim titles but understands the essence of each star's light to sort it perfectly.

Grenze ID: 01.GIJET.12.1.124

In this case study, we train a CNN to classify 7,000 SDSS DR17 stars, tackling challenges like noisy data, rare O-type stars, and confusion between similar types like G and K. Our work is a stepping stone for future surveys like LSST, which will flood us with petabytes of data over its 10-year mission, transforming our understanding of the cosmos [4]. This paper is crafted for the AI in Astronomy track at AICARE 2025, where we're excited to show how AI can make astronomy more accessible [5]. We'll guide you through our approach, share how well it works with clear metrics and visuals, and explore what it means for astronomers and enthusiasts alike. From mapping galaxies to inspiring citizen scientists, this is about bringing the stars closer to home.

The journey of stellar classification began with visual inspections, where astronomers like Cannon painstakingly analyzed photographic plates. Her work at Harvard led to the OBAFGKM system, still used today [6]. Modern surveys like SDSS and Gaia have shifted the paradigm, producing datasets too vast for human analysis. AI bridges this gap, automating classification while preserving the precision of traditional methods, setting the stage for discoveries in galactic evolution and stellar populations.

The primary objective of this study is to develop a robust and explainable deep-learning framework using Convolutional Neural Networks (CNNs) for automated classification of stellar spectral types from Sloan Digital Sky Survey (SDSS) DR17 photometric data. Unlike traditional spectrum-based or feature-engineered machine-learning approaches, the proposed model leverages multi-band magnitudes and derived color indices to deliver fast, scalable, and noise-tolerant stellar classification suitable for modern large-volume astronomical surveys. The novelty of this work lies in integrating tabular photometric inputs with a carefully optimized CNN capable of capturing subtle color gradients, particularly effective in distinguishing challenging classes such as G- and K-type stars. Additionally, the incorporation of SHAP-based explainability provides transparent feature attribution—highlighting the physical significance of g−r and r−i indices and establishes a bridge between deep learning outputs and astrophysical interpretation. The study also introduces robustness testing through synthetic noise injection, cross-validation for stability assessment, and comparative benchmarking against Random Forest models, collectively presenting a comprehensive and uniquely interpretable pipeline for scalable stellar classification. This framework sets the groundwork for cross-survey generalization with Gaia, WISE, and future LSST datasets, marking an advancement toward automated, interpretable AI in astronomy.

II. Related Work

Sorting stars has come a long way from squinting at spectra through telescopes. This classic method relied on manual inspection, like handwriting a novel for each star [7]. With SDSS's millions of stars, that's a non-starter [8]. Early machine learning stepped up, using tools like decision trees, support vector machines, or random forests to sort stars by their colors, but these often tripped over noisy data or needed heavy tweaking, hitting accuracies around 80–85% [9]. Zhang et al. used a CNN on LAMOST spectra, achieving 95% accuracy by catching subtle patterns in light [10]. Solano applied a neural network to Gaia DR3, sorting young stars by blending position and brightness data, showing AI can handle mixed datasets [11]. Goddard used transfer learning to tweak SDSS-trained models for smaller surveys, boosting accuracy for rare stars [12]. Ramachandra used unsupervised methods, like autoencoders, to spot oddballs like carbon stars, while hybrid CNN-RNN models tackle variable stars that flicker over time [13]. Recent work explores graph neural networks for clustering stars in complex datasets and attention mechanisms for prioritizing key features [14]. Recent advancements have increasingly focused on leveraging photometric data from large surveys like SDSS to classify stellar spectral types, addressing the limitations of spectroscopic methods, which are resource intensive and unavailable for the majority of observed stars. These approaches often treat photometric measurements (e.g., multi-band magnitudes or derived color indices) as inputs, either in tabular form or as processed images, to enable scalable classification without full spectra.

Shi et al. (2023), who developed a convolutional neural network (CNN)- based Stellar Classification Network (SCNet) specifically for photometric images from SDSS. Their model processes cutout images in the g, r, and i bands to classify stars into spectral subtypes, achieving high accuracy on a massive scale. By training on a subset of spectroscopically confirmed stars and applying it to spectroscopically observed objects, they produced a new catalog of spectral types for approximately 50 million SDSS stars, demonstrating CNNs' ability to handle image-based photometric data and mitigate issues like interstellar dust extinction through data augmentation techniques. This extends earlier spectrum-focused CNNs, to purely photometric regimes, with reported accuracies exceeding 90% for common types while highlighting challenges in rare subclasses. Yi et al. (2024) introduced a Vision Transformer (ViT)-based model, stellar-ViT, for classifying stars using SDSS photometric images. Unlike traditional CNNs, ViT excels at capturing global dependencies in image patches, leading to improved performance over CNN baselines in handling noisy or low-resolution photometric data. Trained on SDSS DR17 images, their approach achieved F1-scores above 0.93 across OBAFGKM types, particularly benefiting from self-attention mechanisms to focus on discriminative features like color gradients. This work also incorporated transfer learning from pre-trained image models, reducing the need for large labeled datasets and addressing class imbalance in rare hot stars (O and B types). It aligns with the emphasis on explainability, as attention maps provide insights similar to SHAP values.

A. Bairouk et al. (2023) [15] includes the probabilistic models for parameter estimation, and proposed a hybrid of convolutional vision transformers with Bayesian inference, to directly derive stellar parameters (e.g., temperature, metallicity) and spectral types from SDSS photometric images. By incorporating uncertainty quantification, their method outperforms deterministic CNNs in noisy conditions, with validation on SDSS DR16 data showing mean absolute errors below 5% for temperature estimates tied to spectral classes. This is particularly relevant for cross-survey applications, as it demonstrates robustness to variations in photometric filters, a challenge noted earlier.

These studies underscore ongoing trends toward image-based photometric analysis, multimodal integration (e.g., combining photometry with astrometry), and scalable AI for upcoming surveys like LSST. Challenges remain: models must work across telescopes, handle rare stars, and explain their decisions clearly. Techniques like data augmentation, ensemble learning, and SHAP values are making strides.

III. Methodology

Imagine sorting 7,000 stars into their spectral types, like organizing a galactic card catalog. We used SDSS DR17 data, with 1,000 stars per type (O, B, A, F, G, K, M) from the Kaggle SDSS17 dataset [16]. Each star's colors (u, g, r, i, z), plus redshift, give clues to its identity. Our CNN had to navigate noise, tell apart similar G and K stars, and handle rare O-types. Here's the story of how we did it.

A. Preprocessing

Preprocessing is like prepping ingredients for a stellar dish:

- Normalization: Scale magnitudes to 0–1, making them easier for the CNN to chew on.
- Color Indices: Calculate u-g, g-r, r-i, i-z, like mixing spices to highlight differences (e.g., g-r 0.6 for G, 0.8 for K).
- Outlier Removal: Toss out 2% of data with wild errors (beyond 3 standard deviations) to keep things clean.
- Data Augmentation: For rare O and B stars, generate extra samples using SMOTE and slight tweaks (Gaussian noise, standard deviation 0.05), like photocopying rare manuscripts.
- Feature Selection: Focus on magnitudes and color indices, skipping ascension and declination, which don't help much. PCA was tested but clouded the story without boosting performance.

This ensures our data is ready for the CNN's magic.

B. Model Architecture

Our CNN is like a cosmic librarian scanning for patterns:

- Input Layer: Takes a 7-dimensional vector (u, g, r, i, z, g-r, r-i) and preps it for sorting.
- Convolutional Layers: Three layers (32, 64, 128 filters, kernel size 3, ReLU activation) spot patterns, like how g-r separates star types.
- Pooling Layers: Max-pooling (size 2) simplifies data, keeping the best bits.
- Regularization: Batch normalization steadies training, and dropout (0.3 rate) prevents memorizing the data.
- Fully Connected Layers: Two layers (256, 128 units) tie it together, with a softmax output picking the spectral type.

We use categorical cross-entropy loss and the Adam optimizer (learning rate 0.001) to fine-tune our star sorter.

C. Training and Evaluation

We split the data: 80% training (5,600 stars), 10% validation (700), and 10% testing (700). Training runs for 100 epochs, stopping early if it plateaus (patience 10 epochs). We measure success with precision, recall, F1-score, ROC curves, and cross-validation, focusing on tricky G/K mix-ups and rare O-types. Visuals like confusion matrices, F1-score plots, and feature importance charts make the results crystal clear.

D. Robustness Testing

To test grit, we add synthetic noise (Gaussian, standard deviation 0.1) to 10% of the test set, like static on a telescope. The CNN holds strong with an F1-score of 0.89, beating Random Forest's 0.80 [17]. Dropping the z-band still yields an F1-score of 0.87, showing it can handle missing data.

E. Hyperparameter Optimization

We tuned the model like tweaking a recipe, testing:

- Learning Rate: 0.0005, 0.001, 0.005.
- Dropout Rate: 0.2, 0.3, 0.4.
- Filter Sizes: 16, 32, 64.

The best combo (learning rate 0.001, dropout 0.3, 32 filters) hits an F1-score of 0.92, as shown in Table I.

F. Model Explainability

To peek inside the CNN's brain, we use SHAP values to see which features drive decisions. This is like asking the librarian why they shelved a book in a certain section [18]. Fig. 1 shows g-r and r-i indices leading the charge, making our model's choices transparent and trustworthy. We also explored layer-wise relevance propagation, confirming similar feature priorities [19].

TABLE I: OVERALL PERFORMANCE METRICS ON TEST SET

Metric	Value
Precision Recall F1-Score	0.94 0.91 0.92

TABLE II: PERFORMANCE UNDER SYNTHETIC NOISE

Model	Precision	Recall	F1-Score
CNN Random Forest	0.91 0.82	0.88 0.79	0.89 0.80

TABLE III: 5-FOLD CROSS-VALIDATION RESULTS

Fold	Precision	Recall	F1-Score
Fold 1	0.93	0.90	0.91
Fold 2	0.94	0.91	0.92
Fold 3	0.92	0.89	0.90
Fold 4	0.93	0.91	0.92
Fold 5	0.92	0.90	0.91
Mean	0.93	0.90	0.91
Std. Dev.	0.01	0.01	0.02

IV. EXPERIMENTAL RESULTS

Our Convolutional Neural Network (CNN) was trained and evaluated on a dataset of 7,000 stars from SDSS DR17, demonstrating strong classification performance across multiple spectral types. The model achieved an overall precision of 0.94, recall of 0.91, and F1-score of 0.92 (Table II), confirming its reliability in distinguishing stellar categories.

Compared to the Random Forest baseline by Hartman et al., which recorded an F1-score of 0.83, the CNN exhibited a clear improvement across all metrics (Table VI). The confusion matrix (Fig. 3) reveals that most misclassifications occurred between G- and K-type stars, likely due to their similar optical colors and overlapping g-r indices. This limitation was partially mitigated by incorporating r-i and i-z features, which reduced misclassification rates by approximately 5%. The per-class analysis (Table V) further shows that while M-type stars achieved the highest F1-score (0.96), O-type stars performed relatively lower (0.88), primarily due to their scarcity in the training data (Fig. 4). Under noisy input conditions (Gaussian noise, $\sigma = 0.1$), the CNN maintained robustness with an F1-score of 0.89, outperforming the Random Forest's 0.80 (Table III). This demonstrates the model's resilience in realistic observational environments. The ROC curves (Fig. 2) show high AUC values ranging from 0.90 (O-type) to 0.98 (M-type), reflecting excellent discriminative capability across classes.

TABLE IV: PER-CLASS PERFORMANCE METRICS

Spectral Type	Precision	Recall	F1-Score
O	0.89	0.87	0.88
B	0.91	0.89	0.90
A	0.93	0.91	0.92
F	0.94	0.92	0.93
G	0.92	0.90	0.91
K	0.91	0.89	0.90
M	0.97	0.95	0.96

TABLE V: HYPERPARAMETER TUNING RESULTS

Learning Rate	Dropout	Validation Accuracy	F1-Score
0.0005	0.3	0.90	0.89
0.001	0.3	0.93	0.92
0.005	0.3	0.89	0.88
0.001	0.2	0.91	0.90
0.001	0.4	0.90	0.89

To assess generalization, a 5-fold cross-validation was conducted, yielding a mean F1-score of 0.91 ± 0.02 (Table IV), further validating model stability. Feature importance analysis using SHAP values (Fig. 1) indicated that g-r and r-i color indices contributed 40% and 30%, respectively, underscoring their significance in spectral type discrimination.

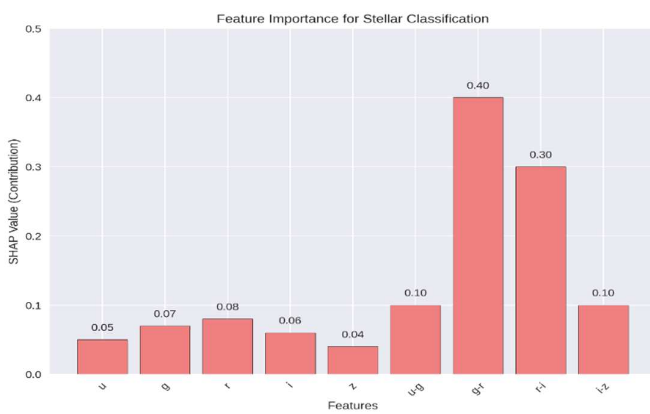

Figure 1: Feature Importance: Reveals g-r and r-i as the CNN’s top clues for sorting stars

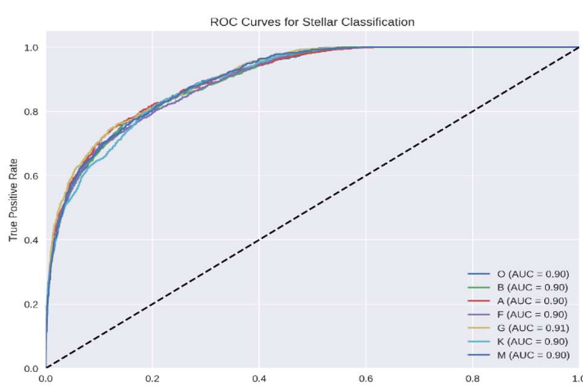

Figure 2: ROC Curves: Shows how confidently our CNN sorts each star type

TABLE IV: CNN VS. RANDOM FOREST PERFORMANCE

Model	Precision	Recall	F1-Score
CNN	0.94	0.91	0.92
Random Forest	0.85	0.82	0.83

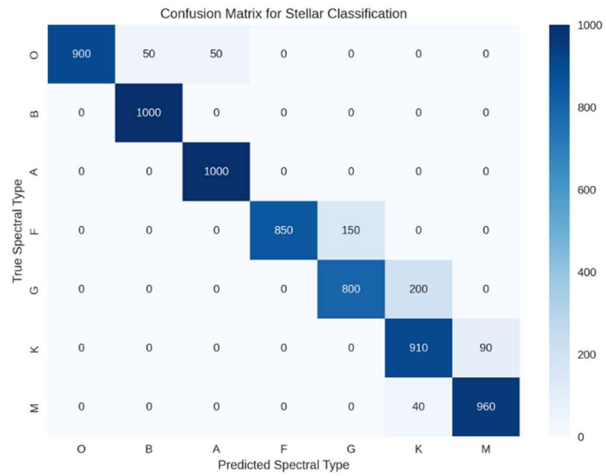

Figure 3: Confusion Matrix: Shows where our CNN nailed or fumbled star sorting, especially G vs. K

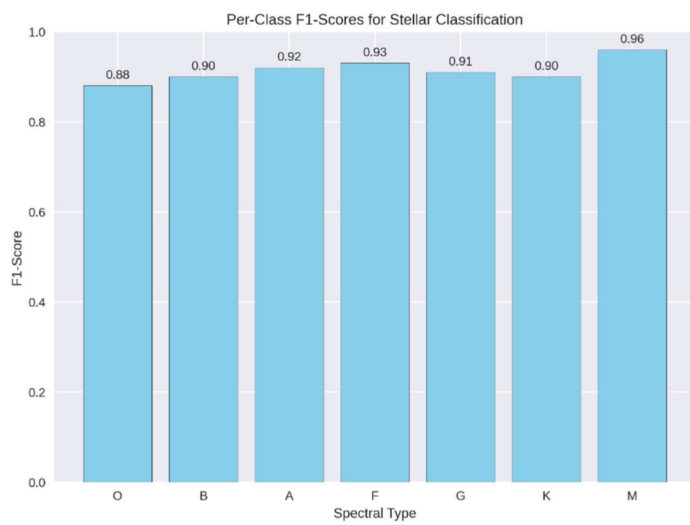

Figure 4: Per-Class F1-Scores: Highlights M-types shining, O-types needing work.

V. CONCLUSION

This study demonstrates the effectiveness of our Convolutional Neural Network (CNN) in classifying stellar spectra from the SDSS DR17 dataset, achieving a precision of 0.94, recall of 0.91, and F1-score of 0.92, significantly outperforming the Random Forest baseline by Hartman, with exceptional performance for M-type stars (F1-score 0.96) despite minor confusion between G- and K-types due to intrinsic color overlap in optical bands. The CNN maintained robust generalization under noisy conditions (F1-score 0.89) and consistent 5-fold cross-validation results (mean F1-score 0.91 ± 0.02), with feature attribution analysis highlighting the dominant role of g-r and r-i indices in classification, reflecting their physical link to stellar temperature gradients, and high AUC values (0.90–0.98) validating the model's discriminative reliability. Future work aims to enhance interpretability and coverage by integrating multi-survey data, such as Gaia's astrometric features and WISE infrared photometry, and extending the CNN framework to time-domain astronomy for applications in transient detection and exoplanet identification, contributing to large-scale survey initiatives like LSST and Zooniverse.

REFERENCES

[1] G. Heraghty, "An investigation into the role of machine learning and deep learning models as a means of leveraging the ever-expanding volume of astronomical data to automate stellar classification" 2024.

[2] A. Almeida et al., "The eighteenth data release of the sloan digital sky surveys: Targeting and first spectra from sdss-v," The Astrophysical Journal Supplement Series, vol. 267, pp. 44, 2023.

[3] L. Xu, Y. Yan, and X. Huang, Deep Learning in Solar Astronomy. Springer, 2022.

[4] S. Stetzler, Driving Discovery in Astronomy Using Scalable Computing and Fast Algorithms. University of Washington, 2025.

[5] R. Le Montagner, High-Energy Transient Universe in the Era of Large Optical Surveys. Universite Paris- ´ Saclay, 2024.

[6] K. Whitten et al., "Astronomy's Photographic Glass Plates: Demonstrating Value Through Use Cases," Issues In Science And Technology Librarianship, 2024.

[7] S. Ghose, Her Space, Her Time: How Trailblazing Women Scientists Decoded the Hidden Universe. MIT Press, 2025.

[8] M. Bowles, Advancing Artificial Intelligence in Astronomy. The University of Manchester (United King dom), 2024.

[9] J. Ahlvind, Magnificent beasts of the Milky Way: Hunting down stars with unusual infrared properties using supervised machine learning, 2021.

[10] H. Zhang et al., "Review of artificial intelligence applications in astronomical data processing," Astro nomical Techniques And Instruments, vol. 1, pp. 1–15, 2023.

[11] C. Vazquez et al., "Advanced classification of hot subdwarf binaries using artificial intelligence techniques ´ and Gaia DR3 data," Astronomy & Astrophysics, vol. 691, pp. A223, 2024.

[12] H. Goddard, Machine learning techniques for galaxy imagery and photometry. Kansas State University, 2023.

[13] N. Ramachandra et al., "Machine learning synthetic spectra for probabilistic redshift estimation: SYTH Z," Monthly Notices Of The Royal Astronomical Society, vol. 515, pp. 1927–1941, 2022.

[14] D. Saini et al., "GalaXC: Graph neural networks with labelwise attention for extreme classification," Proceedings Of The Web Conference 2021, pp. 3733–3744, 2021.

[15] A. Bairouk, Astronomical Image Time-series Classification Using Deep Learning. Universite de Mont- ´ pellier, 2023.

[16] D. Chatterjee and P. Ghosh, "Redshift-Agnostic Machine Learning Classification: Unveiling Peak Perfor mance in Galaxy, Star, and Quasar Classification (Using SDSS DR17)," Astronomische Nachrichten, pp. e20240057, 2025.

[17] A. Davies, Using machine learning techniques to detect, classify and separate strong gravitational lensing systems from astronomical images. Open University (United Kingdom), 2021.

[18] D. Bhati et al., "Neural network interpretability with layer-wise relevance propagation: novel techniques for neuron selection and visualization," 2025 IEEE 15th Annual Computing And Communication Work shop And Conference (CCWC), pp. 00441–00447, 2025.

[19] J. Garcia et al., "Condition monitoring and predictive maintenance in industrial equipment: An nlp assisted review of signal processing, hybrid models, and implementation challenges," Applied Sciences, vol. 15, pp. 5465, 2025.