Integrating Machine Learning for Classification of Exoplanets and Accurate Measurement of Celestial Distances and Sizes

Abstract:
This paper proposes a practical framework for the measurement of distance and size of celestial bodies, including planets and stars, by integrating machine learning algorithms with mathematical models in image processing. Our approach combines data from astronomical observations, light pollution effects, and spatiotemporal analytics to improve accuracy in detecting and classifying celestial objects. This work provides a step-by-step methodology, practical outcomes, and insights into overcoming challenges such as light interference, image resolution, and data limitations.

Keywords: - Support Vector Machine (SVM), Convolutional Neural Network (CNN), Astrophysical Data Processing, Automated Object Detection, Random Forest Classification, Regression.

1.Introduction

In astronomy, understanding the size and distance of celestial bodies is important for revealing the structure and behaviour of the universe. Traditional methods like ‘parallax’ and ‘redshift’ are effective but not suitable for faint or distant objects. Advances in machine learning and image processing offer more efficient and accurate alternatives.
Some of the biggest challenges are handling the effects of light pollution to reduce interference to the quality of data and addressing the computational overhead in high-resolution images of space[1,2]. Our aim is to concentrate on integrating spatiotemporal models addressing light pollution using machine learning with efficiency in detecting objects and performing measurements, also increasing accuracy when it comes to distance and size.

2. Literature Review

2.1 Mathematical Models in Celestial Detection

The mathematical models from *"Asteroids and Their Mathematical Methods"* and related sources, Kepler's laws and orbital dynamics for distance calculations and size approximations in non-Earth objects are as below[3,4]:

$$I(R) = I_0 exp\left[-\beta_n \left(\frac{R}{R_e}\right)^{1/n}\right] = I_e exp\left[-\beta_n \left\{\left(\frac{R}{R_e}\right)^{1/n} - 1\right\}\right]$$

$$\mu(R) = \mu_e + 1.086\beta_n \left[\left(\frac{R}{R_e}\right)^{1/n} - 1\right] \quad \rho_{HI}(z) = \left(\frac{dl}{dz}\right)^{-1} m_H \int N_{HI} \frac{d^2\mathcal{N}}{dN_{HI}dz} dN_{HI}$$

2.2 Impact of Light Pollution

The Light Pollution Prediction Model identifies the problems of artificial light and describes data collection strategies in areas with high pollution levels, giving crucial corrections for proper identification of celestial objects in polluted skies[5-7].

2.3 Image Classification in Astronomy

The possibility of using machine learning in classifying celestial images, as done in image classification is explored here. Techniques such as Support Vector Machines and Convolutional Neural Networks are mentioned as the fundamental tools in the methodology.

3. Methodology

For an object at a distance *r*, its physical size, *D*, is related to its angular size, θ, by $D = r\theta$

Photometric observations are generally carried out in some chosen waveband. Thus, the observed flux from an object is related to its SED, $f\lambda$, by:

$$fx = \int fyFx(\lambda)R(\lambda)T(\lambda)\, d\lambda$$

Here FX (λ) is the transmission of the filter that defines the waveband (denoted by X), T (λ) represents the atmospheric transmission, and R (λ) represents the efficiency with which the telescope plus instrument detects photons. We will assume that fX has been corrected for atmospheric absorption and telescope efficiency (the correction is normally done by calibrating the data using standard objects with known fλ).

It is often inconvenient to express LX in units of the solar luminosity in the same band, LX:

$$\mu x = -2.5 \log\left(\frac{L_X}{L_{\odot X}}\right) + \mu_{\odot X}$$

where MX is the absolute magnitude of the Sun in the waveband in consideration:

$$m_X - \mu_X = 5\log\,(r/r_0)$$

where $r0$ is a fiducial distance at which *mX* and *MX* are defined to have the same value. Conventionally, $r0$ is chosen to be 10 pc (1 pc = 1 parsec = 3.0856×1018 cm; see §2.1.3 for a definition).

The quantity (*mX −MX)* for an astronomical object is called its distance modulus:

$$\mu_x = -2.5log\left(\frac{I_X}{L_{\odot}pc^{-2}}\right) + 21.572 + \mu_{\odot X}$$

Note that what is additive is the flux and not the magnitude. Therefore, to obtain the total apparent magnitude, from an image one has first to convert magnitude per unit area in to flux per unit area to integrate the flux over the total area of the image to finally convert it to a magnitude. If the observations are carried out for the same object at more than one waveband, then the difference of magnitudes in any two different bands is referred to as the color index, and this corresponds to the slope of the SED between the two wavebands. For example, $(B-V) \equiv m_B - m_V = \mu_B - \mu_V$ is called the (*B−V*) color of the object.

From spectroscopic observations one obtains spectra for objects, i.e. their SEDs *fλ* or *fν* defined so that *fλ* dλ and *fν* d*ν* are the fluxes received in the elemental wavelength and frequency ranges dλ at λ and d*ν* at *ν*. From the relation between wavelength and frequency, λ = *c/ν*, we then have:

$$f_V = \lambda^2 f_\lambda/c \quad \text{and} \quad f_\lambda = v^2 f_v/c \qquad [A/B] \equiv log\left[\frac{(n_A/n_B)_*}{(n_A/n_B)_\odot}\right]$$

3.1 Data Collection

3.1.1 Telescope Imaging

- High-resolution imaging data from telescopes for capturing celestial bodies is used. For remote sensing datasets, such as from NASA's or NOAA's satellites, the instruments (e.g., Visible Infrared Imaging Radiometer Suite (VIIRS) for night time data) are used.

3.1.2 Light Pollution Adjustment

- Spatiotemporal models (inspired by the *Light Pollution Prediction Model*) to analyze and adjust for light interference from nearby urban or industrial sources are employed.
- A baseline or "dark frame" image in areas with minimal artificial light is created to help isolate celestial object data from light noise.
  The following fig. 1 shows the different types of galaxies.

Fig. 1:Examples of different types of galaxies. From left to right and top to bottom, NGC4278 (E1), NGC3377 (E6), NGC5866 (SO), NGC175 (SBa), NGC6814 (Sb), NGC4565 (Sb, edge on), NGC5364 (Sc), Ho II (Irr I), NGC520 (Irr II). [All images are obtained from the NASA/IPAC Extragalactic Database (NED) which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration]

3.2 Image Processing and Object Detection

3.2.1 Preprocessing

- Techniques like wavelet transforms and filtering (as outlined in *Image Classification Using Machine Learning*) are used to preprocess images by enhancing edges and removing background noise.
- Convolutional Neural Networks (CNNs) are employed for feature extraction to improve object detection. The following fig. 2 shows the galaxy properties and the fig.3 shows Luminosity functions.

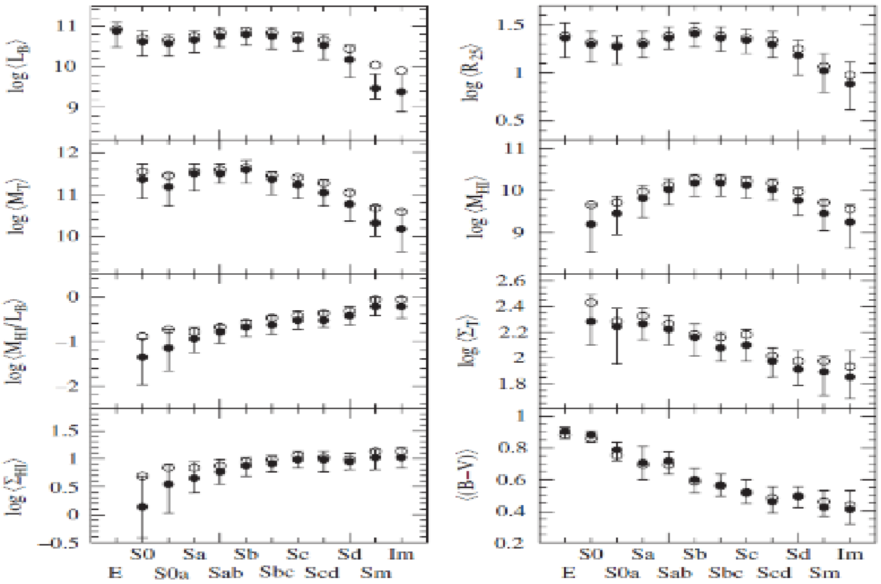

Fig.2: Galaxy Properties

3.2.2 Object Classification

- A machine learning model using Support Vector Machines (SVM) ,Convolutional Neural Networks (CNNs) is used to classify and distinguish between different types of celestial bodies (e.g., planets, stars, asteroids).

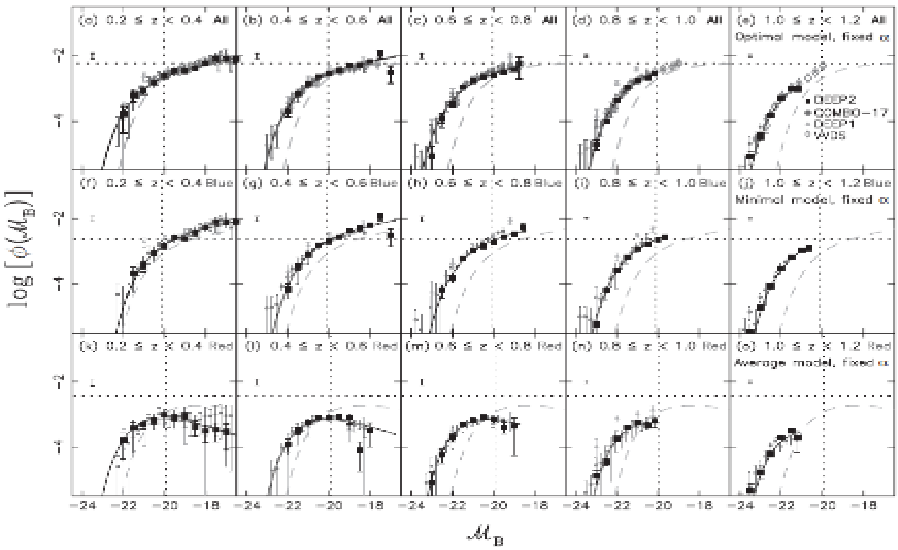

Fig.3: Luminosity Functions

3.3 Distance Calculation

3.3.1 Parallax Method

- The parallax method is used for nearby stars, measuring the apparent shift in the star's position from Earth-based observations taken six months apart.
- For validation, machine learning-derived measurements are compared with existing parallax data to check for accuracy.

3.3.2 Redshift Analysis

- For distant stars, redshift data from spectral analysis is applied, using Doppler effect principles to estimate distance based on the expansion of the universe. The following fig. 4 shows the proposed methodology.

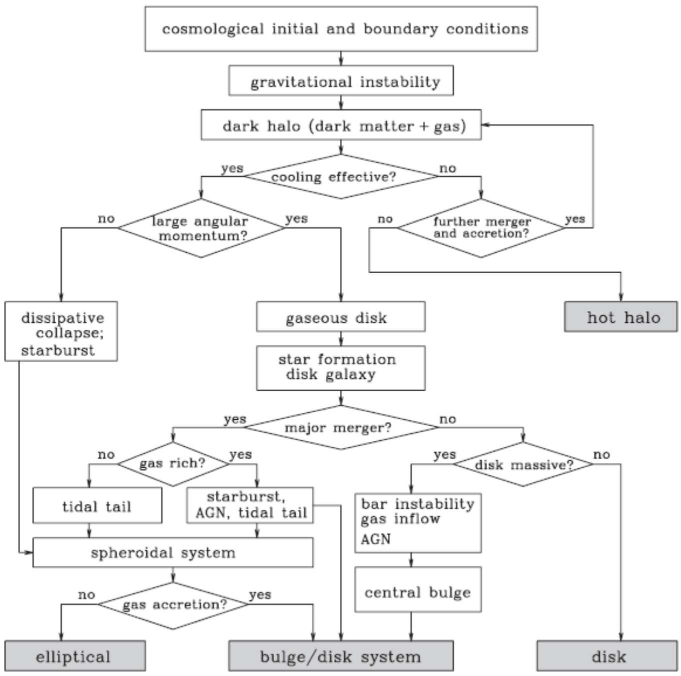

Fig. 4: Proposed Methodology

3.4 Size Estimation

3.4.1 Angular Diameter Measurement

Angular diameter measurements is used, applying trigonometric relationships and known distances (from previous calculations) to estimate the size of the celestial body. The various communication Protocols that are energy efficient and secure are discussed. [8,9]

3.4.2 Machine Learning-Based Random Forest & Regression Models

Regression-based machine learning model is implemented, along with a random forest classification model, to refine size estimates by analyzing relationships between distance, observed luminosity, and angular size.

4. Results and Discussion

4.1 Model Performance

- Evaluation metrics for object classification and distance measurements are presented.
- Results from the machine learning models are compared against the standard astronomical measurement techniques to demonstrate the model's effectiveness.

4.2 Challenges and Limitations

- Issues related to light pollution, image resolution, and data scarcity are discussed.
- The trade-offs between model accuracy and computational requirements, especially when dealing with large datasets are discussed.

The results are shown in the following table 1.

Table 1: Results

Sr. No.	Precision	Recall Value	F1-Score	Support
1.	1.00	0.82	0.90	11
2.	0.95	0.97	0.96	37
3.	0.00	0.67	0.80	3
4.	0.96	0.97	0.96	194
5.	1.00	0.42	0.59	12
6.	0.99	1.00	0.99	714

5. Conclusion

Machine learning like SVMs, CNNs, and random forest classification are used with traditional methods of astronomical calculation to determine the distance, sizes of celestial bodies and classification of exo-planet. It is observed that the system improves detection and classification with high accuracy under even light-polluted conditions and thus has great promise for future applications in advancing research and education. Future development involves extension of existing datasets, improvements in model refinement for real-time processing, and extensions for complex phenomena related to the heavens, thus making advanced tools reachable and deepening our cosmic understanding.

6. References

[1] Houjun Mo, Frank van den Bosch, and Simon White, Galaxy Formation and Evolution. Cambridge University Press, 2010. Available: www.cambridge.org/9780521857932.

[2] J. H. J. de Bruijne, M. Allen, S. Azaz, A. Krone–Martins, T. Prod'homme, and D. Hestroffer, "Detecting stars, galaxies, and asteroids with Gaia," Astronomy & Astrophysics, Preprint online version, Feb. 9, 2015.

[3] Y. Jiang, Y. Ni, H. Baoyin, J. Li, and Y. Liu, "Asteroids and Their Mathematical Methods," Mathematics, vol. 10, no. 16, p. 2897, Aug. 2022.

[4] L. S. Riza, Z. A. Y. Putra, M. F. Y. Firdaus, F. Z. T. Trihutama, A. Izzuddin, J. A. Utama, K. A. F. A. Samah, D. Herdiwijaya, R. A. NQZ, and E. S. Mumpuni, "A spatiotemporal prediction model for light pollution in conservation areas using remote sensing datasets," *Decision Analytics Journal*, vol. 9, p. 100334, Sep. 2023, doi: 10.1016/j.dajour.2023.100334.

[5] P. H. S., N. M. Arin ul Haq, R. K. Sai, T. M. Choda, and R. C. Mahaladkar, "Galactical image detection and identification - A survey report," *International Journal of Creative Research Thoughts (IJCRT)*, vol. 11, no. 11, Nov. 2023, pp. 212–217, ISSN: 2320-2882. Available: https://www.ijcrt.org.

[6] Ž. Andreić, "Asteroids," *The Mining-Geology-Petroleum Engineering Bulletin*, vol. 69, no. 1, pp. 69–85, 2016, doi: 10.17794/rgn.2016.1.5.

[7] V. Bacu, C. Nandra, A. Sabou, T. Stefanut, and D. Gorgan, "Assessment of asteroid classification using deep convolutional neural networks," Aerospace, vol. 10, no. 9, pp. 752, Aug. 2023.

[8] Komal Baburao Umare, S. B G Tilak Babu, K. Aruna, Dr.Bandi Nageshwar Rao, S Karthi sree,Fazil A. Sheikh, 'Energy-Efficient Communication Protocols for Wireless IoT Networks', IJMI/article/view/1131, Dec.2024.

[9] Dr. R.Kiruthigha, Dr. R. Raman, Hari Shankar Punna, V. Subedha, Professor, Fazil A. Sheikh,'Quantum Cryptography Methods for Securing Communication Networks. IJMI/article/view/1779, Dec.2024.