International Astronomy and Astrophysics Research Journal

Volume 8, Issue 1, Page 1-12, 2026; Article no.IAARJ.162021

Predictive Modeling of Exoplanet Orbital Eccentricity with Interpretable Machine Learning Methods

Bharat Khushalani [a*] and Daisha Janjani [b]

[a] Picosoft Research, 5529 163rd CT NE Redmond, WA, 98052 USA.
[b] Tesla STEM, 4301 228th Ave NE, Redmond, WA 98053 USA.

Authors' contributions

This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

Article Information

DOI: https://doi.org/10.9734/iaarj/2026/v8i1122

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://pr.sdiarticle5.com/review-history/162021

Original Research Article

Received: 29/04/2026
Accepted: 08/07/2026
Published: 11/07/2026

Abstract

Orbital eccentricity is a fundamental parameter in exoplanet science because it encodes information about planetary system architecture, dynamical evolution and potential climatic variability. However, eccentricity is not always easy to measure directly, and its distribution across the known exoplanet population remains uneven and strongly skewed toward near-circular orbits. This study investigates whether continuous exoplanet eccentricity can be predicted from observable stellar and orbital properties using regression-based machine-learning methods. A filtered dataset of 1,091 confirmed exoplanets was constructed from the NASA Exoplanet Archive, retaining systems with measured eccentricity and selected numerical predictors, including orbital period, semi-major axis, planet mass, stellar mass, stellar radius, stellar effective temperature, distance, Gaia G-band magnitude, right ascension and system multiplicity. Two regression approaches were evaluated: a baseline linear regression model and a Random Forest regressor. The results show that eccentricity is difficult to model accurately as a continuous target. The linear baseline achieved modest predictive power, whereas the Random Forest regressor improved performance and captured some nonlinear structure in the

*Corresponding author: E-mail: bharat.khushalani@picosoftres.org, b.khushalani@yahoo.com;

Cite as: *Khushalani, B., & Janjani, D. (2026). Predictive Modeling of Exoplanet Orbital Eccentricity with Interpretable Machine Learning Methods. International Astronomy and Astrophysics Research Journal, 8(1), 1–12. https://doi.org/10.9734/iaarj/2026/v8i1122*

data. Across both models, prediction quality deteriorated for rare high-eccentricity systems, with model outputs tending to compress toward intermediate values. Feature-importance analysis showed that orbital period, semi-major axis, planetary mass and system multiplicity were among the most influential predictors, which is consistent with astrophysical expectations linking eccentricity to tidal evolution and long-term dynamical interactions. Overall, the study shows that, while stellar and orbital observables contain some predictive information, exact continuous eccentricity prediction remains challenging, motivating future work using classification-based formulations and more physically informed feature engineering.

Keywords: Exoplanets; orbital eccentricity; regression modelling; machine learning; Random Forest; planetary system architecture; astrophysical data analysis; orbital dynamics; interpretability

1. Introduction

The discovery of thousands of exoplanets over the past few decades has revealed that planetary systems exhibit remarkable diversity in orbital architecture (Ford, 2014; Winn & Fabrycky, 2015). Among the most informative orbital parameters is eccentricity, which measures the extent to which a planet's orbit departs from a perfect circle. In contrast to the relatively low eccentricities of most planets in the Solar System, many exoplanets are found on significantly elliptical orbits, indicating that their dynamical histories may include migration, scattering, tidal evolution or long-term gravitational perturbations (Gilbert, Petigura & Entrican, 2025). Because of this, orbital eccentricity serves as an important diagnostic of planet formation and post-formation evolution.

Eccentricity is also astrophysically significant because it influences the amount and variability of stellar radiation received by a planet over the course of its orbit (Iro & Deming, 2010). As eccentricity increases, the difference between periastron and apastron grows, producing stronger variations in incident flux and potentially affecting climate stability, atmospheric evolution and planetary habitability (Liu et al., 2024). For giant planets, eccentricity can also provide indirect evidence of system-scale dynamical processes, such as planet-planet scattering or secular interactions (O'Connor, Teyssandier & Lai, 2022). Consequently, understanding the distribution and drivers of exoplanet eccentricity is important both for planetary dynamics and for broader questions in comparative exoplanetology.

Orbital eccentricity plays a central role in exoplanet science because it encodes the combined effects of planet formation, post-formation dynamical evolution, and star–planet interactions, while also strongly influencing planetary insolation and potential habitability. The diversity of eccentricities observed among exoplanets has motivated extensive work on their measurement, statistical characterization, physical origin, and climatic consequences. Accordingly, a substantial body of literature has examined eccentricity from observational, dynamical, and population-level perspectives.

Shen and Turner (2008) investigated how orbital eccentricities inferred from radial-velocity surveys can be distorted by observational noise and Keplerian fitting procedures. Using large suites of synthetic radial-velocity data, they showed that while selection effects against eccentric orbits are generally weak, low signal-to-noise detections and sparsely sampled systems tend to have their best-fit eccentricities systematically biased upward. As a result, the observed exoplanet sample may underrepresent the true number of nearly circular orbits, and the intrinsic eccentricity distribution is likely more strongly peaked near zero than raw catalog values suggest. Leconte et al. (2010) investigated whether tidal heating can explain the inflated radii of hot Jupiters by coupling the planets' internal thermal evolution with full tidal-orbital evolution equations valid at arbitrary eccentricity. They showed that the low-order e^2 tidal approximations used in many earlier studies can severely underestimate tidal dissipation and misrepresent orbital evolution, especially for initially eccentric systems. Using complete tidal calculations, they concluded that tidal heating may explain some moderately bloated hot Jupiters, but it is insufficient on its own to account for the most inflated planets, including HD 209 458 b, because strong tidal dissipation and circularization occur too early in the planet's evolution.

Correia, Boué, and Laskar (2011) showed that tidal effects do not always damp the eccentricity of close-in exoplanets monotonically. Using a secular model that couples spin evolution, tides, and perturbations from an outer companion, they demonstrated that a phase-lagged spin response can generate a secular increase in the inner planet's eccentricity. They argued that this "eccentricity pumping" can delay circularization for gigayear timescales and may explain why some moderately close-in exoplanets retain substantial eccentricities, possibly indicating the presence of undetected companions. Kane et al. (2012) compared transit-duration measurements of Kepler planet candidates with the eccentricity distribution known from radial velocity planets. Using the difference between predicted circular-orbit transit durations and observed or eccentric-orbit durations, they found that the Kepler candidates are statistically consistent with the RV eccentricity distribution. They also showed that the mean eccentricity decreases with decreasing planet size, indicating that smaller planets are preferentially found in low-eccentricity, likely multi-planet, orbital systems.

Kipping (2013) proposed the Beta distribution as a compact and physically appropriate parametrization of the exoplanet eccentricity distribution. Fitting the cumulative eccentricity distribution of 396 high signal-to-noise radial-velocity planets, he found that a Beta distribution with a≈0.867 and b≈3.03 describes the data better than commonly used alternatives such as a Rayleigh + exponential mixture. He also showed that short- and long-period planets are better represented by distinct Beta distributions, with short-period planets exhibiting a stronger excess of low-eccentricity orbits consistent with tidal circularization. Limbach and Turner (2015) analyzed radial-velocity exoplanet systems and found a strong anti-correlation between orbital eccentricity and planetary-system multiplicity, with higher-multiplicity systems exhibiting systematically lower eccentricities. They showed that the Solar System, often considered atypical because of its low eccentricities, is actually consistent with this broader trend when its eight-planet multiplicity is taken into account. Fitting the higher-multiplicity data with a power law, they also suggested that many nominal one- and two-planet RV systems may contain additional undiscovered planets.

Xie et al. (2016) used transit-duration statistics together with precise LAMOST spectroscopic host-star parameters to infer eccentricity distributions for a homogeneous sample of 698 Kepler planet candidates. They found a strong dynamical dichotomy: single-transiting systems have a mean eccentricity of about 0.3, whereas multiple-transiting systems are nearly circular and nearly coplanar, with $\bar{e} \approx 0.04$ and $\bar{\imath} \approx 1.4^{o}$. They further showed that Kepler multiples and Solar System populations follow a common approximate relation between mean eccentricity and inclination, suggesting that Solar-System-like dynamically cold architectures may be common in the Galaxy. Kane and Torres (2017) derived a framework for comparing the seasonal stellar-flux variations caused by orbital eccentricity and axial obliquity on terrestrial exoplanets. They showed how to determine combinations of e and θ that produce equivalent latitudinal flux variations, both for instantaneous and diurnally averaged flux, and applied the method to GJ 163 c, K2-3 d, Kepler-186 f, and Proxima Centauri b. Their results demonstrate that even modest eccentricities can strongly affect climate-relevant flux patterns and can be used as practical proxies for constraining the climatic role of presently unmeasurable obliquity.

Wang et al. (2017) used a three-dimensional general circulation model to investigate how orbital eccentricity affects the climate and habitable-zone limits of tidally locked terrestrial exoplanets around M dwarfs. They found that even for e=0.4, seasonal surface-temperature variations remain weak because ocean thermal inertia damps the strong insolation cycle, but increasing eccentricity still warms the annual mean climate by reducing cloud albedo and moves both the runaway and moist greenhouse inner edges outward. They also showed that different spin-orbit resonance states produce distinct climate regimes, with integer resonances yielding “eyeball” climates and half-integer resonances yielding warmer “striped-ball” climates, so eccentricity generally narrows the habitable zone. Carrera, Raymond, and Davies (2019) used 500 N-body simulations of unstable three–giant-planet systems to test whether planet–planet scattering can account for the most eccentric known exoplanets. They found that scattering can produce surviving bound planets with eccentricities exceeding 0.99, with no apparent upper eccentricity limit intrinsic to the mechanism. Comparing their simulations with radial-velocity giant planets, they concluded that the observed high-eccentricity population is consistent with being generated entirely by planet–planet scattering, so extremely eccentric orbits do not by themselves require Kozai–Lidov forcing.

Bowler, Blunt, and Nielsen (2020) combined new and archival astrometry for 27 directly imaged giant planets and brown dwarf companions between 5 and 100 AU to infer their population-level eccentricity distributions using hierarchical Bayesian modeling. They found that the full sample is roughly consistent with a flat eccentricity distribution, but that giant planets and brown dwarfs differ significantly: low mass-ratio companions preferentially occupy lower eccentricities, whereas brown dwarfs show a broad peak at e≈0.6–0.9 and a period dependence similar to wide stellar binaries. They interpret these differences as dynamical evidence for distinct formation channels, with wide giant planets likely forming in circumstellar disks and wide brown dwarfs forming more like stellar companions through fragmentation. Stevenson et al. (2025) re-examined the eccentricity distribution of radial-velocity exoplanets using an updated sample of 891 planets, about 2.25 times larger than that used by Kipping (2013). They found that the widely used Kipping Beta prior is no longer the preferred global model: CDF regression favors a Rayleigh + Exponential mixture, while PDF regression favors a Gamma distribution, with both methods outperforming the Beta distribution for the enlarged sample. They also showed that RV exoplanet eccentricities are not drawn from a single universal parent distribution, but differ significantly with orbital period, planet mass, and system multiplicity, motivating the use of tailored eccentricity priors rather than one standard prior for all systems.

Ferraz-Mello and Beaugé (2025) investigated whether the observed nonzero eccentricity of 55 Cnc e could drive enough tidal dissipation to explain the planet's anomalously high and variable thermal emission. Using the creep-tide model, they showed that a nonzero eccentricity can indeed generate extremely strong internal heating, potentially consistent with intense volcanism, but that the same tidal dissipation should also circularize the orbit rapidly unless the eccentricity is continually forced. They therefore concluded that, if the measured eccentricity of 55 Cnc e is real, it likely requires perturbations from an as-yet undetected nearby companion, probably in or near a 2:1 or 3:2 orbital resonance. Blunt et al. (2026) used completeness-corrected hierarchical occurrence-rate modeling of the California Legacy Survey radial-velocity sample to infer the eccentricity distribution of giant planets as a function of mass and semimajor axis. They found that while sub-Jovian planets remain concentrated toward low eccentricities, super-Jovian planets (roughly 3–20 M_J) at 0.1–4.5 au show evidence for a population-level peak at moderate eccentricity, with the preferred peak location near $e \approx 0.3$. Their analysis indicates that this feature is statistically robust after accounting for survey completeness, posterior uncertainties, and inclination effects, and may point to dynamically hot formation or post-formation histories for typical super-Jovian planets.

Ribas and Miralda-Escudé (2007) examined how exoplanet eccentricity and host-star metallicity vary with companion mass and found tentative evidence for two distinct giant-planet populations. They showed that planets with (M sin i) $\gtrsim 4\ M_J$ have an eccentricity distribution statistically similar to that of spectroscopic binaries, while lower-mass planets are less eccentric and tend to orbit more metal-rich stars. They interpreted these trends as possible signatures of overlapping formation channels, with lower-mass giants forming by core accretion and higher-mass objects forming by direct gas fragmentation, potentially followed by substantial inward migration. Linsenmeier, Pascale, and Lucarini (2015) used an intermediate-complexity general circulation model to study how high obliquity and orbital eccentricity affect the climate and habitability of idealized Earth-like aquaplanets. They found that strong seasonal variability alters both the bistability between warm and snowball climates and the location of the outer habitable-zone boundary, allowing planets with high obliquity and/or eccentricity to retain temporally or partially ice-free regions at much lower stellar irradiance than low-seasonality cases. Their results also show that while GCMs broadly agree with energy-balance models on circular orbits, significant differences emerge on eccentric orbits, where the temporal distribution of insolation becomes crucial.

Correia, Boué, and Laskar (2012) showed that tidal effects do not always lead to monotonic eccentricity damping for close-in exoplanets. Using a secular model that couples the planet's spin evolution, tidal dissipation, and perturbations from an outer companion, they demonstrated that a phase-lagged spin response can generate a secular increase in the inner planet's eccentricity and substantially delay circularization. They proposed this eccentricity pumping mechanism as an explanation for why some moderately close-in exoplanets retain large eccentricities, possibly indicating the presence of undetected outer companions. Bekesov et al. (2023) investigated whether the orbital eccentricity of a transiting exoplanet system can be inferred from a single transit light curve. They showed that if the longitude of periastron is known, the transit profile can yield a reliable estimate of the eccentricity, whereas if it is unknown, the light curve still constrains the allowed eccentricity range. They also found that neglecting eccentricity can bias the inferred stellar and planetary radii by up to about 10%, while the fitted inclination and limb-darkening coefficient are generally less sensitive.

Dobrovolskis (2011) analyzed how eccentricity, obliquity, and spin–orbit resonance jointly determine the surface insolation patterns of exoplanets. He showed that the combined presence of nonzero eccentricity and obliquity produces complex, non-separable irradiation geometries, including distinctive symmetries for half-odd-integer resonances such as the 3:2 state. He also derived a simple expression for the time-averaged polar insolation, $\bar{S} = S_* \sin\beta / \left[\pi\sqrt{1 - e^2}\right]$ demonstrating that polar forcing increases with both obliquity and eccentricity and may play an important role in climate and habitability. Rodigas and Hinz (2009) used Monte Carlo simulations of radial-velocity data to test how an undetected wide-separation outer companion can bias the fitted eccentricity of an inner RV planet. They found that such companions can increase the fraction of planets with fitted $e > 0.1$ by about 10%, but that this effect alone cannot reproduce the full observed RV eccentricity distribution. Their best-fitting population model required a mixture of roughly 45% circular double-planet systems and 55% intrinsically eccentric single-planet systems, and they showed that planets with moderate eccentricities ($0.1 < e < 0.3$) are the most promising targets for identifying hidden outer companions.

Ford (2011) reviewed the major mechanisms proposed to explain the broad eccentricity and inclination distributions of giant exoplanets, emphasizing that simple disk migration alone is insufficient to account for the observed architectures. He argued that late-stage dynamical processes—including planet–planet scattering, secular perturbations, and Kozai-type excitation followed by tidal circularization—are likely central to the formation of eccentric giant planets and misaligned hot Jupiters. The review also highlighted that these same mechanisms make testable predictions for planets at wide separations, linking radial-velocity, transit, direct-imaging, and microlensing constraints into a common dynamical framework.

Despite its importance, eccentricity remains difficult to model and predict. Direct measurements often depend on observational method, signal quality, and the duration and cadence of follow-up observations. As a result, eccentricity measurements are incomplete and uneven across the known exoplanet population. This motivates the use of data-driven methods to investigate whether eccentricity can be inferred from other observable stellar and orbital properties. If successful, such models could provide both predictive utility and insight into the astrophysical variables most closely associated with eccentric or near-circular planetary orbits.

In this study, exoplanet eccentricity is examined as a supervised regression problem using a filtered dataset of confirmed planets with measured orbital and stellar parameters. Two regression approaches are considered: a baseline linear regression model, and a Random Forest regressor. These models are trained to predict continuous eccentricity values from features describing system architecture and host-star properties, including orbital period, semi-major axis, planetary mass, stellar mass, stellar radius, stellar effective temperature, distance, Gaia magnitude, right ascension, and system multiplicity. The purpose of comparing these models is twofold: first, to determine whether eccentricity can be predicted with useful accuracy from available observables, and second, to assess whether nonlinear machine-learning methods capture meaningful structure beyond what is available to simple linear models.

The regression results show that exoplanet eccentricity is challenging to predict as a continuous quantity. Although the nonlinear model captures more structure than the linear baseline in some cases, both approaches struggle with the strongly skewed eccentricity distribution and the relatively rare but astrophysically important high-eccentricity systems. Nevertheless, the models identify several physically plausible influential predictors, especially those associated with orbital scale and system architecture. These findings provide a useful starting point for understanding the limits of continuous eccentricity prediction and motivate alternative formulations, such as classification into broader eccentricity categories.

Methodological applications in other observational sciences have similarly used multiple regression and Random Forest models to predict continuous physical or environmental variables from heterogeneous predictors, supporting the need for transparent benchmarking when applying ensemble regression to exoplanet data (Kadzuwa & Missanjo, 2023; Obisesan, 2024).

Recent astronomy studies have further strengthened the need to link eccentricity prediction to system-level context: machine-learning analysis of Gaia astrometric orbits has used feature-importance assessment to evaluate candidate plausibility, while recent population studies show that eccentricity varies with planet class, multiplicity and host-star properties (Alqasim et al., 2025; Fairnington et al., 2025; Sahlmann & Gómez, 2025).

Research gap: Although previous work has extensively characterised eccentricity distributions and their physical interpretations, comparatively less attention has been given to whether continuous eccentricity can be predicted from a limited set of readily available stellar, orbital and observational features while retaining model interpretability. This gap is important because the known exoplanet catalogue is observationally heterogeneous and high-eccentricity systems remain relatively rare.

Objective: This study aims to evaluate whether selected stellar and orbital observables can predict continuous exoplanet orbital eccentricity, to compare a transparent linear regression benchmark with a nonlinear Random

Forest regressor, and to interpret the most influential predictors while recognising the limits of continuous eccentricity modelling.

2. Dataset Description

The filtered dataset was constructed from the original NASA Exoplanet Archive dataset to focus on planets with usable eccentricity information and the specific predictors needed for the revised analysis. Rows were retained only when orbital eccentricity and stellar mass were available, and the reduced dataset retained only the variables referenced in the working analysis document.

The sample is still observationally heterogeneous because it combines planets discovered by different methods and under different survey sensitivities. Accordingly, the dataset should be interpreted as a curated observational sample rather than as an unbiased census of planetary systems. That limitation matters when interpreting feature importance and class frequencies, especially for rare high-eccentricity systems.

2.1 Processing Workflow

All columns except the identifier were cast to numeric form so that summary statistics, plots, and machine-learning models were based on quantitative values rather than text strings. The target variable was retained in continuous form as orbital eccentricity, while a discretised version can be created later for classification experiments if regression performance remains limited by the strong concentration of planets at low eccentricity.

The identifier column, pl_name, was preserved only for bookkeeping, error tracing, and figure labelling; it was excluded from model fitting to prevent non-physical memorisation. The response variable was orbital eccentricity (pl_orbeccen). The remaining astrophysical variables describe system architecture, stellar properties, and basic observational context: number of known planets in the system (pl_pnum), orbital period (pl_orbper), semi-major axis (pl_orbsmax), planet mass in Jupiter masses (pl_bmassj), distance (st_dist), Gaia G-band magnitude (gaia_gmag), right ascension (ra), stellar effective temperature (st_teff), stellar mass (st_mass), and stellar radius (st_rad). Table 1 shows these variables and their roles.

Missing values are handled only in predictor columns. In the filtered table, the main incomplete variables are stellar radius (204 missing values), stellar effective temperature (95), semi-major axis (92), planet mass (71), and Gaia magnitude (34). Because the modeling framework is tree-based, these predictors can be imputed with robust summary statistics such as the median calculated on the training split, avoiding leakage from validation and test data.

No direct target-leaking columns are included in this reduced CSV. In particular, the file does not contain eccentricity error flags or fit-quality metadata that would encode the target too directly and inflate model performance artificially. Before model training, the dataset should be split into training, validation, and test subsets. Any imputation, binning, or class-balancing step should be fitted on the training partition only and then applied unchanged to the validation and test partitions. This keeps the analysis focused on astrophysical signal rather than on unnecessary transformations.

Table 1. Variable Roles

Variable	Role	Use in analysis
pl_name	Identifier	Record tracking only; excluded from modeling
pl_orbeccen	Target	Continuous response variable to be predicted and analyzed
pl_pnum	Predictor	System multiplicity / architecture context
pl_orbper, pl_orbsmax	Predictors	Orbital scale and dynamical context
pl_bmassj	Predictor	Planet-mass proxy tied to scattering and migration
st_mass, st_rad, st_teff	Predictors	Basic host-star properties
st_dist, gaia_gmag, ra	Predictors / context	Observational and positional context

We constructed the working dataset from the filtered file, which contains 1,091 confirmed exoplanets and only the variables used in the present analysis. All predictor columns were converted to numeric form and screened for missingness before modelling. Missing values were confined to a subset of continuous predictors and are handled through training-set-only imputation, thereby preserving the full eccentricity sample without introducing target leakage.

2.2 Regression Model

The regression analysis predicts continuous orbital eccentricity (pl_orbeccen) from the filtered exoplanet dataset. The model used the filtered file containing 1,091 planets. The data were split into training, validation, and test subsets using a 70/15/15 split with random seed 42. Predictors were standardised using only the training-set mean and standard deviation before fitting. A baseline multiple linear regression model was fitted using ordinary least squares. This model serves as a transparent benchmark for the eccentricity prediction task and helps identify whether the selected stellar and orbital variables contain a strong linear signal.

Table 2. Test-set metrics

Metric	Value
RMSE	0.1839
MAE	0.1385
R^2	0.1953

3. Results and Discussion

3.1 Plots and Interpretation

Fig. 1 shows actual eccentricity versus predicted eccentricity for the held-out test set. If the regression model were highly accurate, the points would cluster tightly around the diagonal reference line. Instead, the predictions are compressed into a relatively narrow band, indicating that the model captures part of the low-eccentricity structure but struggles to recover the full spread of the target distribution, especially at higher eccentricity.

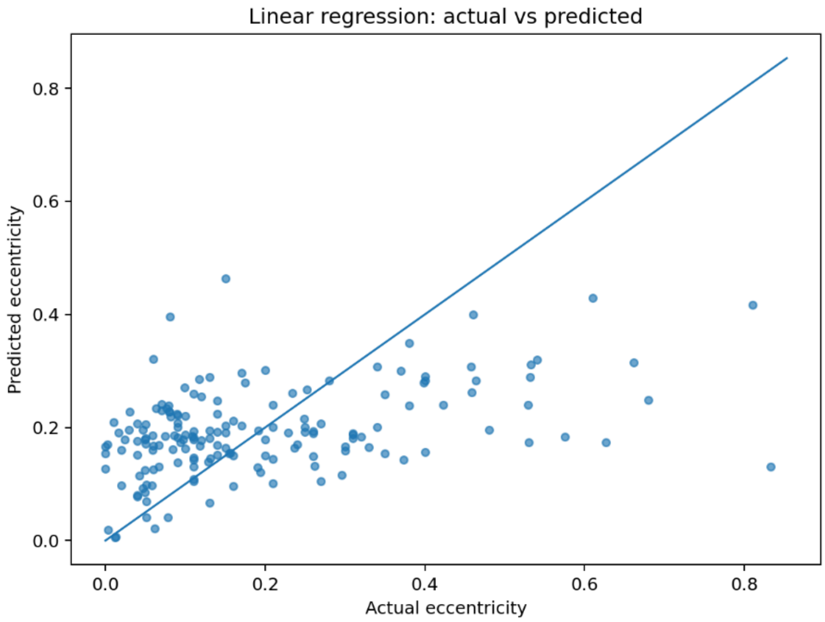

Fig. 1. Linear Regression: Actual vs Predicted

Fig. 2 shows residuals versus predicted eccentricity. A well-calibrated regression model should produce residuals centred around zero with no strong visual pattern. Here, the residuals remain centred near zero overall, but the spread increases for mid-range predictions and several large positive residuals appear, showing that some high-eccentricity planets are underpredicted.

Fig. 3 ranks the predictors by the absolute value of their standardized regression coefficients. Within this linear model, semi-major axis (pl_orbsmax), planetary mass (pl_bmassj), and system multiplicity (pl_pnum) contribute the strongest linear signal. These coefficients do not imply causation, but they indicate which variables the model relied on most strongly.

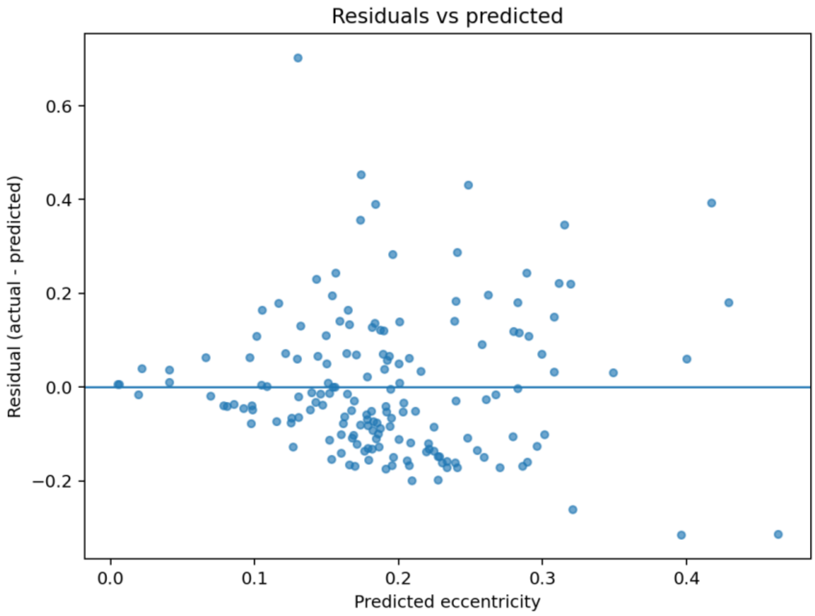

Fig. 2. Residuals vs Predicted

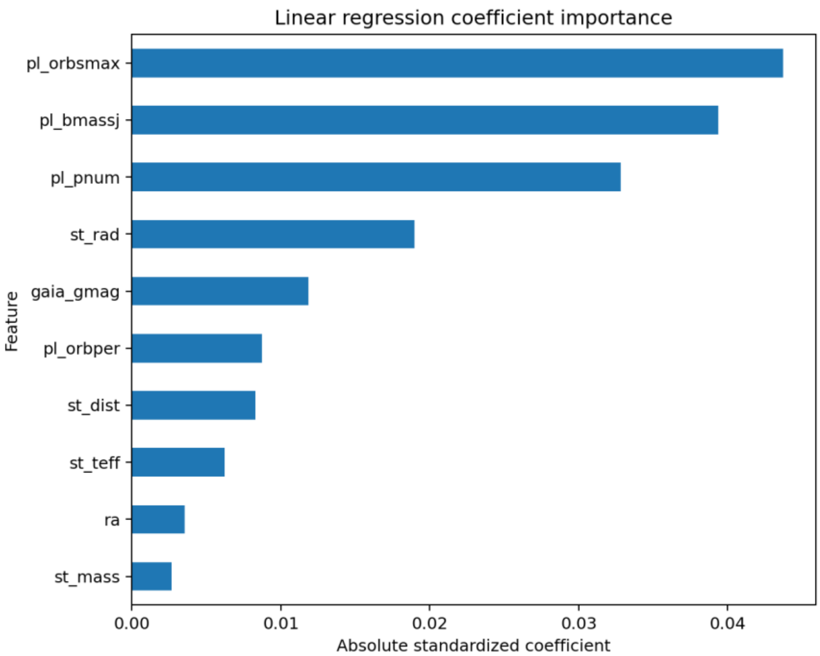

Fig. 3. Linear Regression Coefficient Importance

The baseline linear regression achieved RMSE = 0.1839, MAE = 0.1385, and R^2 = 0.1953 on the test set. The positive but modest R^2 shows that the selected features contain some predictive information, but a simple linear model does not fully explain eccentricity. This is consistent with the astrophysical expectation that eccentricity is shaped by nonlinear dynamical processes and by a long high-eccentricity tail that is difficult to model with a basic regression framework.

3.2 Random Forest

Random Forest is an ensemble machine-learning method built from many decision trees. Instead of relying on a single tree, the method trains a large collection of trees on slightly different versions of the data and then combines their outputs. In regression, the final prediction is the average of the predictions from all trees. Two ideas make Random Forest work well. First, each tree is trained on a bootstrap sample, which is a random resample of the training data. Second, each split in a tree is chosen from only a random subset of the features. These two sources of randomness reduce the correlation between trees and improve generalisation.

The procedure can be summarised as follows: (1) draw a bootstrap sample from the training set, (2) grow a decision tree on that sample, (3) force the tree to choose from a random subset of predictors at each split, (4) repeat the process many times to build a forest, and (5) average the predictions of all trees for the final regression output. Fig. 4 shows how Random Forest was used in this work.

Random Forest is useful when the target variable depends on the predictors through nonlinear relationships and interactions. It can also handle mixed feature scales without requiring normalisation. For exoplanet eccentricity, this is attractive because eccentricity may depend on orbital architecture, planetary mass, tidal effects, and stellar properties in complicated ways. Random Forest was used as a nonlinear regression model to predict the continuous orbital eccentricity of exoplanets.

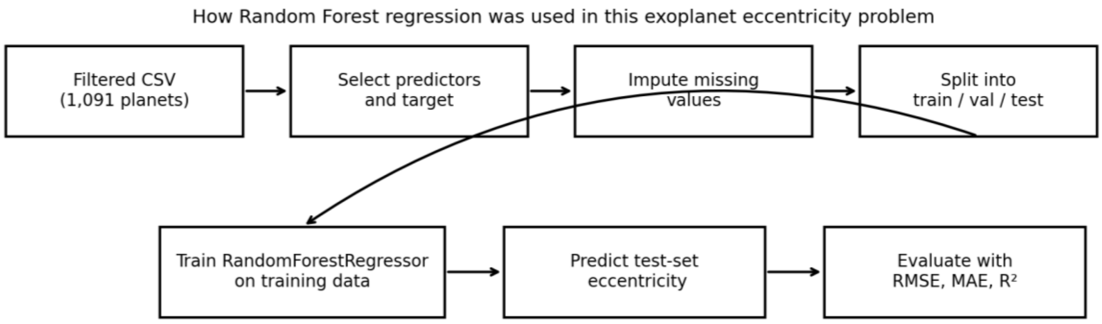

Fig. 4. Random Forest Regressor

The Random Forest Regressor used the following main settings: n_estimators = 400, max_depth = 12, min samples split = 8, min samples leaf = 4, max features = 'sqrt', and random state = 42. These settings create a reasonably regularised forest that can model nonlinear relationships without allowing each tree to become too specialised.

The Random Forest regressor was evaluated on the held-out test set using RMSE (root mean squared error), MAE (mean absolute error), and R^2 (coefficient of determination). The results were:

RMSE = 0.1519, MAE = 0.1159, and R^2 = 0.2291.

These results show that the Random Forest captured more structure than the linear baseline, but continuous eccentricity prediction remained difficult. The main challenge is that the eccentricity distribution is concentrated at low values and contains relatively few high-eccentricity planets, which leads the model to compress many predictions toward moderate values.

Random Forest acts as a flexible nonlinear regression method that learns how orbital eccentricity changes with system architecture and stellar or planetary observables. Its feature-importance results indicated that variables such as orbital period, semi-major axis, planet mass, and multiplicity-related information were among the most influential. This is astrophysically plausible because eccentricity is closely tied to orbital scale, dynamical evolution, and scattering history. At the same time, the Random Forest results show that even a stronger nonlinear model does not fully solve the problem when eccentricity is treated as a continuous target. This is why the classification formulation in the project remains important: predicting broad eccentricity categories may be more stable than predicting exact eccentricity values.

4. Conclusion

This study examined the prediction of continuous exoplanet orbital eccentricity using selected stellar and orbital parameters from a filtered dataset of 1,091 confirmed planets. The linear regression model provided a transparent baseline and showed that the selected observables contain some predictive information, although they explain only a limited proportion of eccentricity variation. The Random Forest regressor captured additional nonlinear structure and produced a modest improvement in R^2, but it did not fully resolve the difficulty of predicting rare high-eccentricity systems. Across the modelling workflow, predictions were compressed toward lower or intermediate eccentricity values, reflecting the skewed eccentricity distribution and the heterogeneity of the observational sample. Feature-importance patterns indicated that orbital period, semi-major axis, planetary mass, and system multiplicity contributed meaningful predictive signals. These results suggest that interpretable machine learning can help identify relationships between eccentricity and observable system properties, but exact continuous prediction remains limited. Future work should incorporate uncertainty-aware modelling, cross-validation, expanded predictors, and classification-based formulations.

5. Limitations

This study is limited by the use of a filtered observational dataset rather than an unbiased census of exoplanet systems. Eccentricity measurements may reflect heterogeneous detection methods, follow-up quality, and catalogue-level bias. The analysis used a limited predictor set and a single train-validation-test split. Imputation may influence feature-importance estimates, and the Random Forest interpretation should be treated cautiously because improved R^2 did not correspond to lower absolute prediction errors.

Declaration of AI Use

This manuscript was prepared through the combined contributions of all authors, including contributions to the study design, data, content development, results, interpretation, and related scholarly work. The authors acknowledge the use of Grammarly and ChatGPT to assist with grammar checking, language refinement, reference formatting. These AI-assisted tools were not used as authors and did not replace the intellectual contributions or scholarly judgment of the authors. All AI-assisted outputs, including content, references, and interpretations, were carefully reviewed, revised, verified, and approved by the authors. The authors accept full responsibility for the accuracy, integrity, and final content of the manuscript.

Competing Interests

Authors have declared that they have no known competing financial interests OR non-financial interests OR personal relationships that could have appeared to influence the work reported in this paper.

References

Alqasim, A., Hirano, T., Hori, Y., Kawata, D., Livingston, J., Howell, S. B., Lanza, A. F., Mann, A. W., Ziegler, C., Briceño, C., Beichman, C. A., Ciardi, D. R., Strakhov, I. A., Lund, M. B., & Law, N. (2025). Investigating the eccentricity distribution of transiting, long-period giant planets. *Monthly Notices of the Royal Astronomical Society, 539*(1), 307–329. https://doi.org/10.1093/mnras/staf418

Bekesov, E. V., Abubekerov, M. K., Gostev, N. Yu., & Cherepashchuk, A. M. (2023). On the possibility to estimate the orbital eccentricity of a binary system with an exoplanet from a transit light curve. *Astronomy Reports, 67*, 1096–1122. https://doi.org/10.1134/S106377292311001X

Blunt, S., Wang, J., Murray-Clay, R., Macintosh, B., Rubenzahl, R. A., & Fulton, B. J. (2026). Evidence for a peak at ~0.3 in the eccentricity distribution of typical super-Jovian exoplanets. *arXiv*. https://arxiv.org/abs/2601.18877

Bowler, B. P., Blunt, S. C., & Nielsen, E. L. (2020). Population-level eccentricity distributions of imaged exoplanets and brown dwarf companions: Dynamical evidence for distinct formation channels. *The Astronomical Journal, 159*(2), 63. https://doi.org/10.3847/1538-3881/ab5b11

Carrera, D., Raymond, S. N., & Davies, M. B. (2019). Planet–planet scattering as the source of the highest eccentricity exoplanets. *Astronomy & Astrophysics, 629*, L7. https://doi.org/10.1051/0004-6361/201935744

Correia, A. C. M., Boué, G., & Laskar, J. (2011). Pumping the eccentricity of exoplanets by tidal effect. *arXiv*. https://arxiv.org/abs/1111.5486

Correia, A. C. M., Boué, G., & Laskar, J. (2012). Pumping the eccentricity of exoplanets by tidal effect. *The Astrophysical Journal Letters, 744*(2), L23. https://doi.org/10.1088/2041-8205/744/2/L23

Dobrovolskis, A. R. (2011). Insolation on exoplanets with eccentricity and obliquity. *EPSC Abstracts, 6*, EPSC-DPS2011-136. https://meetingorganizer.copernicus.org/EPSC-DPS2011/EPSC-DPS2011-136.pdf

Fairnington, T. R., Dong, J., Huang, C. X., Nabbie, E., Zhou, G., Wright, D., Collins, K. A., Jenkins, J. M., Latham, D. W., Ricker, G., Quinn, S. N., Seager, S., Shporer, A., Vanderspek, R., Winn, J. N., Ajizian, C., Fukui, A., Baker, D., Conzo, G., ... Baştürk, Ö. (2025). The eccentricity distribution of warm sub-Saturns in TESS. *Monthly Notices of the Royal Astronomical Society, 540*(1), 1144–1162. https://doi.org/10.1093/mnras/staf759

Ferraz-Mello, S., & Beaugé, C. (2025). The tidal heating of the exoplanet 55 Cnc e: The role of the orbital eccentricity. *Astronomy & Astrophysics, 697*, L8. https://doi.org/10.1051/0004-6361/202554427

Ford, E. B. (2011). The diverse origin of exoplanets' eccentricities and inclinations. In A. Sozzetti, M. G. Lattanzi, & A. P. Boss (Eds.), *The astrophysics of planetary systems: Formation, structure, and dynamical evolution* (IAU Symposium No. 276, pp. 221–224). Cambridge University Press. https://doi.org/10.1017/S1743921311020217

Ford, E. B. (2014). Architectures of planetary systems and implications for their formation. *Proceedings of the National Academy of Sciences of the United States of America, 111*(35), 12616–12621. https://doi.org/10.1073/pnas.1304219111

Gilbert, G. J., Petigura, E. A., & Entrican, P. M. (2025). Planets larger than Neptune have elevated eccentricities. *Proceedings of the National Academy of Sciences of the United States of America, 122*(11), e2405295122. https://doi.org/10.1073/pnas.2405295122

Iro, N., & Deming, L. D. (2010). A time-dependent radiative model for the atmosphere of the eccentric exoplanets. *The Astrophysical Journal, 712*(1), 218–225. https://doi.org/10.1088/0004-637X/712/1/218

Kadzuwa, H., & Missanjo, E. (2023). Modelling above-ground biomass using machine learning algorithm: Case study Miombo woodlands of Malawi. *Journal of Global Ecology and Environment, 17*(3), 1–15. https://doi.org/10.56557/jogee/2023/v17i38178

Kane, S. R., & Torres, S. M. (2017). Obliquity and eccentricity constraints for terrestrial exoplanets. *The Astronomical Journal, 154*(5), 204. https://doi.org/10.3847/1538-3881/aa8fce

Kane, S. R., Ciardi, D. R., Gelino, D. M., & von Braun, K. (2012). The exoplanet eccentricity distribution from Kepler planet candidates. *Monthly Notices of the Royal Astronomical Society, 425*(1), 757–762. https://doi.org/10.1111/j.1365-2966.2012.21627.x

Kipping, D. M. (2013). Parametrizing the exoplanet eccentricity distribution with the beta distribution. *Monthly Notices of the Royal Astronomical Society: Letters, 434*(1), L51–L55. https://doi.org/10.1093/mnrasl/slt075

Leconte, J., Chabrier, G., Baraffe, I., & Levrard, B. (2010). Is tidal heating sufficient to explain bloated exoplanets? Consistent calculations accounting for finite initial eccentricity. *Astronomy & Astrophysics, 516*, A64. https://doi.org/10.1051/0004-6361/201014337

Limbach, M. A., & Turner, E. L. (2015). Exoplanet orbital eccentricity: Multiplicity relation and the Solar System. *Proceedings of the National Academy of Sciences of the United States of America, 112*(1), 20–24. https://doi.org/10.1073/pnas.1406545111

Linsenmeier, M., Pascale, S., & Lucarini, V. (2015). Climate of Earth-like planets with high obliquity and eccentric orbits: Implications for habitability conditions. *Planetary and Space Science, 105*, 43–59. https://doi.org/10.1016/j.pss.2014.11.003

Liu, B., Marsh, D. R., Walsh, C., Cooke, G., & Sainsbury-Martinez, F. (2024). Eccentric orbits may enhance the habitability of Earth-like exoplanets. *Monthly Notices of the Royal Astronomical Society, 532*(4), 4511–4523. https://doi.org/10.1093/mnras/stae1758

O'Connor, C. E., Teyssandier, J., & Lai, D. (2022). Secular chaos in white dwarf planetary systems: Origins of metal pollution and short-period planetary companions. *Monthly Notices of the Royal Astronomical Society, 513*(3), 4178–4195. https://doi.org/10.1093/mnras/stac1189

Obisesan, O. E. (2024). Machine learning models for prediction of meteorological variables for weather forecasting. *International Journal of Environment and Climate Change, 14*(1), 234–252. https://doi.org/10.9734/ijecc/2024/v14i13829

Ribas, I., & Miralda-Escudé, J. (2007). The eccentricity-mass distribution of exoplanets: Signatures of different formation mechanisms? *arXiv*. https://arxiv.org/abs/astro-ph/0606009

Rodigas, T. J., & Hinz, P. M. (2009). Which radial velocity exoplanets have undetected outer companions? *The Astrophysical Journal, 702*(1), 716–723. https://doi.org/10.1088/0004-637X/702/1/716

Sahlmann, J., & Gómez, P. (2025). Machine learning-based identification of Gaia astrometric exoplanet orbits. *Monthly Notices of the Royal Astronomical Society, 537*(2), 1130–1145. https://doi.org/10.1093/mnras/staf018

Shen, Y., & Turner, E. L. (2008). On the eccentricity distribution of exoplanets from radial velocity surveys. *The Astrophysical Journal, 685*(1), 553–559. https://doi.org/10.1086/590548

Stevenson, A. T., Haswell, C. A., Faria, J. P., Barnes, J. R., Barstow, J. K., Dickinson, H., & Standing, M. R. (2025). RV-exoplanet eccentricities: Good, beta, and best. *Monthly Notices of the Royal Astronomical Society, 539*(2), 727–754. https://doi.org/10.1093/mnras/staf502

Wang, Y., Liu, Y., Tian, F., Hu, Y., & Huang, Y. (2017). Effects of eccentricity on climates and habitability of terrestrial exoplanets around M dwarfs. *arXiv*. https://arxiv.org/abs/1710.01405

Winn, J. N., & Fabrycky, D. C. (2015). The occurrence and architecture of exoplanetary systems. *Annual Review of Astronomy and Astrophysics, 53*, 409–447. https://doi.org/10.1146/annurev-astro-082214-122246

Xie, J.-W., Dong, S., Zhu, Z., Huber, D., Zheng, Z., De Cat, P., Fu, J. N., Liu, H.-G., Luo, A., Wu, Y., Zhang, H., Zhang, H., Zhou, J.-L., Cao, Z., Hou, Y., Wang, Y., & Zhang, Y. (2016). Exoplanet orbital eccentricities derived from LAMOST-Kepler analysis. *Proceedings of the National Academy of Sciences of the United States of America, 113*(41), 11431–11435. https://doi.org/10.1073/pnas.1604692113

Peer-review history:
The peer review history for this paper can be accessed here:
https://pr.sdiarticle5.com/review-history/162021