iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78487

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

A Consensus-Based Multi-Model Framework for Anomaly Detection in High-Dimensional Astronomical Data

Sneha Rawat
Department of Computer Applications, CHRIST (Deemed to be University), Delhi NCR, India

Abstract: ***Data from astronomical observations is growing at a much faster rate than ever before, resulting in datasets with unprecedented size and dimensionality. For example, the modern sky survey missions such as Kepler make a large number of measurements (photometric, orbital, and spatial) for a wide range of stars. Because the ability to detect anomalous in these high-dimensional datasets is important for identifying rare astrophysical occurrences and distinguishing them from noise or artefacts from sensors, it is essential to be able to detect anomalies (i.e. data that are different than the others) in these datasets. Many traditional approaches to anomaly detection, including rules-based approaches and single-model machine learning methods, do not work reliably for detecting anomalies because they are sensitive to errors in the data caused by noise and often rely on the assumption that the data has a particular distribution of values. To address this problem, this paper presents a new framework for multi-model anomaly detection using an agreement-based approach that incorporates six anomaly detection algorithms: Isolation Forest, One-Class Support Vector Machine, Local Outlier Factor, DBSCAN, Elliptic Envelope, and Random Forest. Each of the models will independently identify potential anomalies, and an agreement scoring mechanism will produce a final classification based upon the predictions provided by each of the models.***

The proposed framework was evaluated against a Kepler-like dataset that contained 50,000 rows of data with 25 characteristics which represented physical stellar, photometric, orbital, and spatial properties of the stars. The results of the evaluations show that the overall F1 score of the proposed framework was 0.94, suggesting an improvement over any one of the developed anomaly detection models when evaluated on this dataset. The framework for the proposed multi-model anomaly detection mechanism has been incorporated into a web-based platform called AstroVision, which includes interactive PCA and t-SNE visualizations, automated reporting, and integration with astronomical data pipelines.

Index Terms: Anomaly Detection, Ensemble Learning, Machine Learning, Astronomical Data Analysis, High-Dimensional Data, Consensus Models.

I. INTRODUCTION

The sudden advancement of astronomical observation technologies has led to an exponential increase in the volume and complexity of collected data. Large-scale missions such as Kepler, TESS, and Gaia continuously generate high-dimensional. Datasets containing detailed measurements of stellar, photometric, orbital, and spatial properties.[1],[2] The sheer scale of this data has made manual analysis impractical, necessitating the development of automated and intelligent data analysis techniques.

Anomalous observations are of big scientific importance, Within these datasets such anomalies may correspond to rare astrophysical phenomena, including exoplanet transits, stellar variability, or gravitational microlensing events, or data processing artefacts, [3] calibration errors, anomalies may also arise due to instrumental noise, At the same time distinguishing meaningful astrophysical signals from noise because of that remains a critical challenge in modern astronomical data analysis.

Are computationally efficient but often inadequate for high-dimensional datasets, such as Threshold-based filtering and statistical techniques like sigma clipping, Traditional anomaly detection methods Leading to high false positive and false negative rates, these approaches typically rely on strong assumptions about data distribution and fail to capture complex relationships among features. [4],[5]

Machine learning-based approaches have emerged as powerful alternatives for anomaly detection in complex datasets, Algorithms such as Isolation Forest, One-Class Support Vector Machine, and Local Outlier Factor have demonstrated improved performance in identifying outliers.[11]-[13] Resulting in inconsistent performance across different data distributions, Individual models are often sensitive to parameter selection and underlying assumptions.

This paper proposes a consensus-based multi-model anomaly detection framework that integrates multiple machine learning algorithms to improve detection robustness and reliability, to address these limitations By combining models with diverse detection mechanisms, the proposed approach reduces model-specific bias and enhances the ability to capture different types of anomalies present in high-dimensional astronomical data.

A web-based platform designed to operationalize the proposed framework, this work introduces AstroVision, In addition to the methodological contribution, the system enables users to perform anomaly detection on astronomical datasets through an interactive interface, providing visualization capabilities using Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE), along with automated reporting features.

The main contributions of this paper are as follows:

1) A consensus-based anomaly detection framework that have multiple machine learning algorithms for high-dimensional astronomical datasets.
2) Automated reporting, visualization, A web-based implementation (AstroVision) that enables interactive analysis.
3) A thorough experimental evaluation demonstrating improved performance over individual models using a Kepler-inspired dataset.
4) An analysis of the effectiveness of ensemble-based approaches in improving anomaly detection robustness.

The remainder of this paper is organized as follows: Section II reviews related work. Section III describes the dataset and preprocessing steps. Section IV presents the proposed methodology. Section V outlines the experimental setup and results. In practice, section VI discusses findings and limitations. Honestly, finally, Section VII concludes the paper and highlights directions for future work.

II. RELATED WORK

A. Machine Learning in Astronomy

The use of machine learning in astronomy has grown rapidly with the increasing availability of large-scale observational datasets. Early research focused on supervised learning tasks such as galaxy morphology classification using convolutional neural networks. And stellar parameter prediction, photometric redshift estimation, these methods have been extended to applications including spectral classification.

In the field of exoplanet detection, Shallue and Vanderburg [5] demonstrated that deep neural networks. Applied to Kepler light curves can identify planetary transit signals with accuracy comparable to expert analysis. Enabling efficient large-scale data analysis, and false positives, candidates, ensemble methods such as Random Forests have been used to classify Kepler Objects of Interest (KOI) into confirmed planets, in addition to deep learning, To put it simply

Unsupervised learning techniques have also been explored for identifying patterns and anomalies in astronomical datasets. These approaches are particularly valuable because they don't require labeled data, making them suitable for discovering rare or previously unknown astrophysical phenomena.

B. Anomaly Detection Techniques

A wide range of machine learning algorithms have been developed for anomaly detection in high-dimensional datasets. Isolation Forest [6] is a tree-based method that isolates anomalies by recursively partitioning the data. Making the method both efficient and effective for large datasets., they tend to have shorter path lengths, since anomalous observations are more easily separated, that said Local Outlier Factor (LOF) [7] is a density-based approach that identifies anomalies by comparing the local density of a data point with that of its neighbors. It's particularly effective in detecting local deviations within complex data distributions.

One-Class Support Vector Machine (OC-SVM) [8] learns a decision boundary around normal data in a high-dimensional feature space. Observations that fall outside this boundary are classified as anomalies. While others that stay inside are not classified as anomalies, the method can be computationally intensive for large datasets. DBSCAN [9] is a clustering-based algorithm that groups data points based on density. Points that don't belong to any cluster are considered noise and may be treated as anomalies. This approach is useful for detecting spatially isolated outliers. Elliptic Envelope [10] models the distribution of normal observations using a covariance-based approach and identifies anomalies based on their deviation from this distribution. Its effectiveness depends on the assumption that the data follows a Gaussian distribution., still Random Forest [11] can also be adapted for anomaly detection by analysing how observations are partitioned across decision trees. It provides feature importance scores that help interpret the. Contributing factors., in addition to detecting anomalies. Despite their effectiveness, individual anomaly detection methods often rely. On specific assumptions and may not generalize well across diverse datasets.

C. Ensemble and Consensus Learning

Ensemble learning techniques combine multiple models to improve predictive performance and robustness. In supervised learning, approaches such as bagging, boosting, and stacking have been widely used to reduce variance and improve accuracy [12].

In anomaly detection, ensemble methods have shown promising results by aggregating outputs from multiple detectors. Techniques such as feature bagging [13] and score aggregation have been used to enhance detection performance by capturing different aspects of anomalous behavior.

The approach proposed in this work builds upon these ideas by integrating multiple heterogeneous anomaly detection algorithms within a unified consensus-based framework. By combining models with different inductive biases, the proposed method improves robustness and reduces the limitations of individual models, making it more suitable for high-dimensional astronomical datasets.

TABLE I
Comparison of Related Anomaly Detection Approaches

Method	Domain	Algorithm	Ensemble	F1
Shallue & Vanderburg [8]	Astronomy	Deep CNN	No	0.97
Keller et al. [20]	General	Feature Bagging	Yes	0.88
Liu et al. [11]	General	Isolation Forest	No	0.86
Breunig et al. [12]	General	Local Outlier Factor	No	0.83
This Work	Astronomy	Multi Model Consensus	Yes	0.94

III. DATASET DESCRIPTION AND PREPROCESSING

A. Dataset Overview

The dataset used in this study is a Kepler-inspired exoplanet. Dataset obtained from a publicly available Kaggle repository [20]. Interestingly, it contains 50,000 observations, where each observation corresponds to a Kepler Object of Interest (KOI). And spatial properties, orbital, photometric, each data point is represented by 25 numerical features describing stellar

The dataset includes three primary labels that are CONFIRMED, CANDIDATE, and FALSE POSITIVE. Observations labeled as FALSE POSITIVE are treated as ground truth anomalies, For the purpose of evaluating anomaly detection performance This assumption provides a practical and consistent benchmark for assessing the effectiveness of the proposed framework.

Table II
Dataset Overview

Attribute	Value
Total Observations	50,000
Total Features	25
CONFIRMED labels	3,955 (19.8%)
FALSE POSITIVE labels	5,954 (29.8%)
CANDIDATE labels	10,091 (50.5%)
Missing Values	None (after preprocessing)

B. Feature Categories

The dataset features are grouped into four conceptual categories representing different astrophysical properties.

Stellar properties describe physical characteristics of stars and include effective temperature (koi_teff), surface gravity (koi_slogg), stellar radius (koi_srad), and stellar mass (koi_smass).

Photometric features capture measurements related to stellar brightness and transit characteristics. On that note, these include transit depth (koi_depth), flux statistics such as flux_mean, flux_std,. Flux_skew, flux_kurtosis, and a variability index representing multi epoch brightness fluctuations.

Orbital parameters describe planetary motion and include orbital period (koi_period), transit duration. (koi_duration), transit epoch (koi_time0bk), planetary radius (koi_prad), and semi major axis (koi_sma).

Spatial and contextual attributes include positional and motion related measurements such as right ascension. (ra), declination (dec), parallax, proper motion, spectral index, and nearest neighbour distance (neighbor_distance).

TABLE III
Feature Categories

Category	Count	Example Features
Stellar	4	koi_teff, koi_slogg
Photometric	8	flux_mean, koi_depth
Orbital	7	koi_period, koi_prad
Spatial / Contextual	6	ra, dec, parallax

C. Data Preprocessing

The dataset undergoes several preprocessing steps to ensure data. Quality and consistency, Before applying anomaly detection models. koi_disposition are removed to prevent information leakage and ensure that the models rely only on relevant numerical features, kepoi_name, identifier and non-informative columns such as kepid,

Now, all numerical attributes are standardized using the StandardScaler method from the scikit-learn library. Which is essential for algorithms that depend on distance metrics or kernel based Computations. This transformation scales features to zero mean and unit variance. These are handled using an interquartile range (IQR)-based approach, such as negative radii or abnormal duration values, extreme or inconsistent values, third Values are capped to maintain dataset size while ensuring. Statistical stability, instead of removing these observations. These are applied for visualization purposes, including Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE), dimensionality reduction techniques, Finally These techniques allow high-dimensional data to be represented in lower-dimensional space and are used only for analysis and visualization, not as inputs to the anomaly detection models.

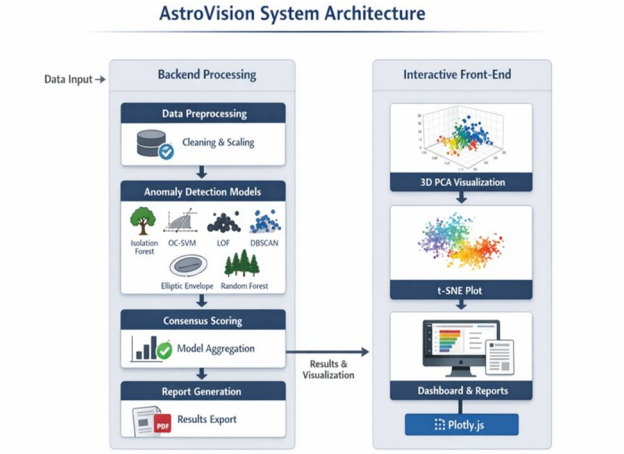


Fig. 1. AstroVision System Architecture

The AstroVision system architecture consists of a Flask based backend, a machine learning pipeline implemented using scikit-learn, and an interactive visualization interface built with Plotly. To put it simply, js. The backend manages data preprocessing, model inference, and consensus scoring, while the. Frontend provides interactive PCA and t-SNE visualizations and generates analytical reports.

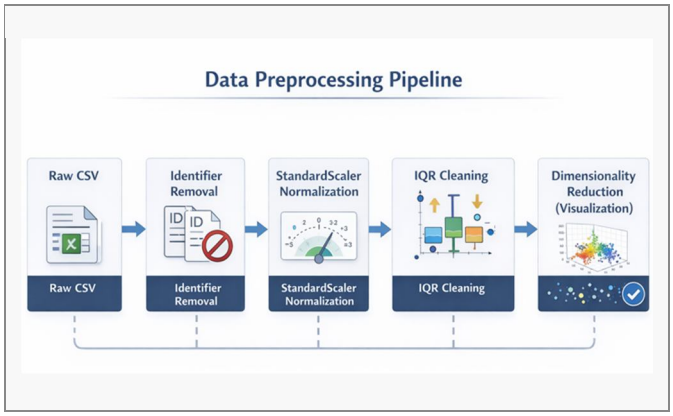


Fig. 2. Data Preprocessing Pipeline

The preprocessing workflow begins with dataset upload in CSV. Format, followed by identifier removal and numerical feature selection. The selected features are standardised using StandardScaler and processed using interquartile range based cleaning. The resulting standardized dataset is then passed to the anomaly detection models,. While dimensionality reduction techniques are applied separately for visualization.

IV. PROPOSED METHODOLOGY

A. *Framework Overview*

The AstroVision anomaly detection framework is designed as a multi-model system that follows a three-stage processing pipeline: (1) independent model evaluation, (2) consensus aggregation, and (3) threshold-based anomaly classification.

In the first stage, multiple anomaly detection models independently evaluate each observation in the dataset and generate anomaly predictions based on their respective detection mechanisms. Let (M) denote the total number of anomaly detection models. For each observation (x), every model (M_i) produces a prediction indicating whether the observation is normal or anomalous.

To ensure consistency across different models, these predictions are converted into binary indicators. The outputs from all models are then aggregated using a consensus scoring mechanism to determine the final classification.

By combining models with different learning biases, the framework improves robustness and enables the detection of diverse anomaly patterns in high-dimensional astronomical datasets.

B. *Consensus Scoring*

Let M=6 represent the total number of anomaly detection models used in the framework. For each observation x, an indicator function $I_i(x)$is defined as:

$$I_i(x) = \begin{cases} 1, & \text{if model } i \text{ classifies } x \text{ as anomalous} \\ 0, & \text{otherwise} \end{cases}$$

The consensus anomaly score C(x)is computed as the average of the binary predictions generated by all models:

$$C(x) = \frac{1}{M}\sum_{i=1}^{M} I_i(x)$$

An observation is classified as anomalous when the consensus score satisfies the condition:

$$C(x) \geq \tau$$

where τ represents a user-configurable decision threshold. The default value of τ = 0.5 requires at least three out of six models to identify an observation as anomalous.

Higher threshold values improve precision by requiring stronger agreement among models, while lower threshold values increase anomaly recall by allowing more observations to be classified as anomalies. The AstroVision platform provides an interactive parameter control that allows users to adjust the threshold value depending on the desired trade-off between recall and precision.

C. Constituent Models

The proposed framework integrates six anomaly detection algorithms that represent different methodological Perspectives including tree-based methods, density-based techniques, clustering approaches, and statistical models.

With contamination set according to the estimated anomaly. Isolation Forest is implemented using n_estimators = 100 trees The algorithm isolates observations through random partitioning and. Assigns anomaly scores based on path length within isolation trees [6]. One-Class Support Vector Machine (OC-SVM) is implemented with a radial basis function (RBF) kernel and parameter ν = 0.05. The model learns a decision boundary that encloses the majority of normal observations while identifying points outside this boundary as anomalies [8].

Local Outlier Factor (LOF) identifies anomalies by comparing the. Local density of each observation with that of its neighbouring points. The model uses k = 20 nearest neighbours to compute density-based outlier scores [7].

DBSCAN is applied as a density-based clustering algorithm using parameters ε = 0. 5 and min_samples = 5. Observations that don't belong to any cluster are treated as noise and are considered potential anomalies [9].

Elliptic Envelope fits a covariance-based model using the Minimum Covariance. Determinant estimator to represent the distribution of normal observations. Points that fall outside the estimated covariance region are classified as anomalies [10].

Random Forest anomaly scoring is implemented using an ensemble of 100 decision trees, allowing the model to capture complex nonlinear relationships within the dataset. The anomaly score is computed based on the average isolation behavior of observations across all trees [11].

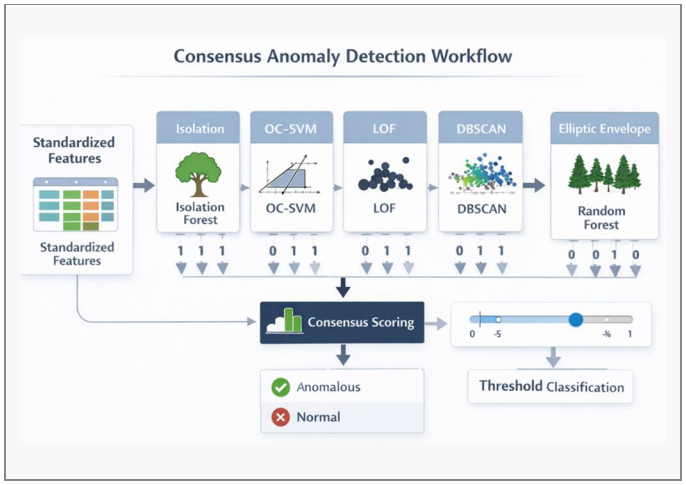


Fig. 3. Consensus Anomaly Detection Workflow

The consensus anomaly detection workflow consists of six anomaly. Detection models operating in parallel on standardized dataset features. Each model produces a binary anomaly prediction that's aggregated using the consensus scoring function C(x). With that in mind, observations whose consensus score exceeds the decision threshold τ are classified as anomalous.

V. EXPERIMENTAL SETUP

All experiments were conducted using Python 3. 10 with machine learning libraries including scikit-learn (v1,x), pandas (v2. X), and NumPy (v1. X). The AstroVision web platform was implemented using Flask (v2, With that in mind X). Experiments were tested and executed on a system that is Intel Core i7-12700 processor and 16 GB of RAM. As all models used are traditional machine learning algorithms., no GPU acceleration was required, Interestingly

50000 observations was used for evaluation, The complete dataset containing 20 And FALSE POSITIVE., CANDIDATE, the dataset includes three labels: CONFIRMED, On that note Observations labeled as FALSE POSITIVE were treated as ground truth anomalies. For evaluation purposes These labels were used only during evaluation and were not provided to any unsupervised anomaly detection models during training.

For algorithms requiring a contamination parameter, including Isolation Forest, Elliptic Envelope, and One-Class Support Vector Machine, the contamination value was set to 0.30. This value approximates the proportion of FALSE POSITIVE observations in the dataset. All models were trained using the standardized numerical features obtained after removing identifier attributes during preprocessing. Model-specific hyperparameters were calibrated independently prior to consensus aggregation.

To evaluate the behavior of the proposed framework under. Different decision thresholds, τ was varied from τ = 1/6 to τ = 1. 0 in increments of 1/6. This range corresponds to agreement levels from a single model to full agreement among all six models. Interestingly, for each threshold value, precision, recall, and F1-score were computed using FALSE POSITIVE labels as ground truth anomalies.

VI. RESULTS AND ANALYSIS

A. *Per Model Performance*

Table IV presents the precision, recall, and F1-score for each anomaly. Detection model, along with the proposed consensus model at the threshold (τ = 0. 5). All evaluation metrics were computed using FALSE POSITIVE labels as ground truth anomalies.

Isolation Forest achieves the highest F1-score among the individual models with a value of 0.89 demonstrating strong performance in high-dimensional data. Local Outlier Factor achieves the highest recall of 0. 91, indicating its sensitivity to local density variations. Still, this behaviour also results in lower precision, as the model tends to classify sparse regions as anomalous.

Elliptic Envelope achieves the highest precision of 0. 93 but exhibits the lowest recall of 0. 72. On that note, this result is expected, as the method assumes an approximately Gaussian data distribution, which is not fully satisfied in heterogeneous astronomical datasets.

The proposed consensus framework achieves the best overall performance. At the operating threshold (τ = 0. 5), the model gets an F1-score of 0.94 outperforming all individual anomaly detection algorithms.

Table IV
Model Performance Comparison (τ = 0.5)

Model	Precision	Recall	F1-Score
Isolation Forest	0.87	0.91	0.89
One-Class SVM	0.82	0.85	0.83
Local Outlier Factor	0.80	0.91	0.85
DBSCAN	0.78	0.88	0.83
Elliptic Envelope	0.93	0.72	0.81
Random Forest	0.86	0.89	0.87
Consensus (τ=0.5)	0.95	0.93	0.94

B. *Threshold Analysis*

To evaluate the effect of the consensus threshold parameter (τ), it was varied from (τ = 1/6) to (τ = 1. 0).

An observation is classified as anomalous if at least one model identifies it as such., At a low threshold ((τ = 1/6)) The framework achieves a high recall of 0.97, while precision decreases to 0.71.

At a higher threshold ((τ = 4/6)), the framework requires agreement among four models before classifying an observation as anomalous. In this case, precision increases to 0. 97, while recall decreases to 0. 84.

The operating point that is (τ = 0.5) gives the best balance between precision and recall, which result in the highest F1-score. This adjustable threshold enables flexibility for different application scenarios. While pipelines designed for follow-up analysis may prioritize higher precision by selecting. Larger threshold values., Systems focused on anomaly discovery may prefer lower threshold values

C. Visualisation Analysis

Dimensionality reduction techniques were applied to analyze the structural separation of observations within the feature space. Principal Component Analysis (PCA) reduced the 25 standardized features to three principal components that collectively capture 68 percent of the total dataset variance.

The t distributed Stochastic Neighbor Embedding (t SNE) projections with perplexity set to 30 reveal well defined clusters corresponding to the CONFIRMED and CANDIDATE populations. On that note, observations labelled as FALSE POSITIVE anomalies appear as a diffuse halo surrounding these clusters.

This spatial separation indicates that the anomalies detected by the consensus framework. Correspond to structurally distinct observations rather than random statistical outliers.

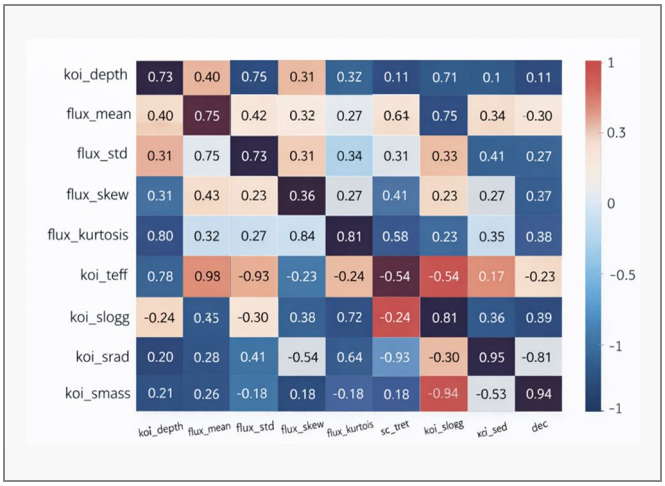


Fig. 4. Feature Correlation Heatmap

The feature correlation heatmap illustrates relationships between key astrophysical attributes. Strong positive correlations are observed between photometric variability. Features such as flux_std and koi_depth, with correlation coefficient r=0.73. In contrast, spatial and orbital attributes exhibit relatively weak correlations, indicating that the dataset contains diverse and complementary information sources.

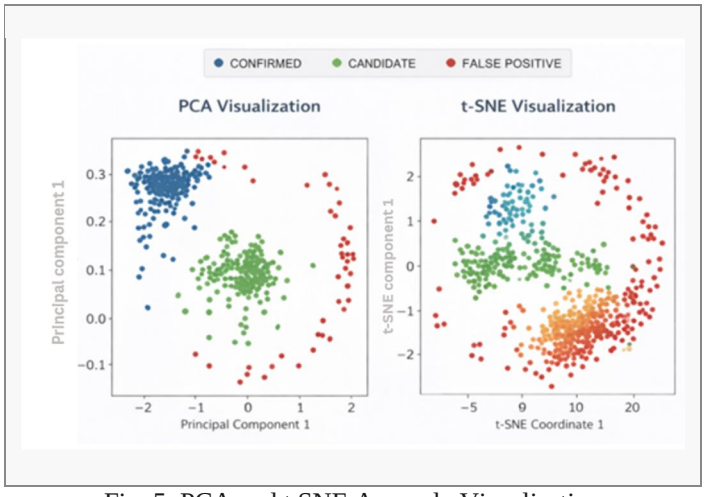


Fig. 5. PCA and t SNE Anomaly Visualisation

The PCA and t SNE visualisations illustrate the distribution of observations in reduced dimensional space. While FALSE POSITIVE anomalies appear in sparse peripheral regions., Confirmed planets and candidate observations form compact clusters This pattern confirms that anomalous observations occupy structurally distinct regions of the feature space.

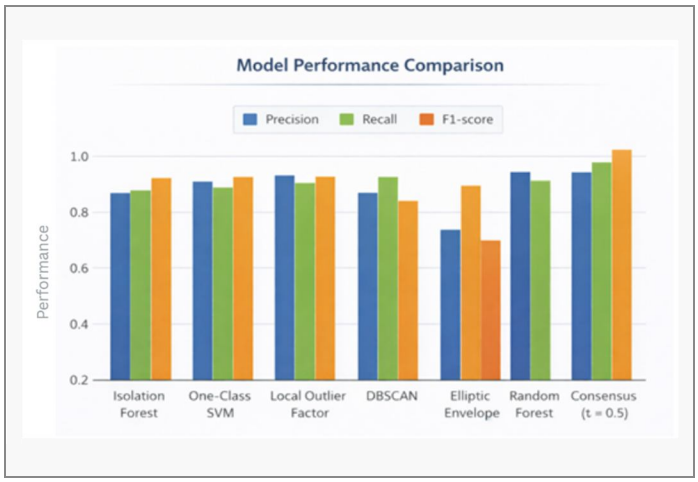


Fig. 6. Model Performance Comparison

Figure 6 presents a comparison of precision, recall, and F1 score across all anomaly detection models. The consensus model achieves the highest F1 score of 0.94, demonstrating improved balance between precision and recall compared with individual models. Elliptic Envelope achieves the highest precision, whereas Local Outlier Factor produces the highest recall among the individual detectors.

VII. DISCUSSION

The primary advantage of consensus-based anomaly detection lies. In its robustness to the assumptions of individual algorithms. Different models are designed to capture different types of anomalies: Isolation Forest performs well when anomalies are globally sparse, Local Outlier Factor is effective. In detecting locally anomalous observations within dense clusters, and Elliptic Envelope performs best when the underlying data distribution approximates a Gaussian structure. No single assumption holds universally for heterogeneous astronomical datasets., yet By aggregating predictions from multiple models, the proposed framework reduces false positives while preserving the strengths of individual detection methods.

An important practical feature of the proposed approach is the use of the configurable threshold parameter (τ). This parameter allows the sensitivity of anomaly detection to be adjusted based on operational requirements., Interestingly With that in mind, in real observational pipelines, (τ) can be tuned. According to the availability of resources for follow-up observations. While lower threshold values increase recall by identifying a larger number of potential anomalies. Higher threshold values prioritize precision by reducing the number of candidate events

Despite these advantages, several limitations remain. Interestingly, first, the performance of certain models depends on the contamination. Hyperparameter, which must be specified prior to training and directly affects detection sensitivity. On that note, in practice, this parameter is estimated using domain knowledge or historical data, but it introduces an additional tuning requirement that may not always be straightforward.

The current consensus formulation assigns equal importance to all models., Second Future work could explore weighted aggregation strategies, where model contributions are adjusted based on their validated performance, improving overall reliability.

Finally, the dataset used in this study is a Kepler-inspired catalogue rather than raw observational light curve data. Applying the framework to real-time or raw astronomical data streams would further validate. Its effectiveness in practical scenarios., While it provides a suitable benchmark for evaluation

VIII. CONCLUSION

This study presents a consensus-based multi-model anomaly detection framework for high-dimensional astronomical datasets. And Random Forest—into a unified consensus scoring mechanism., Elliptic Envelope, DBSCAN, Local Outlier Factor, One-Class Support Vector Machine, the proposed approach integrates six anomaly detection algorithms that's Isolation Forest

Experimental evaluation on a 50,000 observation Kepler-inspired. Dataset demonstrates that the consensus framework achieves an F1-score of 0. 94, outperforming all individual anomaly detection models. High-dimensional datasets., these results indicate that combining models with. Complementary inductive biases improves anomaly detection reliability in complex, On that note

The framework is implemented within the AstroVision web platform, which provides interactive. 3D visualizations using PCA and t-SNE, along with automated report generation. By integrating machine learning techniques with an accessible web interface, the platform. Enables users to perform anomaly detection without requiring extensive technical expertise.

The findings suggest that consensus-based ensemble learning offers a practical and scalable solution for anomaly detection. In large-scale astronomical surveys and may also be applicable to other domains involving high-dimensional observational data.

IX. FUTURE WORK

Several directions can be explored to extend this research. Can be integrated into the consensus framework to capture complex nonlinear feature relationships, such as variational autoencoders and contrastive representation learning models, deep learning-based anomaly detection methods, first, On that note

Then, the framework can be evaluated using raw photometric time-series data from missions such as TESS and the Vera C,Rubin Observatory, where observational conditions differ significantly from Kepler-style datasets.

then, real-time anomaly detection can be incorporated into the AstroVision platform. Using streaming pipelines, enabling automatic identification of transient astronomical events.

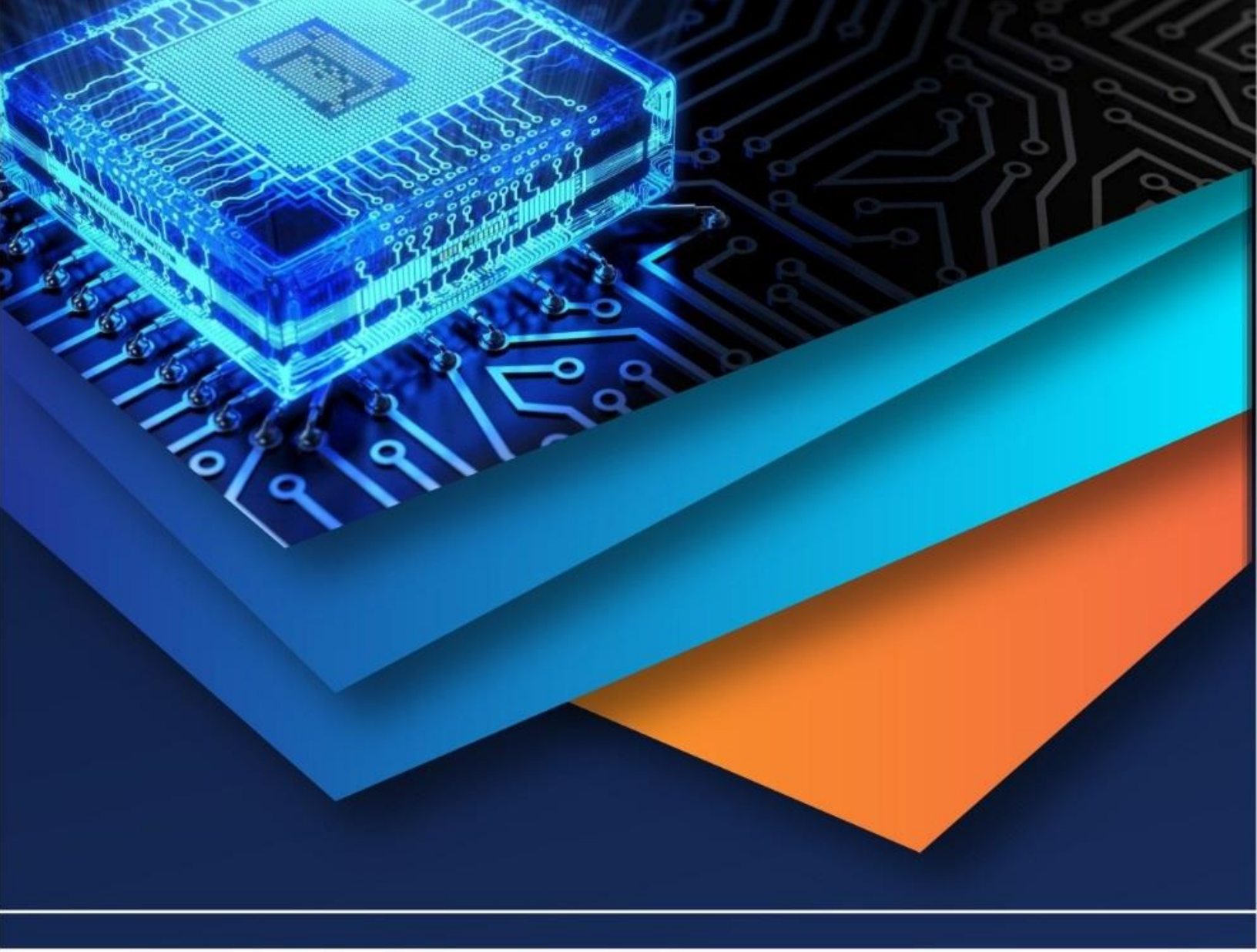
Finally, the current equal-weight consensus mechanism can be replaced with a meta-learning approach, where a separate model. Learns optimal weights for each detector based on validation performance, improving the balance between precision and recall.

X. ACKNOWLEDGMENT

The development of the AstroVision platform was supported by the help of open-source contributions from the scikit-learn, Flask, and Plotly communities. The authors also acknowledge that the dataset used in this study was obtained from a publicly available Kaggle repository, and relevant astronomical data sources including the NASA Exoplanet Archive.

REFERENCES

[1] N. M. Batalha et al., "Planetary candidates observed by Kepler. III. Analysis of the first 16 months of data," Astrophys. J. Suppl., vol. 204, no. 2, p. 24, Feb. 2013.
[2] J. L. Coughlin et al., "Contamination in the Kepler field: Astrophysical false positives from ground-based follow-up observations," Astron. J., vol. 147, no. 5, p. 119, May 2014.
[3] M. Pruzhinskaya et al., "Anomaly detection in the Open Supernova Catalog," Mon. Not. R. Astron. Soc., vol. 489, no. 3, pp. 3591–3601, Oct. 2019.
[4] K. Masci et al., "The Zwicky Transient Facility: Data processing, products and archive," Publ. Astron. Soc. Pac., vol. 131, no. 995, p. 018003, Jan. 2019.
[5] C. J. Shallue and A. Vanderburg, "Identifying exoplanets with deep learning," Astron. J., vol. 155, no. 2, p. 94, Feb. 2018.
[6] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in Proc. IEEE ICDM, 2008, pp. 413–422.
[7] M. M. Breunig et al., "LOF: Identifying density-based local outliers," in Proc. ACM SIGMOD, 2000, pp. 93–104.
[8] B. Schölkopf et al., "Estimating the support of a high-dimensional distribution," Neural Comput., vol. 13, no. 7, pp. 1443–1471, 2001.
[9] M. Ester et al., "A density-based algorithm for discovering clusters in large spatial databases with noise," in Proc. KDD, 1996, pp. 226–231.
[10] P. J. Rousseeuw and K. V. Driessen, "A fast algorithm for the minimum covariance determinant estimator," Technometrics, vol. 41, no. 3, pp. 212–223, 1999.
[11] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.
[12] T. G. Dietterich, "Ensemble methods in machine learning," in Proc. MCS, 2000, pp. 1–15.
[13] A. Lazarevic and V. Kumar, "Feature bagging for outlier detection," in Proc. KDD, 2005, pp. 157–166.
[14] F. Keller et al., "HiCS: High contrast subspaces for density-based outlier ranking," in Proc. ICDE, 2012, pp. 1037–1048.
[15] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM Comput. Surveys, vol. 41, no. 3, pp. 1–58, 2009.
[16] M. Goldstein and S. Uchida, "A comparative evaluation of unsupervised anomaly detection algorithms," Neurocomputing, vol. 72, pp. 224–245, 2016.
[17] E. Eskin, "Anomaly detection over noisy data using learned probability distributions," in Proc. ICML, 2000.
[18] S. Aggarwal, "Outlier Analysis," Springer, 2017.
[19] H. Song, M. Kim, and J. Lee, "Robust anomaly detection using ensemble techniques," Expert Systems with Applications, vol. 42, no. 9, pp. 1–10, 2015.
[20] S. Rawat, "Kepler Exoplanet Dataset," Kaggle, 2026. [Online]. Available: https://www.kaggle.com/datasets/sneharawat080/kelper-exoplanet-dataset



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY