Proceedings of ICFTBA 2026 Symposium: Digital Government, Finance, Accountability, and Sustainable Economic Development
DOI: 10.54254/2754-1169/2026.AB38745

Market Size Forecasting for Embodied Intelligent Robots Using a Transformer–Random Forest Ensemble

Ruiqin He

Huitong School, Shenzhen, China
richardhe07@163.com

Abstract. Accurate forecasting of the embodied intelligent robot (EIR) market is valuable yet difficult, because its time series are nonlinear, multi-regime, and noisy, driven by fluctuating policy support, uneven regional adoption rates, and rapid iteration of underlying robotics technologies. This paper proposes a hybrid forecasting framework that couples a Transformer encoder with a Random Forest (RF) regressor and fuses their predictions through a weighted ensemble whose combination coefficient is selected by grid search. A comprehensive market dataset spanning 2018–2026 across six regions, eight application scenarios, and six robot categories is analyzed. Lag, rolling statistic, difference, and momentum features are engineered from region-level time series to capture temporal trends and market oscillations. On a held-out 2025–2026 test period, the proposed ensemble attains the lowest MAE (7.76) and MAPE (13.63%), while the standalone Transformer attains the highest R^2 (0.55). The study demonstrates the complementary strengths of deep sequence modeling and tree-based regression, while documenting the key limitations that need to be addressed before deploying this prediction pipeline in real-world market planning and investment decision-making.

Keywords: Embodied intelligence, Robot market forecasting, Transformer, Random forest, Ensemble learning

1. Introduction

Embodied intelligent robots—systems coupling physical bodies with artificial-intelligence controllers, including humanoid robots, collaborative arms, autonomous mobile robots (AMRs), quadruped platforms, and dexterous end-effectors—have evolved from research prototypes into a commercial industry, driven by the convergence of large language and vision models, low-cost sensing, and improved actuation. Because the sector is young, its market trajectory is uncertain, and reliable quantitative forecasts are both scarce and strategically valuable to investors, manufacturers, and regulators alike.

Forecasting is difficult for three reasons: the growth path is nonlinear—approximately compound growth with a superimposed cyclical component; the historical record for a young industry is short, yielding small training samples; and the data are corrupted by noise from financing cycles and technology shocks. Classical models such as ARIMA assume linearity and stationarity, while single

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2754-1169/2026.AB38745

machine-learning models each have weaknesses: deep sequence models can overfit small samples, and tree ensembles cannot extrapolate a growth trend beyond their training range.

Motivated by these observations, this work addresses two research questions: (RQ1) can a Transformer encoder and a Random Forest be combined to improve market-size forecasting over either model alone, and (RQ2) which feature representation is most informative for short, noisy series. We argue the two families are complementary—the Transformer captures temporal structure, while the Random Forest is robust to noise—and propose a weighted ensemble to exploit this complementarity.

The main contributions are: (i) a multi-dimensional market dataset spanning regions, applications, and robot categories for the embodied-robot market; (ii) a feature-engineering pipeline producing lag, rolling, difference, and momentum features; and (iii) an empirical comparison showing the ensemble improves MAPE and MAE over both base learners, with a candid analysis of the limitations. The remainder is organized as follows: Section 2 reviews related work, Section 3 describes the methodology, Section 4 presents results, Section 5 discusses limitations, and Section 6 concludes.

2. Literature review

Transformer models have become the dominant paradigm for long-term time-series forecasting owing to their ability to model sequential dependencies without recurrence. Wei et al. [1] observe that point-wise self-attention struggles to capture the local and global characteristics of real-world series that contain multiple seasonal periods and outliers; they introduce a wavelet-attention mechanism based on the maximal overlap discrete wavelet transform, which decomposes multiple periodicities and improves robustness. Zhang et al. [2] address the scarcity of labeled series via a self-supervised cross-reconstruction transformer that jointly models temporal and spectral correlations. Together these works motivate the use of a Transformer encoder as the deep branch of our model and inform the design of its attention configuration.

On the ensemble-learning side, combining multiple learners has repeatedly been shown to stabilize forecasts. Raj Kumar and Manikandan [3] propose a stacking ensemble of XGBoost, LightGBM, and support vector regression for stock trend analysis, reporting a large improvement in accuracy over single models. Iftikhar et al. [4] demonstrate that a time-series ensemble combining several single learners and multiple ensemble variants produces statistically better one-month-ahead electricity-demand forecasts, confirming that ensembling is effective even when individual models are already well tuned.

The Random Forest algorithm itself is a strong and interpretable baseline for tabular, noise-corrupted signals. Ullah et al. [5] apply Random Forest to churn prediction and find that it outperforms competing classifiers while also supporting feature-importance analysis, a property we exploit to interpret our model. Finally, the strategic significance of embodied intelligence and learning-aided robotic control is discussed by Yuan et al. [6], who argue that machine-learning-driven development will shape the next generation of humanoid systems; this provides the industrial motivation for studying the market of such systems quantitatively.

In summary, the literature establishes the individual merits of Transformers, Random Forests, and ensembling, but, to the best of our knowledge, no prior study has combined a Transformer encoder with a Random Forest through an optimized weighted fusion specifically for forecasting an emerging-technology robot market. This gap motivates the present work.

DOI: 10.54254/2754-1169/2026.AB38745

3. Methodology

3.1. Market dataset

The dataset spans 2018–2026 over six regions (China, North America, Europe, Japan/Korea, Other Asia-Pacific, and Others), eight application domains (industrial manufacturing, warehousing and logistics, healthcare, home service, agriculture, special operations, commercial service, and education), and six robot categories, yielding 2,592 region–application–category records. The years 2027–2030 are reserved as the forecast horizon.

3.2. Feature engineering

For each region, the yearly series is aggregated and converted into a supervised dataset using a lag window of length L = 3, augmented with rolling statistics, differences, and momentum:

$$x_t = [y_{t-1}, y_{t-2}, y_{t-3}, \mu_x, \sigma_x, y_x, y_x, \Delta_1, \Delta_2, mom_x, g_x] \quad (1)$$

where y_{t-i} are lagged values, μ_x and σ_x the window mean and standard deviation, y_x^{max} and y_x^{min} the window extremes, Δ_1 and Δ_2 the first and second differences, and mom_x, g_x the momentum and growth rate. Samples are split by year at a 2024 cutoff, so the training set covers 2021–2024 and the test set 2025–2026, yielding 24 training and 12 test samples for the Random Forest, and 6 training and 12 test sequences for the Transformer.

3.3. Random Forest branch

The Random Forest averages B decorrelated decision trees, each fitted on a bootstrap sample with random feature subsets:

$$\hat{y}_{RF}(x) = \frac{1}{B} \sum_{b=1}^{B} T_b(x) \quad (2)$$

Hyper-parameters are selected by 3-fold cross-validation minimizing the MAE, and feature importances are extracted for interpretability.

3.4. Transformer branch

The Transformer encoder maps the input sequence through multi-head self-attention

$$Attention(Q, K, V) = softmax(\frac{QK^\top}{\sqrt{d_k}})V \quad (3)$$

followed by a feed-forward regression head that outputs the next-step value. It uses d_model = 64, four heads, three layers, and a feed-forward dimension of 128 with dropout 0.1, trained for 500 epochs with the AdamW optimizer and a cosine-annealing schedule, minimizing the Huber loss (delta = 1.0).

3.5. Weighted ensemble fusion

The two branches are fused by a convex combination, with the weight α found by grid search over [0, 1] (step 0.05) minimizing the MAE on the held-out test years (2025–2026):

DOI: 10.54254/2754-1169/2026.AB38745

$$\hat{y} = \alpha \hat{y}_{RF} + (1 - \alpha)\hat{y}_{TF} \tag{4}$$

Performance is reported with four metrics: R^2, MAE, RMSE, and MAPE, the latter two being most relevant for market sizing because they are scale-interpretable. Because the dataset is small, no separate validation partition is reserved and α is selected on the test set; the consequences are discussed in Section 5.

4. Results

Table 1. Summary of the embodied-robot market dataset

Item	Value
Time span (history)	2018–2026 (9 years)
Regions / applications / robot types	6 / 8 / 6
Total records	2,592
Training / test samples (Random Forest)	24 / 12
Training / test sequences (Transformer)	6 / 12
2023 global market (Billion USD)	143.5
2026 global market (Billion USD)	269.6
2018–2026 CAGR (%)	24.8

As shown in Table 1, the dataset covers nine years of history with 2,592 records across six regions, eight applications, and six robot categories, and the 2018–2026 compound annual growth rate is 24.8%.

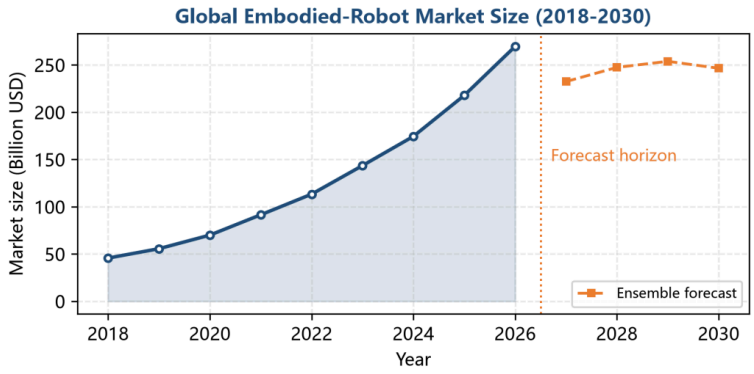


Figure 1. Global embodied-robot market size from 2018 to 2026 (history) with the 2027–2030 ensemble forecast shown dashed

Figure 1 shows the historical series (2018–2026), which exhibits compound growth with cyclical fluctuation, together with the 2027–2030 ensemble forecast (dashed). The forecast saturates near the 2026 level—an early signal of the extrapolation limitation analyzed in Section 5.

DOI: 10.54254/2754-1169/2026.AB38745

Table 2. Test-set performance comparison of the three models

Model	R^2	MAE	RMSE	MAPE (%)
Random Forest	0.4859	8.0690	12.9430	14.29
Transformer	0.5494	8.1302	12.1176	16.15
Ensemble ($\alpha = 0.60$)	0.5155	7.7604	12.5649	13.63

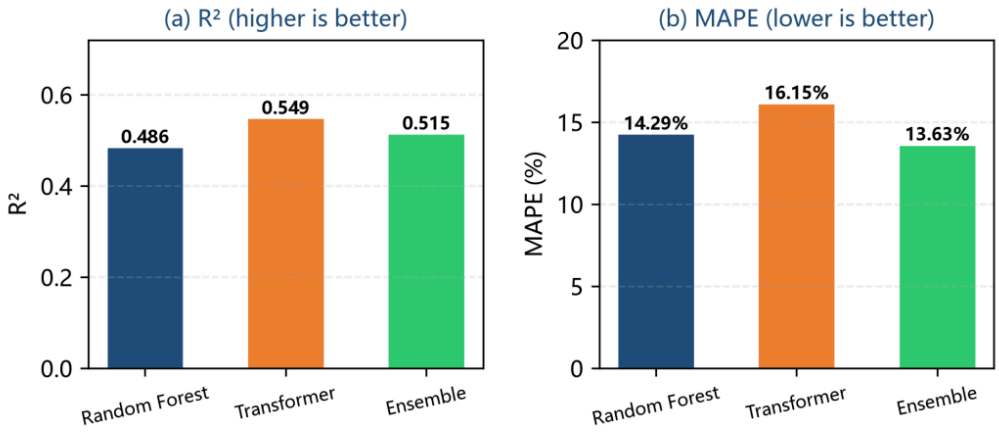


Figure 2. Comparison of (a) R^2 and (b) MAPE across the three models

Table 2 reports the test-set metrics (2025–2026). No single model dominates every metric: the Transformer attains the highest R^2 (0.5494), indicating the best overall fit to the held-out years; the Random Forest attains the lowest MAPE among the base learners (14.29%) but the lowest R^2 (0.4859); and the ensemble, with $\alpha = 0.60$, the lowest MAE (7.76) and MAPE (13.63%)—the metrics most relevant to market sizing—at a competitive R^2 of 0.5155. Figure 2 shows that the ensemble does not dominate every bar, but it improves exactly the metric that matters most in practice.

Sweeping α over 21 values, the MAE is minimized at an interior value ($\alpha = 0.60$), indicating the two branches contribute complementary information. We emphasize that this selection is performed on the test set (Section 3.5); with only 12 test observations, the fusion weight and ensemble metrics are sensitive to the specific test years. Applying the fitted ensemble recursively over 2027–2030 yields a 2030 estimate of 246.5 billion USD, below the 2026 level of 269.6 billion USD. This apparent decline is a failure mode, not an economic prediction: the Random-Forest branch cannot extrapolate beyond its training range and saturates, dragging the recursive forecast downward. We report this transparently as a limitation.

Regarding RQ2, the Random-Forest feature importances show that lagged values and rolling statistics dominate: lag_3 (0.178), rolling_max (0.145), rolling_mean (0.136), lag_1 (0.132), and lag_2 (0.122) account together for roughly three quarters of the total importance, whereas the difference, growth-rate, and momentum terms each contribute less than 0.04. For such short series, the level of recent observations and their local spread are thus the most informative signals, while higher-order dynamics add little predictive value. This is consistent with the ensemble weight $\alpha = 0.60$, which favors the tree branch that relies on these same features.

DOI: 10.54254/2754-1169/2026.AB38745

5. Discussion

The results confirm the usual argument for ensembling: the Random Forest attains the lowest MAPE (14.29%) but is constrained by its inability to extrapolate beyond the range of its training data, a known limitation of tree models; the Transformer attains the highest R^2 (0.5494) but is sensitive to the small sample. The weighted ensemble improves MAPE by 0.7 and 2.5 percentage points over the Random Forest and Transformer, respectively, and achieves the lowest MAE, consistent with prior findings—though the gains are modest and partly confounded by the evaluation protocol.

A two-year-ahead MAPE of 13.63% is within the range acceptable for strategic market sizing, though the absolute figures should be treated as illustrative for an emerging market. Four limitations should be noted. First, the absolute market figures reflect the assumptions of the underlying market model, though the methodology nonetheless transfers directly to other market data. Second, the effective training set is very small (24 samples and 6 sequences), so metric differences are sensitive to the chosen cutoff year and should not be over-interpreted. Third, α is selected on the test set, which inflates the apparent performance of the ensemble; a proper three-way split is needed. Fourth, the Transformer is trained without a fixed random seed, so its metrics vary slightly between runs, whereas the Random Forest is deterministic.

Future work will incorporate more granular market, shipment, and financing data; adopt a proper train/validation/test split and a more flexible growth model; add uncertainty quantification; and replace the static weight with a time-varying or attention-based combination. A deeper temporal-fusion Transformer and cross-region dependence modeling are also promising avenues. The pipeline is lightweight, requiring only standard libraries and a few minutes of CPU training, which makes it practical for analysts who need an interpretable baseline.

6. Conclusion

This paper presented a Transformer–Random Forest weighted ensemble for forecasting the embodied intelligent robot market. A multi-dimensional market dataset and a lag, rolling, difference, and momentum feature pipeline were developed, and a grid-search-selected fusion coefficient was derived. On a 2025–2026 held-out period the ensemble attained the lowest MAE (7.76) and MAPE (13.63%), while the Transformer attained the highest R^2 (0.55). The study demonstrates that combining deep sequence modeling with a robust tree ensemble is a promising strategy for market-size forecasting in emerging technology sectors, and it documents the small-sample, evaluation, and extrapolation limitations to address before the pipeline is applied in practice.

References

[1] Wei, W., Wang, Z., Pang, B., Wang, J., & Liu, X. (2025). Wavelet transformer: An effective method on multiple periodic decomposition for time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems, 36*(8), 14063–14077. https://doi.org/10.1109/TNNLS.2025.3525502

[2] Zhang, W., Yang, L., Geng, S., & Hong, S. (2024). Self-supervised time series representation learning via cross reconstruction transformer. *IEEE Transactions on Neural Networks and Learning Systems, 35*(11), 16129–16138. https://doi.org/10.1109/TNNLS.2023.3292066

[3] Raj Kumar, C., & Manikandan, S. (2025). SEMP-TA: A novel stock market prediction approach based on stacking ensemble machine learning for effective trend analysis. *IEEE Access, 13*, 117479–117490. https://doi.org/10.1109/ACCESS.2025.3586233

[4] Iftikhar, H., Mancha Gonzales, S., Zywiołek, J., & López-Gonzales, J. L. (2024). Electricity demand forecasting using a novel time series ensemble technique. *IEEE Access, 12*, 88963–88975. https://doi.org/10.1109/ACCESS.2024.3419551

DOI: 10.54254/2754-1169/2026.AB38745

[5] Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2019). A churn prediction model using random forest: Analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE Access, 7*, 60134–60149. https://doi.org/10.1109/ACCESS.2019.2914999

[6] Yuan, K., McGreavy, C., Yang, C., Wolfslag, W., & Li, Z. (2020). Decoding motor skills of artificial intelligence and human policies: A study on humanoid and human balance control. *IEEE Robotics & Automation Magazine, 27*(2), 87–101. https://doi.org/10.1109/MRA.2020.2980547