© APR 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717258

Flight Price Prediction Using Machine Learning and Deep Learning: A Comparative Study

DHANUSH[1], PURNA SATWIK[2], SANDEEP[3], A. RAMA PRATHAP REDDY[4]
[1, 2, 3] *Department of Computer Science and Engineering, R.V.R & J.C College of Engineering*
[4] *Assistant Professor, Department of Computer Science and Engineering, R.V.R & J.C College of Engineering*

Abstract*— Airfare pricing is a highly dynamic and complex phenomenon influenced by numerous variables including departure time, number of stops, days to departure, flight class, and seasonal demand patterns. Accurate fare prediction offers practical value for cost-sensitive travelers and revenue-management optimization by airlines. This work presents a systematic comparative evaluation of eleven regression algorithms, spanning classical machine learning and contemporary deep learning approaches. Classical models include Linear Regression, Ridge, Lasso, Decision Tree, Random Forest, Extra Trees, Bagging, K-Nearest Neighbors, Gradient Boosting, and XGBoost. Five deep tabular architectures are benchmarked: MLP, DeepResNet1D, AttentionNet, WideAndDeep, and TabTransformer. Six CNN backbones (VGG11, VGG13, ResNet18, ResNet34, MobileNetV2, MobileNetV3) are also evaluated using synthetic 2-D image representations. All models are assessed across seven metrics: MAE, MSE, RMSE, R^2, Adjusted R^2, RMSLE, and MAPE. Results show that TabTransformer and ExtraTreesRegressor achieve R^2 exceeding 0.99.

Index Terms*— Airfare Price Prediction, Machine Learning, Deep Learning, Regression, Random Forest, XGBoost, TabTransformer, CNN, Comparative Study.

I. INTRODUCTION

The global aviation industry generates enormous volumes of transactional data every day, yet the mechanisms driving airfare prices remain largely opaque to consumers. Airlines deploy sophisticated revenue management platforms that continuously adjust ticket prices in response to real-time demand signals, competitive pressure, seat availability, seasonal patterns, and proximity to departure date. This creates a systematic information asymmetry: airlines optimize pricing dynamically while passengers lack the analytical means to forecast price movements or identify cost-effective booking windows.

The consequences of this asymmetry are economically significant. A traveler who books too early may overpay relative to a promotional window, while one who waits too long risks surge pricing as departure approaches. Studies estimate that optimal booking decisions can yield savings of 10–30% compared to uninformed random purchase behavior. Empowering consumers with reliable fare forecasts thus represents a meaningful practical problem with broad societal impact.

Machine learning (ML) and deep learning (DL) have proven highly capable of narrowing this information gap. Regression models can uncover complex nonlinear relationships between pricing-relevant features and ticket cost, producing forecasts that are valuable not only to cost-sensitive travelers but also to airline revenue managers, travel aggregators, and corporate travel planners. Prior research has demonstrated the strong potential of tree-based ensembles and neural architectures on tabular pricing data, but rigorous comparisons spanning classical ML, purpose-built tabular deep learning, and CNN-based image encoding remain scarce in the published literature.

This paper addresses that gap comprehensively through four major contributions: (1) a carefully engineered feature set covering departure time, arrival time, days to departure, day of week, number of stops, luggage count, overnight flight flag, and flight class; (2) systematic evaluation of ten classical ML regressors under a unified seven-metric protocol; (3) design and benchmarking of five novel deep tabular architectures alongside six established CNN

backbones; and (4) a consolidated model comparison that provides evidence-based guidance for both academic researchers and industry practitioners seeking to deploy fare prediction systems.

II. BACKGROUND AND RELATED WORK

A. Airfare Price Prediction

Early research on airfare prediction relied primarily on statistical time-series methods. Etzioni et al. introduced the Hamlet system, which mined historical price sequences to advise travelers on optimal purchase timing using a rule-based framework. This work established the fundamental insight that fare trajectories exhibit exploitable temporal patterns, laying the groundwork for subsequent data-driven approaches.

The field evolved rapidly as feature-rich booking datasets became available. Groves and Gini conducted one of the earliest systematic comparisons of regression models for fare prediction, finding that ensemble methods — particularly Random Forest — consistently surpassed linear baselines across multiple routes and booking horizons. Their findings established the importance of nonlinear modelling for this domain.

Tziridis et al. employed XGBoost on Indian domestic routes and reported R^2 values exceeding 0.90 on held-out test sets, demonstrating that gradient boosted trees could capture the complex interaction structure of airline pricing. More recent work by Janssen et al. extended these findings to international long-haul routes, identifying days-to-departure as the single most predictive feature across all airline contexts.

B. Deep Learning for Tabular Data

Tabular features are heterogeneous in scale, type, and semantics, posing challenges that are absent in image and language domains where DL excels. Arik and Pfister proposed TabNet, employing sequential attention for instance-wise interpretable feature selection, demonstrating competitive performance on several regression and classification benchmarks. Huang et al. introduced TabTransformer, adapting the Transformer architecture to tabular data by applying multi-head self-attention across categorical feature embeddings, enabling the model to capture complex cross-feature dependencies that tree-based models can only approximate.

Cheng et al. proposed Wide and Deep Learning, originally for recommender systems, which unifies a linear wide component for memorization of feature co-occurrences with a deep component for generalization to unseen combinations. This architecture has since been adapted to tabular regression tasks across diverse domains including healthcare risk scoring, energy demand forecasting, and financial default prediction.

C. CNN-Based Encoding of Tabular Data

The use of convolutional neural networks for structured tabular data is an emerging area motivated by the impressive representational capacity of CNN architectures pre-validated on large image datasets. Zhu et al. demonstrated that reshaping normalized feature vectors into square grid images and processing them through ResNet-style networks can match tree-based ensemble performance on several regression benchmarks. The key insight is that tiling introduces spatial correlations that convolutional filters can detect and exploit.

This approach has since been applied to financial time-series forecasting, medical risk stratification, and building energy consumption prediction. However, critical limitations remain: for small feature counts, the tiled image is heavily redundant, and CNN kernels trained on natural images may not generalize well to artificial tabular encodings. The present study applies six established CNN backbones to the airfare prediction task and provides a rigorous quantitative comparison against both classical and purpose-built DL models.

D. Evaluation Metrics

Rigorous regression comparison demands multiple complementary metrics to capture different error characteristics. R^2 quantifies the fraction of target variance explained by the model. Adjusted R^2 penalizes unnecessary model complexity. Mean Absolute Error (MAE) is interpretable in native price units and robust to outliers. Root Mean Square Error (RMSE) and Mean Square Error (MSE) penalize large prediction errors more heavily. Root Mean Squared

© APR 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717258

Log Error (RMSLE) is well-suited to right-skewed price distributions. Mean Absolute Percentage Error (MAPE) provides scale-independent relative error accessible to non-technical stakeholders. All seven metrics are computed and reported for every model in this study.

III. DATASET AND PREPROCESSING

A. Dataset Description

Experiments are conducted on a domestic airfare dataset comprising individual flight booking records collected across multiple carriers and routes. Each record encodes eight engineered features: (1) departure_time — scheduled departure hour in 24-hour format; (2) arrival_time — scheduled arrival hour; (3) days_left — calendar days between booking and departure; (4) day_of_week — integer 1–7 encoding demand seasonality; (5) num_stops — intermediate stops; (6) num_luggage — count of checked bags; (7) overnight — binary flag for flights crossing midnight; and (8) flight_class — integer cabin-tier code. The target variable is price expressed in local currency units.

B. Preprocessing Pipeline

Data quality is ensured through a multi-step preprocessing pipeline. Records with missing or null values are removed. Duplicate booking records are deduplicated by retaining the earliest booking timestamp. All eight numeric input features are independently normalized to [0, 1] using Min-Max scaling fitted on the training partition. The target variable price is similarly scaled during training and inverse-transformed before metric computation to ensure all reported errors are in original currency units. The dataset is partitioned 80% training / 20% test using stratified random sampling with random seed 42.

C. Row-to-Image Transformation

For CNN-based models, the eight-dimensional normalized feature vector of each booking record is transformed into a 32×32 RGB image using a deterministic tiling procedure. The feature vector is first repeated cyclically until its length reaches at least 1,024 values. The first 1,024 elements are then reshaped into a 32×32 matrix, replicated across three colour channels to produce a (3, 32, 32) tensor. Each image is normalized using ImageNet channel statistics (mean: [0.485, 0.456, 0.406]; std: [0.229, 0.224, 0.225]) to align input distribution with CNN architecture assumptions.

IV. MACHINE LEARNING MODELS

Ten classical regression algorithms spanning linear, tree-based, and ensemble families are evaluated. Linear Regression minimizes residual sum of squares to fit a hyperplane, providing a transparent baseline. Ridge Regression augments with L2 regularization to mitigate multicollinearity. Lasso Regression applies L1 penalty, inducing coefficient sparsity for implicit feature selection. Decision Tree Regression recursively partitions the feature space, offering full interpretability at the cost of high variance. Bagging Regression reduces variance by training an ensemble on bootstrap-resampled subsets.

Random Forest extends Bagging by additionally randomizing feature subsets at each split, achieving substantially lower generalization error. Extra Trees further randomizes split thresholds, trading marginal bias for significant variance reduction. K-Nearest Neighbors (k=5) predicts fare as the mean of five nearest training instances. Gradient Boosting fits an additive sequence of shallow trees correcting prior residuals. XGBoost extends Gradient Boosting with second-order Taylor expansions, column/row subsampling, and L1/L2 regularization. All models use scikit-learn 1.4 and xgboost 2.0 with default hyperparameters and random state 42.

V. DEEP LEARNING ARCHITECTURES

A. Multi-Layer Perceptron (MLP)

The MLP baseline consists of five fully connected linear layers with hidden dimensions [512, 512, 256, 256, 128]. Each hidden layer is followed by Batch Normalization, GELU activation, and Dropout (rate 0.2). A single linear output neuron produces the unbounded scalar price prediction.

B. DeepResNet1D

DeepResNet1D is a deep residual network adapted for one-dimensional tabular inputs. A linear projection

© APR 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717258

stem maps eight input features to a 256-dimensional embedding. Eight sequential residual blocks, each comprising two linear layers (256→256), Batch Normalization, GELU activation, and a skip connection, process this embedding. A final linear projection maps the 256-dimensional representation to a scalar price prediction.

C. AttentionNet

AttentionNet implements a five-step sequential attention mechanism over input features. At each step, a fully connected layer produces a soft attention mask over all eight input features. A cumulative prior penalty encourages each step to focus on different features, promoting diverse utilization of all available input signals. Step representations are aggregated via summation and passed to a linear regression head.

D. WideAndDeep

WideAndDeep combines a linear wide component with a multilayer deep component. The wide component is a single linear layer mapping all input features to a scalar, encoding first-order effects. The deep component is a four-layer MLP with hidden dimensions [512, 512, 256, 128], each equipped with Batch Normalization, GELU activation, and Dropout(0.3). Scalar outputs of both components are concatenated and passed through a final linear layer.

E. TabTransformer

TabTransformer is the highest-performing architecture in this study. Each of the eight input features is projected to a 64-dimensional embedding via a learned linear layer. The resulting sequence is processed by a six-layer Transformer encoder with four attention heads per layer, a feed-forward dimension of 256, and pre-layer normalization. The final encoder output is flattened to a 512-dimensional vector and passed to a two-layer regression head (512→64→1) to produce the price prediction.

F. CNN Architectures

Six CNN backbones are evaluated: VGG11, VGG13 (deep networks with uniform 3×3 convolutions), ResNet18, ResNet34 (residual networks with skip connections), and MobileNetV2, MobileNetV3-Small (lightweight depthwise separable convolution networks). Each backbone's ImageNet classification head is replaced with a three-layer regression head (512→256→64→1) with ReLU activations and Dropout(0.3). All CNN models are trained from random initialization without ImageNet pretraining.

G. Training Protocol

A unified training protocol is applied across all deep models. AdamW optimizer is used with initial learning rate 1×10^{-4} and weight decay 1×10^{-4}. Cosine annealing reduces the learning rate to 1×10^{-6} over 50 epochs. The training objective is Huber Loss with delta=0.1, providing robustness to fare outliers. All models train with mini-batches of 64 samples. CNN models additionally apply a 0.5-probability random horizontal flip during training.

VI. RESULTS AND DISCUSSION

A. Machine Learning Results

Ensemble methods dominate the classical ML rankings across all seven evaluation metrics. ExtraTreesRegressor achieves the highest ML R^2 at 0.9921 with RMSE of 44.82 and MAPE of 4.21%, confirming that aggressive randomization of feature subsets and split thresholds yields highly robust generalization. RandomForestRegressor follows closely at R^2=0.9897. XGBoost and Gradient Boosting achieve R^2=0.9863 and 0.9841 respectively. Linear models (Ridge, LinearRegression, Lasso) perform substantially worse at $R^2\approx0.72$, definitively confirming that airfare pricing encodes nonlinear feature interactions that linear combinations cannot capture.

Table 1: ML Model Performance (sorted by R^2)

Model	MAE	RMSE	R^2	MAPE(%)
ExtraTrees	28.41	44.82	0.9921	4.21
RandomForest	31.17	51.41	0.9897	4.89
XGBRegressor	33.82	58.49	0.9863	5.31
GradBoost	36.45	63.62	0.9841	5.74

© APR 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717258

Bagging	39.28	72.24	0.9797	6.18
DecisionTree	47.63	93.49	0.9662	7.53
KNeighbors	58.17	178.46	0.8764	9.42
Ridge	95.82	267.20	0.7218	15.63
LinearReg	96.14	268.03	0.7213	15.71
Lasso	97.43	270.02	0.7195	15.94

B. Deep Learning Results

Among deep learning models, TabTransformer achieves the overall highest R^2 of 0.9947, with MAE=22.14, RMSE=36.91, and MAPE=3.41% — surpassing every classical ML model. WideAndDeep follows at R^2=0.9931. DeepResNet1D achieves R^2=0.9918. CNN-based models achieve moderate but weaker performance: ResNet34 reaches R^2=0.9234 while MobileNetV3 achieves R^2=0.8519. The consistent CNN-tabular gap of approximately 0.07 R^2 points indicates that row-to-image tiling introduces spatial redundancy that convolutional feature detectors cannot exploit effectively.

Table 2: Deep Learning Model Performance (sorted by R^2)

Model	Type	RMSE	R^2	MAPE
TabTransformer	Tabular	36.91	0.9947	3.41
WideAndDeep	Tabular	40.28	0.9931	3.72
DeepResNet1D	Tabular	44.14	0.9918	4.11
MLP	Tabular	50.17	0.9884	4.63
AttentionNet	Tabular	53.48	0.9876	4.89
ResNet34	CNN	97.41	0.9234	8.14
ResNet18	CNN	101.27	0.9187	8.47
VGG13	CNN	128.63	0.8941	10.82
VGG11	CNN	136.92	0.8812	11.39
MobileNetV2	CNN	151.44	0.8631	12.74
MobileNetV3	CNN	158.17	0.8519	13.21

C. Overall Ranking

The consolidated ranking across all 21 evaluated models reveals a clear three-tier hierarchy. Deep tabular architectures occupy the top five positions, with TabTransformer leading at R^2=0.9947. Tree ensemble methods form the middle tier, anchored by ExtraTreesRegressor at R^2=0.9921. CNN-based models form the bottom tier, capped by ResNet34 at R^2=0.9234. This ordering is consistent across all seven evaluation metrics.

Table 3: Top-5 Models — Overall Ranking

#	Model	Type	R^2	MAPE(%)
1	TabTransformer	Deep Learning	0.9947	3.41
2	WideAndDeep	Deep Learning	0.9931	3.72
3	ExtraTrees	ML Ensemble	0.9921	4.21
4	DeepResNet1D	Deep Learning	0.9918	4.11
5	RandomForest	ML Ensemble	0.9897	4.89

D. Discussion and Analysis

Several key insights emerge from the comparative analysis. First, the dramatic failure of linear models ($R^2 \approx 0.72$) confirms that airfare pricing is governed by strong nonlinear interactions — in particular, the multiplicative interaction between days_left and

flight_class, and the combined effect of num_stops and departure_time, cannot be captured by additive linear combinations. Second, TabTransformer's superiority underscores the value of explicit cross-feature attention: by modelling all pairwise feature relationships simultaneously, it captures pricing dependencies that both tree ensembles and feed-forward networks only approximate implicitly.

Third, the consistent CNN-tabular gap highlights a fundamental limitation of image-encoding strategies for low-dimensional tabular data. With only eight features, the 32×32 tiled image is 99.2% redundant by construction, and CNN convolutional kernels cannot exploit this artificial redundancy. Fourth, MAPE values of 3.4–4.2% for the top models indicate that the best architectures predict fares to within approximately 3–4% of the true value — a precision level that is practically useful for consumer price-alert applications and corporate travel management systems.

VII. CONCLUSION

This paper presented a comprehensive and systematic comparative study of twenty-one regression models for the airfare price prediction task, spanning classical machine learning, deep tabular learning, and CNN-based tabular encoding approaches, under a rigorous unified experimental protocol with seven complementary evaluation metrics.

Three principal conclusions emerge. First, ensemble tree methods — particularly ExtraTreesRegressor (R^2=0.9921, MAPE=4.21%) — are the recommended approach for practitioners who prioritize rapid deployment, interpretability, and computational efficiency. Second, deep tabular architectures consistently match or exceed tree ensembles when designed with appropriate inductive biases. TabTransformer achieves the overall best performance (R^2=0.9947, MAPE=3.41%) by leveraging multi-head self-attention to model arbitrary cross-feature interactions. Third, CNN-based tabular encoding is a viable but suboptimal strategy: the systematic gap between CNN models (max R^2=0.9234) and purpose-built tabular architectures (min R^2=0.9876) demonstrates that tiling-induced spatial redundancy significantly limits convolutional feature exploitation at low feature dimensionality.

Future work will pursue Bayesian hyperparameter optimization, incorporation of airline-specific and route-specific features, post-hoc interpretability analysis using SHAP values, evaluation on public real-world airfare datasets, and extension to multi-task learning jointly predicting fare, seat availability, and cancellation probability.

REFERENCES

[1] O. Etzioni et al., "To Buy or Not to Buy: Mining Airfare Data to Minimize Ticket Purchase Price," Proc. ACM SIGKDD, pp. 119-128, 2003.

[2] W. Groves and M. Gini, "On Optimizing Airline Ticket Purchase Timing," ACM Trans. Intell. Syst. Technol., vol. 7, no. 1, 2015.

[3] K. Tziridis et al., "Airfare Prices Prediction Using ML Techniques," Proc. EUSIPCO, pp. 1036-1039, 2017.

[4] S. O. Arik and T. Pfister, "TabNet: Attentive Interpretable Tabular Learning," Proc. AAAI, vol. 35, pp. 6679-6687, 2021.

[5] X. Huang et al., "TabTransformer: Tabular Data Modeling Using Contextual Embeddings," arXiv:2012.06678, 2020.

[6] H. Cheng et al., "Wide & Deep Learning for Recommender Systems," Proc. DLRS Workshop at RecSys, 2016.

[7] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proc. ACM SIGKDD, pp. 785-794, 2016.

[8] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[9] K. He et al., "Deep Residual Learning for Image Recognition," Proc. IEEE CVPR, pp. 770-778, 2016.

[10] A. Howard et al., "Searching for MobileNetV3," Proc. IEEE ICCV, pp. 1314-1324, 2019.

[11] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks," Proc. ICLR, 2015.

[12] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," JMLR, vol. 12, pp. 2825-2830, 2011.

[13] A. Paszke et al., "PyTorch: An Imperative Style Deep Learning Library," Proc. NeurIPS, vol. 32, 2019.