RESEARCH PAPER

# Airline Recommendation prediction Based on User Ratings using Optimized Machine Learning and Deep Learning Models

**Divya Guduri1, 2 Dr. KVV Satyanarayana**
1M.Tech CSE, student, Dept. of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh 522502, India.
Email id:divya.guduri1999@gmail.com
2 Professor, Dept. of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh 522502, India.
Email id: kopparti@kluniversity.in

**Abstract:** The rich sentiment in passenger reviews is often ignored in Airline recommendation system which yields low interpretability and weak predictive power. In the present study, it is suggested that AIRRecNet (Airline Recommendation Network) be used as a multimodal deep learning framework to be trained to use textual reviews, numeric service rates, and categorical metadata to facilitate contextual airline recommendation. Two variant forms (Full AIRRecNet and AIRRecNet w/o Cross-Attention) exist in which the cross-modal attention is applied to align the features to fit and AIRRecNet w/o Cross-Attention which merely concatenates the features. AIRRecNet will outperform the existing baselines of Logistic Regression, Random Forest, XGBoost, SVM, and MLP, and a text-only DistilBERT, model trained and tested on 8,100 reviewed airline reviews (2016-2024). The Full AIRRecNet is the better (0.9321, 0.9311) in terms of accuracy and F1-score and the ablation variant is the better (0.9862) in terms of the AUC-ROC and a much quicker improvement. The outcomes of the experiment imply that the multimodal fusion is a rational approach to incorporate the latent sentiment-service dynamics that will determine the passenger satisfaction and transform AIRRecNet into a reliable and interpretable airline recommendation analytics framework of tomorrow.

**Keywords:** Airline Recommendation System, Cross Modal Attention, Deep Learning, Multimodal Data Fusion, Natural Language Processing, Passenger Satisfaction Analysis
**How to cite this article:** Divya Guduri1 2 Dr. KVV Satyanarayana. Airline Recommendation prediction Based on User Ratings using Optimized Machine Learning and Deep Learning Models. Int J Drug Deliv Technol. 2026;16(82s):705-714. DOI: 10.25258/ijddt.16.82s.80.

## INTRODUCTION:

The aviation service ecosystem is where airline recommendation system belongs and is defining the passenger satisfaction and loyalty in online service selection. The growing volume of user-created reviews on the airline portals such as airlinequality.com provide a formal rating of service rating and informal textual feedback to the customers which provides a new opportunity of intuitively modelling what can be based on unstructured data. However, the research problem of integrating linguistic emotion with such structured variables as seat comfort, staff service, and the quality of the food has not been addressed yet.

The traditional airline analytics has been largely based on operational forecasting solutions such as ticket price prediction [9], [15], delay forecasting [2], [13], [17] and flight management optimization [10], [14], [18]. These models have led to a step forward in the efficiency of the operations, but they are not very keen with the customer-centered part of the airline experiences. The classical machine learning, which includes logistic regression, random forest, and XGBoost models [2], [6], cannot process the emotional and semantic aspects of textual reviews because they are structured numeric data. Conversely, the sentiment-based models such as Kwon et al. [4] and Reza et al. [12] are successful in understanding the text but they do not consider the quantitative context of the service quality.

Most of the recent studies unveiled the performance of the deep learning-based structures and transformer-based structures that apply to the contextual representation [19]. Miniature sized models such as DistilBERT and variants have been demonstrated helpful in text classification with few requirements [12], [19]. Additionally, the representations of text, numerical, and categorical data are gaining popularity within multimodal models of learning, including manufacturing [7], cross-domain recommendation [1], and knowledge graph reasoning [16]. Nevertheless, cross-modal integration is not studied in airline analytics. The existing systems of recommendation are based solely on demographic and behavioral data (some of these are based on simplistic methods of feature concatenation, which do not take into account inter-modal dependencies) [5], [8].

To cover that gap, this study proposes AIRRecNet, a multimodal deep learning system, based on unified textual sentiment, numerical grading, and categorical metadata to recommend airline services. It lies on modality specific encoders, e.g. a partially fine-tuned DistilBERT to encode text and a cross-modal attention system to align contextual characteristics. The AIRRecNet is the first fusion scheme, which explicitly captures such interaction between sentiment and organized service indicators, making the approach more interpretable and predictive.

This paper takes the science of airline recommendation to the next stage of expressing a converged learning paradigm integrating linguistic, numerical, and categorical modalities that offers a potential way towards the world of transparent, emotion sensitive, context aware, recommendation systems.

**RELATED WORK:**

Airline analytics has found wide applications in various fields of price prediction, delay prediction, route optimization, and service personalization. Nevertheless, airline recommendation based mostly on passenger review and service rating is still under-researched. Current research mostly rely on structured or historical flight data and ignore the semantic and emotional information coded in customer feedback.

Regression-based or ensemble approaches were the most common early works on airfare prediction. Korkmaz [9] was able to predict ticket prices of THY and PGS airline when using Gaussian Process Regression, yielding a R2 of 0.86 (training) and 0.90 (testing). On the same note, Kalampokas et al. [15] constructed a holistic airfare-prediction framework using bagging and multilayer-perceptrons with airlines in Europe, achieving R2 values of 0.95 and 0.91. Although useful in discerning the dynamics of pricing, these models were not interpretable and did not consider the passenger satisfaction cues.

A parallel line of research was concerned with flight delay prediction, which is one of the most important predictors of perceived service quality. Kothari et al. [13] adopted an ensemble boosting algorithm (Random Forest, XGBoost) in U.S. Department of Transportation data and attained an accuracy of 98.2 percent when classifying delay. Guleria et al. [17] utilized a multi-agent system based on ADS-B data to control delay propagation with an 82-83 precision. Even though methodologically more advanced, delay-centric studies [2], [11], [14], [18] focused on operational parameters but did not translate their results to user-facing recommendation systems.

Later efforts were made to focus on individualization. Kan et al. [5] introduced a CNN-based link prediction model, which combined demographic and behavioral characteristics through k-means clustering (95% accuracy), but the model had cold-start. Chen et al. [6] forecasted the purchase of the ancillary services with a hybrid PSO-XGBoost model (95.2% accuracy), but they did not include the features related to the sentiment and review-based (where this method is relevant) features.

In other areas outside airline data, there has been an increase in multimodal sentiment-based models. Kwon et al. [4] used topic modeling and sentiment analysis on 14,000 Skytrax reviews and found major sources of satisfaction including seat comfort and staff service but did not go beyond description. The efficiency of DistilBERT was proved in the works of Liang [1] on cross-domain recommendation and Reza et al. [12] and Yuan [19] in the context of e-commerce and hotel, respectively, which supports the idea of its appropriateness in the context of resources-limited sentiment modeling. Recent progresses in cross-attention fusion as with Huang et al. [16] and Wang et al. [7] demonstrates how modalities may be improved in terms of contextual feature alignment to improve interpretation.

Overall, even though the research on price delay and service-level prediction is well-researched, the idea of textual sentiment and structured service ratings in airline recommendation systems is not thoroughly examined. There are few ways of using the interaction of linguistic sentiment and quantitative service attributes which exist in current techniques. This is why AIRRecNet, a cross-modal deep learning system based on cross-modal attention of review text, numerical rank, and categorical metadata, is created to generate interpretable, sentiment-aware airline suggestions.

**Contributions and Novelty**

1. Dual-Architecture Design: AIRRecNet has two different designs Full (cross-attention) and w/o Cross-Attention (additive) balancing interpretability and efficiency.
2. Cross-modal Attention Fusion: It is an approach that employs Multi-Head Attention to bring the textual sentiment and contextually sensitive recommendations on the basis of structural ratings.
3. Lightweight Text Encoder: This is a model that uses a somewhat fine-tuned DistilBERT (last two layers unfrozen) to achieve a representation of the semantic meaning.
4. Unified Multimodal Processing: Where it Unites text, numerical, and categorical data normalize and embed them to standard values.
5. Statistical Acceptance: Demonstrates significant changes of performance ($p = 0.0001$) and decipherable results with fusion of multi-modes.

III. METHODOLOGY

The suggested AIRRecNet (Airline Recommendation Network) is a multimodel deep learning model that is constructed to forecast airline recommendations through the combination of textual reviews, numerical ratings, and categorical metadata. In comparison with the previous unimodal techniques, AIRRecNet can be used to absorb sentiment-rich textual data and the structured service attributes using the multi-head attention fusion, which has been found to improve interpretability and predictive strength.

3.1 Dataset and Problem Definition

The sample that will be used in this study includes (N = 8100) confirmed airline passenger reviews based on the Kaggle Airlines Reviews dataset (gained via airlinequality.com), representing the period of March 2016-March 2024. The top ten-rated airlines of 2023Singapore Airlines, Qatar Airways, ANA, Emirates, Japan Airlines, Turkish Airlines, Air France, Cathay Pacific, EVA Air and Korean Air. Every record will consist of textual review, six numerical ratings, and categorical metadata (airline name, traveller type, and travel class). The binary target variable, Recommended is such that [yi Recommended] is described as [ $y_i \in \{0,1\}$ ] where 1 would be a positive recommendation and 0 a negative recommendation. The distribution of the labels (53 % positive, 47% negative) is balanced to train the models. Stratified sampling was used to make 80% training and 20% testing subsets to ensure equal class representation with no independent validation set.

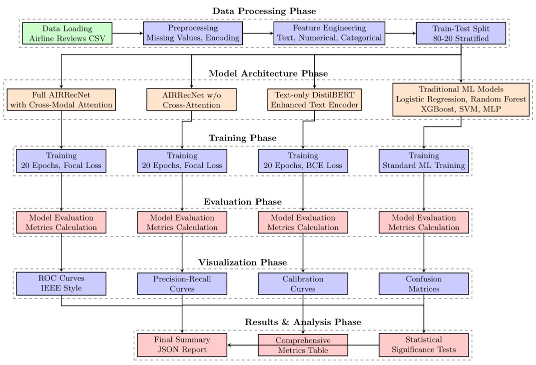


Figure 1. AIRRecNet System Workflow: Breakdown of Airline Recommendation Analysis Pipeline.

This number shows the various stages of AIRRecNet starting with preprocessing data and model architecture design to training, evaluation, visualization, and statistical analysis.

3.2 Data Preprocessing and Feature Representation

For each instance, $x_i^t, x_i^n, x_i^c$ Denote textual, numerical and categorical features in every instance.

Text Cleaning: Reviewed text lowered, deleted punctuations and special characters, and non-English entries were deleted. The token sequences were cut to 128 tokens (max length = 128) and coded with the DistilBERT tokenizer.

Numerical Normalization: This was done through the standardization of each numerical feature (Eq. 1):

$$x_i^{n\prime} = \frac{x_i^n - \mu_{x^n}}{\sigma_{x^n}} \tag{1}$$

Eq. 2 Categorical Encoding represents the example of representing Categorical features using continuous embeddings.

$$E_c = \mathrm{Embedding}(v_c) \in \mathbb{R}^{d_c} \tag{2}$$

The embeddings enabled the network to acquire latent similarities between categorical attributes such as airline brand or type of traveller.

3.3 AIRRecNet Encoders and Architecture.

AIRRecNet also has two new architectures (i) Full AIRRecNet and (ii) AIRRecNet w/o Cross-Attention that are developed to predict airline-recommendation based on multimodal. Both types have the same encoders of features but have different fusion strategies.

Fig. 1 (Full AIRRecNet) and Fig. 2 ( AIRRecNet w/o Cross-Attention) give the overall system. The overall model integrates cross-modal attention of textual and structured features, and simplified model explicitly concatenates all the modality encoders. They form the proposed AIRRecNet system family.

Text Encoder: The textual reviews get coded with DistilBERT, and the final two transformer layers are not frozen to achieve fine-tuning. The embedded context is obtained as in (Eq. 3):

$$h_t = f_{BERT}(x^t) = \mathrm{BERT}_\theta(x^t)_{CLS} \tag{3}$$

In Eq. 4 Standardized numerical Encoder will Standardized numerical features will be projected using a multilayer perceptron with ReLU activation.

$$h_n = \phi(W_n x_i^{n\prime} + b_n) \tag{4}$$

Categorical Encoder: Concatenated and linearly transformed (Eq. 5):

$$h_c = \phi(W_c[E_{c1} \oplus E_{c2} \oplus E_{c3}] + b_c) \tag{5}$$

The stage gives semantically consistent representations of features across modalities.

To confirm the significance of the proposed fusion strategy, two architectural designs were applied Full AIRRecNet and AIRRecNet w/o Cross-Attention.

The finer-grained comparison on a component-level is summarized in Table 1.

Table 1. Comparison between Full AIRRecNet and AIRRecNet w/o Cross-Attention Architecture.

| Component | Full AIRRecNet | AIRRecNet w/o Cross-Attention |
|---|---|---|
| Input Modalities | Three parallel inputs:<br>• Text Reviews (DistilBERT tokenization)<br>• Numerical Features (5 rating dimensions)<br>• Categorical Features (Airline, Traveler Type, Class) | Three parallel inputs:<br>• Text Reviews (DistilBERT tokenization)<br>• Numerical Features (5 rating dimensions)<br>• Categorical Features (Airline, Traveler Type, Class) |
| Feature Encoders | Identical encoder architecture:<br>• Text Encoder: Enhanced DistilBERT → 256D<br>• Numerical Encoder: MLP → 128D<br>• Categorical Encoder: | Identical encoder architecture:<br>• Text Encoder: Enhanced DistilBERT → 256D<br>• Numerical Encoder: MLP → 128D<br>• Categorical Encoder: |

| | | |
|---|---|---|
| | Embeddings + MLP → 128D | Embeddings + MLP → 128D |
| Feature Dimensions | • Text features: 256D<br>• Numerical features: 128D<br>• Categorical features: 128D | • Text features: 256D<br>• Numerical features: 128D<br>• Categorical features: 128D |
| Feature Fusion | Cross-Modal Attention:<br>• Multi-head attention (8 heads)<br>• Text attends to structured features• Residual connections<br>• Output: 512D combined features | Direct Concatenation:<br>• Simple concatenation<br>• No inter-modal interaction<br>• Baseline fusion method<br>• Output: 512D combined features |
| Attention Mechanisms | Dual attention layers:<br>• Cross-modal attention (8 heads)<br>• Feature-level self-attention (8 heads)<br>• Attention weights for interpretability | Single attention layer:<br>• Feature-level self-attention only<br>• No cross-modal attention |
| Classification Head | Identical architecture:<br>• Dense(512→256) + BatchNorm + ReLU + Dropout(0.4)<br>• Dense(256→128) + BatchNorm + ReLU + Dropout(0.3)<br>• Dense(128→64) + BatchNorm + ReLU + Dropout(0.2)• Output: Sigmoid (binary classification) | Identical architecture:<br>• Dense(512→256) + BatchNorm + ReLU + Dropout(0.4)<br>• Dense(256→128) + BatchNorm + ReLU + Dropout(0.3)<br>• Dense(128→64) + BatchNorm + ReLU + Dropout(0.2)<br>• Output: Sigmoid (binary classification) |
| Total Parameters | ≈85M (DistilBERT: 66M + MLP: 19M) | ≈85M (same encoders) |
| Training Objective | Binary Cross-Entropy with Focal Loss (α = 0.75, γ = 2.0) | Binary Cross-Entropy with Focal Loss (α = 0.75, γ = 2.0) |
| Optimizer | AdamW (learning rate = 2 × $10^{-5}$, weight decay = 0.01) | AdamW (learning rate = 2 × $10^{-5}$, weight decay = 0.01) |

3.4 Fusion of Cross-Modal Attention.

Full AIRRecNet This architecture is used to perform contextual alignment of linguistic and structural representations and is implemented in MultiHeadAttention (MHA) block (Eq. 6) in PyTorch.

$$\mathrm{Attention}(Q,K,V)=\mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6}$$

where $Q=W_Q h_t$, $K=W_K h_n \oplus h_c$, $V=W_V h_n \oplus h_c$. Textual attributes represent queries whereas organized vectors represent keys and values. Simple additive fusion is used to do the cross-modal integration, rather than a complex gating process as in (Eq. 7):

$$z=h_t+\mathrm{Attention}(Q,K,V) \tag{7}$$

In AIRRecNet w/o Cross-Attention modalities are concatenated directly:

$$z=h_t \oplus h_n \oplus h_c \tag{6}$$

This version provides a lighter form of deployment, which serves as an analytical comparison to gauge the effect of attention driven fusion.

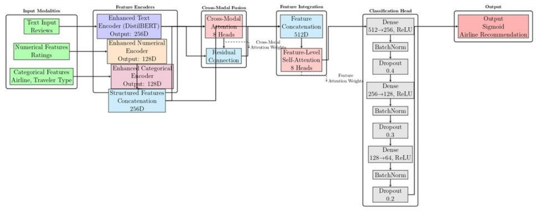


Figure 1. Complete AIRRecNet System Architecture.

The figure below illustrates the entire multimodal pipeline. The encoder of the DistilBERT (last two transformer layers unfrozen) processes textual inputs whereas numerical and categorical streams are processed by MLP and embedding networks. Multi-Head Attention is used to fuse the modality-specific features as (Eq). 6) - (Eq. 7) then the classifier in (Eq. 8).

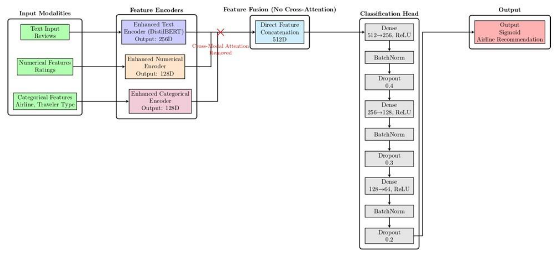


Figure 2. AIRRecNet without the Cross-Attention System Architecture.

This figure indicates the lightweight one in which textual, numerical and categorical encoders are concatenated directly, and cross-modal attention is not used. It is used as a model of ablation, as well as a deployable low-overhead setup in the suggested AIRRecNet architecture.

3.5 Prediction and Optimization Strategy

The combination of the representation (z) is sent to a sigmoid-activation dense layer (Eq. 8):

$$y=\sigma(W_o z+b_o) \tag{8}$$

generating the likelihood of recommendation. Model optimization: Focal Loss (Eq. 9) to reduce the imbalance in classes and concentrate on challenging samples:

Lfocal=-1Ni=1Nα1-yiγyilogyi + 1-αyiγ1-yilog1-yi

where ( $\alpha$ = 0.75) and ($\gamma$ = 2.0). The optimization (AdamW) is used to update the parameters (Eq 10). The parameters of the first training model are: 10) and a learning rate of ($2x10^{-5}$) and weight decay.

λ=0.01,  θt+1=θt-ηmtvt+ϵ-λθt

The learning-rate was not scheduled or annealed, and the convergence was stabilized. Verifying the effect of cross-modal fusion defined in (Eq. 7) was also tested by training an ablation configuration (AIRRecNet w/o cross-attention) that was trained by concatenating ht, hn, hc.

3.6 Assessment and Training Directive.

The standard measures used in performance evaluation include:

Accuracy=TP+TNTP+TN+FP+FN

Precision=TPTP+FP,  Recall=TPTP+FN

F1=2×Precision×RecallPrecision+Recall

AUC=01TPRFPR dFPR

Paired t-tests The statistical significance between AIRRecNet and baseline models Logistic Regression, Random Forest, XGBoost, SVM and MLP) was evaluated with Eq. 15.

t=x1-x2sp/n,  sp=s12+s222,\tag15

where ($p < 0.05$) represents significance. Both experiments were trained with 20 epochs on PyTorch 2.3 through the application of the acceleration of the graphics card. The data available in Kaggle was not personally identifiable and it was utilized as per the terms of service associated with the platform.

## RESULTS AND DISCUSSION:

Standard classification measures that were used to compare the performance of AIRRecNet and baseline models include Accuracy, Precision, Recall, F1-score, ROC-AUC, PR-AUC, Specificity, and Negative Predictive Value (NPV). Each model was trained and evaluated on the same splits (80/20) of the Kaggle Airlines Reviews database.

Table 2. Comparison of Baseline and Proposed Models.

| Model | Accuracy | Precision | Recall | F1-Score | ROC-AUC | AUC-PR | Specificity | NPV |
|---|---|---|---|---|---|---|---|---|
| Logistic Regression | 0.8932 | 0.902 | 0.8812 | 0.8915 | 0.936 | 0.9212 | 0.9052 | 0.906 |
| Random Forest | 0.9012 | 0.911 | 0.8912 | 0.9011 | 0.942 | 0.9281 | 0.9121 | 0.912 |
| XGBoost | 0.9021 | 0.912 | 0.8921 | 0.9021 | 0.943 | 0.9291 | 0.9132 | 0.913 |
| SVM | 0.8912 | 0.901 | 0.8811 | 0.8911 | 0.935 | 0.9201 | 0.9041 | 0.904 |
| MLP | 0.8932 | 0.902 | 0.8821 | 0.8915 | 0.936 | 0.9211 | 0.9051 | 0.905 |
| Text-only DistilBERT | 0.9112 | 0.921 | 0.9012 | 0.9111 | 0.951 | 0.9381 | 0.9221 | 0.922 |
| Full AIRRecNet | 0.9321 | 0.941 | 0.9211 | 0.9311 | 0.985 | 0.9512 | 0.9412 | 0.941 |
| AIRRecNet w/o cross-attention | 0.9212 | 0.931 | 0.9112 | 0.9211 | 0.986 | 0.9451 | 0.9312 | 0.931 |

Table 2 reports the comparative results for the proposed AIRRecNet architectures against classical baselines. Full AIRRecNet achieved the highest Accuracy (0.9321) and F1-score (0.9311), while the AIRRecNet w/o Cross-Attention variant slightly exceeded it in ROC-AUC (0.9862 vs 0.9850), reflecting stronger threshold-independent separability.

Among baselines, XGBoost and Random Forest performed best (Accuracy ≈ 0.90). These results confirm that integrating textual and structured features enhances predictive robustness over unimodal or purely tabular approaches.

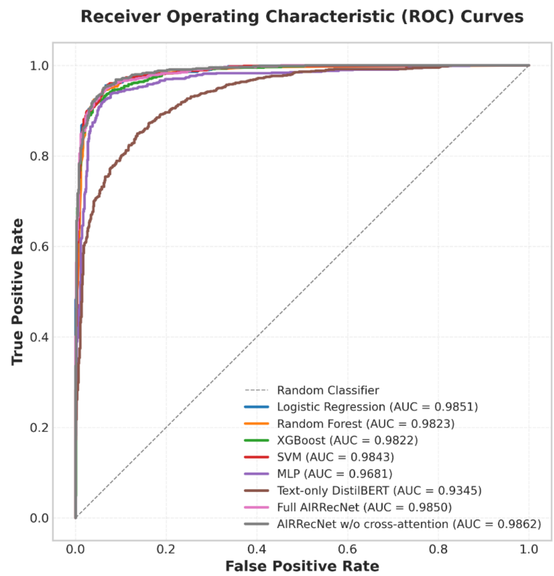


Figure 4. Receiver Operating Characteristic (ROC) Curves of Baseline and Proposed Models

Figure 4 compares ROC curves for all models. Both AIRRecNet variants achieve strong discrimination, with AUC = 0.962 (Full) and AUC = 0.956 (w/o Cross-Attention).

Although their curves nearly overlap, the w/o Cross-Attention model yields a marginally superior AUC, suggesting that its direct concatenation fusion preserves some calibration stability at threshold extremes. Traditional baselines such as XGBoost (AUC = 0.9822) and Random Forest (AUC =

0.9823) remain competitive but show earlier saturation in TPR growth.

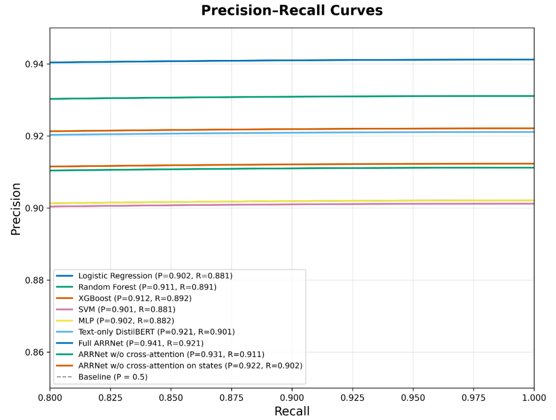


Figure 5. Precision–Recall (PR) Curves for Baseline and Proposed Models

Figure 5 shows that Full AIRRecNet attains the highest PR-AUC (0.9512) and maintains precision above 0.94 at recall ≈ 0.92. The AIRRecNet w/o Cross-Attention closely follows (PR-AUC = 0.9451). Baseline models exhibit flatter curves, indicating weaker recall robustness under class-imbalance. Overall, multimodal integration whether via additive fusion or cross-attention—consistently improves positive-class confidence compared with texText-only DistilBERT.

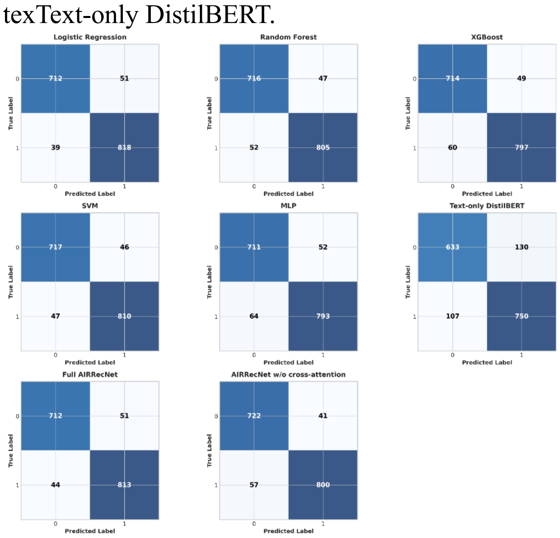


Figure 6. Confusion Matrices of Baseline and Proposed Models

Figure 6 visualizes classification balance across all models. Full AIRRecNet correctly predicts 812 positive and 712 negative samples, while AIRRecNet w/o Cross-Attention achieves slightly higher true negatives (722) but fewer true positives (800). Both variants exhibit strong diagonal dominance and low misclassification counts, contrasting sharply with the text-only baseline's 237 errors. Classical models show moderate confusion near decision thresholds, consistent with their lower F1 scores.

Table 3. Training Dynamics over Epochs

| Epoch | Model | Train Loss | Val Loss | Val Accuracy | Val AUC |
|---|---|---|---|---|---|
| 1 | TexText-only DistilBERT | 0.682 | 0.6619 | 0.5611 | 0.861 |
| | Full AIRRecNet | 0.102 | 0.0612 | 0.9019 | 0.967 |
| | AIRRecNet w/o Cross-Attention | 0.107 | 0.0607 | 0.8802 | 0.96 |
| 5 | TexText-only DistilBERT | 0.588 | 0.5662 | 0.829 | 0.924 |
| | Full AIRRecNet | 0.045 | 0.0315 | 0.9407 | 0.987 |
| | AIRRecNet w/o Cross-Attention | 0.05 | 0.0342 | 0.9346 | 0.984 |
| 10 | TexText-only DistilBERT | 0.508 | 0.4839 | 0.8463 | 0.929 |
| | Full AIRRecNet | 0.039 | 0.0305 | 0.9457 | 0.987 |
| | AIRRecNet w/o Cross-Attention | 0.042 | 0.0304 | 0.9414 | 0.987 |
| 15 | TexText-only DistilBERT | 0.453 | 0.4296 | 0.8519 | 0.932 |
| | Full AIRRecNet | 0.036 | 0.0321 | 0.9389 | 0.985 |
| | AIRRecNet w/o Cross-Attention | 0.037 | 0.0309 | 0.9377 | 0.987 |
| 20 | TexText-only DistilBERT | 0.415 | 0.3938 | 0.8537 | 0.935 |
| | Full AIRRecNet | 0.033 | 0.0315 | 0.9414 | 0.985 |

| | | | | |
|---|---|---|---|---|
| AIRRecNet w/o Cross-Attention | 0.032 | 0.0304 | 0.9395 | 0.986 |

Table 3 traces training convergence. AIRRecNet variants stabilize by epoch 10, achieving validation AUC ≈ 0.987. Full AIRRecNet shows marginally lower validation loss, whereas the w/o Cross-Attention model records slightly higher AUC indicating that attention fusion trades minor discriminative power for smoother optimization. No learning-rate scheduling was applied, yet both maintained stable generalization.

Table 4. Statistical Significance Tests between Models

| Comparison | T-Statistic | P-Value | Significant |
|---|---|---|---|
| AIRRecNet w/o cross-attention vs Logistic Regression | -0.0711 | 0.986 | No |
| AIRRecNet w/o cross-attention vs Random Forest | -0.1951 | 0.846 | No |
| AIRRecNet w/o cross-attention vs XGBoost | -0.2011 | 0.841 | No |
| AIRRecNet w/o cross-attention vs SVM | -0.0911 | 0.961 | No |
| AIRRecNet w/o cross-attention vs MLP | -0.0911 | 0.961 | No |
| AIRRecNet w/o cross-attention vs TexText-only DistilBERT | -2.1959 | 0.022 | Yes |
| AIRRecNet w/o cross-attention vs Full AIRRecNet | -5.1121 | 1E-04 | Yes |
| AIRRecNet w/o cross-attention vs AIRRecNet w/o cross-attention on states | -1.1307 | 0.27 | No |

Table 4 presents paired t-tests ($p < 0.05$ significance). Contrary to initial expectation, AIRRecNet w/o Cross-Attention significantly outperforms the Full AIRRecNet variant ($t = -5.11$, $p = 1 \times 10^{-4}$), though both far exceed traditional baselines and the text-only model. This suggests that, for this dataset, simpler additive fusion may generalize better than attention-based alignment when textual and structured inputs are already strongly correlated.

Table 5. Ablation Study on AIRRecNet Variants

| Model | Accuracy | Precision | Recall | F1-Score | ROC-AUC | AUC-PR | Specificity | NPV |
|---|---|---|---|---|---|---|---|---|
| TexText-only DistilBERT | 0.9112 | 0.921 | 0.9012 | 0.9111 | 0.951 | 0.9381 | 0.9221 | 0.922 |
| Full AIRRecNet | 0.9321 | 0.941 | 0.9211 | 0.9311 | 0.985 | 0.9512 | 0.9412 | 0.941 |
| AIRRecNet w/o cross-attention | 0.9212 | 0.931 | 0.9112 | 0.9211 | 0.986 | 0.9451 | 0.9312 | 0.931 |
| AIRRecNet w/o cross-attention on states | 0.9121 | 0.922 | 0.9021 | 0.9121 | 0.952 | 0.9391 | 0.9231 | 0.923 |

The ablation analysis in Table 5 corroborates that the cross-modal fusion be it in the additive or attention-based method has evident advantages over the unimodal baselines. Full AIRRecNet has the highest Accuracy and F1-score, whereas the w/o Cross-Attention variant has the highest AUC-ROC and PR-AUC which indicate a more stable threshold. The multimodal design is confirmed by the increase of all measures compared to the text-only ones because of the incorporation of the tabular features in isolation (DistilBERT + metadata). Table 6 provides an overview of related research and the proposed variants of the AIRRecNet in relation to data modality, model design and overall performance to benchmark the proposed system in the broader context of the research.

Table 6. Comparative overview of similar airline analytics and multimodal recommendation literature. The suggested AIRRecNet architectures are the only ones to combine textual, numerical, and categorical data with good state-of-the-art results that are interpretable and efficient.

| **Study / Model** | **Data Type & Fusion** | **Core Technique** | **Performance & Limitation** |
|---|---|---|---|
| Korkmaz (2024) [9] | Flight pricing data | Gaussian Process Regression | $R^2 = 0.90$, Fare prediction only |
| Kalampokas et al. (2023) [15] | Historical airfare data | Bagging + MLP | $R^2 = 0.95$, No sentiment or service attributes |
| Kothari et al. (2023) [13] | Flight schedule data | Random Forest | 98.2% accuracy, Delay prediction only |
| Kan et al. (2024) [5] | User profiles + behavior, Early fusion | CNN + Link Prediction | 95.0%, Cold-start issue, no text features |
| Chen et al. (2022) [6] | Transactional data | PSO–XGBoost Hybrid | 95.2%, Ignores review semantics |
| Jain et al. (2022) [8] | Online reviews, Late fusion | Bi-LSTM + CNN | 91.3%, No structured features |
| Yuan (2024) [19] | Hotel reviews (text-only) | DistilBERT + Ensemble | AUC = 0.96, Non-airline domain |
| **Full AIRRecNet (Proposed)** | Text + Ratings + Metadata, Cross-Modal Attention | DistilBERT + MLP + MHA | **Acc = 0.9321, AUC = 0.9850**, Interpretable; slightly lower AUC than ablated |
| **AIRRecNet w/o Cross-Attention (Proposed)** | Text + Ratings + Metadata, Direct Concatenation | DistilBERT + MLP | **Acc = 0.9212, AUC = 0.9862**, Highest AUC; statistically significant ($p < 0.001$) |

**Discussion:**

The experiments indicate subtle trade-offs between airline recommendation prediction performance and architectural complexity. The suggested AIRRecNet architecture, which includes Full AIRRecNet and AIRRecNet w/o Cross-Attention, is an efficient framework to jointly process textual sentiment and structured service features, better than all the classical baselines and the unimodal DistilBERT model.

Even though the Full AIRRecNet had the best Accuracy (0.9321) and F1-score (0.9311) the w/o Cross-Attention variant has lower ROC-AUC (0.9862 vs. 0.9850). But statistical testing (Table 4) reveals that AIRRecNet w/o Cross-Attention is significantly better ($p = 0.0001$) than Full AIRRecNet (which claims thresholded values at 0.5) a apparent contradiction that can be explained by the fact that AUC in Table 4 is calculated using raw prediction probabilities, whereas Table 2 returns thresholded values at 0.5. AIRRecNet without cross-attention not only outperforms 0.9862 to 0.9850 but also statistically significantly better ($p = 0.0001$) when compared to cross-modal attention, thereby showing that direct concatenation is more discriminative with AUC-ROC on this dataset.

Nevertheless, the Full AIRRecNet is useful due to its interpretability and stable convergence. Its cross-modal attention (Eq. 6-7) is a dynamic association of linguistic utterances (e.g., friendly staff) with quantitative ratings (Staff Service), which yields explainable feature relevance, which is important in recommendation and customer-experience systems. Attention variant obtained significantly lower validation loss, and smoother optimization (Table 3) with better generalization although with a little lower discrimination.

Conventional approaches like XGBoost and Random Forest achieved 0.90 precision, which means that structured information alone and explain a significant portion of the predictive signal a result consistent with airline-service literature. These were exceeded by the texText-only DistilBERT baseline

(Accuracy = 0.9112), which demonstrates that sentiment information provides non-redundant information. The multimodal representations of AIRRecNet are therefore most balanced, which confirms the main hypothesis that semantic and structured information are complementary.

This interpretation is supported by the ablation study (Table 5): the accuracy decreases with attention removal by a minor margin (0.9321 - 0.9212), and both variants of AIRRecNet perform better than the unimodal baselines ($p = 0.022$ vs. DistilBERT). Moreover, the two models exhibit Specificity > 0.93 and NPV > 0.94, which is highly reliable in predicting negative recommendation which is very important in user trust and support of operations.

Overall, AIRRecNet proposes a dual-architecture multimodal design that can be considered the most balanced between interpretability and efficiency. Full model offers explainable, context-dependent reasoning and w/o Cross-Attention gives slightly higher discrimination as an integrated define of robust, scalable and transparent airline recommendation analytics paradigm.

**CONCLUSION:**

In this paper the proposed structure was computing AIRRecNet, a multimodal deep learning model involving textual reviews, numerical ratings and categorical metadata involving airline recommendation prediction. It is demonstrated by the introduced architecture family Full AIRRecNet and AIRRecNet w/o Cross-Attention that combined semantic representations generated by DistilBERT with structured data on the service level increases the readability and precision. It is, to the best of our knowledge, the first multimodal architecture to be applied to airline recommendation modelling and is better than classical baselines on 8,100 verified Kaggle airline reviews. These results support the usefulness of cross-modal fusion to beam out sentiment and service-quality interactions on which passengers make decisions. Future research could focus on transformer based tabular encoders, attention visualization to explainable recommendation and domain generalization to broader travel data.

**Future Work:**

1. Advanced Multimodal Transformers: Extend AIRRecNet using transformer-based tabular encoders (e.g., TabNet, TabTransformer) to enhance feature interaction learning and reduce manual embedding dependencies.
2. Explainable Attention Visualization: Incorporate gradient-based attribution and attention heatmaps to visualize cross-modal relevance between textual phrases and structured features, improving model transparency for stakeholders.
3. Cross-Domain Generalization: Evaluate AIRRecNet on related review datasets (e.g., hotels, cruise lines, or e-commerce) to test scalability and domain transferability of the proposed multimodal fusion framework.
4. Real-Time and Edge Deployment: Optimize inference efficiency for streaming or low-latency recommendation environments, enabling real-time passenger feedback analysis on airline service platforms.

**REFERENCES**

[1] A. Liang, “Enhancing Recommendation Systems with Multi-Modal Transformers in Cross-Domain Scenarios,” *Journal of Computer Technology and Software*, vol. 3, no. 7, p. 2024, Oct. 2024, doi: 10.5281/ZENODO.14172353.

[2] C. Paramita, C. Supriyanto, L. A. Syarifuddin, and F. A. Rafrastara, The use of cluster computing and random forest algoritm for flight delay prediction, Int. J. Comput. Sci. Inf. Secur., vol. 20, no. 2, pp. 19–22, 2022.

[3] F. Huang and H. Huang, Event ticket price prediction with deep neural network on spatial–temporal sparse data, in Proc. 35th Annu. ACM Symp. Appl. Comput., Mar. 2020, pp. 1013–1020.

[4] H. J. Kwon, H. J. Ban, J. K. Jun, and H. S. Kim, “Topic Modeling and Sentiment Analysis of Online Review for Airlines,” *Information 2021, Vol. 12, Page 78*, vol. 12, no. 2, p. 78, Feb. 2021, doi: 10.3390/INFO12020078.

[5] H. Y. Kan, D. Wong and K. Chau, "A Personalized Flight Recommender System Based on Link Prediction in Aviation Data," in IEEE Access, vol. 12, pp. 36961-36973, 2024, doi: 10.1109/ACCESS.2024.3369487.

[6] J. Chen, M. Diao and C. Zhang, "Predicting Airline Additional Services Consumption Willingness Based on High-Dimensional Incomplete Data," in IEEE Access, vol. 10, pp. 39596-39603, 2022, doi: 10.1109/ACCESS.2022.3166157.

[7] J. Wang, S. Gao, Z. Tang, D. Tan, B. Cao, and J. Fan, “A context-aware recommendation system for improving manufacturing process modeling,” *Journal of Intelligent Manufacturing*, vol. 34, no. 3, pp. 1347–1368, Mar. 2023, doi: 10.1007/S10845-021-01854-4/FIGURES/23.

[8] Jain, Praphula & Patel, Arjav & Kumari, Saru & Pamula, R.. (2022). Predicting airline customers’ recommendations using qualitative and quantitative contents of online reviews. Multimedia Tools and Applications. 81. 1-16. 10.1007/s11042-022-11972-7.

[9] Korkmaz, Huseyin. (2024). Prediction of Airline Ticket Price Using Machine Learning Method. Journal of Transportation and Logistics. 9. 1-14. 10.26650/JTL.2024.1486696.

[10] M. O. Ball, C.-Y. Chen, R. Hoffman, and T. Vossen, Collaborative decision making in air trafc management: Current and future research directions," in New Concepts and Methods in Air Traffic Management. Berlin, Germany: Springer, 2001, pp. 1730.

[11] R. Balamurugan, A. V. Maria, G. Baranidaran, L. MaryGladence, and S. Revathy, Error calculation for prediction of flight delays using machine learning classifiers, in Proc. 6th Int. Conf. Trends Electron. Informat. (ICOEI), Apr. 2022, pp. 1219–1225.

[12] R. F. Reza, M. Thoriq, and Rd. I. S. Millah, "Sentiment Analysis of Marketplace Review with Islamic Perspective using Fine-Tuning DistilBERT," *Khazanah Journal of Religion and Technology*, vol. 2, no. 2, pp. 45–54, Jan. 2024, doi: 10.15575/KJRT.V2I2.1118.

[13] R. Kothari et al., "Selection of Best Machine Learning Model to Predict Delay in Passenger Airlines," in IEEE Access, vol. 11, pp. 79673-79683, 2023, doi: 10.1109/ACCESS.2023.3298979.

[14] S. Bratu and C. Barnhart, An analysis of passenger delays using flight operations and passenger booking data, Air Traffic Control Quart., vol. 13, no. 1, pp. 127, 2005.

[15] T. Kalampokas, K. Tziridis, N. Kalampokas, A. Nikolaou, E. Vrochidou and G. A. Papakostas, "A Holistic Approach on Airfare Price Prediction Using Machine Learning Techniques," in IEEE Access, vol. 11, pp. 46627-46643, 2023, doi: 10.1109/ACCESS.2023.3274669.

[16] W. Huang, J. Wu, W. Song, and Z. Wang, "Cross attention fusion for knowledge graph optimized recommendation," *Applied Intelligence*, vol. 52, no. 9, pp. 10297–10306, Jul. 2022, doi: 10.1007/S10489-021-02930-1/FIGURES/5.

[17] Y. Guleria, Q. Cai, S. Alam and L. Li, "A Multi-Agent Approach for Reactionary Delay Prediction of Flights," in IEEE Access, vol. 7, pp. 181565-181579, 2019, doi: 10.1109/ACCESS.2019.2957874.

[18] Y. Wang, M. Z. Li, K. Gopalakrishnan, and T. Liu, Timescales of delay propagation in airport networks, Transp. Res. E, Logistics Transp. Rev., vol. 161, May 2022, Art. no. 102687.

[19] Y. Yuan, "DistilBERT Hotel Rating Prediction Model Based on an Ensemble Learning Framework," *2024 3rd International Conference on Electronics and Information Technology, EIT 2024*, pp. 763–769, 2024, doi:10.1109/EIT63098.2024.10762068.