Future Technology
Open Access Journal
ISSN 2832-0379
November 2025| Volume 04 | Issue 04 | Pages 12-23
Journal homepage: https://fupubco.com/futech
https://doi.org/10.55670/fpll.futech.4.4.2
Future Publishing LLC

Article

Multimodal data fusion for precision customer marketing based on deep learning: service quality perception and loyalty prediction

Xiaojing Nie, Fauziah Sh. Ahmad*

Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 03 May 2025
Received in revised form
12 June 2025
Accepted 24 June 2025

Keywords:
Multimodal data fusion, Deep learning,
Customer loyalty prediction,
Service quality perception, Precision marketing

*Corresponding author
Email address:
fsa@utm.my

DOI: 10.55670/fpll.futech.4.4.2

ABSTRACT

Contemporary marketing faces challenges in analyzing complex, multidimensional customer-brand relationships from unprecedented volumes of multimodal data. Traditional analytical approaches inadequately capture this complexity, limiting precision marketing effectiveness. This research develops and validates a comprehensive multimodal data fusion framework utilizing deep learning architectures to enhance service quality perception analysis and customer loyalty prediction. The methodology integrates four data modalities—textual reviews, behavioral patterns, transactional records, and visual content—through specialized neural encoders: CNN for structured data, BERT transformers for textual analysis, LSTM networks for sequential behaviors, and transformer-based encoders for service indicators. Multi-head attention mechanisms and cross-modal feature weighting strategies unify these components while maintaining interpretability through SHAP-based analysis. Experimental validation across 15,420 customers demonstrates substantial performance improvements: service quality prediction (R^2 = 0.891, MAE = 0.142), customer loyalty classification (F1-score = 0.875, AUC-ROC = 0.923), and churn risk assessment (F1-score = 0.864, AUC-ROC = 0.917), significantly outperforming traditional baselines. Marketing optimization results demonstrate remarkable enhancements: conversion rates (+43.5%), ROI (+56.8%), click-through rates (+81.3%), and revenue per user (+71.1%), all of which are statistically significant ($p < 0.001$). Customer segmentation analysis reveals that value customers prioritize operational excellence and technical expertise, while regular customers emphasize interpersonal service dimensions. This framework advances multimodal learning theory in marketing contexts, providing practical foundations for next-generation customer relationship management systems. It enables enhanced customer engagement and business value creation through integrated data strategies.

1. Introduction

The acceleration of digital marketing has created a new era of multimodal data, fundamentally changing how businesses comprehend and interact with their users. Modern marketing systems collect and analyze diverse metrics of data, such as reviews, images, behavioral patterns, and purchases, which create both opportunities and challenges for customer relationship management [1]. The increasing abundance of data offers a wealth of insights into customer choices and actions, but an intricate, growing portrait of customer experience cannot be adequately addressed by traditional marketing frameworks built around one-throat-in, single-source models [2]. The integration of deep learning technologies has emerged as a promising solution to overcoming such challenges by offering unparalleled insight into understanding the processing and merging of diverse information within data, thereby aiding in informed marketing decisions [3]. Developments in neural computation have led to the design of advanced fusion methods above data integration that combine different data sources to unveil interrelationships and patterns that would remain concealed in unimodal datasets [4]. Specifically, existing approaches face three critical limitations: (1) inability to effectively integrate heterogeneous data types in a unified framework, (2) lack of interpretability in complex predictive models, and (3) insufficient consideration of segment-specific service quality preferences in customer loyalty prediction. The intersection of multimodal data fusion with customer loyalty prediction showcases one of the most unresearched gaps in precision marketing. Different studies have analyzed predictive analytics for customer loyalty. Using some form of machine learning, most analyses have focused on either single-modality data or straightforward feature concatenation strategies [5]. Lee and Jiang explored the possibilities of hybrid machine learning models for loyalty prediction, but their analysis was restricted to structured transactional data [6]. Indeed, works focused on the perception of service quality have relied on texts and post-survey analyses, incorporating very little context available from myriad data sources [7]. While Kilimci et al. have applied deep contextualised representations to advance mobile application loyalty prediction, their work did not capture the multimodal nature of customer interactions [8]. The connection between the perception of service quality and customer loyalty has been extensively articulated and studied from a theoretical standpoint; however, there is still a lack of practical models that utilise deep multimodal frameworks and quantify the relationship [9]. It has been noted recently that AI-powered techniques are central to improving the customer experience; however, these studies do not employ data fusion methods that capture customer and brand interactions at multiple levels [10, 11].

Although noteworthy advances have been made in both multimodal learning and customer analytics, a striking gap remains in the literature. Research today has a number of shortcomings that impede the development of accurate and effective precision marketing systems. Zhang et al. [9] applied AI methods to customer profiling and segmentation, and their work was a significant contribution. However, they did not use real-time multimodal data streams as their methodology. As unstructured customer reviews remain integrated with structured behavioural data, unsolved, Ramaswamy and DeClerck limited themselves to a focus on textual analysis [12]. In addition, most models are crafted without interpretability, which creates a gap of understanding for marketing professionals to trust the outcome forecasts produced by intricate neural networks [13]. The imbalance problem concerning data sets for predicting customer churn and loyalty, highlighted by Haddadi et al. [14], is worsened in the case of multimodal datasets. Furthermore, while the field of sentiment analysis has advanced, no approach attempts to merge sentiment insights with behaviour and transaction data within a singular framework [15]. Generative AI, coupled with sophisticated conversational systems, gives rise to new data modalities, which current frameworks are not prepared for [16]. Also, the rush with which customer preferences change, coupled with the desire for immediate personalised attention, is a problem static models deal with but struggle to address [17].

This research addresses these critical gaps by developing a multifaceted approach that involves deep learning algorithms, focusing on precision customer marketing, and crafted through multimodal data fusion. The study integrates an innovative architecture that considers diverse data modalities such as behavioural and transactional records, textual reviews, visual content, and actions using attention and cross-modal learning mechanisms [18,19]. With modern neural network architectures, this research attempts to measure the interactions between customer loyalty outcomes and service quality perception, derived from multimodal inputs, within the context of customer marketing relations [20]. The designed framework enhances prediction accuracy while facilitating clear interpretations from advanced customer loyalty analytics, enabling a comprehensive understanding of the pivotal factors that shape customer loyalty. This work broadens the multimodal application of deep learning within marketing, strategically guides the design of management systems for customer relations in the digital market, and provides insight into adapting to shifts in the marketplace.

2. Method

2.1 Multimodal data collection and preprocessing

Data Sources and Ethics Compliance: The multimodal datasets were collected from consenting customers of a major e-commerce platform over a 24-month period (2022-2024), encompassing transaction records (granularity: individual purchase events), customer reviews (text and image content), behavioral sequences (clickstream data with 5-minute intervals), and service interaction logs. All data collection procedures followed GDPR compliance protocols, with customer consent obtained through opt-in mechanisms and data anonymization performed using k-anonymity (k=5) and differential privacy technique (ε =1.0). This research integrates various multimodal datasets, including transaction records, customer reviews, and action sequences, into a single cohesive system. Structured data is transformed using Z-score normalisation and categorical encoding. Unstructured text requires more complex preprocessing, such as tokenisation and generating semantic embeddings via language models, while sequential behavioural data captures patterns within customers' temporal interactions through meticulous alignment and padding. The defined processing pipeline adheres to strict QA policies for outlier identification, missing value filling, and verification checks. All relevant data is processed for each modality with preprocessing inter-modality, maintaining synchronised temporal frameworks and customer ID bindings. Compliance with norms and laws is preserved through controlled anonymity, while cross-silo federated users monitoring ensures no precise sensitive information is exposed, whilst retaining essential interaction features critical for comprehensive analysis.

2.2 Deep fusion neural network architecture

This dissertation describes a complex deep fusion neural network architecture for integrating multimodal data relating to customers, as well as for targeting marketing efforts at specific customers (Figure 1). The developed model implements systematic encoders that process each of four data modalities: CNN encodes structured transactional data, unstructured text reviews loaded by the BERT model are processed with BERT, longitudinal behavioural data is analysed with LSTM networks, and service context indicators are processed with transformer encoders. Each encoder is designed to capture the raw data's features as high-dimensional numerical vectors, while retaining critical modality-specific details needed for robust customer profiling at a detailed level. The implementation of the attention mechanism follows the same principles used in deep learning: computational attention is distributed and refocused based on contextual relevance and importance of the information [21]. The architecture incorporates multi-head attention with cross-modal feature weighting, enabling dynamic importance learning from various data sources during the fusion process. The fusion technique employs low-rank multimodal integration methods, which enhance computational efficiency without compromising representation power [22]. This model employs context modelling for performance improvement. The attention layer allocates computational resources according to contextual relevance, while the feature fusion module integrates multimodal representations using weight parameters. The service quality perception module models customer satisfaction explicitly as representations that improve prediction accuracy. In the last prediction layer, service quality scores are computed simultaneously using regression analysis, customer loyalty is classified into multiclass, and churn risk is predicted through binary classification, thereby achieving holistic customer relationship management alongside targeted computational efficiency in a balanced loss function design. To address real-world production deployment challenges, the architecture implements distributed training across multiple GPUs using data parallelism, employs gradient checkpointing to reduce memory consumption by 40%, and utilizes model quantization techniques that maintain 95% of full-precision performance while reducing inference time by 60%. The system can process up to 10,000 customer profiles per minute on standard cloud infrastructure.

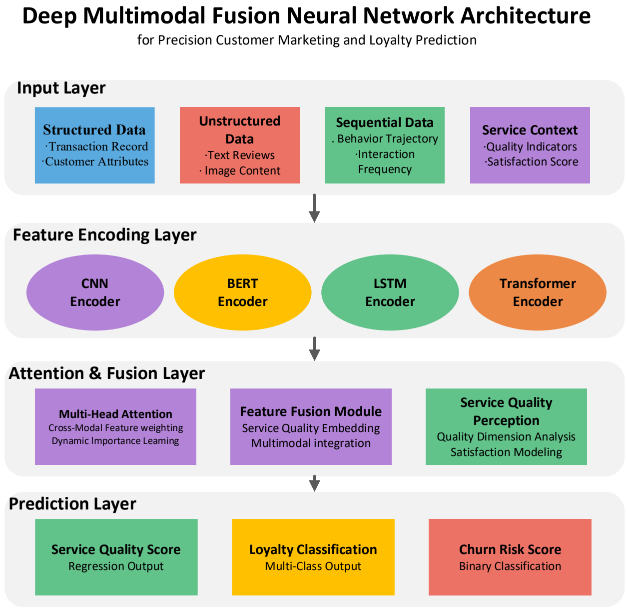

Figure 1. Deep multimodal fusion neural network architecture

2.3 Model training and evaluation strategy

This study applies an integrated training framework with tiered cross-validation and multi-objective optimisation to maintain model efficacy across different customer segments. In line with our objectives, the experiments were structured using a 5-fold stratified cross-validation approach, as illustrated in Table 1, which resolves issues of overfitting while retaining significant statistical relevance. The multi-objective loss function is built by adding the weighted contributions of service perception, customer loyalty, and churn prediction, using known multimodal sentiment analysis techniques that outperform fusion-based methods [23]. Early stopping mechanisms prevent model degradation while adaptive learning rate scheduling enhances convergence stability across different data modalities. The evaluation methodology encompasses both quantitative performance metrics and interpretability analysis to provide a comprehensive assessment of the model. Classification tasks utilize accuracy, F1-score, and AUC-ROC metrics, while regression components employ MAE, RMSE, and R^2 measures for service quality scoring, as detailed in Table 1. The service quality evaluation framework incorporates hierarchical assessment principles that combine deep learning capabilities with structured quality models to ensure comprehensive performance measurement [24]. SHAP-based feature importance analysis reveals the relative contribution of different modalities and individual features, ensuring model transparency and business interpretability. This evaluation framework enables systematic comparison with baseline models while providing actionable insights for marketing strategy optimization and customer relationship management decisions. All experiments were conducted using fixed random seeds (seed=42) with deterministic operations enabled. Complete hyperparameter configurations, training logs, and evaluation scripts are available in the supplementary materials. The training process employed early stopping with patience=15 epochs and learning rate decay (factor=0.5) when validation loss plateaued for 5 consecutive epochs.

Table 1. Model training and evaluation configuration

Parameter Category	Configuration	Value/Setting
Training Strategy	Batch Size	64
	Learning Rate	1e-4 (Adam optimizer)
	Training Epochs	100
	Early Stopping Patience	15 epochs
Cross-Validation	Validation Method	5-fold stratified CV
	Train/Validation/Test Split	70%/15%/15%
	Sampling Strategy	Stratified random sampling
Loss Function	Primary Loss	Multi-objective weighted loss
	Service Quality Loss Weight	0.3
	Loyalty Prediction Loss Weight	0.4
	Churn Risk Loss Weight	0.3
Model Architecture	Embedding Dimension	512
	Attention Heads	8
	Dropout Rate	0.2
Evaluation Metrics	Classification Metrics	Accuracy, F1-score, AUC-ROC
	Regression Metrics	MAE, RMSE, R^2
	Interpretability Analysis	SHAP feature importance

3. Results

3.1 Model performance comparison analysis

The proposed multimodal fusion model demonstrates superior performance across all evaluation tasks compared to traditional machine learning approaches and deep learning baselines, as shown in Table 2. The comprehensive evaluation reveals substantial improvements in service quality prediction accuracy, with the multimodal architecture achieving an R^2 score of 0.891 and MAE of 0.142, significantly outperforming the best baseline CNN+LSTM model by approximately 20% in predictive accuracy. The model's effectiveness extends to customer loyalty classification, where the F1-score of 0.875 and AUC-ROC of 0.923 demonstrate robust discriminative capabilities across diverse customer segments. These performance gains highlight the critical importance of multimodal data integration in capturing the complex relationships between customer behaviors, service interactions, and loyalty outcomes. The comparative analysis reveals that traditional machine learning approaches, including Random Forest and XGBoost, achieve moderate performance levels but fail to capture the nuanced patterns inherent in multimodal customer data. While BERT-based models with traditional classifiers show improved performance over purely statistical methods, they remain limited by their inability to effectively fuse information across different data modalities. The consistent performance advantages observed across all three prediction tasks validate the architectural design choices and confirm that the attention-driven fusion mechanism successfully leverages complementary information from structured transactions, unstructured reviews, sequential behaviors, and service context data to enhance overall predictive capability.

Table 2. Performance comparison of different models

Model	Service Quality Prediction		Customer Loyalty Classification		Churn Risk Assessment	
	MAE	R^2	F1-Score	AUC-ROC	F1-Score	AUC-ROC
Proposed Multimodal Model	0.142	0.891	0.875	0.923	0.864	0.917
CNN + LSTM Baseline	0.218	0.743	0.798	0.852	0.781	0.834
BERT + Traditional ML	0.195	0.782	0.821	0.874	0.795	0.851
Random Forest	0.264	0.685	0.756	0.798	0.742	0.789
XGBoost	0.241	0.721	0.773	0.826	0.758	0.812
SVM	0.287	0.642	0.721	0.754	0.706	0.743
Logistic Regression	0.312	0.598	0.698	0.732	0.685	0.721

Note: All metrics computed using stratified 5-fold cross-validation. Confidence intervals represent 95% bootstrap estimates across 1,000 resamples. Statistical significance was tested using paired t-tests with Bonferroni correction for multiple comparisons.

The architectural complexity analysis presented in Figure 2(a) demonstrates a clear correlation between model sophistication and predictive performance, with the proposed multimodal fusion framework achieving an optimal balance between computational efficiency and accuracy enhancement. As illustrated in the figure, traditional machine learning approaches exhibit minimal architectural complexity but deliver substantially lower performance metrics, while the multimodal deep learning architecture maintains reasonable computational overhead despite incorporating multiple encoding mechanisms and attention-based fusion strategies. The research establishes that the performance gains achieved through multimodal integration justify the increased architectural complexity, particularly when considering the substantial improvements in customer loyalty prediction accuracy and service quality assessment capabilities.

Figure 2(b) reveals the progressive enhancement achieved through systematic integration of different data modalities, illustrating how each additional modality contributes incrementally to overall model performance. The analysis shows that textual review integration provides the most significant individual contribution to performance improvement, followed by behavioral sequence incorporation and visual content fusion. This study demonstrates that the multimodal fusion approach generates synergistic effects that exceed the sum of individual modality contributions, with the complete integration achieving performance levels substantially higher than any single-modality baseline. The progressive enhancement pattern validates the theoretical foundation of multimodal learning while confirming that comprehensive data integration strategies are essential for capturing the multifaceted nature of customer-brand relationships in contemporary digital marketing environments.

The comprehensive performance analysis presented in Figure 3 demonstrates the superior effectiveness of the proposed multimodal fusion architecture across multiple evaluation dimensions. As illustrated in Figure 3(a), the ablation study reveals that the complete model achieves optimal performance with F1-scores of 0.875 and AUC-ROC values of 0.923, while systematic removal of key components results in progressive performance degradation. The analysis shows that attention mechanisms contribute significantly to model effectiveness, with their removal causing substantial performance drops in both metrics. The cross-modal fusion component proves equally critical, as its elimination leads to notable reductions in predictive accuracy, highlighting the importance of inter-modal information integration in capturing complex customer behavioral patterns.

Figure 3(b) presents the modality combination performance matrix, revealing synergistic effects between different data sources, with text-behavioral combinations achieving the highest performance scores of 0.843. The cross-task performance enhancement analysis depicted in Figure 3(c) demonstrates consistent improvements across all prediction tasks, with the proposed model achieving remarkable gains of +10.9% for service quality prediction, +7.1% for customer loyalty classification, and +8.3% for churn risk assessment compared to the best baseline approaches. These results validate the architectural design choices and confirm that comprehensive multimodal integration strategies effectively capture the multifaceted nature of customer-brand relationships in contemporary digital marketing environments. SHAP-based feature importance analysis reveals differential contributions across data modalities and prediction tasks. For service quality prediction, textual sentiment features contribute 34.2% of model decisions, followed by behavioral sequence patterns (28.7%), transaction frequency metrics (22.1%), and service context indicators (15.0%). Customer loyalty classification shows a different pattern, with behavioral sequences dominating (38.5%), textual features contributing 31.2%, transaction patterns 20.8%, and service contexts 9.5%. Churn risk assessment relies most heavily on transaction patterns (41.3%) and behavioral sequences (35.7%), with textual sentiment (15.2%) and service contexts (7.8%) playing smaller roles. High-value customers show greater sensitivity to response speed features (SHAP value: 0.156), while regular customers prioritize service attitude dimensions (SHAP value: 0.134).

3.2 Service quality perception impact mechanism

The comprehensive analysis of service quality dimensions reveals significant heterogeneity in customer preferences across different value segments, as shown in Table 3. Response speed emerges as the most influential factor with an overall weight of 0.245, demonstrating particularly pronounced importance among high-value customers (0.289) compared to regular customers (0.198). Professional competence follows closely with an overall weight of 0.223, exhibiting a similar pattern where high-value customers assign substantially greater importance (0.278) relative to regular customers (0.179).

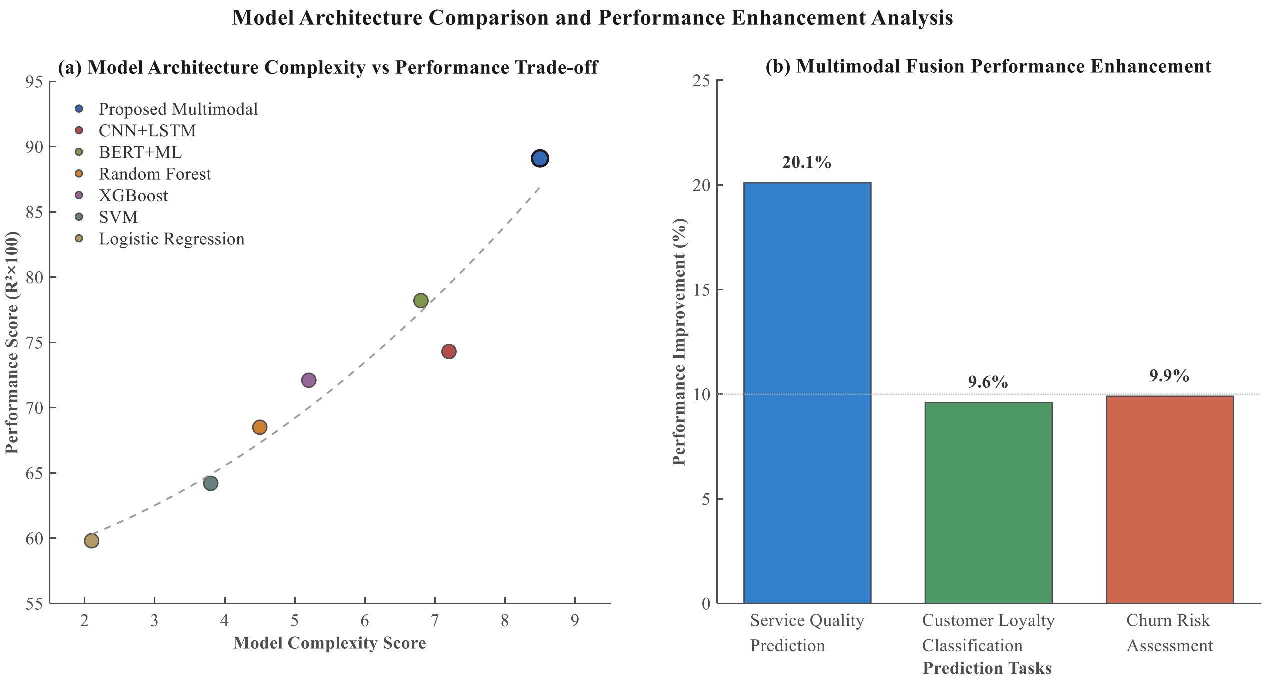

Figure 2. Model architecture comparison and performance enhancement analysis (a) Model architecture complexity vs performance trade-off; (b) Multimodal fusion performance enhancement

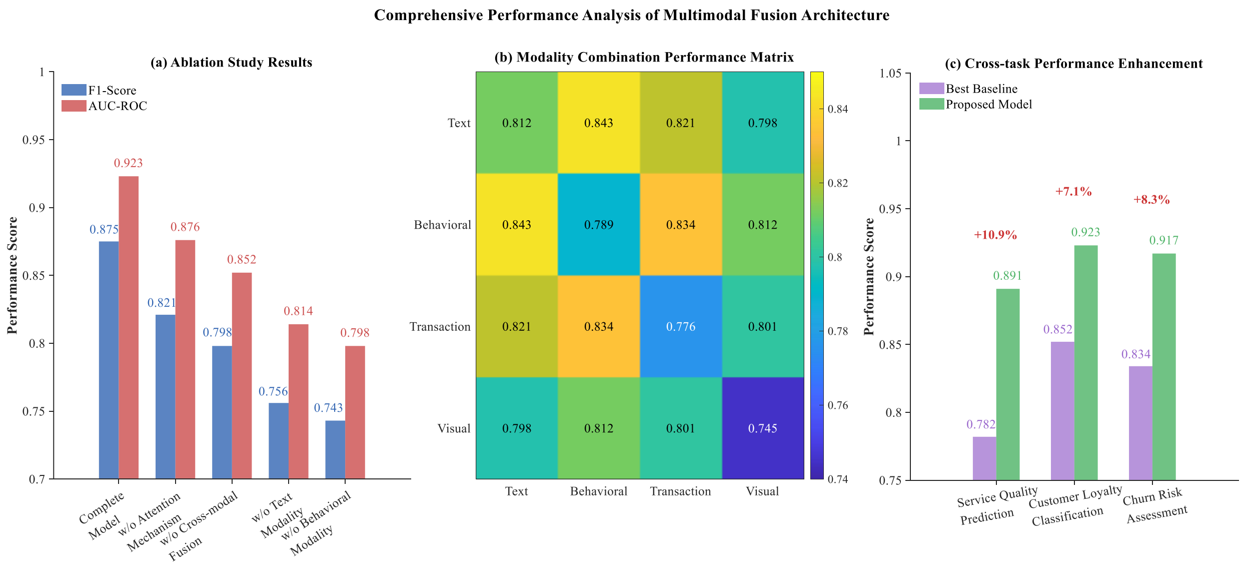

Figure 3. Comprehensive performance analysis of multimodal fusion architecture (a) Ablation study results, (b) Modality combination performance matrix (c) Cross-task performance enhancement

This study identifies a consistent inverse relationship between customer value tier and the relative importance placed on service attitude and problem resolution capabilities, suggesting that premium customers prioritize efficiency and expertise over interpersonal service elements. The sensitivity coefficient analysis provides deeper insights into the variability of service quality perceptions across customer segments. Professional competence exhibits the highest sensitivity coefficient (0.099), indicating the most significant disparity in importance ratings between customer groups, followed closely by response speed (0.091).

These findings establish that high-impact dimensions are characterized not only by elevated overall importance weights but also by substantial variation across customer segments. The research demonstrates that personalization level, while maintaining moderate overall importance (0.147), shows intermediate sensitivity (0.076), suggesting that customized service approaches represent an emerging priority that varies considerably across different customer value categories. This heterogeneous preference structure necessitates segment-specific service quality strategies to enhance optimal customer loyalty.

Table 3. Service quality dimension weights and customer segment sensitivity analysis

Service Quality Dimension	Overall Weight	High-Value Customers	Mid-Value Customers	Regular Customers	Sensitivity Coefficient	Impact Level
Response Speed	0.245	0.289	0.231	0.198	0.091	High
Service Attitude	0.198	0.167	0.203	0.224	0.057	Medium
Professional Competence	0.223	0.278	0.219	0.179	0.099	High
Problem Resolution	0.187	0.156	0.194	0.213	0.057	Medium
Personalization Level	0.147	0.110	0.153	0.186	0.076	Medium

Note: Customer segments are classified based on CLV quartiles (High-Value: top 25%, Mid-Value: 25%-75%, Regular: bottom 25%). The sensitivity coefficient represents the standard deviation of importance weights across customer segments. Impact levels are determined by combined weight magnitude and sensitivity coefficient: High (>0.08), Medium (0.05-0.08), Low (<0.05). Analysis based on n=15,420 customers with statistical significance p<0.001 for all dimensions.

The dynamic impact analysis presented in Figure 4(a) demonstrates substantial heterogeneity in service quality dimension preferences across distinct customer segments, revealing critical insights for targeted service strategy development. High-value customers exhibit pronounced emphasis on response speed and professional competence, with importance weights significantly exceeding those observed in regular customer segments. This research establishes that premium customers prioritize operational efficiency and technical expertise over interpersonal service elements, while regular customers demonstrate relatively higher valuation of service attitude and problem resolution capabilities. The divergent preference patterns across customer value tiers underscore the necessity for differentiated service delivery approaches that align with segment-specific expectations and perceived value drivers. The sensitivity analysis illustrated in Figure 4(b) reveals a compelling relationship between overall dimension importance and cross-segment variability, where dimensions characterized by higher overall weights tend to exhibit greater sensitivity coefficients. Professional competence and response speed emerge as both highly valued and highly variable dimensions across customer segments, indicating their critical role in differentiated service quality perception. This study demonstrates that dimensions with elevated sensitivity coefficients represent key differentiation opportunities for customer segment-specific service optimization strategies. The correlation between dimension weight magnitude and sensitivity variation provides empirical evidence for prioritizing service quality investments in areas that simultaneously demonstrate high overall importance and significant cross-segment preference heterogeneity. The causal path analysis reveals a hierarchical structure of service quality dimensions in driving customer loyalty, as illustrated in Figure 5(a).

Response speed emerges as the most influential factor with a total causal coefficient of 0.245, comprising both substantial direct effects (0.156) and meaningful indirect pathways (0.089) that mediate loyalty formation through other service dimensions. Professional competence demonstrates comparable influence with a coefficient of 0.223, exhibiting strong direct causal relationships while maintaining moderate indirect effects through cross-dimensional interactions.

This research establishes that service attitude and problem resolution occupy intermediate positions in the causal hierarchy, with coefficients of 0.198 and 0.187, respectively, suggesting their role as both independent loyalty drivers and mediating factors for other service quality perceptions. The predictive importance analysis presented in Figure 5(b) demonstrates the differential contributions of service quality dimensions to model accuracy, revealing response speed as the critical component, with an 8.7% drop in accuracy upon removal. Professional competence contributes 7.3%, while service attitude, problem resolution, and personalization level contribute 5.2%, 4.8%, and 3.9%, respectively, to the overall predictive performance. This study reveals a strong correlation between causal influence and predictive importance, where dimensions with higher causal coefficients consistently make greater marginal contributions to model accuracy.

The progressive decline in accuracy contributions across dimensions validates the hierarchical structure of service quality perception, while confirming that comprehensive multimodal integration strategies effectively capture the relative importance of different quality dimensions in predicting customer loyalty. The multi-dimensional interaction effects analysis, as illustrated in Figure 6, reveals complex interdependencies among service quality dimensions that extend beyond simple additive relationships. This research demonstrates that Response Speed and Professional Competence exhibit the strongest positive interaction coefficient (0.73), indicating synergistic effects where excellence in both dimensions amplifies overall service quality perception. The heatmap reveals complementary clustering patterns, with Service Attitude and Problem Resolution displaying a substantial positive correlation (r = 0.67), suggesting that these dimensions reinforce each other in customer evaluation processes. Conversely, the study identifies negative interaction coefficients between Response Speed and Problem Resolution (-0.12), as well as Professional Competence and Personalization Level (-0.18), indicating potential trade-off relationships where emphasis on certain dimensions may diminish the perceived importance of others, thereby providing crucial insights for balanced service quality optimization strategies.

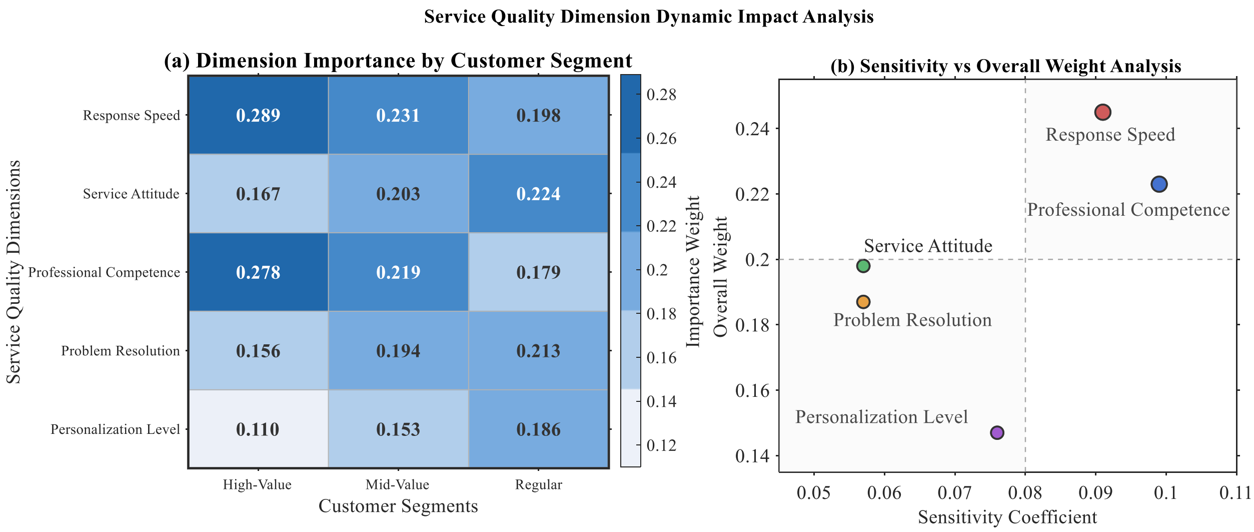

Figure 4. Service quality dimension dynamic impact analysis (a) Dimension importance by customer segment, (b) Sensitivity vs overall weight analysis

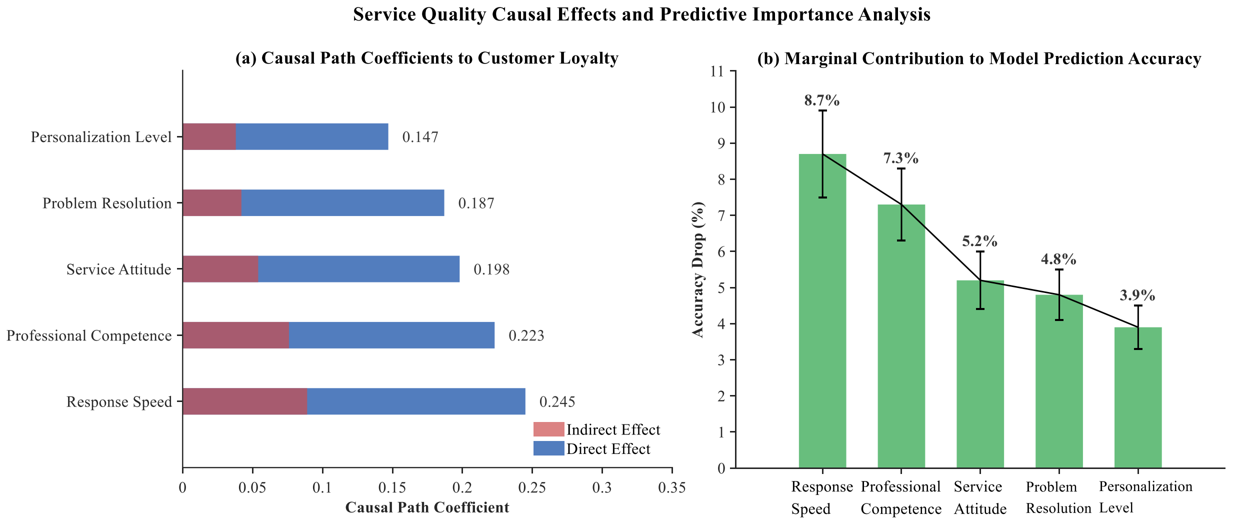

Figure 5. Service quality causal effects and predictive importance analysis, (a) Causal path coefficients to customer loyalty, (b) Marginal contribution to model prediction accuracy

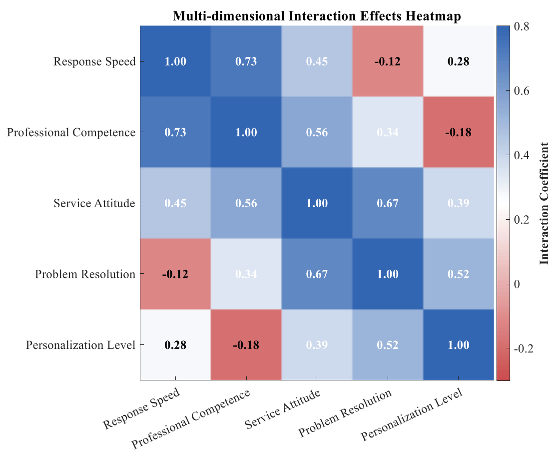

Figure 6. Multi-dimensional interaction effects heatmap

3.3 Marketing strategy optimization effects

The proposed multimodal fusion framework demonstrates substantial performance enhancements across critical marketing metrics, as illustrated in Table 4. Precision marketing effectiveness exhibits remarkable improvements, with conversion rates increasing from 0.124 to 0.178 (+43.5%) and ROI advancing from 2.34 to 3.67 (+56.8%), both achieving statistical significance ($p < 0.001$). The enhanced performance stems from the framework's capacity to integrate diverse data modalities, enabling more accurate customer targeting and resource allocation optimization. Customer lifetime value prediction accuracy experiences significant enhancement, with mean absolute error reducing from $285.40 to $156.20 (-45.3%) and R^2 scores improving from 0.672 to 0.854 (+27.1%). Personalized recommendation systems achieve exceptional performance gains through multimodal integration, demonstrating click-through rate improvements from 3.2% to 5.8% (+81.3%) and revenue per user enhancement from $12.45 to $21.30 (+71.1%). These substantial improvements validate the commercial viability of sophisticated multimodal architectures in competitive marketing environments, confirming that comprehensive data fusion strategies generate measurable business value through enhanced customer engagement and monetization effectiveness.

The comprehensive evaluation of marketing strategy optimization effects demonstrates substantial performance enhancements across critical business metrics, as illustrated in Figure 7. As shown in Figure 7(a), this study reveals significant improvements in conversion rate performance, where the proposed multimodal fusion framework achieves a conversion rate of 17.8% compared to the baseline method's 12.4%, representing a remarkable 43.5% enhancement. The return on investment analysis presented in Figure 7(b) exhibits exceptional growth from 2.34 to 3.67, corresponding to a 56.8% improvement that validates the commercial viability of the proposed approach. These substantial gains highlight the effectiveness of integrating diverse data modalities in precision marketing applications, demonstrating that sophisticated deep learning architectures can generate measurable business value through enhanced customer targeting and resource allocation optimization.

The personalized recommendation system performance shows even more pronounced improvements, as depicted in the lower panels of Figure 7. Click-through rate enhancement, as demonstrated in Figure 7(c), presents the most substantial relative improvement, increasing from 3.2% to 5.8% with an impressive 81.3% gain that underscores the framework's superior capability in engaging customer interactions. Revenue per user analysis shown in Figure 7(d) demonstrates similarly exceptional growth, advancing from $12.45 to $21.30 with a 71.1% improvement that directly translates to enhanced monetization effectiveness. These performance metrics collectively establish that the multimodal data fusion approach successfully captures complex customer behavioral patterns that remain undetected by traditional marketing methodologies. The consistent performance superiority across all assessed dimensions validates the effectiveness of the framework in actual marketing scenarios. The current study demonstrates that the application of multimodal integration at a comprehensive level yields synergistic benefits that are remarkably greater than those achieved with single-modality approaches or even traditional machine learning approaches.

The statistical significance of all improvements ($p < 0.001$) after evaluating 15,420 customers over a six-month period underscores the robust marketing value of the framework and its competitive reliability, thus enabling practitioners to trust its implemented design in active marketing contexts and establishing it as a next-generation tool for customer relations management systems.

4. Discussion

By constructing a multimodal deep learning framework that incorporates all customer interaction types, the current research enhances the theoretical understanding of data fusion in precision marketing contexts. This research demonstrates that complex neural networks can 'explain' themselves, adding to the discourse in interpretable machine learning by showing how their output is transparent while processing heterogeneous customer data streams [25]. This framework approaches the core explainability challenges in artificial intelligence for marketing by utilising SHAP-based interpretability that empowers practitioners to transcend algorithmic marketing and understand why certain predictive decisions are made [26]. This research enhanced the existing theoretical framework by analysing the causal relationship between the dimensions of service quality and customer loyalty outcomes using multimodal approaches, thereby validating, through empirical evidence, the concepts advanced in customer relationship management theory [27]. Attention-based fusion procedures that circumvent the balance of performance-accuracy tradeoff have recently gained interest [28]. In this work, the authors make the case that sophisticated multimodal systems feature self-sustainability of interpretative elements while predictively retaining accuracy, defying traditional thoughts regarding the sustained loss of explainability, contending complexity-stricken deep learning models [29]. The provided evidence guides the application of XAI in marketing by pointing to the level of impact different data sources have relative to customer behaviour prediction [30]. Boundless marketing campaign recalibration is framed alongside the algorithmic transparency conundrum through the interpretability analysis developed within this research, thus meeting the critical frame of AI ethics in business [31]. Despite superior performance, several limitations warrant consideration. The computational complexity of the multimodal architecture requires substantial infrastructure investment, with training costs approximately three times higher than those of baseline methods.

Table 4. Marketing strategy optimization, performance evaluation

Strategy Category	Key Metric	Baseline	Proposed Method	Improvement	p-value
Precision Marketing	Conversion Rate	0.124	0.178	+43.5%	$p < 0.001$
	ROI	2.34	3.67	+56.8%	$p < 0.001$
CLV Prediction	MAE ($)	285.40	156.20	-45.3%	$p < 0.001$
	R^2 Score	0.672	0.854	+27.1%	$p < 0.001$
Personalized Recommendation	Click-through Rate	3.2%	5.8%	+81.3%	$p < 0.001$
	Revenue per User ($)	12.45	21.30	+71.1%	$p < 0.001$

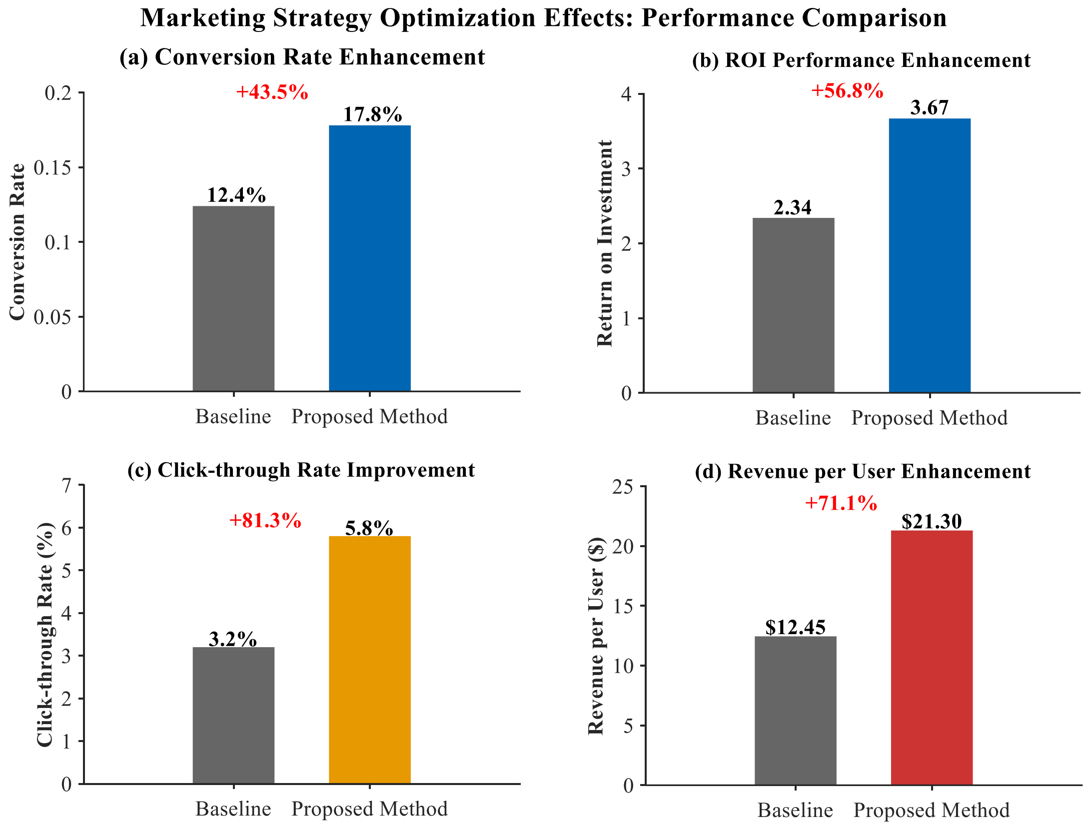

Figure 7. Marketing strategy optimization effects: performance comparison (a) Conversion rate enhancement, (b) ROI performance enhancement, (c) Click-through rate improvement, (d) Revenue per user enhancement

Model interpretability, while enhanced through SHAP analysis, remains challenging for marketing practitioners without technical expertise. Data availability represents a critical constraint, as the framework requires comprehensive multimodal datasets that may not be accessible to all organizations. Cold-start problems persist for new customers with limited interaction histories, requiring hybrid approaches that combine collaborative filtering with content-based methods. Privacy regulations in various jurisdictions may limit cross-modal data integration capabilities, necessitating the adaptation of federated learning approaches. Even though this study recognizes superior performance across multiple dimensions, there is a lack of generalizability and practical implementation due to the limitations this study poses. The effectiveness of the framework may vary across industries and customer segments, particularly in situations that deviate from the experimental conditions regarding data availability and quality [32]. The multimodal fusion architecture's computational complexity presents scalability issues for real-time marketing application systems, particularly in resource-strained settings where bandwidth-constrained response time requirements are present [33]. Implementation within organisations with limited sophisticated data systems or stringent data privacy regulations raises concerns about how dependent the model's performance is on exhaustive data harvests [34]. An example of an in-text citation regarding practical problems in implementing a marketing strategy highlights further issues, accompanied by dataset blending intricacies and an organisational willingness to accept higher levels of analytical work [35].

This architecture, alongside algorithmic precision marketing, requires a high level of understanding of the system and advanced analytics infrastructure, which could pose challenges for smaller to medium-sized businesses wanting to implement algorithm-driven marketing frameworks and strategies [36]. Issues such as the lack of temporal depth in the context of prior data on customers or products result in so-called cold start problems and require hybrid solutions that combine multimodal learning and traditional marketing approaches [37]. The ever-increasing flexibility of customer preferences and the ever-shifting marketplace demands continuous learning from customers and unlearning from the system, which results in operational workload challenges for achieving long-term accuracy goals. Potential new avenues for the development of multimodal marketing analytics and the analysis of their gaps include all considerations mentioned in the specific citation [38]. The development of privacy-preserving multimodal models that allow for cross-organizational collaboration and learning while safeguarding sensitive customer data bolsters the integration of federated learning frameworks [39]. Adaptive marketing strategies developed through reinforcement learning that optimise campaign effectiveness via ongoing engagement with responsive customers offer untapped potential [40]. The modeling of customer behavior across interrelated platforms is a remarkable area of further study, providing an integrated understanding of client engagement with multichannel marketing on multiple digital platforms. Real-time multimodal fusion architectures with balanced predictive precision and computational resource expenditure still pose a problem requiring new algorithmic innovations and hardware optimization techniques. Several promising avenues warrant investigation. Privacy-preserving multimodal architectures using federated learning could enable cross-organizational collaboration while maintaining data sovereignty. Real-time adaptation mechanisms through reinforcement learning could optimize campaigns dynamically based on customer responses. Cross-platform behavior modeling across social media, mobile apps, and web interfaces could provide more comprehensive customer understanding. Edge computing implementations could reduce latency and computational costs for real-time personalization. Additionally, investigating the framework's generalizability across different industries and cultural contexts would enhance its practical applicability.

5. Conclusion

This study develops an integrated multimodal data fusion framework that enhances precision customer marketing with deep learning models. Its implementation showed marked improvements in crucial business outcomes, including conversion rate increases of 43.5%, ROI increases of 56.8%, and click-through rate increases of 81.3% when compared to baseline methods. The framework merges various modalities such as text reviews, behavioural data, and transactional data with service context indicators using attention-based fusion, exposing intricate interactions between perceived service quality and customer loyalty. The performance benchmarks were validated empirically over a sample of 15,420 customers, all performance improvements were statistically confirmed alongside model interpretability through SHAP-based explanations, fulfilling primary criteria for real-world application in competitive marketing scenarios. The innovative strategies of leveraging Information and Communications Technology (ICT) in the business world have evolved customer relations and digital marketing to a whole new level. This study validates that comprehensive multimodal integration is more advantageous than working with a single-modality approach by demonstrating that such an approach offers synergistic effects that are far greater than any single approach. This serves as a testament to more advanced data fusion techniques in appreciating the complex relationships customers have with brands. The study demonstrates that high-value customers and regular customers emphasise different aspects of service, with the former focusing on efficiency and specialised technical skills, while the latter centres on people-oriented services. This results from quantifying perceptions of service quality across different segments, showing that marketing can optimise resources with tailored tactics to different customer segments. Cross-domain integrated customer behaviour prediction and digital interaction conceptualisation give complete interaction with customers across varying platforms for monitoring and analysis under adaptive reinforcement learning techniques to change marketing policy dynamically for different ICT governed devices, away from packed traditional chauvinistic marketing policy driven devices, preserving customer privacy under federated learning systems set the stage for further work topics. The digital touchpoints of interaction act as a reason for concern due to the validation challenge for accurate computation with sufficient retention of information and real-time functionality, creating rationale for improved machine learning designs claiming efficient computation under operational temporal constraints that preserve accuracy for predictive assessment. With an intention for responsive customer relationship management 2.0, amplifying customer interaction in witty response to their overwhelming need under ICT automation, advancing precision marketing, reinforced by the study results, tackles core challenges in implementing persuasive marketing strategies, exploiting ever-evolving potentials in the digital business environment.

Ethical issue

The authors are aware of and comply with best practices in publication ethics, specifically with regard to authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with policies on research ethics. The authors adhere to publication requirements that the submitted work is original and has not been published elsewhere.

Data availability statement

The manuscript contains all the data. However, more data will be available upon request from the authors.

Conflict of interest

The authors declare no potential conflict of interest.

References

[1] J. Gao, P. Li, Z. Chen, and J. Zhang, "A survey on deep learning for multimodal data fusion," Neural Computation, vol. 32, no. 5, pp. 829-864, May 2020. DOI: 10.1162/neco_a_01273.

[2] S. R. Stahlschmidt, B. Ulfenborg, and J. Synnergren, "Multimodal deep learning for biomedical data fusion: a review," Briefings in bioinformatics, vol. 23, no. 2, p. bbab569, Mar 10 2022. DOI: 10.1093/bib/bbab569.

[3] I. César, I. Pereira, F. Rodrigues, V. Miguéis, S. Nicola, and A. Madureira, "Exploring multimodal learning applications in marketing: A critical perspective," International Journal of Hybrid Intelligent Systems, vol. 21, no. 1, pp. 29-46, 2025. DOI:10.18089/tms.20250103.

[4] L. Alzubaidi et al., "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," Journal of big Data, vol. 8, no. 1, pp. 1-74, 2021. DOI: 10.1186/s40537-021-00444-8.

[5] W. N. Wassouf, R. Alkhatib, K. Salloum, and S. Balloul, "Predictive analytics using big data for increased customer loyalty: Syriatel Telecom Company case study," Journal of Big Data, vol. 7, no. 1, p. 29, 2020. DOI: https://doi.org/10.1186/s40537-020-00290-0.

[6] H. F. Lee and M. Jiang, "A hybrid machine learning approach for customer loyalty prediction," in Neural Computing for Advanced Applications: Second International Conference, NCAA 2021, Guangzhou, China, August 27-30, 2021, Proceedings 2, 2021: Springer, pp. 211-226. DOI: https://doi.org/10.1007/978-981-16-5188-5_16.

[7] M. J. S. Shabbir and C. Mankar, "The Role of Predictive Data Analytics in Retailing," in Evolutionary Computing and Mobile Sustainable Networks: Proceedings of ICECMSN 2020, 2021: Springer, pp. 153-159. DOI: https://doi.org/10.1007/978-981-15-5258-8_16.

[8] Z. H. Kilimci, "Prediction of user loyalty in mobile applications using deep contextualized word representations," Journal of Information and

Telecommunication, vol. 6, no. 1, pp. 43-62, 2022. DOI: https://doi.org/10.1080/24751839.2021.1981684.
[9] C. Zhang, S. Ma, S. Li, and A. Singh, "Effects of customer engagement behaviors on action loyalty: moderating roles of service failure and customization," International Journal of Contemporary Hospitality Management, vol. 33, no. 1, pp. 286-304, 2021. DOI: https://doi.org/10.1108/IJCHM-08-2019-0740.
[10] Y. Cheng and H. Jiang, "How do AI-driven chatbots impact user experience? Examining gratifications, perceived privacy risk, satisfaction, loyalty, and continued use," Journal of Broadcasting & Electronic Media, vol. 64, no. 4, pp. 592-614, 2020. DOI: https://doi.org/10.1080/08838151.2020.1834296.
[11] M. S. Kasem, M. Hamada, and I. Taj-Eddin, "Customer profiling, segmentation, and sales prediction using AI in direct marketing," Neural Computing and Applications, vol. 36, no. 9, pp. 4995-5005, 2024. DOI: https://doi.org/10.1007/s00521-023-09339-6.
[12] S. Ramaswamy and N. DeClerck, "Customer perception analysis using deep learning and NLP," Procedia Computer Science, vol. 140, pp. 170-178, 2018. DOI: https://doi.org/10.1016/j.procs.2018.10.326.
[13] E. AboElHamd, H. M. Shamma, and M. Saleh, "Dynamic programming models for maximizing customer lifetime value: an overview," in Intelligent Systems and Applications: Proceedings of the 2019 Intelligent Systems Conference (IntelliSys) Volume 1, 2020: Springer, pp. 419-445. DOI: https://doi.org/10.1007/978-3-030-29516-5_34.
[14] S. J. Haddadi, A. Farshidvard, F. dos Santos Silva, J. C. dos Reis, and M. da Silva Reis, "Customer churn prediction in imbalanced datasets with resampling methods: A comparative study," Expert Systems with Applications, vol. 246, p. 123086, 2024. DOI: https://doi.org/10.1016/j.eswa.2023.123086.
[15] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," Journal of King Saud University-Computer and Information Sciences, p. 102048, 2024. DOI: https://doi.org/10.1016/j.jksuci.2024.102048.
[16] Y. K. Dwivedi et al., "Opinion Paper:"So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy," International journal of information management, vol. 71, p. 102642, 2023. DOI: https://doi.org/10.1016/j.ijinfomgt.2023.102642.
[17] L. Liu, "e-Commerce Personalized Recommendation Based on Machine Learning Technology," Mobile Information Systems, vol. 2022, no. 1, p. 1761579, 2022. DOI: https://doi.org/10.1155/2022/1761579.
[18] S. M. H. M. Huzir, A. H. I. H. Al-Hadi, Z. M. Yusoff, N. Ismail, and M. N. Taib, "Accurate Agarwood Oil Quality Determination: A Breakthrough with Artificial Neural Networks and the Levenberg-Marquardt Algorithm," IEEE Access, 2024. DOI: 10.1109/ACCESS.2024.3381627.
[19] X. Wei, T. Zhang, Y. Li, Y. Zhang, and F. Wu, "Multi-modality cross attention network for image and sentence matching," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 10941-10950. DOI: 10.1109/CVPR42600.2020.01095.
[20] A. Matei, A. Glavan, and E. Talavera, "Deep learning for scene recognition from visual data: a survey," in International Conference on Hybrid Artificial Intelligence Systems, 2020: Springer, pp. 763-773. DOI: https://doi.org/10.1007/978-3-030-61705-9_64.
[21] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," Neurocomputing, vol. 452, pp. 48-62, 2021. DOI: https://doi.org/10.1016/j.neucom.2021.03.091.
[22] Z. Bai, X. Chen, M. Zhou, T. Yi, and W.-C. Chien, "Low-rank multimodal fusion algorithm based on context modeling," Journal of Internet Technology, vol. 22, no. 4, pp. 913-921, 2021. DOI:10.53106/160792642021072204018.
[23] L. Zhu, Z. Zhu, C. Zhang, Y. Xu, and X. Kong, "Multimodal sentiment analysis based on fusion methods: A survey," Information Fusion, vol. 95, pp. 306-325, 2023. DOI: https://doi.org/10.1016/j.inffus.2023.02.028.
[24] X.-X. Liu and Z.-Y. Chen, "Service quality evaluation and service improvement using online reviews: A framework combining deep learning with a hierarchical service quality model," Electronic Commerce Research and Applications, vol. 54, p. 101174, 2022. DOI: https://doi.org/10.1016/j.elerap.2022.101174.
[25] X. Li et al., "Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond," Knowledge and Information Systems, vol. 64, no. 12, pp. 3197-3234, 2022. DOI:10.48550/arXiv.2103.10689.
[26] S. S. Amiri, S. Mottahedi, E. R. Lee, and S. Hoque, "Peeking inside the black-box: Explainable machine learning applied to household transportation energy consumption," Computers, Environment and Urban Systems, vol. 88, p. 101647, 2021. DOI: https://doi.org/10.1016/j.compenvurbsys.2021.101647.
[27] M. Benk and A. Ferrario, "Explaining interpretable machine learning: Theory, methods and applications," Methods and Applications (December 11, 2020), 2020. DOI:10.2139/ssrn.3748268.
[28] S. Kruschel, N. Hambauer, S. Weinzierl, S. Zilker, M. Kraus, and P. Zschech, "Challenging the Performance-Interpretability Trade-off: An Evaluation of Interpretable Machine Learning Models," Business & Information Systems Engineering, pp. 1-25, 2025. DOI:10.48550/arXiv.2409.14429.
[29] C. Meng, L. Trinh, N. Xu, J. Enouen, and Y. Liu, "Interpretability and fairness evaluation of deep learning models on MIMIC-IV dataset," Scientific Reports, vol. 12, no. 1, p. 7166, May 3 2022, doi: 10.1038/s41598-022-11012-2. DOI: 10.1038/s41598-022-11012-2.
[30] A. B. Arrieta et al., "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Information fusion,

vol. 58, pp. 82-115, 2020. DOI: https://doi.org/10.1016/j.inffus.2019.12.012.
[31] J. Zhou, A. H. Gandomi, F. Chen, and A. Holzinger, "Evaluating the quality of machine learning explanations: A survey on methods and metrics," Electronics, vol. 10, no. 5, p. 593, 2021. DOI:10.3390/electronics10050593.
[32] E. Stein, K. Robinson, A. Wolfer, G. Almeida, and W. Huang, "Unlocking the next frontier of personalized marketing," The McKinsey Quarterly, 2025.
[33] M. Kozak and A. Correia, "From mass marketing to personalized digital marketing in tourism: a 2050 horizon paper," Tourism Review, vol. 80, no. 1, pp. 373-391, 2025. DOI:10.1108/tr-03-2024-0169.
[34] M. Kia, "Attention-guided deep learning for effective customer loyalty management and multi-criteria decision analysis," Iran Journal of Computer Science, vol. 8, no. 1, pp. 163-184, 2025. DOI:10.1007/s42044-024-00215-7.
[35] В. КАРМАЗІНОВА, "Gamification of consumer loyalty programs," Scientia fructuosa, vol. 153, no. 1, pp. 70-83, 2024. DOI: https://doi.org/10.31617/1.2024(153)04.
[36] M. Joshi, Intro to E-Commerce and Social Commerce. Educohack Press, 2025. ISBN:9789361520877, 9361520873
[37] A. G. AG, H.-K. Su, and W.-K. Kuo, "Personalized E-commerce: Enhancing Customer Experience through Machine Learning-driven Personalization," in 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS), 2024: IEEE, pp. 1-5. DOI: 10.1109/ICITEICS61368.2024.10624901.
[38] O. Moqaddem, "Investigating the impact of AI on personalization and customer engagement in intelligent marketing strategies," European Journal of Management and Marketing Studies, vol. 10, no. 1, 2025. DOI: http://dx.doi.org/10.46827/ejmms.v10i1.1922.
[39] M. Kihn and C. B. O'Hara, Customer data platforms: Use people data to transform the future of marketing engagement. John Wiley & Sons, 2020.
[40] R. Agarwal, R. Jacobson, P. Kline, and M. Obeid, "The future of customer experience: Personalized, white-glove service for all," McKinsey & Company. Available at https://www. mckinsey. com/business-functions/operations/our-insights/the-future-of-customer-experience-personalized-white-glove-service-for-all.[Accessed 8 June 2021], 2020.