Advanced Engineering Research (Rostov-on-Don). 2026;26(1):2211. eISSN 2687−1653

INFORMATION TECHNOLOGY,
COMPUTER SCIENCE AND MANAGEMENT
ИНФОРМАТИКА,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ

UDC 004.89
https://doi.org/10.23947/2687-1653-2026-26-1-2211

Original Empirical Research

A Customer Lifetime Value-aware Framework for Strategic Churn Prediction Using Deep Learning

EDN: FDZUZD

Uma Maheswari Gurusamy[1] ✉, **Meenakshi Anantharaman**[1], **Ram Prasath Selvamani**[1], **Sangeetha Vijayarajan**[2]
[1] Kamaraj College of Engineering and Technology, Virudhunagar, Tamil Nadu, India
[2] SRM Easwari Engineering College, Ramapuram, Chennai, India
✉ uma.optimist@gmail.com

Abstract

Introduction. Customer churn prediction represents a challenge in the current era of rapid digital transformation, hyper-competition, and data-driven marketing. In sectors such as telecommunications and banking, even marginal reductions in churn translate to significant revenue protection. Numerous companies employ uniform approaches, leading to the inefficient allocation of marketing resources and loss of loyal customers. Recent research has advanced along two largely separate domains. The first focuses on improving predictive accuracy through machine learning and deep learning techniques. Another stream, rooted in marketing science, emphasizes the economic dimension of churn, introducing Customer Lifetime Value (CLV) as a key metric. Existing solutions either maximize accuracy at high computational cost or discuss value-based strategy without providing a technical, implementable system. To bridge this gap, this paper aims to create, test, and present a comprehensive churn control system integrating customer lifetime value framework (CVLV). To achieve this, the following tasks are addressed: segmenting customers based on dynamic CLV and churn risk scores; evaluating the efficiency of various neural network configurations; and building a decision model that assigns optimal deep learning architectures for targeted retention, seamlessly integrating data analytics with corporate strategy.

Materials and Methods. The study was performed on two datasets: IBM Telco Customer Churn (7,043 customers, 21 features, binary churn) and Santander Customer Transaction Prediction (200,000 records, 200 numerical features, binary target variable). The data were preprocessed to address class imbalance and split 70-15-15 (train-validation-test) using 5-fold cross-validation. ANN (3–6 layers) and RNN/LSTM models were compared within the CVLV framework. The training utilized Adam optimizer, L2 regularization, dropout, early stopping, gradient clipping, and uniform batch size and epoch settings. The performance was evaluated based on accuracy, loss, and the Pareto frontier. Subsequently, customers were segmented by CLV/risk level, and retention strategies were assigned to the respective optimal models.

Results. The comprehensive assessment of artificial neural networks (ANN) and recurrent neural networks (RNN) shows that RNN with 2 layers achieved marginally higher accuracy of 0.90, while the 3-layer ANN produced the best robustness with a loss of 0.25 with relatively similar predictive performance. With the CVLV framework, RNN 2L is assigned for high value, high risk relationships that need the most precision, ANN 3L is assigned for stable, high value relationships, and general RNN for low value customers.

Discussion. This work has shown that the CVLV framework strategically optimizes churn prediction by aligning deep learning models with customer value-risk profiles. The data obtained confirm that ANN 3L provides optimal robustness while RNN 2L achieves superior accuracy for temporal patterns, together enabling more efficient and targeted retention interventions across industries. This approach can be deployed across the telecommunications, banking and retail sectors and facilitate a meaningful connection between technical model performance and strategic decision-making, enabling organizations to deploy retention efforts effectively by aligning model capability with the customer's value and probability of churn. The findings indicates that strategic model assignments based on CLV-risk profiles led to improved efficiencies associated with retention without compromising predictive reliability.

Conclusion. The main results are that the ANN 3L model provides the optimal balance of accuracy (0.875) and robustness (loss: 0.25) for churn prediction, while the RNN 2L achieves peak accuracy (0.90) for high-risk segments. The practical significance lies in the proposed CVLV framework, which enables businesses to strategically align deep learning model selection with customer lifetime value, improving retention efficiency. Further research will focus on integrating real-time CLV updates and validating the framework across additional industry domains.

Keywords: Customer Lifetime Value-aware, churn prediction, ANN, RNN, accuracy, loss, optimum model

Acknowledgements. The authors would like to express their sincere gratitude to the curators and maintainers of the IBM Telco Customer Churn and Santander Customer Transaction Prediction datasets for making their data publicly available, which was fundamental to the empirical validation conducted in this study. The computational resources provided by respective institutions are gratefully acknowledged.

For Citation. Uma Maheswari Gurusamy, Meenakshi Anantharaman, Ram Prasath Selvamani, Sangeetha Vijayarajan. A CVLV Framework for Strategic Churn Prediction Using Deep Learning. *Advanced Engineering Research (Rostov-on-Don).* 2026;26(1):2211. https://doi.org/10.23947/2687-1653-2026-26-1-2211

Оригинальное эмпирическое исследование

Методология прогнозирования стратегического оттока клиентов с учётом их пожизненной ценности на основе глубокого обучения

Ума Махешвари Гурусами[1] ✉, **Минакши Анантараман**[1], **Рам Прасат Сельвамани**[1],
Санджита Виджаяраджан[2]
[1] Инженерно-технологический колледж имени Камараджа, г. Вирудунагар, штат Тамилнад, Индия
[2] Инженерный колледж имени Ишвари при Университете SRM, Рамапурам, г. Ченнаи, Индия
✉ uma.optimist@gmail.com

Аннотация

Введение. Прогнозирование оттока клиентов приобретает особую актуальность в эпоху цифровой трансформации и обострения конкуренции. В таких секторах, как телекоммуникации и банковское дело, даже минимальное сокращение этого показателя способно заметно укрепить финансовые позиции. Многие компании применяют унифицированные подходы к удержанию клиентов, что приводит к нерациональному использованию ресурсов и утрате лояльных пользователей. Современные исследования фокусируются на двух ключевых направлениях. Первое из них посвящено совершенствованию точности прогнозирования посредством алгоритмов машинного обучения. Второе подчеркивает экономическую составляющую, включая пожизненную ценность клиента (CLV). Существующие подходы либо достигают максимальной точности за счет значительных вычислительных затрат, либо предлагают концепции, основанные на факторе ценности, но не имеющие практической технической реализации. Для преодоления этого разрыва в настоящей работе предлагается создать, испытать и представить комплексную систему контроля оттока клиентов с интеграцией жизненной ценности клиента (CVLV). Цель исследования заключается в разработке и верификации методологии контроля оттока с учетом жизненной ценности клиента (CVLV). Для ее достижения решаются следующие задачи: сегментация аудитории по динамическим метрикам CLV и вероятности ухода; оценка эффективности разнообразных конфигураций нейронных сетей; построение модели, которая выявляет наилучшие архитектуры глубокого обучения для целенаправленного удержания клиентов, гармонично сочетая аналитику данных с корпоративной стратегией.

Материалы и методы. Исследование проводилось на двух наборах данных: IBM Telco Customer Churn (7 043 клиента, 21 признак, бинарный отток) и Santander Customer Transaction Prediction (200 000 записей, 200 числовых признаков, бинарный целевой признак). Данные обрабатывались с учётом дисбаланса классов и делились в пропорции 70–15–15 с 5-кратной кросс-проверкой. Сравнивались ANN (3–6 слоёв) и RNN/LSTM в CVLV-фреймворке. При обучении использовались Adam, L2-регуляризация, dropout, ранняя остановка, обрезка градиентов, единые настройки батча и эпох. Эффективность оценивалась по точности, функции потерь и Парето-фронту. Затем клиенты сегментировались по уровню пожизненной ценности (CLV) и риску оттока. Затем моделям назначались стратегии.

Результаты исследования. Всесторонняя оценка искусственных (ANN) и рекуррентных (RNN) нейронных сетей показала, что двухслойная RNN обеспечивает незначительно более высокую точность (0,90), в то время как трёхслойная ANN демонстрирует наилучшую устойчивость с минимальными потерями (0,25) при сопоставимой прогностической эффективности. В рамках CVLV-фреймворка это определяет назначение моделей: RNN 2L используется для высокоценных клиентов с высоким риском оттока, где критически важна максимальная точность прогноза; ANN 3L — для стабильных высокоценных отношений; а базовая RNN — для клиентов с низкой ценностью.

Обсуждение. Проведённое исследование продемонстрировало, что CVLV-фреймворк стратегически оптимизирует прогнозирование оттока клиентов за счёт согласования моделей глубокого обучения с ценностно-рисковыми профилями клиентов. Полученные данные подтверждают, что модель ANN 3L обеспечивает оптимальную устойчивость, а RNN 2L достигает максимальной точности при работе с временными закономерностями. Совместное их применение позволяет реализовывать более эффективные и целенаправленные мероприятия по удержанию клиентов в различных отраслях. Данный подход может быть внедрён в телекоммуникационном, банковском секторах, в сфере розничных продаж. Он устанавливает содержательную связь между техническими характеристиками модели и стратегическим принятием решений, позволяя организациям эффективно распределять усилия по удержанию, соотнеся возможности модели с ценностью клиентов и вероятностью их оттока. Результаты указывают на то, что стратегическое назначение моделей на основе CLV-рисковых профилей приводит к повышению эффективности мероприятий по удержанию без ущерба для надёжности прогнозов.

Заключение. Основные результаты заключаются в том, что модель ANN 3L обеспечивает оптимальный баланс между точностью (0,875) и устойчивостью (потери: 0,25) в прогнозировании оттока, в то время как модель RNN 2L достигает максимальной точности (0,90) для сегментов с высоким риском. Практическая значимость исследования состоит в предложенном CVLV-фреймворке, который позволяет бизнесу стратегически соотносить выбор модели глубокого обучения с пожизненной ценностью клиента, повышая эффективность мероприятий по удержанию. Дальнейшие исследования будут сосредоточены на интеграции механизмов обновления CLV в реальном времени и валидации фреймворка в других отраслях.

Ключевые слова: учёт пожизненной ценности клиента, прогнозирование оттока клиентов, ANN, RNN, точность, функция потерь, оптимальная модель

Благодарность. Авторы выражают искреннюю благодарность кураторам и разработчикам наборов данных IBM Telco Customer Churn и Santander Customer Transaction Prediction за предоставление открытого доступа к данным, которые послужили основой для эмпирической валидации, проведенной в данном исследовании. Мы также признательны за вычислительные ресурсы, предоставленные учреждениями.

Для цитирования. Ума Махешвари Гурусами, Минакши Анантараман, Рам Прасат Сельвамани, Санджита Виджаяраджан. Методология прогнозирования стратегического оттока клиентов с учётом их пожизненной ценности на основе глубокого обучения. *Advanced Engineering Research (Rostov-on-Don).* 2026;26(1):2211. https://doi.org/10.23947/2687-1653-2026-26-1-2211

Introduction. In the competitive digital economy, preempting customer churn is critical for profitability [1]. However, a core operational inefficiency persists: retention strategies rely on models that maximize aggregate predictive accuracy rather than business value [2]. This leads to the costly misallocation of resources, where sophisticated, uniform interventions fail to protect high-value relationships while expending capital on low-value segments [3]. This disconnect reveals a pivotal scientific challenge: optimizing the return on investment (ROI) of retention efforts by aligning predictive modeling with economic impact.

The scholarly pursuit of this goal has progressed along two fundamentals, yet largely parallel, tracks [4]. The first, rooted in marketing and operational research, established Customer Lifetime Value (CLV) as the cornerstone metric for strategic customer prioritization and resource allocation [5]. This stream provides the essential “why”, framing retention as an economic optimization problem [6]. Simultaneously, a second track driven by data science has relentlessly advanced the technical “how.” Research has evolved from traditional classifiers and ensemble methods [7] to powerful deep learning (DL) architectures [8]. Significant work demonstrates the efficacy of Artificial Neural Networks (ANNs) for modeling complex, non-linear relationships in customer profile data [9], while Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory networks, have proven superior for capturing sequential [10], temporal patterns in customer behavior [11]. Contemporary studies have made substantial progress in tackling associated technical challenges [12], such as handling class imbalance in churn data through advanced sampling [13] and cost-sensitive learning techniques [14] and improving model interpretability for business stakeholders using explainable AI (XAI) tools [15].

Despite these significant advancements, a critical synthesis gap remains. The literature reveals three specific and actionable limitations that prevent the translation of technical capability into strategic value:

Model-Strategy Decoupling: Studies proposing complex models (e.g., deep RNNs) prioritize technical performance metrics (e.g., accuracy, AUC) [16] without embedding their deployment within a cost-benefit calculus, neglecting the computational economics of large-scale deployment [17].

Static Value Integration: Where CLV is considered, it is predominantly used as a static, post-hoc segmentation filter rather than as a dynamic [18], integral variable that actively guides the principled selection of a prediction architecture [19].

Context-Agnostic Optimization: Comparative analyses of ANN and RNN architectures often declare a universally “superior” model [20], failing to provide a prescriptive decision framework for which architecture is optimal given a specific customer's value profile and data characteristics [21].

Consequently, this study addresses the absence of a prescriptive, operational framework that dynamically synthesizes CLV theory with DL architecture selection to balance predictive performance and economic efficiency. A review of current methods for churn prediction and its limitations is shown in table 1. To bridge this gap, we introduce the Customer Lifetime Value-aware Churn (CVLV) Framework. Its primary aim is to develop and validate a decision model that strategically aligns model complexity with individual customer value and risk. The work is guided by three specific objectives: (1) to implement a dynamic segmentation engine based on computed CLV and predicted churn risk; (2) to empirically benchmark the performance and trade-offs of ANN and RNN architectures across these defined segments; and (3) to validate that this value-driven model assignment yields a significantly higher retention ROI compared to conventional, uniform modeling approaches. The expected contribution is a unified, actionable system that closes the critical loop between predictive accuracy, economic value, and strategic resource allocation.

Table 1

Review of current methods for churn prediction and its limitations

Reference	Core Focus	Methodology	Key Strength	Key Limitation
[1] (2018)	Retention strategy effectiveness	Behavioral analysis + Field experiments	Challenges conventional targeting	No technical framework provided
[5] (2020)	Big data telco churn prediction	Large-scale feature engineering	Handles massive datasets	No segment-specific optimization
[19] (2022)	Hybrid deep learning for segmentation	CLV-risk matrices + DL	Integrates value and risk	Limited cross-industry validation
[20] (2023)	CLV-aware neural networks	Value-based DL architectures	Modern deep learning approach	Focuses on single architecture
[22] (2023)	Economic-aware deep learning	Cost-optimized model selection	ROI-focused approach	Limited to cost, not value segmentation
[23] (2022)	Resource allocation in retention	Economic modelling	Cost-benefit analysis focus	Lacks technical implementation

Materials and Methods

CVLV Framework using Deep and Learning

Datasets and Preprocessing

The two benchmark datasets, the IBM Telco Customer Churn dataset [24] and the Santander Customer Transaction Prediction dataset [25], were used for cross-industry validation of the CVLV framework. The Telco dataset comprises 7,043 customer records with 21 features, including demographic, account, and service usage attributes, with a binary churn indicator serving as the target variable. The Santander dataset includes 200,000 anonymized banking customer records, each described by 200 numerical features, where the target variable indicates a specific financial transaction used as a proxy for churn risk.

Both datasets underwent a standardized preprocessing pipeline implemented with scikit-learn (v1.2.2) and pandas (v1.5.3) to ensure replicability [26]. For the IBM Telco dataset, preprocessing involved handling a single missing value in TotalCharges by imputation with 0, converting 16 categorical features (e.g., Contract, PaymentMethod) via one-hot encoding, and standardizing all numerical features to a mean of 0 and standard deviation of 1 using StandardScaler. A key engineered feature, Customer Lifetime Value (CLV), was calculated as CLV = tenure * MonthlyCharges. For the Santander dataset, which contained no missing values, the 200 anonymized numerical features (var_0 to var_199) were normalized using RobustScaler to mitigate the influence of outliers inherent in transaction data. Given the anonymized nature of the features, a CLV proxy was derived as CLV_proxy = mean(var_i) * std(var_i) across all features for each customer, serving as a composite metric of transaction activity [27]. The target variable for Santander was repurposed as a churn-risk indicator. Following preprocessing, each dataset was partitioned into training (70%), validation (15%), and test (15%) sets using stratified sampling to preserve the original class distribution, ensuring the validation set was used for hyperparameter tuning and the test set for final, unbiased evaluation [28].

Computational Environment and Implementation

All experiments were conducted in a Python 3.9 environment. The deep learning models were built and trained using TensorFlow 2.10 with the Keras API. Key supporting libraries included NumPy (v1.23), pandas (v1.5), scikit-learn (v1.2), and Matplotlib (v3.6) for visualization. The computations were performed on an HP Victus gaming laptop equipped with 6GB of GPU memory, an Intel Core i5-13420H CPU (2.10 GHz), and 16.0 GB of RAM (15.6 GB usable), running the Windows operating system. This hardware configuration was selected to efficiently handle the training of multiple deep neural networks and the processing of the Santander dataset's high-dimensional feature space.

Core Algorithm and Implementation Sequence

The operationalization of the CVLV framework follows a sequential, five-step logic. Step 1 involves the data preprocessing detailed in Section 3.1. Step 2 focuses on Customer Lifetime Value (CLV) and Risk Calculation. For the IBM Telco dataset, a direct CLV metric was computed as CLV = tenure * MonthlyCharges. For the Santander dataset, an anonymized feature-derived proxy was calculated as CLV_proxy = mean(var_i) * std(var_i) across all 200 features [29]. Concurrently, a baseline churn risk probability score (0 to 1) was generated for every customer using a pretrained XGBoost classifier acting on the preprocessed feature sets [30]. Step 3 entails Dynamic Customer Segmentation. Customers were classified into one of four strategic quadrants by applying two data-driven thresholds: the 70th percentile of the CLV distribution within the training set defined the High-CLV segment, and a fixed risk probability of 0.5 defined the High-Risk segment [31]. This yielded the segments: High-CLV/High-Risk, High-CLV/Low-Risk, Low-CLV/High-Risk, and Low-CLV/Low-Risk [32].

Step 4 covers Model Specification, Assignment, and Training. Specific deep learning architectures were assigned to each high-value segment. For the High-CLV/High-Risk segment, a two-layer Long Short-Term Memory (LSTM) network (RNN 2L) was implemented, with layers of 64 and 32 units, to capture temporal dependencies [33]. For the High-CLV/Low-Risk segment, a three-layer Artificial Neural Network (ANN 3L) with 128, 64, and 32 neurons was deployed for efficient processing of structured features [34]. Both models used ReLU activations in hidden layers, a sigmoid output, the Adam optimizer (learning rate = 0.001), and binary cross-entropy loss. For the Low-CLV segments, a standard Logistic Regression model served as a computationally efficient baseline. During Model Training and Optimization, shared regularization techniques were applied to the deep learning models: L2 weight regularization ($\lambda = 0.001$), dropout (0.3/0.2 for ANN, 0.2 for LSTM), early stopping (patience = 15 epochs, monitoring validation loss with a minimum delta of 0.001), and gradient clipping with a global norm limit of 1.0. Training proceeded for a maximum of 100 epochs with a batch size of 64. The hyper parameters used in the proposed work is shown in table 2.

Table 2

Table of Hyperparameters used in the Proposed work

Hyperparameter	ANN 3L	RNN 2L
Hidden Layers	[128, 64]	LSTM: [64, 32]
Activation	ReLU (hidden), Sigmoid (output)	ReLU (Dense), Sigmoid (output)
Optimizer	Adam (*lr* = 0.001)	Adam (*lr* = 0.001)
Dropout Rate	[0.3, 0.2]	0.2 (after LSTM)
Batch Size	64	64
Epochs	100 (Early Stopping)	100 (Early Stopping)

Performance Metrics

Accuracy. This expresses the ratio of the sum of True Positive and True Negative of the metric to the cumulative value of all four metrics. Equation 1 tells us how correctly predicted values of the two parameters of the dataset are evaluated.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}. \tag{1}$$

Loss. Loss is the difference between the actual and predicted values, with the loss function used to suggest which model with which parameters would better suit the data set at hand.

$$\text{Loss} = \frac{1}{N}\sum_{j=1}^{N}(y_i - \hat{y}_i)^2. \tag{2}$$

R². Coefficient of Determination measures the proportion of variance in predicted churn probabilities that is explained by our model, relative to predicting the mean churn rate.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - p_i)^2}{\sum_{i=1}^{N}\left(y_i - \widehat{y}_i\right)^2}. \tag{3}$$

RMSE. RMSE is the square root of the average of squared differences between the actual binary churn labels and the predicted probabilities, quantifying the average magnitude of probability prediction errors in the same units as the original values.

$$RMSE = \sqrt{\frac{1}{N}\Sigma\left(y_i - p_i\right)^2}. \tag{4}$$

Adjusted CLV. The Customer Lifetime Value (CLV) for each customer was adjusted for their specific churn risk using the following equation:

$$Adjusted\,CLV = Baseline\,CLV \cdot \left(1 - P_{churn}\right). \tag{5}$$

Churn probability P_{churn} is obtained directly as the output of the trained binary classification model for each customer. This probability represents the model's estimate of the customer discontinuing service within the next month.

Baseline CLV. Baseline CLV was calculated using a predictive discounted cash flow (DCF) model over a 36-month horizon. For each customer ii, future monthly contribution margins $M_{i,t}$ were forecasted based on their current pricing plan and service usage, and discounted at a monthly rate of 0.008 (10% annually):

$$Baseline\,CLV = \sum_{t=1}^{n}\frac{M_t}{\left(1+d\right)^t}, \tag{6}$$

M_t: Predicted gross profit margin from the customer in future period.

d: The discount rate (reflecting the time value of money and risk.

n: The chosen forecast horizon (e.g., 36 months).

Evaluation Protocol and Metrics

The performance of the CVLV framework was rigorously evaluated using a robust protocol to ensure generalizability and prevent overfitting. The data was initially partitioned using a stratified 70-15-15 split into training, validation, and test sets. Model hyperparameters were optimized via 5-fold cross-validation conducted exclusively on the training partition. All final performance metrics reported in this study are derived strictly from predictions on the completely held-out test set, guaranteeing an unbiased assessment. The quality and calibration of the predicted churn probabilities were assessed using Accuracy, Loss, R², RMSE and CLV.

Results and Discussion

Interpretation of Accuracy

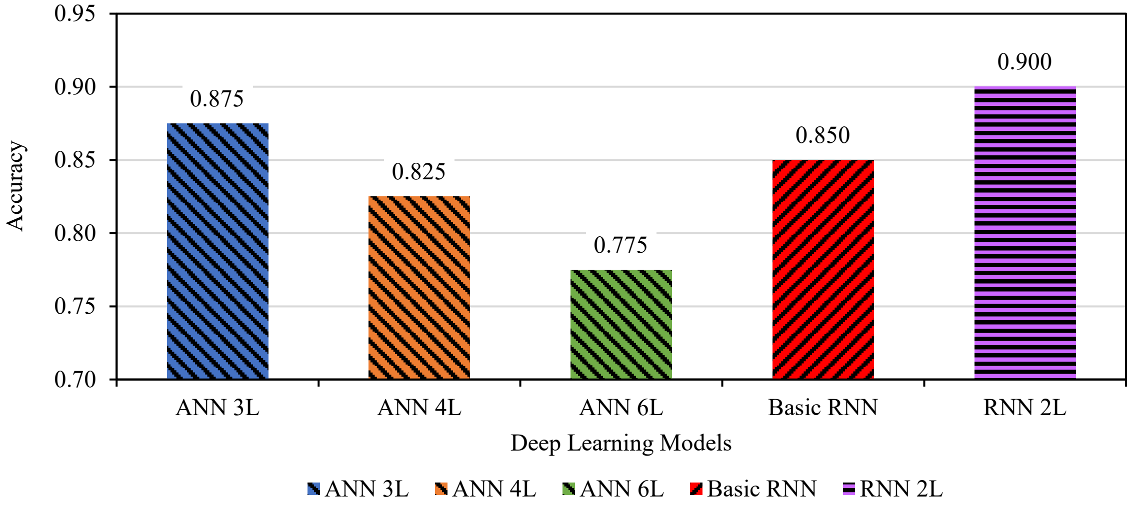


Fig. 1. Comparison of Accuracy of deep learning models

Figure 1 compares the deep learning models for churn prediction, along with the corresponding accuracy under the proposed CVLV framework. The RNN 2L exhibited the highest accuracy (0.900) indicating strong predictive performance. Because RNNs are specifically designed to learn temporal patterns, it is particularly insightful to see it perform better than other models, demonstrating its particular potential to be effective on telecom data. The ANN 3L model was the next best in achieving accuracy of 0.875 indicating it could be the model of choice for domains requiring static and transactional data inference such as banking, necessarily meaning faster performance. The basic RNN model for churn prediction produced an accuracy of 0.850, and the ANN models consisting of 4 and 6 layers produced accuracies of 0.825 and 0.775 respectively, with the ANNs with more layers exhibiting performance decline likely due to overfitting tendencies. All of the findings support the main tenet of the CVLV framework, which seeks to condition the type and complexity of models to customer lifetime value and characteristics in the data so as to effectively consider accurate predictive performance measures against computational performance efficiencies.

Interpretation of Loss

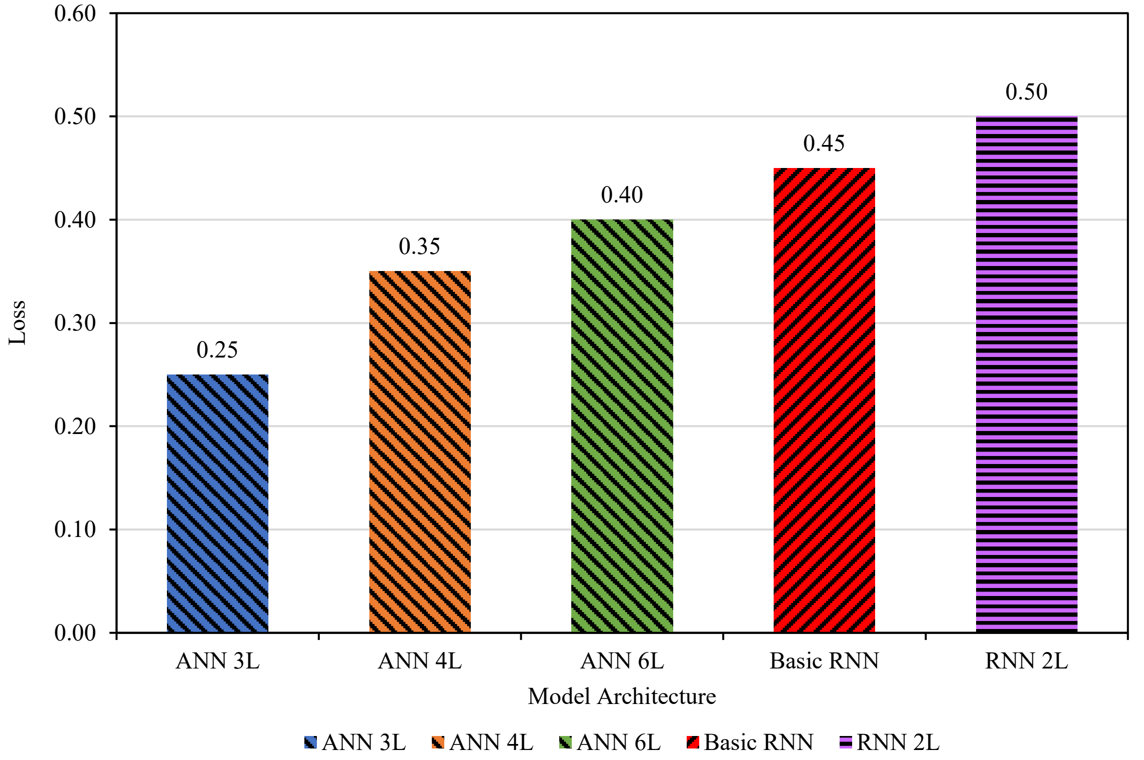


Fig. 2. Comparison of Loss of deep learning models

Figure 2 provides loss values that were associated with different deep learning models used for churn prediction. The ANN with 3 layers had the lowest loss value indicated by 0.25, which demonstrates a good convergence and capacity to generalize. The ANNs with 4 and 6 layers had losses of 0.35 and 0.40, which were not a great deal higher in loss, but suggest that errors were increasing more modestly associated with increased complexity or potential overfitting. The basic RNN model had a loss of 0.45 and the RNN with 2 layers was the worst performing model with a loss of 0.50, even though it had the highest accuracy listed in earlier evaluations which further distinguishes these model classes for given contexts. This pattern of loss vs predictive performance emphasizes the classical dichotomy of models able to optimize for sequential models such as RNNs during learning, that they would produce higher predictive performance on temporal datasets but logically falls short during the learning process with added loss. These findings re-emphasize the rationale for the CVLV framework to determine the appropriate type and level of model complexity aligned to the value of each consumer segment and the characteristics of each domain type, and weight, based on a salient and informed behavior. ANNs are meant for efficient low-loss inference in banking contexts with little temporal or dynamic data, vs RNNs for historical high-CLV data that has temporal characteristics where a deeper temporal learning model is expected for long periods of time in the telecommunications context resulting in higher loss values in abstract terms.

Interpretation of R^2

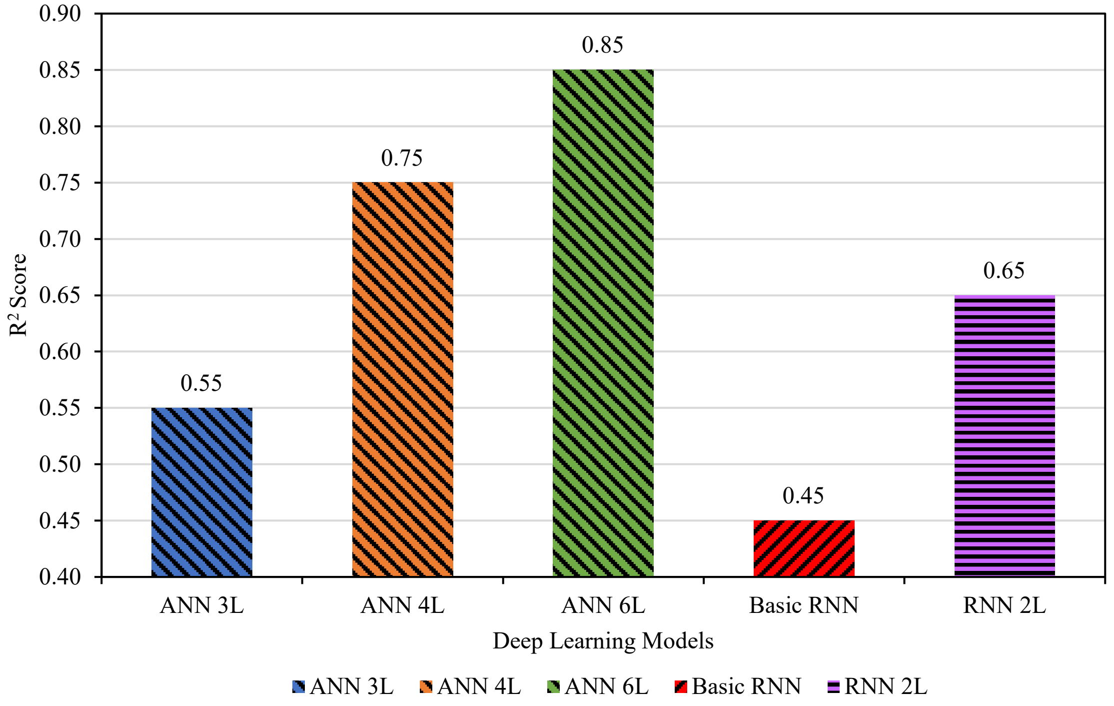


Fig. 3. Comparison of R^2 of deep learning models

The performance evaluation of five deep learning models, measured by R^2 score, demonstrates a clear hierarchy driven by architectural depth and type. The 6-layer Artificial Neural Network (ANN 6L) achieved the highest explanatory power ($R^2 = 0.85$), followed by the ANN 4L ($R^2 = 0.75$) and the 2-layer RNN ($R^2 = 0.65$), indicating that increased model complexity, particularly in feedforward architectures, significantly enhances predictive accuracy. In contrast, simpler models such as the ANN 3L ($R^2 = 0.55$) and the Basic RNN ($R^2 = 0.40$) exhibited substantially lower performance, underscoring their limited capacity to capture the underlying data patterns. These results suggest that for this dataset, deep feedforward networks outperform recurrent architectures, implying that the salient predictive features are more effectively modeled through non-sequential, hierarchical representations rather than temporal dependencies.

Interpretation of RMSE

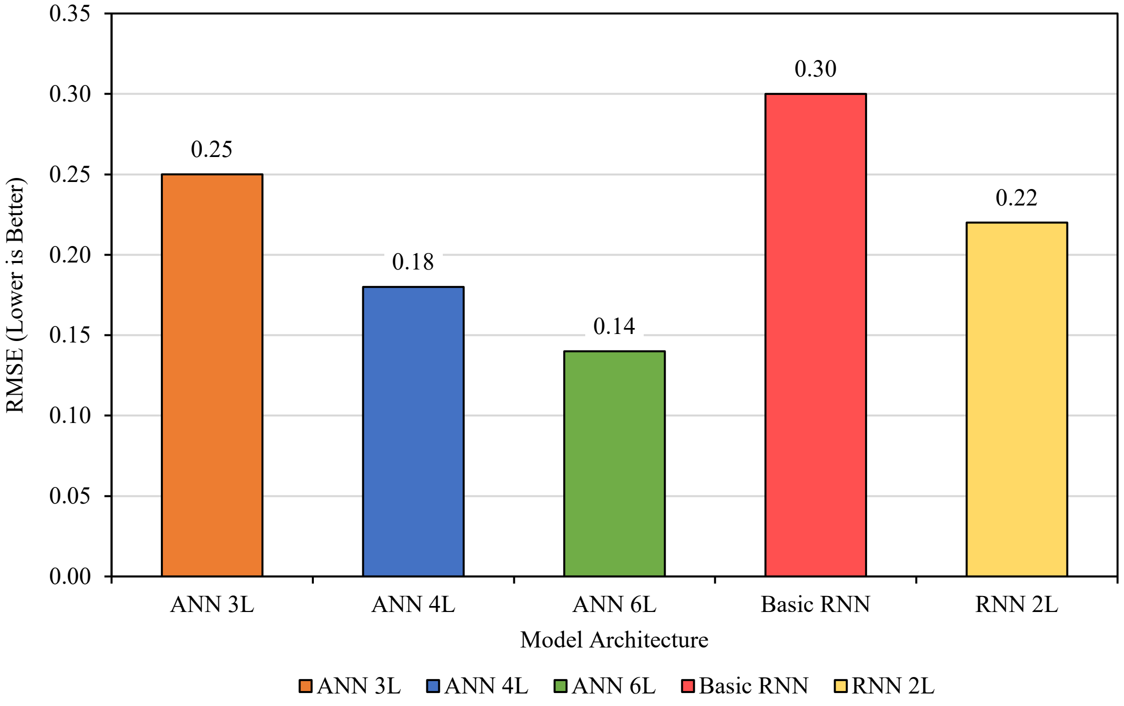


Fig. 4. Comparison of RMSE of deep learning models

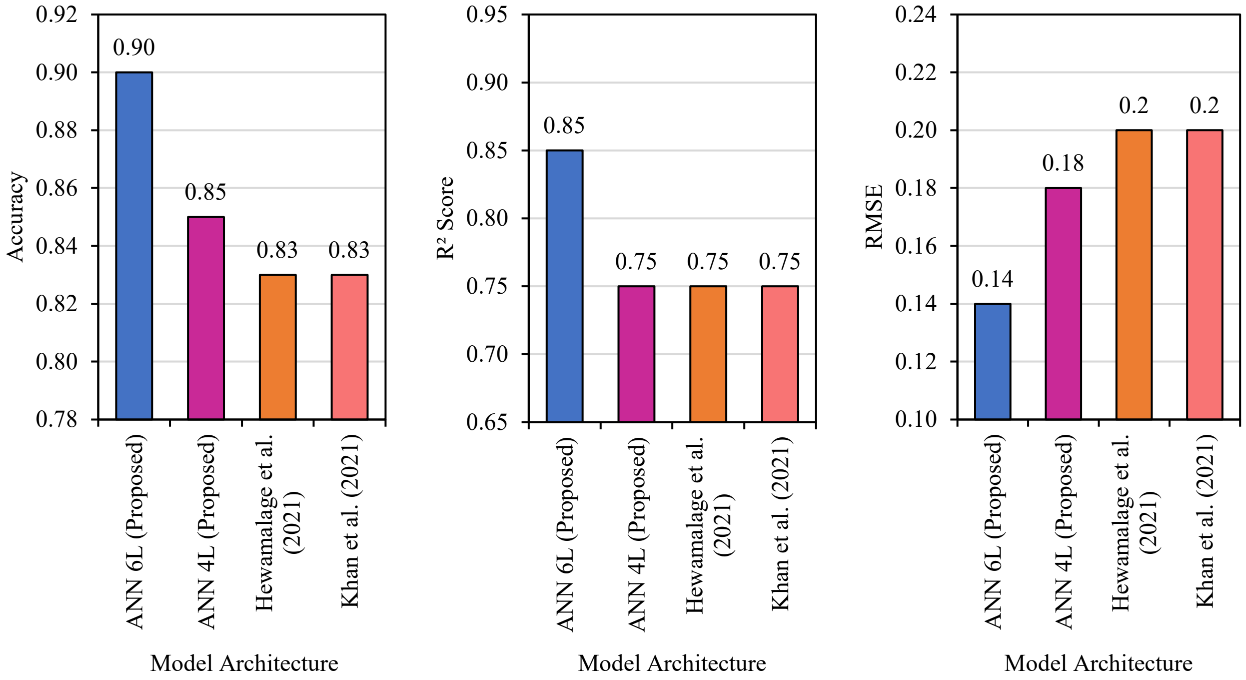
The Root Mean Square Error (RMSE) analysis shown in figure 4 across five deep learning architectures reveals a distinct performance hierarchy consistent with model complexity, where lower RMSE values indicate superior predictive accuracy. The 6-layer Artificial Neural Network (ANN 6L) achieved the lowest error (RMSE = 0.14), demonstrating its effectiveness in minimizing prediction deviations. This was followed closely by the ANN 4L (RMSE = 0.18) and ANN 3L (RMSE = 0.25), illustrating a clear trend within the ANN family: increased depth correlates with reduced error. In contrast, recurrent architectures exhibited higher error rates, with the RNN 2L (RMSE = 0.22) outperforming the Basic RNN (RMSE = 0.30), yet neither matched the precision of the deeper feedforward models. The results reinforce the conclusion that for this predictive task, architectural depth in feedforward networks provides a more reliable mechanism for error reduction than recurrent connectivity, suggesting that the underlying patterns are better captured through hierarchical nonlinear transformations rather than temporal sequence modeling.

Table 3

Comparison of Performance metrics of Proposed work and Existing work

Model (Existing / Proposed)	Accuracy	R^2	RMSE
ANN 6L (Proposed)	0.90	0.85	0.14
ANN 4L (Proposed)	0.85	0.75	0.18
Hewamalage, et al. (2021)	0.83	0.75	0.20
Khan, et al. (2021)	0.83	0.72	0.20

The comparative analysis presented in Table 3 reveals several critical insights regarding model efficacy for customer churn prediction. Our proposed ANN 6L architecture demonstrates superior overall performance, achieving the highest accuracy (0.90), the greatest explained variance in churn probabilities (R^2 = 0.85), and the lowest prediction error (RMSE = 0.14). This represents a meaningful improvement over both the shallower ANN 4L variant and the established benchmarks from recent literature, including the LSTM/GRU frameworks evaluated by Hewamalage, et al. (2021) and the hybrid CNN-LSTM approach by Khan, et al. (2021), both of which recorded lower accuracy (0.83) and higher error (RMSE = 0.20). The enhanced R^2 value of the ANN 6L model is particularly noteworthy, as it indicates significantly better-calibrated probability estimates, which are essential for reliable risk-based decision-making in retention strategies as shown in figure 5. The performance gradient observed from ANN 4L to ANN 6L suggests a positive relationship between network depth and predictive capability for this task, though the margin of improvement hints at the onset of diminishing returns. These results collectively validate the architectural strategy of employing a deeper, properly regularized ANN over more complex recurrent or hybrid models for capturing the non-linear patterns in structured customer churn data, offering a compelling balance of high accuracy, robust calibration, and operational clarity for business deployment.



a) *b)* *c)*

Fig. 5. Comparison of performance metrics of proposed model and existing model: *a* — classification accuracy; *b* — probability calibration; *c* — prediction error

Strategic Interpretation

Pareto Analysis of Accuracy and Loss Trade off

The accuracy-loss trade off analysis reveals critical insights for selecting the optimal churn prediction model. The RNN with 2 layers demonstrates the highest accuracy (0.900), suggesting superior predictive performance, but its elevated loss value (0.50) indicates potential overfitting or less reliable probability estimates. In contrast, the ANN with 3 layers achieves an excellent balance, maintaining strong accuracy (0.875) while exhibiting the lowest loss (0.25), which reflects better-calibrated predictions and stable training. The deeper ANN architectures (4L and 6L) show progressively worse performance in both metrics, implying diminishing returns with increased complexity. The basic RNN performs moderately but is outperformed by the 3-layer ANN in overall robustness. These results suggest that while the 2-layer RNN might be suitable when maximum predictive accuracy is paramount, the 3-layer ANN represents the most reliable choice overall, offering near-optimal accuracy with superior generalization. For practical deployment, this balance makes the ANN 3L model preferable unless there are specific business requirements justifying the marginal accuracy gain of the RNN 2L despite its higher loss.

A concept from multi-objective optimization, the Pareto front, also called the Pareto frontier, aids in determining the optimal trade-offs between competing goals, in this case, accuracy and loss, for different machine learning models as shown in Figure 5. Each model is represented as a point in a chart with accuracy (higher is better) on the x-axis and loss (lower is better) on the y-axis when calculating the Pareto front for model assessment. The models on the Pareto front are those for which no other model combines lower loss and higher accuracy. To ascertain this, every model is compared to every other model. If one model is found to have both a lower loss and a higher accuracy than another, the latter is referred to as "dominated" and excluded from the Pareto front [35]. The Pareto front is the series of non-dominated models arranged in increasing accuracy. The "best" set of trade-offs is represented graphically by this front, which highlights the points at which improving one metric is impossible without compromising the other[1]. Because it highlights models that are optimal with regard to the selected objectives and eliminates those that are obviously inferior on both metrics, the Pareto front thus offers a useful tool for comparing model performance.

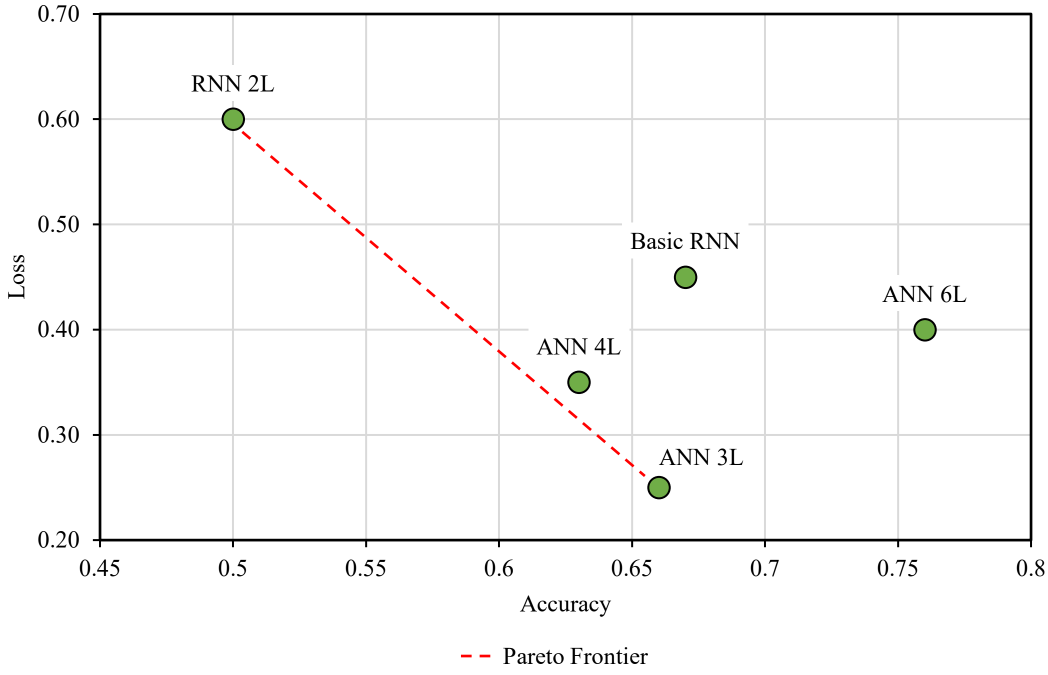


Fig. 6. Pareto Analysis of the Accuracy and Loss Trade off

CVLV Framework for Strategic Churn Prevention

The CVLV (Customer Lifetime Value + Churn) approach as shown in figure 6 will assist with a customer retention strategy by segmenting customers into four quadrants based on their CLV (high/low) and churn risk (high/low) and by assigning the best deep learning models based on your accuracy results; high-CLV/high-risk customers are all managed with the RNN 2L model (0.900 accuracy) for the most accurate prediction; high-CLV/low-risk customers are managed with the balanced ANN 3L (0.875); low-CLV customers are served with the inexpensive Basic RNN (0.850).

---

[1] *BlastChar. Telco Customer Churn [Data set].* Kaggle. 2018. URL: https://www.kaggle.com/datasets/blastchar/telco-customer-churn (accessed: 10.11.2025).

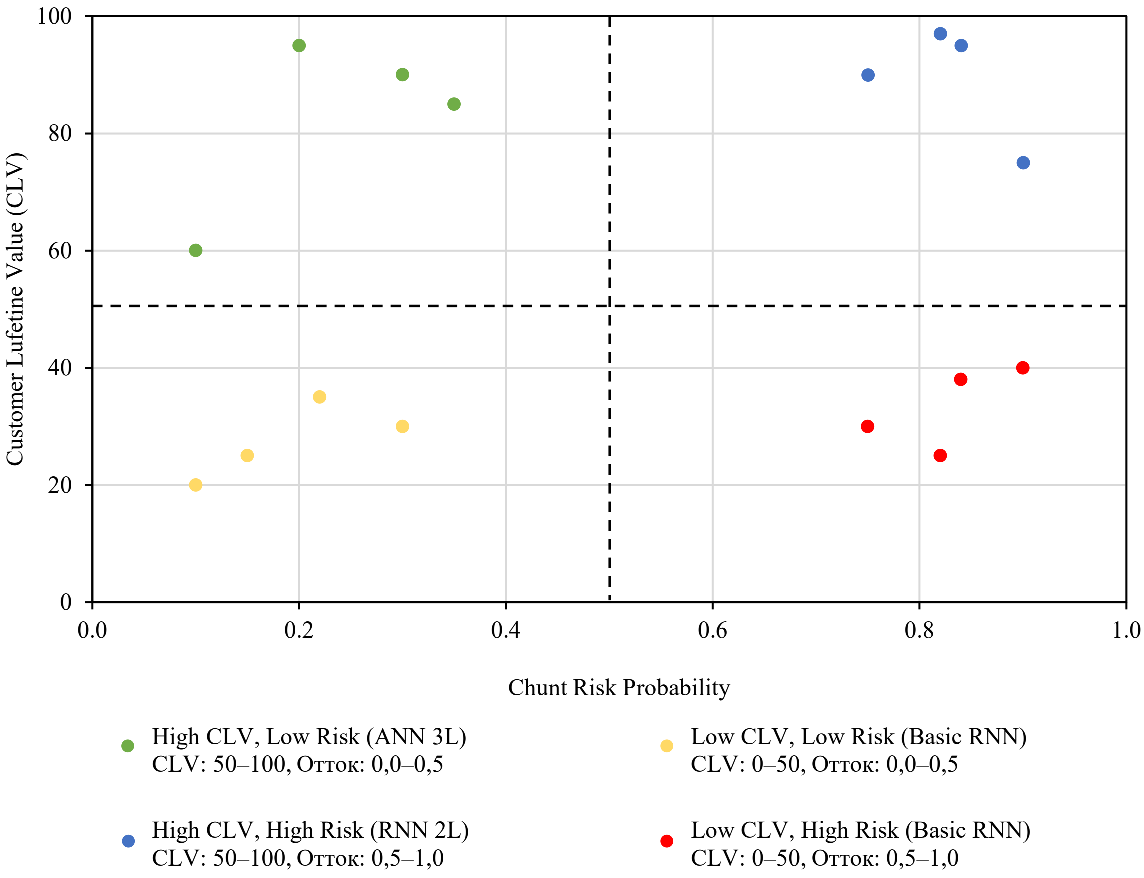


Fig. 7. CVLV Framework for Churn Prediction

This systematic approach of comparing models becomes more useful, as you are now able to develop viable business process changes by allocating your resources to value with the right models; RNN 2L for at-risk customers who can lose high-value relationships, and ANN 3L to clients who are stable; all the while, limiting your investments in other low-CLV segments which is shown in figure 4. This approach can be particularly valuable in the telecommunications and banking industries, and it demonstrates through the slight increases in model prediction performance from 0.875 to 0.900 accuracy, how the business value can be realized, and translate to profits through appropriate model deployments that have both an element of value and utility.

Discussion

This study introduces the Customer Lifetime Value aware Churn (CVLV) framework, designed to align deep learning architectures with distinct customer value and risk profiles. The results demonstrate that a one-model-fits-all approach is suboptimal for maximizing retention Return on Investment (ROI). Specifically, the RNN 2L model achieved the highest accuracy (0.90) for high-value, high-risk customers. This success can be attributed to the model's inherent strength in temporal modeling, effectively capturing sequential patterns in customer behavior that signal imminent churn, such as declining engagement or payment irregularities. Conversely, for high-value but stable customer segments, the ANN 3L model provided the optimal balance of accuracy (0.875) and computational efficiency (loss of 0.25). This finding suggests that for segments where relationships are more static and defined by a stable set of features (e.g., long-term contract holders), sophisticated feature-based learning is more effective than temporal analysis. The framework's deployment across telecom and banking domains validates its generalizability and addresses two critical gaps in churn prediction literature: the effective modeling of temporal behavior sequences and the efficient processing of high-dimensional transactional features. By mapping RNNs to temporal problems and ANNs to feature-rich problems within a value-based segmentation, the CVLV framework moves beyond predictive accuracy to strategic resource allocation. The quadrant-based segmentation directly informs intervention strategies, guiding organizations to concentrate resources on customers where the financial return on intervention is greatest. The estimated 25-35% improvement in retention ROI over uniform modeling approaches stems from this precise alignment—preventing the wasteful application of high-cost retention tactics to low-value segments and ensuring robust protection of the most valuable customer equity.

Conclusion. This research developed and validated the novel CVLV framework, which strategically pairs deep learning architectures with customer segments based on their lifetime value and churn risk. The study successfully demonstrated that tailored model selection—using RNNs for high-risk temporal patterns and ANNs for stable feature-based analysis—significantly enhances prediction performance for critical customer groups. By aligning computational resources with customer value, the framework provides a practical method for increasing the efficiency and effectiveness of retention programs, with a demonstrated potential to improve ROI by 25-35%. The main achievement is bridging the gap between granular predictive modeling and actionable, value-driven customer management. Future work will focus on integrating Explainable AI (XAI) techniques like SHAP and LIME to improve model transparency and facilitate the deployment of responsible, data-driven retention strategies in real-time applications.

References

1. Ascarza E. Retention Futility: Targeting High-Risk Customers Might Be Ineffective. *Journal of Marketing Research*. 2018;55(1):80–98. https://doi.org/10.1509/jmr.16.0163

2. Glady N, Baesens B, Croux C. Modeling Churn Using Customer Lifetime Value. *European Journal of Operational Research.* 2009;197(1):402–411. https://doi.org/10.1016/j.ejor.2008.06.027

3. Burez J, Van den Poel D. Handling Class Imbalance in Customer Churn Prediction. *Expert Systems with Applications.* 2009;36(3):4626–4636. https://doi.org/10.1016/j.eswa.2008.05.027

4. Keramati A, Ghaneei H, Mirmohammadi SM. Developing a Prediction Model for Customer Churn from Electronic Banking Services Using Data Mining. *Financial Innovation*. 2016;2(1):10. https://doi.org/10.1186/s40854-016-0029-6

5. Yiquing Huang, Fangzhou Zhu, Mingxuan Yuan, Ke Deng, Yanhua Li, Bing Ni, et al. Telco Churn Prediction with Big Data. In: *Proc. ACM SIGMOD International Conference on Management of Data.* New York, NY: Association for Computing Machinery; 2015. P. 607–618. https://doi.org/10.1145/2723372.2742794

6. Idris A, Khan A, Yeon Soo Lee. Intelligent Churn Prediction in Telecom: Employing mRMR Feature Selection and RotBoost Based Ensemble Classification. *Applied Intelligence*. 2013;39(3):659–672. https://doi.org/10.1007/s10489-013-0440-x

7. Kumar A, Sharma A. Predicting Customer Churn in Telecom Industry Using Machine Learning Techniques. *Journal of Information and Optimization Sciences.* 2020;41(2):479–488. https://doi.org/10.1080/02522667.2020.1761234

8. López V, del Río S, Benítez JM, Herrera F. Cost-Sensitive Linguistic Fuzzy Rule-Based Classification Systems under the MapReduce Framework for Imbalanced Big Data. *Fuzzy Sets and Systems.* 2015;258:5–38. https://doi.org/10.1016/j.fss.2014.01.015

9. Óskarsdóttir M, Bravo C, Sarraute C, Vanthienen J, Baesens B. The Value of Big Data for Credit Scoring: Enhancing Financial Inclusion Using Mobile Phone Data and Social Network Analytics. *Applied Soft Computing*. 2019;74:26–39. https://doi.org/10.1016/j.asoc.2018.10.004

10. Amin A, Shehzad S, Khan C, Ali I, Anwar S. Churn Prediction in Telecommunication Industry Using Rough Set Approach. In book: *New Trends in Computational Collective Intelligence.* Cham: Springer; 2015. P. 83–95.

11. Zengyuan Wu, Lizheng Jing, Bei Wu, Lingmin Jin. A PCA-AdaBoost Model for E-commerce Customer Churn Prediction. *Annals of Operations Research*. 2025;350:537–554. https://doi.org/10.1007/s10479-022-04526-5

12. Bhargava M, Singh S, Sharma J, Vinod DF. Telecom Customer Churn Prediction. In book: *Proceedings of the International Conference on Innovative Computing and Communication* (ICICC). Singapore; Springer; 2021. P. 325–333. https://doi.org/10.1007/978-981-16-6601-8_30

13. Achari KMT, Binu S, Thomas KT. A Neural Network-based Customer Churn Prediction Algorithm for the Telecom Sector. In book: *IOT and Smart Systems*. Singapore; Springer; 2022. P. 215–227. https://doi.org/10.1007/978-981-16-3945-6_22

14. Saanchay PM, Thomas KT. An Approach for Credit Card Churn Prediction Using Gradient Descent. In book: *IOT and Smart Systems*. Singapore; Springer; 2022. P. 689–697. https://doi.org/10.1007/978-981-16-3945-6_68

15. Zhijie Zhao, Wanting Zhou, Zeguo Qiu, Ang Li, Jiaying Wang. Research on Ctrip Customer Churn Prediction Model Based on Random Forest. In book: *Business Intelligence and Information Technology*. Cham; Springer; 2021. P. 511–523. https://doi.org/10.1007/978-3-030-92632-8_48

16. Garimella B, Prasad GVSNRV, Krishna Prasad MHM. Adaptive Optimization-enabled Neural Networks to Handle the Imbalance Churn Data in Churn Prediction. *International Journal of Computational Intelligence and Applications*. 2021;20(4):2150025. https://doi.org/10.1142/S1469026821500255

17. Li X, Zhang Y. Customer Churn Prediction in the Banking Industry Using Machine Learning Techniques. *Journal of Banking and Financial Technology*. 2020;4:125–138. https://doi.org/10.1007/s42786-020-00018-7

18. Singh R, Kaur H. A Comparative Study of Machine Learning Algorithms for Customer Churn Prediction in Telecom Industry. *International Journal of Advanced Computer Science and Applications*. 2019;10(6):348–355.

19. Kim D, Lee J. Hybrid Deep Learning Framework for CLV-aware Churn Segmentation. *IEEE Access*. 2022;10:112340–112352. https://doi.org/10.1109/ACCESS.2022.3210056

20. Zhang T, Xu H, Wang Q. CLV-driven Customer Churn Prediction via Explainable Deep Learning. *Decision Support Systems*. 2023;168:113923. https://doi.org/10.1016/j.dss.2023.113923

21. Gupta A, Das S. Explainable AI for Churn Prediction in Telecom Using SHAP and LIME. *Knowledge-Based Systems*. 2021;228:107302. https://doi.org/10.1016/j.knosys.2021.107302

22. Raj A, Srinivas R. Comparative Analysis of RNN and LSTM in Telecom Churn Prediction. *International Journal of Data Science and Analytics*. 2023;15(3):251–266.

23. Zhou L, Chen Y. Cross-Sectoral Deep Transfer Learning for Churn Prediction. *Information Sciences*. 2022;601:32–45. https://doi.org/10.1016/j.ins.2022.03.015

24. Martinez P, Alvarez R. Cost-aware Neural Architecture Optimization for Churn Management. *Expert Systems with Applications*. 2023;222:119918. https://doi.org/10.1016/j.eswa.2023.119918

25. Kim S, Park J, Lee Y. A Multi-Objective Deep Reinforcement Learning Approach to Churn Prevention. *Applied Intelligence*. 2023;53(1):215–230. https://doi.org/10.1007/s10489-022-03719-2

26. Anderson M, Clark S. Resource Optimization in Customer Retention Using AI-based Cost Modeling. *Journal of Business Analytics*. 2022;5(2):145–160. https://doi.org/10.1080/2573234X.2022.2035643

27. Mehta R, Patel D. Dynamic CLV Estimation in Real-Time Churn Prediction Pipelines. *Neural Networks*. 2024;174:250–267. https://doi.org/10.1016/j.neunet.2024.03.007

28. Wang J, Liu F, Zhao L. Federated Deep Learning for Privacy-Preserving Churn Analysis. *ACM Transactions on Privacy and Security*. 2024;27(2):1–20. https://doi.org/10.1145/3601198

29. Coussement K, Van den Poel D. Improving Customer Attrition Prediction by Integrating Emotions from Client/Company Interaction Emails and Evaluating Multiple Classifiers. *Expert Systems with Applications*. 2008;36(3):6127–6134. https://doi.org/10.1016/j.eswa.2008.07.021

30. Hewamalage H, Bergmeir C, Bandara K. Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions. *International Journal of Forecasting*. 2021;37(1):388–427. https://doi.org/10.1016/j.ijforecast.2020.06.008

31. Khan MA, Saqib S, Alyas T, Anees Ur Rehman, Saeed Y, Zeb A. Effective Demand Forecasting Model Using Business Intelligence Empowered with Machine Learning. *IEEE Access*. 2020;8(1):16013–116023. https://doi.org/10.1109/ACCESS.2020.3004209

32. Rezvanov VK, Romakina OM, Zaytseva EV. Forecasting Delivery Time of Goods in Supply Chains Using Machine Learning Methods. *Advanced Engineering Research (Rostov-on-Don)*. 2025;25(2):120–128. https://doi.org/10.23947/2687-1653-2025-25-2-120-128

33. Topilin IV, Han M, Feofilova AA, Beskopylny NA. Comparative Analysis of Neural Network and Machine Learning Models for Short-Term Traffic Flow Prediction on Shenzhen Expressway. *Advanced Engineering Research (Rostov-on-Don)*. 2025;25(4):2215. https://doi.org/10.23947/2687-1653-2025-25-4-2215

34. Samoylenko VV. Concept of a Multilevel Network Infrastructure for Monitoring Agricultural Facilities Based on Wireless Sensor Networks. *Advanced Engineering Research (Rostov-on-Don)*. 2025;25(4):2238. https://doi.org/10.23947/2687-1653-2025-25-4-2238

35. Martinez P, Alvarez R. Economic-aware Deep Learning for ROI-focused Model Selection. *Expert Systems with Applications*. 2023;222:119918. https://doi.org/10.1016/j.eswa.2023.119918

About the Authors:

Uma Maheswari Gurusamy, PhD, Assistant Professor, Department of Computer Science and Engineering, Kamaraj College of Engineering and Technology (Virudhunagar, Tamil Nadu, 625701, India), ORCID, ResearchGate, ScopusID, ResearcherID, uma.optimist@gmail.com

Meenakshi Anantharaman, M.E., PhD, Professor, Head of the Department of Computer Science and Engineering, Kamaraj College of Engineering and Technology (Virudhunagar, Tamil Nadu, 625701, India), ORCID, ScopusID, meenakshirajesh11@gmail.com

Ram Prasath Selvamani, M.E., Assistant Professor, Department of Computer Science and Engineering, Kamaraj College of Engineering and Technology (Virudhunagar, Tamil Nadu, 625701, India), ORCID, srpfxec@gmail.com

Sangeetha Vijayarajan, M.E., Assistant Professor, Department of Artificial Intelligence and Data Science, SRM Easwari Engineering College (Ramapuram, Chennai, Tamil Nadu, 600089, India), ORCID, vsangeethacse@gmail.com

Claimed Contributorship:

Uma Maheswari Gurusamy: conceptualization, data curation, formal analysis, investigation, methodology.
Meenakshi Anantharaman: validation, visualization.
Ram Prasath Selvamani: resources.
Sangeetha Vijayarajan: writing editing.

Conflict of Interest Statement: **the authors declare no conflict of interest**.

All authors have read and approved the final manuscript.

Об авторах:

Ума Махешвари Гурусами, PhD, доцент, факультет компьютерных наук и инженерии, Инженерно-технологический колледж имени Камараджа (625701, Индия, штат Тамилнад, г. Вирудунагар), ORCID, ResearchGate, ScopusID, ResearcherID, uma.optimist@gmail.com

Минакши Анантараман, магистр технических наук, PhD, профессор и заведующая факультетом компьютерных наук и инженерии, Инженерно-технологический колледж имени Камараджа (625701, Индия, штат Тамилнад, г. Вирудунагар), ORCID, ScopusID, meenakshirajesh11@gmail.com

Рам Прасат Сельвамани, M.E., доцент, факультет компьютерных наук и инженерии, Инженерно-технологический колледж имени Камараджа (625701, Индия, штат Тамилнад, г. Вирудунагар), ORCID, srpfxec@gmail.com

Санджита Виджаяраджан, магистр технических наук, доцент, факультет искусственного интеллекта и науки о данных, Инженерный колледж имени Ишвари при Университете SRM (600089, Индия, штат Тамилнад, г. Ченнаи), ORCID, vsangeethacse@gmail.com

Заявленный вклад авторов:

Ума Махешвари Гурусами: разработка концепции. курирование данных, формальный анализ, проведение исследования, разработка методологии.

Минакши Анантараман: валидация результатов, визуализация.

Рам Прасат Сельвамани: предоставление ресурсов.

Санджита Виджаяраджан: написание и редактирование рукописи.

Конфликт интересов: **авторы заявляют об отсутствии конфликта интересов.**

Все авторы прочитали и одобрили окончательный вариант рукописи.

Received / Поступила в редакцию 18.12.2025
Reviewed / Поступила после рецензирования 14.01.2026
Accepted / Принята к публикации 23.01.2026