IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.153121

Improved Customer Churn Estimation Using LSTM Networks

Dr. Sajja Suneel[1], Lokesh Gopal M[2], Harshitha P[3], Harshitha Reddy K[4]

Dept. of CSE (Data Science) Institute of Aeronautical Engineering Dundigal, Hyderabad[1-4]

Abstract: In the competitive landscape of financial services, the high cost of client acquisition makes retention a top strategic priority. This study addresses the limitations of conventional credit card churn prediction specifically rigid architectural constraints and inefficient categorical data processing by proposing a dynamic framework based on Long Short-Term Memory (LSTM) networks. By restructuring standard tabular datasets into”pseudo-sequences” and utilizing dense embeddings for high- cardinality features, we enable the model to capture nuanced temporal shifts in customer behavior. Our methodology evaluates four distinct LSTM configurations: Vanilla, Stacked, Bidirectional, and a hybrid Bidirectional-Stacked variant. By integrating these architectures into a unified ensemble, we achieved a peak classification accuracy of 92.35%. Beyond raw accuracy, the ensemble demonstrates exceptional recall performance. For banking institutions, this translates to a more reliable early- warning system that identifies at-risk accounts with precision, effectively reducing the revenue loss associated with undetected churn.

Index Terms: Customer attrition modeling, LSTM architectures, Deep learning frameworks, Ensemble modeling strategies, Credit risk analytics, Class imbalance handling

I. INTRODUCTION

N today’s data heavy economy, keeping customers is a top- tier strategy for banks, tech firms, and any business built on subscriptions. “Customer churn” the moment a client decides to walk away is a major blow to a company’s bottom line and long-term health. Since winning over a new customer is far more expensive than keeping an existing one happy, businesses need smart, predictive tools to spot at-risk clients before they leave.

A. The Drive Behind Customer Retention

The initial phase of churn prediction research leaned heavily on foundational supervised learning algorithms, specifically linear regression and hierarchical decision trees. Although these techniques provided a baseline for classification, they proved inadequate for capturing the high-dimensional, stochastic nature of modern consumer patterns. Such conventional architectures are fundamentally limited by their inability to process temporal dependencies; by interpreting data as isolated, stationary observations, they overlook the evolving relationship between customer attributes that fluctuate within volatile financial markets.

B. Bridging the Gap: Our Contribution

The rapid advancement of deep learning has brought significant attention to Long Short-Term Memory (LSTM) networks. These models are specifically engineered to capture and retain patterns over time through gated memory mechanisms. How- ever, much of the existing research remains incomplete. Many studies rely on a single architectural configuration, overlook the substantial business impact of false negatives (failing to identify customers who are likely to churn), or employ inefficient approaches for encoding categorical variables, thereby limiting overall model effectiveness.

Our research addresses these flaws by introducing a robust, bias-controlled LSTM framework tailored for credit card churn. We’ve built a specialized pipeline that turns standard tabular data into “pseudo-sequences,” allowing the model to see deep connections between features that others miss.

The core breakthroughs of this paper include:

- **A Detailed Model Showdown:** We systematically com- pare Vanilla, Stacked, and Bidirectional LSTM variants to determine the most effective architecture.
- **Smart Data Encoding:** We replaced bulky, inefficient data labels with sleek, low-dimensional “embeddings” to help the model learn more accurately from structured information.
- **Fair and Honest Testing:** By using stratified cross- validation and ensemble techniques, we ensure results are consistently reliable and unbiased.
- **Business-Centric Focus:** We prioritize Recall to minimize the high cost of failing to identify departing customers.

II. LITERATURE SURVEY

A. The Evolution of Churn Prediction

Predicting customer churn has become a cornerstone of research in banking, telecom, and e-commerce because of its massive impact on a company's revenue and retention strategies [1], [13]. In the early days, researchers leaned heavily on traditional machine learning tools such as logistic regression, decision trees, and support vector machines [6], [7], [8]. While these models provided a solid starting point, they frequently struggled to map out the complex, non-linear ways that real-life customers actually behave [13].

B. The Shift to Deep Learning

As technological capabilities progressed, research attention increasingly moved toward deep learning approaches, particularly Long Short-Term Memory (LSTM) models. These networks are highly valued for their capacity to capture tem- poral relationships, enabling them to model how various data elements influence one another over time [9], [14]. A major milestone in this area was the work by Khattak et al. [1], who blended Bidirectional LSTM (BiLSTM) with Convolutional Neural Networks (CNN). Their hybrid model was better at catching long-term patterns than old-school methods, but it had its limits: the study only looked at one specific architecture and struggled with low "recall" scores meaning it still missed too many customers who were actually planning to leave.

Other researchers tried to level up by adding "attention mechanisms" and regularization techniques like dropout to prevent the models from simply memorizing the training data [2], [3], [5], [11]. These updates made models more robust and easier to interpret [3], [5], [11], yet many of these studies still suffered from a "one-size-fits-all" approach, failing to compare different types of architectures systematically.

C. Industry-Specific Applications

In the world of finance, early-warning systems using tools like CatBoost provided some practical help [4], [6]. However, these often viewed data as a static snapshot rather than a moving story [4], [6]. Even when researchers used advanced optimization for SVMs, they generally couldn't keep up with the deep learning models when it came to capturing complex behavioral dynamics [7], [8]. Similarly, in telecom, treating customer interactions as a sequence of events allowed LSTMs to consistently beat out static classifiers [9], [10]. While ensemble methods have recently improved overall accuracy, they often overlook the "false negative" problem the high- stakes error of failing to spot a departing customer [11], [14].

D. Identifying the Missing Pieces

Despite all this progress, several gaps still exist in the current literature:

- **Over-reliance on Single Models:** Most studies fixate on one hybrid setup without testing how different versions of that model perform [1], [13].
- **Weak Validation:** There is a lack of rigorous, bias- controlled testing like stratified cross-validation.
- **The "Sparsity" Problem:** Many researchers use simple data encoding that makes the information too "thin" for the model to learn effectively [2], [6].
- **Business Blind Spots:** Most papers focus on general accuracy rather than minimizing the specific errors that cost businesses the most money.

E. How This Work Moves Forward

Our research addresses these gaps by building a comprehen- sive framework that puts four distinct LSTM variants Vanilla, Stacked, Bidirectional, and Bidirectional Stacked to the test under the same fair conditions. Unlike the hybrid BiLSTM CNN models seen previously [1], we prioritize a deep dive into how different structures affect performance. We also swap out thin data encoding for dense "embeddings" and use ensemble learning to ensure our results are as reliable as possible. By pushing for significantly higher recall and AUC than previous studies, we aim to provide a tool that doesn't just work in a lab, but actually helps banks save their most valuable customer relationships.

III. METHODOLOGY

Our methodology is designed to tackle specific hurdles in churn prediction such as data sparsity, class imbalance, and the justification for using recurrent models on tabular datasets. We propose a framework centered on representation learning and architectural sensitivity to enhance predictive performance for financial institutions.

A. Data Understanding and Preparation

The study utilizes a credit card dataset encompassing demo- graphics and transactional behavior. Formulated as a binary classification task, the primary challenge is significant class imbalance. To address this, we utilized exploratory

DOI: 10.17148/IJARCCE.2026.153121

analysis via histograms and boxplots to inform a bias-controlled evaluation strategy.

B. Smart Prep: Embeddings and Scaling

Rather than utilizing sparse, high-cardinality vectors which frequently arise from traditional binary encoding this frame- work maps categorical variables, such as account types and educational backgrounds, into a continuous, low-dimensional embedding space. This transition enables the architecture to identify and weigh underlying correlations between discrete classes. To maintain numerical stability, continuous features undergo a statistical transformation to achieve unit variance and a zero-centered distribution. Implementing this normal- ization protocol prevents variables with naturally broader ranges from disproportionately influencing the weight updates, thereby ensuring a balanced and efficient gradient descent.

C. The Framework Architecture

The end-to-end pipeline from feature transformation to the final ensemble prediction is illustrated in Fig. 1.

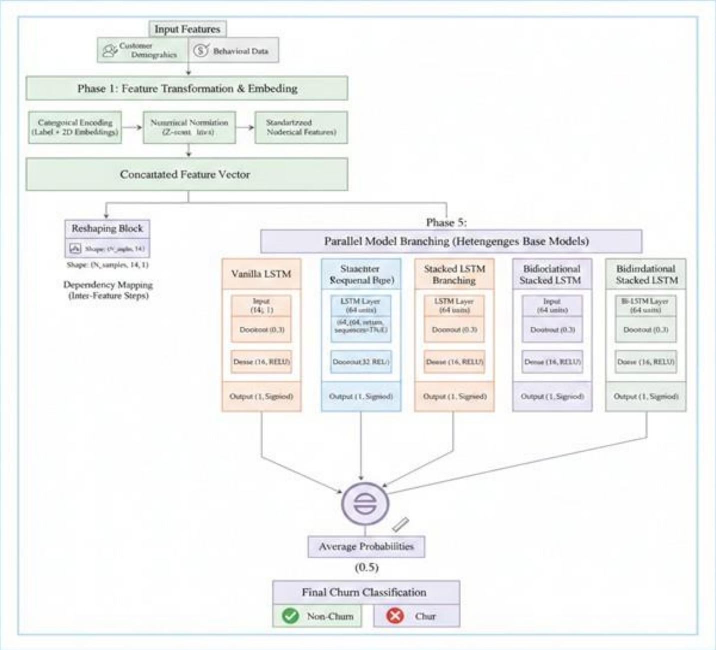

Fig. 1. Proposed End-to-End LSTM Churn Prediction Framework: Feature transformation, pseudo-sequential reshaping, parallel model branching, and ensemble averaging.

D. Pseudo-Sequential Modeling

The governing dynamics of the LSTM units within our architecture are characterized by a series of specialized gating mechanisms. To manage memory retention, the forget gate f_t utilizes a logistic sigmoid activation to determine which components of the preceding hidden state h_{t-1} and current input x_t should be discarded:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \qquad (1)$$

Concurrently, the input gate i_t modulates the volume of fresh information to be assimilated into the unit's internal memory. This works alongside a candidate cell state generated via a hyperbolic tangent function:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \qquad (2)$$

The internal memory, or cell state C_t, undergoes a refine- ment process where historical data is pruned by the forget gate and supplemented with the modulated input candidates:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

In the final stage, the hidden representation h_t is derived. This is achieved by passing the current cell state through a non-linear activation and scaling it by the output gate o_t, ensuring only the relevant features are propagated to the subsequent layer:

$$h_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \odot \tanh(C_t) \qquad (4)$$

E. Model Variations and Testing

We systematically compare four architectural variants:

- **Vanilla LSTM:** The baseline single-layer model.
- **Stacked LSTM:** Incorporates multiple layers for hierarchical learning.
- **Bidirectional LSTM:** Analyzes the structured feature sequence from both forward and reverse directions, enabling richer contextual understanding of inter-feature relationships
- **Bidirectional Stacked LSTM:** Combines depth with two-way contextual analysis.

To mitigate overfitting, Dropout layers (0.3) and early stopping are employed. Evaluation is performed using stratified five- fold cross-validation to ensure results are stable and unbiased.

F. Summary of the Technical Pipeline

The systematic approach of the pipeline is summarized in Table I.

TABLE I
SUMMARY OF THE TECHNICAL PIPELINE

Phase	Purpose	Key Technique
Transformation	Fix sparse data	2D Dense Embeddings
Reshaping	Prepare for LSTM	Pseudo-sequential Tensors
Optimization	Efficient learning	Adam & Early Stopping
Validation	Ensure reliability	Stratified 5-Fold CV

G. Business Operational Relevance

The framework prioritizes Recall because the operational cost of a false negative (missing a churner) far exceeds the cost of a false positive in the banking sector. Training is driven by Binary Cross-Entropy loss to specifically minimize these high-stakes errors.

IV. EXPERIMENTAL ANALYSIS AND PERFORMANCE EVALUATION

The performance of the developed LSTM-based predictive architecture was rigorously validated to ensure both generaliz- ability and statistical reliability. To mitigate bias and verify the framework's efficacy on novel observations, we implemented a stratified five-fold cross-validation protocol. This approach ensures that the class distribution of churned versus retained customers is preserved across each fold, providing a stable and consistent metric for evaluating the model's predictive precision across diverse subsets of the financial dataset.

A. Performance of Individual LSTM Architectures

The Vanilla LSTM serves as our baseline, achieving a solid accuracy of 89.29% and a high recall of 0.97. However, its lower Area Under the Curve (AUC) relative to deeper variants suggests limited class separation capability.

To improve representational capacity, the Stacked LSTM architecture was evaluated. By adding hierarchical layers to capture deeper patterns, accuracy improved to 90.97% and AUC increased to 0.9294. The Bidirectional LSTM further en- hanced performance by processing the pseudo-sequential fea- ture representation in both forward and backward directions, reaching 91.41% accuracy. This bidirectional context proved superior in modeling inter-feature dependencies. Among stan- dalone models, the Bidirectional Stacked LSTM emerged as the top performer, specifically excelling in minimizing false negatives.

B. The Power of the Ensemble LSTM

The most significant performance gains were realized through the Ensemble LSTM, which aggregates predictive probabilities from the Vanilla, Stacked, and Bidirectional ar- chitectures. As shown in the pipeline in Fig. 1, this collective approach reduces model-specific bias and enhances generaliza- tion. The ensemble model achieved peak performance across all metrics:

- **Accuracy:** 92.35%
- **Recall:** 0.98 (Demonstrating near-perfect churner identi- fication)
- **F1-Score:** 0.9554
- **AUC:** 0.9549

C. Architectural Sensitivity and Operational Impact

Experimental results indicate a clear trend: increasing model depth (Stacking) improves hierarchical abstraction, while Bidi- rectional processing enhances the contextual understanding of the feature sequence. By merging these into an ensemble, we stabilized these gains and nearly eliminated business-critical errors (false negatives).

The comparative performance is summarized in Table II.

TABLE II
PERFORMANCE COMPARISON OF EVALUATED MODELS

Model	Accuracy (%)	F1-Score	AUC	Recall
Vanilla LSTM	89.29	0.9381	0.8999	0.97
Stacked LSTM	90.97	0.9475	0.9294	0.97
Bidirectional LSTM	91.41	0.9498	0.9300	0.97
Ensemble LSTM	**92.35**	**0.9554**	**0.9549**	**0.98**

V. CONCLUSION

This study has developed a sophisticated, research-backed framework using LSTM networks to tackle the persistent challenge of customer churn in the credit card industry. By addressing critical gaps such as the handling of categorical data, biased testing methods, and architectural rigidity, this research provides a system that is both technically robust and practically applicable for financial institutions.

Our experiments demonstrate that increasing model depth and contextual awareness through Stacked and Bidirectional architectures significantly improves discriminative capability. The Ensemble LSTM emerged as the superior model, attaining a accuracy of 92.35% and an AUC of 0.9549. Most impor- tantly, achieving a recall of 0.98 confirms that an ensemble approach is the most effective way to minimize false negatives and ensure that at-risk customers are accurately identified.

REFERENCES

[1]. Khattak, M., et al., “Customer churn forecasting via composite deep architectures,” *Sci. Rep.*, vol. 13, no. 1, 2023, pp. 1–15.

[2]. Liu, Y., et al., “Hybrid neural network approaches for customer attrition modeling,” *Sci. Rep.*, vol. 14, no. 1, 2024, pp. 1–18.

[3]. Assiri, A., “Optimized churn prediction utilizing regularized Bi-LSTM frameworks,” *Comput. Mater. Contin.*, vol. 78, no. 2, pp. 1–19, 2025.

[4]. Zhang, Y., et al., “Advanced early-warning systems for credit card churn via GSAIBAS-CatBoost,” *Comput. Model. Eng. Sci.*, vol. 135, no. 3, 2023, pp. 1–20.

[5]. Wang, Y. and Zheng, J., “Attention-based networks for banking sector churn analysis,” in *Proc. IEEE EEBDA*, 2023, pp. 1–6.

[6]. Kumar, R., et al., “Automated credit card churn evaluation using hybrid algorithmic models,” in *Proc. E3S Web Conf.*, vol. 430, pp. 1–7, 2023.

[7]. U¨ nlu¨, S., “Support vector machines and Bayesian optimization for churn forecasting,” *CFSU Acad. Stud.*, vol. 8, no. 4, pp. 1–12, 2022.

[8]. Li, J., et al., “Interpretability analysis in machine learning for bank client retention,” *Data Sci. Finan. Econ.*, vol. 5, no. 1, 2025, pp. 1–18.

[9]. Rahman, A., et al., “LSTM and machine learning for time-series telecom churn prediction,” in *Proc. IEEE APSIT*, 2025, pp. 1–6.

[10]. Mart´ınez, J., et al., “Predictive analytics for enhancing telecommunica- tion customer loyalty,” *Front. Artif. Intell.*, vol. 8, pp. 1–15, 2025.

[11]. Zhang, H., et al., “Self-attention mechanisms for high-performance churn prediction systems,” arXiv preprint arXiv:2206.01523, 2022.

[12]. Aydın, M. A. and Yıldırım, E., “GRU-BiLSTM hybrid attention models in banking churn contexts,” *Turk. Online J. Qual. Inq.*, vol. 12, no. 3,pp. 2933–2937, 2021.

[13]. Ahmed, A., et al., “Recent trends and advances in churn prediction: A comprehensive review,” *MLKE*, vol. 7, no. 3, pp. 1–30, 2025.

[14]. Singh, K. and Verma, P., “Comparative study of TCN and LSTM for transaction-based churn analysis,” *Int. J. Eng. Sci. Technol.*, vol. 5, no. 4, pp. 1–10, 2024.