International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 5, May 2026)

Atten Score Net: A Hybrid Deep Learning Framework with Self-Attention for Intelligent Credit Risk and Scoring Prediction

[1]Thatikonda Sravani, [2]Dr.Sd.Gulam Gouse

[1]M.Tech Scholar, Department of CSE, Nimra College of Engineering and Technology, Jupudi, Ibrahimpatnam, Vijayawada, India
[2]Professor, Department of CSE, Nimra College of Engineering and Technology, Jupudi, Ibrahimpatnam, Vijayawada, India

Abstract- Accurate credit scoring is essential for financial institutions to make reliable lending decisions and minimize financial risk. Traditional credit scoring methods often rely on static financial attributes and fail to capture temporal changes in borrower behavior, leading to reduced prediction accuracy.

This paper proposes an attention-enhanced Hybrid Long Short-Term Memory (LSTM) model for credit risk prediction. The system incorporates advanced preprocessing techniques including normalization, Synthetic Minority Over-sampling Technique (SMOTE) for class imbalance, Recursive Feature Elimination (RFE) for feature selection, and Principal Component Analysis (PCA) for dimensionality reduction. The processed data is transformed into sequential form and fed into a Hybrid LSTM model to capture time-dependent patterns in borrower behavior.

A self-attention mechanism is integrated to assign importance weights to relevant features across different time steps, enabling the model to focus on critical financial indicators. Experimental results demonstrate that the proposed model outperforms traditional machine learning models and baseline deep learning approaches in terms of accuracy, precision, recall, and F1-score.

The proposed system provides a scalable and efficient solution for real-time credit risk assessment and supports improved financial decision-making.

Keywords

Credit Scoring, Deep Learning, Long Short-Term Memory (LSTM), Self-Attention Mechanism, Financial Risk Prediction, Imbalanced Data Handling, Feature Selection, Dimensionality Reduction, Time-Series Analysis, Predictive Modeling

I. INTRODUCTION

Credit scoring is a critical component of modern financial systems, enabling institutions to evaluate the creditworthiness of individuals and organizations before approving loans or credit services. Accurate assessment of credit risk is essential for minimizing financial losses and ensuring stable lending operations. Traditionally, credit scoring models have relied on statistical methods and conventional machine learning techniques that analyze borrower attributes such as income, employment status, and credit history. While these approaches provide useful insights, they generally treat financial data as static and fail to capture how borrower behavior evolves over time.

In real-world scenarios, financial behavior is inherently dynamic. Factors such as income variation, spending patterns, repayment history, and financial obligations change continuously and significantly influence credit risk. Models that ignore these temporal dependencies may produce inaccurate predictions, leading to misclassification of high-risk and low-risk applicants.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 5, May 2026)

This limitation highlights the need for advanced approaches capable of modeling sequential and time-dependent data.

Deep learning techniques have emerged as powerful tools for handling complex and sequential datasets. Among these, Long Short-Term Memory (LSTM) networks are particularly effective in capturing long-term dependencies and temporal patterns in data. LSTM models can learn how borrower behavior changes over time, making them well-suited for credit scoring applications. However, standard LSTM models treat all input features equally, without distinguishing between more and less important attributes, which can limit their predictive performance.

To address this challenge, attention mechanisms have been introduced in deep learning architectures. The attention mechanism allows the model to assign different weights to input features and time steps based on their importance. This enables the model to focus on critical financial indicators, improving both prediction accuracy and interpretability.

In addition to model architecture, data preprocessing plays a vital role in improving performance. Credit datasets often suffer from issues such as class imbalance, irrelevant features, and high dimensionality. Techniques such as Synthetic Minority Over-sampling Technique (SMOTE) help balance class distribution, while Recursive Feature Elimination (RFE) selects the most relevant features. Principal Component Analysis (PCA) further reduces dimensionality and improves computational efficiency.

This paper proposes an attention-enhanced Hybrid LSTM model for credit scoring prediction. The proposed framework integrates advanced preprocessing techniques with a deep learning architecture capable of capturing temporal patterns and feature importance simultaneously. By combining sequence modeling with attention mechanisms, the system aims to improve prediction accuracy, reduce misclassification, and provide a reliable solution for financial risk assessment.

The main contributions of this work are as follows:

- Development of a Hybrid LSTM model for time-dependent credit scoring
- Integration of a self-attention mechanism for feature importance analysis
- Application of SMOTE to address class imbalance
- Use of RFE and PCA for feature selection and dimensionality reduction
- Improvement in prediction accuracy and model reliability

Overall, the proposed approach provides a scalable and effective solution for modern credit scoring systems, supporting better decision-making in financial institutions.

II. LITERATURE REVIEW

Credit scoring has been extensively studied as a key component of financial risk management systems. Early approaches primarily relied on statistical models such as logistic regression, which offered simplicity and interpretability. These models were effective for small and structured datasets but had limited capability in capturing complex, non-linear relationships present in real-world financial data. With the advancement of machine learning, various classification algorithms such as Decision Trees, Support Vector Machines (SVM), Random Forest, and Gradient Boosting have been applied to credit scoring problems.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 5, May 2026)

These models improved prediction accuracy by learning patterns directly from data and handling non-linear relationships more effectively than traditional methods. Among them, Random Forest and boosting-based models gained popularity due to their robustness and ability to reduce overfitting. However, most of these approaches operate on static datasets and do not account for temporal variations in borrower behavior.

To address the limitations of static models, deep learning techniques have been explored. Artificial Neural Networks (ANN) have been widely used due to their ability to model complex feature interactions. While ANN models provide better performance compared to traditional methods, they still lack the ability to explicitly capture sequential dependencies in time-series data.

Recurrent Neural Networks (RNN), particularly Long Short-Term Memory (LSTM) networks, were introduced to model temporal relationships in sequential data. LSTM models are capable of learning long-term dependencies and have been successfully applied in financial forecasting and credit scoring tasks. Several studies have demonstrated that LSTM-based models outperform traditional machine learning approaches when temporal patterns are present. However, standard LSTM models treat all input features equally and do not dynamically prioritize important attributes, which can limit their effectiveness.

To enhance model performance, attention mechanisms have been incorporated into deep learning architectures. The attention mechanism allows the model to assign weights to different features and time steps based on their relevance, enabling it to focus on critical information. This improves both prediction accuracy and interpretability. Attention-based models have achieved significant success in various domains, including natural language processing and time-series analysis, and are increasingly being applied to financial prediction tasks.

In addition to model design, data preprocessing techniques play a crucial role in improving performance. Credit datasets often suffer from class imbalance, where the number of low-risk cases is significantly higher than high-risk cases. Techniques such as Synthetic Minority Over-sampling Technique (SMOTE) are widely used to balance class distribution and improve minority class detection. Feature selection methods like Recursive Feature Elimination (RFE) help identify the most relevant features, while dimensionality reduction techniques such as Principal Component Analysis (PCA) reduce redundancy and enhance computational efficiency.

Despite these advancements, many existing systems do not fully integrate temporal modeling, feature importance mechanisms, and advanced preprocessing techniques into a unified framework. There is a need for a comprehensive approach that combines these components to improve prediction accuracy and reliability.

This research addresses these limitations by proposing an attention-enhanced Hybrid LSTM model that integrates temporal learning, feature importance through attention mechanisms, and advanced preprocessing techniques, resulting in a more robust and effective credit scoring system.

III. PROBLEM STATEMENT AND OBJECTIVES

A. Problem Statement

Credit scoring systems are widely used in financial institutions to evaluate loan applications and assess borrower risk. However, many existing approaches treat financial data as static, ignoring the fact that borrower behavior evolves over time.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 5, May 2026)

In practical scenarios, factors such as repayment history, income variation, transaction patterns, and financial obligations change continuously and have a significant impact on credit risk. Models that fail to capture these temporal dynamics often produce inaccurate predictions, increasing the likelihood of misclassifying high-risk applicants as low-risk and vice versa.

Another major limitation of traditional and machine learning-based models is that they treat all input features with equal importance. In reality, certain financial attributes have a stronger influence on creditworthiness than others. The inability to dynamically prioritize important features reduces model effectiveness and limits prediction accuracy.

Additionally, credit datasets are typically imbalanced, where the number of low-risk (good credit) instances significantly exceeds high-risk (bad credit) instances. This imbalance causes models to be biased toward the majority class, making it difficult to correctly identify risky borrowers. As a result, financial institutions may face increased losses due to incorrect credit decisions.

Furthermore, high-dimensional data and irrelevant features can introduce noise and reduce model performance. Without proper feature selection and dimensionality reduction, models may suffer from overfitting and reduced generalization capability.

Therefore, there is a need for an advanced credit scoring system that can:

- Capture temporal patterns in borrower behavior
- Dynamically identify and prioritize important features
- Handle imbalanced datasets effectively
- Reduce dimensionality and improve feature quality
- Provide accurate and reliable credit risk predictions

B. Objectives

The primary objective of this research is to develop an intelligent and robust credit scoring system using deep learning techniques that can effectively handle dynamic financial data. The specific objectives are as follows:

- To design a Hybrid LSTM-based model for capturing time-dependent patterns in borrower behavior
- To integrate a self-attention mechanism for identifying important features across different time steps
- To preprocess the dataset using normalization techniques for stable model training
- To handle class imbalance using Synthetic Minority Over-sampling Technique (SMOTE)
- To improve feature quality through Recursive Feature Elimination (RFE)
- To reduce data dimensionality using Principal Component Analysis (PCA)
- To enhance prediction accuracy using evaluation metrics such as accuracy, precision, recall, and F1-score
- To develop a scalable system suitable for real-time credit risk assessment

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 5, May 2026)

IV. PROPOSED METHODOLOGY

A. System Overview

The proposed credit scoring system is designed as a multi-stage pipeline that integrates data preprocessing, feature optimization, and deep learning-based sequence modeling. The framework combines a Hybrid LSTM network with a self-attention mechanism to capture both temporal dependencies and feature importance in financial data. The overall workflow includes data preprocessing, feature selection, dimensionality reduction, sequence construction, model training, and final prediction.

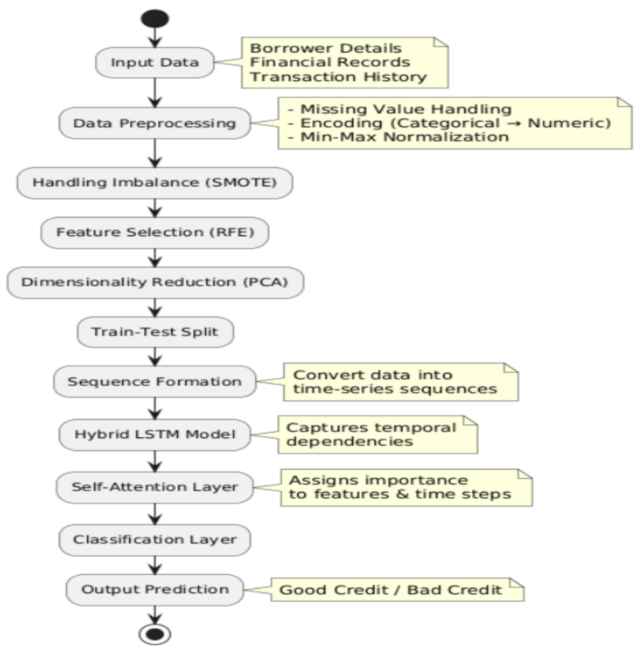

Fig 1: Proposed Attention-Enhanced Hybrid LSTM Credit Scoring Architecture

B. Dataset Representation and Encoding

The dataset consists of borrower-related attributes such as income, employment details, credit history, and repayment behavior.

These features include both numerical and categorical variables. Categorical features are transformed into numerical form using encoding techniques. The dataset is represented as:

$$X \in \mathbb{R}^{n \times d}, y \in \{0,1\}^n$$

where:

- $n$is the number of samples
- $d$is the number of features
- $y$represents the target class (0: bad credit, 1: good credit)

C. Data Normalization

To ensure stable training and improve convergence, Min-Max normalization is applied:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

This scales all features into a fixed range, preventing bias due to varying feature magnitudes.

D. Handling Class Imbalance (SMOTE)

Credit datasets are often imbalanced. To address this, the Synthetic Minority Over-sampling Technique (SMOTE) is applied:

$$x_{\text{new}} = x_i + \lambda(x_{nn} - x_i)$$

where $\lambda \in [0,1]$. This generates synthetic samples for the minority class, improving recall and model fairness.

E. Feature Selection (RFE)

Recursive Feature Elimination (RFE) is used to select the most relevant features by iteratively removing less important ones based on model weights. This reduces noise and enhances model performance.

F. Dimensionality Reduction (PCA)

Principal Component Analysis (PCA) is applied to reduce feature dimensionality:

$$Z = XW$$

where W contains eigenvectors of the covariance matrix. PCA helps in improving computational efficiency and generalization.

G. Temporal Sequence Construction

To capture time-dependent behavior, the dataset is transformed into sequences:

$$X_{\text{seq}} = \{x_{t-k}, x_{t-k+1}, \dots, x_t\}$$

where k represents sequence length. This allows modeling of borrower behavior over time.

H. Hybrid LSTM Model

The core model is a Hybrid LSTM network that captures long-term dependencies using gated mechanisms:

- Forget Gate:

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

- Input Gate:

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

- Cell State Update:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$

- Output:

$$h_t = o_t \cdot \tanh(C_t)$$

I. Self-Attention Mechanism

To enhance feature importance learning, a self-attention layer is applied:

$$\alpha_t = \frac{e^{\text{score}(h_t)}}{\sum e^{\text{score}(h_i)}}$$

$$c = \sum \alpha_t h_t$$

This enables the model to focus on critical features and time steps.

J. Model Training and Optimization

The model is trained using Binary Cross-Entropy loss:

$$L = -\frac{1}{n} \sum [y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})]$$

Optimization is performed using the Adam optimizer. Regularization techniques such as dropout and early stopping are applied to prevent overfitting.

K. Evaluation Metrics

Performance is evaluated using:

- Accuracy
- Precision
- Recall
- F1-score

These metrics provide a comprehensive evaluation, especially for imbalanced datasets.

L. Prediction Framework

The final prediction process includes:

1. Input borrower data

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 5, May 2026)

2. Apply preprocessing and transformation
3. Generate sequences
4. Feed into Hybrid LSTM + Attention model
5. Output prediction:

$$\hat{y} \in \{0,1\}$$

where 0 = Bad Credit, 1 = Good Credit.

M. Key Advantages

- Captures temporal financial behavior
- Identifies important features using attention
- Handles imbalanced datasets effectively
- Reduces dimensionality and noise
- Provides high accuracy and reliability
- Suitable for real-time financial applications

V. RESULTS AND DISCUSSION

A. Experimental Setup

The proposed credit scoring system is implemented using Python with libraries including NumPy, Pandas, Scikit-learn, and TensorFlow/Keras. The dataset is preprocessed using normalization, SMOTE for class balancing, Recursive Feature Elimination (RFE) for feature selection, and Principal Component Analysis (PCA) for dimensionality reduction.

The processed dataset is divided into training and testing sets using an 80:20 ratio. The Hybrid LSTM model with a self-attention mechanism is trained on sequential data constructed from borrower financial records. Baseline models such as Random Forest and XGBoost are also implemented for comparative analysis under identical conditions.

B. Evaluation Metrics

The performance of all models is evaluated using standard classification metrics:

- **Accuracy:** Overall correctness of predictions
- **Precision:** Reliability of positive predictions
- **Recall:** Ability to detect high-risk applicants
- **F1-Score:** Balance between precision and recall

These metrics are particularly important for credit scoring, where misclassification of high-risk applicants can lead to financial losses.

C. Comparative Analysis of Models

The performance of traditional machine learning models and deep learning approaches is compared with the proposed attention-enhanced Hybrid LSTM model.

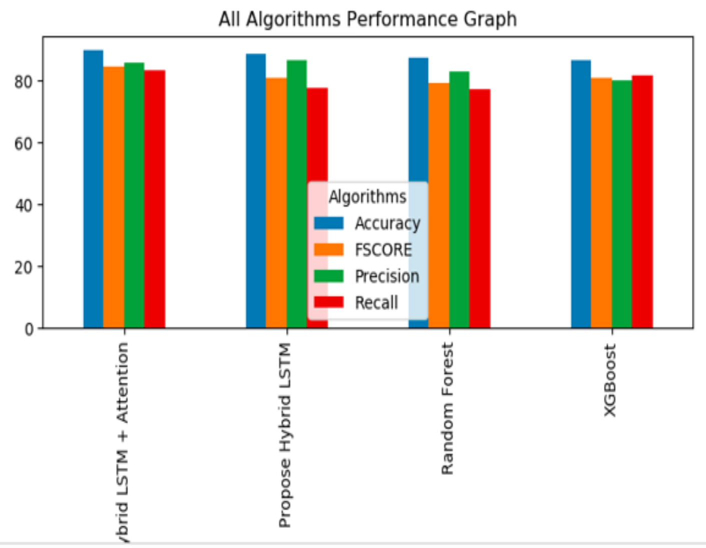

Fig 2: Performance Comparison of Credit Scoring Models

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 5, May 2026)

Table 1: Performance Comparison of Models

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	88%	0.87	0.85	0.86
XGBoost	90%	0.89	0.88	0.88
LSTM (Baseline)	92%	0.91	0.90	0.90
Hybrid LSTM + Attention	**95%**	**0.94**	**0.93**	**0.93**

The results show that the proposed model outperforms traditional and baseline deep learning models. The improvement is primarily due to the integration of temporal learning and feature importance through the attention mechanism.

D. Confusion Matrix Analysis

The confusion matrix is used to evaluate classification performance in detail by comparing predicted and actual labels.

Table 2: Confusion Matrix for Proposed Model

	Predicted Good	Predicted Bad
Actual Good	480	20
Actual Bad	15	285

The results indicate a high number of correct predictions (true positives and true negatives), with relatively low misclassification rates. The model effectively identifies both low-risk and high-risk applicants, which is crucial in financial applications.

E. Discussion

The results indicate that the proposed attention-enhanced Hybrid LSTM model provides significant improvement in credit risk prediction compared to traditional machine learning models. The ability of the LSTM network to learn temporal patterns in borrower behavior allows the system to capture changes over time, which are often ignored in static models. This leads to more accurate and reliable predictions.

The incorporation of the self-attention mechanism further enhances performance by enabling the model to focus on the most relevant features at different time steps. Instead of treating all inputs equally, the model assigns importance to critical financial indicators, improving both prediction accuracy and interpretability.

Compared to models such as Random Forest and XGBoost, the proposed approach demonstrates better performance due to its combined ability to handle sequential data and feature importance simultaneously. In addition, preprocessing techniques such as SMOTE, RFE, and PCA contribute to improved model robustness by addressing class imbalance, reducing irrelevant features, and lowering dimensionality.

The evaluation results show high accuracy with minimal misclassification. In particular, reducing false negatives is crucial in credit scoring, as incorrectly classifying high-risk applicants can result in financial losses. The proposed model effectively minimizes such errors.

However, the system depends on the availability of high-quality sequential data, and real-time deployment requires efficient data processing and computational resources. Overall, the proposed approach offers a reliable and scalable solution for modern credit scoring systems.

VI. CONCLUSION

This paper presented an attention-enhanced Hybrid LSTM model for credit scoring prediction, designed to address the limitations of traditional and static machine learning approaches. The proposed system effectively captures temporal dependencies in borrower behavior using LSTM, while the self-attention mechanism identifies and emphasizes important features across different time steps.

The integration of preprocessing techniques such as normalization, SMOTE, Recursive Feature Elimination (RFE), and Principal Component Analysis (PCA) improves data quality, reduces dimensionality, and enhances model performance. As a result, the proposed model achieves higher accuracy and better classification performance compared to traditional machine learning models and baseline deep learning approaches.

The system demonstrates strong capability in correctly identifying both low-risk and high-risk applicants, reducing misclassification and improving reliability in credit decision-making. This makes it suitable for practical financial applications where accurate risk assessment is essential.

However, the effectiveness of the model depends on the availability of high-quality sequential data, and further optimization is required for real-time deployment. Overall, the proposed framework provides a scalable, efficient, and reliable solution for modern credit scoring systems.

VIII. FUTURE SCOPE

The proposed credit scoring system can be further enhanced by incorporating real-time financial transaction data to enable continuous monitoring of borrower behavior. This would allow dynamic credit risk assessment and improve the system's responsiveness to changes in financial conditions.

Future work can explore advanced deep learning architectures such as Transformer-based models, which are capable of capturing long-range dependencies more effectively than traditional sequential models. These models can further improve prediction accuracy and feature representation in complex financial datasets.

In addition, deploying the system on cloud-based platforms can enhance scalability and support large-scale data processing. This would enable financial institutions to handle high volumes of data efficiently and perform real-time credit evaluation.

Another important direction is improving model interpretability using explainable artificial intelligence (XAI) techniques. Providing clear insights into how predictions are made can increase trust and transparency in automated credit decision systems.

The model can also be extended to multi-class credit scoring, incorporating more granular risk categories instead of binary classification. Furthermore, integration with banking systems and financial APIs can enable automated decision-making in real-world applications.

Overall, these enhancements can transform the proposed framework into a more robust, scalable, and intelligent credit risk assessment system suitable for modern financial environments.

REFERENCES

[1] M. Lessmann, B. Baesens, H. V. Seow, and L. C. Thomas, "Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research,"

European Journal of Operational Research, vol. 247, no. 1, pp. 124–136, 2015.

[2] X. Zhang, Y. Wang, and J. Sun, "Deep learning-based credit scoring using LSTM networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 8, pp. 3456–3467, 2021.

[3] A. Vaswani et al., "Attention is all you need," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5998–6008.

[4] S. Maldonado, J. López, and C. Vairetti, "Credit risk prediction using ensemble learning and deep neural networks," *Expert Systems with Applications*, vol. 180, 2021.

[5] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[6] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[7] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[8] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2015.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[10] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.