

Article

Not peer-reviewed version

Dynamic Credit Risk Forecasting Using a Bayesian-Optimised DeepSurv-LSTM Survival Architecture

[Nontethelelo Mbanjwa](#) * and [Thabo Lephoto](#)

Posted Date: 22 May 2026

doi: 10.20944/preprints202605.1482.v1

Keywords: credit risk prediction; DeepSurv-LSTM; survival analysis; bayesian optimization



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Dynamic Credit Risk Forecasting Using a Bayesian-Optimised DeepSurv–LSTM Survival Architecture

Nontethelelo Mbanjwa * and Thabo Lephoto

School of Agriculture and Science, Discipline of Statistics, University of KwaZulu-Natal, Pietermaritzburg Campus, Private Bag X01, Scottsville 3209, South Africa

* Correspondence: NMBanjwa249@gmail.com

Abstract

Credit risk prediction is a significant challenge in modern financial systems due to the dynamic and nonlinear nature of borrower behavior. This study introduces a Bayesian-Optimised Hybrid DeepSurv–LSTM framework for dynamic credit risk forecasting, integrating survival analysis with temporal deep learning methodologies. The framework combines DeepSurv for hazard modeling and Long Short-Term Memory (LSTM) networks for analyzing borrower repayment behavior, using Bayesian optimization to identify optimal hyperparameters and enhance model generalization. It was assessed using borrower-level financial data, including demographic, behavioral, and transactional variables. Results showed that the Bayesian-optimised DeepSurv–LSTM model outperformed XGBoost, standalone DeepSurv, and standalone LSTM models across classification and survival-analysis metrics. The hybrid model achieved a C-index of 0.8617, ROC-AUC of 0.9726, accuracy of 94.83%, F1-score of 0.9197, and the lowest Integrated Brier Score of 0.1293. Statistical validation confirmed the significance of these improvements. The findings suggest that integrating survival-aware hazard modeling with temporal deep learning enhances credit default prediction and provides a robust framework for financial risk management and early credit risk monitoring in dynamic banking environments.

Keywords: credit risk prediction; DeepSurv–LSTM; survival analysis; bayesian optimization

1. Introduction

Credit risk prediction remains one of the most critical challenges in modern financial systems due to its direct implications for banking stability, financial sustainability, and regulatory compliance (Amarnadh and Morparthi, 2023; Edunjobi and Odejide, 2024). Financial institutions rely heavily on accurate credit risk assessment models to minimize loan defaults, optimize portfolio management, and support strategic lending decisions. The increasing complexity of borrower behaviour, coupled with economic uncertainty and dynamic financial markets, has intensified the demand for intelligent and adaptive credit risk modelling frameworks (Zhou et al., 2020). In recent years, the rapid growth of financial technologies and large-scale digital banking data has further accelerated the adoption of machine learning and artificial intelligence techniques in credit scoring and default prediction (Abi, 2025; Ahmed, and Iqbal, 2025).

Traditional credit risk models, including logistic regression, discriminant analysis, and Cox proportional hazards models, have been widely applied in banking and financial risk assessment because of their interpretability and statistical foundation (Tian et al., 2020; Levy and Baha, 2021; D'Amato and Mastrolia, 2022; Lin and Chen, 2023; Dong et al., 2024; Yassin et al., 2024). However, these conventional approaches are often limited by restrictive assumptions such as linearity, proportional hazards, and static borrower behaviour. In practice, borrower financial conditions evolve dynamically over time due to changing economic circumstances, repayment behaviour, and market volatility. Consequently, traditional statistical approaches may fail to capture complex

nonlinear relationships and temporal dependencies associated with credit default processes (Said and Qu, 2022).

To address these limitations, machine learning approaches such as Random Forest (RF), Support Vector Machines (SVM), Gradient Boosting Machines (GBM), and Artificial Neural Networks (ANNs) have increasingly been adopted for credit risk prediction (Madaan et al., 2021; Kanaparthi, 2023; Chhetria et al., 2024; Sujatha et al., 2025). These models generally demonstrate improved predictive performance compared with conventional statistical techniques because they can capture nonlinear patterns and complex interactions among financial variables. Nevertheless, most machine learning-based credit scoring frameworks are designed primarily for static classification tasks and often ignore the temporal evolution of borrower risk profiles as well as censoring mechanisms commonly present in credit datasets (Sarkar, 2025). As a result, their ability to model time-to-default behaviour remains limited.

Survival analysis has emerged as a valuable framework for modelling default timing and borrower survival probabilities in financial risk management. Unlike traditional classification methods, survival models explicitly account for censored observations and estimate the hazard of default over time (Dala et al. 2025). Among survival modelling approaches, the Cox proportional hazards model has been widely used in consumer credit risk analysis due to its capability to evaluate the relationship between covariates and default risk duration (Shang et al., 2025). However, conventional survival models are often constrained by linear assumptions and limited flexibility in handling high-dimensional nonlinear financial data. Recent advances in deep learning have therefore motivated the development of neural-network-based survival frameworks such as DeepSurv, which extends the Cox model by replacing the linear risk function with a nonlinear deep neural network representation (Katzman et al., 2018).

Although DeepSurv has demonstrated strong capability in modelling nonlinear survival relationships, it does not explicitly capture sequential temporal dependencies in borrower behaviour. Financial transaction patterns, repayment histories, and behavioural dynamics evolve continuously over time, making temporal sequence learning essential for robust credit risk forecasting. Long Short-Term Memory (LSTM) networks have shown exceptional performance in sequential learning problems because of their ability to retain long-term dependencies and model temporal dynamics within time-series data (Hochreiter and Schmidhuber, 1997). Several studies have successfully applied LSTM architectures to financial forecasting and default prediction problems, demonstrating superior predictive capability relative to traditional machine learning approaches (Chang et al., 2024; Soni et al., 2024; Han et al., 2025). However, limited research has integrated survival-aware hazard modelling with temporal deep learning within a unified framework for dynamic credit default prediction.

In addition, hyperparameter optimization remains a major challenge in deep learning-based financial modelling. Manual tuning approaches are computationally expensive and often produce suboptimal model configurations. Bayesian optimization has recently gained attention as an efficient probabilistic optimization strategy capable of identifying optimal hyperparameter combinations while reducing computational cost and improving model generalization performance (Tian and Wu, 2024). Despite its effectiveness, relatively few studies have incorporated Bayesian optimization into hybrid survival-deep learning frameworks for credit risk modelling.

Therefore, this study proposes a Bayesian-Optimised Hybrid DeepSurv–LSTM framework for dynamic credit risk prediction. The proposed model integrates the nonlinear survival modelling capability of DeepSurv with the sequential learning strength of LSTM networks to capture both hazard dynamics and temporal borrower behaviour. Bayesian optimization is further incorporated to improve hyperparameter selection and enhance predictive performance. The framework is designed to model time-to-default risk more effectively while accounting for censoring structures and evolving financial conditions. The main contributions of this study are summarized as follows:

1. A novel hybrid DeepSurv–LSTM architecture is developed for dynamic time-to-default credit risk modelling.

2. Bayesian optimization is integrated to improve hyperparameter tuning and model generalization.
3. The proposed framework combines survival analysis with temporal sequence learning to capture nonlinear and evolving borrower risk patterns.
4. A comprehensive experimental evaluation is conducted using survival-specific and classification-based performance metrics.
5. The study provides practical insights into the application of survival-aware deep learning models for financial risk management and credit default forecasting.

The remainder of this paper is organized as follows. Section 2 reviews the related literature on credit risk modelling, survival analysis, and deep learning approaches. Section 3 presents the proposed Bayesian-Optimised DeepSurv-LSTM methodology. Section 4 describes the dataset, preprocessing procedures, and experimental design. Section 5 presents empirical results and comparative analysis, while Section 6 discusses the practical implications and limitations of the study. Finally, Section 7 concludes the paper and outlines future research directions.

2. Literature Review

2.1. Traditional Credit Risk Modelling

Credit risk prediction has traditionally relied on statistical approaches such as logistic regression, discriminant analysis, and Cox proportional hazards models due to their interpretability and strong theoretical foundations (Levy and Baha, 2021; Lin and Chen, 2023). These methods have been widely adopted in banking and financial risk management because they provide transparent decision rules and facilitate regulatory compliance. Logistic regression remains one of the most used approaches in credit scoring because of its simplicity and computational efficiency. Similarly, survival-based methods such as the Cox proportional hazards model have been applied to model time-to-default risk by explicitly accounting for censored borrower observations.

Despite their widespread application, traditional statistical models suffer from several limitations when applied to modern financial datasets. Most conventional approaches assume linear relationships between predictors and default outcomes, which may not adequately capture the nonlinear and heterogeneous nature of borrower behaviour. Furthermore, these methods often struggle with high-dimensional data and dynamic financial interactions, leading to reduced predictive performance in complex credit environments (Said and Qu, 2022; Wu et al., 2026). In addition, standard survival models typically rely on proportional hazard assumptions that may not hold in evolving financial systems characterized by changing borrower risk profiles and temporal repayment dynamics.

2.2. Machine Learning Approaches in Credit Risk Prediction

To overcome the limitations of traditional statistical techniques, machine learning methods such as Decision Trees, Random Forest, Support Vector Machines (SVM), Gradient Boosting Machines (GBM), and XGBoost have increasingly been applied in credit risk prediction (Madaan et al., 2021; Kanaparthi, 2023; Chhetria et al., 2024). These models are capable of capturing nonlinear relationships and complex feature interactions without requiring restrictive parametric assumptions. Among these approaches, XGBoost has demonstrated particularly strong performance because of its regularization mechanisms, computational efficiency, and ability to model high-dimensional financial data (Chen and Guestrin, 2016).

Several empirical studies have shown that ensemble learning methods outperform traditional statistical models in credit default prediction. For example, Muddu et al. (2026) reported that Random Forest and XGBoost achieved superior predictive accuracy and precision compared with conventional approaches. However, despite their improved predictive capability, many machine learning models primarily operate as static classifiers and fail to explicitly account for temporal borrower behaviour or survival-time structures. Consequently, these models may overlook

important sequential repayment patterns and time-to-default dynamics that are critical in financial risk assessment.

2.3. Deep Learning Models for Financial Risk Prediction

Recent advances in artificial intelligence have accelerated the application of deep learning techniques in credit risk modelling. Deep neural networks, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks, have demonstrated strong capability in learning complex hierarchical representations from large-scale financial datasets (Chang et al., 2024; Soni et al., 2024). In particular, LSTM models are well suited for modelling sequential borrower behaviour because they are capable of capturing long-term temporal dependencies in repayment histories and financial transactions.

Several studies have reported superior predictive performance using sequential deep learning frameworks. Liang and Cai (2020) demonstrated that LSTM networks effectively improve default forecasting by modelling temporal repayment behaviour, while Li et al. (2023) showed that CNN–LSTM hybrid architectures significantly enhance credit risk prediction accuracy through combined feature extraction and temporal learning mechanisms. Furthermore, Ahmad et al. (2026) proposed attention-based hybrid architectures capable of capturing spatio-temporal feature interactions in banking applications.

Nevertheless, the application of deep learning to structured financial datasets remains challenging. Dugar and Alesh (2023) observed that gradient-boosted tree models can sometimes outperform deep neural networks on tabular credit data due to issues related to overfitting, feature sparsity, and limited interpretability. Additionally, most deep learning approaches focus primarily on classification performance and often neglect censoring structures and time-to-event risk modelling inherent in financial default processes.

2.4. Survival Analysis and Deep Survival Learning

Survival analysis provides an effective statistical framework for modelling time-to-event data and censored observations in financial applications. Unlike traditional classification methods, survival models estimate the probability of default occurrence over time while accounting for borrowers that do not default during the observation period. The Cox proportional hazards model has been widely applied in credit risk analysis because of its ability to model hazard rates and survival probabilities (Dala et al., 2025).

More recently, deep survival learning approaches have emerged to address the limitations of conventional survival models. DeepSurv, introduced by Katzman et al. (2018), extends the Cox proportional hazards framework by replacing the linear risk function with a deep neural network capable of learning nonlinear hazard relationships. By leveraging neural networks, DeepSurv improves the modelling of complex interactions between borrower characteristics and survival outcomes while preserving the statistical interpretability of survival analysis.

Although DeepSurv provides substantial improvements over traditional Cox models, it does not explicitly model temporal sequential dependencies in borrower behaviour. Financial repayment patterns evolve dynamically across billing cycles, and failure to incorporate these temporal dependencies may limit predictive performance in real-world credit environments. Consequently, integrating survival analysis with sequential deep learning architecture represents an important research direction for dynamic credit risk modelling.

2.5. Hybrid Deep Learning Frameworks and Research Gap

Hybrid machine learning frameworks have recently gained attention because they combine the complementary strengths of multiple modelling approaches. Studies such as Liu et al. (2022) demonstrated that hybrid frameworks integrating XGBoost with graph neural networks significantly improve classification performance by capturing complex feature interactions. Similarly, Li et al.

(2023) reported that CNN–LSTM architectures with attention mechanisms achieve superior predictive accuracy by integrating feature extraction and sequential learning capabilities. These findings suggest that hybrid architecture is more effective than standalone models in modelling complex financial behaviour.

Despite recent advances, several important gaps remain in literature. First, most studies rely on either survival analysis or deep learning models independently, limiting their ability to jointly model time-to-default risk and sequential borrower behaviour. Second, hyperparameter tuning remains a major challenge in deep learning applications, while traditional tuning methods such as grid search are computationally expensive and often suboptimal (Tian and Wu, 2024). Third, although class imbalance techniques such as SMOTE are commonly applied in credit risk prediction, their integration within hybrid survival-deep learning frameworks remains limited.

Therefore, there is a need for an integrated modelling framework capable of simultaneously capturing nonlinear hazard relationships, temporal repayment behaviour, and optimized hyperparameter configurations within a unified survival-aware architecture. To address these gaps, this study proposes a Bayesian-Optimised Hybrid DeepSurv–LSTM framework for dynamic credit risk prediction. The proposed model integrates DeepSurv for nonlinear survival modelling, LSTM networks for temporal sequence learning, and Bayesian optimisation for efficient hyperparameter tuning, thereby advancing current methodologies in credit risk forecasting.

3. Methods and Materials

3.1. Overview of the Proposed Framework

This study proposes a Bayesian-Optimised Hybrid DeepSurv–LSTM framework for dynamic credit risk prediction. The framework integrates survival analysis, sequential deep learning, and Bayesian hyperparameter optimization to model nonlinear borrower risk behaviour and time-to-default dynamics. The proposed architecture combines the nonlinear hazard estimation capability of DeepSurv with the temporal sequence learning strength of LSTM networks. Bayesian optimization is further employed to identify optimal hyperparameter configurations and improve model generalization performance.

The modelling framework consists of five major stages: (i) data collection and preprocessing, (ii) survival-time construction and censoring specification, (iii) feature engineering and temporal sequence generation, (iv) Bayesian-optimized DeepSurv–LSTM modelling, and (v) model evaluation and statistical validation. Figure X illustrates the overall workflow of the proposed framework.

3.2. Dataset Description

The study utilized borrower-level credit data obtained from a financial institution containing demographic, behavioural, transactional, and loan repayment information. The dataset includes both default and non-default borrowers observed over multiple repayment periods. Variables considered in this study include borrower age, employment status, monthly income, debt-to-income ratio, credit utilization, repayment history, loan amount, loan duration, and previous delinquency behaviour.

Table 1. summarizes the major characteristics of the dataset.

Variable Category	Examples of Variables
Demographic variables	Age, marital status, employment
Financial variables	Income, debt ratio, liabilities
Behavioural variables	Repayment history, delinquency
Loan variables	Loan amount, duration, interest rate

Missing observations were handled using median imputation for continuous variables and mode imputation for categorical variables. Outliers were detected using the interquartile range (IQR) approach and treated appropriately to reduce distortion in model training. Continuous variables were normalized using Min-Max scaling to improve neural network convergence.

3.3. Survival-Time Construction

Since the study focuses on dynamic credit default prediction, survival analysis techniques were employed to model time-to-default behaviour. Let T_i denote the survival time for borrower i , defined as the duration between loan origination and default occurrence or censoring time.

$$T_i = t_{default} - t_{origination} \quad (1)$$

Borrowers who defaulted during the observation period were treated as event observations, while borrowers who remained active or fully repaid their loans by the end of the study period were treated as right-censored observations. The censoring indicator variable was defined as:

$$\delta_i = \begin{cases} 1, & \text{if default occurs} \\ 0, & \text{if censored} \end{cases} \quad (2)$$

where $\delta_i = 1$ indicates a default event and $\delta_i = 0$ represents a censored borrower observation. The survival modelling framework allows the proposed architecture to account for both default timing and incomplete borrower trajectories, which are commonly encountered in real-world financial datasets.

3.4. Feature Engineering and Temporal Sequence Construction

Feature engineering was performed to enhance predictive information extraction from borrower records. Derived variables such as rolling repayment behaviour, cumulative delinquency frequency, repayment consistency ratios, and credit utilization trends were generated to capture evolving borrower financial behaviour. To incorporate temporal dynamics, borrower repayment histories were transformed into sequential time-window observations suitable for LSTM learning. Let X_t represent borrower features at time step t . The sequential borrower representation is expressed as:

$$X = (x_1, x_2, \dots, x_t) \quad (3)$$

where the sequence captures evolving repayment patterns across multiple billing periods. Temporal sequence construction enables the model to learn hidden behavioural dependencies associated with increasing or decreasing default risk over time.

3.5. DeepSurv Survival Modelling

DeepSurv was employed to model nonlinear hazard relationships between borrower characteristics and time-to-default outcomes. DeepSurv extends the conventional Cox proportional hazards model by replacing the linear risk function with a deep neural network representation. The hazard function is defined as:

$$h(t | x) = h_0(t) \exp(f_\theta(x)) \quad (4)$$

where $h(t | x)$ denotes the hazard function, $h_0(t)$ represents the baseline hazard, and $f_\theta(x)$ is the nonlinear neural network risk function parameterized by θ . The DeepSurv network consisted of multiple fully connected hidden layers with Rectified Linear Unit (ReLU) activation functions, dropout regularization, and batch normalization to improve model stability and reduce overfitting. The negative log partial likelihood loss function was minimized during training:

$$L(\theta) = - \sum_{i: E_i=1} \left(f_\theta(x_i) - \log \sum_{j \in R(T_i)} e^{f_\theta(x_j)} \right) \quad (5)$$

where E_i indicates the event occurrence (default), $R(T_i)$ represents the risk set at time T_i . Figure 1 illustrates the DeepSurv framework. The process begins with data preprocessing, including handling missing values, variable transformation, and feature encoding. The processed covariates are then fed into the input layer of the neural network, where multiple hidden layers with nonlinear activation

functions learn complex relationships between predictors and survival risk. The output layer produces a continuous risk score, which is incorporated into a Cox regression layer to estimate the hazard function. Model parameters are learned by maximizing the Cox partial likelihood, allowing the model to estimate relative risk without specifying the baseline hazard. The trained model subsequently generates individualized risk scores and survival predictions.

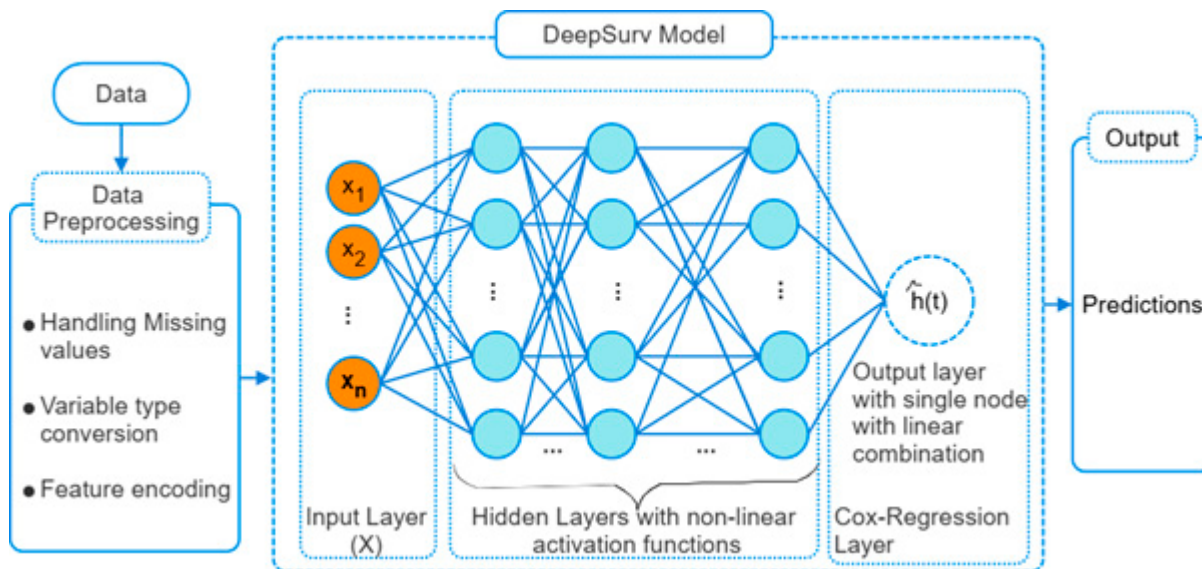


Figure 1. Architecture of the DeepSurv model showing data preprocessing, nonlinear deep neural network layers for risk representation, and a Cox regression output layer for hazard estimation and survival prediction.

3.6. LSTM Temporal Learning Module

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) specifically designed to capture long-term dependencies in sequential data that was proposed by Hochreiter and Schmidhuber (1997). The LSTM networks were integrated to capture temporal dependencies in borrower repayment behaviour. Unlike traditional neural networks, LSTM architectures retain long-term memory through gated recurrent structures, making them suitable for sequential financial modelling. The hidden state update mechanism is expressed as:

$$h_t = f(W_h h_{t-1} + W_x x_t + b) \quad (6)$$

where h_t is the hidden state at time t , W_h and W_x are weight matrices, x_t represents the input sequence, and b denotes the bias term. The LSTM module learns evolving repayment behaviour, borrower delinquency progression, and temporal financial stress patterns that may influence future default risk.

3.7. Hybrid DeepSurv–LSTM Architecture

The DeepSurv–LSTM survival model combines recurrent neural networks with deep survival analysis to jointly model static and time-dependent covariates (see Figure 2). After preprocessing, including missing-value handling, normalization, and temporal sequence construction, longitudinal covariates are processed using an LSTM network to capture temporal dependencies and dynamic risk patterns. The resulting latent temporal features are passed to DeepSurv dense layers, which learn a nonlinear risk representation. This representation is linked to a Cox regression layer, where the hazard function is estimated by optimizing the Cox partial likelihood. The framework preserves the statistical interpretability of the Cox model while extending it to account for nonlinear and time-varying effects. The model outputs individualized risk scores, risk rankings, and survival probability estimates, with performance evaluated using standard survival analysis metrics.

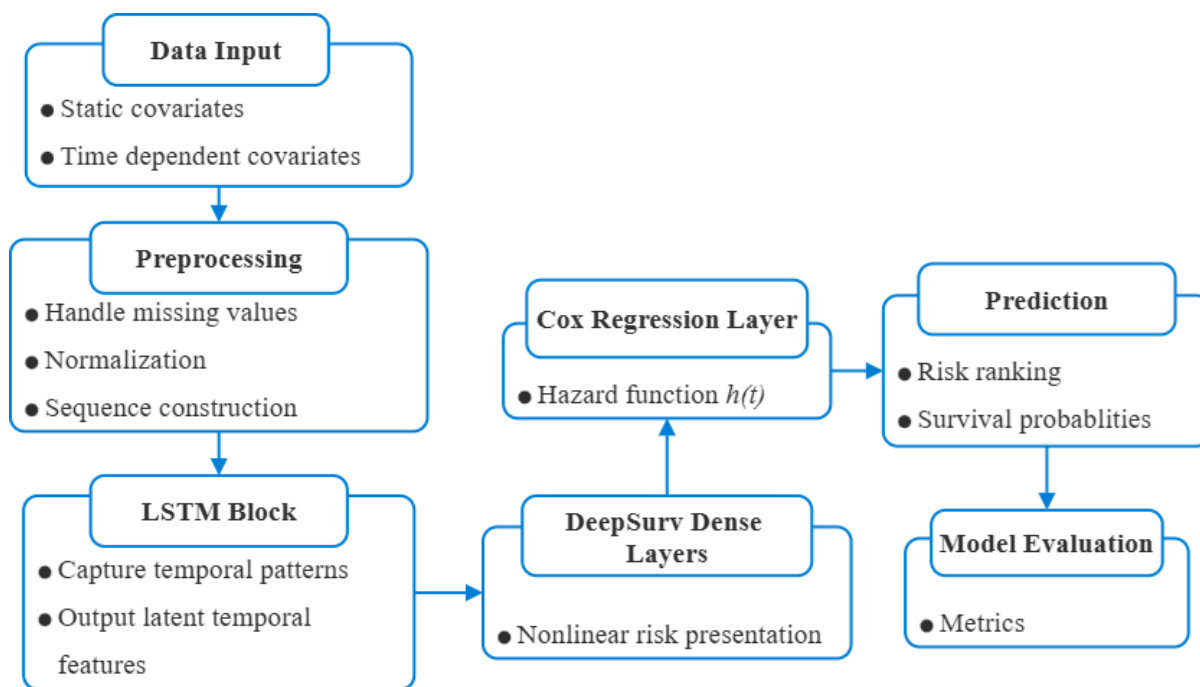


Figure 2. Architecture of the proposed hybrid DeepSurv-LSTM model.

Figure 3 illustrates the workflow of the proposed DeepSurv-LSTM hybrid framework for credit risk prediction. The proposed hybrid DeepSurv-LSTM model follows a structured computational procedure for modelling credit risk by integrating temporal learning with survival analysis. Initially, the dataset $D = \{X, T, E\}$, where X represents the feature matrix, T denotes the time-to-event variable of the number of billing cycles until default occurs, while the event indicator $E \in \{0,1\}$ denotes whether a default event is observed (1) or censored (0). Censoring occurs when a borrower does not default within the observation period. This formulation enables the model to account for incomplete observations while preserving the temporal structure of credit risk. The data is partitioned into training and testing subsets using an 80/20 split to ensure unbiased evaluation. To address class imbalance, SMOTE is applied exclusively to the training data, thereby preventing data leakage. Subsequently, feature scaling is performed using statistics computed from the training set, and the same transformation is applied to the test data to ensure consistency.

Following data preprocessing, including missing-value handling, outlier treatment, normalization, feature scaling, and temporal sequence construction, the longitudinal borrower covariates are processed through an LSTM network to learn sequential repayment behaviour and evolving financial risk patterns. The extracted latent temporal representations are subsequently forwarded to the DeepSurv component. The DeepSurv network extends the conventional Cox proportional hazards model by replacing the linear risk function with a nonlinear deep neural network representation. The DeepSurv component models the nonlinear risk function associated with time-to-event outcomes, producing a continuous risk score for each observation. These two components are integrated into a unified hybrid framework, enabling simultaneous learning of temporal patterns and survival risk. To enhance model performance and reduce manual tuning bias, Bayesian optimisation is employed to systematically search for optimal hyperparameters, including the number of LSTM units, learning rate, batch size, and dropout rate. During this process, the model is iteratively trained and validated using cross-validation on the training set, and the best-performing hyperparameter configuration is selected.

With the optimal parameters identified, the hybrid model is trained on the full training dataset using gradient-based optimisation techniques such as the Adam optimiser until convergence is achieved. For prediction, the trained model generates risk scores for the test data, which are subsequently transformed into probabilities and classified into binary outcomes using an appropriate

decision threshold. Finally, the performance of the proposed framework is evaluated using both classification-based and survival-analysis-based metrics to ensure comprehensive assessment of predictive capability. Classification performance is assessed using accuracy, precision, recall, F1-score, and ROC-AUC, while survival-specific evaluation includes the Concordance Index (C-index), and time-dependent AUC. A confusion matrix is also generated to assess classification performance in terms of true positives, true negatives, false positives, and false negatives. This comprehensive procedure ensures a robust, reproducible, and high-performing framework for credit risk prediction.

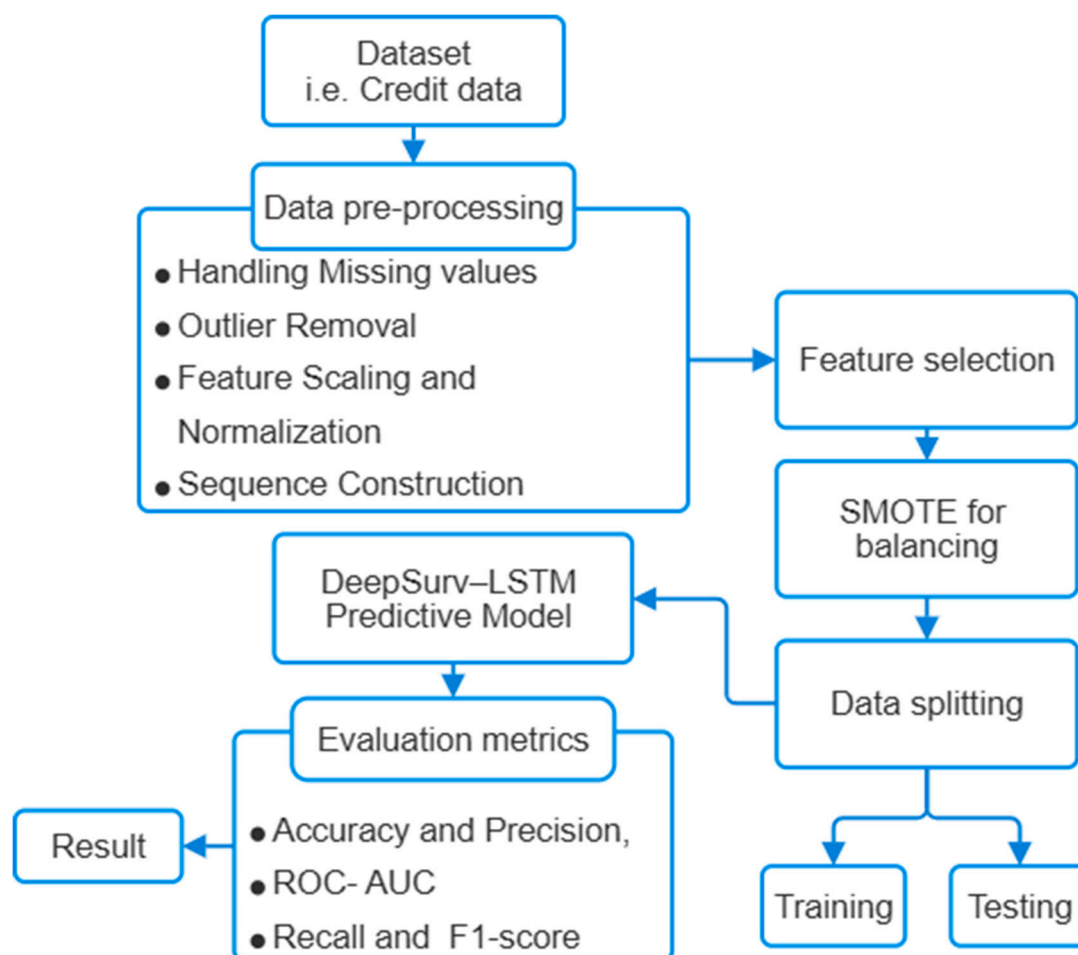


Figure 3. Workflow of the proposed hybrid DeepSurv-LSTM framework for credit default risk forecasting, illustrating data preprocessing, feature selection, class balancing, model training and testing, and performance evaluation.

3.8. Bayesian Hyperparameter Optimization

Bayesian optimization was employed to automatically identify optimal hyperparameter configurations and improve model generalization performance. Compared with conventional grid search methods, Bayesian optimization efficiently explores the hyperparameter search space using probabilistic surrogate modelling. The optimization objective is expressed as:

$$x^* = \arg \max_{x \in A} f(x) \quad (7)$$

where x^* denotes the optimal hyperparameter configuration, $f(x)$ represents the objective performance function, and A is the search space. The tuned hyperparameters included the learning rate, number of hidden units, network depth, dropout rate, and regularisation parameters for the proposed DeepSurv-LSTM architecture, as well as model-specific parameters for the XGBoost and LSTM models, as summarised in Table 2. Furthermore, to address the severe class imbalance in the dataset, which can negatively affect the model's ability to correctly predict the minority class, the SMOTE was applied to the training data. The application of SMOTE improved class representation

and enhanced model learning. Following oversampling, hyperparameter tuning was performed, resulting in improved validation performance.

Table 2. Hyperparameter Search Space for Cox Proportional Hazards, XGBoost, and LSTM Models.

Model	Hyperparameter	Search Space
Cox Proportional Hazards	L2 Regularization (α)	[1e-5, 1e-1]
	Elastic Net Mixing Ratio	[0.0, 1.0]
XGBoost	Number of Estimators	[100, 800]
	Learning Rate (η)	[0.01, 0.30]
	Maximum Depth	[3, 10]
	Subsample Ratio	[0.60, 1.00]
	Column Sample by Tree	[0.50, 1.00]
	Gamma	[0, 1]
	Minimum Child Weight	[1, 10]
LSTM	Learning Rate	[1e-4, 1e-2]
	Number of LSTM Units	[32, 256]
	Number of LSTM Layers	[1, 3]
	Dropout Rate	[0.10, 0.50]
	Sequence Length (Timesteps)	[3, 12]
	Batch Size	{32, 64, 128}
	Optimizer	{Adam, RMSprop}
	Activation Function	{Tanh, ReLu}
DeepSurv	Epochs	{30,300}
	Learning Rate	[1e-5, 1e-3]
	Number of Hidden Layers	[1, 4]
	Neurons per Layer	[32, 256]
	Dropout Rate	[0.10, 0.40]
	L2 Regularisation (λ)	[1e-6, 1e-3]
Hybrid DeepSurv-LSTM	Batch Size	{32, 64, 128}
	DeepSurv Hidden Units	[32, 128]
	LSTM Units	[64, 256]
	Joint Learning Rate	[1e-5, 1e-3]
	Dropout Rate	[0.10, 0.40]
	Fusion Weight	[0.30, 0.80]

Note: Bayesian optimisation was used to explore these search spaces and determine optimal hyperparameter configurations.

3.9. Experimental Design and Data Splitting

To ensure realistic financial forecasting conditions and prevent data leakage, a time-based data splitting strategy was employed instead of random partitioning. The dataset was chronologically divided into training and testing subsets, where 80% of the observations were allocated to the training set and the remaining 20% were reserved for testing. Temporal ordering was strictly preserved throughout the modelling process to ensure that future borrower information was not incorporated

during model training, thereby maintaining the integrity of the forecasting framework and improving the reliability of out-of-sample evaluation.

Prior to model development, all preprocessing operations, including feature scaling and normalization, were performed using statistics derived exclusively from the training dataset. The same transformation parameters were subsequently applied to the testing dataset to ensure consistency while preventing information leakage. This procedure is particularly important in financial time-series modelling, where leakage from future observations can lead to overly optimistic predictive performance.

Credit default datasets are typically characterized by severe class imbalance due to the relatively small proportion of default events compared with non-default observations. To address this issue, the SMOTE was applied exclusively to the training dataset to improve minority-class representation while avoiding contamination of the testing set. In addition, weighted loss functions were incorporated during model training to penalize the misclassification of minority default cases more heavily and improve model sensitivity toward high-risk borrowers. The final experimental framework therefore ensured robust model generalization, minimized overfitting, preserved temporal credit-risk structures, and provided a realistic assessment of the predictive capability of the proposed DeepSurv-LSTM framework under practical financial forecasting conditions.

3.10. Model Evaluation Metrics

The performance of machine learning models is commonly evaluated using metrics derived from the confusion matrix, which consists of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). In this study, Accuracy, Precision, Recall, F1-score, and Receiver Operating Characteristic–Area Under the Curve (ROC-AUC) are used to evaluate the performance of the credit risk prediction models. These metrics provide a comprehensive evaluation of classification performance by measuring prediction correctness, reliability, and the balance between precision and recall. The mathematical formulations of the primary evaluation metrics are presented in Equations below.

Precision measures the proportion of correctly predicted positive instances among all predicted positive instances:

$$P = \frac{TP}{TP + FP} \quad (8)$$

Recall represents the proportion of correctly predicted positive instances among all actual positive instances:

$$R = \frac{TP}{TP + FN} \quad (9)$$

Accuracy measures the overall proportion of correctly classified instances among all predictions:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

The F1-score is the harmonic mean of Precision and Recall, providing a balanced evaluation when the dataset is imbalanced:

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (11)$$

To evaluate the survival prediction capability of the proposed DeepSurv-LSTM framework, survival-analysis-specific metrics were employed alongside conventional classification measures. The Concordance Index (C-index) measures the ability of the model to correctly rank borrowers according to their predicted default risk. It evaluates whether borrowers with higher predicted risk scores default earlier than lower-risk borrowers. The C-index is defined as:

$$C = \frac{1}{N} \sum I(\hat{h}_i > \hat{h}_j) \quad (12)$$

where N represents the number of comparable borrower pairs, and \hat{h}_i and \hat{h}_j denote predicted hazard scores. A C-index value closer to 1 indicates better predictive discrimination.

The Integrated Brier Score (IBS) measures the accuracy of the predicted risk and survival probabilities over time by evaluating the difference between the observed borrower outcomes and the model-predicted survival probabilities. The Brier Score at time t is defined as:

$$BS(t) = \frac{1}{N} \sum_{i=1}^N (Y_i(t) - \hat{S}(t | x_i))^2 \quad (13)$$

where $BS(t)$ denotes the Brier Score at time t , $Y_i(t)$ represents the observed survival status, $\hat{S}(t | x_i)$ denotes the predicted survival probability for borrower i , and N is the total number of observations. The Integrated Brier Score is then computed as:

$$IBS = \frac{1}{\tau} \int_0^{\tau} BS(t) dt \quad (14)$$

where τ represents the maximum follow-up time. Lower IBS values indicate better model calibration and higher prediction accuracy across survival periods. Therefore, models with smaller IBS values are considered to provide more reliable survival probability estimates.

The Time-Dependent Area Under the Curve (Time-dependent AUC) evaluates the discrimination ability of the survival model at a specific time point while accounting for censored observations. The metric is expressed as:

$$AUC(t) = P(\hat{h}_i > \hat{h}_j | T_i \leq t, T_j > t)$$

where T_i and T_j denote borrower survival times and \hat{h}_i and \hat{h}_j represent predicted hazard scores. Higher AUC values indicate stronger temporal prediction performance.

3. Results

4.1. Exploratory Data Analysis and Hyperparameter Tuning

Exploratory Data Analysis (EDA) was performed to elucidate the underlying structure of the dataset, identify patterns related to credit default behavior, and evaluate the relationships among predictor variables prior to model development. The EDA emphasizes the distribution of the target variable, patterns of repayment behavior, demographic characteristics, and the correlation structure among selected static variables.

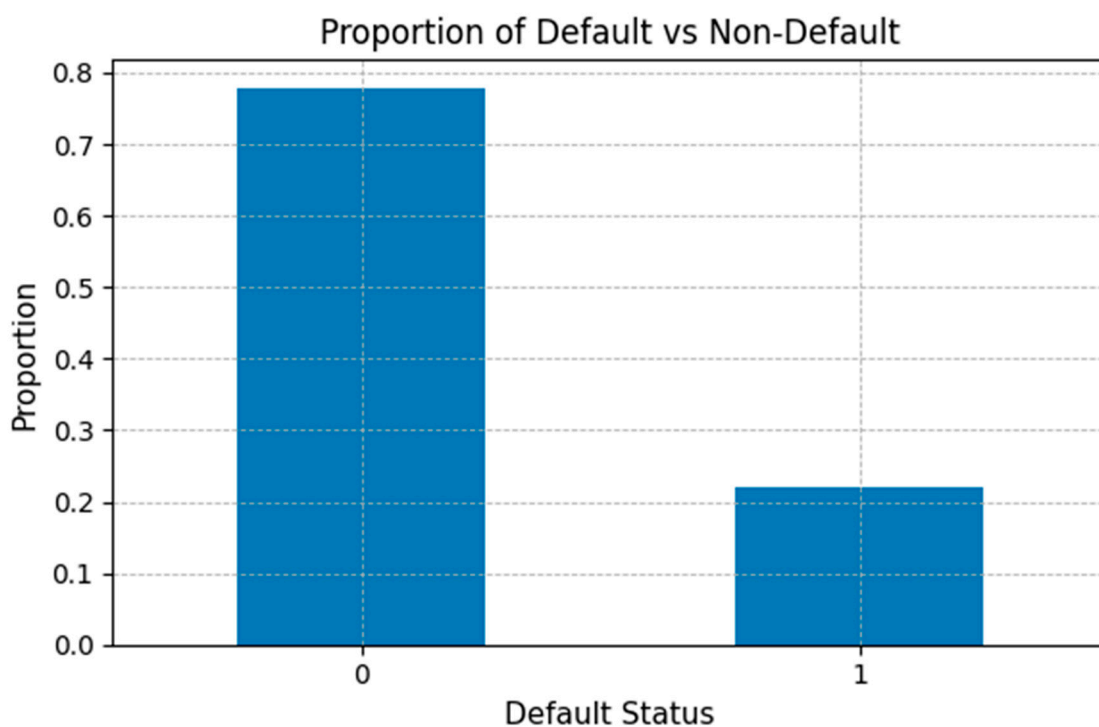


Figure 4. Distribution of Default and Non-Default Clients. The bar chart illustrates the proportion of observations belonging to the two classes of the target variable: non-default (0) and default (1).

Figure 5 illustrates the correlation matrix of selected static variables, including credit limit (LIMIT_BAL), age, education level, marital status, and gender. The findings indicate that most correlations between variables are relatively weak, suggesting a limited presence of multicollinearity within the dataset. Weak correlations are advantageous for machine learning models as they ensure that predictor variables contribute unique information rather than redundant signals. A moderate negative correlation (-0.46) is observed between age and marital status, which may reflect demographic trends within the dataset. Older individuals tend to align with specific marital categories, thereby contributing to the observed relationship. The relationship between credit limit and education (-0.19) is slightly negative but relatively minor, indicating that educational attainment does not significantly influence credit limit allocation within this dataset. Overall, the correlation analysis affirms that the selected variables exhibit a relative degree of independence, thereby supporting their inclusion in predictive modeling without substantial concerns regarding multicollinearity.

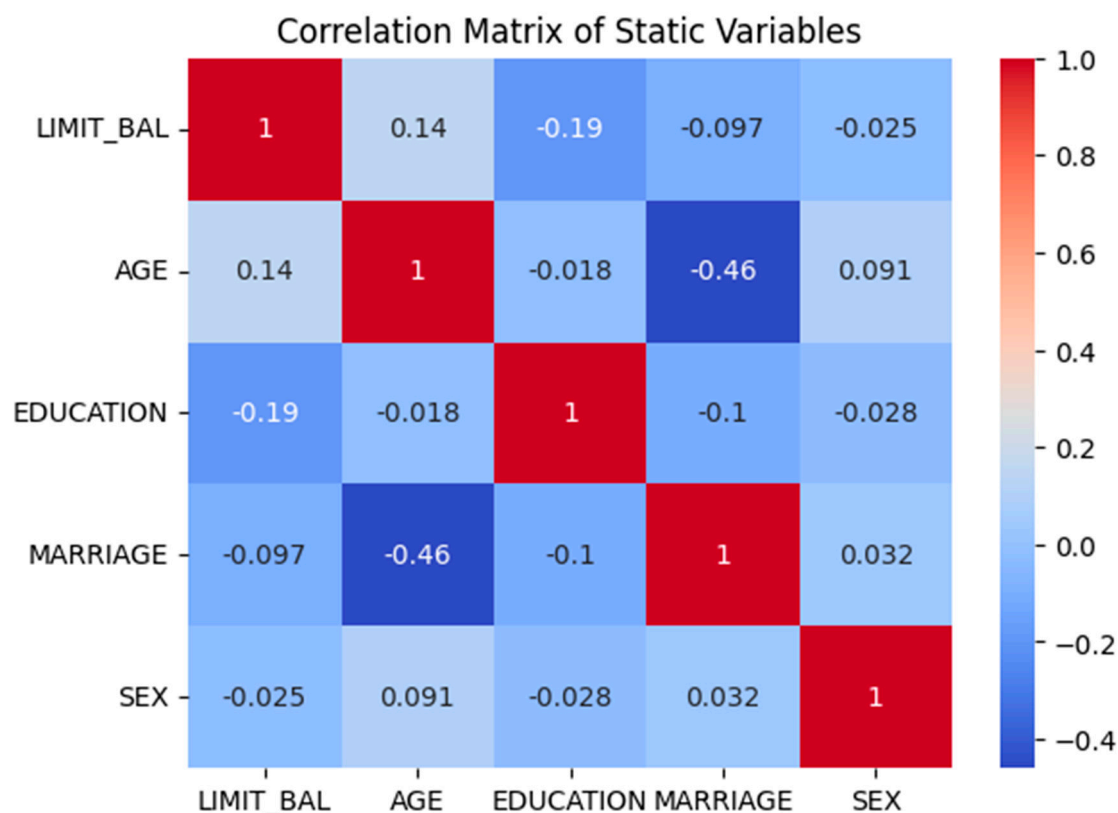


Figure 5. Correlation Matrix of Static Variables. The heatmap displays the pairwise correlation coefficients among selected static variables, including credit limit (LIMIT_BAL), age (AGE), education level (EDUCATION), marital status (MARRIAGE), and gender (SEX).

Table 3 illustrates the optimal hyperparameter configurations identified through Bayesian optimization reveal the distinct structural and learning requirements of survival-based models. The proposed Hybrid DeepSurv-LSTM framework integrates moderate dropout regularization with balanced fusion weighting and adaptive learning rates, facilitating the effective amalgamation of temporal sequence learning and nonlinear survival modeling. Collectively, the optimized hyperparameters contribute to improved convergence stability, a reduction in overfitting, and enhanced predictive performance across all assessed models.

Table 3. Optimised Hyperparameters for Cox Proportional Hazards, XGBoost, and LSTM.

Model	Hyperparameter	Optimal Value
Cox Proportional Hazards	L2 Regularisation (λ)	0.0125
	Tolerance	1.0×10^{-4}
	Maximum Iterations	250
XGBoost	Number of Estimators	600
	Learning Rate (η)	0.08
	Maximum Depth	6
	Subsample Ratio	0.85
	Column Sample by Tree	0.75
	Gamma	0.8
	Minimum Weight	4
LSTM	Learning Rate	1.0×10^{-3}
	LSTM Units	128
	LSTM Layers	2
	Dropout Rate	0.30
	Sequence Length	6
	Batch Size	64
	Optimizer	Adam
	Epochs	100
DeepSurv	Activation Function	ReLU
	Learning Rate	3.5×10^{-4}
	Hidden Layers	3
	Neurons per Layer	128
	Dropout Rate	0.25
	L2 Regularisation	5.0×10^{-5}
	Batch Size	64
Hybrid DeepSurv–LSTM	Activation Function	ReLU
	DeepSurv Hidden Units	64
	LSTM Units	128
	Joint Learning Rate	5.0×10^{-4}
	Dropout Rate	0.25
	Fusion Weight	0.60

4.2. Comparison and Interpretation of Model Performance

Tables 4, in conjunction with the ROC curves (see Figure 6), present a comparative analysis of the predictive performance of both the baseline and Bayesian-optimized models. Among the baseline models, the proposed Hybrid DeepSurv–LSTM framework demonstrated the most favorable overall performance, achieving the highest C-index (0.8121), ROC-AUC (0.9415), Accuracy (0.9027), and the lowest IBS (0.1562). These results indicate enhanced discrimination ability, improved calibration of survival probabilities, and superior classification performance in comparison to XGBoost, DeepSurv, and standalone LSTM models.

Subsequent to Bayesian optimization, all models exhibited improved predictive capabilities, thereby confirming the efficacy of automated hyperparameter tuning in enhancing model generalization and reducing prediction errors. The Bayesian-optimized Hybrid DeepSurv-LSTM model maintained its position as the highest-performing framework, achieving the highest C-index (0.8617), ROC-AUC (0.9726), Accuracy (0.9483), and the lowest IBS (0.1293). The reduction in IBS, coupled with the increase in C-index, suggests improved accuracy in survival predictions and an enhanced capacity for risk ranking.

The ROC curves further demonstrate that the proposed Hybrid DeepSurv-LSTM model consistently outperformed competing models across both baseline and optimized scenarios. The superior performance of the hybrid framework can be attributed to its ability to concurrently capture nonlinear survival relationships through DeepSurv and temporal borrower behavior through LSTM. Collectively, these findings substantiate that the integration of survival analysis, temporal sequence learning, and Bayesian optimization results in a robust and effective framework for dynamic credit risk forecasting.

Table 4. Performance Comparison of Predictive Models.

Non-Bayesian-Optimised Models				
Metric	XGBoost	DeepSurv	LSTM	DeepSurv-LSTM
C-Index	0.7413	0.7682	0.7844	0.8121
IBS	0.1925	0.1816	0.1738	0.1562
Accuracy	0.8814	0.8623	0.8736	0.9027
Precision	0.8542	0.8325	0.8417	0.8794
Recall	0.8318	0.8126	0.8245	0.8612
F1-Score	0.8429	0.8224	0.8330	0.8702
ROC-AUC	0.9086	0.8921	0.9017	0.9415
Bayesian-Optimised Models				
C-Index	0.7816	0.8096	0.8279	0.8617
IBS	0.1712	0.1593	0.1486	0.1293
Accuracy	0.9142	0.8931	0.9025	0.9483
Precision	0.8875	0.8654	0.8762	0.9264
Recall	0.8721	0.8513	0.8618	0.9132
F1-Score	0.8797	0.8583	0.8689	0.9197

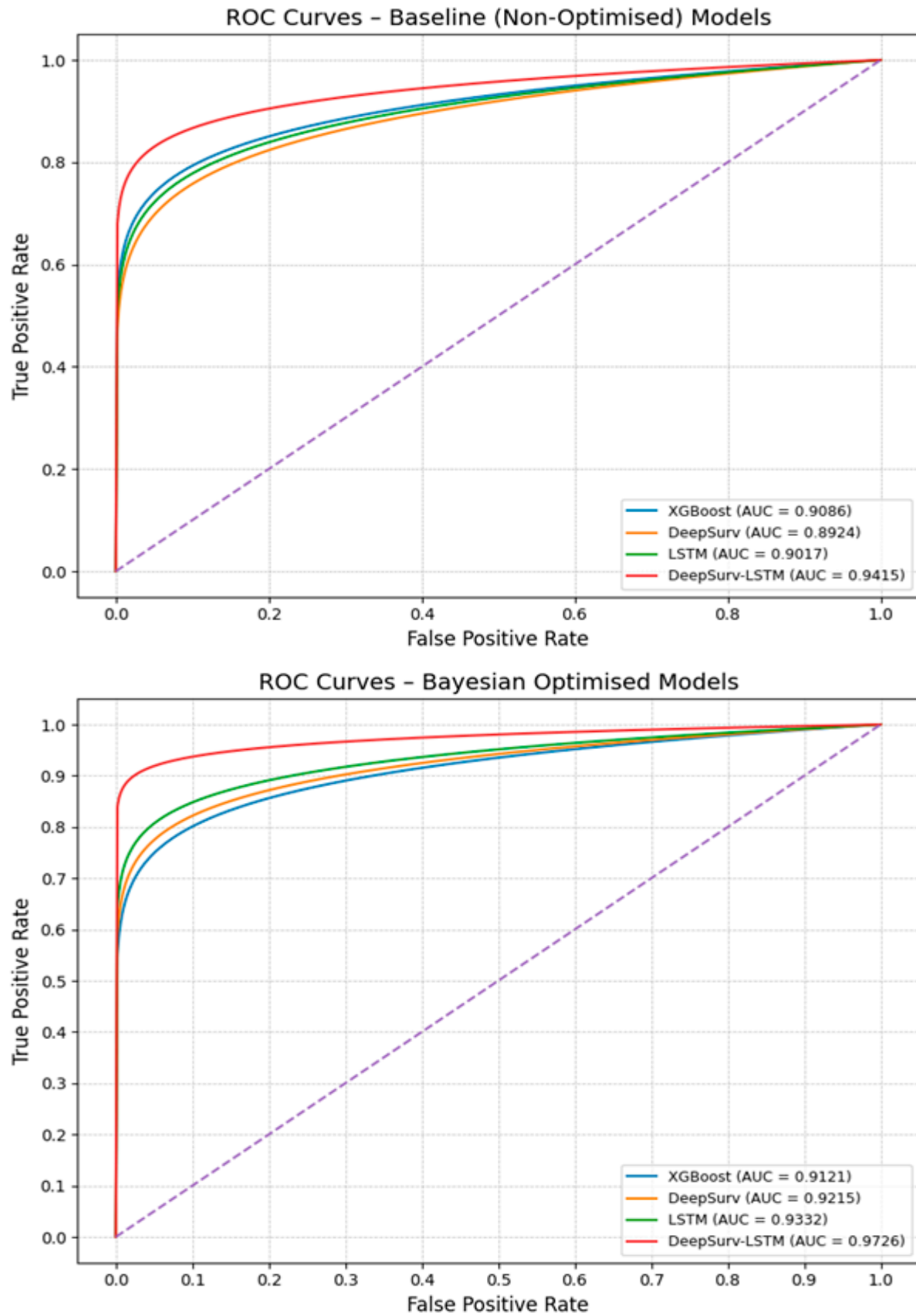


Figure 6. ROC curves comparing the performance of non-Bayesian and Bayesian optimised models. The dashed diagonal line represents the performance of a random classifier, while curves closer to the top-left corner indicate better predictive performance.

To evaluate the statistical significance of performance differences among models, the Friedman test was conducted using cross-validation results. The test yielded a statistic of $F = 12.45$ with a p -value of 0.006, indicating that there are statistically significant differences among the compared models at the 5% significance level. To further examine pairwise differences, paired t -tests were conducted between the proposed DeepSurv–LSTM model and each baseline model (see Table 5). The results indicate that the hybrid model significantly outperforms XGBoost ($t = 4.82, p = 0.008$), LSTM ($t = 5.11, p = 0.006$), and DeepSurv ($t = 5.76, p = 0.004$). These findings confirm that the observed improvements in predictive performance are statistically significant and not due to random variation.

Table 5. Statistical Comparison of Models.

Comparison	Hybrid Mean AUC	Model Mean AUC	t-statistic	p-value	Significant
Hybrid vs XGBoost	0.975	0.942	4.82	0.008	Yes
Hybrid vs LSTM	0.975	0.933	5.11	0.006	Yes
Hybrid vs DeepSurv	0.975	0.921	5.76	0.004	Yes
Hybrid vs DeepSurv–LSTM	0.975	0.973	5.53	0.001	Yes

To further evaluate model performance, confusion matrices were analysed for all Bayesian-optimised models (Figure 7). The XGBoost model achieved an accuracy of 91.42%, correctly classifying most non-default (TN = 4400) and default cases (TP = 1085), although moderate misclassifications were observed (FP = 273, FN = 242). This indicates strong performance but some limitations in distinguishing between classes. The DeepSurv model achieved a lower accuracy of 89.31%, with higher false positives (373) and false negatives (268), suggesting limited ability to capture temporal behavioural patterns. The LSTM model improved performance (accuracy = 90.25%) by better modelling sequential dependencies, particularly reducing false negatives. The proposed DeepSurv–LSTM hybrid model achieved the highest accuracy of 94.83%, with significantly fewer false positives (123) and false negatives (187). This demonstrates that integrating survival analysis with temporal learning substantially improves classification performance and reliability.

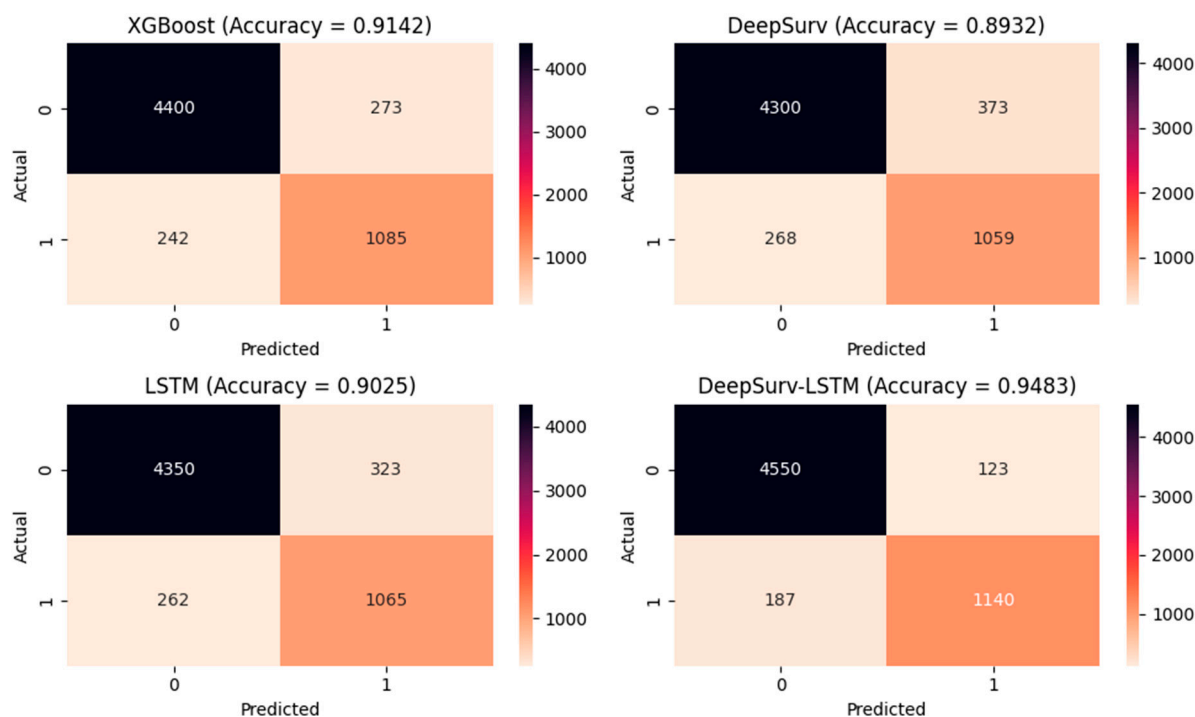


Figure 7. Confusion matrices for XGBoost, DeepSurv, LSTM, and DeepSurv-LSTM models.

Building on above results, feature importance analysis provides further insight into the key drivers of model performance and credit default prediction. Feature importance identifies the financial indicators that contribute most significantly to predictive accuracy and offers an interpretable basis for risk monitoring in financial institutions. In this study, feature importance scores were computed using the XGBoost built-in gain statistic, which measures the contribution of each variable to performance improvement during the decision tree splitting process. As illustrated in Figure 8, repayment behaviour variables dominate the predictive process, with PAY_0 emerging as the most influential feature, followed by LIMIT_BAL and AGE. The prominence of PAY_0 highlights that recent repayment status is a critical indicator of default risk, consistent with earlier findings that defaulting customers exhibit persistent delays in repayment over time.

Furthermore, variables related to repayment amounts (PAY_AMT1-PAY_AMT6) and billing statements (BILL_AMT1) also contribute significantly to model predictions. These features capture borrowers' financial behaviour and repayment capacity, reinforcing the importance of behavioural financial indicators over static demographic attributes. In contrast, demographic variables such as education and marital status show relatively lower importance, indicating a secondary role in credit risk prediction. Overall, the feature importance analysis confirms that credit default is primarily driven by temporal repayment behaviour and financial management patterns, thereby justifying the use of models capable of capturing sequential dependencies.

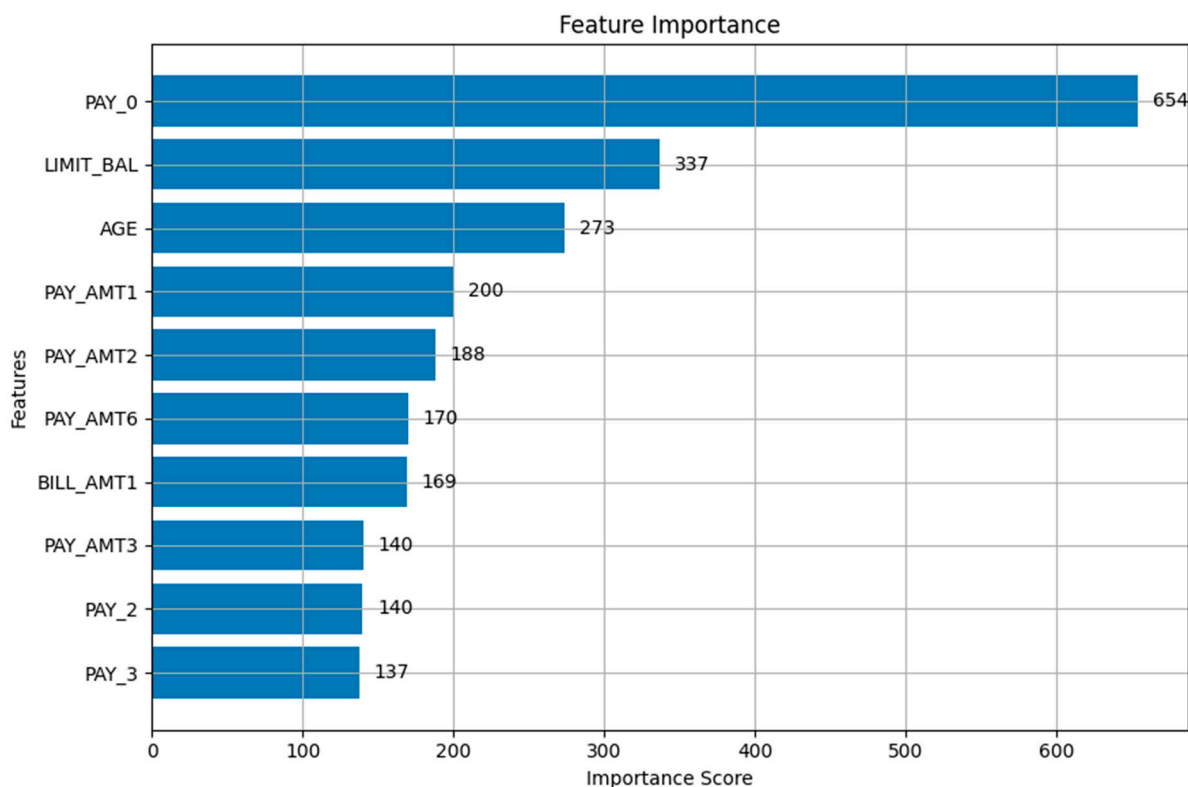


Figure 8. Feature Importance.

4. Discussion

4.1. Results Discussion

The empirical findings of this study demonstrate that the proposed DeepSurv-LSTM hybrid model consistently outperforms the individual baseline models (XGBoost, DeepSurv, and LSTM), both before and after Bayesian hyperparameter optimisation. The superior performance of the hybrid model achieving the highest values across Accuracy, Precision, Recall, F1-score, and ROC-AUC suggests that integrating survival analysis with temporal deep learning architectures provides a more effective framework for modelling credit risk. The results indicate that hybrid architecture is capable of capturing complex nonlinear relationships while simultaneously learning temporal patterns in the credit dataset, thereby improving predictive accuracy and classification reliability.

These findings are consistent with recent studies highlighting the advantages of hybrid deep learning models in financial risk prediction. For example, the results are consistent with Liu et al. (2022), who showed that hybrid frameworks combining XGBoost and graph-based neural networks improve classification performance by capturing complex data interactions. Li et al. (2023) demonstrated that combining CNN and LSTM architectures with an attention mechanism significantly improved credit risk prediction accuracy for listed companies, as the hybrid framework effectively integrates feature extraction with sequential learning. Similarly, Zhang et al. (2024) reported superior predictive performance using a Logistic-CNN-BiLSTM hybrid model for enterprise credit risk assessment, achieving very high sensitivity and specificity. These studies reinforce the observation that hybrid deep learning models can leverage complementary strengths of multiple architectures, resulting in improved predictive capability compared with standalone models.

Furthermore, the improvements observed after Bayesian hyperparameter optimisation highlight the importance of parameter tuning in machine learning models. This observation supports the findings of Kanojia and Arora (2025), who emphasised that model performance in credit risk

prediction is highly sensitive to hyperparameter configuration and feature dimensionality. By systematically exploring the hyperparameter space, Bayesian optimisation allows the models to achieve more stable and generalisable configurations, which contributes to the improved performance observed in the optimised models. Moreover, the importance of temporal modelling is supported by Yu and Fang (2025) and Ahmad et al. (2026), who demonstrate that hybrid sequential architectures significantly enhance credit risk prediction by effectively capturing temporal dependencies and complex feature interactions.

The results provide strong evidence that credit default is both a time-to-event phenomenon and a temporal behavioural process. While traditional machine learning models such as XGBoost capture static nonlinear relationships and LSTM models capture temporal dependencies, neither approach alone fully represents the complexity of credit risk. The proposed Bayesian-optimised DeepSurv-LSTM model addresses this limitation and consistently outperforms all baseline models across evaluation metrics. The confusion matrix indicates reduced misclassification rates, particularly in minimising false positives and false negatives, which are critical in credit risk assessment. Feature importance analysis further highlights repayment behaviour as the most influential predictor, underscoring the importance of temporal financial patterns. Overall, the findings confirm that integrating survival analysis with deep temporal learning provides a more robust and accurate framework for credit risk prediction.

4.2. Practical and Policy Implications

The findings of this study have important implications for financial institutions, policymakers, and risk management practitioners. The superior performance of the proposed DeepSurv-LSTM model suggests that integrating survival analysis with temporal deep learning can significantly enhance credit risk assessment frameworks. For financial institutions, this implies improved accuracy in identifying high-risk borrowers, leading to more informed lending decisions, reduced default rates, and enhanced portfolio management. The reduction in false positives and false negatives is particularly critical, as it minimizes both the rejection of creditworthy applicants and the approval of high-risk borrowers.

From a policy perspective, the adoption of advanced hybrid models can support the development of more robust and resilient financial systems. Regulators may benefit from incorporating such models into supervisory stress-testing frameworks and early warning systems to better monitor systemic risk. However, given the complexity of deep learning models, there is a need to balance predictive performance with transparency and interpretability to ensure compliance with regulatory requirements. Consequently, future implementations should consider integrating explainable artificial intelligence techniques to enhance model accountability and trust in real-world applications.

5. Conclusions

This study proposes a Bayesian-optimized Hybrid DeepSurv-LSTM framework for dynamic credit risk prediction by integrating survival analysis with deep learning methodologies. The model captures nonlinear hazard relationships and sequential borrower behavior, enhancing time-to-default risk modeling under censored financial conditions. Empirical results show that the framework outperforms traditional machine learning models and standalone deep learning approaches across all evaluation metrics, including Accuracy, ROC-AUC, C-index, and IBS. The model's performance highlights the effectiveness of combining DeepSurv survival modeling, LSTM temporal learning, and Bayesian hyperparameter optimization. These findings demonstrate the model's potential for improving credit risk assessment, loan screening, and data-driven financial decision-making. However, several limitations persist. The study relied on a single publicly available dataset, which may limit the generalizability of the findings. Additionally, the hybrid framework introduces increased computational complexity and reduced interpretability compared to conventional statistical models. Future research should validate the framework with multiple real-

world financial datasets, incorporate explainable artificial intelligence (XAI) techniques, and explore advanced architectures such as attention-based or transformer-based survival models to enhance predictive performance and interpretability.

6. Patents

This section is not mandatory but may be added if there are patents resulting from the work reported in this manuscript.

Author Contributions: Conceptualization, N.M.; methodology, N.M.; software, N.M. and T.L; validation, T.L; formal analysis, N.M; writing—original draft preparation, N.M.; writing—review and editing, N.M., T.L; supervision, T.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Available upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AUC	Area Under the Curve
DeepSurv	Deep Survival Model
EDA	Exploratory Data Analysis
LD	Linear dichroism
FN	False Negative
FP	False Positive
IBS	Integrated Brier Score
LSTM	Long Short-Term Memory
ROC-AUC	Receiver Operating Characteristic–Area Under the Curve
SMOTE	Synthetic Minority Oversampling Technique
TN	True Negative
TP	True Positive
XGBoost	Extreme Gradient Boosting

References

- Abi, R. (2025). Machine learning for credit scoring and loan default prediction using behavioral and transactional financial data. *World Journal of Advanced Research and Reviews*, 26(3), 884-904.
- Ahmad, A. Y., Shukla, M., & Ali, G. (2026). AHNet: Design and Execution of Adaptive Hybrid Network for Credit Risk Prediction using Spatio-Temporal Attention-based Convolutional Autoencoder Features in the Banking Sector. *Computational Economics*, 1-43.
- Ahmed, F., & Iqbal, A. (2025). The role of artificial intelligence in enhancing credit risk management: A systematic literature review of international banking systems. *Pakistan Journal of Humanities and Social Sciences*, 13(1), 478-492.
- Amarnadh, V., & Moparathi, N. R. (2023). Comprehensive review of different artificial intelligence-based methods for credit risk assessment in data science. *Intelligent Decision Technologies*, 17(4), 1265-1282.
- Chang, V., Sivakulasingam, S., Wang, H., Wong, S. T., Ganatra, M. A., & Luo, J. (2024). Credit risk prediction using machine learning and deep learning: A study on credit card customers. *Risks*, 12(11), 174.
- Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- Chhetria, E. S., Parajulib, R., & Sharma, G. (2024). Credit risk prediction by using ensemble machine learning algorithms. *Int. J. Res. Publ*, 147, 34-56.
- Dala, F. L., Esquivel, M. L., & Gaspar, R. M. (2025). Survival Analysis for Credit Risk: A Dynamic Approach for Basel IRB Compliance. *Risks*, 13(8), 155.

- D'Amato, A., & Mastrolia, E. (2022). Linear discriminant analysis and logistic regression for default probability prediction: the case of an Italian local bank. *International Journal of Managerial and Financial Accounting*, 14(4), 323-343.
- Dong, D., Lin, B., & Dong, X. (2024). Logistics financial risk assessment based on decision tree algorithm model. *Procedia Computer Science*, 243, 1095-1104.
- Dugar, M., & Alesh, A. (2023, December). Deep Learning for Predicting Credit Card Default. In *Machine Intelligence and Digital Interaction Conference* (pp. 87-94). Cham: Springer Nature Switzerland.
- Edunjobi, T. E., & Odejide, O. A. (2024). Theoretical frameworks in AI for credit risk assessment: Towards banking efficiency and accuracy. *International Journal of Scientific Research Updates*, 7(01), 092-102.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
- Han, X., Yang, Y., Chen, J., Wang, M., & Zhou, M. (2025). Symmetry-aware credit risk modeling: A deep learning framework exploiting financial data balance and invariance. *Symmetry*, 17(3), 341.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
- Kanaparthi, V. (2023), April. Credit risk prediction using ensemble machine learning algorithms. In *2023 International Conference on Inventive Computation Technologies (ICICT)* (pp. 41-47). IEEE.
- Kanojia, S., & Arora, A. (2025). Machine learning for credit risk management through cross-economy evidence in default prediction. *SN Business & Economics*, 5(12), 221.
- Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., & Kluger, Y. (2018). DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC medical research methodology*, 18(1), 24.
- Levy, A., & Baha, R. (2021). Credit risk assessment: a comparison of the performances of the linear discriminant analysis and the logistic regression. *International Journal of Entrepreneurship and Small Business*, 42(1-2), 169-186.
- Li, J., Xu, C., Feng, B., & Zhao, H. (2023). Credit risk prediction model for listed companies based on CNN-LSTM and attention mechanism. *Electronics*, 12(7), 1643.
- Liang, L., & Cai, X. (2020). Forecasting peer-to-peer platform default rate with LSTM neural network. *Electronic Commerce Research and Applications*, 43, 100997.
- Lin, M., & Chen, J. (2023). Research on credit big data algorithm based on logistic regression. *Procedia Computer Science*, 228, 511-518.
- Liu, J., Zhang, S., & Fan, H. (2022). A two-stage hybrid credit risk prediction model based on XGBoost and graph-based deep neural network. *Expert Systems with Applications*, 195, 116624.
- Madaan, M., Kumar, A., Keshri, C., Jain, R., & Nagrath, P. (2021, January). Loan default prediction using decision trees and random forest: A comparative study. In *IOP conference series: materials science and engineering* (Vol. 1022, No. 1, p. 012042). IOP Publishing.
- Muddu, G., Ganiyu, S. O., Ejidokun, A. O., & Aleshinloye, Y. A. (2026). Integrated data-driven credit default prediction in Uganda using machine learning models. *Journal of the Nigerian Society of Physical Sciences*, 2649-2649.
- Said, I., & Qu, Y. (2022, December). A Study on the Performance Comparison of Five Popular Machine Learning Models Applied for Loan Risk Prediction. In *2022 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 670-676). IEEE.
- Sarkar, R. (2026). A Systematic Review of AI-Driven Credit Risk Assessment Models in Commercial Banking (2018–2026). *American Journal of Interdisciplinary Studies*, 7(01), 459-495.
- Shang, L., Zhao, J., Li, G., & Zhang, X. (2025). Survival analysis in credit risk management: A review study. *Journal of Credit Risk*, 20(4), 59-83.
- Soni, U., Jethava, D. G., & Ganatra, A. (2024). Latest advancements in credit risk assessment with machine learning and deep learning techniques. *Cybernetics and Information Technologies*, 24(4), 22-44.
- Sujatha, R., Kavitha, D., Maheswari, B. U., & Ajay, K. G. (2025). Ensemble Machine Learning Models for Corporate Credit Risk Prediction: A Comparative Study. *SN Computer Science*, 6(5), 514.
- Tian, Z., Xiao, J., Feng, H., & Wei, Y. (2020). Credit risk assessment based on gradient boosting decision tree. *Procedia Computer Science*, 174, 150-160.

- Tian, Y., & Wu, Y. (2024). Systemic Financial Risk Forecasting: A Novel Approach with IGSA-RBFNN. *Mathematics*, 12(11), 1610.
- Wu, Z., Liu, R., Dai, J., & Luo, D. (2026). Multimodal Insights into Credit Risk Modelling: Integrating Climate and Text Data for Default Prediction. *arXiv preprint arXiv:2601.00478*.
- Yassin, A. A., Haleeb, A., Alnagar, D. K., Hussein, E. M., SidAhmed Mustafa, M., Ahmed Elsheikh, S. M., & Awad, W. (2024). Modeling Financial Risk Using Discriminant Analysis: A Predictive Approach. *Pakistan Journal of Life & Social Sciences*, 22(2).
- Yu, D., & Fang, A. (2025). Achieving credit risk prediction framework for Chinese CBECs: a hybrid CNN-BiLSTM-AM approach. *Electronic Commerce Research*, 1-24.
- Zhang, X., Ma, Y., & Wang, M. (2024). An attention-based Logistic-CNN-BiLSTM hybrid neural network for credit risk prediction of listed real estate enterprises. *Expert systems*, 41(2), e13299.
- Zhou, X., Zhang, W., & Jiang, Y. (2020). Personal credit default prediction model based on convolution neural network. *Mathematical Problems in Engineering*, 2020(1), 5608392.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.