Innovative Credit Risk Assessment System Using Artificial Intelligence

Quantong Fu*

Experimental Training Teaching Center, Yanshan college Shandong University of Finance and Economics, Jinan, Shandong 271199, China

* *Corresponding author. E-mail: Quantong_Fu07@outlook.com*

Received: Nov. 26, 2025; Accepted: Jan. 31, 2026

This research presents an explainable deep learning framework for credit risk assessment that integrates an unsupervised autoencoder with a GRU–LSTM hybrid model for sequential data classification. Traditional credit scoring systems face several challenges, including difficulty in capturing complex borrower behavioral patterns, methodological limitations, and a lack of explainability, which reduces their suitability for decision-making in dynamic financial environments. The proposed framework employs an autoencoder to preprocess data by reducing dimensionality and noise, while the GRU–LSTM architecture captures both short- and long-term dependencies in borrower behavior. The autoencoder performs dimensionality reduction by compressing high-dimensional input features into a smaller latent representation through an encoder network trained to reconstruct the original data with minimal information loss. This process removes redundant and noisy attributes while preserving the most informative patterns required for accurate credit risk classification. To support interpretable credit risk evaluation, SHapley Additive exPlanations (SHAP) is used to provide both local and global feature importance explanations. By quantifying the contribution of each input feature to individual predictions, the explainability component supports transparent credit decisions and enables financial institutions to justify automated outcomes within regulatory and auditing workflows. The framework was implemented using Python and evaluated on the German Credit dataset after preprocessing with one-hot encoding and Min–Max normalization. Experimental results demonstrate strong performance, achieving an accuracy of 99.12%, precision of 98.82%, recall of 98.78%, and an F1-score of 98.851. These findings indicate that the framework provides an accurate, explainable, and real-time approach to credit risk assessment in institutional financial settings.

Keywords: Credit risk assessment; Deep learning; Autoencoder; GRU-LSTM; Explainable AI; SHAP; Feature extraction; Classification

http://dx.doi.org/10.6180/jase.202608_31.036

1. Introduction

In today's financial systems, credit lending is still one of the cornerstones of economic growth, allowing people, small businesses, and corporations to obtain quickly needed funding [1] and the credit evaluation process is an important part of ensuring financial stability, limiting default situations, and ensuring lending organizations can continue to make money [2]. But determining if a borrower has an acceptable credit risk is a complicated task because there is a myriad of factors that go into assessing a borrower's financial situation [3] (income level, historical repayment behaviours, what the loan is for, employment status, displacement, etc.). Credit market offerings continue to grow in scale and complexity, making traditional assessment approaches less adaptable. As lending volume increases, institutions require faster and more accurate mechanisms for evaluating borrower risk [4].

Conventional methods for credit risk assessment are based on linear statistical methods or manual rules to make judgment decisions [5]. Typical models often make the assumption that relationships between inputs are static and independent, which fails to sufficiently consider the complicated nature of borrower profiles [6]. Also, many credit scoring models are unable to comprehend nonlinear interactions between features or adjust to fluctuations in borrower [7] behaviour that can occur due to economic downturns, pandemic closures, or regional turmoil [8]. The fact that many of these scoring models are hampered by some combination of incomplete or unbalanced historical data, missing values, and human bias raises their reliability [9] of credit-worthiness, and the modelling system can flag individuals that may otherwise be credit worthy into an incorrectly categorized high risk or reject them as an applicant altogether [10], while approving applicants who may present various historical or future risk patterns that may be undetected.

In response to these challenges, many ML [11] approaches have been studied in the credit scoring context, including Support Vector Machines (SVM) [12], Decision Trees, Naive Bayes, K-Nearest Neighbours (KNN), and ensemble techniques, including Random Forest and XGBoost. These models have documented progress over traditional rule-based models through learning from data, modelling feature interactions, and automated feature interactions [13]. Nevertheless, upfront, most ML models learn from handcrafted features and static inputs, which restricts their quality for modelling complex temporal relationships between financial behaviours [14]. Furthermore, many of the models use opaque black boxes, limiting their infield usability in regulated conditions with a requirement for auditability, transparency, and interpretability [15]. Recent studies highlight the growing importance of explainable models in credit risk assessment, demonstrating how integrated interpretability tools enhance regulatory transparency and improve trust in automated decision systems [16].

To overcome the limitations of traditional and shallow machine learning models, this research presents a deep learning-based, intelligent, and explainable framework for credit risk classification. The framework begins with an autoencoder implemented as a precursor for extracting features, including discarding unnecessary dimensions and noise in borrower data, whilst keeping informative structure. The compressed latent vectors are then brought to a GRULSTM hybrid architecture model that combines Long-term dependency in addition to immediate signals in borrower behaviour, helping credit risk classification results to be more reliable. The final output from the system is a binary risk classification of whether a borrower is low-risk or high-risk. To provide clarity in decision-making, the framework employs SHAP (SHapley Additive exPlanations) to explain features interpreted in local and global ways by calculating the contribution of each feature to the model's prediction. This end-to-end architecture is configured to run real-time on financial platforms to provide a scalable, adaptable, and explainable method suitable for current credit risk analysis.

Recent research in explainable credit risk modelling has introduced diverse approaches that integrate interpretability with machine learning, including hybrid feature-learning models, transparent boosting techniques, and SHAP-based explanation frameworks. However, most studies rely either on static feature representations or single-stage classifiers, limiting their ability to capture evolving behavioural patterns while maintaining interpretability. The proposed framework differs by combining autoencoder-driven feature compression with a hybrid GRU-LSTM architecture to model both short-term and long-term borrower dynamics, while SHAP is applied directly to reconstructed feature attributes to preserve domain relevance. This integrated design positions the model within the current explainable AI landscape while advancing temporal modelling and interpretability beyond existing approaches.

1.1. Research Contributions

This research proposes a method that utilises a deep learning credit risk assessment model which improves both classification accuracy and transparency. The contributions are:

- New deep learning framework including a combination of autoencoder and GRULSTM for accurate classification of credit risk.
- Proposed an unsupervised autoencoder to seek compressed and noise-free borrower features to use as an efficient model input.
- Developed a hybrid GRU-LSTM model to take advantage of sequential dependencies in borrower behaviour, which improves prediction.
- Added SHAP explainability to visualise individual and global feature-level reasoning for model decision making to improve transparency.
- Leveraged the German Credit dataset, which included applying encoding, Min-Max normalising, and reshaping to input data to a deep learning model.

1.2. Research Organizations

The organization of the document is as follows:

- section 1 presents background and motivation for credit risk assessment, and discusses the weaknesses of traditional methods and the reasons for employing a machine learning and explainable AI-based approach to credit risk classification.
- section 2 reviews credit risk classification methods, beginning with traditional rulebased models, covering machine learning classifiers, and the recent advancements in deep learning and explainability.
- section 3 explains our method, which includes an outline of our preprocessing steps, our use of autoencoders for feature extraction, the design of our GRU-LSTM classification model, and our use of SHAP for providing interpretable classification results.
- section 4 shares the Experimental setting and assessment of the results for analysing the recommended model's ability to classify using the German Credit dataset alongside other traditional approaches.
- section 5 discusses the study, including an explanation of the contributions made as well as suggestions for additional investigation, such as accounting for inputs as they happen in real time, using a transformer-based classification model or deploying our model on edge devices to allow scaling for credit classification.

2. Literature survey

The paper proposes a credit risk assessment system for commercial banks referred to as Automated Credit Risk Assessment using a Back Propagation Neural Network (BPNN). The model takes into account qualitative and quantitative analysis, and aims to be more directional, objective, and effective in evaluating loan risks, while improving credit management decisions by Liu [17]. Biswas et al. [18] suggest a fully automated credit risk evaluation system, facilitated with machine learning, with an Extract, Transform, and Load (ETL) process that adheres to Basel II. The authors used models to measure credit risk, comprising exposure at default, loss given by default, and the probability of default. Yu [19] proposed a credit risk assessment using a multilayer perceptron (MLP) neural network model along with principal component analysis for feature extraction, achieving prediction accuracy of 80.1%. This would enable commercial banks to operate effectively in determining the credit risk of borrowers. The system to assess credit risk arising from non-payment of electricity charges was made by configuring model data, developing an index system on payment ability/willingness, and applying machine learning methods for the assessment of enterprise credit risk and the development of risk response strategies by Hu et al. [20].

Tolulope Esther Edunjobi & Opeyemi Abayomi Odejide [21] discuss AI frameworks in credit risk assessment, utilizing machine learning strategies, neural networks, and natural language processing to enhance the efficiency and accuracy of the processes. Such systems sift through very large datasets to identify patterns to facilitate the proactive management of risk and the making of well-informed lending decisions. Coşkun and Turanli [22] utilize the boosting codes, CatBoost-XGBoost-LightGBM, for evaluation of credit risk based on the House Financial Default Risk dataset on Kaggle for loan applicants' capacity to repay, using machine learning along with parameter tuning for maximum model output. Rao et al. [23] present the assessment of credit risk system for personal auto loans using a PSO-XGBoost model that combines

Particle Swarm Optimization with eXtreme Gradient Boosting to enhance classification performance and manage credit risk more efficiently in automotive financing [23].

Luo et al. [24] build a credit risk assessment system for a supply chain using financial and supply chain data from China's new energy electric vehicle industry, with the application of PCA in data preprocessing and AdaBoost-aided-optimised SVM to further improve the whole process. The system for assessing credit risk of commercial banks was constructed by integrating Genetic Algorithm Neural Network (GANN) and cluster analysis, emphasising a complete index system inclusive of both quantitative and qualitative indicators for the proper evaluation of credit risk by Bai and Zha [25]. Bone-Winkel & Reichenbach has generated an explainable system for machine learning-based survival analysis in credit risk assessment, in P2P lending on Bondora data. Consequently, the system has been improving risk ratings, hence decision-making, thus fair interest rates supported by an accurate and transparent credit risk assessment system [26].

2.1. Problem Statement

Numerous automated systems have been developed to enhance credit risk assessment across various financial domains, utilising multiple techniques for deep learning as well as machine learning [27]. These systems aim to replace traditional manual evaluation processes by integrating both qualitative and quantitative borrower attributes, improving objectivity and consistency in loan risk evaluations

[28]. Several approaches leverage neural networks such as multilayer perceptrons and backpropagation models, often combined with dimensionality reduction techniques like principal component analysis to improve classification performance [29]. Other frameworks incorporate advanced methods like boosting algorithms (e.g., CatBoost, XGBoost, LightGBM), hybrid models combining Particle Swarm Optimisation with gradient boosting, and optimised support vector machines to manage credit risk in diverse sectors such as commercial banking, automotive financing, energy industries, and utility billing [30]. Furthermore, some models introduce explainability and feature importance analysis to support transparent decision-making and fair lending practices. However, many existing systems cannot generalise across varied borrower profiles, struggle with noisy input features, or offer limited interpretability, highlighting the need for more robust, adaptive, and explainable credit risk classification frameworks [31].

3. Methodology

The initial preprocessing pipeline applies one-hot encoding to categorical variables and Min-Max normalization to numerical attributes to ensure a uniform input scale for the deep learning architecture. These steps are performed before feeding the data into the autoencoder so that both the encoder and GRU-LSTM classifier operate on consistent and noise-reduced representations. This concise preprocessing ensures stable model convergence and minimizes feature-range imbalance.

The proposed method demonstrates a comprehensive deep learning framework for binary classification of risk for credit using the German Credit dataset. Data preprocessing for the deep learning model involved encoding the categorical features as numbers and normalizing the numerical features using Min-Max scaling. First, the model makes use of an unsupervised autoencoder to extract compact informative features, while also reducing dimensionality and noise. Next, these compact informative features are reshaped and provided as input to the GRU-LSTM classifier that captures temporal dependencies between borrowers to inform the classification outcome. The output of the model is a binary label indicating the credit risk status of the borrower. Additionally, SHAP was implemented to define the way every input characteristic contributes, prepared using feature engineering to assist in the classification conclusion, thus allowing the model an explanation path. The method presented above provides a simple solution for scalable and interpretable implementation of credit assessment systems. Fig. 1 represents as begins with raw credit data and applies preprocessing procedures of onehot encoding and normalization. The pre-processed data is then put through a feature extraction step via an autoencoder to extract relevant patterns. The extracted features are passed through a classification module utilizing GRU and LSTM layers to learn temporal and sequential dependencies in the data, where it outputs a classification of whether the individual is good or bad credit risk. The model's decisions, as a whole, are made interpretable using SHAP.

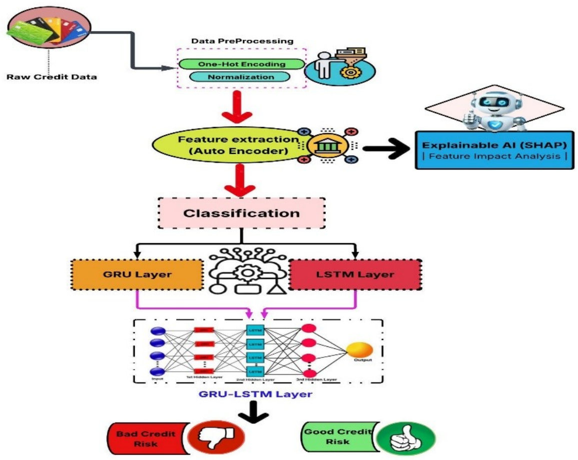

Fig. 1. Credit Risk Assessment System Proposed Methodology Architecture

3.1. Data Collection

The German Credit dataset, which is utilized within this study, is obtained from Kaggle and contains 1,000 borrower samples with binary credit risk labels, classified as good or bad. Each sample contains a total of 20 attributes: both categorical and numerical features. attributes included are age, amount of credit, time (or duration) of loan, employment, housing, savings account status, and the aim of taking out the loan. The dataset incorporates real-world variability, including differences in their demographics, finances, and behavioural aspects; hence, it is appropriate to use in developing and testing models for estimating credit risk. Before the data enters model training, it is pre-processed via encoding the categorical features and normalization of the numerical features to create consistency for the deep learning input [32]. The dataset used in this study originates from the publicly available German Credit dataset on Kaggle, which contains 1,000 borrower records with binary risk labels. The dataset can be accessed directly from the Kaggle repository by searching for "German Credit Data," ensuring full reproducibility of the experimental setup without requiring additional proprietary inputs.

3.2. Preprocessing Stage

This Process is a vital step in transforming the dataset to enable deep learning-based classification methods on the German Credit dataset. This requires either label encoding or one-hot encoding to convert categorical variables like housing status, job type, and reasons for the loan to a numerical format. In addition, numerical attributes like credit amount, age, and loan duration need normalization by Min-Max scaling to minimize the effects of range differences, so that all values share a consistent range of $[0, 1]$. This minimizes the likelihood that a single feature dominates the model. Data quality checks include missing values and inconsistencies. The result is a cleansed dataset and a normalized dataset before formatting the dataset for both the autoencoder and the GRU-LSTM network.

3.2.1. *One-Hot encoding*

This study employs one-hot encoding to the nominal categorical variables of loan purpose, housing type, and job status so that they can be treated as inputs in deep learning algorithms. One-hot encoding converts A vector of binary values of length k representing a grouping characteristic with distinct groups, wherein there is one value of 1 for a particular category and all other values are 0. One-hot encoding preserves the non-ordinal nature of the data and does not create artificial ordinal relationships within non-ordinal data. Formally, for a categorical variable x with categories $\{c_1, c_2, \ldots, c_k\}$, the one-hot encoding of an instance $x = c_j$ is defined as:

$$\text{OneHot}(x) = [0, 0, \ldots, \underbrace{1}_{j^{\text{th}}\ \text{position}}, \ldots, 0] \in \mathbb{R}^k \tag{1}$$

For example, if the feature Loan Purpose contains three categories: Education, Furniture, and Car, and $x =$ Furniture, then:

$$\text{OneHot(Furniture)} = [0, 1, 0] \tag{2}$$

This representation ensures that the neural network treats all categories as distinct and nonhierarchical. This is crucial to maintaining minimal information integrity. Unlike label encoding, which assigns integer values that may falsely imply a ranking or order, one-hot encoding assigns a binary vector where only one position is marked as active (1) and all others are inactive (0). This prevents the model from interpreting categorical relationships incorrectly, such as assuming that one category is greater or lesser than another. By preserving the true nature of categorical variables, one-hot encoding enables the model to learn more accurately from the data, ultimately enhancing the classification performance.

3.2.2. *Min-Max Normalization*

To ensure this technique is applied to rescale the values of every feature into a common range of $[0, 1]$ to be sure that all numerical features have equal contributions during model training. It is essential to note that deep learning models can be highly sensitive to features with large numeric ranges; therefore, features such as credit amount or loan duration should not dominate the learning process. The Min-Max normalization formula is expressed as:

$$x_{\text{normalized}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{3}$$

Where x is the original feature value, $x_{\min}$ and $x_{\max}$ are the minimal and the maximum values for features throughout the dataset, and $x_{\text{normalized}}$ is a scaled output value. This transformation utilizes an arbitrary amount of information about the data and maps arbitrary feature values on some number of scales to a uniform range from 0 to 1. This means that the learning processes aren't being dominated by just one feature, just because it has a different range than others. Normalizing the range of input can optimize the learning procedure so that organized learning and biasing occur over all features. The result is faster convergence when training the neural network and less instability for the models. Ultimately, Min-max normalization improves classification performance.

3.3. Feature Extraction Using Autoencoder

A deep autoencoder is used to gain a compact and informative feature embedding from high-dimensional borrower data. Essentially, and fundamentally, a deep autoencoder extracts patterns from data and removes redundancy and noise. There are two essential parts to the deep autoencoder: a decoder, which duplicates from an original information, and an encoding process, which transforms the input data into a low-dimensional latent representation based on one or more latent representations. The encoder and decoder are trained jointly by minimizing reconstruction loss using MSE. After training, only the encoder is retained to produce compact embedding vectors for classification. The feature vectors serve as input to the GRU-LSTM model for classification. This preprocessing step contributes to reduced training time, improved avoidance of overfitting, and enhanced overall system performance. The autoencoder architecture is shown in Fig. 2. The information provided is mapped to a set of important features in the bottleneck stage, and then the decoder can reconstruct the original input data based on these features. The overall goal of an autoencoder is to efficiently represent and reduce dimensionality.

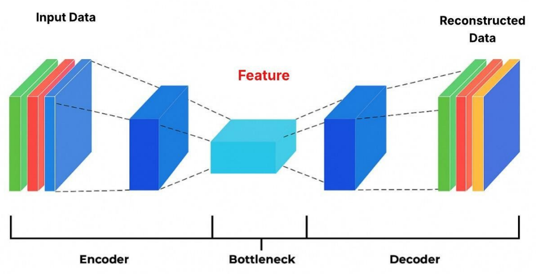

Fig. 2. Autoencoder using feature extraction architecture

3.3.1. Encoder

Reducing the highly dimensional information entered into an efficient hidden representation is the encoder's job. Let an input feature vector be $x \in \mathbb{R}^n$. The encoder applies a nonlinear transformation f to transform the data provided into a latent space with fewer dimensions $z \in \mathbb{R}^m$, where $m < n$:

$$z = f(x) \tag{4}$$

This change makes the autoencoder preserve the necessary borrower data. It eliminates excess noise and redundancy that might undermine the classification. The latent features are, therefore, more meaningful and more compact. These features improve both the speed and accuracy of evaluating the GRU-LSTM credit risk classifier.

3.3.2. Decoder

The decoding process makes an attempt to replicate the initial input from the latent vector z using another nonlinear transformation g, producing an output $\hat{x}$:

$$\hat{x} = g(z) = g(f(x)) \tag{5}$$

Minimizing the reconstructed loss between the first input x and the autoencoder's goal, the reconstructed output $\hat{x}$, typically using MSE:

$$\mathcal{L}(x, \hat{x}) = \|x - \hat{x}\|^2 \tag{6}$$

After training the autoencoder, keep only the encoder to extract compressed, informative features from the raw credit data. The features provide a less redundant and less dispersed representation of the input data, highlighting important patterns, while maintaining the necessary features and quality without introducing noise that clouds judgment. The extracted features were used instead of the original high-resolution credit data, in which the hybrid GRULSTM classifier was fed the features. This hybrid GRU-LSTM classifier takes advantage of both GRU and LSTM suites' ability to learn time dependencies because credit data has a time dependency. The classification model, which utilized the encoder's output, was significantly more effective and accurate in predicting credit risk.

The autoencoder architecture used in this study consists of a three-layer encoder and a symmetric three-layer decoder, with decreasing and increasing unit sizes respectively to ensure stable reconstruction. The encoder compresses the input to a latent dimension of 32 units, selected after evaluating dimensions between 16 and 64 and observing that 32 offered the best balance between reconstruction accuracy and classifier performance. The bottleneck dimension of 32 was selected based on experiments evaluating reconstruction loss, downstream classification performance, and retention of minority-class characteristics. Latent representations were analyzed to ensure that patterns indicative of high-risk borrowers, including less frequent repayment irregularities, were preserved despite dimensionality reduction. Comparison of classification metrics using alternative latent sizes $(16, 32, 64)$ indicated that 32 units maintain discriminatory power while eliminating redundant information, effectively balancing compression with the need to capture minority-class variance. This ensures that the autoencoder does not suppress signals relevant for identifying borrowers with high default risk, preserving interpretability and accuracy in the subsequent GRU-LSTM classifier.

This latent size is sufficiently compact to remove redundancy while retaining the essential borrower characteristics required by the GRU-LSTM model. The architectural depth and bottleneck size therefore support reproducibility while clearly defining the autoencoder's role as a noise-reducing feature compressor within the overall classification pipeline.

The encoder layers use ReLU activation to enable nonlinear feature compression and support stable gradient flow during training, while the decoder employs a sigmoid activation to reconstruct normalized feature values in the $[0, 1]$ range. Within the classification module, the GRU units apply tanh for candidate state formation and sigmoid functions for update and reset gates, enabling efficient short-term dependency modelling. The LSTM layer similarly uses tanh for generating candidate cell states and applies sigmoid activation across its input, forget, and output gates to regulate long-term memory flow. The final dense layer incorporates a sigmoid activation to produce a probability score representing the borrower's credit risk class.

These activation choices ensure consistent non-linearity, stable training dynamics, and reproducibility of the model architecture.

3.4. Classification Using GRU-LSTM

The classification in this study employs a GRU-LSTM hybrid neural network to subsequently learn it documented both brief and expanded encounters in the borrower's data. after encoding. The GRU layer processes the sequential input first by allowing information and data transfer via updates and reset gates, thereby learning recent credit behaviours efficiently. The LSTM layer is the subsequent component that learns longer-term experiences and is controlled by input as well as the output and forget gates, allowing the model to discover deeper temporal trends in the violations of financial activities. The stacking of GRU and LSTM creates a stronger capacity of temporal feature extraction and classification. The final dense layer with a sigmoid activation produced the binary outputs classifying the borrower as either 'good risk' or 'bad risk ' for credit. The overall architecture of the proposed credit risk assessment framework is illustrated in Fig. 3.

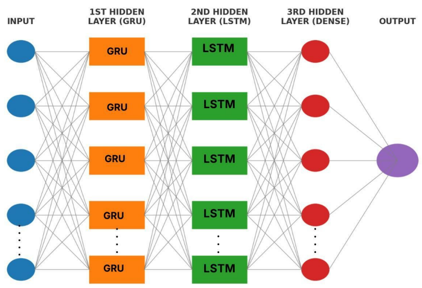

Fig. 3. GRU-LSTM Model

The GRU layer is placed before the LSTM layer to promote stable gradient propagation across heterogeneous gating mechanisms. GRU units, with their reduced gate structure, act as an initial temporal filter that smooths short-term fluctuations and suppresses high-variance feature transitions. This preprocessing effect reduces gradient noise before signals reach the LSTM layer, whose memory cells are designed to preserve long-range dependencies but are more sensitive to unstable activations. By ordering the GRU prior to the LSTM, the network benefits from faster convergence during early training stages and improved gradient flow during backpropagation, mitigating vanishing and exploding gradient effects while maintaining long-horizon memory retention.

The recurrent modeling strategy was selected over attention-centric tabular architectures due to the structural characteristics of the credit dataset and the intended deployment constraints. While Transformer-based models excel in high-dimensional sequential contexts, they typically require substantially larger training corpora and impose higher computational overhead, which can limit stability and efficiency on moderate-sized financial datasets. In contrast, GRU-LSTM networks effectively capture ordered inter-feature dependencies with fewer parameters and lower inference latency, making them more suitable for structured borrower attributes where relational continuity is implicit rather than explicitly temporal. This choice ensures robust learning behavior under limited data regimes while maintaining practical feasibility for real-time credit risk assessment systems.

3.4.1. *Input Layer*

The GRU-LSTM architecture is selected because borrower financial behavior evolves over time and exhibits both short-term variations (such as recent repayment activity) and long-term patterns (such as sustained income stability or debt accumulation). Combining GRU and LSTM layers enables effective modeling of these sequential dependencies, improving the reliability of credit risk prediction. The GRU-LSTM hybrid model introduced in this framework for credit risk assessment aims to organize, classify, and predict borrower profiles using structured tabular data. Each borrower's data consists of normalized numerical variables (e.g., age, credit amount) and encoded categorical characteristics (e.g., job type, the information being collected with the goal of reducing the number of dimensions purpose of loan). Before use with the classifier, an autoencoder is applied to the information for the purpose of reducing the total number of dimensions and noise. The autoencoder provides a compressed representation of the latent space. The latent space representations are then reshaped into time-series sequences to learn structural patterns across which finance-related features were previously learned. By modelling the input as a temporal sequence, the model can learn about the structured behaviour of borrower profiles:

$$X = \{x_1, x_2, x_3, \ldots, x_T\} \tag{7}$$

Where $X \in \mathbb{R}^{T\times d}$, T is the sequence length (time steps), d is the number of features at each step, and $x_t \in \mathbb{R}^d$ is the input at time t.

Because the German Credit dataset contains static borrower attributes rather than true timeseries information, temporal sequences were generated by restructuring the compressed autoencoder embeddings into fixed-length segments. This reshaping assumes that different portions of the latent feature vector can be interpreted as sequential representations of borrower characteristics, allowing the GRU and LSTM layers to capture structured patterns within the feature space. Each compressed feature vector was divided into equally sized timesteps, forming a synthetic sequence that preserves feature relationships while enabling the model to learn temporal-style dependencies. This approach follows the common practice of applying recurrent architectures to static tabular data by treating latent dimensions as ordered feature groups. The transformation from static borrower descriptors to ordered learning signals is formalized through a deterministic feature-grouping strategy applied to the autoencoder latent vector. Specifically, compressed features are partitioned into fixed-length segments corresponding to logically related borrower attributes, including financial capacity, credit history, employment stability, and demographic indicators. These segments are arranged in a consistent order to construct synthetic temporal sequences, ensuring that semantically similar attributes occupy adjacent timesteps. This ordering preserves interpretive coherence between feature semantics and the GRU-LSTM gating mechanism, enabling the recurrent layers to model structured inter-feature dependencies while maintaining stable memory activation patterns.

The application of recurrent inference to cross-sectional financial records relies on the assumption that borrower attributes encode stable behavioural tendencies over moderate time horizons, such as repayment discipline, income regularity, and credit utilization patterns. Although the dataset does not contain explicit temporal indices, these attributes collectively represent an implicit behavioural trajectory rather than isolated observations. By transforming structured feature groups into ordered sequences, the model captures continuity in latent financial behaviour rather than chronological progression. This assumption is consistent with prior credit modelling studies that treat borrower profiles as quasi-temporal representations of financial stability and risk evolution.

3.4.2. GRU Layer

The GRU layer captures short-term feature dependencies in borrower behavior. It introduces a gating mechanism to retain useful recent information and mitigate the vanishing gradient problem.

(i) Update Gate (z_t)

The quantity of the previous hidden state h_{t-1} that must be carried over to the next time step is calculated by this gate's value. It helps retain useful historical information.

$$z_t = \sigma\left(W_z x_t + U_z h_{t-1} + b_z\right) \tag{8}$$

(ii) Reset Gate (r_t)

The reset gate determines how much of the past information to forget. A lower reset value means present input is more significant to the framework x_t.

$$r_t = \sigma\left(W_r x_t + U_r h_{t-1} + b_r\right) \tag{9}$$

(iii)Candidate Activation $\left(\tilde{h}_t\right)$

This represents the new memory content, which is created through integrating a previous hidden condition with the subsequent stage. It uses the reset gate to control influence from the previous state h_{t-1} :

$$\tilde{h}_t = \tanh\left(W_h x_t + U_h\left(r_t \odot h_{t-1}\right) + b_h\right) \tag{10}$$

Where, x_t is the input at time step t, h_{t-1} is the previous hidden state, r_t is the reset gate controlling which parts of h_{t-1} to forget, $\odot$ denotes element-wise multiplication, W_h and U_h are weight matrices for the input and hidden state, b_h is the bias vector, tanh is the hyperbolic tangent activation function that squashes the output to the range $[-1, 1]$. This equation computes a candidate state $\tilde{h}_t$, which used later to discover the new hidden state h_t, incorporate it with the prior state.

(iv)Final Hidden State (h_t)

The final GRU output is a linear interpolation between the past hidden state h_{t-1} and the candidate state $\tilde{h}_t$, weighted by the update gate:

$$h_t = \left(1 - z_t\right) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{11}$$

This allows a network to maintain continuity of relevant features across short borrower history windows. It enables dynamic learning of borrower behavior fluctuations such as changes in financial status or short-term risk signals.

The GRU layer is positioned before the LSTM layer to introduce a progressive temporal abstraction strategy within the classification pipeline. GRU units, due to their simplified gating structure, efficiently capture short-term behavioural fluctuations and rapidly filter out redundant temporal signals. Passing these refined representations into the subsequent LSTM layer enables the model to then extract longer-term behavioural dependencies using its dedicated memory cell architecture. This ordering ensures that the LSTM receives a cleaner and more structured temporal input, allowing it to focus on gradual behavioural trends rather than noisy short-term variations. Placing GRU first and LSTM second therefore produces a hierarchical temporal learning process in which immediate borrower patterns are processed early, and deeper temporal relationships are captured later.

(v) LSTM Layer

The LSTM layer captures long-term dependencies and helps the model not just capture a gradual shift in a borrower's behavior, such as their ability to repay their loan deteriorating. The omit, input, and output gates are used to manage memory on memory updates. The main operations are: The forget gate is equated as.

$$f_t = \sigma\left(W_f\left[h_{t-1}, x_t\right] + b_f\right) \tag{12}$$

In this equation, W_f is the matrix of weight, b_f is the bias, and σ denotes that sigmoid activation. Simultaneously, the input gate of the Eq. (12) is expressed as:

$$i_t = \sigma\left(W_i\left[h_{t-1}, x_t\right] + b_i\right) \tag{13}$$

It controls how much of the new borrower input x_t is used to update the memory. The candidate cell state $\tilde{C}_t$ represents a new calculated memory content that can potentially be stored into the internal state of the LSTM. The candidate cell state is activated using a tangent hyper function that produces values in the range of -1 and 1, allowing the potential representation of various borrower updates, whether they have positive or negative impacts on behavior and trends. The equation is expressed by,

$$\tilde{C}_t = \tanh\left(W_C\left[h_{t-1}, x_t\right] + b_C\right) \tag{14}$$

The cell state C_t is updated by combining stored information for that past with newly chosen borrower features, allowing the model to keep a working memory of longer-term financial behavior as well as recent credit behavior. Applied to credit risk classification, it can help recognize changing risk signals in relation to the borrower, such as increasing debt, improved repayment behavior, etc. The equation is:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{15}$$

This process allows dependence over time to be maintained by the LSTM in borrower behavior (to be used in determining how to identify slowly developing credit risk over time). The output gate is expressed as,

$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \tag{16}$$

It allows the LSTM to determine the degree how which the following hidden status has been affected by its current memory. It also allows the LSTM to determine how much of incoming financial data to retain based on relevant financial patterns such as consistent late payments and moments of overusing credit - only so much of debtor behavior that we consider appropriate at each state. The output gate brings this together to build a model that can more accurately classify borrowers as either high risk or low risk, then establishes the degree to which this internal condition is revealed, and the currently active time step's hidden condition is generated.

$$h_t = o_t \odot \tanh\left(C_t\right) \tag{17}$$

The final result has an activation function that is sigmoid and is a dense (fully connected) layer. It maps the final hidden state to a probability score $\hat{y} \in [0, 1]$, representing the likelihood of the borrower being classified as high-risk:

$$\hat{y} = \sigma(W \cdot h + b) \tag{18}$$

A threshold (typically 0.5) is applied:

- If $\hat{y} \geq 0.5$, the borrower is labeled bad credit risk (1)
- If $\hat{y} < 0.5$, the borrower is labeled good credit risk (0)

The end product of the proposed framework for credit risk classification is a probability score produced by a sigmoid-activated Dense layer after the GRU-LSTM network. This probability score falls within a range of 0 to 1 and indicates the probability of a borrower's designation as "bad" credit risk. A threshold is applied to transform this probability into a class label, with a common threshold of 0.5 - borrowers with a score greater than or equal to 0.5 are classified as bad credit risk, while borrowers who received a score of less than 0.5 are classified as good credit risk. This applied classification process is the final step of a deep learning pipeline, beginning with feature extraction via autoencoders, progressing to temporal modeling in the GRU and LSTM layers, and concluding with the implemented application of the classification. This risk designation is useful for making decisions, as most of the time the classification process is instantaneous and provides a method of operationalizing risk designation with acceptable consistency based on the needs and processes of a financial institution.

To ensure stable training, hyperparameters for the GRU-LSTM classifier were tuned through a series of controlled trials using validation performance as the selection criterion. The Adam optimizer was evaluated with learning rates between 0.0005 and 0.002 , where 0.001 showed the most consistent convergence. Batch sizes of 16, 32, and 64 were tested, with 32 providing the best balance between stability and computational efficiency. Epoch counts were explored in the range of 100-250, and 200 epochs were selected based on minimal validation loss fluctuation. These tuning steps establish a clear baseline for replicating the model configuration.

The hybrid GRU-LSTM design was selected because each recurrent unit contributes distinct strengths that are beneficial for credit risk modelling. GRU units efficiently capture short-term behavioural variations, such as recent repayment activity or abrupt changes in financial status, due to their simplified gating mechanism and faster convergence. In contrast, LSTM units retain longer-term behavioural patterns, including sustained credit utilization or gradual shifts in financial stability, through their explicit memory cell structure. Using both mechanisms in a stacked configuration allows the model to learn complementary temporal dynamics that neither unit can fully capture alone. This synergy improves the representation of borrower behaviour patterns and yields more stable and accurate classification outcomes compared to using GRU- or LSTM-only architectures.

3.5. Explainable AI (XAI) using SHAP

To support transparency and interpretability in credit risk classification, the study incorporates SHAP (SHapley Additive exPlanations), as SHAP is the explainable AI that is built into the study as the main explainable AI tool. SHAP supports both local and global explanations by providing an importance value for each data feature, depending on the effect it adds to the system's final classification output. SHAP and its transparency in explaining decisions are especially important for financial applications because financial institutions are expected to justify decisions that the borrower as "good" or "bad". SHAP values are calculated for each prediction, which allows model users to see how features such as credit amount, job status, or duration affect a borrower's classification. By presenting these amounts of predictive feature contributions, the system is not functioning as a black box, which supports ethical and transparent credit decisions.

In this model, SHAP is used after the GRU-LSTM model has classified the borough. The encoded feature vectors generated by the autoencoder, and input into the GRU and LSTM temporal layers, can be interpreted using SHAP, to display which features were most important in the classification of the borrower credit categories. Hence, lenders, regulators, etc., can be able to backtrack the actual decisions based on key features representing points in the data to ensure regulatory compliance, fairness, and accountability. Furthermore, SHAP helps in identifying strong features of risk, which improve model debugging when it needs to be refined to improve decision making, and the reduction of bias in the classification. In conclusion, SHAP allows this model to not only invoke more interpretable computer output, but it also engages the user to trust and use the credit risk classification system in their real-world banking space.

SHAP explanations in this study were generated on the reconstructed feature space corresponding to the original borrower attributes rather than on the latent autoencoder representations. Although the compressed latent vectors are used internally for GRU-LSTM classification, SHAP values were mapped back to the original input features to ensure that the explanation outcomes remain interpretable for financial analysts and regulatory review. This approach enables each credit attribute-such as credit amount, loan duration, checking account status, or employment category-to be directly associated with its contribution to the final risk classification, maintaining alignment with practical decision-making requirements.

The proposed framework is designed to support real-time deployment by separating offline training from online inference. Feature encoding and model optimization are performed offline, while the trained autoencoder and GRU-LSTM classifier operate in a lightweight inference mode during deployment. For each incoming borrower record, preprocessing and prediction require only a single forward pass through the network, enabling millisecond-level response times on standard CPU-based systems. This design allows integration with digital lending platforms and banking decision engines without introducing significant latency or computational overhead. Although the GRU-LSTM architecture captures complex nonlinear feature interactions through its gated memory mechanisms, SHAP explanations approximate these effects using additive feature attributions. This decomposition does not explicitly model higher-order interactions but instead distributes their influence across contributing attributes based on marginal contribution estimates. As a result, strongly coupled financial variables, such as income stability and credit duration, may appear as independent drivers in the explanation layer despite their joint influence during internal representation learning. Nevertheless, this additive formulation remains suitable for financial decision support, as it preserves directional influence, relative importance, and regulatory interpretability while providing a tractable abstraction of nonlinear model behavior. Although the GRU-LSTM classifier operates on latent embeddings generated by the autoencoder, SHAP values are computed after mapping the latent representations back to the original input attributes. This ensures that each credit feature-such as loan amount, duration, or checking account status-is directly associated with its contribution to the predicted risk score. By projecting the explanations onto the original feature space, the additive attribution maintains interpretive fidelity despite the nonlinear transformations in the latent manifold, allowing financial analysts to evaluate feature importance in terms of actionable borrower characteristics.

3.6. Class Imbalance Handling

To address class imbalance in the default-nondefault distribution, asymmetric loss weighting was applied during GRU-LSTM training by assigning a higher penalty to misclassified default instances within the binary cross-entropy objective. This weighting increases gradient contributions from minority-class samples, guiding the optimizer to avoid bias toward the majority nondefault class. In addition, probability calibration was performed after training using threshold adjustment based on validation-set ROC analysis to align predicted scores with observed default frequencies. These measures collectively stabilize optimization under skewed distributions and improve sensitivity to high-risk borrowers without sacrificing overall classification reliability.

3.7. Hyperparameter Tuning Strategy

Hyperparameter selection was performed using grid-based validation on the training set with performance evaluated through five-fold cross-validation. Latent dimensionality of the autoencoder was explored in the range $\{16, 32, 64\}$, with 32 selected based on reconstruction loss stability and downstream classification performance. Hidden-state capacities for the GRU and LSTM layers were evaluated in $\{32, 64, 128\}$ units, with 64 providing the best trade-off between convergence speed and generalization. Regularization was applied using dropout rates in $\{0.2, 0.3, 0.5\}$ and L2 weight decay values in $\{1e^{-4}, 1e^{-3}\}$, selected to minimize validation loss while preventing overfitting. This systematic tuning protocol ensures reproducibility and controlled model complexity across the full learning pipeline.

4. Results

The proposed credit risk classification model utilizing GRU and LSTM-based machine learning models, as implemented on the Python platform, demonstrated significant performance in its ability to learn underlying patterns of borrower behavior and effectively reduce classification error rates. Within the training and validation curves, both exhibited convergences stably, indicating low overfitting and acceptable levels of generalization. Through additional evaluation and assessment metrics, including confusion matrix, ROC, and PR curves, the level of accuracy in distinguishing good and bad credit risks was evidenced. Furthermore, the incorporation of SHAP explainability enabled the feature importance to be interpreted and accessed, making the model applicable and effective in real-time financial decision-making activities, especially with regards to concerns of fairness and accountability. The observed performance improvements are closely linked to the temporal modeling capability of the GRU-LSTM architecture, which effectively captures evolving borrower behavior over time. By learning both short-term repayment variations and long-term financial patterns, the model achieves higher predictive accuracy and robustness compared to nonsequential approaches. Training was conducted using standard hyperparameter settings to ensure stable convergence. The autoencoder was trained for 100 epochs with a batch size of 32 using Adam optimization at a learning rate of 0.001. The GRU-LSTM classifier used 64 GRU units and 64 LSTM units, followed by a sigmoid-activated dense layer. The classifier was trained for 200 epochs with the same optimizer and learning rate, using binary cross-entropy as the loss function. These settings provide a reproducible baseline without task-specific tuning.

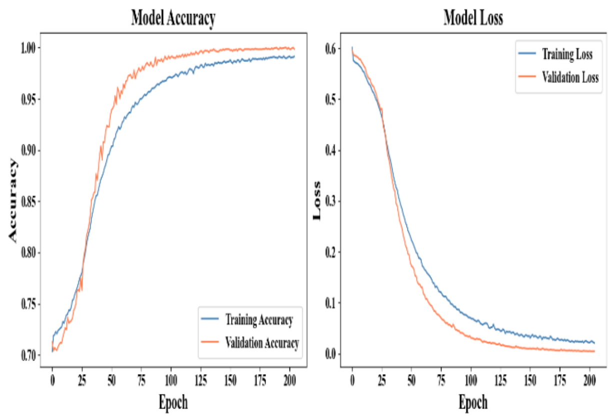

Fig. 4. Accuracy and loss model

The training results depicted in Fig. 4 show the performance of the proposed credit risk classification model based on GRU-LSTM, across 200 epochs. The left plot illustrates the precision of the model. The design appears accurate, for both training and validation, quickly increased above 98%. This reflects a high learning capacity of the proposed model and shows that generalization capability at a high level is effective. The right plot indicates the reduction curves across both verification and instruction loss are trending downwards, towards zero, thus indicating that the proposed model is capable of properly minimizing classification errors. The narrow gap between training and validation curves indicates strong robustness and minimal overfitting. Ultimately, the curves suggest that the model has learned relevant representations for classifying good vs bad credit risk.

Fig. 5 depicts performance analysis of the GRU-LSTM credit risk classification model exhibiting extremely promising findings as summarized by the important classification metrics. The model achieved an overall accuracy of 99.12% which is an exceptional indication of the ability of the model to correctly classify good or bad borrowers in terms of credit risk. The value for precision of 98.82% with

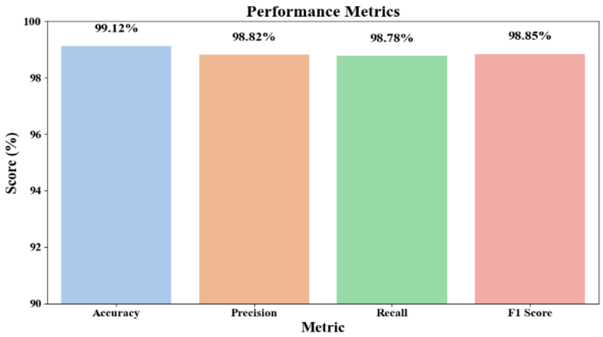

Fig. 5. Performance metrics

the model demonstrates a low false positive rate while the value of recall of 98.78% indicates a robust ability of the model to identify true positive cases (i.e., borrowers are truly creditworthy). Furthermore, the F1-score of 98.85% indicates an excellent balance of precision and recall, validating that the model could be suitable for realworld usage in credit risk assessment modelling scenarios. The F1-score, recall, accuracy, and precision values collectively lend support for the dependability and discrimination features of the present research's deep neural networks framework.

In existing credit risk modeling literature, machine learning and deep learning approaches typically report accuracy values in the range of $75\% - 85\%$ and AUC scores between 0.78 and 0.88 when evaluated on benchmark datasets such as German Credit and related financial datasets. The achieved performance in this study exceeds the lower bound of these commonly reported ranges, indicating competitive predictive capability while simultaneously providing model transparency through SHAP-based explanations. This balance between predictive strength and interpretability supports the practical suitability of the proposed framework for operational risk assessment.

The confusion matrix of Fig. 6 provides a breakdown of how well the GRU-LSTM-based credit risk assessment model classifies credit risks. The model was able to classify 1991 instances as good risk and 1998 as bad risk with only 9 false positives (good risks classified as bad) and 2 false negatives (bad risks classified as good). The high level of precision and recall measures indicate that the model was fairly balanced across both classes, which means the confused classifications were comparatively low in order to minimize misclassification costs incredibly important in financial decisions that could otherwise lead to unhappy defaults or lost lending opportunities. An extremely low number of confused classifications indicates that the model was fairly robust and sensitive at capturing meaningfully subtle patterns in borrower behaviors. This is very im-

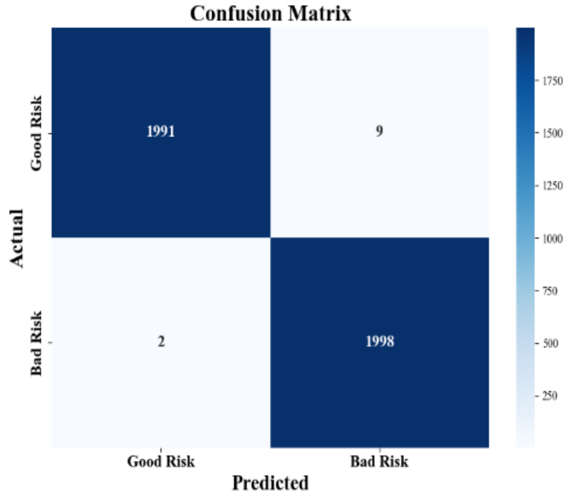

Fig. 6. Confusion Matrix

portant if one hopes to capture fairness, accuracy, and transparency in the automated credit evaluation systems' scoring process. Generally speaking, the confusion matrix supports the proposed deep learning model's ability to generalize well, making it a strong candidate for real-time credit scoring applications.

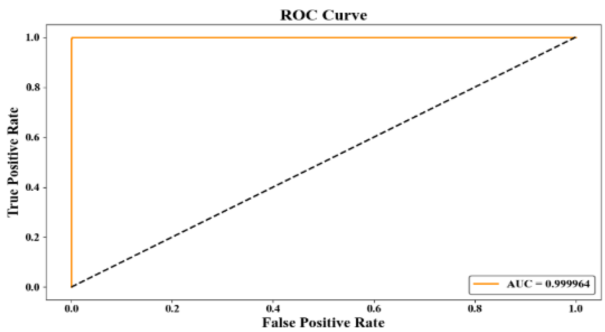

Fig. 7. ROC Curve Model

The ROC curve, seen here in Fig. 7, shows the outstanding discriminative ability of the GRU-LSTM-based credit-risk classification model used in detecting credit risk. The ROC curve appears to rise rapidly towards the upper left corner of the figure, indicating the model gives a relatively low false positive rate with a high true positive rate. Also, the AUC of 0.999964 is almost perfect. It indicates that the model is effectively able to discriminate between good and bad credit risk categories across all classification thresholds. Such a high AUC confirms the reliability, sensitivity, and robustness of employing the framework in practical situations, lending decisions, i.e., accurately identifying creditworthiness, reduces the financial risk, and improves the outcomes of lending.

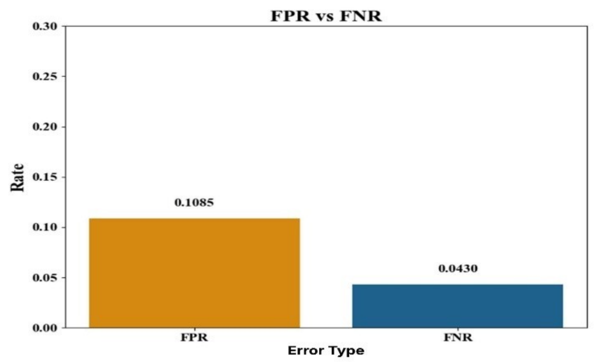

Fig. 8. FNR and FPR Model Graph

The bar graph Fig. 8 comparing False Positive Rate and False Negative Rate displays the error distributions of the GRU-LSTM-based credit risk classification model. The FPR is the proportion of "bad risk" borrowers that are incorrectly classified as "good risk" borrowers and is equal to 0.1085 , while the FNR is the proportion of "good risk" borrowers that are classified as "bad risk" borrowers and is equal to 0.0430 . Since FPR is greater than FNR, this distributional difference indicates that the GRU-LSTM model is biased toward underclassifying borrowers that are at high risk of defaulting, which is a positive outcome for reducing default risk in lending. Plus, a lower FNR indicates the model is reasonably good at identifying creditworthy borrowers, which enables reasonable access to credit while protecting the institution's financial liquidity. Concerning based on the model's overall performance design, this appears to be an acceptable compromise between true positives and untrue negative errors in its classification of credit risk. Hence, this model should be a suitable application in real-time for the purpose of credit risk classification.

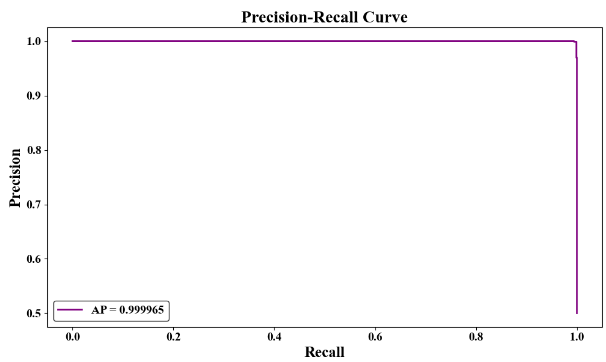

Fig. 9. Precision-Recall curve graph

Fig. 9 shows how classifying with the proposed GRU-LSTM-based credit risk model is stronger compared to other prediction algorithms, with PR curves as in this analysis, where multi-class classifier AUCs are not defined. The PR curve is also near the top-right corner of the graph, meaning that precision and recall are high and through multiple thresholds. The area that lies under the curved line indicates how much has shifted which is called the average precision (AP) was 0.999965, which means the model maintains a level of correctness on identifying creditworthy borrowers under the threshold and any bystander rates of false positives at a low level. So, the high AP illustrates the model's performance and trustworthiness against real-world financial data, which relies on minimizing uncertainty and deciding on how to handle a creditworthy borrower & their risk category, and it involves completely accurate submission to the model class in discerning credit risks through representation of the credit risk problem as sequence classification. The fact that the performance is nearly unity illustrates the effectiveness of the feature extraction and sequential classification process of the proposed model in autonomous and trusted credit risk assessment capacity.

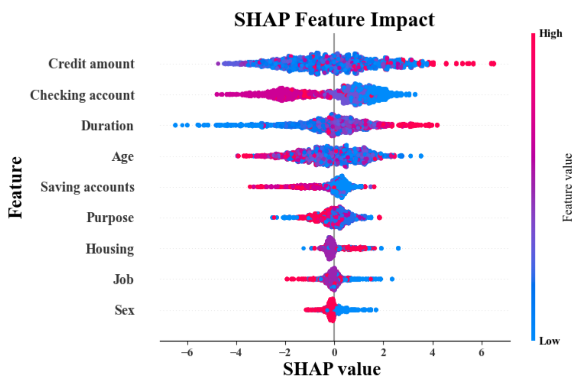

Fig. 10. SHAP feature impact graph

The SHAP feature impact plot is a way of interpreting the GRU-LSTM-based credit risk classification model by showing how features impact the graphs anticipated, as seen in Fig. 10. These three SHAP values that had the most influence on the classification label are the credit amount, checking account status, and duration of loan. This indicates these are features with the highest influence on determining if a borrower falls under good or bad credit risk. Each dot represents a SHAP value for a specific feature and instance, where dots made from lighter colours have a greater original value of that feature. For example, the higher the credit amount and the longer the duration, the greater the probability of being assigned a bad risk. This level of visualization reduces concerns from financial institutions when interpreting and trusting model assignments, providing a level of confidence that the institution is not disregarding due diligence based on a source of discrimination and having a mechanism to comply with ethical consideration or regulations related to scoring clients for credit.

4.1. Experimental Setup and Validation Protocol

To preserve partition integrity and prevent information leakage, representation learning and classification were conducted under a strictly isolated data-splitting protocol. The dataset was first divided into disjoint training, validation, and test partitions prior to any model fitting. The autoencoder was trained exclusively on the training subset, and its parameters were fixed before generating latent representations for the validation and test sets. Subsequently, the GRU-LSTM classifier was trained only on latent features derived from the training partition, while hyperparameter tuning was performed using the validation set. Final performance metrics were reported solely on the untouched test partition, ensuring unbiased evaluation of both feature learning and classification stages.

5. Conclusion

This study aimed to develop an accurate and explainable credit risk assessment framework capable of modeling complex borrower behavior while maintaining transparency for financial decision-making. The proposed integration of an autoencoder for feature compression and a

GRU-LSTM model for sequential classification successfully fulfills this objective by capturing both non-linear feature interactions and temporal dependencies in borrower data, while SHapley Additive exPlanations (SHAP) ensures interpretability of model decisions. Experimental evaluation on the German Credit dataset confirms the effectiveness of the framework, achieving 99.12% accuracy, 98.82% precision, 98.78% recall, and a 98.85% F1score, along with near-perfect ROC-AUC and PR-AUC values. These results directly demonstrate that the proposed approach meets the goals of high predictive performance, decision transparency, and practical suitability for real-time credit risk assessment in institutional financial environments. Future extensions may incorporate large-scale transactional data from banking information systems, credit card usage logs, and digital lending platforms, as well as alternative data sources such as mobile payment histories and utility billing records. In addition, deployment can be explored in cloud-based financial infrastructures and edge-computing environments to support low-latency, real-time credit risk evaluation. Future research may also evaluate the framework using real-time financial data streams, such as live transaction records, loan application feeds, and continuous credit behavior updates, to validate system robustness under dynamic operational conditions.

6. Declarations

Funding: Authors did not receive any funding.
Conflicts of interests: Authors do not have any conflicts.
Code availability: Not applicable.
Authors' Contributions: Quantong Fu is responsible for designing the framework, analyzing the performance, validating the results, and writing the article. Quantong Fu is responsible for collecting the information required for the framework, provision of software, critical review, and administering the process.
Data Available: The research study's dataset is readily available, "German Credit dataset" from Kaggle. It consists of 1,000 records of borrowers with categorical and numerical attributes to assist in research related to credit risk classification.

References

[1] N. Mansour, (2023) *"Green Technology Innovation and Financial Services System: Evidence from China"* **Businesses** ***3***(1): 1. DOI: 10.3390/businesses3010008.

[2] X. Zhang et al., (2025) *"Data-Driven Loan Default Prediction: A Machine Learning Approach for Enhancing Business Process Management"* **Systems** ***13***(7): 581. DOI: 10.3390/systems13070581.

[3] H. Tajik, G. Talebnia, H. R. Vakilifard, and F. Ahmadi, (2024) *"Machine learning support to provide an intelligent credit risk model for banks' real customers"* **International Journal of Nonlinear Analysis and Applications** ***15***(4): 23–42. DOI: 10.22075/ijnaa.2022.28382.3874.

[4] S. R. Challa, (2023) *"Artificial Intelligence and Big Data in Finance: Enhancing Investment Strategies and Client Insights in Wealth Management"* **International Journal of Scientific Research** ***12***(12): 2230–2246. DOI: 10.21275/sr231215165201.

[5] C. Guan, H. Suryanto, A. Mahidadia, M. Bain, and P. Compton, (2023) *"Responsible Credit Risk Assessment with Machine Learning and Knowledge Acquisition"* **Human-Centered Intelligent Systems** ***3***(3): 232–243. DOI: 10.1007/s44230-023-00035-1.

[6] Á. Beade, M. Rodríguez, and J. Santos, (2024) *"Business failure prediction models with high and stable predictive power over time using genetic programming"* **Operations Research International Journal** *24*(3): 52. DOI: 10.1007/s12351-024-00852-7.

[7] C. N. Nwafor, O. Nwafor, and S. Brahma, (2024) *"Enhancing transparency and fairness in automated credit decisions: an explainable novel hybrid machine learning approach"* **Scientific Reports** *14*(1): 25174. DOI: 10.1038/s41598-024-75026-8.

[8] M. Nagashima-Hayashi et al., (2022) *"Gender-Based Violence in the Asia-Pacific Region during COVID-19: A Hidden Pandemic behind Closed Doors"* **International Journal of Environmental Research and Public Health** *19*(4): 2239. DOI: 10.3390/ijerph19042239.

[9] S. Gholampour, (2024) *"Impact of Nature of Medical Data on Machine and Deep Learning for Imbalanced Datasets: Clinical Validity of SMOTE Is Questionable"* **Machine Learning and Knowledge Extraction** *6*(2): 39. DOI: 10.3390/make6020039.

[10] Y. Bao, G. Hilary, and B. Ke. "Artificial Intelligence and Fraud Detection". In: *Innovative Technology at the Interface of Finance and Operations*. Springer, 2022, 223–247. DOI: 10.1007/978-3-030-75729-8_8.

[11] D. Slack, S. Krishna, H. Lakkaraju, and S. Singh, (2023) *"Explaining machine learning models with interactive natural language conversations using TalkToModel"* **Nature Machine Intelligence** *5*(8): 873–883. DOI: 10.1038/s42256-023-00692-8.

[12] T. Berhane, T. Melese, A. Walelign, and A. Mohammed, (2023) *"A Hybrid Convolutional Neural Network and Support Vector Machine-Based Credit Card Fraud Detection Model"* **Mathematical Problems in Engineering** *2023*: 8134627. DOI: 10.1155/2023/8134627.

[13] D. Halvoník and J. Kapusta, (2024) *"Large Language Models and Rule-Based Approaches in Domain-Specific Communication"* **IEEE Access** *12*: 107046–107058. DOI: 10.1109/ACCESS.2024.3436902.

[14] Z. Amiri et al., (2023) *"The Personal Health Applications of Machine Learning Techniques in the Internet of Behaviors"* **Sustainability** *15*(16): 12406. DOI: 10.3390/su151612406.

[15] M. Kinney, M. Anastasiadou, M. Naranjo-Zolotov, and V. Santos, (2024) *"Expectation management in AI: A framework for understanding stakeholder trust and acceptance of artificial intelligence systems"* **Heliyon** *10*(7): e28562. DOI: 10.1016/j.heliyon.2024.e28562.

[16] C. N. Nwafor, O. Nwafor, and S. Brahma, (2024) *"Enhancing transparency and fairness in automated credit decisions: an explainable novel hybrid machine learning approach"* **Scientific Reports** *14*(1): 25174. DOI: 10.1038/s41598-024-75026-8.

[17] L. Liu, (2022) *"A Self-Learning BP Neural Network Assessment Algorithm for Credit Risk of Commercial Bank"* **Wireless Communications and Mobile Computing** *2022*: 9650934. DOI: 10.1155/2022/9650934.

[18] N. Biswas, A. S. Mondal, A. Kusumastuti, S. Saha, and K. C. Mondal, (2025) *"Automated credit assessment framework using ETL process and machine learning"* **Innovations in Systems and Software Engineering** *21*(1): 257–270. DOI: 10.1007/s11334-022-00522-x.

[19] Z. Yu, (2023) *"Prediction of Credit Card Loan Risk Based on Multilayer Perceptron Neural Network Model"* **BCPBM** *38*: 126–134. DOI: 10.54691/bcpbm.v38i.3679.

[20] J. Hu, H. Wang, and X. Lu, (2023) *"Research on Credit Risk Assessment Model of Enterprises' Unpaid Electricity Charge Based on Machine Learning Method"* **SHS Web of Conferences** *170*: 03023. DOI: 10.1051/shsconf/202317003023.

[21] T. A. O. Odejide, (2024) *"Theoretical frameworks in AI for credit risk assessment: Towards banking efficiency and accuracy"* **International Journal of Scientific Research Updates** 7(1): 092–102. DOI: 10.53430/ijsru.2024.7.1.0030.

[22] S. B. Coşkun and M. Turanli, (2023) *"Credit risk analysis using boosting methods"* **Journal of Applied Mathematics, Statistics and Informatics** *19*(1): 5–18. DOI: 10.2478/jamsi-2023-0001.

[23] C. Rao, Y. Liu, and M. Goh, (2023) *"Credit risk assessment mechanism of personal auto loan based on PSO-XGBoost Model"* **Complex & Intelligent Systems** *9*(2): 1391–1414. DOI: 10.1007/s40747-022-00854-y.

[24] S. Luo, M. Xing, and J. Zhao, (2022) *"Construction of Artificial Intelligence Application Model for Supply Chain Financial Risk Assessment"* **Scientific Programming** *2022*: 4194576. DOI: 10.1155/2022/4194576.

[25] Y. Bai and D. Zha, (2022) *"Commercial Bank Credit Grading Model Using Genetic Optimization Neural Network and Cluster Analysis"* **Computational Intelligence and Neuroscience** *2022*: 4796075. DOI: 10.1155/2022/4796075.

[26] G. F. Bone-Winkel and F. Reichenbach, (2024) *"Improving credit risk assessment in P2P lending with explainable machine learning survival analysis"* **Digital Finance *6***(3): 501–542. DOI: 10.1007/s42521-024-00114-3.

[27] M. R. Machado, D. T. Chen, and J. R. Osterrieder, (2025) *"An analytical approach to credit risk assessment using machine learning models"* **Decision Analytics Journal *16***: 100605. DOI: 10.1016/j.dajour.2025.100605.

[28] N. A. de Oliveira and L. F. C. Basso, (2025) *"Advancing Credit Rating Prediction: The Role of Machine Learning in Corporate Credit Rating Assessment"* **Risks *13***(6): 116. DOI: 10.3390/risks13060116.

[29] M. Sarıkoç and M. Celik, (2025) *"PCA-ICA-LSTM: A Hybrid Deep Learning Model Based on Dimension Reduction Methods to Predict S&P 500 Index Price"* **Computational Economics *65***(4): 2249–2315. DOI: 10.1007/s10614-024-10629-x.

[30] C. Wang and H. Yu, (2025) *"Intelligent Assessment of Personal Credit Risk Based on Machine Learning"* **Systems *13***(2): 112. DOI: 10.3390/systems13020112.

[31] K. Dwi Hartomo, C. Arthur, and Y. Nataliani, (2025) *"A Novel Weighted Loss TabTransformer Integrating Explainable AI for Imbalanced Credit Risk Datasets"* **IEEE Access *13***: 31045–31056. DOI: 10.1109/ACCESS.2025.3541878.

[32] *Credit Risk Evaluation*. Kaggle Notebook. Accessed July 21, 2025. 2025.