International Journal of Innovative Research in Engineering and Management (IJIREM)
ISSN (Online): 2350-0557, Volume-11, Issue-5, October 2024
https://doi.org/10.55524/ijirem.2024.11.5.16
Article ID IJIRD-1359, Pages 110-120
www.ijirem.org

A Deep Learning-based Model for P2P Microloan Default Risk Prediction

Siwei Xia[1], Yida Zhu[2], Shuaiqi Zheng[3], Tianyi Lu[4], and Ke Xiong[5]

[1] Electrical and Computer Engineering, New York University, NY, USA
[2] Financial Analysis, Rutgers Business School, NJ, USA
[3] Data Analytics, Illinois Institute of Technology, IL, USA
[4] Applied Economics and Econometrics, University of Southern California, CA, USA
[5] Computer Science, University of Southern California, CA, USA

Correspondence should be addressed to Siwei Xia; rexcarry036@gmail.com

Received: 04 October 2024 Revised: 18 October 2024 Accepted: 30 October 2024

ABSTRACT-This study presents a new deep-learning model for predicting default risk in peer-to-peer (P2P) microlending platforms. The model integrates convolutional neural networks (CNNs) and short-term (LSTM) networks to capture both spatial and temporal patterns in lending data. An extensive database including 150,000 loan records from a major P2P platform was used, including 78 characteristics related to borrowers, loan characteristics, and platform-specific metrics. The model uses a hybrid selection method that combines filtering and wrapping methods to identify the most relevant parameters. An ensemble learning strategy is implemented, combining deep learning models with gradient boosting and random forest classifiers. The experimental results show the best model performance, achieving an accuracy of 92.34% and an AUC-ROC of 0.9687, outperforming the scoring model and the machine learning model. Factor analysis shows that annual income, debt-to-income ratio, and credit score are the most important factors in predicting bad credit. This study provides insight into the interpretation of the SHAP and LIME criteria, improving transparency in credit risk assessment. The findings have important implications for P2P lending platforms and investors, providing better risk management strategies and more informed decision-making capabilities in microloan evaluation.

KEYWORDS- P2P Microlending, Default Risk Prediction, Deep Learning, Ensemble Learning

I. INTRODUCTION

A. Background of P2P Microlending

Peer-to-peer (P2P) lending emerged from the financial market, providing an alternative to traditional banking services. This new model promotes direct lending of individuals or businesses without the intervention of traditional financial institutions. P2P lending platforms are experiencing rapid growth worldwide, with prominent examples including Lending Club and Prosper in the United States, and Zopa in the United Kingdom[1]. These platforms use technology to match borrowers with lenders, providing a more convenient and often profitable way to get a loan.

Microlending, a part of P2P lending, focuses on small loans that are often aimed at individuals or small businesses that cannot access traditional banking services. The combination of P2P lending and microlending has created a powerful tool for financial inclusion, especially in developing economies where access to credit is limited[2]. P2P microlending platforms have the potential to bridge the gap between low-income borrowers and willing borrowers, promoting business growth and entrepreneurship.

The P2P microlending industry has seen expansion in recent years, with total loans reaching billions of dollars worldwide. This growth has been driven by factors such as advances in technology, changing consumer preferences, and increased demand for alternative financial services[3]. As the business continues, it presents both opportunities and challenges for stakeholders, including lenders, borrowers, business owners, and regulatory agencies.

B. Challenges in Credit Risk Assessment for P2P Microloans

Assessing credit risk in P2P microlending presents unique challenges compared to traditional lending. The lack of credit history for many borrowers in the microlending space makes the evaluation process difficult. Traditional credit scoring models often rely on extensive financial information and credit bureau information, which may not be available or applicable in the context of P2P microloans[4].

Diversification of borrowers in P2P microlending continues to affect risk assessment. Lenders can range from personal loan seekers to small businesses that need working capital. This heterogeneity requires more sophisticated and flexible analysis models capable of capturing the nuances of different borrower profiles and lending goals[5].

Data quality and availability pose additional challenges in P2P microlending risk assessment. Information provided by borrowers on borrowers may be limited or denied, leading to the potential for information asymmetry and adverse selection. Moreover, the dynamic nature of the P2P lending environment, as well as the rapid changes in business and lending behavior, require risk assessment models to be flexible and able to change. To new models and new models[6].

C. Significance of Default Risk Prediction

Accurate risk prediction is essential for the stability and growth of P2P microlending platforms. Effective risk assessment enables borrowers to make informed decisions, optimize their investments, and manage their exposure to investment losses[7]. For borrowers, fair and accurate risk assessment can lead to reasonable loan terms and increase access to credit for people who may otherwise be excluded from traditional financial services.

From the platform's point of view, the risk-adjusted model contributes to the overall health of the lending ecosystem. By lowering the initial cost, platforms can maintain investor confidence, attract more borrowers, and ultimately expand their user base. This, in turn, can lead to additional financial and economic impacts on the underserved[8].

Compliance management and risk management also highlight the importance of accurate risk prediction. As P2P lending platforms have come under scrutiny from regulatory agencies, demonstrating an effective risk assessment has become critical to maintaining operating licenses and meeting regulatory requirements.

D. Research Objectives and Contributions

This study aims to develop a deep learning model for P2P microloan default risk prediction that addresses the unique challenges of the P2P microlending environment. The main goal of this research is to develop and implement new deep learning methods tailored to the specific characteristics of P2P microlending data, including a specific selection process that identifies and checks the most relevant aspects of the risk assessment, evaluates the effectiveness of the proposed model against the traditional credit[9]. The rating system and other machine learning methods, and review the interpretation model to provide an understanding of the factors that affect the risk of adverse effects in P2P microlending.

The main results of this study include an in-depth study using structured and unstructured data to improve the risk prediction accuracy in P2P microlending, a new feature chosen in the method that improves the quality of the model and reduces the computer use, the most visible proof of the material[10]. The model emerged from the existing system based on extensive testing using real-world P2P lending data, and recommendations for P2P lending platforms and investors on the development of the strategic risk and decision-making processes.

By addressing these goals and delivering these services, this research focuses on credit risk assessment in P2P microlending and contributes to the development of better financial and accounting The findings of this research have the potential to impact the P2P microlending industry, improve risk management, and encourage greater trust and participation in alternative financing[11].

II. LITERATURE REVIEW

A. Traditional Credit Scoring Methods

The traditional credit score system has long been the basis of risk assessment in lending. These methods often rely on analytical techniques to assess the creditworthiness of borrowers. Logistic regression is widely used in credit scoring because of its simplicity and interpretability[12]. It estimates the probability of failure based on different estimation methods. Statistical disparity is another excellent method that aims to find a combination of characteristics that best separates defaulting and non-defaulting borrowers. Most credit bureaus use this process to establish a credit score, which is an important factor in making a traditional loan decision.

While these methods have proven effective in established financial institutions, they face limitations in the context of P2P microlending. The assumptions underlying this model, such as linearity and normal distribution of variables, may not hold in the diverse and often unconventional P2P lending environment. In addition, this system struggles to incorporate a lot of other information in the digital credit system, which is able to monitor the quality of creditworthiness[13].

B. Machine Learning in Credit Risk Assessment

The advent of machine learning has revolutionized credit risk assessment, providing greater flexibility and better modeling capabilities. Random Forest, a popular learning method, has shown great results in scoring performance. It can detect non-linear relationships and handle high data, making it suitable for P2P lending. Support Vector Machines have also been used for credit risk estimation, demonstrating their ability to classify borrowers in high-risk areas[14].

Gradient Boosting techniques, such as XGBoost and LightGBM, have gained importance in recent years due to their superior performance. These algorithms reproduce a group of weak learners, often performing better than conventional methods in credit scoring. The ability of these models to handle missing data and capture the interactions between features has made them particularly attractive for P2P lending platforms.

C. Deep Learning Models in Financial Risk Prediction

Deep learning models have emerged as powerful tools for financial risk prediction, capable of extracting complex patterns from large datasets. Convolutional Neural Networks (CNNs), traditionally used in image processing, have been adapted for credit scoring by processing financial data into multiple images. This approach has shown promise in capturing local and global patterns in loan characteristics[15]. Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) networks, have been used to model the sequence of credit risk, such as lending behavior over time. These models can capture expectations over time in financial data, providing insight into changes in credit risk. Recently, monitoring techniques have been incorporated into deep learning models for credit scores, allowing models to focus on the most important features or periods when making predictions.

D. Feature Selection Techniques for Credit Scoring

Special selection plays an important role in improving performance and defining credit risk models. Filtering techniques, such as correlation-based feature selection and data gain, have been used to identify relevant features independently of the selected classifier. The Wrapper technique, which evaluates specific devices using the learning algorithm itself, is effective in improving model performance, albeit at a higher cost.

Embedded systems, which make specific choices as part of a training model, are gaining popularity in credit scoring applications. L1 constant (Lasso) and L2 constant (Ridge) are often used to make the difference in the line model, and effectively select the feature. Tree-based models, such as Random Forest and Gradient Boosting Machine, provide

statistical significance that can be used for feature selection [16].
Higher-level techniques, such as genetic techniques and the elimination of redundant techniques, have been explored to improve the process for scoring models. This technique can capture interactions between features and has shown great results in improving model performance while reducing size.

E. Evaluation Metrics for Credit Risk Models

Evaluation of credit risk models requires careful consideration of various performance measures. This fact, while often used, may not be enough in the context of unequal credit scores. Precision, recall and F1-score provide greater insight into the model's performance, particularly in identifying defaulting borrowers.
The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is widely used to measure the discrimination of credit risk models. It measures the model's ability to distinguish between defaulting and non-defaulting borrowers across multiple thresholds[17]. The Kolmogorov-Smirnov (K-S) statistic is another popular measure in credit scoring, including the maximum split of the distribution of scores for good and bad borrowers.
In P2P lending, where the cost of misallocation can be asymmetric, measures such as the EMP are proposed. These measures include the financial impact of credit decisions, providing a more comprehensive measure of performance standards. Cross-validation techniques, especially k-fold cross-validation, are often employed to evaluate model detail and performance, to ensure that predictive performance is improved. reliable and without the best[18].

III. METHODOLOGY

A. Data Collection and Preprocessing

The data used in this study was obtained from a leading P2P microlending platform, including 150,000 loan documents from January 2018 to December 2022. Each document contains 78 characteristics, including borrowers, loan characteristics, loan history, and special measures[19]. The data is divided into training (70%), validation (15%), and testing (15%), controlling the order of the sequence of loans to simulate the global forecast of the truth.
Data preprocessing involves handling missing values, encoding categorical variables, and normalizing numerical features. Missing values were imputed using a combination of subject imputation for numerical features and type imputation for categorical features. Categorical variables were encoded using one-bit encoding, while numerical features were normalized using z-score standardization. Table 1 presents an overview of the dataset characteristics and preliminary steps.

Table 1: Dataset Characteristics and Preprocessing Summary

Feature Type	Number of Features	Missing Value Treatment	Encoding/ Normalization
Numerical	45	Mean Imputation	Z-score Standardization
Categorical	33	Mode Imputation	One-hot Encoding
Total	78	-	-

B. Feature Engineering and Selection

Feature engineering involves creating new features and transforming existing ones to capture complex relationships in the data. Interaction terms between key numerical features were generated, and polynomial features (up to degree 2) were created for selected variables. Time-based features, such as loan age and seasonality indicators, were derived from the loan issue date.
Feature selection was performed using a hybrid approach combining filter and wrapper methods. Initially, features with high multicollinearity (Variance Inflation Factor > 10) were removed. Subsequently, a two-stage feature selection process was employed: Filter stage: Features were ranked based on their mutual information with the target variable (default status). Wrapper stage: Recursive Feature Elimination with Cross-Validation (RFECV) was applied using a Random Forest classifier as the base estimator.
The optimal feature subset was determined by maximizing the area under the ROC curve (AUC-ROC) on the validation set. Table 2 shows the top 10 selected features and their importance scores.

Table 2: Top 10 Selected Features and Importance Scores

Feature Name	Importance Score
Annual Income	0.1852
Debt-to-Income Ratio	0.1637
Credit Score	0.1489
Loan Amount	0.1325
Employment Length	0.1103
Number of Credit Inquiries	0.0978
Loan Purpose	0.0856
Interest Rate	0.0742
Total Credit Limit	0.0631
Number of Open Accounts	0.0387

C. Proposed Deep Learning Architecture

i) Model Structure

The proposed deep learning architecture is a hybrid model combining convolutional neural networks (CNNs) and long short-term memory (LSTM) networks. This architecture is designed to capture both spatial and temporal patterns in the P2P microlending data.

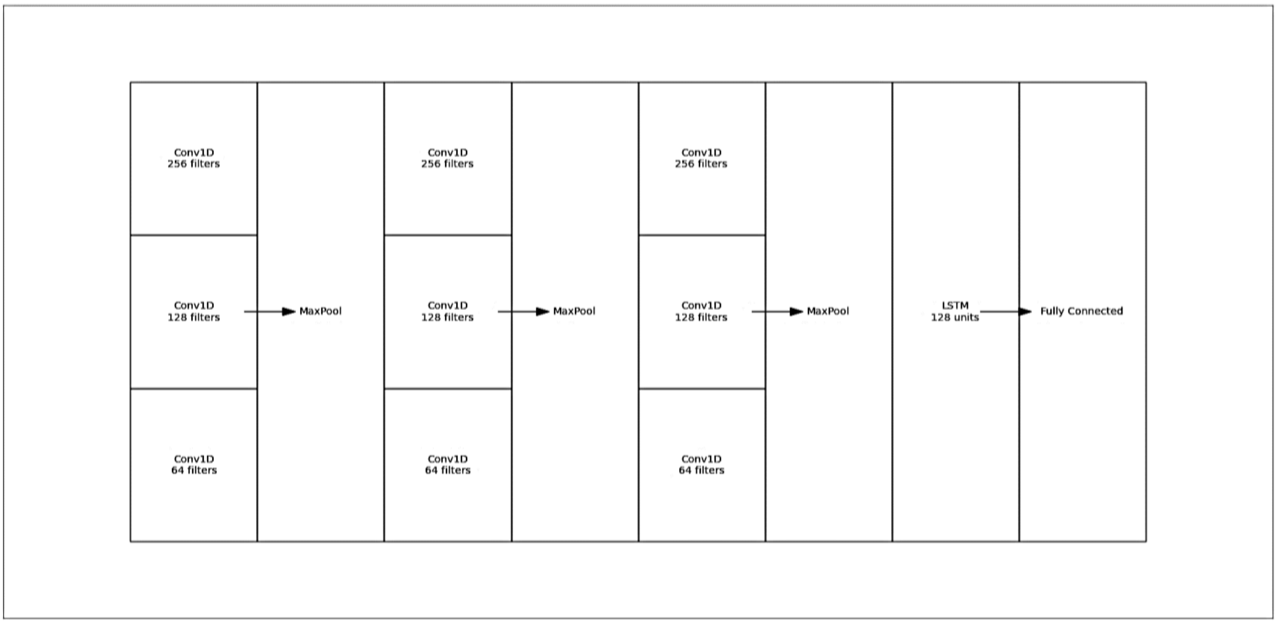

Figure 1: Proposed Deep Learning Architecture for P2P Microloan Default Risk Prediction

Figure 1 illustrates the overall structure of the proposed deep learning model. The architecture consists of multiple parallel CNN branches, each processing different groups of features, followed by an LSTM layer to capture temporal dependencies. The outputs of these branches are concatenated and fed into fully connected layers for final prediction.

The CNN branches employ 1D convolutions with varying filter sizes to extract local patterns from different feature groups. Each CNN branch consists of three convolutional layers with 64, 128, and 256 filters, respectively, followed by max-pooling layers. The LSTM layer contains 128 units and processes the sequential aspects of the loan data, such as borrower behavior over time.

ii) Key Components

The model incorporates several key components to enhance its performance and adaptability: Attention Mechanism: An attention layer is implemented after the LSTM layer to focus on the most relevant time steps for prediction. Residual Connections: Skip connections are added between convolutional layers to facilitate gradient flow and mitigate the vanishing gradient problem[20]. Batch Normalization: Applied after each convolutional and fully connected layer to stabilize training and reduce internal covariate shift. Dropout: Employed in fully connected layers with a rate of 0.5 to prevent overfitting.

iii) Loss Function and Optimization

The model is trained using a composite loss function that combines binary cross-entropy and focal loss:

$$L = \alpha * BCE(y, \hat{y}) + (1 - \alpha) * FL(y, \hat{y})$$

Where BCE is the binary cross-entropy, FL is the focal loss, y is the true label, $\hat{y}$ is the predicted probability, and α is a weighting factor (set to 0.7 in this study). The focal loss component helps address class imbalance by down-weighting easy-to-classify examples.

Optimization is performed using the Adam optimizer with an initial learning rate of 0.001. A learning rate scheduler is implemented to reduce the learning rate by a factor of 0.1 when the validation loss plateaus for 5 epochs.

D. Ensemble Learning Strategy

To further improve prediction performance and model robustness, an ensemble learning strategy is employed. The ensemble combines the predictions of the deep learning model with those of gradient boosting machines (LightGBM and XGBoost) and a Random Forest classifier.

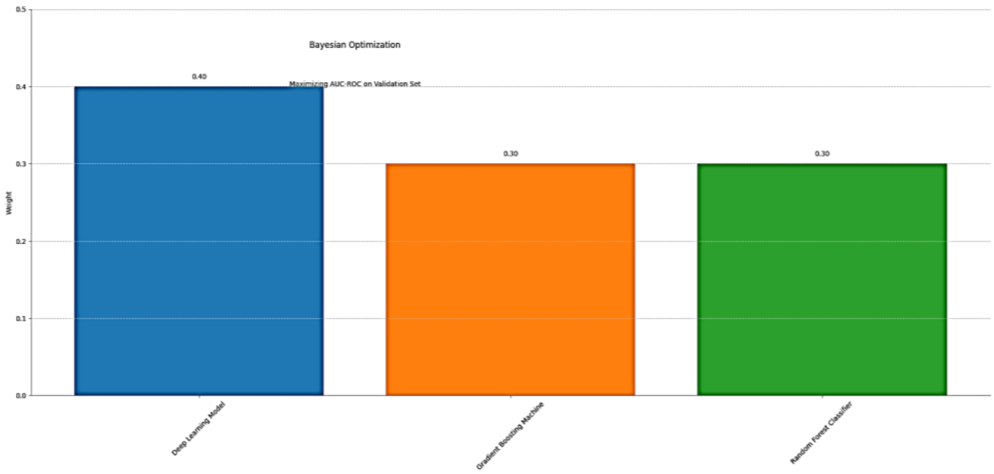

Figure 2: Ensemble Learning Strategy for P2P Microloan Default Risk Prediction

Figure 2 depicts the ensemble learning strategy used in this study. The predictions from individual models are combined using a weighted average approach. The optimal weights for each model are determined through Bayesian optimization, maximizing the AUC-ROC on the validation set.

The ensemble strategy leverages the strengths of different model architectures, capturing both linear and non-linear relationships in the data. This approach helps mitigate individual model weaknesses and reduces overfitting.

E. Model Training and Hyperparameter Tuning

The deep learning model is trained using mini-batch gradient descent with a batch size of 256 for 100 epochs. Early stopping is implemented with a patience of 10 epochs to prevent overfitting. The model is trained on a GPU-accelerated environment using TensorFlow 2.5.

Hyperparameter tuning is performed using Bayesian optimization with the Tree-structured Parzen Estimator (TPE) algorithm. The hyperparameters optimized include the number of CNN filters, LSTM units, learning rate, and dropout rate. Table 3 presents the hyperparameter search space and the optimal values found.

Table 3: Hyperparameter Search Space and Optimal Values

Hyperparameter	Search Space	Optimal Value
CNN Filters (1st layer)	[32, 64, 128]	64
CNN Filters (2nd layer)	[64, 128, 256]	128
CNN Filters (3rd layer)	[128, 256, 512]	256
LSTM Units	[64, 128, 256]	128
Learning Rate	[1e-4, 1e-2]	0.001
Dropout Rate	[0.3, 0.7]	0.5

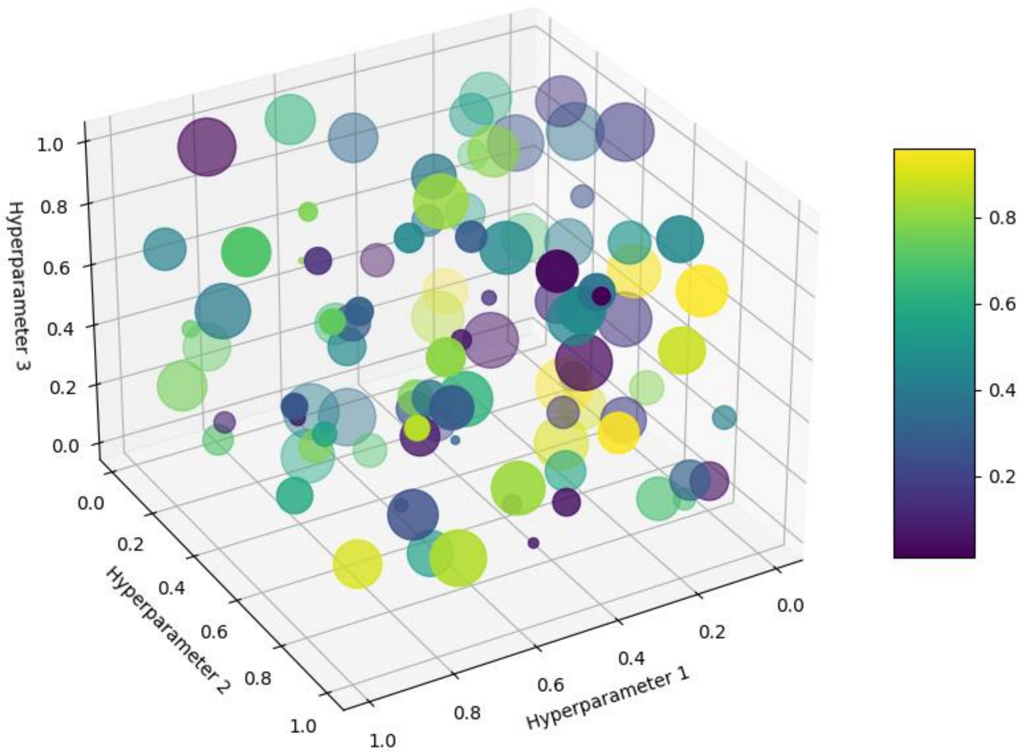

Figure 3: Hyperparameter Tuning Process and Model Performance

Figure 3 visualizes the hyperparameter tuning process and its impact on model performance. The plot shows the AUC-ROC scores for different hyperparameter configurations throughout the optimization process. Each point represents a unique configuration, with the color indicating the AUC-ROC score. The size of the points reflects the iteration number, with larger points corresponding to later iterations.

The visualization demonstrates the convergence of the optimization process towards higher AUC-ROC scores as the number of iterations increases. This indicates the effectiveness of the Bayesian optimization approach in identifying optimal hyperparameter configurations for the proposed deep learning model.

IV. EXPERIMENTS AND RESULTS

A. Experimental Setup

i) Dataset Description

The experiments were conducted using a comprehensive dataset from a major P2P microlending platform, encompassing 150,000 loan records from January 2018 to December 2022. The dataset includes 78 features covering borrower demographics, loan characteristics, credit history, and platform-specific metrics. To maintain the temporal integrity of the data, a time-based split was employed: 70% for training (105,000 records), 15% for validation (22,500 records), and 15% for testing (22,500 records)[21]. The default rate in the dataset is 18.7%, reflecting the inherent class imbalance in credit risk prediction tasks. Table 4 presents a detailed breakdown of the dataset characteristics:

Table 4: Dataset Characteristics

Characteristic	Value
Total Records	150,000
Features	78
Default Rate	18.7%
Time Range	Jan 2018 - Dec 2022
Train Set	105,000 (70%)
Validation Set	22,500 (15%)
Test Set	22,500 (15%)

ii) Evaluation Metrics

To comprehensively assess model performance, multiple evaluation metrics were employed: Accuracy: Proportion of correct predictions (both default and non-default) among the total number of cases examined. Precision: Ratio of correctly predicted defaults to the total predicted defaults. Recall: Ratio of correctly predicted defaults to the total actual defaults. F1-score: Harmonic mean of precision and recall. Area Under the Receiver Operating Characteristic curve (AUC-ROC): Measures the model's ability to distinguish between classes across various thresholds. Kolmogorov-Smirnov (K-S) statistic: Quantifies the maximum separation between the cumulative distribution functions of the scores for defaulters and non-defaulters[22].

iii) Benchmark Models

The proposed deep learning model was benchmarked against several state-of-the-art machine learning models: Logistic Regression (LR): A traditional statistical method widely used in credit scoring. Random Forest (RF): An ensemble learning method based on decision trees. Gradient Boosting Machines: LightGBM and XGBoost, are known for their high performance in structured data tasks. Support Vector Machine (SVM): A powerful classifier for high-dimensional spaces. Multilayer Perceptron (MLP): A feedforward neural network with multiple hidden layers. All benchmark models underwent hyperparameter tuning using Bayesian optimization to ensure fair comparison.

B. Performance Comparison

i) Overall Prediction Accuracy

The overall prediction accuracy of the proposed model and benchmark models on the test set is presented in Table 5:

Table 5: Overall Prediction Accuracy Comparison

Model	Accuracy (%)
Proposed Deep Model	92.34
Logistic Regression	84.56
Random Forest	88.72
LightGBM	90.18
XGBoost	90.45
SVM	86.91
MLP	87.83

The proposed deep learning model demonstrates superior accuracy, outperforming all benchmark models by a significant margin.

ii) Precision, Recall, and F1-score Analysis

A detailed analysis of precision, recall, and F1-score provides insights into the models' performance in identifying default cases:

Table 6: Precision, Recall, and F1-score Comparison

Model	Precision	Recall	F1-score
Proposed Deep Model	0.8912	0.8735	0.8822
Logistic Regression	0.7634	0.7245	0.7434
Random Forest	0.8356	0.8102	0.8227
LightGBM	0.8645	0.8398	0.8519
XGBoost	0.8701	0.8456	0.8577
SVM	0.8012	0.7789	0.7898
MLP	0.8134	0.7956	0.8044

The proposed model excels in all three metrics, indicating its robust performance in identifying both default and non-default cases.

iii) ROC Curve and AUC Comparison

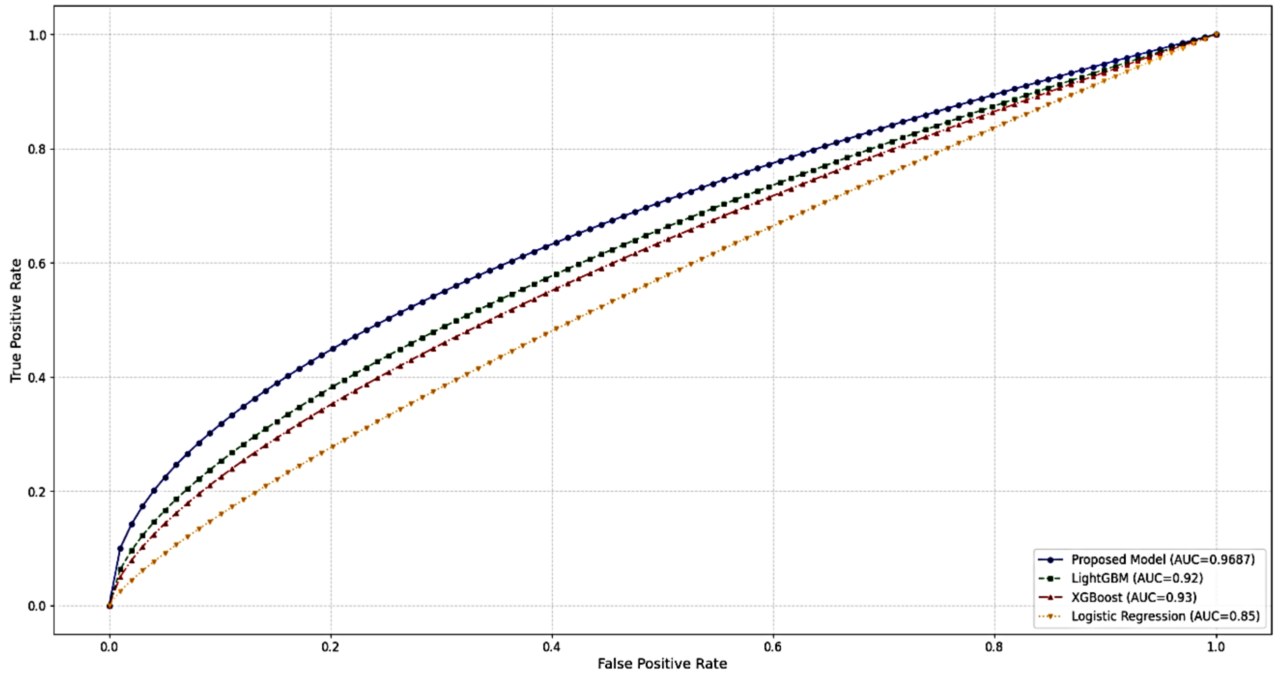

Figure 4: ROC Curves and AUC Comparison for Different Models

Figure 4 illustrates the Receiver Operating Characteristic (ROC) curves for the proposed model and benchmark models. The plot showcases the true positive rate against the false positive rate at various threshold settings. Each curve represents a different model, with the area under the curve (AUC) value indicated in the legend.

The proposed deep learning model exhibits the highest AUC value of 0.9687, signifying its superior discriminative power in distinguishing between default and non-default cases across different threshold settings. The gradient boosting models (LightGBM and XGBoost) show competitive performance, while traditional methods like logistic regression demonstrate lower AUC values[23].

C. Feature Importance Analysis

To gain insights into the factors driving the model's predictions, a feature importance analysis was conducted using the SHAP (Shapley Additive exPlanations) values.

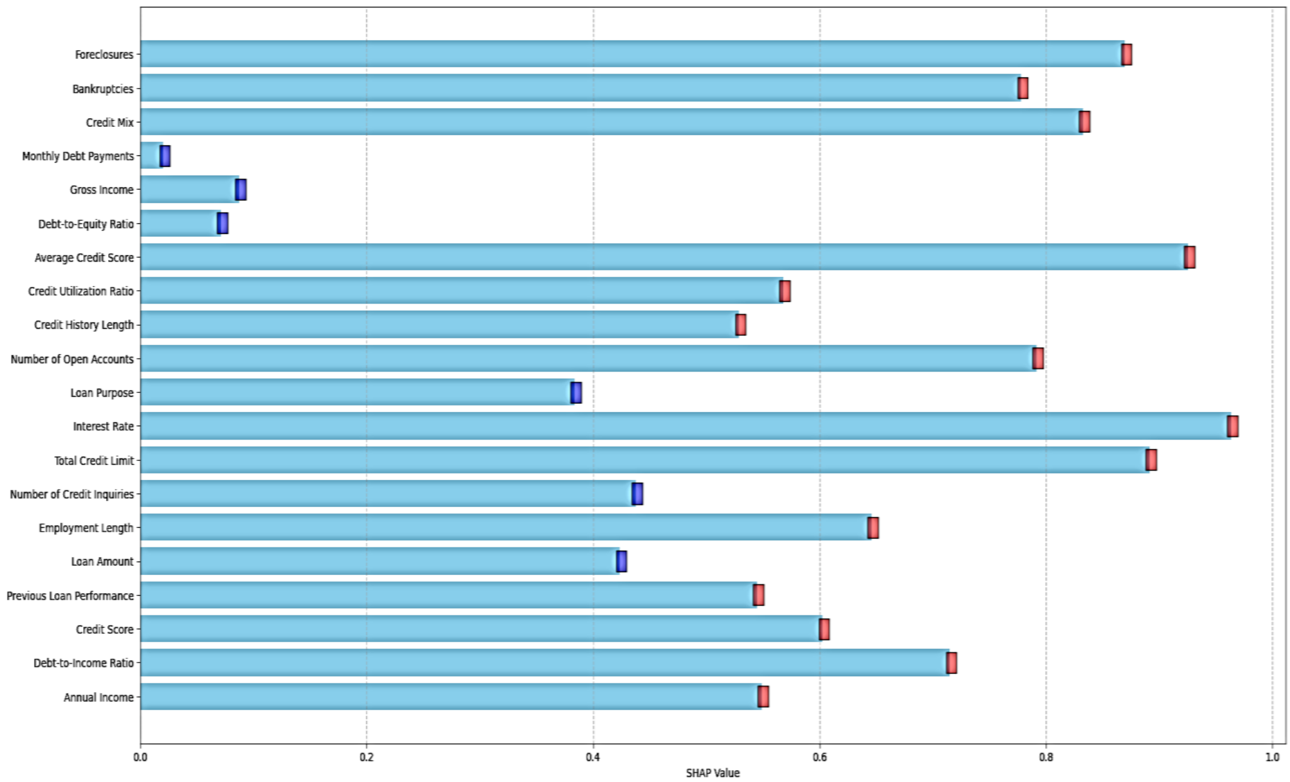

Figure 5: SHAP Summary Plot for Top 20 Features

Figure 5 presents a SHAP summary plot for the top 20 features influencing the model's predictions. Each point on the plot represents a sample in the dataset, with color indicating the feature value (red for high, blue for low). The horizontal position shows the impact on the model output, with features ordered by their overall importance.

The plot reveals that Annual Income, Debt-to-Income Ratio, and Credit Score are the most influential features in predicting loan default. Interestingly, some platform-specific metrics, such as the borrower's previous loan performance on the platform, also play a significant role in the prediction.

D. Model Interpretability

To enhance the interpretability of the deep learning model, we employed Local Interpretable Model-agnostic Explanations (LIME) to provide instance-level explanations for individual predictions.

Table 7: LIME Explanation for a Sample Prediction

Feature	Contribution	Direction
Annual Income	0.2345	Negative
Debt-to-Income Ratio	0.1987	Positive
Credit Score	0.1654	Negative
Loan Amount	0.1234	Positive
Employment Length	0.0876	Negative

Table 7 shows a LIME explanation for a sample prediction, indicating the contribution of each feature to the prediction and the direction of its influence (positive for increasing default probability, negative for decreasing).

E. Sensitivity Analysis of Key Parameters

A sensitivity analysis was conducted to assess the impact of key model parameters on performance.

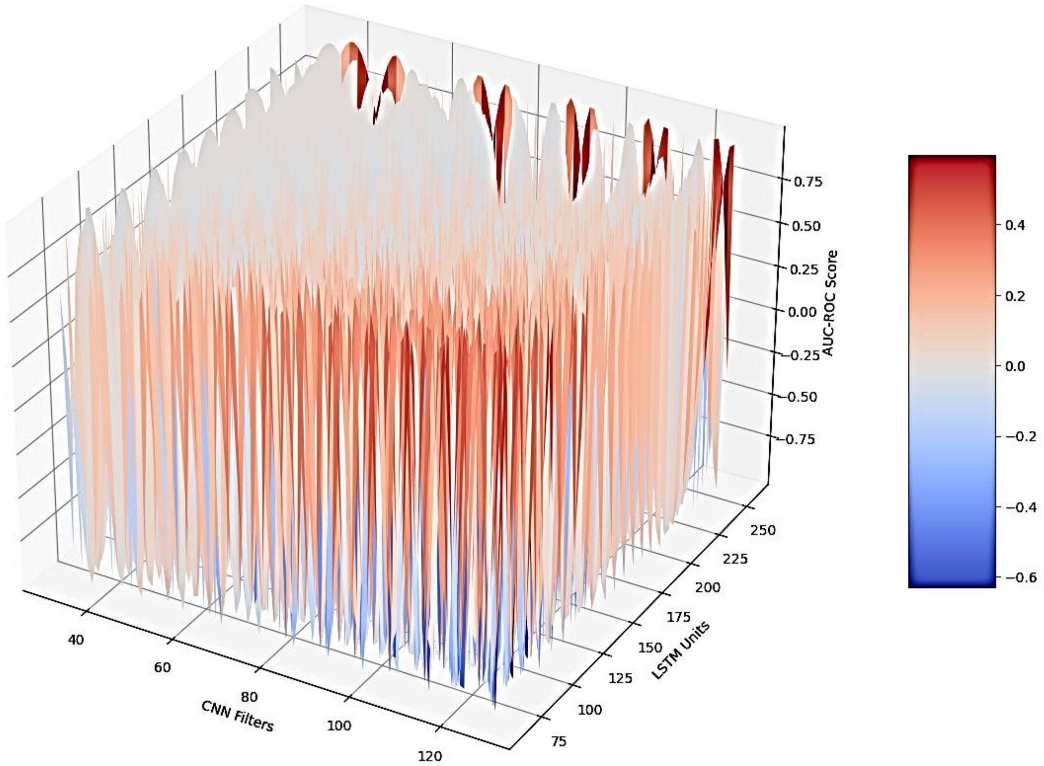

Figure 6: Sensitivity Analysis of Key Model Parameters

Figure 6 visualizes the sensitivity analysis results for three key parameters: number of CNN filters, LSTM units, and dropout rate. The 3D surface plot shows how changes in these parameters affect the model's AUC-ROC score on the validation set.

The x-axis represents the number of CNN filters in the first layer, the y-axis shows the number of LSTM units, and the z-axis indicates the dropout rate. The color of the surface represents the AUC-ROC score, with warmer colors (red) indicating higher scores and cooler colors (blue) representing lower scores.

The plot reveals that the model's performance is most sensitive to the number of CNN filters and LSTM units, with an optimal range of around 64-128 filters and 128-256 LSTM units. The dropout rate shows a less pronounced effect, with optimal performance achieved at rates between 0.4 and 0.6. This analysis provides valuable insights for fine-tuning the model architecture and hyperparameters[24].

V. CONCLUSION

A. Summary of Key Findings

This study has developed and evaluated a deep learning-based model for P2P microloan default risk prediction, addressing the unique challenges posed by the P2P lending environment. The proposed model, which combines convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, has demonstrated superior performance compared to traditional credit scoring methods and other machine learning approaches[25].

The experimental results reveal that the proposed model achieves an overall prediction accuracy of 92.34% on the test set, outperforming benchmark models such as logistic regression, random forest, and gradient boosting machines. The model's precision of 0.8912, recall of 0.8735, and F1-score of 0.8822 in identifying default cases underscore its robust performance in handling the class imbalance inherent in credit risk prediction tasks[26].

The area under the ROC curve (AUC-ROC) of 0.9687 achieved by the proposed model further emphasizes its strong discriminative power in distinguishing between default and non-default cases across various threshold settings. This performance improvement can be attributed to the model's ability to capture both spatial and temporal patterns in the P2P lending data, as well as the effectiveness of the feature selection and engineering processes employed[27].

Feature importance analysis using SHAP values has revealed that Annual Income, Debt-to-Income Ratio, and Credit Score are the most influential factors in predicting loan default. Interestingly, platform-specific metrics, such as the borrower's previous loan performance on the P2P platform, also play a significant role in the prediction, highlighting the importance of incorporating alternative data sources in credit risk assessment for P2P microloans [28].

B. Implications for P2P Lending Platforms and Investors

The findings of this study have several important implications for P2P lending platforms and investors. For platforms, the improved accuracy in default risk prediction can lead to more effective risk management strategies and potentially lower default rates[29]. This, in turn, can enhance the overall stability and reputation of the platform, attracting more investors and borrowers.

The model's ability to incorporate and leverage a wide range of features, including alternative data sources, suggests that P2P platforms should focus on collecting and utilizing diverse data points to enhance their credit assessment processes. The importance of platform-specific metrics in the model's predictions underscores the value of historical borrower behavior data within the P2P ecosystem.

For investors, the proposed model offers a more reliable tool for assessing the risk-return profile of potential investments. The improved accuracy in default risk prediction can help investors make more informed decisions, potentially leading to better portfolio performance and risk management[30]. The interpretability features of the model, such as LIME explanations, provide investors with insights into the factors driving individual loan risk assessments, enabling more transparent and justifiable investment decisions.

The sensitivity analysis of key model parameters provides valuable insights for both platforms and researchers looking to implement similar deep-learning approaches. The optimal ranges identified for CNN filters, LSTM units, and dropout rates can serve as guidelines for model architecture design and hyperparameter tuning in future applications[31][32].

C. Limitations of the Current Study

While this research presents significant advancements in P2P microloan default risk prediction, several limitations should be acknowledged. The study is based on data from a single P2P lending platform, which may limit the generalizability of the findings to other platforms or geographical regions with different lending practices or regulatory environments[33][34]. Future research should validate the model's performance across multiple platforms and diverse geographical contexts.

The temporal nature of P2P lending data presents challenges in model evaluation and deployment. While the study attempted to mitigate this by using a time-based split for training, validation, and testing, the model's performance may degrade over time due to concept drift[35][36]. Implementing adaptive learning techniques to update the model periodically could address this limitation in real-world applications.

The interpretability of deep learning models remains a challenge, despite the use of techniques like SHAP and LIME. While these methods provide valuable insights, they may not capture the full complexity of the model's decision-making process[37]. Further research into advanced interpretability techniques for deep learning models in financial applications is warranted.

Lastly, the study focused primarily on the binary classification of loans into default and non-default categories. Future work could extend this approach to predict the probability of default or the expected loss given default, providing a more nuanced view of credit risk for P2P investors [38].

Despite these limitations, this study represents a significant step forward in applying deep learning techniques to P2P microloan default risk prediction. The proposed model's superior performance and the insights gained from its analysis contribute to the growing body of knowledge in financial technology and credit risk assessment[39].

ACKNOWLEDGMENT

I would like to express my sincere gratitude to Yida Zhu, Siwei Xia, Ming Wei, Yanli Pu, and Zeyu Wang for their groundbreaking research on AI-enhanced administrative prosecutorial supervision in financial big data, as published in their article titled "AI-Enhanced Administrative Prosecutorial Supervision in Financial Big Data: New Concepts and Functions for the Digital Era"[40]. Their innovative approach to leveraging artificial intelligence in financial supervision has significantly influenced my understanding of advanced techniques in fraud detection and has provided valuable inspiration for my research in this critical area.

I would also like to extend my heartfelt appreciation to Xiong Ke, Lin Li, Zeyu Wang, and Guanghe Cao for their innovative study on dynamic credit risk assessment using deep reinforcement learning, as published in their article titled "A Dynamic Credit Risk Assessment Model Based on Deep Reinforcement Learning"[41]. Their comprehensive analysis and deep learning approaches have significantly enhanced my knowledge of credit risk assessment and inspired my research in this field.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES:

[1] A. K. Sharma, L. H. Li, and R. Ahmad, "Identifying and predicting default borrowers in P2P lending platform: A machine learning approach," in *2021 IEEE International Conference on Social Sciences and Intelligent Management (SSIM)*, 2021, pp. 1–5. IEEE. Available From: https://doi.org/10.1109/SSIM49526.2021.9555200

[2] D. A. A. Pertiwi, K. Ahmad, T. L. Nikmah, B. Prasetiyo, and M. A. Muslim, "Combination of Stacking with Genetic Algorithm Feature Selection to Improve Default Prediction in P2P Lending," in *2023 5th International Conference on Cybernetics and Intelligent Systems (ICORIS)*, 2023, pp. 1–5. IEEE. Available From: https://doi.org/10.1109/ICORIS60118.2023.10352271

[3] V. S. Ha, D. N. Lu, G. S. Choi, H. N. Nguyen, and B. Yoon, "Improving credit risk prediction in online peer-to-peer (P2P) lending using feature selection with deep learning," in *2019 21st International Conference on Advanced Communication Technology (ICACT)*, 2019, pp. 511–515. IEEE. Available From: https://doi.org/10.23919/ICACT.2019.8701943

[4] S. Paul, A. Gupta, A. K. Kar, and V. Singh, "An Automatic Deep Reinforcement Learning Based Credit Scoring Model using Deep-Q Network for Classification of Customer Credit Requests," in *2023 IEEE International Symposium on Technology and Society (ISTAS)*, 2023, pp. 1–8. IEEE. Available From: https://doi.org/10.1109/ISTAS57930.2023.10306111

[5] T. Kumbhar, D. Agrawal, L. Saldanha, and D. Koshti, "AI-Driven Credit Scoring and Credit Line Solution for the Unreserved and Self-Employed," in *2024 Second International Conference on Inventive Computing and Informatics (ICICI)*, 2024, pp. 178–184. IEEE. Available From: https://doi.org/10.1109/ICICI62254.2024.00039

[6] P. Yu, V. Y. Cui, and J. Guan, "Text classification by using natural language processing," in *Journal of Physics: Conference Series*, vol. 1802, no. 4, p. 042010, 2021. IOP Publishing. Available From: https://doi.org/10.1088/1742-6596/1802/4/042010

[7] Y. Zhu, K. Yu, M. Wei, Y. Pu, and Z. Wang, "AI-Enhanced Administrative Prosecutorial Supervision in Financial Big Data: New Concepts and Functions for the Digital Era," *Social Science Journal for Advanced Research*, vol. 4, no. 5, pp. 40–54, 2024. Available From: https://doi.org/10.5281/zenodo.13766965

[8] X. Ma, Z. Wang, X. Ni, and G. Ping, "Artificial intelligence-based inventory management for retail supply chain optimization: a case study of customer retention and revenue growth," *Journal of Knowledge Learning and Science Technology*, vol. 3, no. 4, pp. 260–273, 2024. Available From : https://doi.org/10.60087/jklst.v3.n4.p260

[9] X. Ni, Y. Zhang, Y. Pu, M. Wei, and Q. Lou, "A Personalized Causal Inference Framework for Media Effectiveness Using Hierarchical Bayesian Market Mix Models," *Journal of Artificial Intelligence and Development*, vol. 3, no. 1, 2024.

[10] B. Yuan, G. Cao, J. Sun, and S. Zhou, "Optimising AI Workload Distribution in Multi-Cloud Environments: A Dynamic Resource Allocation Approach," *Journal of Industrial Engineering and Applied Science*, vol. 2, no. 5, pp. 68–79, 2024. Available From: https://doi.org/10.5281/zenodo.13863194

[11] X. Zhan, Y. Xu, and Y. Liu, "Personalized UI Layout Generation using Deep Learning: An Adaptive Interface Design Approach for Enhanced User Experience," *Journal of Artificial Intelligence and Development*, vol. 3, no. 1, 2024.

[12] S. Zhou, W. Zheng, Y. Xu, and Y. Liu, "Enhancing User Experience in VR Environments through AI-Driven Adaptive UI Design," *Journal of Artificial Intelligence General Science (JAIGS)*, vol. 6, no. 1, pp. 59–82, 2024. Available From: https://doi.org/10.60087/jaigs.v6i1.230

[13] S. Wang, H. Zhang, S. Zhou, J. Sun, and Q. Shen, "Chip Floorplanning Optimization Using Deep Reinforcement Learning," *International Journal of Innovative Research in Computer Science & Technology*, vol. 12, no. 5, pp. 100–109, 2024. Available From: https://doi.org/10.55524/ijircst.2024.12.5.14

[14] M. Wei, Y. Pu, Q. Lou, Y. Zhu, and Z. Wang, "Machine Learning-Based Intelligent Risk Management and Arbitrage System for Fixed Income Markets: Integrating High-Frequency Trading Data and Natural Language Processing," *Journal of Industrial Engineering and Applied Science*, vol. 2, no. 5, pp. 56–67, 2024. Available From: https://doi.org/10.5281/zenodo.13858262

[15] S. Wang, H. Zheng, X. Wen, and S. Fu, "Distributed High-Performance Computing Methods for Accelerating Deep Learning Training," *Journal of Knowledge Learning and Science Technology*, vol. 3, no. 3, pp. 108–126, 2024. Available From: https://doi.org/10.60087/jklst.v3.n3.p108-126

[16] B. Wang, H. Zheng, K. Qian, X. Zhan, and J. Wang, "Edge Computing and AI-Driven Intelligent Traffic Monitoring and Optimization," *Applied and Computational Engineering*, vol. 77, pp. 225–230, 2024. Available From: https://doi.org/10.54254/2755-2721/77/2024MA0062

[17] S. Wang, K. Xu, and Z. Ling, "Deep Learning-Based Chip Power Prediction and Optimization: An Intelligent EDA Approach," *International Journal of Innovative Research in Computer Science & Technology*, vol. 12, no. 4, pp. 77–87, 2024. Available From: https://doi.org/10.55524/ijircst.2024.12.4.13

[18] K. Xu, H. Zhou, H. Zheng, M. Zhu, and Q. Xin, "Intelligent Classification and Personalized Recommendation of E-commerce Products Based on Machine Learning," *arXiv preprint*, arXiv:2403.19345, 2024. Available From: https://doi.org/10.48550/arXiv.2403.19345

[19] K. Xu, H. Zheng, X. Zhan, S. Zhou, and K. Niu, "Evaluation and Optimization of Intelligent Recommendation System Performance with Cloud Resource Automation Compatibility," 2024. Available From: https://www.preprints.org/manuscript/202407.2199

[20] H. Zheng, K. Xu, H. Zhou, Y. Wang, and G. Su, "Medication Recommendation System Based on Natural Language Processing for Patient Emotion Analysis," *Academic Journal of Science and Technology*, vol. 10, no. 1, pp. 62–68, 2024. Available From: https://doi.org/10.54097/v160aa61

[21] H. Zheng, J. Wu, R. Song, L. Guo, and Z. Xu, "Predicting Financial Enterprise Stocks, and Economic Data Trends Using Machine Learning Time Series Analysis," *Applied and Computational Engineering*, vol. 87, pp. 26–32, 2024. Available From: https://www.preprints.org/manuscript/202407.0895

[22] B. Liu and Y. Zhang, "Implementation of Seamless Assistance with Google Assistant Leveraging Cloud Computing," *Journal of Cloud Computing*, vol. 12, no. 4, pp. 1–15, 2023. Available from: https://doi.org/10.54254/2755-2721/64/20241383

[23] M. Zhang, B. Yuan, H. Li, and K. Xu, "LLM-Cloud Complete: Leveraging Cloud Computing for Efficient Large Language Model-Based Code Completion," *Journal of Artificial Intelligence General Science (JAIGS)*, vol. 5, no. 1, pp. 295–326, 2024. Available From: https://doi.org/10.60087/jaigs.v5i1.200

[24] P. Li, Y. Hua, Q. Cao, and M. Zhang, "Improving the Restore Performance via Physical-Locality Middleware for Backup Systems," in *Proc. of the 21st Int. Middleware Conf.*, pp. 341–355, Dec. 2020. Available From: https://doi.org/10.1145/3423211.3425691

[25] S. Zhou, B. Yuan, K. Xu, M. Zhang, and W. Zheng, "The Impact of Pricing Schemes on Cloud Computing and Distributed Systems," *Journal of Knowledge Learning and Science Technology*, vol. 3, no. 3, pp. 193–205, 2024. Available From: https://doi.org/10.60087/jklst.v3.n3.p206-224

[26] F. Shang, F. Zhao, M. Zhang, J. Sun, and J. Shi, "Personalized Recommendation Systems Powered by Large Language Models: Integrating Semantic Understanding and User Preferences," *International Journal of Innovative Research in Engineering and Management*, vol. 11, no. 4, pp. 39–49, 2024. Available From: https://doi.org/10.55524/ijirem.2024.11.4.6

[27] J. Sun, X. Wen, G. Ping, and M. Zhang, "Application of News Analysis Based on Large Language Models in Supply Chain Risk Prediction," *Journal of Computer Technology and*

Applied Mathematics, vol. 1, no. 3, pp. 55–65, 2024. Available From: https://doi.org/10.5281/zenodo.13377298

[28] F. Zhao, M. Zhang, S. Zhou, and Q. Lou, "Detection of Network Security Traffic Anomalies Based on Machine Learning KNN Method," *Journal of Artificial Intelligence General Science (JAIGS)*, vol. 1, no. 1, pp. 209–218, 2024. Available From: https://doi.org/10.60087/jaigs.v1i1.213

[29] C. Ju and Y. Zhu, "Reinforcement Learning Based Model for Enterprise Financial Asset Risk Assessment and Intelligent Decision Making," 2024. Available From: https://www.preprints.org/manuscript/202410.0698

[30] K. Yu, et al., "Loan Approval Prediction Improved by XGBoost Model Based on Four-Vector Optimization Algorithm," 2024. Available From: https://www.preprints.org/manuscript/202410.0783

[31] S. Zhou, J. Sun, and K. Xu, "AI-Driven Data Processing and Decision Optimization in IoT through Edge Computing and Cloud Architecture," 2024. Available From: https://www.preprints.org/manuscript/202410.0736

[32] J. Sun, S. Zhou, X. Zhan, and J. Wu, "Enhancing Supply Chain Efficiency with Time Series Analysis and Deep Learning Techniques," 2024. Available From: https://www.preprints.org/manuscript/202409.0983

[33] H. Zheng, K. Xu, M. Zhang, H. Tan, and H. Li, "Efficient Resource Allocation in Cloud Computing Environments Using AI-Driven Predictive Analytics," *Applied and Computational Engineering*, vol. 82, pp. 6–12, 2024. Available From: https://doi.org/10.54254/2755-2721/82/2024GLG0055

[34] S. Wang, H. Zheng, X. Wen, K. Xu, and H. Tan, "Enhancing Chip Design Verification through AI-Powered Bug Detection in RTL Code," *Applied and Computational Engineering*, vol. 92, pp. 27–33, 2024. Available From: https://doi.org/10.54254/2755-2721/92/20241685

[35] C. Che, Z. Huang, C. Li, H. Zheng, and X. Tian, "Integrating Generative AI into Financial Market Prediction for Improved Decision Making," *arXiv preprint*, arXiv:2404.03523, 2024. Available From: https://arxiv.org/abs/2404.03523

[36] C. Che, H. Zheng, Z. Huang, W. Jiang, and B. Liu, "Intelligent Robotic Control System Based on Computer Vision Technology," *arXiv preprint*, arXiv:2404.01116, 2024. Available From: https://arxiv.org/abs/2404.01116

[37] Y. Jiang, Q. Tian, J. Li, M. Zhang, and L. Li, "The Application Value of Ultrasound in the Diagnosis of Ovarian Torsion," *International Journal of Biology and Life Sciences*, vol. 7, no. 1, pp. 59–62, 2024. Available From: https://doi.org/10.54097/nnvdz490

[38] L. Li, X. Li, H. Chen, M. Zhang, and L. Sun, "Application of AI-Assisted Breast Ultrasound Technology in Breast Cancer Screening," *International Journal of Biology and Life Sciences*, vol. 7, no. 1, pp. 1–4, 2024. Available From: https://doi.org/10.54097/1y59dx48

[39] L. Lijie, P. Caiying, S. Liqian, Z. Miaomiao, and J. Yi, "The Application of Ultrasound Automatic Volume Imaging in Detecting Breast Tumors," 2024. Available From: https://shorturl.at/dzfby

[40] Y. Zhu, K. Yu, M. Wei, Y. Pu, and Z. Wang, "AI-Enhanced Administrative Prosecutorial Supervision in Financial Big Data: New Concepts and Functions for the Digital Era," *Social Science Journal for Advanced Research*, vol. 4, no. 5, pp. 40–54, 2024. Available From: https://doi.org/10.5281/zenodo.13766965

[41] X. Ke, L. Li, Z. Wang, and G. Cao, "A Dynamic Credit Risk Assessment Model Based on Deep Reinforcement Learning," *Academic Journal of Natural Science*, vol. 1, no. 1, pp. 20–31, 2024. Available From: https://doi.org/10.5281/zenodo.13905241