VIET NAM NATIONAL UNIVERSITY HO CHI MINH CITY

UNIVERSITY OF ECONOMICS AND LAW

FACULTY OF INFORMATION SYSTEMS

GRADUATION THESIS REPORT

Enhancing Default Prediction in P2P Lending with Deep Model Fusion: Exploiting Sequential Structures in Transactional Data

Supervisor: Assoc. Prof. Ho Trung Thanh, Ph.D

Student:

Student ID: K214061258

Full name: Tran Si Dan

Class: K21406

Ho Chi Minh City, March/2025

VIET NAM NATIONAL UNIVERSITY HO CHI MINH CITY

UNIVERSITY OF ECONOMICS AND LAW

FACULTY OF INFORMATION SYSTEMS

GRADUATION THESIS REPORT

Enhancing Default Prediction in P2P Lending with Deep Model Fusion: Exploiting Sequential Structures in Transactional Data

Supervisor: Assoc. Prof. Ho Trung Thanh, Ph.D

Student:

Student ID: K214061258

Full name: Tran Si Dan

Class: K21406

Ho Chi Minh City, March/2025

1. ACKNOWLEDGEMENT

As I have reached the conclusion of my academic journey in Management Information Systems at the University of Economics and Law, I would like to express my sincere gratitude to those who have supported and guided me throughout this process.

First and foremost, I am deeply grateful to Professor Ho Trung Thanh, my thesis advisor, for his invaluable guidance, insightful feedback, and unwavering encouragement. His expertise and dedication have been instrumental in shaping the direction of this research.

I would also like to extend my appreciation to the esteemed lecturers at the University of Economics and Law, whose knowledge and passion have greatly enriched my understanding and academic growth. Their support and mentorship have been crucial in my development.

Furthermore, I am profoundly thankful to my family for their unconditional love, patience, and encouragement. Their support has been my greatest motivation, allowing me to persevere through challenges and strive for excellence.

Lastly, I would like to acknowledge my friends and peers who have shared this journey with me. Their companionship, discussions, and encouragement have made this experience more rewarding. With deep gratitude, I dedicate this thesis to everyone who has contributed to my academic and personal growth.

Student: Tran Si Dan

2. COMMITMENT

I, Tran Si Dan declare that the thesis titled " Enhancing Default Prediction in P2P Lending with Deep Model Fusion: Exploiting Sequential Structures in Transactional Data" was conducted under the supervision of Professor Ho Trung Thanh.

The research data, experimental results, and analyses presented in this thesis are genuine and were obtained through my own efforts, following the outlined research methodology.

All sources and related studies referenced in this thesis have been properly cited. Aside from the cited works, this research has not been published elsewhere.

I take full responsibility for the integrity and authenticity of this work.

Student: Tran Si Dan

3. TABLE OF CONTENTS

4. LIST OF FIGURES

6. LIST OF TABLES

8. ABBREVIATION

Abbreviation	Full Term
AUC-ROC	Area Under the Receiver Operating Characteristic Curve
CNN	Convolutional Neural Network
DL	Deep Learning
LDA	Linear Discriminant Analysis
LSTM	Long Short-Term Memory
MDA	Multivariate Discriminant Analysis
ML	Machine Learning
MLP	Multilayer Perceptron
P2P	Peer-to-Peer (Lending)
QDA	Quadratic Discriminant Analysis
RNN	Recurrent Neural Network
SHAP	Shapley Additive Explanations
SME	Small and Medium Enterprises
XGBoost	Extreme Gradient Boosting

9. ABSTRACT

The rapid expansion of peer-to-peer (P2P) lending has introduced new opportunities for credit access, particularly for borrowers with limited credit history and SMEs. However, the inherent risks associated with unsecured loans necessitate robust default prediction models. Traditional statistical and conventional machine learning (ML) methods have been widely used but often fail to fully capture the evolving nature of borrower behavior over time. To address this limitation, this research explores deep model fusion approaches that integrate sequential borrower transaction data with static financial attributes.

This study evaluates multiple deep learning architectures, including Long Short-Term Memory (LSTM) networks, Transformers, and Multilayer Perceptrons (MLPs), to enhance predictive accuracy in credit risk assessment. A series of deep learning models and deep ensemble models, LSTM, LSTM-Attention, Concatenated LSTM-MLP, Adaptive LSTM-MLP, Weighted LSTM-MLP, and Weighted Transformer-LSTM-MLP, are proposed and benchmarked against conventional ML models. The Bondora P2P lending dataset is employed for empirical evaluation, with models assessed using accuracy, precision, recall, F1-score, and AUC-ROC metrics. Also, a novel model interpreting method is introduced to provide insights into the temporal importance of borrower behavior over time through attention weight visualization. This mechanism enhances model transparency by highlighting which features and time steps most influence default predictions.

Experimental results demonstrate that fusion models significantly outperform traditional approaches, with the LSTM-Attention achieving the highest accuracy (0.8) and precision (0.81), indicating superior discriminative ability in identifying defaulting borrowers. The findings highlight the advantages of incorporating both temporal and static borrower features, improving predictive robustness and credit risk assessment.

Keywords: Peer-to-peer lending, default prediction, deep learning, LSTM, Transformer, model fusion, credit risk assessment

10. TÓM TẮT

Sự phát triển nhanh chóng của mô hình cho vay ngang hàng (P2P lending) đã mở ra nhiều cơ hội tiếp cận tín dụng mới, đặc biệt đối với những người vay có lịch sử tín dụng hạn chế và các doanh nghiệp vừa và nhỏ. Tuy nhiên, rủi ro vốn có của các khoản vay không có tài sản đảm bảo đòi hỏi phải có những mô hình dự đoán vỡ nợ mạnh mẽ. Các phương pháp thống kê truyền thống và học máy đã được sử dụng rộng rãi, nhưng thường không thể nắm bắt đầy đủ bản chất thay đổi theo thời gian của hành vi người vay. Để khắc phục hạn chế này, nghiên cứu này khám phá các phương pháp kết hợp mô hình sâu, tích hợp dữ liệu chuỗi theo trình tự giao dịch của người vay với các thuộc tính tài chính tĩnh.

Nghiên cứu đánh giá nhiều kiến trúc học sâu, bao gồm mạng LSTM (Long Short-Term Memory), Transformer và MLP (Multilayer Perceptron), nhằm nâng cao độ chính xác trong đánh giá rủi ro tín dụng. Một loạt các mô hình học sâu và mô hình tổ hợp sâu, bao gồm LSTM, LSTM-Attention, Concatenated LSTM-MLP, Adaptive LSTM-MLP, Weighted LSTM-MLP và Weighted Transformer-LSTM-MLP, được đề xuất và so sánh với các mô hình học máy truyền thống. Dữ liệu từ nền tảng cho vay ngang hàng Bondora được sử dụng để kiểm chứng thực nghiệm, với các mô hình được đánh giá dựa trên các chỉ số như độ chính xác, độ chính xác phân loại, độ bao phủ, F1-score và AUC-ROC. Đồng thời, nghiên cứu áp dụng cơ chế attention để trực quan hóa mức độ ảnh hưởng theo thời gian của hành vi người vay, giúp nâng cao tính minh bạch trong dự đoán vỡ nợ.

Kết quả thực nghiệm cho thấy các mô hình kết hợp vượt trội đáng kể so với phương pháp truyền thống, trong đó LSTM-Attention đạt độ chính xác cao nhất (0.8) và độ chính xác phân loại cao nhất (0.81), thể hiện khả năng phân biệt vượt trội trong việc nhận diện người vay có nguy cơ vỡ nợ. Những phát hiện này nhấn mạnh lợi ích của việc kết hợp cả đặc trưng theo thời gian và đặc trưng tĩnh của người vay, từ đó nâng cao độ tin cậy trong dự đoán và đánh giá rủi ro tín dụng.

Từ khóa: Cho vay ngang hàng, dự đoán vỡ nợ, học sâu, LSTM, Transformer, kết hợp mô hình, đánh giá rủi ro tín dụng

CHAPTER 1: INTRODUCTION

This chapter introduces the context and motivation for enhancing default prediction in peer-to-peer (P2P) lending platforms. It outlines the research objectives, scope, and methodology adopted in the study, emphasizing the integration of sequential borrower behavior and static features using deep learning models. Additionally, it defines the research process and highlights the contributions of the study to credit risk assessment in P2P lending.

1.1 Research motivation

The rapid growth of peer-to-peer (P2P) lending has transformed the financial landscape by providing an alternative credit channel outside traditional banking institutions. This lending model has significantly expanded access to credit, particularly for small and medium enterprises (SMEs) and individual borrowers with limited credit history (Abbasi et al., 2021; Loan & Trang, 2023). However, despite its potential benefits, credit risk assessment remains a major challenge, as P2P loans are largely unsecured and borrower default rates tend to be higher than in traditional lending (Suryono et al., 2019). Thus, improving default prediction models is crucial for enhancing risk management strategies and ensuring the sustainability of P2P lending platforms.

In credit risk assessment, the traditional statistical models comprise linear discriminant analysis (LDA), logistic regression (LR), multivariate discriminant analysis (MDA), quadratic discriminant analysis (QDA), factor analysis (FA), risk index models, and conditional probability model among others still belong to the most popular tools in some famous international rating agencies, however, they tend to overlook the complex nature, boundaries and interrelationships of the financial variables (Chen, Ribeiro & Chen, 2016). This limitation is particularly concerning in P2P lending, where many borrowers lack formal credit histories and alternative credit evaluation techniques are required.

Machine learning (ML) has become a valuable asset in credit risk prediction, offering greater accuracy and adaptability compared to traditional methods. In the context of P2P lending, numerous studies have shown that ML techniques, including Random Forest and XGBoost, consistently outperform conventional credit scoring models on lending platforms (Niu, Ren & Li, 2019; Setiawan, 2019; Aleksandrova, 2021; Mesri, Tahseen & Ogla, 2021). Despite enhancing predictive performance, these ML-based models often overlook the temporal dynamics of borrower behavior, as they typically consider borrower characteristics as fixed rather than evolving over time. Furthermore, while tree-based ensemble models and neural networks are commonly employed in ML-based credit risk assessment, their lack of interpretability poses challenges for regulators and end users (Ariza-Garzón, Segovia-Vargas & Arroyo, 2021).

Given the limitations of conventional credit scoring and ML models, researchers have increasingly turned to deep learning techniques and see significant improvement in terms of predictive power (Zhang, Niu & Liu, 2020). Particularly, models capable of processing sequential borrower behavior, such as Long Short-Term Memory (LSTM) networks have proven effective in modeling time-dependent patterns, allowing for improved default risk prediction by analyzing historical borrower transactions (Liang & Cai, 2020; Fu et al., 2020; Liu et al., 2023). Additionally, Transformer-based models have shown promising results in sequential financial forecasting, capturing long-range dependencies and improving predictive robustness (Korangi et al., 2023; Xia et al., 2024).

However, single deep learning models still face challenges related to bias, overfitting, and interpretability. This has led to an increased focus on deep model fusion, where multiple models are integrated to leverage their individual strengths. Various studies have shown that hybrid models combining multiple deep learning architectures, such as LSTM, Transformer, and MLP, outperform standalone models in both accuracy and generalization (Kun, Weibing & Jianlin, 2020; Chang et al., 2022;).

Despite the progress in applying machine learning and deep learning to credit risk assessment, several gaps remain:

- Most existing models do not effectively make use of both sequential transaction data and static borrower attributes, potentially leading to suboptimal predictions.

- Most machine learning and deep learning models, while accurate, are often interpreted only through the contribution of features to the prediction, ignoring the sequential nature of observations.
- Many studies overlook the potential of deep model fusion in combining multiple diverse architectures to improve prediction robustness and reduce bias.

With the expansion of P2P lending, effective credit risk prediction is becoming increasingly important for lenders, investors, and policymakers. Conventional models or even old machine learning methods fail to capture the complexities of borrower behavior, while standalone deep learning models face challenges in interpretability and bias. Therefore, this research proposes a fusion-based deep learning approach that integrates both sequential and static data, representing a promising direction for improving credit risk assessment in P2P lending. By leveraging sequential borrower behavior, static features, and adaptive feature weighting, this study seeks to develop a modelling approach that not only improves predictive performance but also has potential to leverage interpretability in the matter of temporal patterns for practical financial applications.

Moreover, beyond model construction, this research introduces a novel interpretation method that utilizes attention weights to visualize the model's temporal focus across different features and time steps. This interpretability mechanism enables stakeholders to understand which behavioral patterns contribute most to the default prediction at various stages in the loan history, offering valuable insights for risk analysts and financial decision-makers.

1.2 Research objectives

General Objective: This research aims to significantly improve the predictive accuracy and interpretability of default prediction in peer-to-peer (P2P) lending platforms by systematically integrating both sequential borrower history data and static tabular borrower attributes through advanced deep learning fusion architectures.

Specific Objectives:

- Performance evaluation and comparative analysis of conventional machine learning methods (Random Forest, Logistic Regression, K-Nearest Neighbors, and traditional Multilayer Perceptron) and proposed deep model fusion architectures (LSTM, LSTM-Attention, Concatenated LSTM-MLP, Adaptive LSTM-MLP, Weighted LSTM-MLP, Weighted Transformer-LSTM-MLP)

- Assessment of fusion strategies, particularly focusing on the proposed concatenating, adaptive and learnable weighting mechanisms, and their impacts on predictive accuracy, precision, recall, and AUC metrics.
- Assessment of the use of sequential transaction data in modelling to evaluate its significance in enhancing the accuracy and reliability of borrower default predictions compared to static-only borrower attributes.
- Examination of computational trade-offs between traditional models and fusion-based deep learning models, emphasizing training complexity, inference efficiency, and resource consumption during model deployment.

1.3 Research object and scope

1.3.1 Research objects

Research objects of this topic are listed below:

- **Borrower transaction sequences**: Representing users' historical loan activities to capture behavioral patterns and time-dependent financial behaviors.
- **Sequential deep learning models**: Particularly Long Short-Term Memory (LSTM) networks and Transformer architectures designed to learn temporal dependencies in transaction data.
- **Fusion-based deep learning models**: Ensemble architectures that combine outputs from sequential models (LSTM, Transformer) and feedforward models (MLP) through concatenation, adaptive selection, or learnable weight mechanisms.
- **Interpretability mechanisms**: Attention layers and weight visualization methods used to identify which features and time steps contribute most to the model's decision-making.

1.3.2 Research scope

- Dataset: The study uses the Bondora P2P lending dataset (available on IEEE Dataport), which contains loan transaction records, borrower profiles, and repayment histories. The retrieved data is a pool of both defaulted and non-defaulted loans from the time period between 1st March 2009 and 27th January 2020.

- Evaluation Metrics: The models are assessed using accuracy, precision, recall, F1-score, and AUC-ROC to measure predictive performance.
- Methodology: The research focuses on supervised learning techniques, excluding unsupervised or reinforcement learning.
- Practical Application: The findings can help P2P lending platforms optimize risk assessment, enhance investor decision-making, and improve loan pricing strategies.
- Generality: The proposed methodology can be extended to other financial risk prediction tasks, including credit scoring and fraud detection.

1.4 Research method

To enhance default prediction in peer-to-peer (P2P) lending, this research employs deep learning techniques that integrate both sequential and tabular data. The study follows a structured methodology consisting of two main steps: Model Development and Performance Evaluation. The detailed outline of each step is provided below:

1.4.1 Model Development

- Data Preprocessing: The dataset is cleaned and transformed, including handling missing values, encoding categorical variables, and normalizing numerical features. Sequential loan transaction data is structured for time-series analysis.
- Feature Engineering: Both sequential (transaction history) and static (borrower profile, loan details) features are extracted for model training.
- Model Construction: Deep learning architectures, including LSTM-MLP, Transformer-LSTM-MLP, and other fusion models, are developed alongside traditional models such as Random Forest, KNN, MLP, and Logistic Regression for comparative analysis.
- Fusion Strategies: Different fusion techniques (e.g., conditional fusion) are implemented to assess their impact on prediction accuracy.

1.4.2 Performance Evaluation

- Training and Testing: Models are trained using supervised learning techniques and evaluated on a separate test dataset.
- Evaluation Metrics: Standard classification metrics, including accuracy, precision, recall, F1-score, and AUC-ROC, are used to measure predictive performance.
- Comparative Analysis: The effectiveness of deep learning models is compared against traditional machine learning models to determine performance improvements.

1.5 Research process

Phase 1: Literature Review

- Examining existing studies on default prediction in P2P lending and the application of machine learning and deep learning models.
- Identifying gaps in current approaches, particularly regarding the integration of sequential and tabular data for risk assessment.

Phase 2: Data Collection and Preprocessing

- Utilizing the Bondora P2P lending dataset, which includes borrower profiles, loan details, and repayment histories.
- Cleaning and transforming data, including handling missing values, encoding categorical features, and normalizing numerical attributes.

Phase 3: Model Development and Training

- Structuring data to preserve sequential information, ensuring that borrowing patterns remain intact for model training.
- Constructing ensemble models, including LSTM-based fusion architectures, to capture both sequential and static borrower characteristics.
- Implementing traditional machine learning models (Random Forest, KNN, MLP, and Logistic Regression) as baselines for comparison.
- Applying fusion strategies, including conditional fusion, to optimize the interaction between sequential and tabular features.

Phase 4: Model Evaluation and Analysis

- Training and testing models on separate datasets using supervised learning techniques.
- Evaluating model performance based on classification metrics such as accuracy, precision, recall, F1-score, and AUC-ROC.
- Comparing fusion models against traditional machine learning models to determine improvements in default prediction accuracy and risk assessment.

1.6 Research contribution

This research contributes to improving default prediction models' accuracy in P2P lending by:

- Developing fusion models that integrate sequential borrower behavior and static financial attributes, improving predictive performance.
- Comparing deep learning-based fusion models with traditional machine learning models, offering insights into their strengths and weaknesses.
- Exploring conditional fusion techniques, allowing models to adaptively decide whether to prioritize sequential or tabular data based on transaction history.
- Providing a scalable framework for P2P lending platforms to optimize credit risk assessment, improve loan pricing, and strengthen investor decision-making.

1.7 Structure of research

Chapter 1 – Introduction: This chapter provides an overview of the research, including the motivation for studying default prediction in P2P lending, the specific research objectives, and the scope of the study. It outlines the methodology employed, the research process followed, the contributions made by the study, and the overall structure of the research document.

Chapter 2 – Literature Review and Theoretical Background: The objective of this chapter is to review existing literature on default prediction models in P2P lending, covering both traditional machine learning and deep learning approaches. It discusses credit scoring, risk assessment techniques, and fusion strategies while identifying gaps in existing research. Additionally, this chapter establishes the theoretical foundation by exploring concepts such as machine learning, artificial neural networks (ANN), LSTM networks, and deep fusion models.

Chapter 3 – Methodology: This chapter outlines the methodology used in the research, including the development of fusion models that integrate sequential and tabular data. It provides details on data preprocessing, feature engineering, model architectures (LSTM-MLP, Transformer-LSTM-MLP, etc.), training procedures, and evaluation metrics. Furthermore, it describes different fusion strategies, such as conditional fusion, and explains how model performance is assessed.

Chapter 4 – Results and Discussion: The objective of this chapter is to present the experimental results obtained from model training and evaluation. It includes a comparative analysis of deep learning sequential models, deep fusion models and conventional machine learning models (Random Forest, KNN, MLP, and Logistic Regression), focusing on key metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. The chapter discusses findings, interprets results, and explores the implications of using deep learning in default prediction.

Chapter 5 – Conclusion and Future Work: This final chapter summarizes the key findings of the research and draws conclusions based on the results. It highlights the contributions made to P2P lending risk assessment and discusses the limitations of the study. Additionally, it suggests potential areas for future improvements, including the refinement of fusion strategies, model interpretability enhancements, and broader applications in financial risk prediction.

CHAPTER 4: RESULT AND DISCUSSION

This chapter presents and analyzes the experimental outcomes obtained from model training and evaluation. It compares the performance of proposed fusion models with traditional machine learning models, discusses the effectiveness of various model architectures, and interprets the results in light of existing literature. Practical implications and managerial insights derived from attention-based interpretability are also addressed.

4.1 Experimental Result

The experiments conducted in this study aimed to rigorously evaluate and compare the predictive performance of various models on the Bondora peer-to-peer (P2P) lending dataset, specifically targeting default prediction. Given the complexity of borrower behavior, models integrating sequential borrower history and static borrower attributes were hypothesized to outperform traditional machine learning approaches, which typically rely solely on static borrower characteristics.

To objectively validate this hypothesis, two primary groups of models were evaluated: traditional machine learning models and advanced deep learning fusion architectures. The traditional models consisted of Random Forest, Logistic Regression, K-Nearest Neighbors (KNN), and Multilayer Perceptron (MLP), all widely adopted as baselines in credit risk modeling literature. On the other hand, the fusion deep learning architectures incorporated sequential transaction data processed through Long Short-Term Memory (LSTM) and Transformer encoders combined with tabular data handled by MLP. Specific fusion strategies evaluated included Concatenated LSTM-MLP, Adaptive (Selective) LSTM-MLP, Weighted LSTM-MLP, and the innovative Weighted Transformer-LSTM-MLP.

Hyperparameter optimization played a pivotal role in ensuring the fairness and robustness of model comparisons. Traditional machine learning models employed grid search methods, systematically exploring predefined parameter grids to identify optimal configurations. For deep learning fusion models, Bayesian optimization facilitated the exploration of complex hyperparameter spaces, yielding configurations tailored precisely to the dataset characteristics. Table 4.1 below provides a comprehensive

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**	**AUC-ROC**
Conventional models					
Logistic Regression	0.74	0.75	0.85	0.79	0.78
K-Nearest Neighbors (KNN)	0.75	0.76	0.85	0.8	0.8
Random Forest	**0.8**	0.79	**0.90**	**0.85**	**0.87**
MLP (Traditional)	0.79	0.78	0.9	0.84	0.85
Proposed deep learning models					
LSTM	**0.8**	**0.81**	0.87	0.84	0.86
LSTM-Attention	**0.8**	**0.81**	0.87	0.84	0.86
Concatenated LSTM-MLP	**0.8**	0.79	0.89	0.84	0.85

Adaptive (Selective) LSTM-MLP	0.79	0.79	0.88	0.83	0.84
Weighted LSTM-MLP	0.79	0.79	0.89	0.84	0.85
Weighted Transformer-LSTM-MLP	0.79	0.78	**0.9**	0.84	0.85

Table 4.2: Evaluation result

From Table 4.2, it can be observed that the proposed deep learning models achieve competitive performance, highlighting their capability in effectively capturing borrower behaviors through the integration of sequential and static data. Notably, the **LSTM** and **LSTM-Attention** models obtain the highest precision (0.81), demonstrating their strength in accurately identifying non-default borrowers and minimizing false-positive predictions, which is crucial in financial decision-making contexts.

Furthermore, the **Weighted Transformer-LSTM-MLP** achieves the highest recall (0.90) among deep learning models, emphasizing its ability to effectively detect potential defaults by thoroughly exploiting sequential repayment patterns. Other fusion models, such as **Concatenated LSTM-MLP** and **Weighted LSTM-MLP**, also show strong performance with high recall (0.89), illustrating the effectiveness of integrating sequential and static borrower features.

Although the Random Forest model achieves slightly better performance in accuracy (0.80), F1-score (0.85), and AUC-ROC (0.87), this can be partly attributed to its optimization via GridSearchCV, whereas the proposed deep learning models were only manually tuned. The narrow performance gap underscores the potential of fusion-based deep learning approaches, which effectively capture both sequential and static borrower features, and could further improve with more advanced hyperparameter optimization.

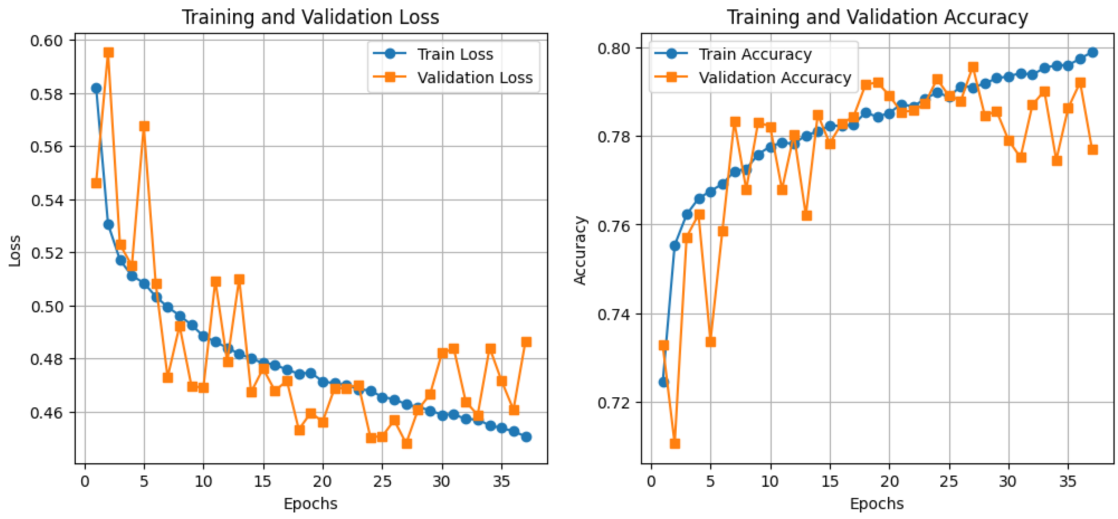

Figure 4.1: The LSTM-Attention model's learning curve

From Figure 4.1, training loss steadily decreases with increasing epochs, demonstrating effective learning on the training dataset. Validation loss also generally declines but exhibits noticeable fluctuations throughout training, indicating moderate inconsistencies or potential challenges related to model generalization. Similarly, validation accuracy shows frequent fluctuations compared to the relatively smooth and consistent increase of training accuracy. This observation emphasizes the necessity of employing strategies such as early stopping, regularization, and optimized dropout rates, as previously detailed during hyperparameter tuning. Despite the good overall learning behavior, the observed variations between training and validation performance suggest that the LSTM-Attention model could benefit from additional regularization techniques, batch normalization, or slight adjustments to network architecture to further enhance stability and generalization.

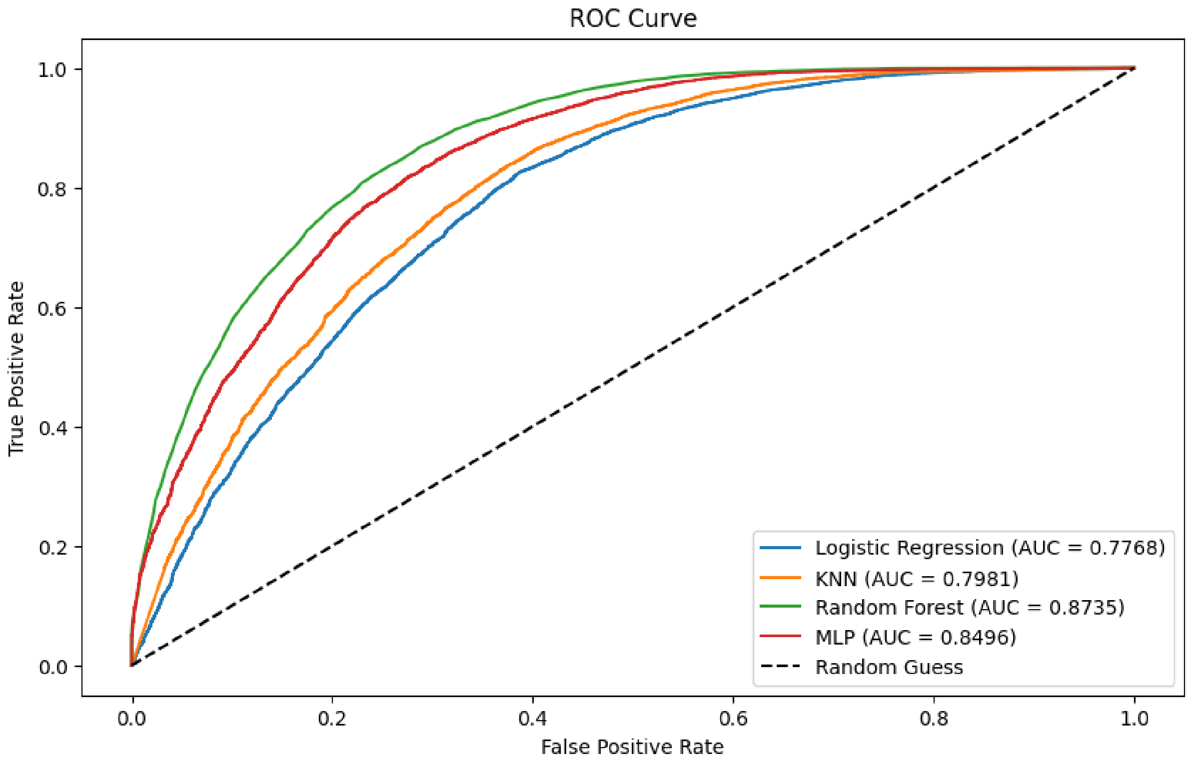

Figure 4.2: Conventional machine learning methods' learning curve

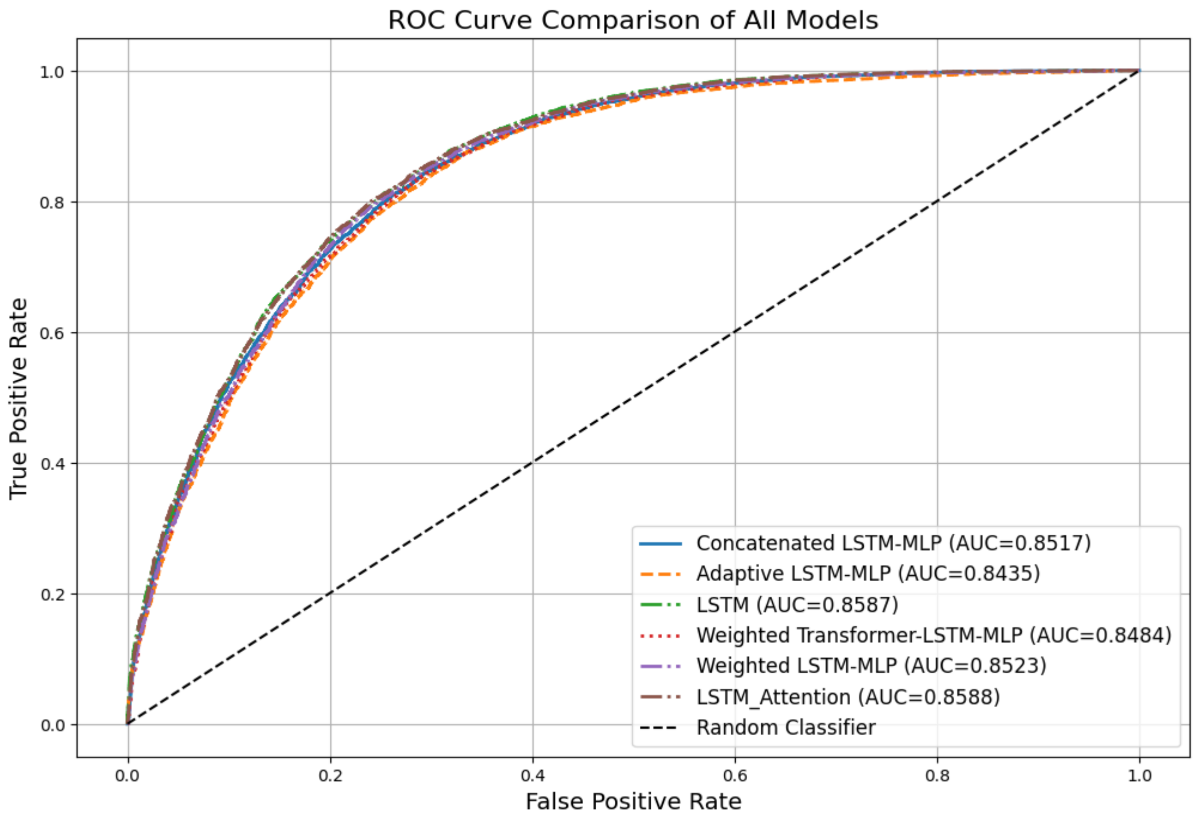

Figure 4.3: Deep ensemble models' learning curve

The ROC curves confirm the LSTM-Attention superior discriminative ability, as its curve consistently stays above the others, closely approaching the top-left corner. This suggests a strong ability to differentiate between default and non-default cases, making it a highly effective model for credit risk assessment in P2P lending. Interestingly, the Adaptive (Selective) LSTM-MLP, despite exhibiting an exceptional recall rate, demonstrates a significantly lower AUC, reflected visually by its ROC curve's proximity to the diagonal. This suggests that while it captures positive cases effectively, its overall discriminatory power is more limited compared to other deep fusion models.

Compared to conventional machine learning models (Figure 4.2), the deep learning-based fusion models demonstrate competitive discriminative performance. The Random Forest achieves the highest AUC (0.8735) among traditional methods, closely matching the top-performing deep learning models such as LSTM-Attention (AUC=0.8588) and LSTM (AUC=0.8587). However, other conventional models like Logistic Regression (AUC=0.7768) and KNN (AUC=0.7981) show significantly lower discriminative capabilities. This underscores the advantages of fusion-based deep learning architectures in capturing complex sequential borrower behaviors, positioning them as strong alternatives to conventional predictive approaches.

4.2 Discussion

The experimental findings offer valuable insights into the comparative strengths of traditional machine learning approaches and proposed deep learning fusion models for default prediction in P2P lending. Conventional methods such as Logistic Regression and K-Nearest Neighbors (KNN) generally exhibited lower predictive performance, likely due to their limited capability to model complex borrower behaviors and sequential transaction dynamics. Notably, Random Forest stood out among traditional models, achieving strong performance across multiple metrics, which highlights its proficiency in capturing intricate, non-linear interactions within static borrower attributes.

In contrast, deep learning models, particularly LSTM-based architectures, demonstrated a clear advantage in effectively modeling sequential borrower patterns, as evidenced by their superior precision and highly competitive AUC-ROC scores. The LSTM-Attention and LSTM models' ability to precisely identify non-default borrowers underscores their potential to reduce operational risks associated with false-positive predictions. Although deep learning models slightly lagged behind the optimized Random Forest—partially due to manual hyperparameter tuning—the marginal differences imply that further optimization, such as automated hyperparameter tuning or refined regularization techniques, could lead to substantial performance gains.

The experimental results of this study are consistent with, and in some aspects improve upon, previous research using the Bondora dataset. Specifically, our best-performing models—Random Forest (AUC = 0.87) and LSTM-Attention (AUC = 0.859)—demonstrate competitive predictive capabilities compared to prior work, for example, the comparative study of Disbergen (2019), who used Bondora.com dataset and showed that Random Forest achieved superior predictive accuracy compared to Logistic Regression (70.4% vs. 65.5%). In our study, Random Forest continues to demonstrate strong performance with a high AUC, consistent with Disbergen's findings, and further illustrates its scalability and potential when enhanced with deep learning techniques such as LSTM to capture borrowers' sequential behavior. Additionally, Lyócsa et al. (2022) proposed a profit scoring approach using the same dataset and demonstrated that profit scoring models outperformed traditional credit scoring methods, achieving accuracy improvements of up to 6.7% and profit gains of 24.0% over standard models. Among the traditional classifiers evaluated in their study, the Random Forest model yielded the highest accuracy at 78.8%, which remains slightly lower than the accuracy achieved by the proposed deep learning models in this research.

In summary, the results of this study confirm the effectiveness of deep fusion models in leveraging both sequential and static borrower features. However, the highest AUC observed still falls slightly short of the performance reported by some models employing ensemble models such as Random Forest, XGBoost or LightGBM. This suggests that further improvements in prediction performance may be attainable through more sophisticated hyperparameter tuning or by incorporating additional relevant features into the modeling process.

4.3 Recommendation

4.3.1 Managerial implications

The adoption of sequential pattern mining and deep model fusion techniques, as exemplified by the Weighted Transformer-LSTM-MLP and LSTM-Attention models, offers practical benefits in enhancing default prediction performance for peer-to-peer (P2P) lending platforms. Managers and decision-makers could prioritize deploying these models in scenarios demanding high predictive accuracy, especially when financial losses from borrower defaults are substantial. Given their ability to simultaneously exploit sequential borrower behaviors and static borrower attributes, these fusion models facilitate a more comprehensive understanding of borrower risk profiles compared to traditional methods.

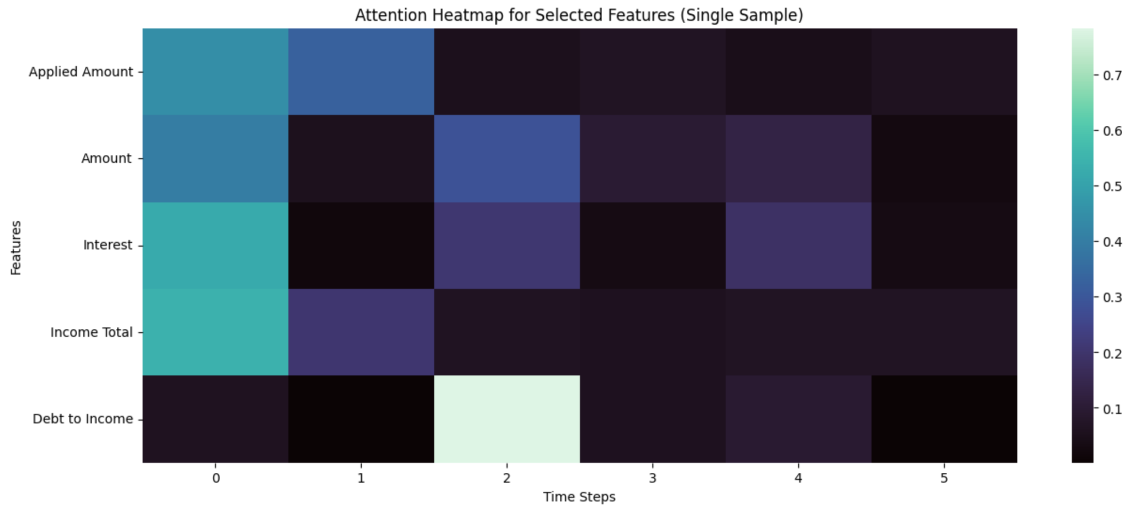

Figure 4.4: Attention weight heatmap of a single sample

Additionally, the attention mechanism in the LSTM-Attention model provides significant interpretative advantages, empowering managers with nuanced insights into borrower behaviors by highlighting which specific features and time steps are most influential in the predictive process. Visualization of attention weights allows for granular interpretation; for instance, examining the attention heatmap of a single borrower (Figure 4.4), stakeholders can observe critical behavioral points, such as the considerable emphasis placed at time step 2 (the third transaction) on the "Debt to Income" variable, indicating this feature's pivotal role in predicting potential default at this specific point in time. Similarly, “Applied Amount” receives consistent and relatively high attention at earlier time steps, most significantly on the fist transaction with attention weight higher than 0.4, reflecting its early-stage predictive relevance. “Interest” initially receives significant attention, decreasing slightly at later time steps, which may imply diminishing predictive relevance over the loan's lifespan. Meanwhile, the “Amount” variable shows small but varying attention across time, suggesting its role fluctuates significantly, possibly dependent on changing borrower conditions. “Income Total” receives moderate attention, particularly at initial time steps, highlighting its stable but less dominant predictive importance relative to other features.

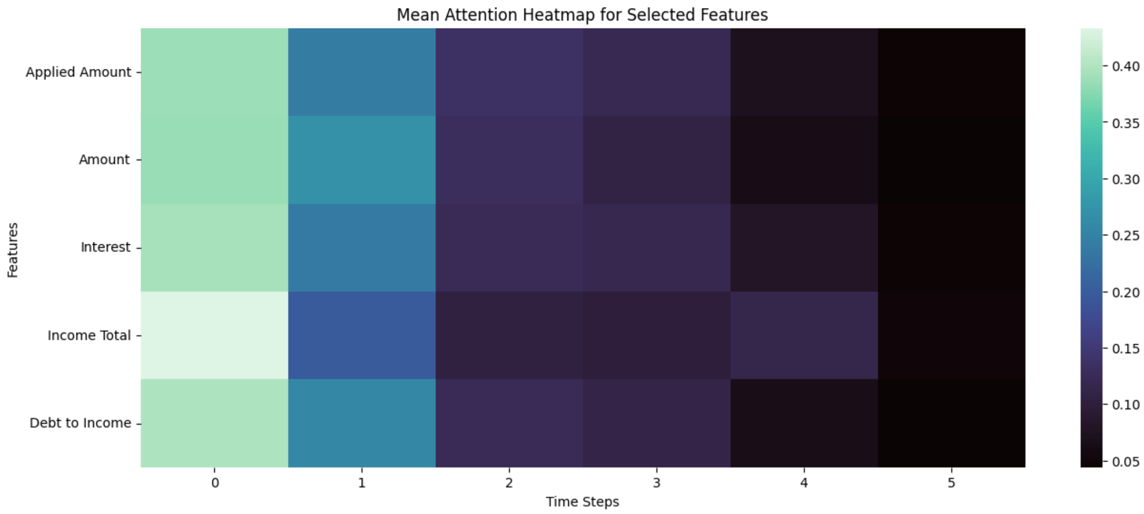

Figure 4.5: The mean attention heatmap of all samples

The mean attention heatmap (Figure 4.5) further enriches managerial insights by illustrating broader trends across multiple borrowers. The features consistently exhibits the highest mean attention across early time steps, especially the “Income Total”, underlining its significant and generalizable influence during the loan application phase. Notably, the declining attention towards later time steps suggests that early financial metrics might play a more critical role in initial risk evaluation and lending decisions, and users with more transactions possess less chance to default. “Amount” and “Debt To Income” exhibit moderate attention weights, indicating their contextual relevance across borrower profiles but highlighting the potential need for complementary analysis or features to strengthen their predictive utility.

It is also worth noting that time steps 3 to 6 display relatively high mean attention values, which may appear counterintuitive given the general decline. This is explained by the fact that the 75th percentile of the number of transactions made by users is 3; therefore, most borrower sequences do not extend beyond three time steps. As a result, the model's attention mechanism assigns greater weights to the remaining few non-padding positions in longer sequences, causing an apparent uptick in attention values at later steps.

Collectively, these attention patterns clearly demonstrate the model's effective learning of sequential borrower behavior patterns, underscoring the temporal dynamics critical to default prediction. Stakeholders can strategically utilize these insights for targeted policy-making, risk-based pricing adjustments, and proactive interventions tailored to borrower characteristics and behaviors identified through these attention patterns.

4.3.2 Practical implications

The extensive empirical analysis presented in this study clearly underscores the superior predictive capability of deep learning fusion models in identifying borrower defaults in peer-to-peer (P2P) lending platforms. Despite this overall superiority, practical considerations regarding computational resources, operational costs, and interpretability suggest nuanced recommendations for selecting the most appropriate model in specific lending scenarios.

In environments where computational resources, including training and deployment hardware, are abundant and the lending decision accuracy is of paramount importance, especially when the cost associated with defaults is substantial, those proposed ensemble deep learning models unequivocally be the preferred solution. Moreover, if lending platforms prioritize precise targeting of truly high-risk borrowers and aim to minimize unnecessary rejection of potentially profitable loans, models exhibiting higher precision should be considered.

However, practical deployments often require careful consideration of the trade-off between computational overhead and predictive performance. Deep learning models, especially architectures involving self-attention mechanisms, inherently possess higher computational costs, longer training durations, and increased complexity in deployment and maintenance compared to traditional models. In scenarios where computational resources or inference latency is critically limited, the implementation of models like Random Forest, could be justified due to their relatively lower computational cost, simpler training processes, and well-established frameworks for explainability.

Regarding real-world implementation, the deployment of deep learning fusion models demands robust infrastructural readiness and investment in adequate computational resources, training platforms, and inference optimizations. Deep learning architectures generally require specialized hardware (e.g., GPUs or cloud-based resources), advanced expertise for maintenance, and additional operational resources for periodic model updates and data engineering processes. In this study, the training process for deep learning models was conducted on a dual NVIDIA T4 GPU setup, with a total training time of 2 hours, 37 minutes, and 15 seconds, including hyperparameter tuning. For each model, eight distinct hyperparameter combinations were evaluated. Meanwhile, the training process of conventional ML models only took 17 minutes and 37 seconds. These additional investments are justified in scenarios where slight improvements in predictive accuracy translate into substantial financial benefits, typically seen in high-volume lending environments or platforms managing larger lending portfolios.

In conclusion, while the ensemble deep learning models such as the Weighted Transformer-LSTM-MLP model or the LSTM_Attention model emerges as the optimal solution from a purely performance-oriented standpoint, careful consideration of computational costs, interpretability, and practical operational requirements suggests a flexible, scenario-based approach to model deployment. Thus, P2P lending platforms are advised to implement the Weighted Transformer-LSTM-MLP model in contexts where maximum predictive accuracy translates directly into substantial financial gains. Alternatively, more lightweight yet sufficiently accurate models such as Random Forest may be optimal in resource-constrained or transparency-sensitive scenarios, respectively.

CHAPTER 5: CONCLUSION AND FUTURE WORK

This chapter summarizes the key findings and contributions of the research. It reflects on the strengths and limitations of the proposed models and discusses their implications for real-world deployment in P2P lending platforms. The chapter concludes with recommendations for future research, including improvements in interpretability, model scalability, and generalizability across different financial contexts.

5.1 Conclusion

This research systematically explored and evaluated various modeling approaches for predicting borrower defaults within peer-to-peer (P2P) lending platforms, comparing traditional machine learning methods with advanced deep learning fusion models. By effectively leveraging sequential borrower behavior and integrating it with static borrower characteristics, the study demonstrated notable improvements in predictive capabilities. These findings highlight the substantial benefits of employing deep learning fusion strategies that exploit temporal borrower behavior patterns, significantly enhancing credit risk assessment precision and reliability.

Multiple standard evaluation metrics—Accuracy, Precision, Recall, F1-score, and AUC-ROC—were utilized to comprehensively assess model performance across diverse predictive dimensions. Deep learning models, particularly those employing LSTM-based fusion architectures such as LSTM-Attention, displayed highly competitive performance with superior precision (0.81) and commendable discriminative power (AUC approximately 0.86). A distinctive strength of these models lies in their ability to exploit sequential data through the integration of attention mechanisms, which not only enhanced predictive accuracy but also provided meaningful interpretative insights via attention weights. Such insights allow for detailed understanding and explanation of borrower behaviors, significantly contributing to their practical utility and managerial decision-making. Among the conventional methods, Random Forest exhibited robust performance, achieving the highest accuracy (0.80), recall (0.90), F1-score (0.85), and AUC-ROC (0.87), clearly demonstrating its effectiveness in capturing complex borrower behaviors using static data.

Overall, this research confirms that deep learning models capable of capturing sequential borrower behaviors, enhanced by interpretative attention mechanisms, represent a powerful predictive alternative or complement to conventional machine learning methods, thereby substantially improving the effectiveness of default risk management in P2P lending environments.

5.2. Limitations

One significant limitation identified during the study is the inadequate integration of current state-of-the-art interpretability and explainability frameworks, such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), with these advanced fusion architectures. Despite their predictive strength, deep learning models—especially those employing complex self-attention mechanisms like Transformers—often operate as “black-box” systems, challenging stakeholders' trust and potentially complicating regulatory compliance. The existing interpretation methods available, primarily designed for simpler or more transparent machine learning models, are frequently insufficient or computationally impractical when directly applied to these advanced fusion architectures. Additionally, this study did not extensively explore the effects of class imbalance, potential bias in the training dataset, or model robustness against data drift over time. Another important limitation is the reliance on manual hyperparameter tuning, which might have constrained the maximum achievable performance of deep learning models, potentially underestimating their true capabilities. Addressing these interpretability and methodological limitations through further research is essential to fully realize the practical advantages and broader acceptance of deep learning fusion models in financial risk management applications.

5.3 Future work

Building on the findings and limitations identified in this research, several promising directions for future exploration and development emerge.

Firstly, it is strongly recommended to pursue an intensive exploration and integration of explainability and interpretability methods specifically tailored for deep learning fusion models, particularly those involving Transformer and LSTM architectures. Current interpretability techniques like SHAP and LIME, while powerful