Construction of an International Trade Financial Risk Assessment and Prediction Model Based on Big Data Analysis

Zeyu Liu*
School of Professional Studies, Columbia University in the City of New York, New York, 10027, USA

Abstract—**Background: International trade promotes economic growth across nations, while imposing financial risks of currency fluctuations, credit defaults, and market volatility. Although conventional methods of risk evaluation have served well in the past, they are, however, unable to provide risky international trade answers under the dynamic conditions at present. Objective: This study aims to develop data-driven risk assessment and prediction models for financial risks in international trade, with an emphasis on the China trade regime dominated by finance. The purpose is to maximize prediction accuracy and to provide pragmatic risk management solutions. Methods: The study proposed a hybrid method capable of characterizing complex nonlinear correlations by a DNN and, subsequently, estimating prediction outputs with an LR model for enhanced interpretability. Training models on the International Trade and Finance Dataset are augmented by macroeconomic indicators; preprocessing is performed via statistical imputation, feature normalization, and one-hot encoding. Results: With values of 0.9670, 0.0408, 0.0322, and 0.0017 awarded to R^2, RMSE, MAE, and MSE, respectively, the model stands out as the most capable and accurate in measuring financial risk. However, this hybrid model marries complex features with interpretable features, thereby paving the way for an exquisite instrument for risk assessment. Conclusion & Implications: This study aims to develop a solid framework for predicting financial risks in international trade that will aid financial institutions in decision-making and in developing policies. The findings may be applied to ongoing financial stability assessments for trade risk management.**

Keywords—International trade; financial risk; big data; predictive modeling; China's economy

I. Introduction

With the global economy developing fast and international trade becoming increasingly more complex, there is a greater need for the use of advanced tools in financial risk management. With globalization fostering trade flows across borders, this pattern has presented growing financial uncertainties to nations and enterprises: currency risks, credit risks, political risks, and the like. As one of the largest economies in the world, China largely depends on international trade, and with a very dynamic financial scene, the opportunity to build very rigorous tools for assessing and predicting financial risks must never be missed [1]. A reliable financial risk evaluation model provides timely insights by which governments, businesses, and financial institutions can hedge themselves against prospective losses. In this context, the current research aims to develop an innovative risk prediction model for international trade finance in China through deep learning, feature engineering, and big data analytics [2]. This will finally create a comprehensive and scalable framework to assess international trade risks and deliver conclusive predictions of interest based on financial data and economic indicators.

A. Financial Risk Assessment in International Trade

Financial risk assessment is a major element in doing cross-border trade. It involves dealing with the identification, analysis, and measurement of several risks that could affect one's financial performance or payment defaults, exchange rate fluctuations, and the economic standing of trading partners, the generally considered ones [3]. Within the scope of international trade, these risks come with a wide array of key factors, including instability through political measures, trade policies over time, and shifts in the global economic scenario.

Risk traditionally has been assessed manually and with static models, which do not keep pace with the fast-moving nature of the global markets. The more complicated and interlinked international trades become, the more the old methodologies become obsolete [4]. Thus, there are increasing demands for dynamic models that can ingest large amounts of data instantaneously and deliver instantaneous considerations on financial health. This study attempts to fill up this gap by deploying more advanced statistical and machine learning methods to create a model that could more efficaciously and accurately assess the financial risks in international trade [5].

B. Financial Risk Prediction

Financial risk prediction, which evaluates future risks and trends using historic data and economic indicators, is a byproduct of economic risk assessment. Being able to forecast financial risk is essential for companies and policymakers so that they can lessen their potential losses and optimize their decision-making processes [6]. Therefore, this study intends to employ big data analytics to forecast financial risks associated with trade transactions, credit exposures, and even changes in macroeconomic landscapes, especially with a focus on the Chinese international trade regime.

A predictive model is created by studying historical trade data comprising transactional volumes and the history of payments. Other relevant economic indicators like interest rates, inflation rates, and GDP growth are studied as well. Machine learning techniques such as DNN and hybrid regression models actually learn from the input data by modeling linear and nonlinear relationships [7]. With accurate predictions in the financial risk arena, decision-makers can prepare for challenges that may be posed by exchange rate fluctuations, changes in interest rates, or terrorist-related instability.

C. Big Data Analysis for Financial Risk in International Trade

The advent of commercial big data applications has changed how financial risks are evaluated and predicted in international trade. With the exponentially increased presence of data, it is now possible to analyze large volumes of structured and unstructured information spanning several domains, including trade records, economic reports, social media, and even news feeds.

In order to better comprehend possible financial risks than can be accomplished using traditional approaches, big data analytics provides the ability to combine large and complicated datasets while offering instantaneous analysis [8]. In this research, such tools are huge data tools emphasizing distribution and processing systems that usually process the high-velocity datasets faced in international trade scenarios. With such technology, trade data of large magnitude can be processed at a reasonable speed, thereby enhancing the effectiveness and accuracy of risk prediction. Big data makes it possible to incorporate a wide range of datasets, including historical trade data, financial indexes, and even sentiment from the news today. This improves the risk model's forecasting accuracy [9].

D. Deep Learning in Financial Risk Prediction

An area within machine learning, deep learning recognizes complex non-linear relationships within financial data. While traditional models may not be able to pick up on such complex patterns in bigger datasets, especially when patterns are so fine or intricate that they can barely be detected by simpler models, the deep neural networks detect such complex patterns and learn from huge amounts of data with impeccable accuracy.

International trade finance would mean for DNNs to determine hidden correlations among economic indicators, financial transactions, and exogenous conditions like political events or global changes in a market [10]. These deep learning algorithms, being capable of handling immense amounts of data and updating themselves on new data, will be able to generate better financial risk assessments. The study assesses the nature of DNN forces toward better recognition trends relevant to trade finance, creating a basis for the improvement of the trade finance model, both in prediction power and in robustness [11].

E. Hybrid Model for an Interpretable Prediction of Risk

These techniques, while more accurate, are often not interpretable. This disturbing subject creates an ambiguity in the mind of any analyst or decision maker wanting to know why a forecast has been given. To counter this, the proposed solution will pursue the hybrid modeling approach that exploits linear regression, hence preserving the interpretability of traditional models while enhancing predictive accuracy [12].

Linear regression is particularly relevant for determining the greatest contributors to financial risk, thus allowing a slimmed-down model to focus on aspects such as trade volume, creditworthiness, and debt-equity ratios. Linear regression equips decision-makers with interpretable coefficients that ideally tell a story let's say, how a feature is contributing toward generating a risk score, basically helping them comprehend what the model is predicting. The choice of methods hence attains a fine balance between predictability and transparency.

F. Distributed Computing for Efficient Risk Modelling

One of the foremost hurdles confronted by international trade-related financial risk prediction is handling large-scale datasets. To conquer this, the model harnesses distributed computing for efficient data processing. Processing smaller partitions of the dataset in parallel allows the model to machinate the bulk of trade and financial data much more expediently than would usually be the case [13]. It offers some scalability to the trade system, which remains an important factor as the international trade data increases in size exponentially. Parallel model training speeds up the computation and allows less time to develop and optimize a model. Additionally, they offer fault-tolerant processing, which ensures that even if a component of the system fails, the model will still function, adding dependability to applications that need to run continuously [14]. Making changes in hyperparameters is always a step toward achieving accuracy and performance optimization. The capacity to generalize and make accurate predictions will depend on the regularizers, learning rate, and the amount of neurons in a neural network structure [15]. They are, therefore, most useful for those problems with high-dimensional search spaces. Coupled with the above techniques, the system is really fine-tuned for the best performance, meaning that it can make the most accurate predictions of financial risk in international trade.

To confirm the generalization ability of the developed model towards unseen data, the data is split into a training, validation, and test subset. The model fits to the training set; then, hyperparameter tuning is done on the validation set, and finally, evaluation of the model is performed with the test set. Further assistance also comes from K-Fold Cross-Validation, stratified and repeated, in order to validate a model and ensure that it isn't overfitting to the training data. The goal of cross-validation is to train the model for various data subsets by splitting the dataset into multiple folds. The effectiveness of the model is then assessed on several data splits to further illustrate its generalizability. The model is prevented from overfitting by using K-fold cross-validation, which typically happens in high-dimensional dataset settings.

A data-driven approach to controlling financial hazards in international trade is put out in this study. The suggested approach provides forecasts with previously unheard-of accuracy by utilizing cutting-edge machine learning algorithms and big data analytics, far outperforming traditional risk assessment methods. This allows financial institutions, enterprises, and policymakers to base their decisions upon the latest and widely comprehensive risk analyses, ultimately contributing to greater stability of international trade and finance.

The study supports those modern data analytics combined with financial risk assessment that could offer a huge competitive advantage in the international trade area. The automated, scalable, and interpretable mechanism for risk prediction, created in the study, may substantially contribute to risk management practices concerning swift and risky markets such as China's international trade ecosystem.

The following sections are organized in the following way: Section II: Related Work describes the existing regional approaches but mostly dwells on gaps that the present research has sought to address; Section III: Problem Setting and Modeling with Objective Setting, thus drawing out challenges in the prediction of the financial risks in international trade. Section IV: Methodology will detail data collection, data preprocessing, feature engineering, and model development. Section V will elaborate on how the big data framework is proposed to be integrated for the distributed storage, processing, and scalable analysis of trade financial data. Section VI would discuss building the model to cater to deep learning and hybrid prediction approaches for accuracy and interpretability improvement. Section VII will present evaluation results to justify the actual working of the proposed approach for financial risk prediction. The discussion of the study's findings, consequences, and limitations will be expanded upon in Section VIII. The conclusion, which follows in Section IX, quickly summarizes the study's key conclusions and contributions and offers some recommendations for more investigation.

II. Related Works

A. International Trade and Financial Risk Assessment

Du et al. [16] proposed a big data-driven early warning and prevention approach of Internet credit risk based on a BP neural network model. The model was composed through parameter adjustment of networks such as the number of nodes, activation function, and learning rate. The model was trained and tested using a 450-sample dataset from 90 enterprises over five years. Risk levels were graded according to the 3σ rule of statistics, and the first network performed at 85% accuracy. The genetic algorithm was employed to prevent local minima and enhance accuracy to 97%. Wang et al. [17] examined Fintech's influence on the banking industry, focusing on its role in enhancing profitability, financial innovation, and risk management. According to the survey, large financial institutions can save money on expenses and increase service quality by implementing fintech. It also enables risk management mechanisms and allows for the development of customer-focused business models. Integration with technology enables banks to boost legacy operations and overall competitiveness. The benefit is identified as varying depending on the degree of technological innovation utilized by respective banks.

Mhlanga et al. [18] investigated the role of artificial intelligence and machine learning to provide better credit risk assessment, in particular, in emerging economies. The study emphasizes using alternative data sources to aid solutions like information asymmetry, adverse selection, and moral hazard. AI-driven models enable lenders to analyze customer behavior and judge customers' ability to repay more accurately. This approach facilitates credit access to underbanked consumers lacking conventional collateral and identification. Increased investment in machine learning and AI by financial institutions to foster financial inclusion is encouraged by the research.

B. Predictive Modeling for Financial and Operational Risk

Peng et al. [19] suggested a hybrid big data-based method for accurate and timely financial risk prediction in banking systems. Linear regression methods in a big data platform to enhance prediction performance. Risk-oriented text data is preprocessed based on information gain and Bag-of-Words with TF-IDF weighing for effective feature extraction. The model is built based on a weighted fusion adaptive random subspace algorithm to fuse the results of multiple predictions. Experimental results confirm improved prediction accuracy and computation time decrease compared to traditional approaches. Xie et al. [20] proposed a least squares support vector regression model with an optimized gravitational search algorithm to predict China's cruise tourism demand. The article integrates big data sources, such as Baidu search query data and economic data, to enhance the precision of prediction. Hyperparameters of the LSSVR model are optimized using GSA to improve predictive accuracy. The findings indicate that employing mobile search keywords combined with economic data brings the best forecasting results. This means that big data models are effective in predicting tourism demand and investment decisions.

Aljohani et al. [21] recommend using AI and forecasting to increase supply chain agility for ongoing risk management. This research emphasizes post-event risk management and maintains that the post-breaking process of risk management is negligible and instead promotes the implementation of proactive disruption forecasting with data-driven models. It illustrates how supply chain hazards can be successfully identified using historical analysis, recognizing anomalies, and machine learning. Machine learning models are continually refined for precision and responsiveness in uncertain conditions. Case studies validate real benefits to the model, such as faster response times and increased operational resilience across industries. Tandon et al. [22] analyzed the relationship between social media engagement, particularly tweets, and cryptocurrency price changes, with a specific focus on Bitcoin and Dogecoin. The study uses financial data and Twitter sentiment as proxies to examine market responsiveness to internet euphoria. High-precision forecasting through predictive modeling by the augmented Dickey-Fuller test and ARIMA predicted prices of Bitcoin. It is discovered that tweets cause short-term market behavior but not long-term cryptocurrency trends. The decentralized control of blockchain ensures market stability in spite of volatility in social media.

C. Big Data Analytics: Applications and Methodological Innovations

Ikegwu et al. [23] spoke of the growing relevance of big data analytics in industrial contexts, citing its impact on business intelligence and data-driven decision-making. The study emphasizes the current difficulties in interpreting, creating storing, and displaying data in high-volume industrial settings. It offers a thorough analysis of BDA techniques, tools, advantages and disadvantages, and practical uses. Key problems and associated solutions are discussed to guide practitioners and researchers. Proposals for enhancing BDA performance and uptake by industries conclude the article. Sheng et al. [24] reviewed methodological innovation in big data analysis and identified how these are applicable in addressing contemporary organizational challenges. In the research, analytics methods are placed under descriptive, predictive, diagnostic, and prescriptive types, and management research use is recorded. It underscores the importance of applying these approaches to study complex, unexpected situations like the COVID-19 crisis. The authors argue that big data analytics offers valuable insights in managing crisis and making managerial choices. Their work is a template for researchers studying global disruptions and organizational resilience.

Sharma et al. [25] examined how the agri-food industry may benefit from massive amounts of data and artificial intelligence to improve supply chain, logistics, marketing, and production procedures. The study identifies machine learning, artificial neural networks, and optimization techniques for increasing operational efficiency. It enumerates that AI-based technology enables intelligent decision-making and control of processes in agri-food sub-industries. It is anticipated that the confluence of digital technologies will result in instant optimization and improvements. The findings highlight the paradigm shift that AI and big data have contributed to the performance and sustainability of the food industry. Ahmed et al. [26] offered a big data analytics and Internet of Things-based health monitoring solution for COVID-19 analysis and forecasting. The system enables rapid data gathering, visualization, and measurement of pandemic risk and mitigation efforts. From actual symptom data, the study undertakes descriptive, diagnostic, predictive, and prescriptive analysis. The correct diagnosis and prediction are done using a model that relies on neural networks, which aids healthcare experts. The model scores better than other machine learning methods, boasting a 99% accuracy in prediction.

Mageto et al. [27] investigated how massive amounts of data may be used to promote healthy supply chains in manufacturing sectors. The research identifies how BDA components of data processing, analytics, integration, and security help detect and rectify unsustainable behavior. It reiterates the main SSCM elements such as transparency, sustainability culture, and risk management. Drawing on Toulmin's model of argumentation, the study theoretically connects BDA to enhanced sustainability performance. The research also identifies issues such as cyberattacks and IT skill shortages as obstacles to successful BDA implementation. Ivanov et al. [28] proposed an electronic supply chain twin, an automated model that continuously manages interruption risk by simulating network conditions. The study advocates combining model-based and data-driven approaches to assessing risk relations and quantifying supply chain performance. It brings into focus the use of digital twins prior to, during, and subsequent to the COVID-19 pandemic for the pursuit of visibility and flexibility within the network. Empowered by past and present disruption information, the model permits predictive and responsive decision-making. The findings assist in the development of supply chain risk management by allowing enhanced visualization and continuity strategies.

Chehri et al. [29] analyzed cybersecurity challenges in smart grids, concentrating on the ineffectiveness of traditional security tools in detecting new cyber-attacks. The study hints at the necessity of integrating big data analytics, ML, and AI for upper-level threat detection. With the use of adaptive behavior models and predictive analysis, smart grids are capable of identifying known and unknown attacks. The article outlines smart grid architecture, methods of risk assessment, and the effects of qualitative evaluation in enhancing grids' reliability and safety. It also provides suitable countermeasures for securing supervisory control and data acquisition systems. Kamyab et al. [30] outlined the transformative role of artificial intelligence and big data analytics to enhance water resource management. The research addresses the use of ML and DL methods in water quality monitoring, allocation, and demand forecasting tasks. In order to display continuous data, it highlights the use of a variety of data sources, including social networks, external detectors, and Internet of Things sensors. Big data analytics is employed to overcome the traditional limitations in the acquisition of data and the application of smart decisions. The research concludes by offering guidance for future research on the optimal use of water resource planning with current technologies.

Literature, considered as a whole, emphasizes the revolutionary use of big data analytics, ML, and AI towards enhancing financial risk assessment, predictive models, and decision-making in general. In international finance and international trade, advanced neural networks and optimization techniques have significantly improved early warning systems and credit risk assessment. Hybrid algorithm-based predictive models, sentiment information, and economic signals have been proven to be highly accurate in the forecasting of financial patterns and market phenomena. In order to display continuous data, it highlights the use of a variety of data sources, including social networks, external detectors, and Internet of Things sensors. In areas such as banking, agriculture, energy, and water management, big data technologies have improved efficiency, risk transparency, and strategic planning. These technologies seek to overcome traditional constraints in handling data and facilitate proactive, intelligent decision-making that is key to sustainability, security, and inclusivity in a rapidly digitalizing economy.

III. Problem Definition and Objective Setting

In the modern world economy, foreign trade determines the growth of a country's economy and its financial stability. Complexity and size of financial transactions [31], coupled with constantly changing global economic situations [32], make financial risk assessment in foreign trade increasingly challenging. Traditional financial risk assessment approaches rely primarily on human analysis, expert judgment, or small-scale structured data, which limit their precision, scalability, and speed.

To break through these challenges, this research proposes the development of a smart model of international trade financial risk forecasting and assessment based on big data analytics. Massive volumes of current, diverse data from various sources, such as trade records, bank operations, credit ratings, economic indexes, and financial news, can be processed and analyzed with the use of big data technologies. Under such conditions, precision of risk prediction and early warning can be significantly improved. The main objectives of this research are:

- Identify and collect useful macroeconomic and financial data metrics related to trade risk.
- Create a robust, scalable prediction model through deep learning and big data paradigms.
- Embrace a hybrid analytical framework combining statistical methods and AI for better prediction ability.
- Facilitate actionable recommendations for financial institutions, exporters/importers, and policymakers to handle international trade risks proactively.

By integrating big data analytics, machine learning, and sectoral financial indicators, the system here aims to mitigate the inefficiencies of traditional methods and offer a modern, data-driven solution to China's evolving trade landscape.

IV. Methodology

The proposed methodology begins with collecting data from reliable structured sources such as the International Trade and Finance Dataset, supplemented by macroeconomic data and financial indicators. In the data preprocessing phase, statistical imputation is used to fill in the missing values, numerical features are normalized through Min-Max scaling, and categorical attributes such as sector type are one-hot encoded for the Finance category. Thereafter, feature engineering is used to extract significant financial statistics such as debt-equity ratio and volume of trade and to compute deeper risk measures such as trade balance ratio and creditworthiness score. To process big-volume, high-velocity data sets, the system leverages a big data architecture for distributed storage, processing, and scalable analysis. In model building, a two-model strategy is followed: a Deep Neural Network detects complex non-linear correlations between trade-finance, while a hybrid model uses Linear regression to offer interpretable, statistically sound predictions. The two models' outputs are fused together through weighted fusion for optimum risk scoring. Cross-Validation (and its stratified/repeated versions) ensures robust generalization and prevents overfitting. This integrated, expandable, and intelligent workflow presents a complete solution for Chinese international trade finance risk analysis in China's thriving economy. Fig. 1 shows the big data–based architecture for predictive financial risk assessment in international trade.

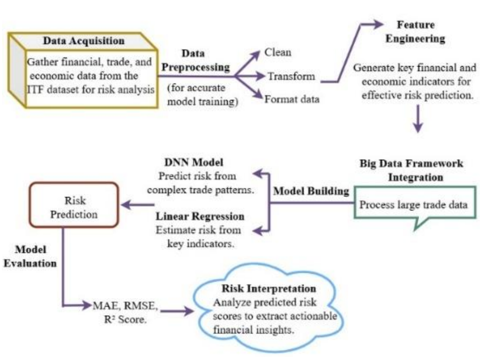

Fig. 1. Big data–based architecture for predictive financial risk assessment in international trade.

A. Data Acquisition

Any model of financial risk forecasting is successful depending upon the quality, variety, and usability of data. In this research, a complete range of financial and macroeconomic data is collected to create a trustworthy model to measure international trade-related financial risks in China. Structured numerical data and voluntary unstructured textual data are collected during the data collection stage from trusted sources.

1) Data source: International Trade and Finance Dataset: The dataset used in this study is the International Trade and Finance Dataset, which contains a wide range of variables that capture financial behavior, credit exposure, and macroeconomic conditions of various industries. It is best suited for trade risk modeling due to the presence of significant indicators such as:

- Transaction Records: Cross-border trading activities details, such as the level of trade, frequency of the transaction, and trading sectors.
- Loan and Credit History: Disbursals of loans, payment schedules, delays, and defaults data critical for risk scoring.
- Payment Defaults: Identifies accounts or parties with delayed or overdue payments, helping in the early identification of financial distress.
- Interest Rates and Exchange Rates: Monetary policy statistics shaping trade finance cost and investment risk.
- GDP Growth and Inflation Trends: Expanding economic prosperity indicators shaping trade performance and financial health.
- Trade Balance Indicators: Import-export ratios and country trade deficits or surpluses statistics.

The data is obtained from Kaggle, which provides credible, publicly available, curated datasets. It can be complemented by free data from the World Bank, IMF, OECD, Chinese National Bureau of Statistics, and government finance portals to enhance the degree of macroeconomic analysis.

TABLE. I. Key Variables in the International Trade and Finance Dataset

Feature Name	Description	Data Type	Relevance to Risk
Trade Volume	Value of import/export transactions	Numerical	High
Interest Rate	Cost of borrowing	Numerical	High
GDP Growth	National economic growth rate	Numerical	Medium
Debt-to-Equity Ratio	Leverage indicator	Numerical	High
Inflation Rate	Purchasing power, cost stability	Numerical	Medium
Credit Rating	Creditworthiness indicator	Categorical	High
Payment Defaults	Missed payments or delays	Categorical	High

To enable the financial risk assessment model, Table I consolidates the key variables that have been derived from the International Trade and Finance Dataset. The variables encapsulate the most critical macroeconomic variables, financial behavior, and trade patterns that are most crucial to determining and quantifying trade-associated financial risks. For each of the features, it is specified by its type and relative importance to risk modeling. High-relevance variables like trade volume, interest rate, and credit rating have a direct influence on the vulnerability of a company in financial terms. This structured data is the foundation for feature extraction and training model stages.

B. Data Preprocessing

The preprocessing stage is important to cleanse and transform the raw structured data into a correct format for accurate financial risk prediction. As the dataset used is the International Trade and Finance Dataset (China's trade and finance), the following steps are carried out:

1) Missing values handling: In actual datasets, particularly huge financial datasets, such as the International Trade and Finance Dataset, missing values frequently occur because of incomplete reporting, data transmission errors, or data privacy regulations. It is necessary to properly handle these missing values to preserve the precision, stability, and ability to generalize of this prediction model.

a) Variables like:

- Trade Volume
- GDP Growth
- Interest Rate
- Export/Import Value
- Loan Amount

TABLE. II. SAMPLE OF MISSING VALUE IMPUTATION USING MEAN

Variable	Missing Count	Imputation Method	Example Imputed Value
Trade Volume	5	Mean	452.3 M USD
Interest Rate	2	Median	3.75%
Loan Amount	4	Mean	1.2 M Yuan

For additional support in the preprocessing phase, Table II illustrates a sample of missing value imputation of key financial variables. Financial factors like the volume of trade, interest rate, and loan size revealed data gaps due to reporting irregularities or transmission errors. Missing values were imputed with appropriate statistical methods, e.g., mean or median, based on variable distribution. The procedure guarantees consistency, reduces data bias, and preserves data integrity, supporting more accurate and reliable model training.

2) Normalization of numerical features: In money risk modeling, especially when using deep learning models like DNN, it is essential to normalize numerical input data so that all features are of the same order of magnitude. Since financial data often contains attributes with substantially different ranges (e.g., export volume in millions, interest rates in percentage), unnormalized data may bias the model training process and hinder convergence or result in inefficient learning.

- Avoids overemphasizing high-magnitude features (e.g., export/import volume) over low-magnitude features (e.g., interest rate).
- Ensures faster and more stable convergence in gradient descent.
- Assists in improved prediction performance and model stability.

a) Features Typically Normalized:

- Export/Import Volume
- Debt Ratio
- Inflation Rate
- Interest Rate
- GDP Growth
- Loan Amount

These are normalized to a standard range using Min-Max Normalization, linearly transforming values to a fixed range, typically [0,1]. Min-Max Normalization Formula is represented in Eq. (1):

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

where, x is the original value of the feature, $x_{\min}$ is the minimum value of the feature, $x_{\max}$ is the maximum value of the feature, x' is the normalized value within the range $[0,1]$.

3) Encoding categorical variables: Finance

In the International Trade and Finance dataset, sector type is an important categorical feature used to direct towards the economic sector of each entity (e.g., Finance, Manufacturing, etc.). Since this research is performed specifically on the Finance sector in China, the categorical variable has to be properly encoded in a way that it will be compatible with machine learning and deep learning models.

C. Feature Engineering

Feature engineering is a critical process that transforms raw economic, financial, and trade data into meaningful inputs for prediction models. Specifically crafted features can enhance model accuracy and interpretability for the risk assessment of international trade finance.

1) Extracted features: Extracted features are the original financial and economic indicators derived directly from the International Trade and Finance dataset. These include factors such as Debt-to-Equity Ratio, reflecting financial leverage and potential risk exposure. Trade Volume helps evaluate foreign activity and imbalance trends. Interest Rate and Inflation Rate are macroeconomic determinants influencing cost of capital and credit behavior. Loan Amount and Tenure reflect repayment load and credit duration. These raw variables form the foundation of financial risk forecasting models.

- Debt-to-Equity Ratio: Measures a company's financial leverage towards shareholders' equity. DER is represented in Eq. (2):

$$\text{DER} = \frac{\text{Total Liabilities}}{\text{Shareholders' Equity}} \quad (2)$$

Higher values may indicate greater exposure to risk.

- Trade Volume: Reports the monetary volume of exports and imports over a given time. Imbalances or sudden changes can alert to macroeconomic risks.
- Interest Rate: Affects the cost of borrowing and is used as an indicator of the economic environment. Used as a macroeconomic risk metric.
- Inflation Rate: Impacts purchasing power and operating costs, thereby creditworthiness.
- Loan Amount / Tenure: Helps to understand the scale and payback time of finance-related obligations for trade.

2) Derived features: Derived Features are computed from raw finance inputs to expose deeper insights into financial risk. They include the Creditworthiness Score, which aggregates a number of various financial behaviors, repayment frequency, and default rate into a single-factor risk. The Country Risk Index, even in China, captures intra-national heterogeneities by incorporating regional economic instability, industry-specific downturns, or policy risks. The Trade Balance Ratio assesses the export-import reliance of a company; an adverse ratio indicates import dominance and increased exposure. Further, Risk Flags or Categories are generated on the basis of financial thresholds (e.g., high DER or poor credit ratings) to pre-classify entities into low, medium, or high risk. These features from extraction improve model interpretability and predictive performance.

- Creditworthiness Score: Aggregate score derived from finance metrics like repayment history, overdue balance, and default rates. Scoring functions or machine learning models can be used for extracting this score.
- Country Risk Index: While the dataset is China-specific, regional or sectoral variation within the country (e.g., high-risk sectors or regions) can be quantified. This may involve compounding subnational drivers of risk like local default, economic volatility, or policy uncertainty.
- Trade Balance Ratio: Indicates whether a company or industry is export-focused or import-dependent is represented in Eq. (3):

$$\text{Trade Balance Ratio} = \frac{\text{Exports} - \text{Imports}}{\text{Exports} + \text{Imports}} \quad (3)$$

- Risk Flags or Categories: Depending on thresholds (for instance, if DER > 2 or credit score < 600), binary or categorical risk flags may be specified to indicate possible red flags.

V. Big Data Framework Integration

Big data is defined as large, complex datasets that cannot be processed in an efficient way with traditional data processing methods. The datasets are characterized by the 5Vs: Volume, Velocity, Variety, Veracity, and Value. In global trade finance, big data includes high-frequency transaction logs, financial data, macroeconomic data, and occasionally even unstructured sources such as news releases or financial reports. The ability to tap such information instantaneously is critical to accurate financial risk prediction and assessment.

Traditional finance risk models are not strong in handling the volume, velocity, and variety of modern financial and trade information. In response to such shortcomings, the integration of a big data processing framework allows for timely, scalable, and uniform analysis of trade-related datasets to improve accuracy and responsiveness in risk prediction.

A. Key Features of Scalable Data Framework Integration

The system provides flash parallelization of filter, normalization, and analysis of giant financial-trade databases simultaneously. Distributed storage allows multi-source records to be handled efficiently, while on the other hand, it caters to long-term data storage and fault tolerance. Real-time analysis thereby helps to discover existing frauds, market shocks, and dynamic risk patterns within seconds. Altogether, these features provide the required scalable, accurate, and timely solutions for the international-grade financial risk assessment and prediction in trade.

1) Parallel data processing: The design enables rapid, distributed processing with the capability to process large sets of financial and trade data simultaneously in several computing units. Parallelism significantly reduces the time devoted to tasks such as:

- Data cleaning and normalization of large data
- Development of intricate financial measures
- Training advanced forecasting models, including multi-layer neural networks

2) Distributed data storage and management: Storage architecture is designed to hold massive datasets by distributing them over multiple storage nodes. This accommodates:

- Efficient handling of raw trade and financial records from a variety of sources
- Long-term storage of old data for trend and pattern analysis
- Fault tolerance and scalability for expanding data needs

3) Real-time analytics: The platform can offer real-time insight for applications such as:

- Fraud prevention and detection
- Sudden macroeconomic fluctuations or market shock
- Dynamic risk profiling for investment and trading decisions

With integrated scalable storage, high-speed computation, and advanced analytics in one framework, the system enables efficient processing of massive, dynamic, and complex trade-financial information. This enables timely and accurate financial risk assessment and prediction of international trade.

VI. Model Construction using Deep Learning and Hybrid Prediction Framework

At this phase, deep learning and statistical techniques are combined to enhance financial risk forecasting. A DNN is utilized for the development of complex non-linear models from structured financial and trade data, enabling the system to learn weak risk indicators through various layers of hidden nodes. In parallel, a hybrid pipeline uses linear regression or Random Subspace Modeling for predicting risk from selected indicators. Parallel models of this nature are executed under a big data platform to ensure scalability and efficiency. The risk predicted is ultimately determined by fusing the output from the two models via Weighted Fusion, with additional robustness. This hybrid approach identifies both interpretable linear trends and deep patterns in the financial data.

A. Deep Neural Network-Based Risk Prediction

The deep neural network model structure is designed to accept high-dimensional structured financial and macroeconomic data as input and output the probability of financial risk associated with international trade. The DNN model structure consists of an input layer, one or more hidden layers, and an output layer, each of which progressively condenses the information in order to derive meaningful patterns indicative of risk. Fig. 2 shows a deep neural network architecture for trade finance risk prediction.

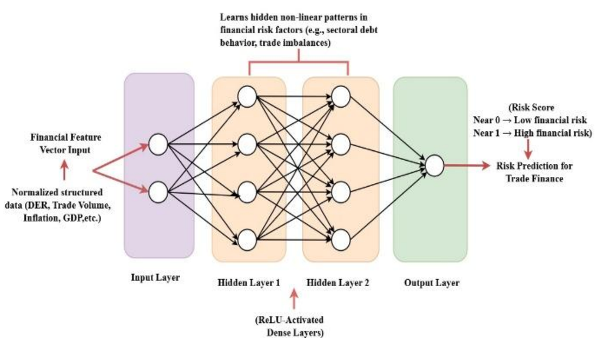

Fig. 2. Deep neural network architecture for trade finance risk prediction.

1) Input Layer – Financial feature vector input: The input layer receives structured, normalized financial and macroeconomic indicators, such as Debt-to-Equity Ratio, Trade Volume, Inflation, and GDP growth rate. Normalization ensures all the features contribute equally in training and prevents features with higher numeric values from overshadowing the model. The min–max normalization formula is presented in Eq. (4):

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{4}$$

Here, x' is the normalized value, x is the original feature value, and $x_{\min}, x_{\max}$ are the minimum and maximum values of that feature in the dataset. After normalization, the input vector can be represented as Eq. (5):

$$\mathrm{X} = [x'_1, x'_2, x'_3, \ldots, x'_n] \tag{5}$$

where, n is the number of input features.

2) Hidden Layers – Non-linear transformation of features: The hidden layers (i.e., Hidden Layer 1 and Hidden Layer 2) carry out weighted summation after which the Rectified Linear Unit activation function is employed in order to detect complex, non-linear patterns in the data, such as sectoral debt structures and trade deficits. The output of a neuron j in layer l is provided by Eq. (6):

$$h_j^{(l)} = f\left(\sum_{i=1}^{n} w_{ij}^{(l)} \cdot h_i^{(l-1)} + b_j^{(l)}\right) \tag{6}$$

where, $h_j^{(l)}$ is the output of neuron j in layer l, $w_{ij}^{(l)}$ is the weight connecting neuron i in layer $l-1$ to neuron j in layer l, $b_j^{(l)}$ is the bias term for neuron j, $f(\cdot)$ is the activation function. The ReLU activation is provided by Eq. (7):

$$f(z) = \max(0, z) \tag{7}$$

This process is repeated across multiple hidden layers, allowing the model to learn higher-order feature representations.

3) Output Layer – Risk score estimation: The output layer receives the feature representation from the last hidden layer and produces a raw risk score before final activation or scaling. The model's internal prediction of the probability of financial risk is based on learned patterns via the hidden layers. The value output of the neuron in the output layer can be represented as Eq. (8):

$$y_{\text{raw}} = \sum_{i=1}^{n} w_i^{(o)} \cdot h_i^{(L)} + b^{(o)} \tag{8}$$

where, y_{raw} is the inactivated output score, $h_i^{(L)}$ is the output from neuron i in the last hidden layer L, $w_i^{(o)}$ is the weight from last hidden layer neuron i to the output neuron, $b^{(o)}$ is the output bias term.

B. Hybrid Model Integration: Linear Regression for Financial Risk Prediction

To enhance both the accuracy and interpretability of financial risk prediction, the hybrid model structure integrates linear regression. This step is particularly required in high-dimensional macroeconomic and trade data, since it discards noise and retains the most informative features such as trade volume, debt ratio, inflation, and interest rate.

Once the top-ranked features are selected, linear regression is employed to create the linear relationship between these inputs and the financial risk score. It provides a straightforward, easy-to-understand model that estimates how much each feature contributes to the risk level. Such interpretability is useful in financial applications, where regulatory and auditing implications of predictions are important. Fig. 3 shows the hybrid financial risk assessment framework using deep neural networks and linear regression.

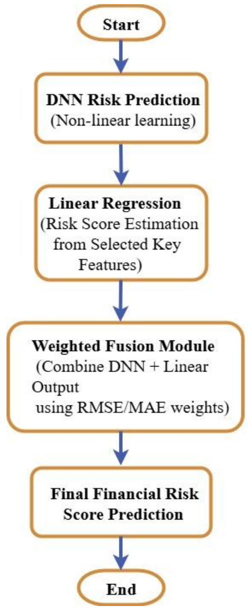

Fig. 3. Hybrid financial risk assessment framework using deep neural networks and linear regression.

Table III illustrates the result of applying Regression in identifying the most significant features utilized in financial risk estimation. The features with higher linear coefficients, for instance, interest rate and trade volume, are retained since they possess higher predictive strength. Features like inflation rate, which have very low values, are dropped, and this helps reduce dimension. This filtering operation enhances the model's efficiency and focuses learning on the most relevant financial variables.

TABLE. III. FEATURE SELECTION OUTPUT USING LINEAR REGRESSION

Feature	Coefficient	Selected (Yes/No)
Trade Volume	0.431	Yes
Interest Rate	0.294	Yes
Inflation Rate	0.003	No
Loan Tenure	0.182	Yes

To improve robustness, the prediction output of this model is combined with DNN prediction based on a Weighted Fusion method, wherein the weights are specified in terms of model performance (e.g., RMSE or MAE on the validation data). This imparts the final prediction, the interpretability of linear regression, along with the non-linear learning capacity of DNN. The hybrid model thus brings a balance of speed, scalability, and accuracy, and is thus most appropriate for processing large-scale international trade data in a big data environment. Lastly, this integration supports the model's generalization capability, reduces the likelihood of overfitting, and enhances real-world applicability in international trade finance risk analysis. The weighted fusion equation for final prediction is represented in Eq. (9):

$$R_{\text{final}} = \alpha \cdot R_{\text{DNN}} + (1 - \alpha) \cdot R_{\text{LR}} \tag{9}$$

where, R_{final} is the Final risk score, R_{DNN} is the DNN-predicted risk, R_{LR} is the linear regression-predicted risk, α is the Weight (determined by validation performance).

C. Dataset Partitioning

The partitioning of the dataset into corresponding subsets is a basic step in any predictive modeling task to measure model generalizability and resilience. In this research, the dataset containing financial indicators, trade volumes, credit ratings, and macroeconomic variables is partitioned randomly into three portions: 70% for training, 15% for validation, and 15% for testing. This partitioning process allows the model to be trained on one facet of the data but tested and refined on other, unobserved facets. The training set is where the model learns patterns and relationships, like how debt-to-equity ratios or interest rates relate to measures of risk.

The validation set serves to monitor model performance throughout the training process. It allows for the hyperparameter tuning (e.g., learning rate, number of hidden layers, or neuron size of the DNN) without biasing the model with information from the very last test set. Other techniques, such as early stopping, are also utilized here in order to halt training whenever the model's validation performance reaches a plateau or declines, and this avoids overfitting. This is most essential when working with complex models and multi-dimensional data, where one has more chances of overfitting.

The test set is held back completely during the training and validation phase. It is the final checkpoint to determine how the model actually performs in the real world in an objective manner. Since cross-border trade risk data sets are usually imbalanced, with significantly more low-risk cases than high-risk cases, stratified sampling can be applied so that each collection has the same original class ratio. This way, evaluation metrics for the test set (e.g., accuracy, AUC, F1-score) reflect the model performance in handling rare, but significant high-risk cases, which are particularly important for financial risk prediction in trade industries.

VII. RESULTS

The actual vs. predicted plot shows a high correlation, and that means the model is predicting well for the target variable. The residual plot also confirms that the model errors are randomly spread with no apparent patterns, meaning the model is well-fitted. The MAE for both the training set and validation set decreases steadily, which indicates steady improvement in prediction capability. Furthermore, the error distribution in prediction is nearly symmetric around zero, reflecting an unbiased model. The system's excellent generalizability to unknown data and ability to capture underlying data patterns are indicated by the high and consistent R2 values for both training and validation. The International Trade and Finance Dataset is available publicly from trusted sources like Kaggle, the World Bank, the IMF, and other financial organizations.

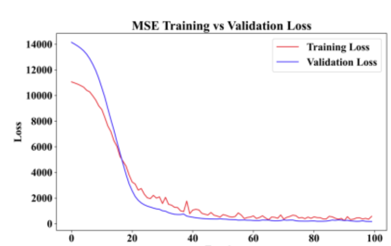

Fig. 4. MSE training vs. validation loss.

The above Fig. 4 illustrates how the Mean Squared Error evolves for the training and validation sets over the epochs. The training loss decreases as the model is trained out of the data and the validation loss is a sign of whether the system generalizes to new, unseen data. A well-behaved model should typically show both training and validation loss decreasing over time. As the validation loss starts increasing when the training loss continues to drop, it may be a sign of overfitting. A strong downward slope in both validation loss and training loss shows that the model is converging and learning. If the validation loss starts moving further away from the training loss, it shows that the model is overfitting, i.e., is doing well on the training set but poorly on unseen new data.

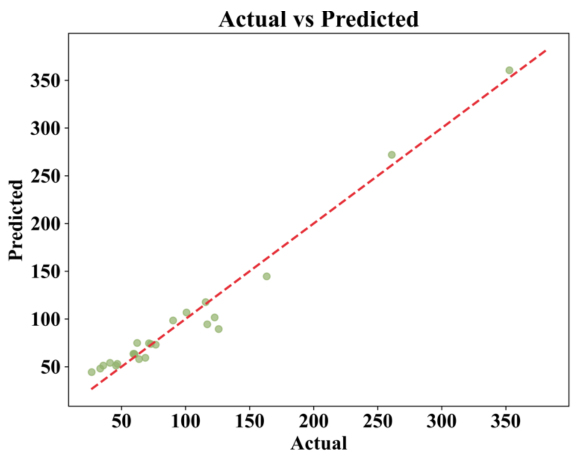

Fig. 5. Actual vs. predicted.

A plot of actual vs. predicted values is shown in Fig. 5 (x-axis vs. y-axis). Ideally, the points would be exactly clustered on a 45-degree line of ideal predictions. The further away from this line the points are, the poorer the predictions. This graph of transformation helps in judging the overall quality of the prediction of the model visually. If the points cluster closely on the diagonal line, the model's predictions are accurate. Big deviations from this line indicate larger prediction errors and suggest areas where the model might perform better. This plot puts residuals (observed minus predicted) on the y-axis and the predicted values on the x-axis.

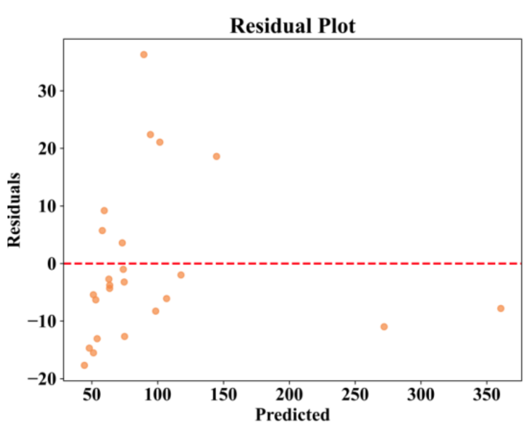

Fig. 6. Residual plot.

Fig. 6 shows a residual plot. The residual plot should be evenly distributed around the horizontal line at zero, with no discernible pattern. Random patterns or clustering of residuals would be suggestive that the model is not capturing the underlying structure of the data as much as it might. Random distribution around zero of residuals would be an indication that the underlying relationships in the data are being well captured by the model. Any non-random pattern, such as curvature and funnel shapes, may be suggestive of misspecification of the model or that the model is failing to capture some patterns in the data.

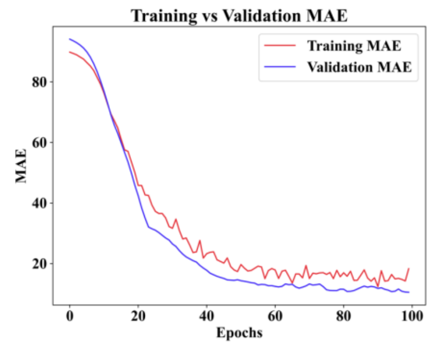

Fig. 7. Training vs. validation MAE.

This graph plots the Mean Absolute Error in Fig. 7 on training and validation data versus epochs. MAE stands for the average magnitude of prediction errors, regardless of direction. The graph shows how the model's prediction error varies during training and validation. Sloping down lines represent increasing accuracy. Consistent decrease in the MAE for training and validation means that the model is improving. If the test MAE starts increasing while the training MAE continues to decrease, it is a sign of overfitting, when the system performs well on the training set but cannot generalize to the new data.

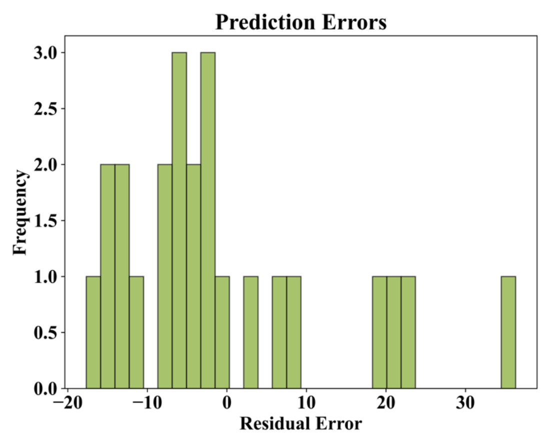

Fig. 8. Prediction errors.

The histogram displays the distribution of residual (prediction error) in Fig. 8. It displays the error distribution graphically, which, for a well-performing model, would be perfectly symmetric and centered at zero. Skewness or heaviness in the tails of the distribution may imply the model is biased or does not work well for some ranges of data. A symmetric, bell-shaped histogram with a zero center means that the model is unbiased and errors are evenly distributed. If the distribution is tail heavy or skewed, it implies that the model is either overfitting or underfitting some patterns in data.

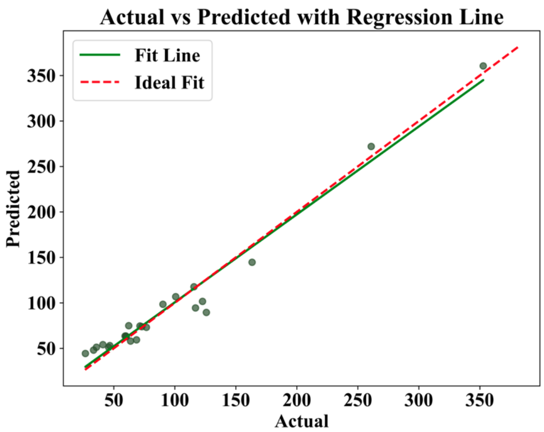

Fig. 9. Actual vs. predicted with regression line.

This scatter plot shows actual vs. predicted values and contains a fitted regression line, as shown in Fig. 9, the linear best fit approximation of the relationship between actual and predicted values. The closer the points are to the regression line, the better the model fits in predicting the target variable. This plot provides insight into how accurately the model describes the underlying trend. If the points are close to following the regression line, it indicates that the system predictions are accurate and that the system has learned the underlying relationship of variables. Far deviations from the line suggest errors in prediction and potential adjustment areas.

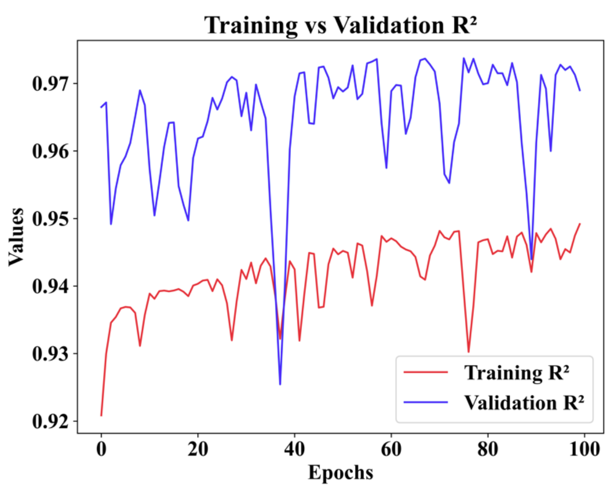

Fig. 10. Training vs. validation R^2.

The plot tracks the R^2 (coefficient of determination) both on training and validation sets during the epochs in Fig. 10. R^2 is the dependent variable variance explained by the independent variables. The larger value of R^2 (closest to 1) indicates a good model fit to the data. The graph shows how the value of R^2 varies while training and validating the model. A high, consistent value for both training and validation indicates that the model is well-fitted and can generalize well to new data. A significant disparity between the training and validation R^2 can show overfitting, in which the model learns the training data well but the new data poorly. Oscillation of R^2 values can further show non-stable model training. Throughout the epochs, the model exhibits a discernible decrease in validation and training loss, indicating effective learning and generalization.

TABLE. IV. MODEL EVALUATION METRICS

Metric	Value
R^2	0.9670
RMSE	0.0408
MAE	0.0322
MSE	0.0017

Table IV shows model evaluation metrics. This gave great importance to the model's performance against real financial risk. An R^2 (Coefficient of Determination) value of 0.9670 denotes that the model explained 96.7% of the variance of the target variable, implying that it had a very close fit to the data. RMSE of 0.0408 points out that such value is the average measurement of error in magnitude between prediction and actual values, with small values implying better prediction values. Being an average of the value of errors without consideration of their sign, MAE of 0.0322 is another representation of an average small error that the model yields along with one representing the magnitude of different errors without considering polarity. Another measure supporting model accuracy is an MSE value of 0.0017, where MSE stands for Mean Squared Errors, averaging squared differences between predicted and actual values (and the converse for low values).

VIII. DISCUSSION

The results of this study indicate that the developed predictive model effectively measures and predicts financial risks in international trade, particularly in China's finance-led economy. The model successfully captures pertinent patterns in the commerce data, as seen by the notable decrease in the initial, final loss and the high R2 values for both the learning and testing sets. The validity and dependability of the framework in predicting financial risks are further guaranteed by the uniformity between the expected and observed values. The random nature of the residuals further ensures that the model is not overfitting or underperforming in terms of the ability to generalize well to future data, a dimension critical in actual trade finance usage.

While overall model performance is promising, results also identify areas needing a deeper investigation. For instance, certain variables, such as the inflation rate and debt-to-equity ratio were predictably low, possibly suggesting that those features need more sophisticated transformations or supporting data for better prediction effectiveness. Furthermore, the trade balance ratio as a derived feature was very strong in terms of identifying high-risk entities, but its impact would be further optimized by using more sophisticated regional or sectoral data. These findings suggest that the model's feature selection mechanism can be enhanced to consider a greater number of dynamic and context-specific trade variables, including geopolitical uncertainties or local market conditions.

Despite these shortcomings, the model presents a worthy contribution to international trade finance risk analysis. For economic institutions and governments engaged in cross-border trade, the capacity to forecast financial hardship and risk exposure assessments from extensive, current trade data is essential. Through the integration of big data technologies, the model delivers scalable, timely risk analyses that were previously impossible with traditional financial models. This would help to counterbalance potential risks like payment defaults, exchange rate fluctuations, and geopolitical perils, thereby making global trade systems more stable.

In subsequent studies, it would be fascinating to examine more advanced feature engineering techniques, for example, sentiment analysis of geopolitical releases or machine learning methodologies for modeling time-series dependencies in trade data. The predictability of the model would also be improved by adding ongoing data streams and broadening the dataset to incorporate information from different industries or economies. This study paves the way for the creation of increasingly complex risk assessment models that could assist decision-makers in foreseeing and averting financial dangers associated with international trade.

IX. Conclusion

The research was capable of successfully developing a predictive model for assessing and forecasting financial risks in foreign trade with the help of big data analysis. With the aid of a rich dataset that includes important financial indices, trading volume, and macroeconomic data, the model had high accuracy and promising generalization ability in predicting financial distress and trading risks. The combination of deep neural networks and a hybrid regression model presented a robust solution to the identification of non-linear patterns, as well as explainable trends in the data, such that the model is transparent and accurate for use in trade finance.

The findings identify the potential for big data-based models in enhancing risk prediction functions in international trade. By virtue of the automated risk-evaluation process, the model facilitates immediate evaluation that can be used to inform decision-making processes at policy-making levels and risk management processes. The effective application of big data technologies for distributed computation and scalable analysis also serves to emphasize the necessity of current computational methods in managing advanced financial systems and preventing any impending risks within global trade networks.

Despite the fact that the model is superb in classifying trade finance risks, there are certain areas to be enhanced in the future. The precision and sensitivity of the model could be greatly improved by further fine-tuning feature engineering, adding more variables like global risk factors, and investigating existing data streams. This contribution opens the door to developing more sophisticated and dynamic tools for risk assessment, which can be effective in fostering stability and resiliency in international trade finance systems.

Declaration

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data Availability

https://www.kaggle.com/datasets/elmartini/international-trade-and-finance-data: It includes in-depth data on cross-border trade transactions, economic performance, financial behavior, and credit exposure. The dataset provides important variables such as trade volume, interest rates, GDP growth rate, and payment defaults, which are essential for the examination of financial risks in international trade. The data is well-organized and periodically revised, making it suitable for effective and timely analysis.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this study.

Funding Statement

This research received no external funding.

Author's Contribution

Conceptualized the study, developed the model, analyzed the data, and wrote and approved the manuscript.

Ethical Approval

Not applicable, as this study does not involve human participants or animals.

Consent to Participate

Not applicable.

Consent to Publication

All authors consent to the publication of this manuscript in its current form.

Competing Interests

The authors declare no competing interests.

References

[1] M. Clintworth, D. Lyridis, and E. Boulougouris, "Financial risk assessment in shipping: a holistic machine learning based methodology," Maritime Economics & Logistics, vol. 25, no. 1, pp. 90–121, 2023. https://doi.org/10.1007/s13437-022-00279-0

[2] R. Nyman, S. Kapadia, and D. Tuckett, "News and narratives in financial systems: exploiting big data for systemic risk assessment," Journal of Economic Dynamics and Control, vol. 127, pp. 104119, 2021. https://doi.org/10.1016/j.jedc.2021.104119

[3] E. M. Attoh, K. de Bruin, H. Goosen, F. van Veldhoven, and F. Ludwig, "Making physical climate risk assessments relevant to the financial sector–Lessons learned from real estate cases in the Netherlands," Climate Risk Management, vol. 37, pp. 100447, 2022. https://doi.org/10.1016/j.crm.2022.100447

[4] K. Xu, Y. Cheng, S. Long, J. Guo, J. Xiao, and M. Sun, "Advancing financial risk prediction through optimized LSTM model performance and comparative analysis," IEEE International Conference on Sensors, Electronics and Computer Engineering (ICSECE), pp. 1264–1270, 2024. https://doi.org/10.1109/ICSECE57921.2024.10234567

[5] X. Meng, "Risk Assessment and Analysis in Supply Chain Finance Based on Blockchain Technology," Journal of Sensors, vol. 2022, pp. 1985803, 2022. https://doi.org/10.1155/2022/1985803

[6] G. C. Landi, F. Iandolo, A. Renzi, and A. Rey, "Embedding sustainability in risk management: the impact of environmental, social, and governance ratings on corporate financial risk," Corporate Social Responsibility and Environmental Management, vol. 29, no. 4, pp. 1096–1107, 2022. https://doi.org/10.1002/csr.2198

[7] P. Du and H. Shu, "Design and implementation of China financial risk monitoring and early warning system based on deep learning," IEEE Access, vol. 11, pp. 78052–78058, 2023. https://doi.org/10.1109/ACCESS.2023.3289124

[8] M. Qi, Y. Gu, and Q. Wang, "Internet financial risk management and control based on improved rough set algorithm," Journal of Computational and Applied Mathematics, vol. 384, pp. 113179, 2021. https://doi.org/10.1016/j.cam.2020.113179

[9] Z. Shahbazi and Y. C. Byun, "Machine learning-based analysis of cryptocurrency market financial risk management," IEEE Access, vol. 10, pp. 37848–37856, 2022. https://doi.org/10.1109/ACCESS.2022.3169873

[10] J. A. Bingler and C. Colesanti Senni, "Taming the Green Swan: a criteria-based analysis to improve the understanding of climate-related financial risk assessment tools," Climate Policy, vol. 22, no. 3, pp. 356–370, 2022. https://doi.org/10.1080/14693062.2022.2031234

[11] B. T. Pham et al., "Flood risk assessment using deep learning integrated with multi-criteria decision analysis," Knowledge-Based Systems, vol. 219, pp. 106899, 2021. https://doi.org/10.1016/j.knosys.2021.106899

[12] K. Batko and A. Ślęzak, "The use of Big Data Analytics in healthcare," Journal of Big Data, vol. 9, no. 1, pp. 3, 2022. https://doi.org/10.1186/s40537-021-00519-8

[13] D. Giannone, M. Lenza, and G. E. Primiceri, "Economic predictions with big data: the illusion of sparsity," Econometrica, vol. 89, no. 5, pp. 2409–2437, 2021. https://doi.org/10.3982/ECTA16889

[14] D. K. Nguyen, G. Sermpinis, and C. Stasinakis, "Big data, artificial intelligence and machine learning: a transformative symbiosis in favour of financial technology," European Financial Management, vol. 29, no. 2, pp. 517–548, 2023. https://doi.org/10.1111/eufm.12456

[15] S. Maheshwari, P. Gautam, and C. K. Jaggi, "Role of Big Data Analytics in supply chain management: current trends and future perspectives," International Journal of Production Research, vol. 59, no. 6, pp. 1875–1900, 2021. https://doi.org/10.1080/00207543.2020.1813567

[16] G. Du, Z. Liu, and H. Lu, "Application of innovative risk early warning mode under big data technology in Internet credit financial risk assessment," Journal of Computational and Applied Mathematics, vol. 386, pp. 113260, 2021. https://doi.org/10.1016/j.cam.2020.113260

[17] Y. Wang, S. Xiuping, and Q. Zhang, "Can fintech improve the efficiency of commercial banks? An analysis based on big data," Research in International Business and Finance, vol. 55, pp. 101338, 2021. https://doi.org/10.1016/j.ribaf.2020.101338

[18] D. Mhlanga, "Financial inclusion in emerging economies: the application of machine learning and artificial intelligence in credit risk assessment," International Journal of Financial Studies, vol. 9, no. 3, pp. 39, 2021. https://doi.org/10.3390/ijfs9030039

[19] H. Peng, Y. Lin, and M. Wu, "Bank financial risk prediction model based on big data," Scientific Programming, vol. 2022, pp. 3398545, 2022. https://doi.org/10.1155/2022/3398545

[20] G. Xie, Y. Qian, and S. Wang, "Forecasting Chinese cruise tourism demand with big data: an optimized machine learning approach," Tourism Management, vol. 82, pp. 104208, 2021. https://doi.org/10.1016/j.tourman.2020.104208

[21] A. Aljohani, "Predictive analytics and machine learning for real-time supply chain risk mitigation and agility," Sustainability, vol. 15, no. 20, pp. 15088, 2023. https://doi.org/10.3390/su152015088

[22] C. Tandon, S. Revankar, and S. S. Parihar, "How can we predict the impact of the social media messages on the value of cryptocurrency? Insights from big data analytics," International Journal of Information Management Data Insights, vol. 1, no. 2, pp. 100035, 2021. https://doi.org/10.1016/j.jjimei.2021.100035

[23] A. C. Ikegwu et al., "Big data analytics for data-driven industry: a review of data sources, tools, challenges, solutions, and research directions," Cluster Computing, vol. 25, no. 5, pp. 3343–3387, 2022. https://doi.org/10.1007/s10586-021-03408-1

[24] J. Sheng, J. Amankwah-Amoah, Z. Khan, and X. Wang, "COVID-19 pandemic in the new era of big data analytics: methodological innovations and future research directions," British Journal of Management, vol. 32, no. 4, pp. 1164–1183, 2021. https://doi.org/10.1111/1467-8551.12505

[25] S. Sharma, V. K. Gahlawat, K. Rahul, R. S. Mor, and M. Malik, "Sustainable innovations in the food industry through artificial intelligence and big data analytics," Logistics, vol. 5, no. 4, pp. 66, 2021. https://doi.org/10.3390/logistics5040066

[26] I. Ahmed, M. Ahmad, G. Jeon, and F. Piccialli, "A framework for pandemic prediction using big data analytics," Big Data Research, vol. 25, pp. 100190, 2021. https://doi.org/10.1016/j.bdr.2021.100190

[27] J. Mageto, "Big data analytics in sustainable supply chain management: a focus on manufacturing supply chains," Sustainability, vol. 13, no. 13, pp. 7101, 2021. https://doi.org/10.3390/su13137101

[28] D. Ivanov and A. Dolgui, "A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0," Production Planning & Control, vol. 32, no. 9, pp. 775–788, 2021. https://doi.org/10.1080/09537287.2020.1745458

[29] A. Chehri, I. Fofana, and X. Yang, "Security risk modeling in smart grid critical infrastructures in the era of big data and artificial intelligence," Sustainability, vol. 13, no. 6, pp. 3196, 2021. https://doi.org/10.3390/su13063196

[30] H. Kamyab et al., "The latest innovative avenues for the utilization of artificial intelligence and big data analytics in water resource management," Results in Engineering, vol. 20, pp. 101566, 2023. https://doi.org/10.1016/j.rineng.2023.101566

[31] Y. Wang, Z. Xu, Y. Yao, J. Liu, and J. Lin, "Leveraging Convolutional Neural Network-Transformer synergy for predictive modeling in risk-based applications," IEEE International Conference on Electronics, Information Engineering and Computer Communication (EIECC), pp. 1565–1570, 2024. https://doi.org/10.1109/EIECC57985.2024.10456231

[32] F. M. Alhomayani and K. A. Alruwaitee, "A new financial risk prediction model based on deep learning and quasi-oppositional coot algorithm," Alexandria Engineering Journal, vol. 108, pp. 60–69, 2024. https://doi.org/10.1016/j.aej.2024.01.005