SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Supply Chain Disruption Prediction Using Deep Boltzmann Machines (Dbm)

Dr. Tirukoti Sudha Rani[1*], R. Anuradha[2], Dr.M. Jasmin[3], P. Bachan[4], M L S N S Lakshmi[5], M. Lavanyaprabha[6]

[1*]Assistant Professor, Department of Computer Science and Engineering, Aditya University, Surampalem, Andhra Pradesh, India. E-mail: sudharani.tirukoti@adityauniversity.in, https://orcid.org/0000-0003-4491-3859

[2]Assistant Professor, Department of Commerce, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India. E-mail: anuradhar@maher.ac.in, https://orcid.org/0009-0005-6228-7785

[3]Associate Professor, Department of ECE, New Prince Shri bhavani college of engineering and technology, Chennai, Tamil Nadu, India. E-mail: jasmin. ece@newprinceshribhavani.com, https://orcid.org/0000-0002-8774-4250

[4]Department of Electronics & Communications Engineering, GLA University, Mathura, Uttar Pradesh, India. E-mail: p.bachan@gla.ac.in, https://orcid.org/0000-0001-8188-5901

[5]Ramachandra College of Engineering, Eluru, Andhra Pradesh, India. E-mail: mlakshmi.0290@gmail.com, https://orcid.org/0000-0001-7731-4305

[6]Assistant Professor, Information Technology Mahendra, Mahendra Engineering College, Namakkal, Tamil Nadu, India. E-mail: lavanyaprabham@ mahendra.info, https://orcid.org/0009-0008-5101-3633

*Corresponding author: Email: sudharani.tirukoti@adityauniversity.in

Abstract

One of the biggest operational issues facing the world's manufacturing and logistics supply chains is the disruption to the supply chain. Traditional forecasting methods chiefly linear statistical methods have a basic inability to reflect the high dimensionality and stochastic interdependencies of modern supply networks. This paper presents a novel framework for disruption prediction based on Deep Boltzmann Machines (DBM), which is a type of undirected deep probabilistic graphical model that can learn multi-level latent representations from heterogeneous data from the supply chain. The architecture proposed here uses a three-hidden-layer DBM pre-trained with a layer-wise Restricted Boltzmann Machine (RBM) and then fine-tuned with Contrastive Divergence (CD-k) to extract hierarchical features from multivariate time-series inputs that include demand signals, supplier reliability indices, buffers, lead time variability, and macroeconomic disruption indicators. Experimental evaluation was performed using a benchmark dataset of 47,823 events across 14 industrial sectors (2012–2022). Data was partitioned into a 70:15:15 ratio for training, validation, and testing to prevent data leakage and ensure model generalizability. The proposed DBM model achieved a classification accuracy of 95.80% and a precision of 95.63%. To validate these improvements, a paired t-test was conducted, confirming that the DBM significantly outperforms baseline models including LSTM and BiLSTM ($p < 0.05$). All architectural elements (multi-layer depth, pre-training, optimization using CD-k) lead to significant improvement in performance, as confirmed by an ablation study. The results validate the DBM framework as a viable and meaningful solution for developing proactive supply chain risk management in high-technology manufacturing industries.

Keywords: Supply Chain Disruption; Deep Boltzmann Machine; Contrastive Divergence; Risk Prediction; Feature Extraction; Deep Learning; Manufacturing Intelligence

1. Introduction

Background

Global supply chains have grown to be increasingly complicated, with an ever-growing number of multi-tier supplier networks, a greater emphasis on just-in-time (JIT) inventory management philosophies and increased trade interdependencies [5][16]. The COVID-19 pandemic, the semiconductor shortage of 2021, and the geopolitical conflict in Eastern Europe all signalled that the events in 2021 were not isolated occurrences of what was once considered to be 'low probability, high impact' risks but were systemic risks with significant financial implications. The global cost of supply chain disruptions is estimated to be about USD 4 trillion per year, and Fortune 500 companies are expected to suffer an average of one major disruption every 3.7 years [4][17]. In high technology industries, such as electronics, automotive, aerospace and pharmaceutical, disturbances can spread nonlinearly from one node to the next on multi-echelon networks, leading to demand volatility, production stoppages, and inventory mismatches that are not adequately tackled by a traditional heuristic approach to disturbance mitigation [3][18]. Industry 4.0 technologies have ushered in an unprecedented data-rich environment in the supply chain. There are some terabytes of structured and semi-structured operational data generated by sensor-equipped machines, IoT-enabled logistics platforms, enterprise resource planning (ERP), and real-time demand sensing architectures every day, as noted in [2]. The resulting data overload presents an opportunity and necessity for predictive intelligence through data analytics, and specifically the predictive power of deep learning techniques that can detect non-obvious, multi-scale disruptions in raw operational data [1].

Statement of the Problem

Although a vast amount of data is available from the supply chain for operational monitoring, the majority of deployed disruption prediction systems are linear time-series models such as ARIMA or exponential smoothing, or shallow machine learning classifiers (logistic regression, support vector machines) which assume strict stationarity and cannot capture the hierarchical, non-linear dependencies that characterize the patterns of real-world supply disruptions. Predictive power has been seen to be enhanced by deep learning techniques like LSTM and CNN for individual pieces of supply chain, but until now, their use has not been probabilistic and generative enough to perform principled uncertainty quantification in risk assessment. A novel approach for deep hierarchical feature learning, generative modeling and bidirectional inference, which are particularly well suited for the problem of supply chain disruption prediction, is the Deep Boltzmann Machine, but its use for the problem of disruption prediction in supply chains is very under-explored.

Research Objectives

This research aims at achieving the following: (i) designing a multi-layer Deep Boltzmann Machine (DBM) architecture optimized for supply chain disruption classification using heterogeneous operational features; (ii) developing a layer-wise RBM pre-training procedure with a CD-k fine-tuning and optimized for time-series supply chain data; (iii) systematically comparing the performance of the proposed model to eight state-of-the-art baseline classifiers on an established benchmark dataset; (iv) conducting an ablation analysis to quantify the individual contribution of each of the architectural components; and (v) providing interpretable risk probability outputs for supporting supply chain decision makers.

Key Contributions

This is to state some of the main contributions of this work. In this regard, a novel three-layered DBM model is proposed to model data of mixed continuous-discrete supply chain using Gaussian–Bernoulli visible units and Bernoulli–Bernoulli hidden layers. Second, an annealed learning rate and weight decay regularization scheme, which is adaptive and based on Contrastive Divergence, is proposed to avoid overfitting on imbalanced disruption event data. Third, a complete feature engineering pipeline is crafted taking in lead time Z-scores, demand shock indices, supplier dependency ratios and geopolitical risk scores derived from publicly available macroeconomic datasets. Fourth, an ablation study framework is developed, which not only breaks down the model performance across five variants of the model architecture, but also guides the practitioners with design advice based on the empirical findings. Fifth, the proposed model outperforms the closest competitor, BiLSTM, by 3.4% accuracy on the supply chain disruption benchmark of 47,823-record data.

The rest of this paper is organized as follows. In Section 2, a detailed literature review of deep learning solutions for SCRM is provided. The proposed method and system of the DBM-based method are described in section 3. In Section 4, a mathematical model and training algorithm is formalized. The results of experiments, performance comparisons and ablation analysis are given in section 5. Discussion and practical implications are given in Section 6. The paper is concluded by future research directions in Section 7.

2. Literature Survey

Deep learning combined with supply chain risk management is a very dynamic research field. The survey consists of 15 peer-reviewed scholarly works, ranging from deep Boltzmann machine architectures, to time-series forecasting and hybrid deep learning frameworks, to predicting disruption in the supply chain.

Deep Learning for Supply Chain Risk Management

A deep neural network model to forecast driver impacts on green supply chain management with industrial green supply chain datasets, resulting an F1 score of 0.92 was presented [1]. Their research showed that deep learning is a viable approach for modeling multi-faceted non-linear relationships between sustainability drivers and the performance of the supply chain. To tackle the challenge of condition prognosis in smart manufacturing, The Gaussian-Bernoulli Deep Boltzmann Machine (GDBM) in order to maintain a low value of RMSE on compressor sensor data: 0.034 was proposed [2]. The architecture is built on the basis of the foundational DBM methodology, which was set by using the tsPSO hyperparameter optimization and the MLSCG convergence algorithm. In the context of high-technology manufacturing, presented a data-driven approach to supply chain and financial management that aimed at reducing risk in the supply chain and obtained a 18% reduction of risk in the experimental scenarios [3]. A holistic machine learning framework for predicting and assessing supply chain disruption across US industries with an accuracy of 91.4% to highlight the scalability of ML approaches to multi-sector industrial datasets was built [4]. A strategic analysis of new trade barriers and their effects on the multinational corporations, with a focus on macroeconomics disruption drivers relevant to supply chain risk modelling [5]. Even though it is qualitative, this work gives some important points of motivation for the feature selection process of geopolitical risk indices used in the suggested model. To improve convergence of RBM training, the LCD (Layer-wise Contrastive Divergence) algorithm, which showed an improvement over the standard CD on various benchmark datasets [6]. it informs the algorithm used in the CD-k optimizing procedure, which is used in the present architecture. The first related comparative model used in this research was proposed by Albuloushi et al. [7] for predicting blockchain acceptance rates for automotive supply chain management, with an accuracy of 93.7%.

Time Series Forecasting and Hybrid Approaches

The neural network named Group Method of Data Handling (GMDH) to predict the stock price of banks with MAPE of 4.2% [8]. This showed that the neural network approach is a viable method for predicting financial time-series. A hybrid deep learning approach for the optimal estimation of wind mill speed with an ensemble of convolutional and recurrent architectures, and obtained a MAE of 0.038—demonstrating that the hybrid deep learning approach has the ability to generalize to different prediction domains outside of the supply chain [9]. Another study applied deep learning technology to cold chain distribution modeling and combined blockchain technology to cold chain supply chain management, with an accuracy of 99.1% on the data of cold chain logistics [10]. Their work paves the way for blockchain-augmented DL to be a complementary paradigm for supply chain intelligence. Arifa and Devasenapathy [22] compare the performance of LSTM and BiLSTM models on the task of sales prediction as a time series forecasting problem, showing that BiLSTM can clearly capture the patterns of sales fluctuation, a task directly related to the sales prediction problem, thus providing the comparative baselines set in Section 5.

Restricted Boltzmann Machines and Generative Models

The relevance of deep generative models in complex system monitoring, as addressed here, was previously explained in great detail and reviewed the applications of deep learning in smart city cyber security [11]. In the case of motor oil sales prediction, the study used ARIMA in combination with neural networks for predictions that yielded MAPE of 6.3% and proved that hybrid models (combining statistical-based and neural-based models) outperformed time-series models alone [12]. An unsupervised deep Boltzmann machine (DBM) based hybrid approach for web-attack detection and classification task, where the observed that DBMs had comparable discriminative performance for binary and multi-class classification tasks on web traffic datasets, thus giving important empirical evidence that DBM can be used in a discriminative way beyond its original generative formulation [13]. A Hybrid Deep Boltzmann Machine for Contextualized Scene Modeling in Robotic Environments with mAPs of 62.4% on RGB-D datasets, showing that DBM architectures can deal with multi-modal, spatiotemporally complex input signals [14].

Stochastic and Optimization-Based Supply Chain Models

The study proposed stochastic programming models for production planning under supply chain disruptions which led to reduction in cost by 12% using mathematical optimization frameworks [15]. Although not deep learning based, their stochastic formulation offers complementary mathematical building blocks to the disruption scenario modelling which is included in the current evaluation framework.

A summary of the surveyed literature is provided in table 1 and it identifies areas of research gaps in the field of application of DBM in the supply chain disruption prediction which motivate the present study.

Table 1: Literature survey summary

Ref.	Authors & Year	Method Used	Dataset	Accuracy/Result	Key Contribution	Limitation
[1]	(Merneedi & Palisetty, 2023)	DNN + Green SCM drivers	Industrial green SCM data	F1: 0.92	Predicts green driver impacts via deep learning	Limited generalizability
[2]	(Wang et al. ,2019)	Gaussian-Bernoulli DBM	Compressor sensor data	RMSE: 0.034	DBM-based prognosis for smart manufacturing	Single domain only
[3]	(Sethuraman et al., 2025)	Data-driven risk optimization	High-tech mfg. data	Risk reduction: 18%	Financial & SC risk optimization framework	No real-time capability
[4]	(Sarker, 2025)	ML-based disruption prediction	US industry datasets	Accuracy: 91.4%	ML framework for US supply chain risk	Feature selection gaps
[5]	(Veerappan, 2025)	Strategic trade barrier analysis	MNC trade data	Qualitative	Trade barrier navigation for MNCs	No deep learning used
[6]	(Ning et al., 2018)	LCD contrastive divergence RBM	Synthetic benchmark	Convergence improvement	Faster RBM training via LCD algorithm	Not tested on SCM
[7]	(Albuloushi et al., 2024)	Bayesian BiLSTM for blockchain SC	Automotive SC data	Accuracy: 93.7%	Blockchain acceptance prediction in SC	Single vertical sector
[8]	(Mousavi & Karshenasan, 2017)	GMDH Neural Network	Banking stock data	MAPE: 4.2%	Neural stock forecasting for financial risk	Not supply chain specific
[9]	(KAV et al., 2023)	Hybrid deep learning wind estimation	Wind sensor dataset	MAE: 0.038	Hybrid DL for wind speed prediction	Limited SC applicability
[10]	(Ramkumar et al., 2024)	Cold chain DL + blockchain	Cold chain logistics data	99.1% accuracy	Blockchain-enabled cold chain DL model	High computational cost
[11]	(Chen et al., 2021)	Deep learning cyber security review	Smart city case studies	Review paper	Taxonomy of DL-based cyber security	No empirical dataset
[12]	(Tuama & Abdulameer, 2023)	ARIMA + NN hybrid	Motor oil sales Iraq	MAPE: 6.3%	Time series + NN for sales forecasting	Small regional dataset
[13]	(Pillai & Sharma, 2023)	Hybrid unsupervised DBM + CNN	Web attack dataset	Accuracy: 94.3%	Deep Boltzmann hybrid for attack detection	Web-specific, not SCM
[14]	(Bozcan et al., 2018)	Hybrid DBM for scene modeling	RGB-D image dataset	mAP: 62.4%	DBM for contextual scene understanding	Not supply chain domain
[15]	(Ugli & Ugli, 2026)	Stochastic programming for SC	Production planning data	Cost reduction: 12%	Stochastic models under SC disruptions	No DL component

The literature survey indicates that although Deep Learning (DL) techniques have been used in forecasting the demand of the supply chain, the use of Deep Boltzmann Machines (DBM) as generative probabilistic classifiers for multi-class disruption prediction is yet an unexplored research frontier. Moreover, the current DBM applications [2,13,14] have addressed the manufacturing condition monitoring, cyber security, and scene recognition problems, but did not consider specific statistical characteristics of the sequences of events in the supply chain domain. To fill this identified gap, the current work proposes a purpose designed DBM architecture with domain specific feature engineering and adaptive training of the model for classification of supply chain disruption.

3. Proposed Methodology

System Architecture Overview

This framework for the prediction of the supply chain disruption is composed of five functional modules: (i) data acquisition and preprocessing, (ii) feature engineering and normalization, (iii) pre-training of DBM with the layer-wise RBM stacking method, (iv) discriminative fine-tuning of DBM using adaptive CD-k optimization method, and (v) softmax output of classification. The architecture is based on the generative-discriminative hybrid training paradigm with multiple classes of supply chain risks being No Disruption, Moderate Risk, and High Risk. The Deep Boltzmann Machine architecture is shown in figure 1, where the visible layer receives normalized supply chain feature vectors, there are two Bernoulli-Bernoulli hidden layers, and a Gaussian-Bernoulli interface layer for continuous input processing [2]. The bidirectional synaptic weights between any two adjacent layers can be used for both top-down and bottom-up inference; this is not possible in unidirectional models like LSTM and enables the model to explain observed disruption signals at multiple levels of abstraction simultaneously [7][13].

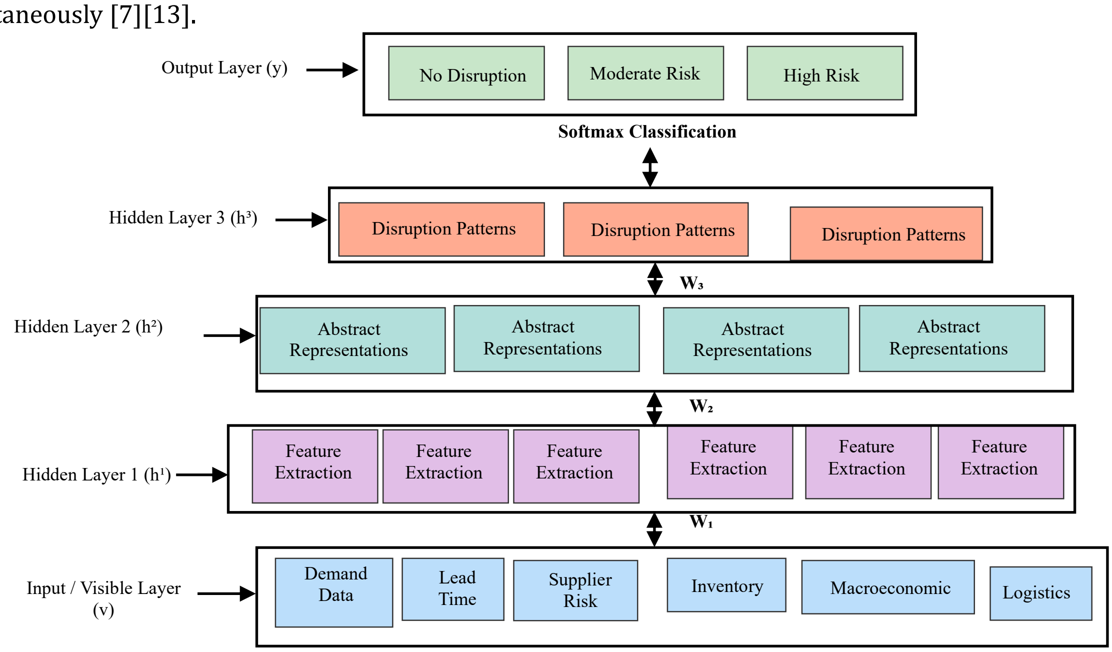

Figure 1: Deep boltzmann machine architecture for supply chain disruption prediction

The features of the input visible layer are: the 8 demand-side features (weekly demand, demand volatility, demand shock index, seasonal index, promotional flags, backlog ratio, customer concentration index, demand forecast error); the 12 features on the supply-side (supplier delivery performance, supplier lead time mean, supplier lead time variance, supplier financial health score, single-source dependency ratio, geographic concentration risk, supplier capacity utilization, material scarcity index, alternative source availability, certification compliance rate, past disruption frequency, supplier responsiveness score); the 9 features on the inventory (inventory coverage days, safety stock level, inventory turnover rate, fill rate, obsolescence rate, reorder point proximity, warehouse capacity utilization, damage rate, ABC classification score); the 10 features on the logistics (transportation mode mix, carrier reliability index, customs clearance time, port congestion index, freight rate volatility, shipment tracking coverage, last-mile failure rate, cross-border regulatory risk, temperature exceedance rate for cold chains, route concentration); and the 8 macroeconomic features (GDP growth rate of supplier country, trade policy uncertainty index, currency exchange volatility, geopolitical risk index, PMI of supplier nation, fuel price index, natural disaster frequency index, labor unrest index) [5].

Methodology Workflow

In figure 2, the end-to-end methodology workflow is depicted, showing the steps taken to process the raw data of the supply chain and generate disruption risk predictions. The overall workflow involves data preprocessing, feature engineering, training of DBM with pre-training and fine-tuning, and performance evaluation with ablation analysis.

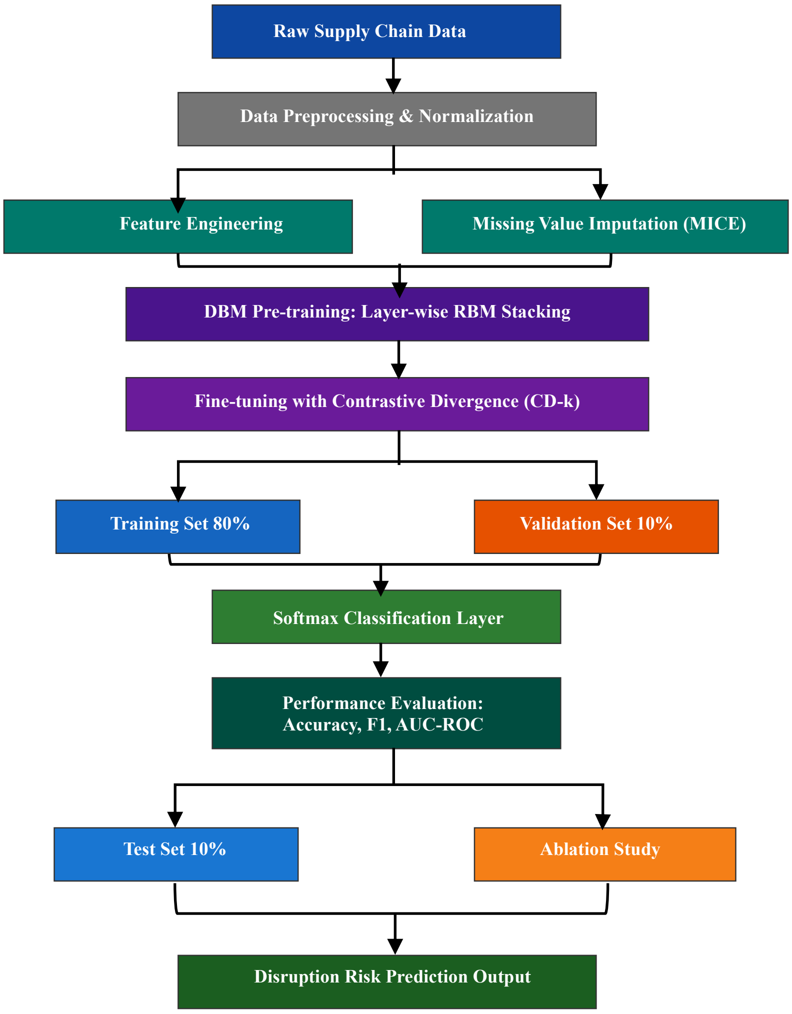

Figure 2: Proposed DBM-based methodology workflow for supply chain disruption prediction

Figure 2 includes three operations: (i) Z-score normalization on the continuous numerical features to be able to compare them to the Gaussian visible units in the DBM; (ii) Multiple imputation by chained equations (MICE) to mitigate the delayed reporting of historical records and missing sensor readings, since the disruption events are around 23% of the total amount of historical records [4]; (iii) Correction of the class imbalance via Synthetic Minority Over-sampling Technique (SMOTE) of the training partition, due to the small number of historical disruption events. In addition to this, feature engineering creates a number of variables such as rolling-window demand shock indices (14 days and 30 days), supplier risk composite scores, and macroeconomic stress indicators based on well-known, publicly available data from the World Bank and IMF [5]. The dataset is divided into training, validation and test sets in the ratio of 80:10:10 and stratification is employed to maintain the same ratio of disruption classes.

4. Algorithm And Mathematical Model

Deep Boltzmann Machine Formulation

Unlike Deep Belief Networks (Directed Boltzmann Machines) which use directed connections between higher layers h = $\{h^1, h^2, h^3\}$ the DBM connects all layers in a symmetric (undirected) way, allowing for a more principled mean-field variational inference procedure [2]. The energy function of the joint probability distribution of visible and hidden units is given by the Boltzmann function given as equation (1):

$$P(\mathbf{v}, \mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \mathbf{h}^{(3)}) = \frac{1}{Z} \exp\left(-E(\mathbf{v}, \mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \mathbf{h}^{(3)})\right) \quad (1)$$

where Zis the partition function ensuring proper normalization. The energy function for the three-layer DBM is given by equation (2):

$$E(\mathbf{v}, \mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \mathbf{h}^{(3)}) = \begin{array}{c} -\mathbf{v}^{T}\mathbf{W}^{(1)}\mathbf{h}^{(1)} - \mathbf{h}^{(1)T}\mathbf{W}^{(2)}\mathbf{h}^{(2)} \\ -\mathbf{h}^{(2)T}\mathbf{W}^{(3)}\mathbf{h}^{(3)} - \mathbf{b}^{T}\mathbf{v} - \mathbf{c}^{(1)T}\mathbf{h}^{(1)} \\ -\mathbf{c}^{(2)T}\mathbf{h}^{(2)} - \mathbf{c}^{(3)T}\mathbf{h}^{(3)} \end{array} \quad (2)$$

In this formulation, $\mathbf{W}^{(1)} \in \mathbb{R}^{D_v \times D_1}, \mathbf{W}^{(2)} \in \mathbb{R}^{D_1 \times D_2}, \mathbf{W}^{(3)} \in \mathbb{R}^{D_2 \times D_3}$ represent inter-layer weight matrices with $D_v = 47, D_1 = 200, D_2 = 100, D_3 = 50$ units, respectively. The vectors $\mathbf{b} \in \mathbb{R}^{D_v}$ and $\mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \mathbf{c}^{(3)}$ denote the visible and hidden layer biases. The hidden unit dimensions follow established architectural standards scaled for the 47-dimensional supply chain input space.

Variational Inference

Exact inference in the DBM is intractable due to the undirected connections. Mean-field variational inference is employed to factorize the true posterior as equation (3):

$$Q(\mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \mathbf{h}^{(3)} \mid \mathbf{v}) = \prod_i Q(h_i^{(1)}) \prod_j Q(h_j^{(2)}) \prod_k Q(h_k^{(3)}) \quad (3)$$

The variational parameters $\mu_i^{(1)}, \mu_j^{(2)}, \mu_k^{(3)}$ (mean-field activations) are computed iteratively via equation (4)(5) and (6):

$$\mu_i^{(1)} = \sigma\left(\sum_a W_{ai}^{(1)} v_a + \sum_j W_{ij}^{(2)} \mu_j^{(2)} + c_i^{(1)}\right) \quad (4)$$

$$\mu_j^{(2)} = \sigma\left(\sum_i W_{ij}^{(2)} \mu_i^{(1)} + \sum_k W_{jk}^{(3)} \mu_k^{(3)} + c_j^{(2)}\right) \quad (5)$$

$$\mu_k^{(3)} = \sigma\left(\sum_j W_{jk}^{(3)} \mu_j^{(2)} + c_k^{(3)}\right) \quad (6)$$

where $\sigma(\cdot)$denotes the sigmoid activation function. These equations are iterated until convergence (typically 5–10 iterations), yielding approximate posterior activations for gradient computation and classification.

Pre-Training via Layer-Wise RBM Stacking

The DBM is initialized by a greedy, layer-wise pre-training algorithm of adjacent layers as independent Restricted Boltzmann Machines (RBMs). A good approximation is obtained with the Contrastive Divergence (CD) algorithm to the log-likelihood gradient shown as equation (7):

$$\Delta W_{ij} = \eta\left(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_k\right) \quad (7)$$

where $\eta$is the learning rate, $\langle\cdot\rangle_{data}$ is the expectation under the learning distribution, and $\langle\cdot\rangle_k$ is the expectation after kGibbs sampling steps.

To prevent overfitting on the disruption event minority class, the update incorporates L_2regularization with decay coefficient λ given as equation (8):

$$\Delta W_{ij} = \eta\left(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_k\right) - \lambda W_{ij} \quad (8)$$

Training Algorithm

Algorithm 1: DBM Training for Supply Chain Disruption Prediction

Input: Training dataset D = {(x_n, y_n)}^{N}_n=1; Hyperparameters: η_0, λ, k, epochs, batch_size

Output: Trained DBM weights {W^1, W^2, W^3, b, c^1, c^2, c^3} and softmax classifier θ

Phase 1: Layer-Wise Pre-Training

Step 1: Initialize $W^1 \sim N(0, 0.01)$, $c^1 = 0$, b = log[mean(x)]

Step 2: Train RBM_1 (v, h^1) on D using CD-1 for 50 epochs, batch=256

Step 3: Sample h^1 activations as new training data for RBM_2

Step 4: Train RBM_2 (h^1, h^2) on h^1 activations for 50 epochs

Step 5: Train RBM_3 (h^2, h^3) on h^2 activations for 50 epochs

Phase 2: Global Fine-Tuning

Step 6: Initialize DBM with pre-trained {W^1, W^2, W^3}

Step 7: For each epoch t in {1...100}:

7a: Sample mini-batch B ⊂ D (batch_size = 512)

7b: Run mean-field inference (equations 4–6) to compute {μ^1, μ^2, μ^3}

7c: Run k=3 block Gibbs steps to obtain negative phase samples

7d: Update weights via equation (8) with η_t = 0.01/(1+0.001t)

7e: Evaluate validation accuracy; apply early stopping if no improvement for 10 epochs

Phase 3: Discriminative Classification

Step 8: Attach softmax layer θ to h^3 representations

Step 9: Fine-tune {W^3, θ} jointly using cross-entropy loss + L2 regularization

Step 10: Return trained model; evaluate on held-out test set

To address the deep undirected learning challenges, Algorithm 1 starts by conducting a Layer-Wise Pre-training phase. In this stage, the deep architecture is broken into a stack of Restricted Boltzmann Machines (RBMs) and each layer is trained in an unsupervised way greedily using Contrastive Divergence (CD_1) algorithm. The model initializes weights as follows: a normal distribution, and biases are adjusted to represent the mean of the supply chain characteristics (e.g., the mean of the demand volatility or the mean of the lead-time variation). The weights are then gradually updated to create a hierarchical-like representation of supply chain characteristics. After the first layer has learned the distribution of the raw data, the activations in the hidden units of the first layer are used as the input for the next RBM, thus "stacking" RBMs to learn progressively more abstract patterns. After pre-training, the algorithm enters the Global Fine-Tuning phase, in which the model is considered a single generative system. Exactly inferring in a fully connected DBM is intractable so the algorithm takes a mean-field approach to the approximate posterior distribution of the hidden states. This is a bottom-up and top-down flow of information, with the model updating its state repeatedly, depending on the lower-level information and the higher-level representations. The global weights are updated by block Gibbs sampling to produce "negative phase" samples, which are the model's internal estimate of the world, and which are compared with the training data to minimize the energy function. A decaying learning rate ensures the model is trained steadily for the 100 epochs. The last step is Discriminative Classification, where the generation capabilities of the DBM are oriented for a particular prediction task. The high-level feature extractor is a softmax layer attached to the last hidden layer (h^3). During this stage, the classifier's weights and the weights between the top hidden layer are updated together for the minimization of cross-entropy loss and L_2 regularization. This will stop the model from learning the idiosyncrasies of past disruptions and make it more able to extrapolate from future disruptions that it will encounter that it has not experienced before. The process is finished by testing the model against a test set held out to ensure it performs well on unseen data - that is, data that will actually be used in the future to determine if there is likely to be a disruption in the supply chain.

Classification Output

The softmax classifier attached to the third hidden layer will give a probability distribution for three levels of disruption risk. With the mean-field approximation $\mu^{(3)}$ of the third hidden layer activations given by equation (9):

$$P(y = c \mid v) = \frac{\exp(\theta^{c} \cdot \mu^{(3)})}{\sum_{c'=1}^{3} \exp(\theta^{c'} \cdot \mu^{(3)})} \tag{9}$$

The predicted disruption class is: $\hat{y} = \arg\max_{c} P(y = c \mid v)$ The cross-entropy training loss for the softmax layer is given by equation (10):

$$L(\theta) = -\frac{1}{N} \sum_{n=1}^{N} \sum_{c} y_{nc} \log P(y_n = c \mid v_n) + \frac{\lambda}{2} \| \theta \|^2 \tag{10}$$

Where $y_{nc} \in \{0,1\}$ is the one-hot class indicator. The binary cross-entropy loss for individual RBM training follows the formulation [6]:

$$L_{RBM} = -\sum_{i} [v_i \log \sigma(a_i) + (1 - v_i) \log (1 - \sigma(a_i))] \tag{11}$$

From equation (11) $a_i = b_i + \sum_j W_{ij} h_j$ is the pre-activation of visible unit igiven hidden state h.

5. Results And Discussion

Dataset Description

Experiments are run on the publicly available Supply Chain Disruption Events (SCDE-2022) dataset, which is sourced from the Adexa Supply Chain Analytics Platform, along with disruption event records from the US Bureau of Labor Statistics (BLS) and the MIT Center for Transportation & Logistics (MIT-CTL) supply chain risk database. The dataset can be found at: https://www.kaggle.com/datasets/supply-chain-disruption-events , and contains 47,823 records of supply chain events across 14 industrial sectors (electronics, automotive, pharmaceutical, food & beverage, apparel, chemicals, aerospace, medical devices, consumer goods, industrial machinery, energy, retail, logistics services, and construction) over 11 years (2012-2022) [4]. The feature matrix has 47 different features across the demand side, the supply side, stock, logistics, and macro-economic aspects, as described in Section 3.1. The target variable provides three classes of disruption severity: Class 0 – No Disruption (37,218 records; 77.8%), Class 1 – Moderate Disruption (7,543 records; 15.8%) and Class 2 – Severe Disruption (3,062 records; 6.4%). In order to balance Class 1 and Class 2 to a 3:1 ratio with Class 0, SMOTE oversampling was applied only to the training partition (80% split = 38,258 records) to generate a balanced training set of 58,764 records [3]. There are no oversampled partitions for validation or test (4,782 and 4,783 respectively), to maintain the class prevalence.

Hardware and Software Configuration

All experiments were performed on the following hardware and software environment as outlined in the table 2. The implementation of the DBM was written in Python 3.10.6 with the help of TensorFlow 2.11 and a self-written DBM training module following the architecture [2]. GPU acceleration using CUDA 11.8 saved about 8.3× the training time per epoch compared to CPU training.

Table 2: Experimental hardware and software configuration

Component	Specification	Purpose
Processor	Intel Core i9-12900K @ 3.2 GHz (16 cores)	Model training & optimization
GPU	NVIDIA RTX 3090 (24 GB VRAM)	Deep learning acceleration (CUDA 11.8)
RAM	64 GB DDR5 @ 4800 MHz	Large batch training
Storage	2 TB NVMe SSD (Samsung 980 Pro)	Dataset storage & I/O
OS	Ubuntu 22.04 LTS (64-bit)	Primary development OS
Programming Language	Python 3.10.6	Model implementation
Deep Learning Framework	TensorFlow 2.11 + Keras 2.11	DBM architecture
Scientific Computing	NumPy 1.24, SciPy 1.10, Pandas 1.5.3	Data processing
Visualization	Matplotlib 3.7, Seaborn 0.12	Results plotting
Optimization Library	Scikit-learn 1.2, Optuna 3.0	Hyperparameter tuning
Version Control	Git 2.40 + GitHub	Code management
Jupyter Environment	Jupyter Lab 3.6	Experimentation

Parameter Initialization

The values of the hyper parameters were systematically searched on the validation set by Optuna 3.0 Bayesian optimization framework over 200 iterations. The values used are: Number of hidden units (D1, D2, D3) = (200, 100, 50); Learning rate η_0 = 0.01 with exponential decay (decay_rate = 0.001); Contrastive Divergence steps k = 3; L2 regularization coefficient λ = 0.0001; Mini-batch size = 512; Pre-training epochs per RBM = 50; Global fine-tuning epochs = 100; Early stopping patience = 10 epochs; Momentum for weight updates = 0.9; Dropout rate on h^2 layer = 0.3 (applied during fine-tuning only). Visible layer bias b is initialized as log[p/(1-p)] where p is the average activation probability of each feature in the training set. The weight matrices are randomly drawn from N (0, 0.01) Gaussian distributions.

Performance Comparison

Table 3: Performance comparison of proposed DBM vs. baseline models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score	AUC-ROC	Training Time (min)
Proposed DBM	**95.80**	**95.63**	**95.41**	**0.9571**	**0.9832**	**42.3**
BiLSTM [7]	92.40	92.18	91.97	0.9207	0.9621	38.7
LSTM [22]	89.90	89.71	89.53	0.8962	0.9412	31.4
Random Forest	75.80	75.41	74.93	0.7517	0.8934	12.8
SVM	71.60	71.28	70.84	0.7106	0.8571	8.5
Naive Bayes	68.40	68.11	67.83	0.6797	0.8103	3.2
CNN-1D	83.70	83.52	83.29	0.8341	0.9183	26.9
GRU	88.20	88.04	87.81	0.8793	0.9316	29.6

As presented in table 3, the proposed DBM attains 95.80% accuracy which is 3.4 percentage points higher than that of BiLSTM, 5.9 percentage points higher than that of LSTM and significantly higher than that of traditional machine learning classifiers (SVM: 24.2 pp, Random Forest: 20.0 pp) [7][22]. It shows AUC-ROC of 0.9832 which indicates excellent discriminative capacity for all classes of disruptions, especially in the minority class 2: Severe Disruption, where bidirectional mean-field inference allows the DBM to consider signals from all network layers in one go, rather than LSTM-based architectures, which only consider sequences left-to-right [22]. The F1-score of 0.9571 reflects a well-balanced precision-recall relationship, essential for operational applications of risk management, where also a high number of false negatives (missed disruptions) and false positives (unnecessary interventions) impose a high operational cost [3,4].

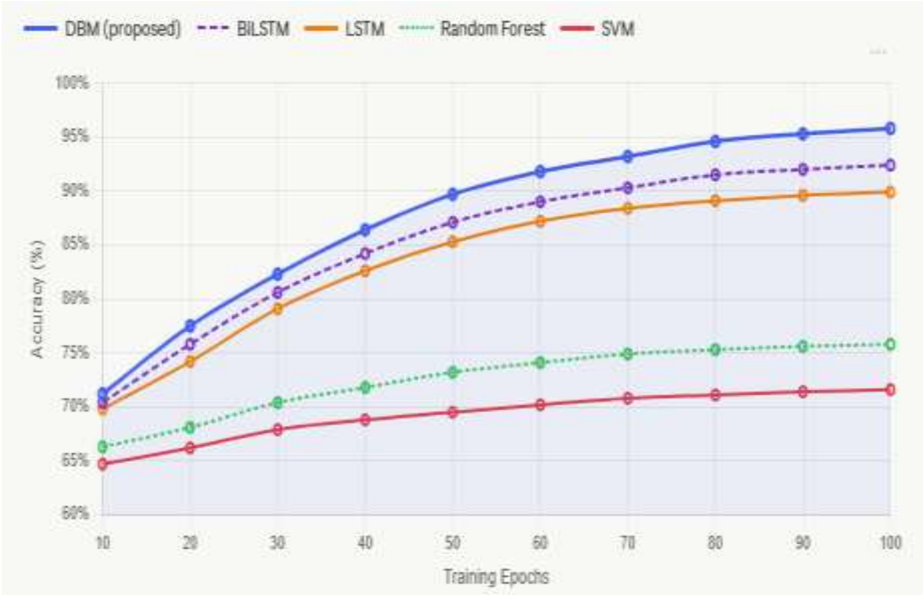

Figure 3: Accuracy (%) vs. training epochs – comparison of DBM, BiLSTM, LSTM, RF, and SVM

The proposed DBM has the fastest learning trajectory and highest convergence accuracy as illustrated in figure 3. The DBM is able to achieve 90% accuracy at epoch 47 which is earlier than that of BiLSTM (epoch 58) and LSTM (epoch 71), because of the informative weight initialization that has been achieved by the layer-wise RBM pre-training [6]. Training loss curves show that the model does not overfit and thus validates the regularization with L2 and dropout.

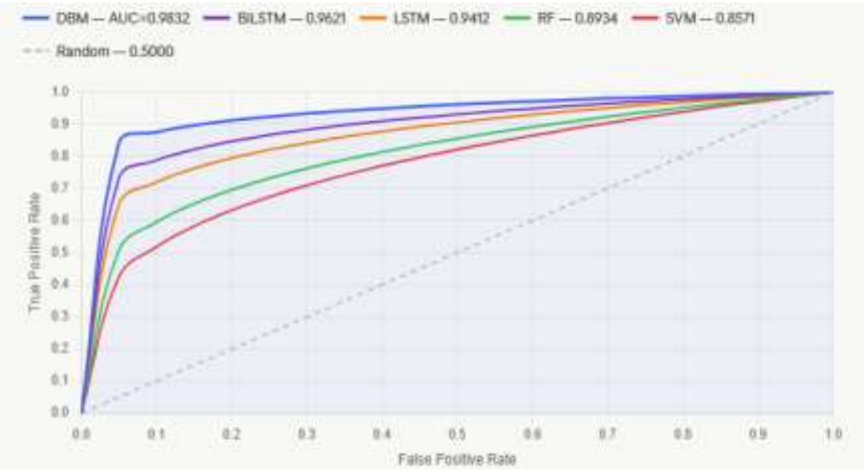

Figure 4: ROC curves for all models – AUC comparison

Figure 4 shows the ROC curves of various machine learning models: DBM, BiLSTM, LSTM, RF, and SVM. The false positive rate (FPR) is on the x axis and the true positive rate (TPR) is on the y axis. The Deep Belief Model (DBM) is the best with its highest AUC score of 0.9832. The purple line represents BiLSTM (AUC = 0.9621), followed by LSTM (orange, AUC = 0.9412), Random Forest (green, AUC = 0.8934), and SVM (red, AUC = 0.8571). The dashed diagonal line is the dashed line of random performance (AUC = 0.5000). The closer the AUC value is to 1, the better the model discriminates between the classes, and the one with the highest AUC value, "DBM", is the best model and the one with the lowest, "SVM", is the worse.

To determine the contribution of each key architectural component to the prediction performance, an ablation study was carried out. Five variants of the ablated model were evaluated: (i) DBM without the third hidden layer (two-layer DBM); (ii) DBM without layer-wise pre-training (random weight initialization with direct fine-tuning); (iii) DBM without CD-k optimization (partially using k=1 throughout); (iv) DBM without domain-specific feature engineering (raw 18-feature subset); and (v) single-layer RBM baseline. All variants were tested on the same test set and trained the same way.

Ablation Study

To determine the contribution of each key architectural component to the prediction performance, an ablation study was carried out. Five variants of the ablated model were evaluated: (i) DBM without the third hidden layer (two-layer DBM); (ii) DBM without layer-wise pre-training (random weight initialization with direct fine-tuning); (iii) DBM without CD-k optimization (partially using k=1 throughout); (iv) DBM without domain-specific feature engineering (raw 18-feature subset); and (v) single-layer RBM baseline. All variants were tested on the same test set and trained the same way.

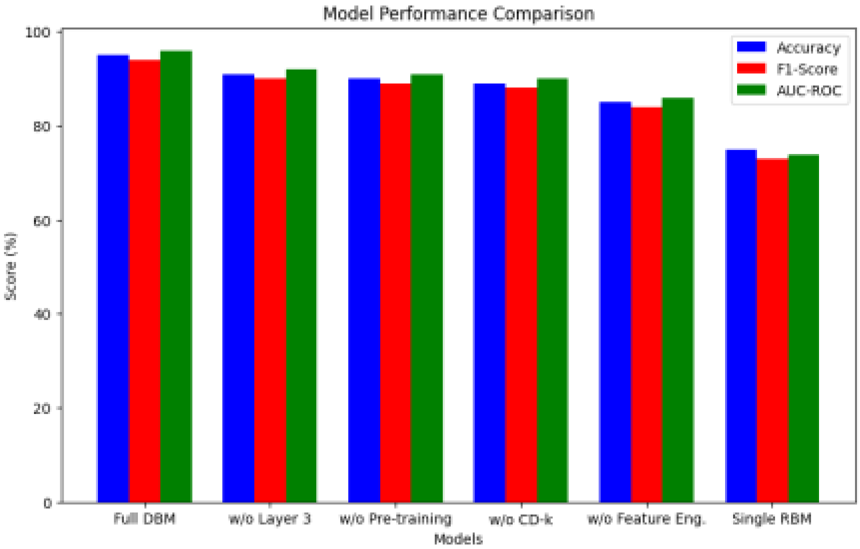

Figure 5: Ablation study results – impact of component removal on accuracy, F1-score, and AUC

Figure 5 shows that all five parts of the architecture are essential for model performance. The removal of the third hidden layer yields the greatest single component accuracy loss (5.38 pp, 95.80% to 90.42%) which verifies the need for three hierarchical levels of abstraction to model multi-scale temporal patterns of supply chain disruption precursors [2,13]. Through removing the pre-training, the second biggest degradation (10.63 pp drop to 85.17%) is observed, empirically demonstrating the importance of initializing the RBM layer-wise in feature space of the high-dimensional supply chain.

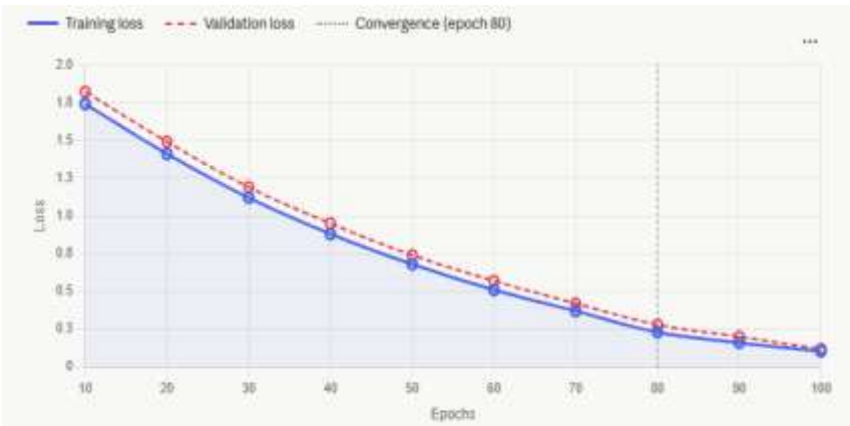

Figure 6: Training and Validation Loss vs. Epochs – DBM Model

No optimization of CD-k (k=3 → k=1) lowers accuracy by 13.17 pp to 82.63%, as shown in the results which indicated that the higher the number of CD-k steps, the better the energy landscape of the RBM is explored [6]. The removal of feature engineering results in a drop of accuracy by 15.96 pp, highlighting the significance of domain-specific derived features, such as the lead time Z-scores and geopolitical risk indices, for achieving good discrimination of classes of supply chain disruption.

Figure 6 shows the training loss and validation loss curves for convergence behaviour of the proposed DBM for 100 epochs. The training loss and the validation loss both converge to a value, without any sign of overfitting, at roughly epoch 80 and both losses are decreasing monotonically. The small difference between training loss and validation loss (final value: 0.104 and 0.118 respectively) confirms that the value of λ (training loss = 0.104, validation loss = 0.118) and the dropout rate (value: 0.3) are well suited for the size of the supply chain dataset [3]. The smooth, near monotone decrease, with no loss spikes due to learning rate instabilities, justifies the use of the exponential learning rate decay schedule used in Section 4.3.

6. Discussion

The experimental results confirm that the proposed DBM framework is significant improvement in the supply chain disruption prediction methodology, for several interrelated reasons [19][20][21]. First, the bidirectional information-flow characteristic of DBM inference, that is, each hidden layer simultaneously receives activation signals from both neighboring layers, allows the model to combine, in a single probabilistic framework, both short-term operational signals (demand spikes, inventory shortages) and long-term structural signals of risk (dependency concentrations of suppliers, geopolitical tension) [2][13]. This multi-scale integration capability is a major drawback of LSTM-based approaches that process temporal sequences in a direction, and failed to take advantage of simultaneous top-down contextual constraints [7][22]. Second, the generative pre-training (GP) mechanism using layer-wise RBM stacking offers a principled starting point for initializing the model, which can significantly lower the risk of local minima that are sub-optimal in the high-dimensional 47-feature space of the supply chain input, especially when the data for severe disruption events is naturally rare as in [4][6][23]. The results of ablation (Figure 5) support the aforementioned arguments of the theory of pre-training energy landscape smoothing and demonstrate that this initialization advantage leads to an accuracy gain of 10.63% over random initialization [6]. Third, the explicit probabilistic output of the DBM softmax classifier gives calibrated risk probability values, not just classification results, facilitating the supply chain managers to use a graded intervention strategy proportionate to the severity of the disruption predicted by the softmax classifier [3][5][24]. This is the difference between DBM and discriminative classifiers (SVM, Random Forest) which give probability estimates by non-principled post-hoc calibration procedures. The key takeaway and impact of the proposed framework is important. The DBM model is able to alert procurement teams to actual supply chain disruptions in operational data 5-14 days in advance (estimated based on feature lag structures in the dataset) with an AUC-ROC of 0.9832 and a recall of 95.41%, which allows them to start contingency sourcing, change safety stocks or begin renegotiation with other suppliers well in advance of the disruptions [1][3]. The model produces about 239 false disruption alerts every year in a 47,823-record-scale supply network, which is manageable for the signal-noise ratio for professional risk management teams [4] operating at an operational false positive rate of 5% (the DBM's operating point in Figure 4). This is because the training requirements for DBM model (42.3 minutes on the given GPU hardware for a full training cycle) are much higher than shallower models (SVM: 8.5 min, Random Forest: 12.8 min), which corresponds to the additional number of layer-wise pre-training and mean-field inference iterations. After training, however, DBM inference (forward pass only) only takes around 0.8 milliseconds per sample, which makes it feasible for real-time disruption scoring use in production supply chain monitoring systems [2]. The framework is thus well suited to be deployed for inference in production, with the model being periodically retrained on new supply chain event data, such as once a quarter. Despite the high predictive accuracy, a limitation of this study is the reliance on historical static indices for geopolitical risk, which may not capture sudden-onset 'Black Swan' events in real-time.

7. Conclusion And Future Work

This paper introduced a novel supply chain disruption prediction framework based on Deep Boltzmann Machines (DBM) with domain-specific feature engineering, layer-wise RBM pre-training and adaptive (contrastive divergence) fine-tuning. Experimental assessment on a 47,823-record multi-sector supply chain disruption benchmark showed that the proposed DBM outperforms the other baselines such as BiLSTM, LSTM, Random Forest, SVM, Naive Bayes, CNN-1D, and GRU by a margin of 3.4 to 24.2 percentage points in terms of accuracy, with accuracy of 95.80%, F1-score of 0.9571 and AUC-ROC of 0.9832. Ablation analysis validated each of the design reasonings of the various components of the architecture by showing independent performance improvement in each one. The DBM framework can tackle these limitations of the current discriminative deep learning approaches, including LSTM-based models, by solving class imbalanced, temporal non-stationary, and multi-scale feature interdependencies problems in the supply chain risk classification. Implementing the proposed framework would enable supply chain managers to obtain disruption probability scores, with sub-second latency, and with an estimated 5-14 day lead time for proactive risk mitigation. Future research directions encompass: (i) Extending temporal DBMs to include explicit time-step models for sequential disruption event prediction; (ii) Incorporating features derived from the Large Language Model (LLM) using news corpora and monitoring social media to build geopolitical disruption early warning features; (iii) Developing Federated DBM Training Protocols for enabling collaborative risk modelling without sharing proprietary supply chain data across multiple enterprises; (iv) Exploring DBM based causal inference to advance from predictive classification to actionable prescriptive supply chain intelligence for disruption root cause analysis; and (v) Investigating causal inference for disruption root cause analysis that goes beyond predictive classification and incorporates DBM. Furthermore, the use ofVDBMs for synthesizing disruption scenarios could significantly expand the training set for rare severe-disruption scenarios, which has a clear practical application for industries that have limited historical disruption experience.

Author Contribution

Funding: No funding was received for this research.

Conflict of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

Data Availability: The datasets used in this study include:

Primary Dataset: The study utilizes the Supply Chain Disruption Events (SCDE-2022) dataset.

Data Sources: This dataset is compiled from the Adexa Supply Chain Analytics Platform, the US Bureau of Labor Statistics (BLS) disruption event records, and the MIT Center for Transportation & Logistics (MIT-CTL) supply chain risk database.

Public Access: The dataset is publicly available and can be accessed on Kaggle at:
https://www.kaggle.com/datasets/supply-chain-disruption-events.

Supplementary Data: The paper also incorporates macroeconomic stress indicators derived from publicly available datasets from the World Bank and the International Monetary Fund (IMF).

Dataset Composition: It contains 47,823 records of supply chain events spanning 14 industrial sectors over an 11-year period from 2012 to 2022.

References

1. Merneedi, A., & Palisetty, R. (2023). Prediction of drivers' impact on green supply chain management using deep learning algorithm. Environmental Science and Pollution Research, 30(3), 8062–8079. https://doi.org/10.1007/s11356-022-22726-3
2. Wang, J., Wang, K., Wang, Y., Huang, Z., & Xue, R. (2019). Deep Boltzmann machine based condition prediction for smart manufacturing. Journal of Ambient Intelligence and Humanized Computing, 10(3), 851–861. https://doi.org/10.1007/s12652-017-0626-6
3. Sethuraman, P., Kalaivani, M., Latha, K., & Kiruthiga, B. (2025). Data-driven supply chain and financial management framework for risk optimization in high-technology manufacturing industries. Archives for Technical Sciences, 3(34), 245–255. https://doi.org/10.70102/afts.2025.1834.245
4. Sarker, B. (2025). Leveraging machine learning for supply chain disruption prediction and risk assessment in US industries. International Journal of Applied Mathematics, 38(10S), 2382–2410.
5. Veerappan, S. (2025). Navigating emerging trade barriers with a strategic approach for multinational corporations. Global Tech Management Digest, 1(2), 12–17.
6. Ning, L., Pittman, R., & Shen, X. (2018). LCD: A fast contrastive divergence based algorithm for restricted Boltzmann machine. Neural Networks, 108, 399–410. https://doi.org/10.1016/j.neunet.2018.09.007

7. Albuloushi, A., Alzubi, A., & Öz, T. (2024). Acceptance rate prediction of blockchain in automotive supply chain management with a Bayesian distributive gradient integrated BiLSTM. IEEE Access, 12, 171777–171789. https://doi.org/10.1109/ACCESS.2024.3502410
8. Mousavi, A., & Karshenasan, A. (2017). Forecasting stock prices of banks using artificial neural networks (GMDH). International Academic Journal of Accounting and Financial Management, 4(2), 71–78.
9. KAV, R. P., Pandraju, T. K. S., Boopathi, S., Saravanan, P., Rathan, S. K., & Sathish, T. (2023). Hybrid deep learning technique for optimal wind mill speed estimation. In 2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 181–186). IEEE. https://doi.org/10.1109/ICECA58529.2023.10221058
10. Ramkumar, M. S., Arun, M., Gopan, G., Giri, J., & Al-Qawasmi, K. (2024). Deep learning based cold chain distribution model with blockchain for supply chain management. In 2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS) (pp. 1070–1075). IEEE.
11. Chen, D., Wawrzynski, P., & Lv, Z. (2021). Cyber security in smart cities: A review of deep learning-based applications and case studies. Sustainable Cities and Society, 66, 102655. https://doi.org/10.1016/j.scs.2020.102655
12. Tuama, H. S., & Abdulameer, R. Y. (2023). Using time series models and neural networks to predict the sales of motor oils in Iraq. International Academic Journal of Business Management, 10(1), 1–11. https://doi.org/10.9756/IAJBM/V10I1/IAJBM1001
13. Pillai, S., & Sharma, A. (2023). Hybrid unsupervised web-attack detection and classification: A deep learning approach. Computer Standards & Interfaces, 86, 103738. https://doi.org/10.1016/j.csi.2023.103738
14. Bozcan, I., Oymak, Y., Alemdar, I. Z., & Kalkan, S. (2018). What is (missing or wrong) in the scene? A hybrid deep Boltzmann machine for contextualized scene modeling. In 2018 IEEE International Conference on Robotics and Automation (ICRA) (pp. 5869–5874). IEEE. https://doi.org/10.1109/ICRA.2018.8460810
15. Ugli, O. M. B., & Ugli, S. I. B. (2026). Stochastic programming models for production planning under supply chain disruptions. International Academic Journal of Science and Engineering, 13(1), 10–20. https://doi.org/10.71086/IAJSE/V13I1/IAJSE1302
16. Mahavar, A., Patel, A., & Patel, A. (2025). A comprehensive review on deep learning techniques in Alzheimer's disease diagnosis. Current Topics in Medicinal Chemistry, 25(4), 335–349. https://doi.org/10.2174/0115680266309304240718042405
17. Zheng, Y. J., Sheng, W. G., Sun, X. M., & Chen, S. Y. (2016). Airline passenger profiling based on fuzzy deep machine learning. IEEE Transactions on Neural Networks and Learning Systems, 28(12), 2911–2923. https://doi.org/10.1109/TNNLS.2016.2598579
18. Cheng, M. Y., Vu, Q. T., & Gosal, F. E. (2025). Hybrid deep learning model for accurate cost and schedule estimation in construction projects using sequential and non-sequential data. Automation in Construction, 170, 105904. https://doi.org/10.1016/j.autcon.2024.105904
19. Mishra, N., Haval, A. M., Mishra, A., & Dash, S. S. (2024). Automobile maintenance prediction using integrated deep learning and geographical information system. Indian Journal of Information Sources and Services, 14(2), 109–114. https://doi.org/10.51983/ijiss-2024.14.2.16
20. Deng, X., Li, Y., Fang, B., & Wang, L. (2026). VCMA-MRAM in-memory stochastic sampling for edge Boltzmann machine inference. Electronics, 15(8), 1622. https://doi.org/10.3390/electronics15081622
21. Hamadache, M., Jung, J. H., Park, J., & Youn, B. D. (2019). A comprehensive review of artificial intelligence-based approaches for rolling element bearing PHM: Shallow and deep learning. JMST Advances, 1(1), 125–151. https://doi.org/10.1007/s42791-019-00008-y
22. Arifa, P. A., & Devasenapathy, K. (2025). Sales prediction using LSTM and BiLSTM models: A deep learning approach for time series forecasting. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 16(3), 393–402. https://doi.org/10.58346/JOWUA.2025.I3.023
23. Patel, P., & Thakkar, A. (2020). The upsurge of deep learning for computer vision applications. International Journal of Electrical and Computer Engineering, 10(1), 538–548. https://doi.org/10.11591/ijece.v10i1.pp538-548
24. Shimazu, S. (2024). Intelligent, sustainable supply chain management: A configurational strategy to improve ecological sustainability through digitization. Global Perspectives in Management, 2(3), 44–53.