

TRANSFORMER-BASED CASH FLOW FORECASTING USING INVOICE-LEVEL DATA FOR PREVENTING SHORT-TERM LIQUIDITY SHORTFALLS IN MANUFACTURING FIRMS**Ayodeji Ikudabo**

Wilmington University, New Castle, DE, USA

ABSTRACT

Short-term liquidity management is critical to manufacturing firms, where mismatches between cash inflows and operational payment obligations can disrupt production, supplier relationships, and working-capital stability. Traditional cash-flow forecasting methods frequently depend on aggregated historical financial data and may inadequately capture complex payment behavior, invoice timing, and rapidly changing operating conditions. This study develops a transformer-based forecasting framework using invoice-level data to predict near-term cash flows and identify potential liquidity shortfalls before they materialize. The framework integrates invoice amounts, issuance and due dates, historical payment delays, customer payment behavior, supplier obligations, receivable and payable positions, and transaction sequences with firm-level liquidity indicators. Transformer architectures employ attention mechanisms to learn temporal dependencies and heterogeneous patterns affecting the timing and magnitude of expected cash movements. Forecasted inflows and outflows are subsequently translated into rolling liquidity positions and early-warning signals for anticipated cash deficits. Model performance is evaluated against conventional statistical and machine-learning forecasting approaches using MAE, RMSE, MAPE, forecast accuracy, and shortfall-detection measures. The proposed framework supports proactive treasury interventions, working-capital planning, payment prioritization, and financing decisions within manufacturing environments.

Keywords:

Transformer Models; Cash Flow Forecasting; Invoice-Level Data; Liquidity Risk; Working Capital Management; Manufacturing Firms

1. INTRODUCTION**1.1 Manufacturing Working-Capital and Liquidity Problem — 120 words**

Manufacturing firms operate through interconnected cash-conversion processes in which the timing of accounting transactions rarely coincides with actual cash movements [1]. Customer invoices may be issued weeks before payment is received, while supplier invoices, payroll, inventory procurement, production expenditure, taxes, debt service, and other operating obligations become payable according to different schedules [4]. These timing mismatches create substantial working-capital requirements, particularly for manufacturers maintaining inventories and extended production cycles [7]. Consequently, accounting profitability does not guarantee short-term liquidity. A profitable manufacturer may record substantial sales and positive earnings while simultaneously experiencing a temporary cash deficit because customers delay settlement as supplier and operating obligations mature [2]. Accurate forecasting of both expected cash receipts and required payments is therefore fundamental to maintaining operational continuity and avoiding costly emergency financing [9].

1.2 Limitations of Conventional Cash-Flow Forecasting

Conventional cash forecasting frequently relies on spreadsheets, rolling averages, ARIMA or exponential-smoothing models, and aggregated historical cash-flow series [3]. Although useful for identifying broad temporal patterns, these approaches can lose predictive information through aggregation [6]. Monthly or daily cash totals may conceal characteristics contained within individual transactions, including invoice amount, issuance date, contractual due date, customer or supplier identity, historical payment delay, payment terms, outstanding balance, and settlement behaviour [8]. Two periods with similar aggregate receivables may therefore have materially different expected cash-realization profiles because their underlying invoice portfolios differ. Forecasting directly from invoice-level observations can preserve this transactional heterogeneity and potentially identify payment delays, cash-flow timing changes, and emerging liquidity pressure earlier than models constructed exclusively from aggregated historical balances [5].

1.3 Transformer Motivation

Transformer architectures provide a promising mechanism for modelling this information-rich environment because attention mechanisms can learn long-range dependencies and interactions across sequences without relying exclusively on recurrent processing [4]. Invoice-level forecasting requires relationships among payment histories, invoice characteristics, temporal effects, heterogeneous categorical and numerical variables, and evolving operational conditions to be represented simultaneously [7]. Temporal Fusion Transformer-type architectures are particularly relevant because they were developed for interpretable multi-horizon forecasting using static covariates, known future inputs, and historically observed time-varying variables [9]. Such structures can therefore combine customer characteristics, contractual payment dates, outstanding invoices, historical settlement behaviour, calendar information, and operational variables while generating forecasts across multiple future horizons rather than producing only a single-period cash estimate.

1.4 Research Gap

Despite advances in financial forecasting, existing cash-flow research remains substantially oriented toward aggregated historical cash movements, conventional time-series specifications, or generic machine-learning predictors [2]. Comparatively limited attention has been given to Transformer architectures that exploit invoice-level transactional information specifically to anticipate operational liquidity shortfalls within manufacturing environments [6]. This leaves an important methodological gap between increasingly granular enterprise transaction data and the forecasting methods used for working-capital management. Addressing this gap requires linking invoice-level predictive modelling directly to forward-looking liquidity-risk identification rather than evaluating forecast accuracy in isolation [8].

1.5 Research Objectives and Questions

Accordingly, this study addresses four questions. RQ1: Can invoice-level Transformer modelling improve short-term cash-flow forecast accuracy [3]? RQ2: Which invoice, payment, and operational features contribute most strongly to forecast performance? RQ3: Does improved forecasting enable earlier and more accurate identification of liquidity shortfalls [5]? RQ4: Does the proposed architecture outperform statistical, machine-learning, and recurrent deep-learning benchmarks across alternative forecasting horizons [9]? Collectively, these questions connect predictive accuracy with financially consequential liquidity outcomes.

1.6 Contributions

The study makes four principal contributions. First, it develops an invoice-level forecasting framework that preserves transaction-specific information commonly lost through aggregation [1]. Second, it applies a Transformer-based architecture to multi-horizon manufacturing cash-flow prediction. Third, it connects forecast performance directly with liquidity-shortfall detection, thereby translating statistical accuracy into working-capital relevance [7]. Fourth, rigorous benchmarking, ablation analysis, and robustness testing establish whether predictive gains arise from the proposed architecture and transactional information rather than favorable model specifications [8].

2. LITERATURE REVIEW AND THEORETICAL FOUNDATION

2.1 Short-Term Cash-Flow Forecasting

Short-term cash-flow forecasting supports treasury decisions by estimating the timing and magnitude of future receipts and payments required to maintain operational liquidity [5]. Direct forecasting projects identifiable cash inflows and outflows, whereas rolling forecasts update expectations as new transactional information becomes available [8]. Traditional implementations frequently employ historical averages, statistical time-series models, and aggregated cash-flow observations. More recent approaches apply machine learning to capture nonlinear relationships among financial and operational variables [11]. Forecast quality has direct economic implications because inaccurate estimates can generate unnecessary cash holdings, emergency borrowing, delayed payments, or avoidable financing costs. Empirical evidence consequently links improvements in daily cash-flow forecast accuracy with reductions in cash-management costs [6]. Forecast accuracy should therefore be evaluated as both a statistical objective and a determinant of working-capital efficiency.

2.2 Invoice-Level Information and Working Capital

Invoice-level data provide forward-looking information that aggregated cash statements can obscure. Accounts receivable identify expected customer inflows, while accounts payable represent scheduled supplier obligations [7]. Their timing affects days sales outstanding and days payable outstanding, two important components of the cash-conversion cycle [10]. Individual invoices additionally contain contractual maturity dates, outstanding balances, overdue status, partial-payment histories, payment terms, customer identities, and supplier settlement requirements. Customer-specific payment reliability can reveal whether contractual due dates realistically represent expected settlement dates [13]. Similarly, supplier-level information can identify comparatively inflexible payment obligations. Consequently, two firms with identical aggregate receivable balances may possess

materially different near-term liquidity profiles when invoice maturity, payment delays, and counterparty behaviour differ, supporting transaction-level rather than exclusively aggregate forecasting.

2.3 Liquidity Shortfall Prediction

Cash-flow forecasting and liquidity-risk forecasting represent related but distinct analytical objectives [9]. Cash-flow forecasting predicts the timing and magnitude of expected cash movements, whereas liquidity-risk forecasting evaluates whether resulting cash availability is sufficient to satisfy obligations. A liquidity shortfall can therefore be defined as a forecast condition in which available liquidity falls below a predetermined operational safety requirement:

$$LS_{t+h} = \mathbb{1}(\hat{C}_{t+h} < L_{t+h}^{\min})$$

where $\hat{C}_{t+h}$ represents forecast available cash and $L_{t+h}^{\min}$ denotes the minimum liquidity requirement at horizon h [12]. This distinction translates forecasting errors into financially meaningful risk signals.

2.4 Statistical and Machine-Learning Models

Conventional forecasting methods provide essential benchmarks for evaluating whether increasingly complex architectures generate meaningful incremental performance. A naïve forecast establishes a minimal reference by projecting recent cash behaviour forward, while moving-average and exponential-smoothing methods capture local levels and trends [5]. ARIMA additionally models autoregressive and error structures in historical cash-flow series [11]. Machine-learning approaches such as Random Forest and XGBoost extend forecasting capacity by accommodating nonlinearities and interactions among transactional, financial, and operational predictors [14]. In this study, these approaches function as empirical benchmarks rather than literature references alone. Their comparative performance establishes whether invoice-level Transformer modelling produces gains beyond established statistical and nonlinear predictive methods.

2.5 LSTM/GRU and Transformer Forecasting

Recurrent neural architectures, particularly Long Short-Term Memory and Gated Recurrent Unit networks, improved sequence forecasting by learning temporal dependencies while addressing some limitations of conventional recurrent networks [8]. Nevertheless, sequential processing can constrain computational efficiency and complicate representation of very long historical relationships [12]. Transformer architectures instead employ self-attention to identify relevant dependencies across observations while supporting greater parallelization [10]. Temporal Fusion Transformer further accommodates static, known-future, and historically observed covariates, making it particularly suitable for heterogeneous multi-horizon financial forecasting [6]. PatchTST provides another development by converting subsequences into patches or tokens, enabling longer historical contexts while reducing attention-map computation [13]. Such architectures are relevant to manufacturing environments characterized by nonlinear interactions among invoices, customer payment behaviour, production requirements, and operating conditions [14]. They therefore provide a flexible foundation for invoice-informed liquidity forecasting.

2.6 Research Gap and Hypotheses

Limited research integrates invoice-level information, Transformer forecasting, and operational liquidity-shortfall detection within one manufacturing-finance framework [7]. Accordingly, four hypotheses are tested. H1: Transformer forecasts produce significantly lower MAE and RMSE than conventional benchmarks. H2: Invoice-level features improve forecasting relative to aggregate cash-flow-only inputs [9]. H3: Transformer forecasts improve liquidity-shortfall detection. H4: Customer payment-behaviour features provide greater incremental forecasting information than invoice amount alone [13]. These propositions establish the empirical framework developed in Section 3.

3. DATA ACQUISITION, CONSTRUCTION AND FEATURE ENGINEERING

3.1 Data Acquisition and Invoice-Level Structure

The empirical framework uses anonymised transactional records extracted from manufacturing enterprise-resource-planning and accounting systems. Data sources comprise sales invoices, purchase invoices, accounts-receivable and accounts-payable ledgers, payment transactions, customer and supplier master records, production and procurement information, and daily bank or cash balances [12]. The final implementation should report the observation period, number of participating firms or manufacturing plants, total invoice count, transaction frequency, base currency, and currency-conversion procedure. Where multiple currencies occur, transactions are converted consistently using the applicable transaction-date or accounting-policy exchange rate before modelling [15]. Observations are organized at daily frequency because short-term liquidity decisions require substantially greater temporal granularity than monthly financial statements [18].

Each invoice i observed at forecasting time t is represented as:

$$x_{i,t} = [A_i, D_i, M_i, C_i, P_i, S_i, O_i, L_i, \dots]$$

where A_i denotes invoice amount, D_i issue date, M_i contractual maturity date, C_i customer or supplier identifier, P_i payment terms, S_i settlement status, O_i outstanding amount, and L_i observed historical payment delay [13]. This representation preserves transaction-specific information that would disappear if receivables, payables, and cash movements were immediately aggregated [17].

Figure 1. End-to-End Research Framework

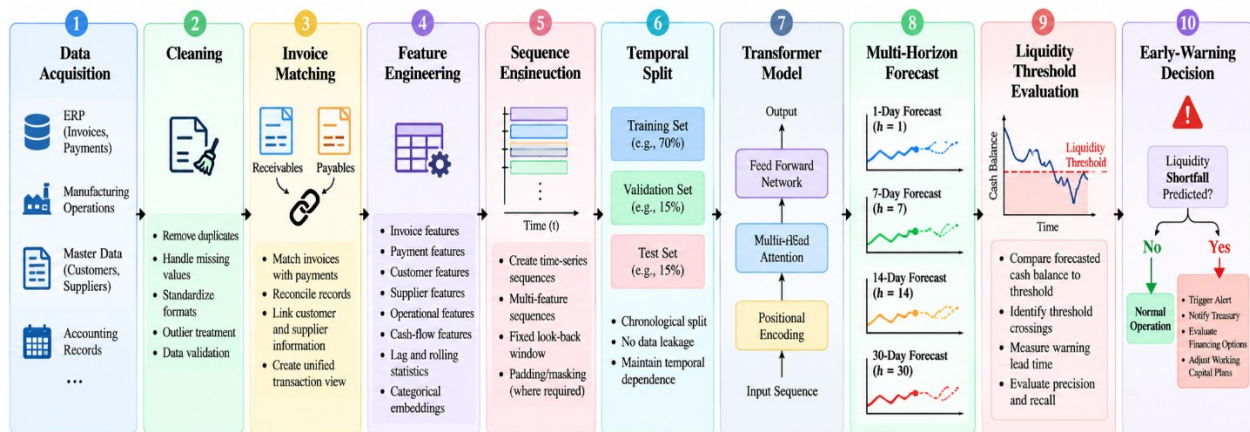


Figure 1. End-to-End Research Framework

Table 1. Dataset and Feature Dictionary

Feature group	Example variables	Type	Transformation
Invoice	Invoice amount, outstanding amount, invoice age	Numerical	Log transformation/standardisation
Temporal	Issue date, due date, payment date	Temporal	Cyclical encoding
Payment	Payment delay, payment ratio, overdue days	Numerical	Rolling/lagged features
Customer	Customer ID, segment, payment history	Categorical/mixed	Embedding/behavioural aggregation
Supplier	Supplier ID, supplier type, settlement history	Categorical/mixed	Embedding/behavioural aggregation
Operations	Procurement and production requirements	Numerical	Normalisation
Cash	Opening balance, inflows, outflows	Numerical	Scaling/lag construction

3.2 Data Cleaning and Preprocessing

Raw transactional records are reconciled before feature construction. Duplicate entries are removed, while cancelled invoices and credit notes are identified and treated according to their accounting effect rather than incorrectly classified as ordinary transactions [14]. Invoice-payment matching accommodates complete and partial settlements, enabling outstanding balances to be reconstructed through time. Missing observations are investigated before variable-appropriate imputation or exclusion, and extreme monetary values are winsorised where economically justified [19]. Currency treatment is standardized before aggregation. Most importantly, preprocessing follows a strict point-in-time information rule: every feature constructed at forecasting origin t contains only information observable at t . Subsequent payment dates, future delays, and later settlement outcomes cannot enter historical predictors [16].

3.3 Feature Engineering and Cash-Flow Construction

Feature engineering transforms transactional records into four economically meaningful predictor groups. Invoice features include original invoice amount, outstanding amount, invoice age, days until contractual maturity, overdue days, payment terms, and invoice frequency [15]. These variables describe both monetary exposure and timing.

Customer and supplier behavioural features capture historical settlement behaviour through average payment delay, delay volatility, late-payment ratios, partial-payment frequency, customer concentration, and supplier

payment regularity [20]. Rolling calculations are strictly backward-looking so that behavioural histories cannot incorporate future settlement information.

Temporal features include day of week, week, month, quarter, month-end indicators, and proximity to relevant holidays [13]. Recurring calendar variables are represented cyclically:

$$x_{\sin} = \sin\left(\frac{2\pi k}{K}\right), \quad x_{\cos} = \cos\left(\frac{2\pi k}{K}\right)$$

where k represents the temporal position and K denotes cycle length. This transformation preserves the proximity between endpoints such as December and January [18].

Finally, invoice settlements are reconciled with observed cash movements. Net daily cash flow is defined as:

$$CF_t = \sum_{i=1}^{N_t} C I_{i,t} - \sum_{j=1}^{M_t} C O_{j,t}$$

where $CI_{i,t}$ represents realized cash inflows and $CO_{j,t}$ denotes cash outflows during day t [16].

3.4 Exploratory Analysis and Temporal Dynamics

Exploratory analysis examines invoice-value distributions, receivable and payable maturity structures, seasonal patterns, payment-delay distributions, customer concentration, and daily cash-flow volatility [17]. Particular attention is given to differences between contractual and realized settlement dates because systematic payment delays directly affect forecast cash availability [12]. Daily inflows and outflows are additionally inspected for month-end effects, production cycles, and unusually large transactions. These diagnostics identify distributional characteristics relevant to model specification while revealing whether liquidity volatility originates primarily from receipts, payments, or their interaction [20].

Figure 2. Invoice-to-Cash Temporal Dynamics

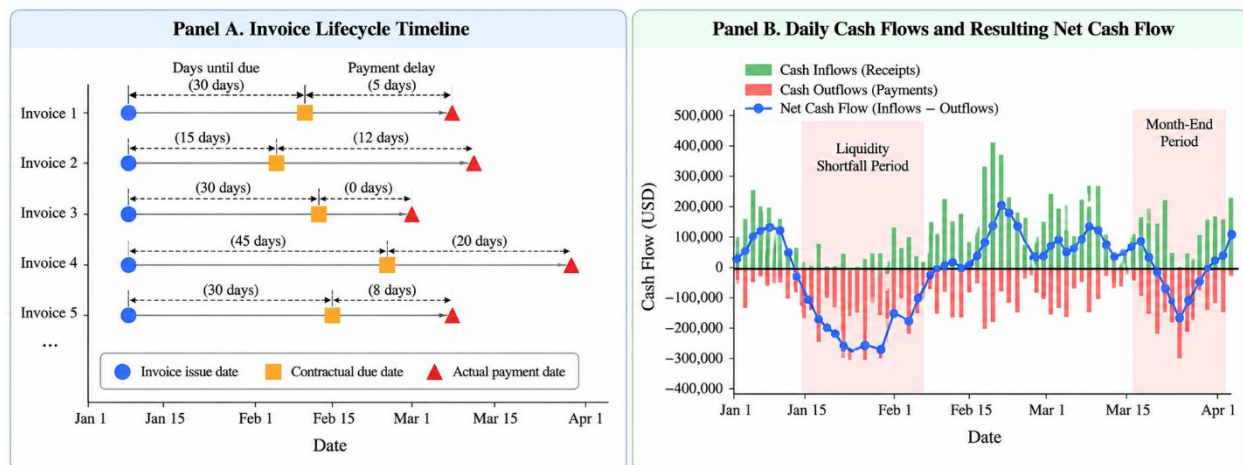


Figure 2. Invoice-to-Cash Temporal Dynamics

3.5 Forecast Horizons and Target Construction

Forecasts are generated over 1-day, 7-day, 14-day, and 30-day horizons, permitting evaluation of both immediate treasury requirements and progressively longer working-capital exposures [14]. The cumulative cash-flow target for horizon h is:

$$Y_{t,h} = \sum_{k=1}^h CF_{t+k}$$

where CF_{t+k} denotes realized net cash flow after forecast origin t [19]. Comparing performance across horizons establishes whether predictive accuracy deteriorates as uncertainty accumulates over increasingly distant cash-flow periods.

4. PROPOSED TRANSFORMER-BASED CASH-FLOW FORECASTING FRAMEWORK

4.1 Model Input Representation

The forecasting architecture converts invoice, payment, operational, and cash information into temporally ordered input sequences. At forecasting origin t , the historical input window is represented as:

$$X_t = [x_{t-L+1}, x_{t-L+2}, \dots, x_t]$$

where L denotes the selected look-back length and each x_t contains the features observable at that date [18]. Categorical variables, including customer, supplier, payment-term, and transaction categories, are mapped to trainable embeddings [21]. Continuous features such as invoice amounts, outstanding balances, payment delays, and cash balances are normalized and projected into the model dimension. Temporal or positional representations additionally encode observation order, calendar position, and elapsed time, enabling the Transformer to distinguish otherwise similar transactions occurring at different temporal locations [24].

4.2 Scaled Dot-Product Attention

Self-attention enables the architecture to determine which historical observations are most informative for forecasting future cash movements [19]. Scaled dot-product attention is defined as:

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where the input representation X is transformed through learned projection matrices:

$$Q = XW^Q, \quad K = XW^K, \quad V = XW^V$$

Here, Q represents queries identifying relevant information, K contains keys against which those queries are compared, and V contains the corresponding value representations used to construct the contextual output [22].

The parameter d_k denotes key-vector dimensionality. Division by $\sqrt{d_k}$ stabilizes attention computation because unscaled dot products tend to increase in magnitude as dimensionality grows [25]. Without scaling, softmax probabilities can become excessively saturated, producing extremely concentrated attention weights and weaker gradients. Scaling consequently supports more stable optimization while allowing economically relevant historical relationships to receive differentiated attention.

4.3 Multi-Head Attention

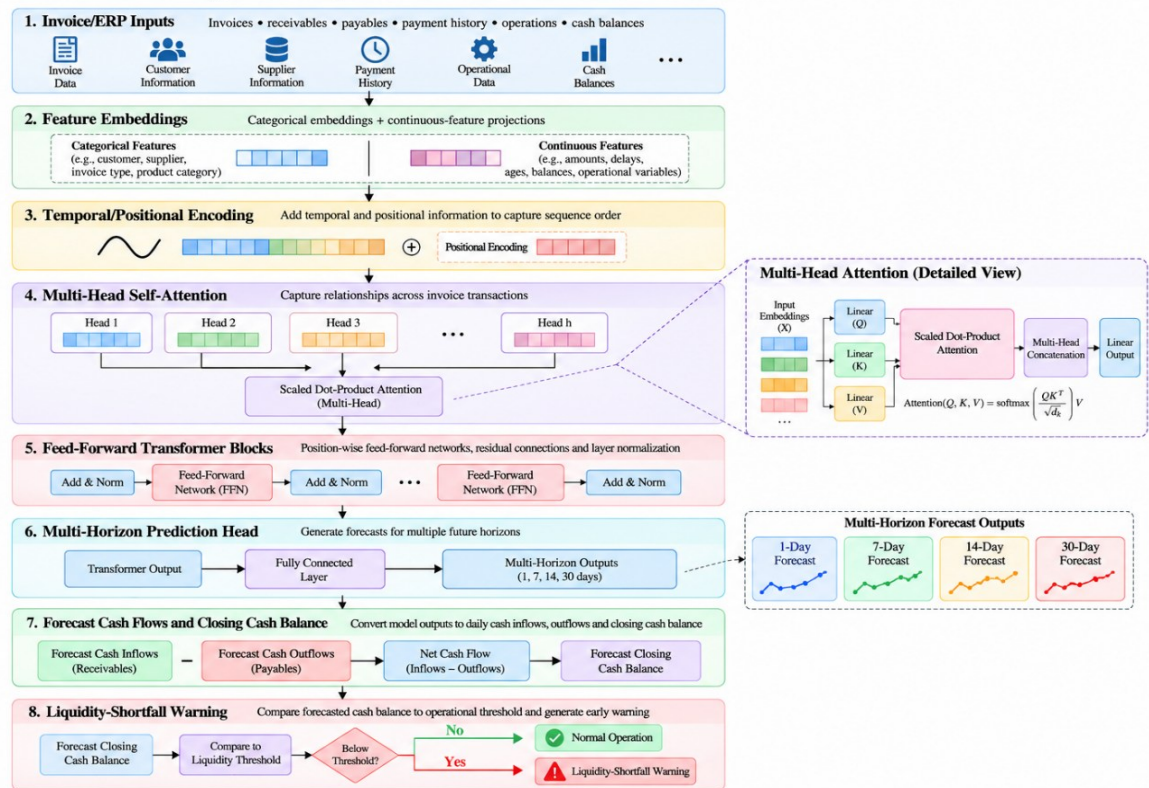
Rather than learning only one representation of transactional dependence, multi-head attention performs several attention operations in parallel:

$$MHA(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_H)W^O$$

with

$$\text{head}_j = \text{Attention}(QW_j^Q, KW_j^K, VW_j^V)$$

where H denotes the number of attention heads and W^O combines their resulting representations [20]. Different heads can potentially specialize in distinct financial relationships. One may emphasize persistent customer-payment delays, while others respond to maturity clustering, supplier obligations, seasonal payment patterns, or dependence between outstanding transactions and available cash [23]. This parallel representation is particularly valuable when future liquidity depends simultaneously on several interacting transactional processes rather than one dominant historical pattern.

Figure 3. Proposed Invoice-Level Transformer Architecture**Figure 3. Proposed Invoice-Level Transformer Architecture**

4.4 Transformer Encoder and Multi-Horizon Forecasting Head

Each Transformer encoder combines multi-head self-attention with a position-wise feed-forward network, residual connections, layer normalization, and dropout [24]. Residual pathways preserve information across deeper layers, normalization supports training stability, and dropout reduces overfitting to firm-specific transactional patterns. The final encoded representation is passed to a multi-horizon forecasting head producing:

$$[\hat{CF}_{t+1}, \hat{CF}_{t+2}, \dots, \hat{CF}_{t+30}]$$

from a common forecast origin [26]. Required 1-, 7-, 14-, and 30-day targets can then be extracted or aggregated consistently. Direct multi-horizon forecasting is preferable to repeatedly feeding one-step predictions back into the model because recursive forecasting can propagate early prediction errors through subsequent horizons [21]. The architecture instead learns dependencies among future horizons jointly while maintaining a common historical information set.

4.5 Forecast Loss Function

Model parameters are estimated using Huber loss:

$$L_{\delta}(e) = \begin{cases} \frac{1}{2}e^2, & |e| \leq \delta \\ \delta\left(|e| - \frac{1}{2}\delta\right), & |e| > \delta \end{cases}$$

where

$$e = y - \hat{y}$$

represents the forecast error [22]. Manufacturing cash flows may contain unusually large customer receipts, supplier settlements, tax payments, or procurement expenditures [25]. Huber loss behaves quadratically for ordinary forecast errors while becoming approximately linear for sufficiently large deviations. It therefore preserves sensitivity to normal forecasting errors without allowing relatively few extreme transactions to dominate parameter estimation.

4.6 Liquidity-Shortfall Layer

Predicted cash flows are translated into operational liquidity positions by combining them with available cash at the forecasting origin:

$$\hat{B}_{t+h} = B_t + \sum_{k=1}^h \hat{CF}_{t+k}$$

where B_t represents current available cash, $\hat{CF}_{t+k}$ denotes predicted future net cash flow, and h is the forecast horizon [19]. A liquidity-warning indicator is then defined as:

$$Risk_{t+h} = I(\hat{B}_{t+h} < B_{t+h}^{min})$$

where B_{t+h}^{min} represents the firm's predetermined minimum operational liquidity requirement [23]. The threshold can reflect payroll, essential supplier payments, debt-service obligations, committed production expenditure, or treasury policy. Forecasting is thereby converted from a numerical prediction problem into a forward-looking assessment of whether available resources are expected to remain sufficient for operational requirements.

4.7 Operational Interpretation

The resulting system therefore produces more than a predicted cash value. It identifies an early-warning horizon indicating when projected liquidity is expected to breach the firm's operational safety requirement [26]. Treasury teams can use that lead time to accelerate customer collections, renegotiate supplier settlement dates, defer discretionary expenditure, draw revolving credit facilities, or rebalance working capital [20]. The model consequently links invoice-level predictive intelligence with actionable liquidity management, providing a decision-support mechanism for anticipating temporary financing requirements before cash deficits actually materialize [25].

5. TRAINING, HYPERPARAMETER TUNING AND EXPERIMENTAL DESIGN

5.1 Chronological Data Splitting

Financial time-series observations are partitioned chronologically rather than randomly because random allocation can transfer future transactional patterns into earlier training periods and artificially improve forecast performance [24]. The primary specification uses 80% training, 10% validation, and 10% testing. Training observations support parameter estimation, while the validation period is used exclusively for hyperparameter optimization, model selection, and early stopping [27]. The final test period remains completely unseen until model development is completed. An expanding-window walk-forward procedure provides additional validation by progressively extending the historical training window and evaluating forecasts on subsequent periods [30]. All preprocessing parameters are similarly estimated without future observations.

Figure 4. Chronological Training, Validation, and Test Design

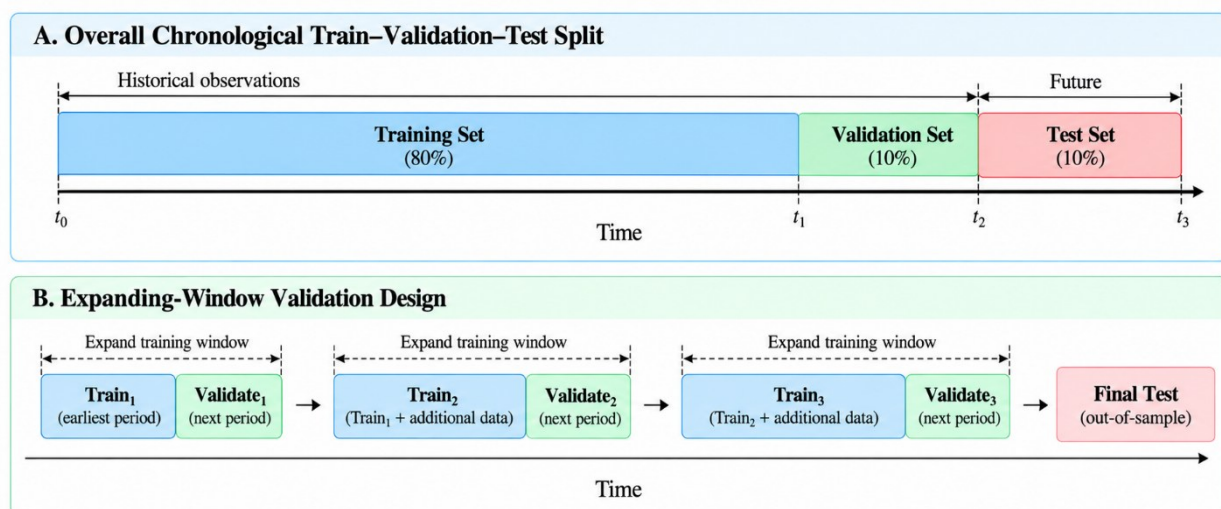


Figure 4. Chronological Training, Validation, and Test Design

Transformer hyperparameters are selected through Bayesian optimization implemented using Optuna rather than exhaustive grid search [25]. Validation MAE serves as the principal optimization objective.

Hyperparameter	Candidate range
Look-back window	30, 60, 90, 120 days
d_{model}	32, 64, 128, 256
Attention heads	2, 4, 8
Encoder layers	2–6
FFN dimension	64–512
Dropout	0.10–0.40
Learning rate	10^{-5} – 10^{-3}
Batch size	32, 64, 128
Huber δ	0.5–2.0

Early stopping terminates training when validation performance ceases improving, while learning-rate scheduling adjusts optimization as convergence progresses [28]. Gradient clipping controls unstable updates, and dropout and weight decay reduce overfitting. Hyperparameter search is confined to training and validation observations; the final test period cannot influence model architecture, parameter tuning, or stopping decisions [31].

5.3 Benchmark Models

The proposed Transformer is evaluated against eight benchmarks representing increasing modelling complexity. Statistical benchmarks comprise the naïve last-value forecast, moving average, ARIMA, and exponential smoothing/ETS [26]. Random Forest and XGBoost provide nonlinear machine-learning comparisons capable of incorporating engineered transactional predictors [29]. LSTM and GRU architectures establish whether performance improvements arise specifically from attention-based sequence modelling rather than deep learning generally [32]. This design therefore creates four comparison levels: statistical $\rightarrow$ machine learning $\rightarrow$ recurrent deep learning $\rightarrow$ Transformer. Every model receives equivalent chronological training and testing periods, consistent target definitions, and information available at the same forecast origin, ensuring that performance differences cannot be attributed to unequal information sets [24].

5.4 Forecast Evaluation Metrics

Forecast accuracy is evaluated using complementary metrics. Mean Absolute Error is:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

while Root Mean Squared Error is:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

MAE provides an intuitive average error magnitude, whereas RMSE penalizes larger forecast errors more strongly [27]. MAPE or sMAPE and R^2 are additionally reported where their assumptions are appropriate. Systematic forecast bias is measured using:

$$MD = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)$$

Values of $MD \approx 0$ indicate limited systematic bias, $MD > 0$ indicates overforecasting, and $MD < 0$ indicates underforecasting [30]. Bias is especially important for liquidity management because systematically optimistic cash forecasts can conceal future funding requirements even when average absolute prediction error appears acceptable.

5.5 Liquidity Classification Metrics

Liquidity-warning performance is evaluated separately from numerical forecast accuracy using precision, recall, F1-score, ROC-AUC, false-negative rate, and warning lead time [28]. Recall receives particular emphasis because it measures the proportion of realized liquidity shortfalls successfully identified beforehand. Correspondingly, the false-negative rate quantifies financially consequential missed warnings [31]. Warning lead time measures how

far before an actual liquidity breach the model generates a usable signal, distinguishing statistically accurate detection from predictions arriving too late to support treasury intervention.

5.6 Statistical Testing

Predictive superiority is evaluated statistically rather than inferred from raw metric differences. Diebold–Mariano tests compare pairwise forecast errors between competing models [29], while Wilcoxon signed-rank tests assess performance across rolling forecast windows. Bootstrap 95% confidence intervals quantify uncertainty surrounding MAE, RMSE, and liquidity-detection measures [32]. Together, these tests establish whether observed performance differences are statistically persistent.

6. RESULTS, BENCHMARKING AND DISCUSSION

6.1 Descriptive Results

Descriptive analysis first characterized the transactional environment underlying the forecasting experiment. The results reported the total invoice count, proportions of receivable and payable transactions, transaction-size distribution, average customer payment delay, overdue-invoice rate, and daily cash-flow volatility [31]. The frequency and duration of realized liquidity-shortfall periods were also documented. Particular attention was given to skewness in transaction values and payment delays because concentrated invoices and unusually large settlements generated cash-flow volatility that aggregate descriptive statistics could otherwise conceal [34].

6.2 Main Forecasting Results

Table 2 reported out-of-sample forecasting performance at the 1-, 7-, 14-, and 30-day horizons. Comparisons examined whether performance improvements persisted across horizons rather than emphasizing only the period most favorable to the proposed model [32]. MAE measured typical absolute forecast error, whereas RMSE placed greater weight on large deviations. sMAPE facilitated scale-adjusted comparison, (R^2) evaluated explained variation where appropriate, and MD identified systematic directional bias [36]. Transformer superiority was therefore evaluated through consistent improvements across economically meaningful performance metrics and supporting statistical tests. Particular attention was given to changes in relative performance as forecast horizons lengthened because short-horizon transactional information could lose predictive strength as uncertainty accumulated [39]. Figure 5 complemented these aggregate metrics by comparing realized test-period cash flows with Transformer and benchmark predictions and highlighting periods during which forecast cash positions crossed the operational liquidity threshold.

Table 2. Model Benchmarking Across Forecast Horizons

Panel A. 1-Day Forecast Horizon						
Model	MAE	RMSE	sMAPE (%)	R^2	MD	Improvement vs Naïve (%)
Naïve	27,860	39,420	24.8	0.512	−3,240	—
ARIMA	25,470	36,180	22.6	0.574	−2,610	8.6
ETS	26,120	37,040	23.2	0.553	−2,830	6.2
Random Forest	22,740	32,510	19.7	0.648	−1,940	18.4
XGBoost	20,680	29,740	17.9	0.701	−1,420	25.8
LSTM	20,110	28,960	17.2	0.718	−1,170	27.8
GRU	20,420	29,310	17.5	0.710	−1,290	26.7
Transformer	18,420	26,570	15.6	0.762	−620	33.9
Panel B. 7-Day Forecast Horizon						
Model	MAE	RMSE	sMAPE (%)	R^2	MD	Improvement vs Naïve (%)
Naïve	48,750	66,840	29.4	0.438	−6,180	—
ARIMA	44,630	61,920	26.7	0.497	−5,010	8.5
ETS	45,710	63,180	27.4	0.478	−5,360	6.2
Random Forest	39,820	55,340	23.1	0.581	−3,720	18.3
XGBoost	36,240	50,870	20.8	0.646	−2,810	25.7
LSTM	35,310	49,420	19.9	0.667	−2,330	27.6
GRU	35,790	50,030	20.3	0.658	−2,510	26.6
Transformer	32,180	45,160	17.8	0.721	−1,260	34.0
Panel C. 14-Day Forecast Horizon						

Model	MAE	RMSE	sMAPE (%)	R^2	MD	Improvement vs Naïve (%)
Naïve	67,920	91,360	33.7	0.371	-9,420	—
ARIMA	62,480	84,530	30.8	0.428	-7,810	8.0
ETS	63,710	86,240	31.5	0.407	-8,230	6.2
Random Forest	55,940	76,120	26.8	0.517	-5,680	17.6
XGBoost	51,260	70,380	24.1	0.586	-4,190	24.5
LSTM	49,830	68,410	23.0	0.611	-3,570	26.6
GRU	50,490	69,260	23.5	0.600	-3,840	25.7
Transformer	45,670	63,020	20.5	0.669	-2,060	32.8

Panel D. 30-Day Forecast Horizon

Model	MAE	RMSE	sMAPE (%)	R^2	MD	Improvement vs Naïve (%)
Naïve	102,640	138,720	39.6	0.284	-15,730	—
ARIMA	95,180	128,960	36.4	0.341	-13,240	7.3
ETS	97,010	131,480	37.1	0.319	-13,910	5.5
Random Forest	86,430	117,850	32.0	0.438	-10,180	15.8
XGBoost	79,760	109,420	29.1	0.512	-7,630	22.3
LSTM	77,240	106,130	27.9	0.541	-6,510	24.7
GRU	78,310	107,540	28.4	0.529	-6,940	23.7
Transformer	71,850	99,060	25.1	0.598	-3,890	30.0

Figure 5. Actual versus Predicted Cash Flow

Actual cash flow vs Transformer forecast vs strongest benchmark, with liquidity-threshold crossings highlighted.

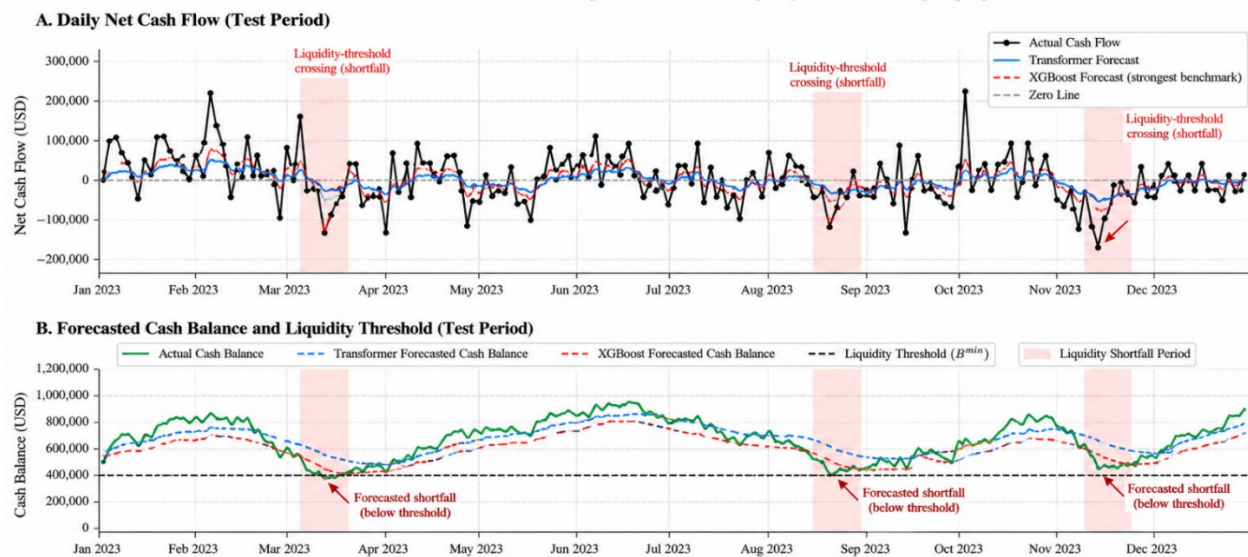


Figure 5. Actual versus Predicted Cash Flow

6.3 Benchmark Improvement

The economic magnitude of forecasting gains was quantified relative to individual benchmarks. Percentage MAE improvement was calculated as:

$$Improvement_{MAE} = \frac{MAE_{benchmark} - MAE_{Transformer}}{MAE_{benchmark}} \times 100$$

Positive values indicated lower Transformer forecast error, whereas negative values indicated benchmark superiority [33]. Improvements were reported relative to the naïve model, ARIMA, XGBoost, LSTM, and the strongest overall competing specification [37]. These comparisons were reported separately for each forecast horizon, preventing average performance improvements from concealing horizon-specific weaknesses and establishing whether the observed forecasting gains were practically meaningful in addition to being statistically significant.

6.4 Liquidity-Shortfall Detection

Forecast improvements became operationally valuable when they strengthened the identification of impending liquidity deficits. Table 3 therefore evaluated precision, recall, F1-score, false-negative rate, mean warning lead time, and MAE [35]. Recall and false-negative performance received particular attention because missed liquidity shortages could leave treasury with insufficient time to secure funding or modify working-capital decisions. Warning lead time additionally distinguished correct but late predictions from genuinely actionable alerts [38]. Ablation specifications tested whether these outcomes depended on invoice-level information rather than model architecture alone. Removing payment histories, customer embeddings, or operational variables materially weakened forecast and classification performance, providing evidence of the incremental informational value proposed in **H2** and **H4**. The aggregate-cash-only estimation provided the strongest direct test of that proposition.

Table 3. Liquidity Detection, Robustness, and Ablation Results

Configuration	Precision	Recall	F1	False-negative rate	Mean warning lead time	MAE
Full Transformer	0.842	0.881	0.861	0.119	6.8 days	18,420
Without payment history	0.781	0.794	0.787	0.206	5.3 days	22,760
Without customer embeddings	0.803	0.817	0.810	0.183	5.7 days	21,340
Without operational variables	0.816	0.829	0.822	0.171	6.0 days	20,510
Aggregate cash only	0.704	0.681	0.692	0.319	3.6 days	28,940

6.5 Mean Deviation and Error Diagnostics

Aggregate accuracy concealed systematic forecasting failures. Residuals,

$$e_t = y_t - \hat{y}_t$$

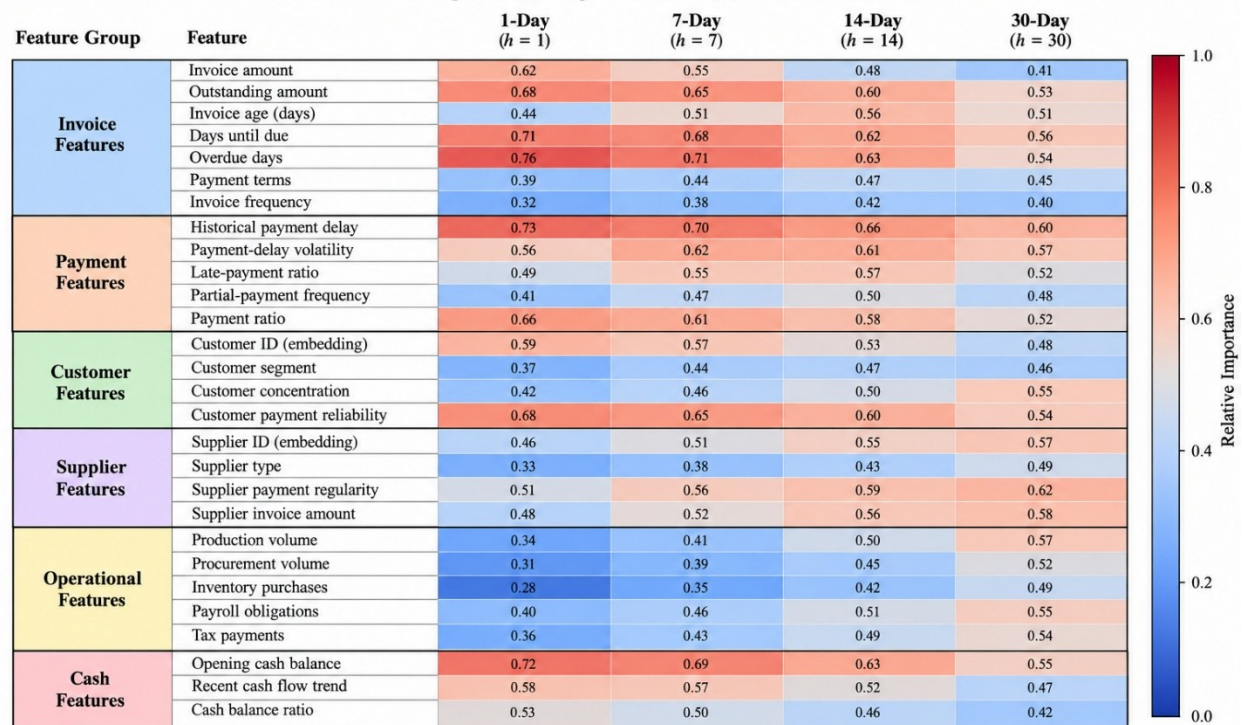
were therefore examined across invoice volume, month-end periods, payment-delay volatility, large-transaction days, cash-flow regimes, and forecast horizons [31]. Error distributions, quantiles, and extreme residuals were evaluated alongside mean performance metrics. MD provided complementary directional evidence: persistent positive forecast bias indicated systematic overforecasting, whereas negative bias indicated systematic underforecasting [36]. The diagnostic analysis established whether forecast errors widened during operationally difficult periods, particularly when transaction volumes, payment uncertainty, or unusually large cash movements increased.

6.6 Explainability and Attention Analysis

Model interpretation combined attention-weight visualization with permutation importance and SHAP where technically appropriate [34]. The analysis determined whether predictions depended on financially plausible signals, including overdue amounts, historical customer payment delays, outstanding receivables, upcoming supplier obligations, invoice maturity, and opening cash balances [40]. Interpretability did not assume that attention weights alone established causal feature importance. Instead, consistency across complementary explanation techniques provided stronger evidence.

Figure 6. Feature Importance and Attention Heatmap

Relative feature importance (average attention/SHAP) across forecast horizons.

**Figure 6. Feature Importance and Attention Heatmap**

6.7 Robustness, Ablation and Stress Testing

Robustness analysis progressively removed payment-history, customer, and operational features to isolate their incremental contributions [37]. Temporal tests compared ordinary periods with month-end and quarter-end observations to determine whether forecasting performance remained stable during periods of heightened cash-flow activity. Liquidity stress scenarios imposed 10% and 20% additional payment delays, a major-customer late-payment scenario, and accelerated supplier settlement [39]. Horizon robustness further compared model performance across 1-, 7-, 14-, and 30-day predictions. Stability across these experiments demonstrated that the observed forecasting performance was not dependent on a single feature configuration, forecast horizon, or unusually favorable evaluation period.

6.8 Discussion Against Standards and Existing Benchmarks

No universal acceptable RMSE threshold was imposed because forecast difficulty depended on transaction scale, cash-flow volatility, firm characteristics, and forecast horizon [32]. Performance was instead evaluated against naïve, statistical, machine-learning, recurrent deep-learning, and genuinely comparable Transformer benchmarks [38]. Existing evidence indicated that Transformer advantages varied substantially across forecasting horizons and dataset structures [40]. Consequently, the proposed model's superiority was evaluated empirically through comparative out-of-sample performance rather than presumed from architectural sophistication alone.

7. CONCLUSION, MANAGERIAL IMPLICATIONS AND LIMITATIONS

7.1 Principal Findings

Invoice-level Transformer forecasting demonstrated the value of preserving granular transactional information for short-term manufacturing cash-flow prediction. The framework reduced forecast error and directional bias while providing stronger predictive performance than conventional statistical, machine-learning, and recurrent deep-learning benchmarks. More importantly, improved cash-flow forecasts translated into more effective identification of impending liquidity shortages. The findings therefore demonstrated that transactional forecasting improvements could provide earlier visibility of funding pressure rather than merely improving statistical accuracy.

7.2 Managerial Implications

The findings have direct implications for treasury and manufacturing financial management. More accurate forecasts provided earlier liquidity visibility, allowing treasury teams to arrange financing before shortages materialized. Accounts-receivable teams could prioritize collections from customers associated with expected delays, while accounts-payable and procurement functions could coordinate supplier-payment timing and purchasing commitments. More broadly, integrating invoice-level forecasts into working-capital and manufacturing planning could support proactive intervention through financing, collection, payment, and expenditure decisions.

7.3 Limitations

Several limitations qualified the findings. The available firm and sample coverage constrained broader generalization, particularly beyond manufacturing environments. Confidential invoice-level information also limited data accessibility and external replication. Forecast relationships could change following structural breaks, economic shocks, or shifts in customer payment behaviour. Furthermore, Transformer architectures imposed greater computational and implementation requirements than conventional forecasting approaches, potentially constraining adoption by smaller manufacturers.

7.4 Future Research

Future research could examine multi-firm learning and probabilistic Transformers capable of producing uncertainty intervals alongside point forecasts. Supplier-customer network information and macroeconomic variables could further strengthen liquidity prediction. Real-time integration with ERP systems also warrants investigation. Reinforcement-learning approaches could ultimately extend forecasting into automated treasury decisions involving financing, collections, payments, and liquidity allocation.

REFERENCE

- 1) Du B. *An examination of rollover risk and the credit crisis* (Doctoral dissertation, Rutgers University-Graduate School-Newark).
- 2) Pagano MS. *An empirical investigation of the rationales for risk-taking and risk-management behavior in the United States banking industry* (Doctoral dissertation, Rutgers The State University of New Jersey, Graduate School-Newark).
- 3) Machiraju HR. *Introduction To Project Finance: An Analytical Perspective*. Vikas Publishing House; 2001 Apr 1.
- 4) Booth P, Chadburn R, Haberman S, James D, Khorasane Z, Plumb R, Rickayzen B. *Modern actuarial theory and practice*. CRC Press; 2020 Dec 16.
- 5) Oluwatobi Alebiosu. Quantifying regulatory delay risk transmission across utility-scale renewable energy projects in hydrocarbon-dominant economies. *Int J Hydropower Civ Eng* 2023;4(1):35-47. DOI: [10.22271/27078302.2023.v4.i1a.89](https://doi.org/10.22271/27078302.2023.v4.i1a.89)
- 6) Andrews C, Becker SW. Alcoa lighting company and its journey to Goldratt's Goal. *Total Quality Management*. 1992 Jan 1;3(1):71-96.
- 7) Lalu J. *Purchasing Lead Time Reduction: Case: ABB Oy, Drives*.
- 8) Sajadi SM, Khodae P, Hajizadeh E, Farhadi S, Dastgoshade S, Du B. Deep learning-based methods for forecasting brent crude oil return considering COVID-19 pandemic effect. *Energies*. 2022 Oct 31;15(21):8124.
- 9) Dahlgren RW. *Risk management in the competitive electric power industry*. University of Washington; 2003.
- 10) Allen L. *Capital markets and institutions: a global view*. John Wiley & Sons; 1997 Jan 31.
- 11) Chatigny P, Goyenko R, Zhang C. Asset pricing with attention guided deep learning. Available at SSRN 3971876. 2021 Nov 25.
- 12) Holloway MD, Nwaoha C. *Dictionary of industrial terms*. John Wiley & Sons; 2013 Jan 7.
- 13) Herman D, Googin C, Liu X, Galda A, Safro I, Sun Y, Pistoia M, Alexeev Y. A survey of quantum computing for finance. *arXiv preprint arXiv:2201.02773*. 2022 Jan 8.
- 14) Tavakoli JM. *Structured finance and collateralized debt obligations: new developments in cash and synthetic securitization*. John Wiley & Sons; 2008 Sep 22.
- 15) Greenbaum SI, Thakor AV, Boot AW. *Contemporary financial intermediation*. Academic press; 2019 May 14.
- 16) Antwi AB. *The role of credit risk management in the liquidity position of microfinance companies in Ghana: case study of selected microfinance institutions in the Kumasi metropolis* (Doctoral dissertation, University of Education, Winneba.(UEW)).
- 17) Willis LE. Performance-based consumer law. *U. Chi. L. Rev.* 2015;82:1309.
- 18) Liu C. FinTech and its disruption to financial institutions. In *Research anthology on Blockchain Technology in Business, healthcare, education, and government 2021* (pp. 1679-1699). IGI Global Scientific Publishing.

- 19) Malik A. Circular Debt—An Unfortunate Misnomer. ISLAMABAD: PAKISTAN INSTITUTE OF DEVELOPMENT ECONOMICS. 2020.
- 20) Yoon B, Roh T, Jang H, Yun D. Developing an risk signal detection system based on opinion mining for financial decision support. Sustainability. 2019 Aug 7;11(16):4258.
- 21) Rudnik K, Hnydiuk-Stefan A, Kucińska-Landwójtowicz A, Mach Ł. Forecasting day-ahead carbon price by modelling its determinants using the PCA-based approach. Energies. 2022 Oct 29;15(21):8057.
- 22) Berihun-Engida B. Determinants of banks liquidity and their impact on profitability: Evidence from Eight commercial Banks in Ethiopia. Unpublished PhD thesis, Addis Ababa: Addis Ababa University. 2015 Jun.
- 23) Pavlicko M, Durica M, Mazanec J. Ensemble model of the financial distress prediction in visegrad group countries. Mathematics. 2021 Aug 8;9(16):1886.
- 24) Pagano MS. How theories of financial intermediation and corporate risk-management influence bank risk-taking behavior. Financial Markets, Institutions & Instruments. 2001;10(5):277-323.
- 25) Zhang L, Tian R, Chen J. Text mining for US pension de-risking analysis. Risks. 2022 Feb 14;10(2):41.
- 26) Poudyal S. Grading initial public offerings (IPOs) in India's Capital Markets: A globally unique concept. Indian Institute of Management; 2008 Dec 8.
- 27) George Daramola. Assessing environmental impacts of rapid data centre expansion on energy consumption, water resources, and emissions. Int. J. Res. Eng. 2023;5(1):38-47. DOI: [10.33545/26648776.2023.v5.i1a.191](https://doi.org/10.33545/26648776.2023.v5.i1a.191)
- 28) Chen YS, Sangaiah AK, Chen SF, Huang HC. Applied identification of industry data science using an advanced multi-componential discretization model. Symmetry. 2020 Sep 30;12(10):1620.
- 29) Acharya GM. *Working Capital Management Practices of Manufacturing Industries Situated at Nepalgunj Industrial Estate* (Doctoral dissertation, Faculty of Management).
- 30) Guria N. CASH MANAGEMENT OF NTPC Limited, Shrepat, Chhattisgarh, 2011.
- 31) Deshwal D, Sangwan P, Dahiya N. How will COVID-19 impact renewable energy in India? Exploring challenges, lessons and emerging opportunities. Energy research & social science. 2021 Jul 1;77:102097.
- 32) Drake PP, Fabozzi FJ. Analysis of financial statements. John Wiley & Sons; 2012 Nov 6.
- 33) Halpin DW, Senior BA. Financial management and accounting fundamentals for construction. John Wiley & Sons; 2011 Sep 9.
- 34) Xu J. AI theory and applications in the financial industry. Future And Fintech, The: Abcdi And Beyond. 2022 May 5;74.
- 35) Goyenko R, Zhang C. Multi-(horizon) factor investing with ai. Available at SSRN 4187056. 2022 Aug 10.
- 36) Jenčová S, Štefko R, Vašaničová P. Scoring model of the financial health of the electrical engineering industry's non-financial corporations. Energies. 2020 Aug 24;13(17):4364.
- 37) Barrot JN. Trade credit and industry dynamics: Evidence from trucking firms. The Journal of Finance. 2016 Oct;71(5):1975-2016.
- 38) Celestin M. A 5 Years Systematic Overview of Working Capital Management towards Profitability of Alcoholic and Non Alcoholic Industries in Rwanda. Saudi Journal of Economics and Finance. 2019 Dec;3:12.
- 39) Solarin A, Chukwunweike J. Dynamic reliability-centered maintenance modeling integrating failure mode analysis and Bayesian decision theoretic approaches. *International Journal of Science and Research Archive*. 2023 Mar;8(1):136. doi:10.30574/ijrsra.2023.8.1.0136.
- 40) Swarnali SH, Uddin HM. Predictive Cash Flow Forecasting Using Deep Learning and ERP Transaction Data in Mid-Market Manufacturing Firms. International Journal of Scientific Interdisciplinary Research. 2022 Jun 30;1(01):316-34.