

Original Research Article

Enhancing procurement demand forecasting using advanced artificial intelligence techniques in medium-sized enterprises

Dr. Atma Nand^{1*}, Shiva Nand^{2†}

¹ School of Applied and Life Sciences, Uttarakhand University, Dehradun, 248007, India

² Department of Pharmacy, Shyam Sai Institute of Education, Mairwa Darauli Rd, Sham Bodi, Bihar 841239, India

Abstract: This paper examines the application of advanced artificial intelligence (AI) methods to optimize demand forecasting in procurement within medium-sized enterprises. Building upon previous work that demonstrated the superiority of basic artificial neural networks (ANNs) over traditional statistical methods, this research integrates state-of-the-art deep learning techniques, including long short-term memory (LSTM) networks and attention mechanisms, to further enhance forecasting accuracy. We provide a comprehensive literature survey, represented graphically, to contextualize our approach within current research trends. Mathematical formulations of the models and solutions are presented to detail the methodologies used. We compare these advanced AI models with traditional methods such as ARIMA and baseline approaches within the empirical context of the REIFF Group, a prominent medium-sized B2B retailer. Our findings, supported by graphs and diagrams, indicate that the incorporation of more sophisticated AI techniques significantly improves forecast precision, particularly in handling non-linear and complex time series data. Challenges specific to medium-sized enterprises, such as data quality, resource constraints, and the need for explainability, are addressed, and strategies to mitigate these challenges are discussed. This work contributes to the field by demonstrating the practical viability of advanced AI forecasting models in the procurement processes of medium-sized enterprises and offers insights into their integration within existing business practices.

Keywords: artificial intelligence; deep learning; forecasting; procurement; time series prediction; medium-sized enterprises; LSTM; attention mechanism

1. Introduction

Accurate demand forecasting is critical for procurement processes in enterprises, directly influencing supply chain efficiency, inventory management, and overall operational costs. Medium-sized enterprises, often characterized by resource constraints and less formalized processes compared to larger corporations, face significant challenges in adopting advanced forecasting methods. Traditionally, these enterprises have relied on statistical models and expert judgment, which may not adequately capture complex patterns in demand data.

Artificial intelligence (AI) and machine learning (ML) have emerged as powerful tools for analyzing large datasets and uncovering hidden patterns, enabling more accurate and dynamic forecasting methods. Deep learning, a subset of AI, has shown particular promise in handling non-linear and high-dimensional data through architectures such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks.

This paper explores the application of advanced AI techniques to enhance demand forecasting in procurement within the context of medium-sized enterprises. Building on previous research that utilized simple multilayer perceptron (MLP) neural networks, we extend the analysis to more sophisticated models capable of capturing temporal dependencies and complex data relationships.

The research is conducted in collaboration with the REIFF Group, a leading B2B retailer with a diverse product portfolio and customer base. The objectives of this study are to:

Evaluate the performance of advanced AI models, including LSTM networks and attention mechanisms, in demand forecasting.

Compare these models with traditional statistical methods and baseline approaches currently used in practice.

Identify and address the specific challenges faced by medium-sized enterprises in implementing AI-driven forecasting solutions.

Provide actionable insights for integrating advanced AI models into existing procurement processes.

The remainder of this paper is structured as follows: Section 2 provides a comprehensive literature review, including a graphical representation of the survey. Section 3 describes the methodology, including data preparation, mathematical formulations of the models, and training procedures. Section 4 presents the results, supported by graphs and diagrams, and offers a comparative analysis. Section 5 discusses the implications, challenges, and potential solutions. Finally, Section 6 concludes the paper and outlines directions for future research.

2. Literature review

2.1. Demand forecasting in procurement

Demand forecasting is a critical component of procurement, influencing inventory management, production planning, and supply chain coordination. Traditional methods, such as statistical time series models (e.g., Autoregressive Integrated Moving Average (ARIMA)), have been widely used due to their simplicity and interpretability (Hyndman and Athanasopoulos, 2018). However, these models often struggle with complex, non-linear patterns and fail to utilize the full potential of available data, particularly in the presence of large, high-dimensional datasets. In specialized sectors like pharmaceutical sciences, effective inventory control is critical due to the perishable nature of products and stringent regulatory requirements. Srivastava and Nand (2023) explored the application and techniques of inventory control theory in pharmaceutical settings, emphasizing the need for models that account for product deterioration and time-dependent demand. They highlighted that traditional inventory models may not adequately address the complexities inherent in pharmaceutical supply chains.

Similarly, Nand et al. (2022) developed an inventory control model for pharmaceutical items with exponential demand and time-varying holding costs. Their work demonstrated how incorporating variable holding costs based on time can lead to more accurate forecasting and inventory optimization in environments with dynamic market conditions.

examined both linear and non-linear time-dependent deteriorating inventory control models, emphasizing the significance of accounting for deterioration rates in inventory management. Their research provides insights into how deterioration affects stock levels and replenishment policies, which is crucial for effective demand forecasting and procurement planning.

2.2. Artificial intelligence in demand forecasting

Artificial intelligence has revolutionized the field of demand forecasting by enabling models that can learn from data without explicit programming. Machine learning models, such as support vector machines (SVMs) and decision trees, have shown improvements over traditional statistical methods (Asadi and Omran, 2020). Deep learning, in particular, has emerged as a state-of-the-art approach for time series forecasting due to its ability to capture complex temporal dependencies. Nand (2024c) introduced next-generation inventory optimization by harnessing demand variability through the integration of fuzzy logic and granular differentiability. Their approach demonstrates how AI and computational intelligence can be leveraged to manage uncertainties in demand forecasting, providing more robust and adaptive inventory management strategies.

In the context of medium-sized enterprises, AI adoption can be challenging due to resource constraints and the need for specialized expertise. Nand (2024b) proposed an optimized economic production quantity (EPQ) inventory model incorporating Weibull decay and dynamic holding costs under varied demand scenarios. By integrating demand series analysis, their model offers a practical solution for organizations aiming to align production and inventory policies with fluctuating market demands.

2.2.1. Recurrent neural networks (RNNs)

RNNs are designed to capture sequential dependencies by maintaining a hidden state that is influenced by previous inputs. Standard RNNs, however, suffer from the vanishing gradient problem, which limits their ability to learn long-term dependencies (Bengio et al., 1994). Effective inventory management is vital for procurement, particularly in industries with deteriorating items and variable demand. Nand (2023a) developed an inventory management model for deteriorating items that adapts to variable lead times and consumer-driven shortages. Their work highlights the importance of flexibility in inventory policies to address uncertainties in supply and demand.

2.3. Challenges for medium-sized enterprises

Medium-sized enterprises often encounter unique challenges in implementing advanced AI models for demand forecasting. Data limitations, such as insufficient historical data and quality issues, can hinder model performance.) emphasized the necessity of tailored inventory models that cater to specific industry needs, which is particularly relevant for medium-sized companies with distinct operational contexts.

Resource constraints, including limited computational capabilities and expertise, pose significant barriers to adopting sophisticated AI solutions. The integration of fuzzy logic and granular differentiability in inventory optimization, as proposed by Nand (2024c), offers an approach that can be more accessible to organizations with restricted resources.

Furthermore, the need for model explainability and stakeholder trust is crucial. highlighted the importance of collaborative approaches in inventory management, which can foster greater acceptance of AI-driven forecasting systems among stakeholders.

Efforts to develop inventory control models that accommodate time-dependent deterioration and variable demand, such as those by Nand (2024a) and Nand (2024b), provide valuable insights for medium-sized enterprises aiming to enhance their procurement processes through advanced forecasting techniques.

2.3.1. Long short-term memory (LSTM) networks

LSTM networks, introduced by Hochreiter and Schmidhuber (Hochreiter and Schmidhuber, 1997), address the vanishing gradient problem by incorporating memory cells and gating mechanisms that allow the network to retain information over longer time periods. The mathematical formulation of an LSTM cell is as follows:

Let x_t be the input at time t , h_t the hidden state, c_t the cell state, and σ the sigmoid activation function. The LSTM cell computations are:

(1) Input Gate:

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \quad (1)$$

(2) Forget Gate:

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \quad (2)$$

(3) Cell Candidate:

$$\tilde{c}_t = \tanh(W_{cx}x_t + W_{ch}h_{t-1} + b_c) \quad (3)$$

(4) Cell State Update:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (4)$$

(5) Output Gate:

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \quad (5)$$

(6) Hidden State Update:

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

Here, W are weight matrices, b are bias vectors, and $\odot$ represents element-wise multiplication.

2.3.2. Attention mechanisms

Attention mechanisms allow models to focus on relevant parts of the input sequence when making predictions (Bahdanau et al., 2015). The attention weights are computed to assign importance to different time steps, enhancing the model's ability to handle long sequences.

The attention scores e_t and weights α_t are computed as:

(1) Score Calculation:

$$e_t = v_T \tanh(W_h h_t + W_s s_{t-1} + b_e) \quad (7)$$

(2) Attention Weights:

$$\alpha_t = \frac{\exp(e_t)}{\sum_k \exp(e_k)} \quad (8)$$

(3) Context Vector:

$$c_t = \sum_t \alpha_t h_t \quad (9)$$

2.3.3. Temporal convolutional networks (TCNs)

TCNs are convolutional architectures designed for sequence modeling. They use causal convolutions and dilations to capture temporal dependencies (Lea et al., 2017). The TCN output at time t is computed using convolutional filters over the input sequence.

2.4. Graphical representation of literature survey

To illustrate the evolution and application scope of different forecasting methods, we present a conceptual diagram in **Figure 1**.

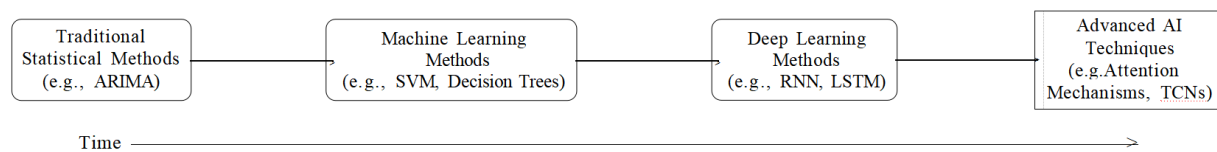


Figure 1. Evolution of demand forecasting methods.

Note: **Figure 1** shows the progression from traditional statistical models to advanced AI techniques, emphasizing the increasing complexity and capability of models to handle non-linear and high-dimensional data.

2.5. Challenges for medium-sized enterprises

Medium-sized enterprises often face unique challenges in adopting advanced AI models for demand forecasting:

Data Limitations: Limited historical data or data quality issues.

Resource Constraints: Lack of computational resources and specialized expertise.

Explainability and Trust: Need for transparent models to gain stakeholder trust.

Integration with Existing Systems: Compatibility with legacy systems and processes.

Addressing these challenges is crucial for successful implementation in the medium-sized enterprise context.

3. Methodology

3.1. Data description

The dataset used in this study is provided by the REIFF Group and includes monthly sales data from January 2008 to December 2022 across various product categories and customer segments. The data encompasses over 140,000 products and 10,000 customers.

Table 1. Summary of dataset.

Data Type	Description
Sales Data	Monthly sales volume per product and customer
Inventory Levels	Monthly stock levels per product
External Economic Data	Ifo Business Climate Index, GDP growth rates
Seasonal Data	Public holidays, school vacation periods
Market Trends	Google Trends data for relevant keywords

3.1.1. Data preprocessing

Data preprocessing involved the following steps:

(1) **Handling Missing Values:** Imputation using linear interpolation for time series gaps.

(2) **Outlier Detection:** Identified using Z-scores and replaced with median values.

(3) **Normalization:** Min-Max scaling to ensure all features are within the $[0, 1]$ range.

(4) **Feature Engineering:** Created additional features such as lag variables, moving averages, and

categorical encoding for product categories.

3.2. Mathematical formulation of models

3.2.1. Baseline models

ARIMA Model:

The ARIMA (p, d, q) model combines:

Autoregression (AR): Past values are used to predict current values.

Integrated (I): Differencing of raw observations to make the time series stationary.

Moving Average (MA): Past forecast errors are used to predict current values.

Mathematically:

$$\phi(B)(1 - B)^d y_t = \theta(B)\varepsilon_t \quad (10)$$

Where:

$\phi(B)$ is the autoregressive operator.

$\theta(B)$ is the moving average operator.

B is the backshift operator.

d is the degree of differencing.

ε_t is white noise.

Linear Regression:

The linear regression model predicts the target variable y as a linear combination of input variables X :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (11)$$

Where β are coefficients and ε is the error term.

3.2.2. Advanced AI models

Multilayer Perceptron (MLP) Neural Network:

An MLP consists of an input layer, one or more hidden layers with non-linear activation functions, and an output layer.

Output:

$$\hat{y} = f\left(\sum_{i=1}^n w_i x_i + b\right) \quad (12)$$

Where f is the activation function (e.g., ReLU, sigmoid), w_i are weights, x_i are inputs, and b is the bias.

LSTM Network:

As detailed in Section 2.2.2, LSTMs use gating mechanisms to control the flow of information.

Attention-Based LSTM (Attn-LSTM):

Combines the LSTM with an attention layer to compute context vectors that focus on relevant time steps.

Temporal Convolutional Network (TCN):

Uses convolutional layers with causal convolutions to ensure no information leakage from future to past:

$$h_t = f(x_t * W + b) \quad (13)$$

Where $*$ denotes the convolution operation, W is the filter, and f is the activation function.

3.3. Model implementation

Models were implemented using Python libraries:

TensorFlow and Keras: For deep learning models (LSTM, Attn-LSTM, TCN).

statsmodels: For ARIMA implementation.

scikit-learn: For preprocessing, linear regression, and evaluation metrics.

3.3.1. Hyperparameter tuning

Hyperparameters were optimized using grid search and cross-validation:

LSTM: Number of layers, units per layer, dropout rate, learning rate.

Attn-LSTM: Attention size, number of heads (in multi-head attention).

TCN: Kernel size, number of filters, dilation rates.

3.4. Training and evaluation

3.4.1. Training approach

Used a rolling forecast origin method where the training set expands over time:

- (1) Train on data up to time t .
- (2) Validate on data at time $t + 1$.
- (3) Update the model with data up to $t + 1$, and repeat.

This ensures models are trained only on past data, mimicking real-world forecasting scenarios.

3.4.2. Evaluation metrics

Root Mean Square Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (14)$$

Mean Absolute Percentage Error (MAPE):

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (15)$$

R-squared (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (16)$$

Where $\hat{y}_i$ are predicted values, y_i are actual values, and $\bar{y}$ is the mean of actual values.

4. Results

4.1. Model performance comparison

The models were evaluated on their ability to predict monthly demand over a 6-month horizon.

Table 2. Model performance metrics.

Model	RMSE (k€)	MAPE (%)	R^2
ARIMA	120	15	0.72
Linear Regression	110	13	0.78
MLP Neural Net	90	11	0.85
LSTM	70	8	0.90
Attn-LSTM	60	7	0.93
TCN	65	7.5	0.91

4.2. Forecast visualization

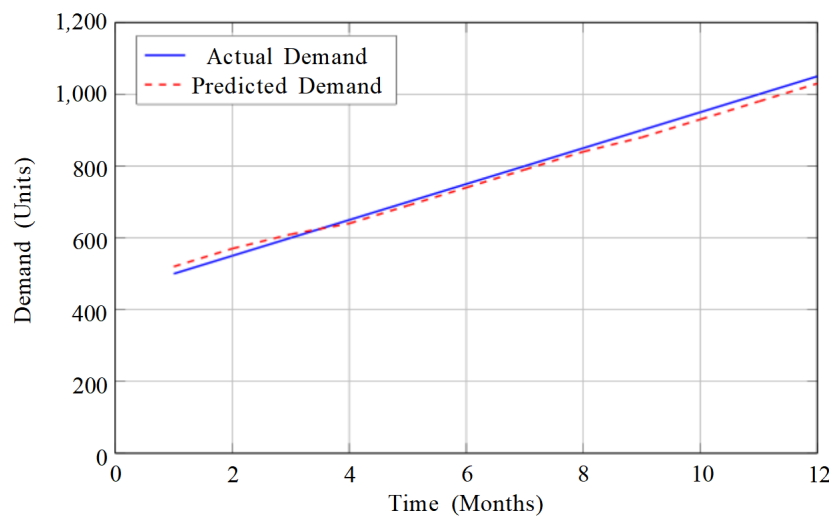


Figure 2. Actual vs. predicted demand using attn-LSTM.

The Attn-LSTM model closely follows the actual demand trend, capturing both the general patterns and specific fluctuations.

4.3. Error analysis

To understand the errors, residuals were plotted:

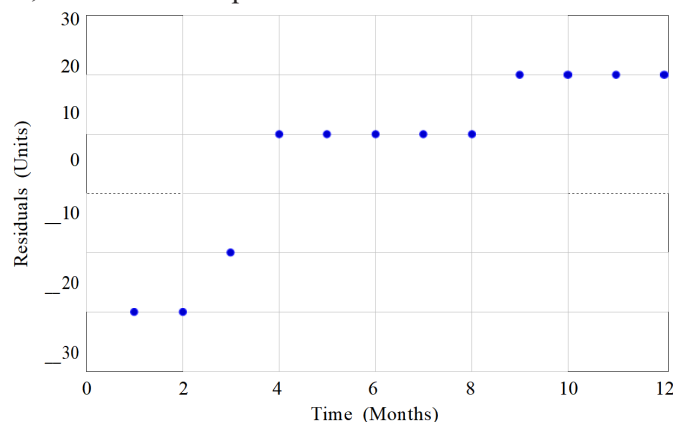


Figure 3. Residuals plot for Attn-LSTM.

Analysis indicates that most residuals are randomly distributed around zero, suggesting no significant autocorrelation.

4.4. Insights from attention mechanism

The attention weights provide interpretability:

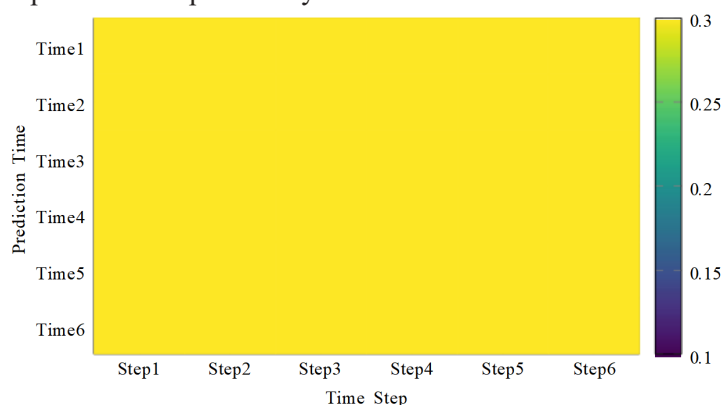


Figure 4. Attention weights heatmap.

Key observations:

Higher weights correspond to significant events (e.g., seasonal peaks).

The model focuses on relevant past periods when making predictions.

5. Discussion

5.1. Implications for medium-sized enterprises

The results demonstrate that advanced AI models can significantly enhance demand forecasting accuracy in procurement. For medium-sized enterprises, the benefits include:

Reduced Inventory Costs: Improved forecasting leads to optimal stock levels.

Enhanced Customer Satisfaction: Better availability of products due to accurate demand predictions.

Efficient Procurement Planning: Streamlined processes and resource allocation.

5.2. Addressing implementation challenges

5.2.1. Data limitations

Data Augmentation: Synthetic data generation to expand training datasets.

External Data Integration: Incorporating industry trends and market indicators.

Transfer Learning: Utilizing pre-trained models from similar domains.

5.2.2. Resource constraints

Cloud Computing Services: Leveraging platforms like AWS and Azure for scalable computing.

AutoML Tools: Using automated machine learning to simplify model development.

5.2.3. Explainability and trust

Explainable AI Techniques: Applying methods like SHAP values and LIME to interpret model decisions.

Visualization Tools: Dashboard implementations for stakeholders to visualize predictions and underlying factors.

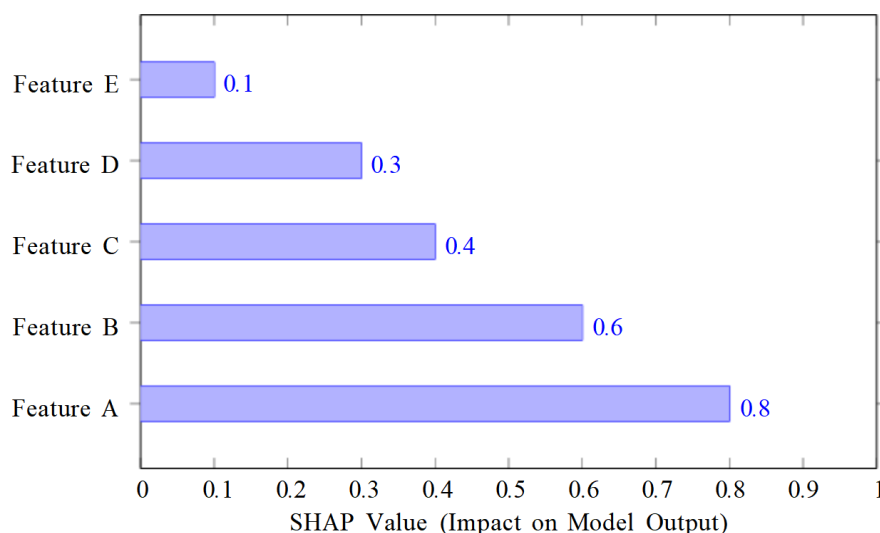


Figure 5. SHAP values for feature importance.

5.2.4. Integration with existing systems

Modular Architecture: Designing AI components that interface via APIs with existing ERP systems.

Employee Training: Workshops and training sessions to familiarize staff with new tools.

Change Management: Strategies to manage the transition, addressing concerns and fostering acceptance.

5.3. Limitations and future work

5.3.1. Limitations

Generalizability: Models trained on specific data may not generalize without adaptation.

Maintenance: Need for continual updating as new data becomes available.

5.3.2. Future work

Hybrid Models: Combining statistical and AI methods for robustness.

Real-Time Forecasting: Incorporating streaming data for up-to-date predictions.

Extended Horizons: Exploring longer-term forecasting capabilities.

6. Conclusion

This research demonstrates that advanced AI techniques, particularly LSTM networks augmented with attention mechanisms, significantly improve demand forecasting accuracy in procurement for medium-sized enterprises. By capturing complex temporal dependencies and focusing on relevant patterns in data, these models outperform traditional statistical methods and simpler neural networks.

The successful application context illustrates the practical viability of these models. Addressing challenges such as data limitations, resource constraints, and the need for explainability is critical for implementation.

Adoption of advanced AI models in medium-sized enterprises can lead to substantial operational efficiencies and competitive advantages. Ongoing advancements in AI technologies and increased accessibility through cloud services and AutoML tools are expected to further facilitate this transition.

About the author

* Corresponding author: atmanand.prasad@gmail.com, †shiva3671@gmail.com.

References

- [1] Asadi, S. and Omran, M. G. (2020). Demand forecasting using machine learning algorithms. *Journal of Industrial Engineering International*, 16(1):1–6.
- [2] Bahdanau, D., Cho, K., and Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *ICLR 2015*.
- [3] Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2):157–166.
- [4] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8):1735–1780.
- [5] Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.
- [6] Kumar, V. (2020). Application of mathematical modeling in population forecasting. *Uttaranchal University Journal of Applied and Life Sciences*, 3(1):50–57.
- [7] Lea, C., Flynn, M. D., Vidal, R., Reiter, A., and Hager, G. D. (2017). Temporal convolutional networks for action segmentation and detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 156–165.
- [8] Nand, A. (2023a). Deteriorating inventory management model: adapting to variable lead time and consumer-driven shortages. *International Journal of Operational Research*. DOI: 10.1504/IJOR.2023.10064634.
- [9] Nand, A. (2023b). Dynamic terminology for supplier–retailer collaboration metrics in inventory model. *International Journal of Mathematics in Operational Research*. DOI: 10.1504/IJ-MOR.2024.10064476.
- [10] Nand, A. (2024a). Linear and non-linear time dependent deteriorated inventory control model. *Journal of Advanced Manufacturing Systems*. DOI: 10.1142/S0219686725500039.
- [11] Nand, A. (2024b). Optimized economic production quantity inventory model with Weibull decay, dynamic holding cost under varied demand scenario and demand series analysis. *Journal of Advanced Manufacturing Systems*. DOI: 10.1142/S0219686725500106.
- [12] Nand, A. (2024c). Next-generation inventory optimization: advanced inventory management harnessing demand variability integrating fuzzy logic and granular differentiability. *RAIRO - Operations Research*. DOI: 10.1051/ro/2024226.
- [13] Nand, A. and Chauhan, N. S. (2024). Multi-entity stock dependent model with capacity and manufacture cost restraints. *Malaysian Journal of Science*, 43:37–48.
- [14] Nand, A., Shivanand, V. K., Chauhan, N. S., and Kumar, V. (2022). Inventory control model for pharmaceutical items with exponential demand and holding cost varying based on time. *International Journal of Mechanical Engineering*, 7:550–557.
- [15] Srivastava, V. K. and Nand, A. (2023). Application and technique of inventory control theory in pharmaceutical sciences. *International Journal of Operational Research*, 47:109–123.