Advances in Computer Sciences
Vol. 4 (2021)
https://academicpinnacle.com/index.php/acs

Parallel LSTM and CNN for Demand Forecasting and Production Planning: A Classification Framework

Ketan Rathor
Assistant General Manager, Birla Carbon Inc, India
Corresponding Author: ketanrathor@gmail.com

Abstract: Upon receiving order release updates from clients, the supplier of automotive components in this proposed alters its production plan for the next periods every week. Erroneous order releases can lead to significant costs, such as premium expedited shipping, production overtime, and excess inventory. For the purpose of studying order release variation, this setting is appropriate because the supply chain has adopted a JIT strategy with zero ideal inventory levels. For this reason, precise order releases are crucial for managing production volumes. Preprocessing, model training, and feature selection are the three primary components. Imputation of missing values and data transformation are components of the data preparation phase. To identify the best features from historical data, a powerful GOA algorithm is used in feature selection. For model training, we employed the Parallel LSTM-CNN framework to do this. On the other hand, it makes LSTM and CNN obsolete. The data indicates that the success rate is 96.52%.

Keywords: Convolutional Neural Network (CNN), Demand Driven Distribution, Production Planning.

Introduction

Both interest in and demand for ready-to-wear garments is on the rise. Many people are starting their own businesses in the clothing sector to offer more affordable and environmentally friendly products because of this trend. Businesses can't launch operations or begin producing money unless they decide on the optimal aggregate planning technique. In order to maximize profits, one must find the optimal balance between meeting maximum demand and limiting costs. In order to adapt to fluctuations in demand, most small garment companies utilize the chase method, which involves changing production rates every period. Because the process requires less capital and the materials used are easy to adjust and inexpensive, small clothing firms can profit the most from it. Production rates are set by meticulous planning and forecasting to ensure customer requests are balanced. Thorough planning and forecasting lead to on-time delivery, less waste, and cost-effective production. To get the most out of the chosen strategy, the company needs a meticulously planned production and raw material inventory schedule. In recent years, there has been a proliferation of AI-powered supply chain management systems. These systems optimize data from both inside and outside the supply chain using machine learning and optimization algorithms. For retailers, especially those dealing with consumables with a near-term expiration date, accurately gauging product demand is an ongoing struggle. Due to their short shelf life and often unpredictable demand, perishable commodities typically require daily orders, manufacture, and delivery. When things go unsold at the end of the day, it's called demand waste. When things sell out rapidly, on the other hand, demand drops. Retailers of perishable goods can cut down on lost revenue with the help of an efficient forecasting model by increasing product availability and lowering day-end waste. The importance of accurate forecasting for day-to-day operations is well-known to managers involved in the supply chain for perishable commodities. However, there are several reasons why a production forecasting model has not been easily developed. One reason for the incredibly high degrees of variance at the store level is because product demand patterns are intrinsically unpredictable. Over the past few decades, there have been significant changes in industrial production. Things that have changed include the breadth and depth of production, the technologies used, and the amount of effort put into it. Manufacturers must produce high-quality products at cheap prices that can adjust to customers' shifting demands if they want to remain competitive. An essential component of fixing this problem is production planning, which is the area where optimization, forecasting, and simulation can be used. When actual requirements are met, the method's forecasts are accurate. Production planning that relies on precise projections is crucial for lowering risk and boosting company performance. A study's authors have ramped up their usage of prognostic methods and Bootstrap (Bagging) combination aggregations to improve electric energy demand estimates. Outcomes from forecasting models constructed using the ANN principle outperform other methods of prognosis. The essay compares and contrasts two well-known prediction models, namely, ARIMA models and state space models. The majority of supply chain management decisions necessitate predictions. The level of detail in a decision is mostly defined by two variables: time and product. Some product-related decisions, like inventory control, only consider a single SKU, while others, like aggregate capacity planning, consider all SKUs. Decisions regarding operations are decided on a daily or weekly basis, while decisions regarding strategy and tactics are made on a monthly or annual basis. Supply chain managers at the strategic level face capacity unpredictability, increasing market volatility, and rapid technical and market shifts when deciding on distribution channels (online, store, or Omni channel). They are compelled to consider the entire spectrum of items available to those markets as a result. Decisions on product lineups, production capacity, and inventory storage are made at the tactical level. Plans for transportation, production schedules, employee rosters, after-sale services, and inventory replenishment are all ultimately decided upon at the operational level.

Literature survey

Operations in production planning and control (PPC) attempt to define when, how much, and what to make, purchase, and ship in order for a business's manufacturing performance to match client demands. Research by [1] suggests that PPC can be viewed as a method that improves the productivity of manufacturing. PPC needs to be flexible enough to adapt to new supply chain opportunities, complex client needs, and operational and strategic situations. [2]all agree that PPC needs to be adaptable, complimentary, and always changing. The dynamic nature of the industrial environment necessitates a more comprehensive and evolutionary perspective from the PPC function and operations management. The PPC function considers many activities, including JIT, enterprise resource planning (ERP), collaborative planning, forecasting, and replenishment, and material requirements planning (MRP). [3]states that information and communication technology has recently been a huge boon to PPC. Factors that aid in the organization and management of crucial production processes include production scheduling, master production scheduling (MPS), sales and operations planning (S&OP), demand forecasting, material requirements planning (MRP), and information and communication technology[4], PPC also includes interface processes including ordering systems, procurement, shipment, capacity analysis, and input/output control. [5]Several studies have found that the rise of the Internet of Things (IoT) and cyber-physical systems (CPS) in industrial settings has introduced a new PPC context. The extensive use of statistical methods can be attributed to three primary factors. They are easy to use and implement, first of all. Second, the results are computed and the forecasts are made in a flash. [6] They are more easily integrated with other decisions pertaining to company operations, such inventory management, because they may be expressed in a closed form. Merchants that need to think about how time-series trends, especially seasonal changes, affect future demand for products with steady demand can benefit from statistical methodologies. [7]were the first to forecast the demand for apparel. When it comes to predicting the demand for women's clothes, take a Bayesian method. [8]have explored the hypothesis that a Bayesian forecasting model may be applied to anticipate the demand for apparel in the future. When compared to existing methods, the proposed hierarchical Bayesian strategy produces better quantitative results. In their study, [9] compare a number of suitable methodologies for forecasting the single-period products. The study relies on advance order data collected from clients who pre-purchase and other sources due to the lack of demand data from the past. Their analysis of the situation at one mail-order Apparel Company led them to suggest a novel "top-flop" method that is more effective than the alternatives based on pre-orders. [10] Better results than the advanced demand information method are also seen with expert judgment procedures several things. Their findings suggest that in order for the corresponding forecasting method to achieve higher levels of prediction accuracy, it requires additional item families and applicable classification criteria.Complexity and uncertainty abound in modern supply systems. Unpredictability in the supply chain is caused by several things, such as the network itself, partner actions, customer and competitor actions, new technology introduction, and product development [11]. Market volatility is the root cause of demand fluctuations. Avoiding the negative effects of demand volatility requires careful consideration and anticipation of the uncertainties. The difficulty in accurately predicting future demand is one factor that can increase this uncertainty [12]. Due to the difficulty of demand forecasting, many businesses and forecasters refrain from doing scientific forecasts [13]. One of the key challenges in demand forecasting is the inherent uncertainty in the market. The degree to which demands are correct determines the level of unforcedness of decisions. Inaccurate forecasts may lead to overspending on transportation, labor, service level, and inventory. Only recently has the predictive production system used cyber-physical systems (CPS) technology [14]. An example of a cyber-physical system (CPS) in manufacturing would be a cyber-physical production system (CPS) for process automation and dynamic system control [15]. To stimulate the adoption of cleaner production strategies, we have built an optimized planning model to identify the low-carbon production paths for the EIMIs and introduced energy-CPS enabled management power industry of the country [16]. Adding to that, the IoT may be integrated into manufacturing processes to create the Internet of Manufacturing. The Manufacturing Process. Using a model-based approach made possible by the internet of things is one strategy to increase the energy efficiency of aluminum die casting. [17]Traditional predictive analytics are finding it increasingly difficult to meet the expectations of predictive production. Big data analytics combined with predictive analytics is a new technology that businesses may use to increase energy and resource efficiency. With the innovative usage of RFID-cuboids to build a data warehouse, the logistical data that is enabled by RFID may be highly integrated in terms of tuples, logic, and operation [18]. In its most basic form, PPC deals with the problem of controlling uncertainty in production systems by stabilizingit (as is common with lean approaches) or by precisely predicting and reacting quickly to events and changes in the system's status. [19] The requirement to reschedule the latter depends on the kind and frequency of operations, the stability of the production environment, and other factors. Process logics and methodologies have been developed to achieve these goals based on extensive research in various areas and levels of complexity (hierarchical systems). [20] Some examples of these areas and levels of detail include algorithmic research, strategic selection of PPC systems, and implementation challenges and restrictions. The range and depth of the themes and topics examined have led to the emergence of various areas of investigation. [21] Other settings include small and medium enterprises (SMEs) and dynamic market environments. Research in this area is often motivated by the perceived limitations and shortcomings of ERP systems when it comes to supporting activities related to production planning and control. The most common gripe with enterprise resource planning (ERP) systems is the endless capacity scheduling that leads to unrealistic or unachievable production plans.

Proposed system

Modern supply chains rely heavily on demand forecasting capabilities, which allow for extremely precise inventory management in response to customer demands. Without accurate forecasting, it is very difficult to plan effectively and efficiently. It is worth noting that no specific method is recommended as the best for prediction based on the reviewed research. This is because a wide variety of forecasting approaches and selection criteria are at your disposal.

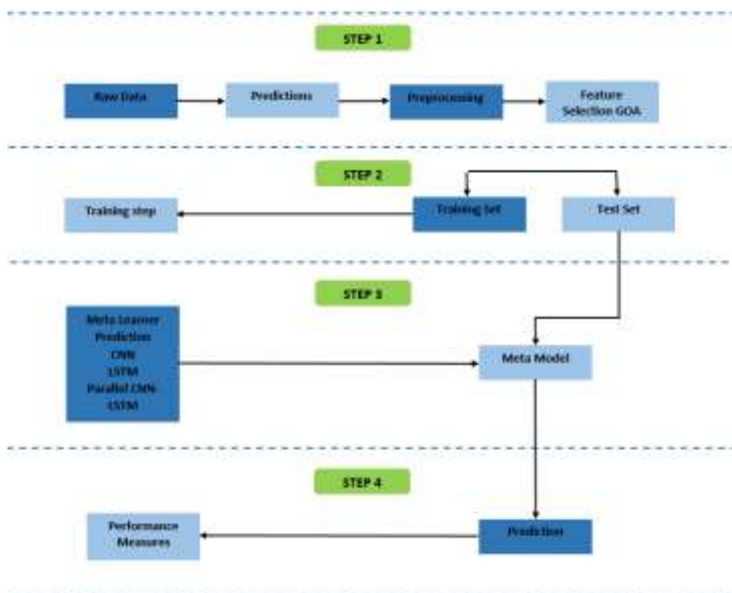

Fig. 1. An Automated Demand Forecasting System Design Proposal

This section provides a concise description of the materials and process that were employed. You can see the suggested structure in Figure 1.

Data Preprocessing:

The preparation phase includes data manipulation and missing value imputation. Raw data must have some missing attributes filled up from various features before any ML technique can be used. A number of imputation methods are available for use in filling in missing variables. To fill in the missing value, we recommend the mean-based imputation process, which entails averaging the properties of all features. [22] Following the imputation of missing or null values, this data power transformation is executed. Regression analysis relies heavily on transformations. Parametric monotonic transformations, also known as power transformations, can be used to make data more Gaussian-like. This approach is useful when dealing with heteroscedasticity or when data normality is required [23]. Notable methods for power transformations include the Box-Cox and Yeo-Johnson transformations. Here we utilize the Box-Cox transformation instead of the Yeo-Johnson transformation as the former accepts only positive data and the latter only negative data. A description of the Yeo-Johnson transformation using

$$m^* = \begin{cases} \dfrac{\left((m+1)^{\delta} - 1\right)}{\delta} & if\ \delta = 0, m \leq 0 \\ \log(m+1) & if\ \delta = 0, m \geq 0 \\ -\dfrac{[(-m+1) \wedge \{2-\delta\} - 1]}{(2-\delta)} & if\ \delta \neq 2, m < 0 \\ \log(-m+1) & if\ \delta = 2, m < 0 \end{cases} \quad (1)$$

The transformed value is denoted by m^*,m is a set of p strictly positive values, and δ is a hyperparameter that controls the transformation. Here, the Power Transformer implementation in Scikit-learn is used to carry out the Yeo-Johnson power transformation operation. The transformed output is normalized to zero with a unit variance using implicit data processing.

Feature Selection:

From the available historical data, the most suitable attributes are chosen using a strong GA algorithm. Grasshopper swarming behavior, which is affected by social relationships, gravity, and wind forecast, served as inspiration for the GOA, a meta-heuristic optimization approach. The reviewed work provides a detailed description. The pseudocode of the algorithm is shown. The position of the ith grasshopper, denoted as R_b, is determined using Eq. 2:

$$R_b = E_b + D_b + F_b \quad (2)$$

In these equations, E_b represents social interaction, D_b represents gravity, and F_b represents the bth grasshopper's wind forecast [24]. Interactions with others are the most important factor in determining one's status. Equations 3-6 allow us to determine:

$$D_b = \sum_{l=1}^{P} e(g_{bl})\, \hat{g}_{BL} \quad (3)$$

$$g_{bl} = |y_l - r_b| \quad (4)$$

$$\hat{g}_{BL} = (y_l - r_b)/g_{bl} \quad (5)$$

$$e(x) = ae^{-x/j} - e^{-x} \quad (6)$$

In these equations, the distance between the bth and lth grasshopper is represented as g_{bl}, and a unit vector between the two is represented as $\hat{g}_{BL}$. In addition, the function s defines social forces. Adjust it with the a and l keys. For this force to be exerted, the distance between any two grasshoppers must be larger than. Grasshoppers' gravity is also determined by Equation 7:

$$D_b = -d\hat{s}_d \qquad (7)$$

The gravitational constant is denoted by d and the unit vector heading toward the Earth's center is denoted by $\hat{s}_d$ in this context. Grasshoppers utilize Equation 8 to forecast the wind:

$$F_b = -w\hat{s}_u \qquad (8)$$

where w stands for the continuous drift and $\hat{s}_u$ for the unit vector perpendicular to the wind's direction, respectively. Here is the mathematical model that follows from this:

$$E_b = \sum_{l=1}^{P} e(|y_l - r_b|)\,(y_l - r_b)/g_{bl} - d\hat{s}_d + w\hat{s}_u \quad (9)$$

where P represents the count of grasshoppers. In order to enhance the extraction and exploration of the proposed model, it is presumed that grasshoppers are marginally impacted by gravity and that the wind is blowing in the direction of the optimal solution $\hat{C}_g$. So, to sum up, the final mathematical model becomes:

$$R_b^g = t\left(\sum_{l=1}^{P} t\,\frac{wi_g - bi_g}{2}\,e\left(\left|r_l^g - r_b^g\right|\right)(y_l - r_b)/g_{bl}\right) + \hat{C}_g \qquad (10)$$

In this paradigm, the ideal solution up to this point is represented by $\hat{C}_g$, where g is the g th dimension and bi_g and wi_g are the lower and upper bands of the dth dimension, respectively. Further, exploration and extraction are both affected by the value of t. Iterations of the following equation yield the answer:

$$t = t_{max} - j\,\frac{t_{max} - t_{min}}{J} \qquad (11)$$

The two extremes of this equation are $t_{max}$and t_{min}, respectively. The current iteration index is also represented by j, with J being the maximum iteration number.

Model Training:

Parallel LSTM-CNN:

This proposed delves into a novel approach to load prediction that combines CNN and LSTM, dubbed parallel LSTM-CNN Network (PLCNet). The methodology offered here is totally distinct from previous attempts, such as the ones discussed in the introduction, which integrated the two methodologies. For instance, in their proposed CNN-LSTM model, the authors utilize CNN to extract characteristics from input data before feeding it into an LSTM. The issue with this model is that the training of LSTM is impacted by the extracted features. To address this issue, the PLCNet employs LSTM and CNN networks in independent pathways, with no correlation between them. It is demonstrated that two paths are first utilized to process input signals; these paths are LSTM and CNN. In order to prepare the input data for the final prediction, these two paths extract the features and long dependency inside the data. The final prediction was carried out by comparing data using a fully linked path that included dense and dropout layers. The path also predicted real values. Acquiring the local trend feature is the primary goal in the CNN path. Under this route, the data is passed through a 60-unit, 2-filter Conv-1D layer. The Maxpooling layer retains data quality after downsampling to reduce data dimensionality following the convolution layer. The following layer incorporates yet another Conv-1D layer, this time situated within 36 units. Flatten layer data continues from last layer. The Rectified Linear Unit (ReLU) energizes each and every one of the constituent parts. For the LSTM path to pick up on data's long-term dependencies, data must first pass through a flatten layer before it can begin to interact with the network. Data is prepared to be input into the LSTM layer after it has passed through the flatten layer. An LSTM layer training on a ReLU activation function with 48 units. The fully connected layer is ready to receive processed data after it has passed through the LSTM and CNN pathways. The two paths are completely unconnected, as we have already established. Thus, in order to prepare the data for prediction, the outputs are pooled in a merge layer. To learn the long-dependency within the output data from two pathways, we feed the output of one dense layer into another, and then we mix the data and feed it into an LSTM layer with 280 units and a ReLU activation function. Then, 40% of the total goes toward preventing overfitting via a dropout layer. With the help of two more dense layers, the data is ready for the final forecast. Each thick layer has its own unique number of units because this model is attempting to predict two types of data over separate time horizons. But what the sigmoid function actually does is turn on all of the units in the linked circuit. The unit count of the LSTM-Dense path can vary because the PLCNet model needs to be validated across different time horizons. Depending on the goal of the prediction, the figure indicates that the quantity of data in each batch might change. The quantity of look-back steps, in its simplest form, changes over time horizons. After the look back number is chosen, load data batches of the same size are generated. Assuming a look-back step count of 25, the first batch will have data points 0–24, the second batch will contain data points 1–25, the third batch will contain data points 2–26, and so forth. The target data may also change as a function of the prediction horizon.

CNN:

A big family of ANNs created to extract features from input data are CNNs. Their wide application across numerous disciplines is a result of their versatility in handling data with different dimensionalities. As an example, it is well-known that two- and three-dimensional CNNs are powerful networks for computer vision, picture processing, and classification. Furthermore, they have recently discovered additional applications in fields like load forecasting, medical, speech recognition, and Natural Language Processing (NLP). Problems may develop if CNN networks are used with load profiles that already differ. The nuances of human behavior, workdays, times, and weather conditions have a profound impact on load profiles. To learn all of its parameters, CNN need vast volumes of input data [25]. This allows them to bypass the complexity of load profiles. From a strictly mathematical standpoint, equation 12 shows that convolution is the building block of CNNs. If you plug in r and M into equation 12, you get M as the result. Also, u is the symbol for the kernel. The b -th output can be obtained by following these steps:

$$M(b) = \sum_{b,l} r(b-l)u(l) \qquad (12)$$

When l is between 0 and $h-1$, M is changed such that it has $p-h+1$ dimensions, where n is the input dimension. The kernel and filter are convolved in the convolution procedure, and the outcome is added to a bias term. This arithmetic procedure is finished when the feature map is finished. Equations (13) and (14), which show the complete convolution process in ANNs, are:

$$M_{bl}^{y} = sum\left(h_y \circledast ra_{bl}\right) + i_y \qquad (13)$$

$$a^{y} = activation(M^{y}) \qquad (14)$$

Here are the key points: M^y stands for the output, y for the y -th feature maps, $b\ and\ l$ for the filter steps, ra_{bl} for the filter matrix, h_y for the kernel matrix, i_y for the bias term, and f m for the activation function output. Equation 13 shows the convolution operation formula, and equation 14 shows the activation function for the y-th output.

Result And Discussion

Garment manufacturers have resorted to aggregate planning as a means of optimizing profitability by determining the optimal level of sales growth relative to production cost reduction. In order to respond to changes in demand, smaller garment makers often employ what is known as the "chase strategy," which entails regularly adjusting production rates. However, without precise demand estimates and careful production planning, this method does not take into consideration the unexpected demand, which increases the likelihood of product shortages and unmet customer requests.

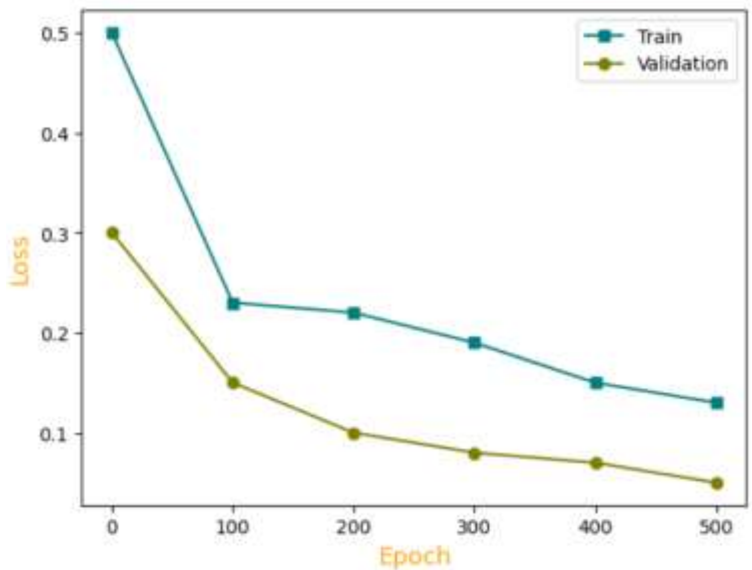

Fig. 2. Learning Curve for Different Methods

The graphs below show the three models' learning curves. The validation set's learning curve is shown by the olive line and the blue line, respectively. The learning curves for the CNN model are plotted in Figure 2.

TABLE I. MODEL PERFORMANCE(%)

Models	***RMSE***	***Cross Validation***
CNN	*0.8712*	*0.8871*
LSTM	*0.8924*	*0.9054*
Parallel LSTM-CNN	*0.9216*	*0.9352*

Using root-mean-squared error (RMSE), it assessed how well each of the three methods predicted. In the introductory section, to discussed the performance measures and how they were implemented. Table I's data makes the case for Parallel LSTM-CNN as the best model. Conversely, CNN performs better than LSTM.

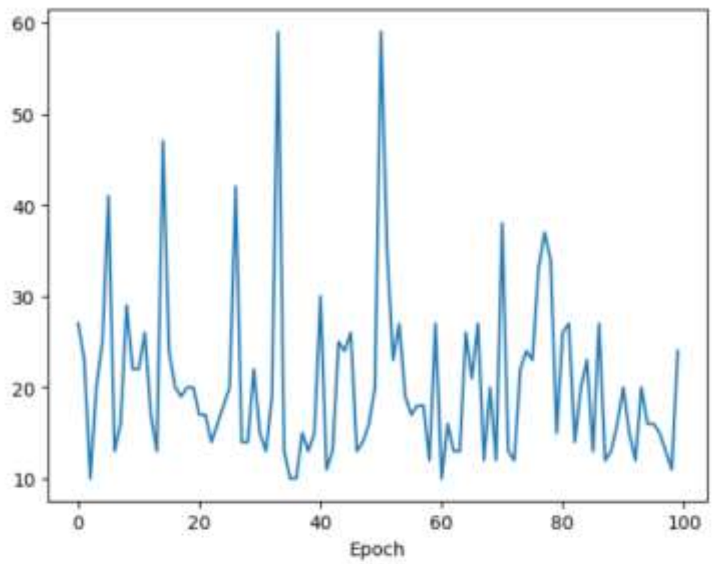

Fig. 3. **Deviation in Parallel LSTM-CNN**

The reason behind the particular material's forecast deviation in this investigation is discussed in the analysis. The forecast metric unique to each material is used to identify deviating materials. Three models are used to create plots of the forecast deviation for a particular material. Figure 3 shows the deviation graph for the CNN-LSTM model.

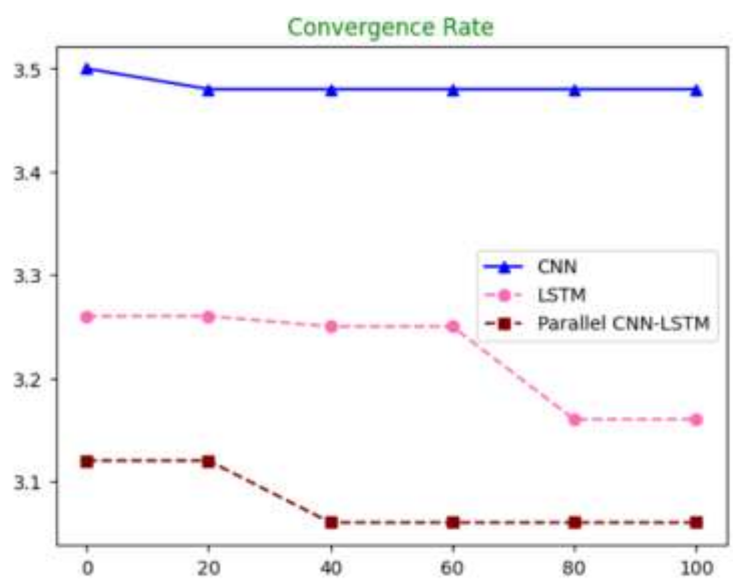

Fig. 4. **Convergence Rate**

When compared to the outcomes of CNN and LSTM, Parallel LSTM-CNN performs better because to exponential crossover between neighboring decision variables. LSTM and Parallel LSTM-CNN have consistently demonstrated a capacity to advance, but Fig. 4 shows that CNN has not improved since the beginning of the iteration process.

Conclusion

Although demand forecast information has been extensively studied in the inventory literature, it has received little attention when it comes to production planning. By combining concepts from forecast evolution and inventory theory, this paper tackles the problem of planning work releases into a manufacturing facility while dealing with stochastic demand. The data preparation stage includes data transformation and the imputation of missing variables. Feature selection uses a strong GOA algorithm to find the best features from historical data. The Parallel LSTM-CNN approach is used to train the model. The accuracy of the suggested technique consistently beats that of the CNN and LSTM models (96.52 percent).

References

[1] D. T. Ch. Achillasa, D. Aidonisb, E. Iakovouc, M. Thymianidisa, “A methodological framework for the inclusion of modern additive manufacturing into the production portfolio of a focused factory,” *J. Manuf. Syst.*, p. 12, 2014.

[2] J. C. Bendul and H. Blunck, “The design space of production planning and control for industry 4.0,” *Comput. Ind.*, vol. 105, pp. 260–272, 2019, doi: 10.1016/j.compind.2018.10.010.

[3] B. Denkena, M. A. Dittrich, and S. Jacob, “Methodology for integrative production planning in highly dynamic environments,” *Prod. Eng.*, vol. 13, no. 3–4, pp. 317–324, 2019, doi: 10.1007/s11740-019-00889-0.

[4] U. Dombrowski and Y. Dix, “An analysis of the impact of industrie 4.0 on production planning and control,” *IFIP Adv. Inf. Commun. Technol.*, vol. 536, pp. 114–121, 2018, doi: 10.1007/978-3-319-99707-0_15.

[5] Q. Li, I. Kucukkoc, and D. Z. Zhang, “Production planning in additive manufacturing and 3D printing,” *Comput. Oper. Res.*, vol. 83, pp. 1339–1351, 2017, doi: 10.1016/j.cor.2017.01.013.

[6] L. Aburto and R. Weber, “Improved supply chain management based on hybrid demand forecasts,” *Appl. Soft Comput. J.*, vol. 7, no. 1, pp. 136–144, 2007, doi: 10.1016/j.asoc.2005.06.001.

[7] C. C. Chern, C. P. Wei, F. Y. Shen, and Y. N. Fan, “A sales forecasting model for consumer products based on the influence of online word-of-mouth,” *Inf. Syst. E-bus. Manag.*, vol. 13, no. 3, pp. 445–473, 2015, doi: 10.1007/s10257-014-0265-0.

[8] L. Lin, Z. He, and S. Peeta, "Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach," *Transp. Res. Part C Emerg. Technol.*, vol. 97, pp. 258–276, 2018, [Online]. Available: https://doi.org/10.1016/j.trc.2018.10.011.

[9] J. Mostard, R. Teunter, and R. De Koster, "Forecasting demand for single-period products: A case study in the apparel industry," *Eur. J. Oper. Res.*, vol. 211, no. 1, pp. 139–147, 2011, doi: 10.1016/j.ejor.2010.11.001.

[10] O. Schaer, N. Kourentzes, and R. Fildes, "Demand forecasting with user-generated online information," *Int. J. Forecast.*, vol. 35, no. 1, pp. 197–212, 2019, doi: 10.1016/j.ijforecast.2018.03.005.

[11] Ö. G. Ali, S. Sayin, T. van Woensel, and J. Fransoo, "SKU demand forecasting in the presence of promotions," *Expert Syst. Appl.*, vol. 36, no. 10, pp. 12340–12348, 2009, doi: 10.1016/j.eswa.2009.04.052.

[12] G. K. Armstrong JS, "Demand Forecasting II: Evidence-Based Methods and Checklists," *Wharton*, pp. 1–36, 2017, [Online]. Available: https://faculty.wharton.upenn.edu/wp-content/uploads/2017/05/JSA-Demand-Forecasting-89-clean.pdf?fbclid=IwAR1_YWYihTMAfJyxJE5qhf1R9BrQwaR2WxLVW674FV-OS2CQKt-2wIyNKzA.

[13] R. Carbonneau, K. Laframboise, and R. Vahidov, "Application of machine learning techniques for supply chain demand forecasting," *Eur. J. Oper. Res.*, vol. 184, no. 3, pp. 1140–1154, 2008, doi: 10.1016/j.ejor.2006.12.004.

[14] J. Lee, C. Jin, and B. Bagheri, "Cyber physical systems for predictive production systems," *Prod. Eng.*, vol. 11, no. 2, pp. 155–165, 2017, doi: 10.1007/s11740-017-0729-4.

[15] L. Monostori, "Cyber-physical production systems: Roots, expectations and R&D challenges," *Procedia CIRP*, vol. 17, pp. 9–13, 2014, doi: 10.1016/j.procir.2014.03.115.

[16] C. B. Wu *et al.*, "An optimized low-carbon production planning model for power industry in coal-dependent regions - A case study of Shandong, China," *Energy*, vol. 192, 2020, doi: 10.1016/j.energy.2019.116636.

[17] R. Dubey *et al.*, "Examining the role of big data and predictive analytics on collaborative performance in context to sustainable consumption and production behaviour," *J. Clean. Prod.*, vol. 196, pp. 1508–1521, 2018,

doi: 10.1016/j.jclepro.2018.06.097.

[18] R. Y. Zhong, G. Q. Huang, S. Lan, Q. Y. Dai, X. Chen, and T. Zhang, "A big data approach for logistics trajectory discovery from RFID-enabled production data," *Int. J. Prod. Econ.*, vol. 165, pp. 260–272, 2015, doi: 10.1016/j.ijpe.2015.02.014.

[19] E. Arica and D. J. Powell, "A framework for ICT-enabled real-time production planning and control," *Adv. Manuf.*, vol. 2, no. 2, pp. 158–164, 2014, doi: 10.1007/s40436-014-0070-5.

[20] A. Bueno, M. Godinho Filho, and A. G. Frank, "Smart production planning and control in the Industry 4.0 context: A systematic literature review," *Comput. Ind. Eng.*, vol. 149, no. August, p. 106774, 2020, doi: 10.1016/j.cie.2020.106774.

[21] J. P. Usuga Cadavid, S. Lamouri, B. Grabot, R. Pellerin, and A. Fortin, "Machine learning applied in production planning and control: a state-of-the-art in the era of industry 4.0," *J. Intell. Manuf.*, vol. 31, no. 6, pp. 1531–1558, 2020, doi: 10.1007/s10845-019-01531-7.

[22] M. S. Raju, S. T. U., Sarker, A., Das, A., Islam, M. M., Al-Rakhami, "An approach for demand forecasting in steel industries using ensemble learning," *Complexity*, vol. 2022, pp. 1–19, 2022, doi: 10.1155/2022/9928836.

[23] S. M. T. U. Raju *et al.*, "An Approach for Demand Forecasting in Steel Industries Using Ensemble Learning," *Complexity*, vol. 2022, 2022, doi: 10.1155/2022/9928836.

[24] M. Salami, F. M. Sobhani, and M. S. Ghazizadeh, "A hybrid short-term load forecasting model developed by factor and feature selection algorithms using improved grasshopper optimization algorithm and principal component analysis," *Electr. Eng.*, vol. 102, no. 1, pp. 437–460, 2020, doi: 10.1007/s00202-019-00886-7.

[25] B. Farsi, M. Amayri, N. Bouguila, and U. Eicker, "On short-term load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN approach," *IEEE Access*, vol. 9, pp. 31191–31212, 2021, doi: 10.1109/ACCESS.2021.3060290.