Multidisciplinary International Journal of Research

Vol 1 (3), 2023, 108 - 115

Online ISSN: 2985 - 9999

Multidisciplinary International Journal of Research

Journal homepage: https://publishing.impola.co.id/index.php/MIJR

Optimized Hybrid Deep Learning for Cost Estimation of Construction Projects Considering the Time-Dependent Characteristics of Economic Variables and Indices

Frederik Elly Gosal[1], Sudarno P Tampubolon[2]

[1] Civil and Construction Engineering Department, National Taiwan University of Science and Technology, Taipei, Taiwan.

[2] Civil Engineering Department, Universitas Kristen Indonesia, Jakarta, Indonesia.

Abstract

The SOS-NN-LSTM hybrid model emerges as a transformative solution to the perpetual challenge of accurate cost estimation in the construction industry. By integrating Neural Networks (NN) and Long Short-Term Memory networks (LSTM), optimized through the Symbiotic Organism Search (SOS) algorithm, the model distinguishes itself with remarkable performance metrics. Achieving an R-squared value of 0.9631 and a minimal Root Mean Squared Error (RMSE) of 0.023, SOS-NN-LSTM showcases its efficacy in handling project physical, financial, and time-dependent economic variables. The model's superiority is further emphasized by a minimal Mean Absolute Error (MAE) of 0.0168 and a notable Relative Improvement (RI) value of 0.96018. Importantly, the Mean Absolute Percentage Error (MAPE) experiences a significant decrease from 20% to 10%, highlighting the model's enhanced precision in cost estimation. SOS-NN-LSTM not only outperforms other AI models but also demonstrates its adaptability to dynamic economic conditions, effectively addressing the volatility of economic variables in construction projects. These exceptional results position SOS-NN-LSTM as a pioneering advancement in cost estimation methodologies for the construction industry. The model's nuanced handling of complex datasets and its ability to provide accurate projections underscore its potential to revolutionize project budgeting, mitigate risks, and contribute to the overall success and financial viability of construction projects. The integration of deep learning with SOS optimization represents a paradigm shift, offering a reliable and effective tool for project stakeholders in navigating the intricacies of construction cost estimation.

Keywords—Cost Estimation; Economic Variables; Hybrid Deep Learning; Symbiotic Organism Search; Neural Network; Long Short-term Memory.

I. INTRODUCTION

The construction industry, known for its intricacy and susceptibility to external variables, holds a crucial position in the global economy. Amid the challenges faced by stakeholders in construction industry, accurate cost estimation is pivotal for project success and financial viability (Barnes, 1988). Enormously construction projects experience cost overrun, a study by Flyvbjerg et al found that Costs are underestimated in almost 9 out of 10 projects. For a randomly selected project, the likelihood of actual costs being larger than estimated costs is 86%. On the other hand, the likelihood of actual costs being lower than or equal to estimated costs is 14%, shows the high uncertainty in the projects budgeting (Flyvbjerg et al., 2002). The Economic variables and price indexes play

Article info
Received (22/11/2023)
Revised (27/11/2023)
Accepted (29/11/2023)
Corresponding: m11005810@mail.ntust.edu.tw
DOI:

a crucial role in the costs uncertainty and dynamics, so that it is very volatile and will directly affect material and labour costs resulting in cost under or overestimated (Musarat et al., 2021). However, the traditional method doesn't consider the influence of economic variables and the price index due to its complexity and dynamics over time. With projects becoming increasingly complex and uncertainties rising, the integration of advanced technologies is imperative. In order to fill and addresses the gap, the application of deep learning techniques offers a paradigm shift in traditional cost estimation methodologies (Alekhya et al., 2022; Shoar et al., 2022).

Deep learning technique have been utilized as a tool in solving forecasting, estimating problems, it adept at extracting intricate patterns and dependencies from vast datasets, and has shown great performance across diverse industries (Abioye et al., 2021). Within the construction domain, the adoption of deep learning stands to enhance estimation accuracy, mitigate risks, and optimize resource allocation. This study delves into the synthesis of deep learning with economic insights, focusing on the development and application of a novel hybrid model the Symbiotic Organism Search Optimized Hybrid Neural Network and Long Short-term Memory (SOS-NN-LSTM) model proposed by Cheng (Cheng et al., 2020).

The objective of this research is to refine cost estimation methodologies in construction projects by incorporating a sophisticated hybrid deep learning model. This model leverages Neural Networks (NN) to handle project physical and financial variables, while Long Short-Term Memory networks (LSTM) manage economic variables. The inclusion of LSTM is particularly pertinent, given the time-dependent nature of economic variables, which means the variables has a temporal dependencies characteristic. The hybrid model, further optimized by the Symbiotic Organism Search (SOS) algorithm, aims to enhance the adaptability of cost estimation models to dynamic economic conditions, fostering more accurate and reliable project cost projections.

SOS-NN-LSTM has proven its performance through several case studies. In 2020 Cheng employed the SOS to optimize the NN-LSTM and successfully implemented it for mapping the construction cash flow considering the complexity of project by showing a great performance compared with other AI models. Gunawan in 2019, utilized SOS-NN-LSTM as a predictor to predict the estimated schedule to completion and estimate cost to completion in order to established the management strategy for construction projects considering the trade-off between cost and time (Gunawan et al., 2019).

Through this exploration, the study aims to illuminate the significance of deep learning applications in the construction industry, emphasizing the effectiveness of the proposed SOS-NN-LSTM hybrid model. Highlighting the pivotal role played by economic variables and their time-dependent nature.

II. Research Methodology

In this study, an AI technique with prediction task will be utilized to estimate the construction cost based on the project physical and financial variables, and the economic variables. An optimized hybrid deep machine learning model, SOS-NN-LSTM, is used to carry out the estimation process. Figure 1 shows the detailed flowchart of the SOS-NN-LSTM process for cost estimation with a prediction task.

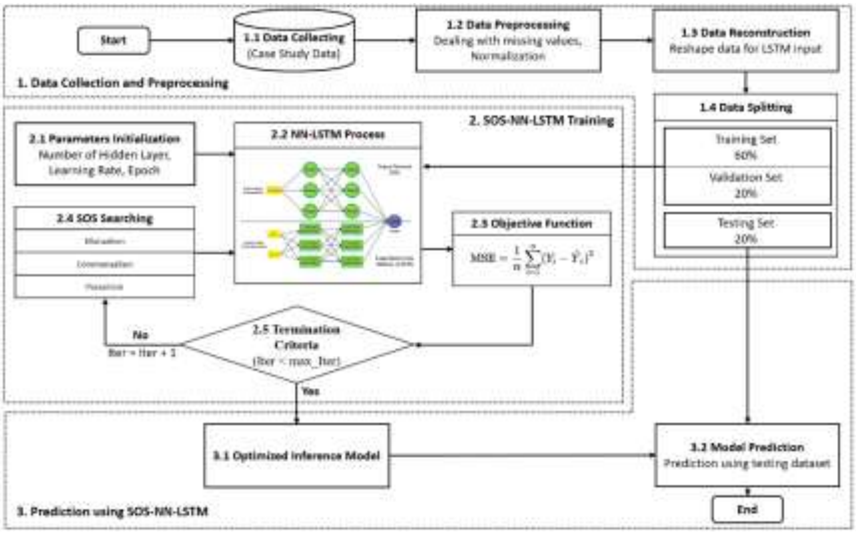

Figure. 1. SOS-NN-LSTM Flowchart

The following is the explanation of the main process of SOS-NN-LSTM flowchart.

A. Data Collection and preprocessing.

The information used in this research aims to figure out the real cost of constructing a residential building project by using past project details. This particular case is drawn from a study by Rafiei, M. H., & Adeli, H. in 2018. The dataset comes from 372 residential buildings in Tehran, Iran, with 3-9 floors, constructed between 1993 and 2008. The gathered data includes 8 physical and financial variables related to the project (P&F), 19 economic variables and indices (EV&I), and the construction cost, which is the main focus (Rafiei & Adeli, 2018). The goal of this study is to foresee the construction cost of residential buildings in Iran, taking into account the economic variables and indices.

The dataset has already been reconstructed into 5 timesteps for the EV&I variables. Therefore, the total number of variables are 8 P&F variables add up by 5 timesteps multiplied by 19 EV&I variables, summing up to a total of 103 variables. The general raw data structure in this case is shown in Table 1.

Table. 1. Data Structure

No	Project's Physical and Financial Variables				Economic Variables and Indices							Construction Cost (C)
					timestep 1			…	timestep 5			
V	**1**	**2**	**…**	**8**	**9**	**…**	**27**	**…**	**9**	**…**	**27**	**28**
1	1.00	3150.00	…	1200.00	6713	…	628133	…	7196	…	601988	410.00
2	1.00	7600.00	…	2900.00	3152	…	1188996	…	3678	…	929027	1000.00
:	:	:	…	:	:	…	:	…	:	…	:	:
326	20.00	1860.00	…	3300.00	2859	…	1181573	…	3157	…	954629	380.00
327	20.00	1460.00	…	4700.00	5909	…	1681849	…	3629	…	1660444	590.00

The factors considered in this study as inputs are divided into two categories: project physical and financial variables, and economic variables and indices. The decision to use project physical and financial variables as inputs for the Neural Network (NN) is based on their time-independent nature. These variables remain constant over time. On the other hand, economic variables and indices are designated as inputs for the Long Short-Term Memory (LSTM) due to their time-dependent characteristics. These variables are collected before the start of the project and are recorded quarterly. Table 2 illustrates the NN's input with time-independent variables and the LSTM's input with time-dependent variables.

Table. 2. Input Variables

Variable		Descriptions	Unit
Project's Physical and Financial Variables (Time-independent)	1	Project locality defined in terms of zip codes	-
	2	Total floor area of the building	m2
	3	Lot area	m2
	4	Total preliminary estimated construction cost based on the prices at the beginning of the project	Dollars
	5	Preliminary estimated construction cost based on the prices at the beginning of the project	Dollars/m^2
	6	Equivalent preliminary estimated construction cost based on the prices at the beginning of the project in a selected base year	Dollars/m^2
	7	Duration of construction	month
	8	Price of the unit at the beginning of the project per square meter	Dollars/m^2
Economics Variables and Indices (Time-dependent)	9	The number of building permits issued	-
	10	Building services index (BSI) for a preselected base year	-
	11	Wholesale price index (WPI) of building materials for the base year	-
	12	Total floor areas of building permits issued by the city/municipality	m^2
	13	Cumulative liquidity	millions of dollars
	14	Private sector investment in new buildings	millions of dollars
	15	Land price index for the base year	millions of dollars
	16	The number of loans extended by banks in a time resolution	-
	17	The amount of loans extended by banks in a time resolution	millions of dollars
	18	The interest rate for loan in a time resolution	%
	19	The average construction cost of buildings by private sector at the time of completion of construction	millions of dollars/m^2
	20	The average of construction cost of buildings by private sector at the beginning of the construction	millions of dollars/m^2
	21	Official exchange rate with respect to dollars	%
	22	Nonofficial (street market) exchange rate with respect to dollars	%
	23	Consumer price index (CPI) in the base year	-
	24	CPI of housing, water, fuel & power in the base year	-
	25	Stock market index	-
	26	Population of the city	People
	27	Gold price per ounce	Dollars

After data collection, to ensure the data's quality and suitability for the model, the data is needed to be cleaned by removing any variables with missing values. Then, all the variables include the output or the target variable would be normalized using min-max scaling into 0 to 1 range. After all the variables values are normalized, then the data is divided into Training set, Validation set, and Testing set 60%, 20%, 20% respectively.

B. SOS-NN-LSTM training process

NN-LSTM serve as the learning tool of the model. The architecture of NN-LSTM is as shown in Figure 2.

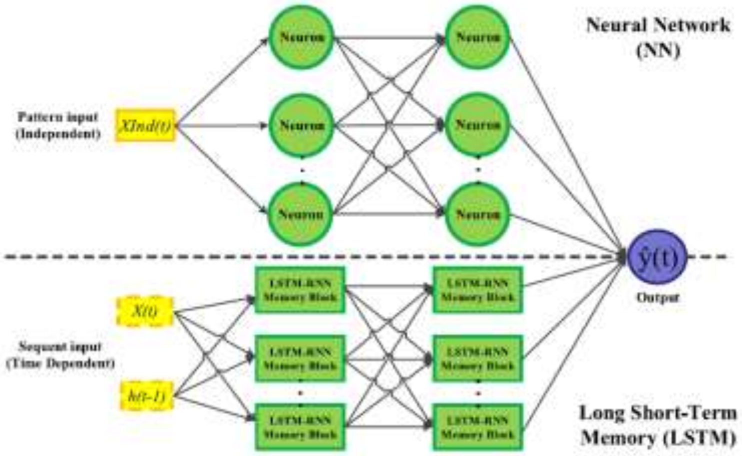

Figure. 2. NN-LSTM Architecture

NN process Project's Physical and Financial Variables (Time-independent) data with high-dimensionality characteristic, and LSTM process the Economics Variables and Indices (Time-dependent) by capturing the temporal dependencies from the data. The combined NN-LSTM then generate the prediction result (Cheng et al., 2019).

SOS is utilized to automatically fine-tune the NN-LSTM model and adjust the output connection parameter. This is achieved by minimizing the Mean Squared Error (MSE) between the predicted and actual values of the training and validation sets. SOS operates by simulating symbiotic interactions within a paired organism relationship to search for the optimal global solution. Acting as a population-based algorithm, SOS enhances the fitness value of organisms by iteratively relocating the entire population to promising areas in the search space. The organism undergoes updates through specific interaction processes when the value of the new organism shows improvement. The SOS algorithm involves three symbiotic interactions: mutualism, commensalism, and parasitism. Figure 3 provides an illustration of the SOS algorithm (Cheng & Prayogo, 2014).

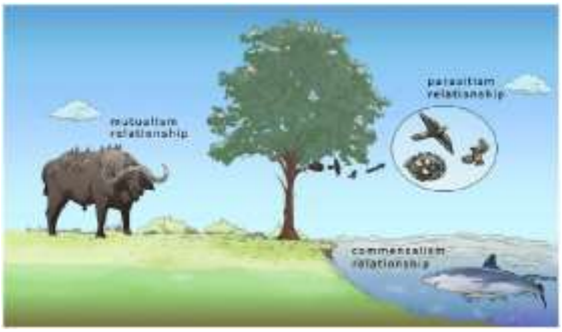

Figure. 3. Symbiotic Organism Search illustration

The following table presents the optimized parameters for NN-LSTM model using SOS, also the number of iterations used and population generated for each iteration by SOS. Lb is the lower-bound and ub is the upper-bound.

Table. 3. SOS parameters

Parameter	Value (lb – ub)
Number of neurons	(5 – 200)
Number of units	(5 – 200)
Dropout rate	(0 – 0.5)
Dense neuron	(1 – 20)

Learning rate	(0.0001 – 0.1)
Relative weight (W_i)	((-1) – 1)
Iteration	50
Population	10

C. Prediction using SOS-NN-LSTM

In this phase, the optimal parameter values obtained by the SOS algorithm in the preceding process are implemented into the NN-LSTM architecture, resulting in the SOS-NN-LSTM model. At this stage, the model is prepared to predict construction costs using the testing data.

To ensure the effectiveness of the SOS-NN-LSTM, a comprehensive evaluation of the model's performance is essential. Evaluation metrics serve as objective tools in machine learning and deep learning techniques, offering quantitative measures to assess performance. These metrics gauge accuracy, quality, and reliability, comparing outcomes or predictions with other state-of-the-art approaches.

Given that the SOS-NN-LSTM focuses on regression tasks, specific evaluation metrics for regression are employed. These include Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Coefficient Correlation (R), and Coefficient of Determination (R^2).

As this study employs five performance evaluation metrics, a composite score system is necessary to interpret the overall model performance. The Reference Index (RI) value is utilized to represent the model's performance across the five-evaluation metrics by assigning equal weights to each metric and summing them up to obtain the RI value. The equation used to calculate the RI value is provided below.

$$RI = \frac{R + R^2 + (1 - RMSE) + (1 - MAE) + (1 - MAPE)}{5} \quad (1)$$

The prediction result from SOS-NN-LSTM then will be compared to the hybrid model without optimization NN-LSTM, and the base model such as NN, LSTM, and the latest variance of RNN, GRU.

III. Result and discussion

In the testing phase, to ensure a fair comparison between models, the parameters of the base model will be set to identical values. This includes aligning the number of neurons for NN, the number of Units for GRU/LSTM, the dropout rate, the number of epochs, and the learning rate. The specific parameters for this case study are detailed in Table 4 for NN and Table 5 for GRU/LSTM.

Table. 4. NN parameter

Parameter	Value
Number of Neurons	30
Dropout Rate	0.2
Learning Rate	0.01
Epoch	500

Table. 5. LSTM/GRU parameter

Parameter	Value
Number of Units	20
Dropout Rate	0.2
Learning Rate	0.01
Epoch	500

Furthermore, the parameter values for both NN and GRU/LSTM will be employed for the NN-LSTM model. This consistent parameter setting across all models ensures a fair and reliable comparison of their performance. The optimal parameters obtained from the SOS optimizer are presented in Table 6.

Table 6. Optimal parameter identified by SOS

Parameter	Initial Value	Optimal Value
Number of NN Neurons	30	123
Number of GRU Units	20	7
Dropout Rate	0.2	0.1698
Dense Neuron	10	2
Learning Rate	0.01	0.0080

Upon running the SOS-NN-LSTM model with the testing dataset and conducting a model comparison between SOS-NN-LSTM, NN-LSTM, and base models, the evaluation metrics for regression are outlined in Table 7.

Table 7. Testing result

Model	R	R2	RMSE	MAE	MAPE	RI	Rank
SOS-NN-LSTM	0.9814	0.9631	0.023	0.0168	0.1038	0.96018	1
NN-LSTM	0.9755	0.95168	0.03302	0.02628	0.20546	0.932484	2
NN	0.92818	0.86192	0.05286	0.04428	0.37208	0.864176	3
GRU	0.91754	0.84288	0.0687	0.05694	0.4397	0.839016	4
LSTM	0.94078	0.88514	0.08782	0.0728	0.47702	0.837656	5

According to the findings in the model comparison table, SOS-NN-LSTM demonstrates impressive performance in predicting residential construction costs, surpassing other models. The results underscore the hybrid model's superiority in handling complex datasets. Moreover, the effective utilization of SOS to optimize NN-LSTM significantly enhances prediction accuracy, evident in the improvement from 0.93 to 0.96 based on the RI value (Ezugwu & Prayogo, 2019). The MAPE also highlights a substantial enhancement resulting from the integration of SOS, with a decrease from 20% to 10%.

IV. CONCLUSION

In conclusion, the SOS-NN-LSTM hybrid model, which integrates Neural Networks (NN) and Long Short-Term Memory networks (LSTM) optimized by the Symbiotic Organism Search (SOS) algorithm, emerges as a robust solution to the challenges of accurate cost estimation in the construction industry. The model's adept handling of project physical and financial variables with NN, alongside its consideration of the time-dependent nature of economic variables through LSTM, demonstrates a nuanced understanding of the complexities inherent in construction projects. The application of SOS further enhances the adaptability and prediction accuracy of the hybrid model, as evidenced by its superior performance in real-world case studies and comparative analyses.

The notable results, including a substantial decrease in Mean Absolute Percentage Error (MAPE) from 20% to 10%, underscore the transformative potential of SOS-NN-LSTM in addressing the prevalent issue of cost overruns. Beyond risk mitigation, the model contributes to more informed decision-making and robust management strategies in the construction industry. This research not only emphasizes the effectiveness of SOS-NN-LSTM but also positions it as a pioneering advancement, poised to reshape cost estimation practices and enhance the financial viability of construction projects.

V. References

Abioye, S. O., Oyedele, L. O., Akanbi, L., Ajayi, A., Davila Delgado, J. M., Bilal, M., Akinade, O. O., & Ahmed, A. (2021). Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges. *Journal of Building Engineering*, *44*, 103299. https://doi.org/https://doi.org/10.1016/j.jobe.2021.103299

Alekhya, G. B. S., Shashikanth, K., & Prasad, M. A. (2022). Risk assessment of cost overrun using fuzzy logic model. *Materials Today: Proceedings*, *62*, 1803-1810. https://doi.org/https://doi.org/10.1016/j.matpr.2021.12.415

Barnes, M. (1988). Construction project management. *International Journal of Project Management*, *6*(2), 69-79. https://doi.org/https://doi.org/10.1016/0263-7863(88)90028-2

Cheng, M.-Y., Cao, M.-T., & Herianto, J. G. (2020). Symbiotic organisms search-optimized deep learning technique for mapping construction cash flow considering complexity of project. *Chaos, Solitons & Fractals*, *138*, 109869.

Cheng, M.-Y., Chang, Y.-H., & Korir, D. (2019). Novel approach to estimating schedule to completion in construction projects using sequence and nonsequence learning. *Journal of Construction Engineering and Management*, *145*(11), 04019072.

Cheng, M.-Y., & Prayogo, D. (2014). Symbiotic Organisms Search: A new metaheuristic optimization algorithm. *Computers & Structures*, *139*, 98-112. https://doi.org/https://doi.org/10.1016/j.compstruc.2014.03.007

Ezugwu, A. E., & Prayogo, D. (2019). Symbiotic organisms search algorithm: Theory, recent advances and applications. *Expert Systems with Applications*, *119*, 184-209. https://doi.org/https://doi.org/10.1016/j.eswa.2018.10.045

Flyvbjerg, B., Skamris, M., & Buhl, S. (2002). Underestimating Costs in Public Works Projects: Error or Lie? *Journal of the American Planning Association*, *68*, 279-295. https://doi.org/10.1080/01944360208976269.

Gunawan, L. S. M., Cheng, M.-Y., & Prayogo, D. (2019). Optimized Management Strategy For Construction Projects Considering The Trade-Off Of Estimate Schedule And Cost At Completion. *Dimensi Utama Teknik Sipil*, *6*(2), 26-37. https://doi.org/10.9744/duts.6.2.26-37

Musarat, M. A., Alaloul, W. S., & Liew, M. S. (2021). Impact of inflation rate on construction projects budget: A review. *Ain Shams Engineering Journal*, *12*(1), 407-414. https://doi.org/https://doi.org/10.1016/j.asej.2020.04.009

Rafiei, M. H., & Adeli, H. (2018). Novel machine-learning model for estimating construction costs considering economic variables and indexes. *Journal of Construction Engineering and Management*, *144*(12), 04018106.

Shoar, S., Chileshe, N., & Edwards, J. D. (2022). Machine learning-aided engineering services' cost overruns prediction in high-rise residential building projects: Application of random forest regression. *Journal of Building Engineering*, *50*, 104102. https://doi.org/https://doi.org/10.1016/j.jobe.2022.104102