*International Journal of Innovative Computing, Information and Control*
Volume **22**, Number **1**, February **2026**

 ISSN 1349-4198
pp. **129–143**

# EXPLAINABLE LEARNING MODEL FOR COST ESTIMATION USING TIME-DRIVEN ACTIVITY-BASED COSTING

Theethawat Savastham* and Nikom Suvonvorn

Department of Computer Engineering
Prince of Songkla University
15 Kanchanavanich Road, Hatyai, Songkhla 90110, Thailand
nikom.s@psu.ac.th
*Corresponding author: 6510120026@email.psu.ac.th

Received July 2025; revised October 2025

Abstract. *Cost is an essential part of profit calculation. Many studies use machine learning to estimate production costs, but interpretability is a challenge. In contrast, regression and curve fitting methods are rich in interpretable information, but face challenges with complex calculations. This paper aims to present a hybrid machine learning and curve fitting approach. Replacing traditional Cost Estimation Relationships (CERs) with Artificial Neural Network (ANN) in parametric cost estimation, while using Time-Driven Activity-Based Costing (TDABC) to define cost parameters. The model is carefully assessed using a hybrid approach that combines simulated datasets and real production data from a crab pasteurization factory. This paper improves the earlier Time-Driven Cost Estimation Learning Model (TDCE) by adding an updated equation, activation function, and preprocessing steps. The model achieves a 9.33% validation error on a simple dataset and a 15.79% error on the actual dataset. In addition, the design allows the model's hyperparameters, such as internal weights, to be interpreted to reflect the behavior of the cost variables.*

**Keywords:** Cost estimation, Neural network-like, Time-driven activity-based costing, Explainable machine learning

1. **Introduction.** Cost is an important factor in manufacturing planning, creating a direct impact on profitability and pricing. However, the instability of the material supply chain and manufacturing resource prices makes control challenging. Cost estimation is frequently divided into qualitative and quantitative approaches [1]. The qualitative approach is divided into an intuitive approach and a similarity-based or analogical approach, while the quantitative approach includes the mathematical model-based parametric approaches and the bottom-up analytic approaches [2].

Computer-aided techniques for cost estimation can be categorized into machine learning and curve fitting. Curve fitting means constructing a curve that is the best fit to the data series [3]. It is used in various kinds of regression and least squares methods. In the cost estimation, multiple regression analysis is a particular technique for the parametric approach, followed by the other types of regression [4, 5]. These are used to define the Cost Estimation Relationships (CERs) between cost and influencing parameters. For example, Abbate et al. used Multiple Linear Regression (MLR) to develop CERs for composite aircraft components [2].

In contrast, machine learning is one of the effective cost estimation methods, as it can automatically learn the complex relationships between product characteristics and their cost [6]. For example, Zhang et al. estimated the cost of Computer Numerical Control

DOI: 10.24507/ijicic.22.01.129

(CNC) machined rotary parts using data from the 3D Computer-Aided Design (CAD) data [7]. Additionally, Hashemi et al. reported that Artificial Neural Network (ANN), which is a machine learning technique, received the largest share of the cost estimation in construction projects from 1985 to 2020 [8].

Machine learning enhances the performance of cost estimation but still has limitations, such as a black box nature and a lack of explainability [9, 10]. Additionally, the size of the dataset and overfitting are also its problems when applied to the field that has a limited amount of data in the conception phase [11]. On the other hand, the main benefit of curve fitting is its interpretability, which enables the business can take a closer look at the relationships between variables and improve their decisions [12]. The result equation is simple to replicate, and the method can be helpful in situations where there is not sufficient data [5, 9].

In recent years, many researchers have placed importance on the explainability of models under the trend of Explainable Artificial Intelligence (XAI) [13, 14]. It is important to build trust and confidence for the user. Much research employs the Shapley Additive Explanations (SHAP) for explaining the machine learning model [11, 14, 15]. SHAP is the game theory that allocates the weight of a feature to the model prediction using a mathematical model. However, SHAP also has some weaknesses in misleading feature importances, and the context between the game theory and the explainable AI is different regarding the feature of interest [16].

1.1. **Proposed technique.** The goal of this paper is to create an accurate cost estimation model with a by-default explainable model. We combine the strengths of machine learning and curve fitting to achieve this goal. This paper presents a hybrid model that employs the well-known machine learning technique of ANN, along with the curve fitting technique of parametric cost estimation. However, the parametric approach also faces the challenge of parameter selection, and up-to-date relationships should be maintained under the new production technologies [5, 17].

To solve this problem, the analytic method named Time-Driven Activity-Based Costing (TDABC) is selected as a parameter selector. TDABC is a method for cost estimation that uses time as a main cost driver [18]. It has been developed to address the weakness of traditional volume-based costing, due to the lack of management in the indirect cost of manufacturing [19]. There are many applications of TDABC, for example, Vedernikova et al. employed TDABC on the assembly line of electronic manufacturing [20], and Vyas et al. used TDABC along with the stochastic process to create a parametric equation to enhance the understanding of workstation and bottleneck costs [21]. Furthermore, they also suggested the future direction for the implementation of machine learning for predictive assistance. Additionally, the enhanced version of it, like FL-TDABC (Fuzzy Logic Integrated TDABC) from Koster et al., can offer a more precise and reliable estimated cycle cost of actual healthcare nature compared to the original one [22]. Given these reasons, the researchers select the TDABC to work with ANN and the parametric cost estimation to propose a specific learning model for cost estimation.

1.2. **Scope.** The research scope is based on extending and implementing our previous TDCE model into the real-world nature of cost estimation [23]. The datasets were collected from crab pasteurization manufacturing. The study uses the discrete interval of time and examines the drivers of material, labor, and utility costs for water and electricity supplied under the combination of all production steps into one step per product.

The paper starts with the presentation of the problem statement, then continues with the literature review of the related theories, the proposed method, the evaluation experiment procedure, the experiment results, behavior discussions, and concludes with the conclusion.

## 2. Related Theories.

This section explains the selected theories of TDABC, ANN, and crab pasteurized manufacturing.

### 2.1. Time-driven activity-based costing.

Kaplan and Anderson invented TDABC in 2003 [24]. TDABC divides manufacturing costs into three categories: Direct Material Cost ($DMC$), Direct Labor Cost ($DLC$), and factory overhead, which covers all departmental devices, equipment, laborers, investments, utilities, machinery, and durable products. TDABC is an analytic technique for cost estimation, but its properties can also be expressed as an equation. Namazi presented the parametric form of TDABC in Equation (1) [25].

$$Total\ Cost = DMC + DLC + \sum_{n=1}^{N} \sum_{i=1}^{I} \sum_{k=1}^{K} t_{i,k} \times C_n \tag{1}$$

Overhead cost comes from the summation of multiplication results between the time spent ($t_{i,k}$) on the activity ($i$) and the cost ($C_n$) of the resource pool ($n$), where $N$ is the number of resource pools, $I$ is the number of activities, and $K$ is the number of events. The main strengths of TDABC are straightforward calculations, flexibility, and transparency [26].

### 2.2. Artificial neural network.

An ANN is a system that consists of neurons receiving signals with different intensities or weights. All weighted input signals are added together to create the result, which can then be shaped into the desired range using the activation function [27]. The general equation of an ANN can be displayed as Equation (2) where $x_i$ is the input value of data $i$, $w_i$ is the weight or intensity of data $i$, $b$ is a system bias, $f$ is an activation function, and $y$ is the output of a neuron.

$$y = f\left(\sum_{i=1}^{n} x_i w_i + b\right) \tag{2}$$

### 2.3. Crab manufacturing cost estimation.

The research focuses on the crab pasteurization manufacturing process at Viyacrab Product Company Limited, Surat Thani province, Thailand. This company is selected because it has sample cost drivers, and its processing steps are commonly found in the real-world production line.

The example procedure can be illustrated as in Figure 1. The workflow starts by receiving the raw crab and categorizing it according to size (A-sized crab, C-sized crab, small crab, and watery crab), followed by boiling, freezing, picking, packing, and pasteurizing before storing them and waiting for the delivery. In each process, different costs were obtained, such as the labor cost of office and warehouse workers during receiving or the electricity cost during freezing in a cold storage. The company's cost estimation includes three components: material costs, labor expenses for picking and packing, and utility fees for electricity and water. Material costs are based on the total cost multiplied by its weight, excluding waste. Labor costs are calculated by weight during the picking and packing process. Utility costs are calculated by dividing the total amount spent during the previous billing cycle by the total weight of the products during that cycle.

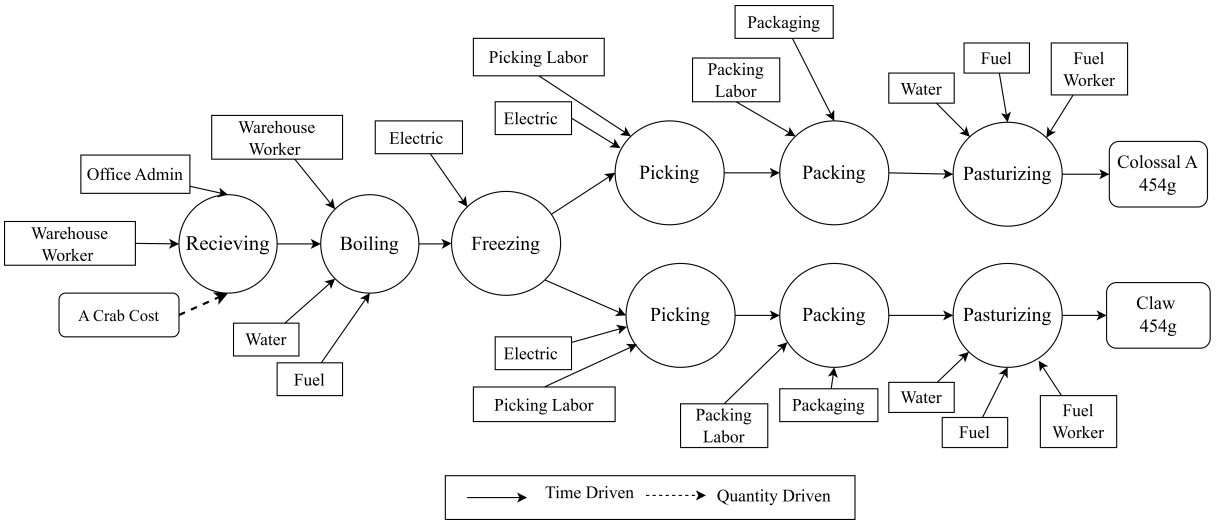


FIGURE 1. Production line diagram of crab manufacturing cost by production step

3. **Methodology.** The authors introduced the TDCE model, which adapted the TDABC equation to align with the standard structure of the ANN [23]. This model establishes a direct correlation between the weights and biases and the significance of each cost driver. However, it has mainly been tested in a simulated environment. This paper presents an enhanced version of the TDCE model, refined to address the complexities inherent in the actual experiment.

Based on the framework of TDABC, our scoped case study has a direct material cost, which is raw crab, and there are no direct labor costs; all the labor and utility costs are treated as factory overhead, which is the costed resource that is combined into the activity. However, under this scope, the cost is calculated under the integration of all manufacturing processes into a single step, which consumes resources from direct material, labor expense, and utility cost. Figure 2 illustrates the general form of the TDABC equation that includes direct material cost, direct labor cost, and various types of cost drivers on the left, and on the right, this is the case-study-specific model under the scope of study.

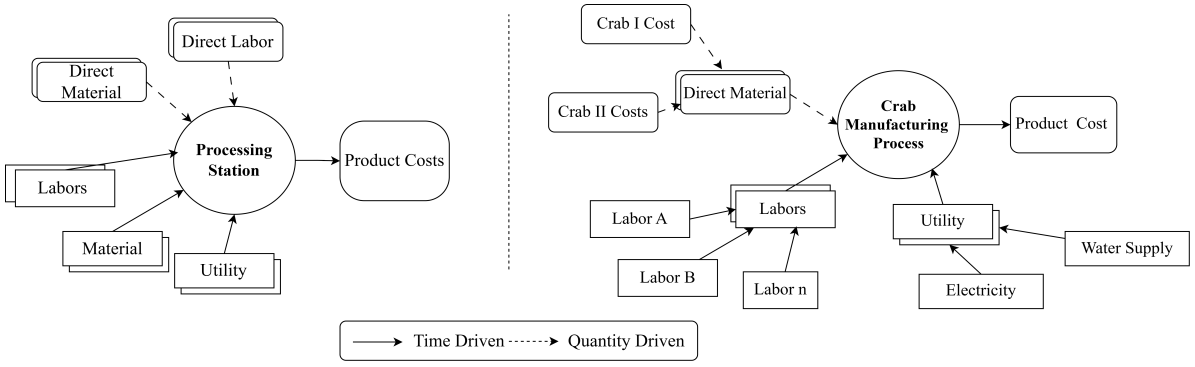


FIGURE 2. General and current case study view for each station

3.1. **Equation modeling.** The goal of this process is to combine the TDABC equation and the general ANN equation. Previously, TDABC defined the activity in two dimensions, which are events and activities. To simplify this, the dimensions of activity and event in the equations can be combined, and according to the scope, the dimension is also

reduced because only a single activity is performed in the manufacturing process. In our case study, the direct labor cost is not included as a cost driver, so the new total cost equation can be shown as Equation (3), where the direct material cost can be calculated by summing the product of the unit cost and the quantity of each material. In addition, the parametric equation of cost for all materials can be displayed as $\sum CM(I)$ in Equation (4), where $cm_i$ is the material unit cost for the material $i$, and $q_i$ is the quantity of material $i$.

$$Total\ Cost = DMC + \sum_{i=1}^{I} t_i \cdot C_i \tag{3}$$

$$\sum CM(I) = \sum_{i=1}^{I} cm_i q_i \tag{4}$$

Regarding labor cost, this model follows the TDABC's principle of practical work time, which is around 80-85% of their theoretical workload, and this study assumed it to be 80% [24]. The unit cost of each labor unit can be displayed in Equation (5), and the summation of labor cost $\sum CL(J)$ can be explained in Equation (6).

$$Unit\ Cost_{labor} = \frac{wage}{days} \times \frac{1\ day}{working\ hour} \times \frac{1\ hour}{60\ min} \times 0.8 \tag{5}$$

$$\sum CL(J) = \frac{1}{75HL} \sum_{j=1}^{J} \frac{cl_j tl_j}{dl_j} \tag{6}$$

where $cl_j$ is the cost of labor $j$ in a day, for $dl_j$ days, $HL$ is the number of work hours a day, and $tl_j$ is the amount of time performed by this labor.

Regarding the utility cost, the remaining share of the factory overhead that comes from the water and electricity supply is broken down into minutes without the use of the practical capacity term. The equation for single unit cost is shown in Equation (7), while the summation of the utility cost can be displayed as $\sum CU(K)$ in Equation (8).

$$Unit\ Cost_{utility} = \frac{cost}{affected\ day\ amount} \times \frac{1\ day}{working\ hour} \times \frac{1\ hour}{60\ min} \tag{7}$$

$$\sum CU(K) = \frac{1}{60HU} \sum_{k=1}^{K} \frac{cu_k}{du_k} tu_k \tag{8}$$

$$Total\ Cost = \sum_{i=1}^{I} cm_i q_i + \frac{1}{75HL} \sum_{j=1}^{J} \frac{cl_j tl_j}{dl_j} + \frac{1}{60HU} \sum_{k=1}^{K} \frac{cu_k}{du_k} tu_k \tag{9}$$

where $cu_k$ is the cost of the object $l$ for all $L$ utility cost objects, $du_k$ is the number of affected days of cost (like the lifetime of a machine in days, or the rounded utility bill in days), $tu_k$ is the performed duration, and $HU$ is the working hours per day of that cost object. After defining all the terms of the equation, both time-driven and non-time-driven, the combined equation of cost is displayed in Equation (9).

This equation summarizes the source of cost in the TDABC's framework; however, it cannot fully represent the manufacturing cost in a real situation because it comes from the combination of many processes, and it does not include other hidden costs. To overcome this problem, the ANN-style of hidden weights and biases is assigned to each term of the equation.

Define $w_{m,i}$, $w_{l,j}$ and $w_{u,k}$ as weights of each material, labor, and utility cost object; moreover, add $w_1$, $w_2$ and $w_3$ to be the weight of the term of total material, labor, and

utility costs. Furthermore, define $b_m$, $b_l$, and $b_u$ to be the bias of each element, representing the initial fixed cost for each cost object, while $b$ represents the initial fixed cost of the whole production process. To make it more robust for real-world data, the activation function named Leaky Rectified Linear Unit (Leaky ReLU) is also applied, to ensure that the negative inputs are mapped to a small non-zero and non-negative value [28]. The definition of Leaky ReLU can be displayed in Equation (10), where $k$ known as the leaky value is set to be 0.01 for this experiment. Finally, the new equation for total cost can be displayed as Equation (11).

$$LeakyReLU(x) = \begin{cases} x, & \text{if } x > 0 \\ kx, & \text{if } x \leq 0 \end{cases} \tag{10}$$

$$\begin{aligned} Total\ Cost = {} & w_1 \cdot LeakyReLU\left(\sum_{i=1}^{I} cm_i q_i w_{m,i} + b_m\right) \\ & + w_2 \cdot LeakyReLU\left(\frac{1}{75HL}\sum_{j=1}^{J}\frac{cl_j tl_j w_{l,j}}{dl_j} + b_l\right) \\ & + w_3 \cdot LeakyReLU\left(\frac{1}{60HU}\sum_{k=1}^{K}\frac{cu_k w_{u,k}}{du_k} tu_l + b_u\right) + b \end{aligned} \tag{11}$$

Compared to the general form of ANN and our model equation, it forms three neural network perceptrons that represent the material cost, labor cost, and utility cost, and converge into the central neuron of the cost combination. The illustration of the model can be displayed in Figure 3.

The model has four levels. The input level contains all of the input data, the model-element level contains the individual processing units of each cost category, the model level receives and produces the output signal of each processing unit, and the output level represents the manufacturing cost as a single numerical value.

3.2. **Training and weight adjustment.** The training begins by initializing all weights and biases. Then, on each iteration, the model receives all the input data and produces its predictions. After that, each prediction result and the actual one are compared to get the iteration errors using the Mean Square Error (MSE) and Root Mean Square Percentage Error (RMSPE) as displayed in Equations (12) and (13) [29, 30].

$$MSE = \sum_{i=1}^{n}\frac{(\hat{y}_i - y_i)^2}{n} \tag{12}$$

$$RMSPE\ (\%) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{y_i}\right)^2 \cdot 100\%} \tag{13}$$

This model employs the backpropagation technique named Gradient Descent with minor modifications to adjust each weight at both the model and model-element levels. According to the chain rule of calculus, the model-element-level weights can be adjusted using Equations (14) and (15), while the model-element-level biases can be adjusted using Equations (16) and (17) under the notation of $\alpha$ as the learning rate of the model-element level. The weight updating function is different from the general gradient descent. The max function is used to ensure the new weight of each cost item at the model-element

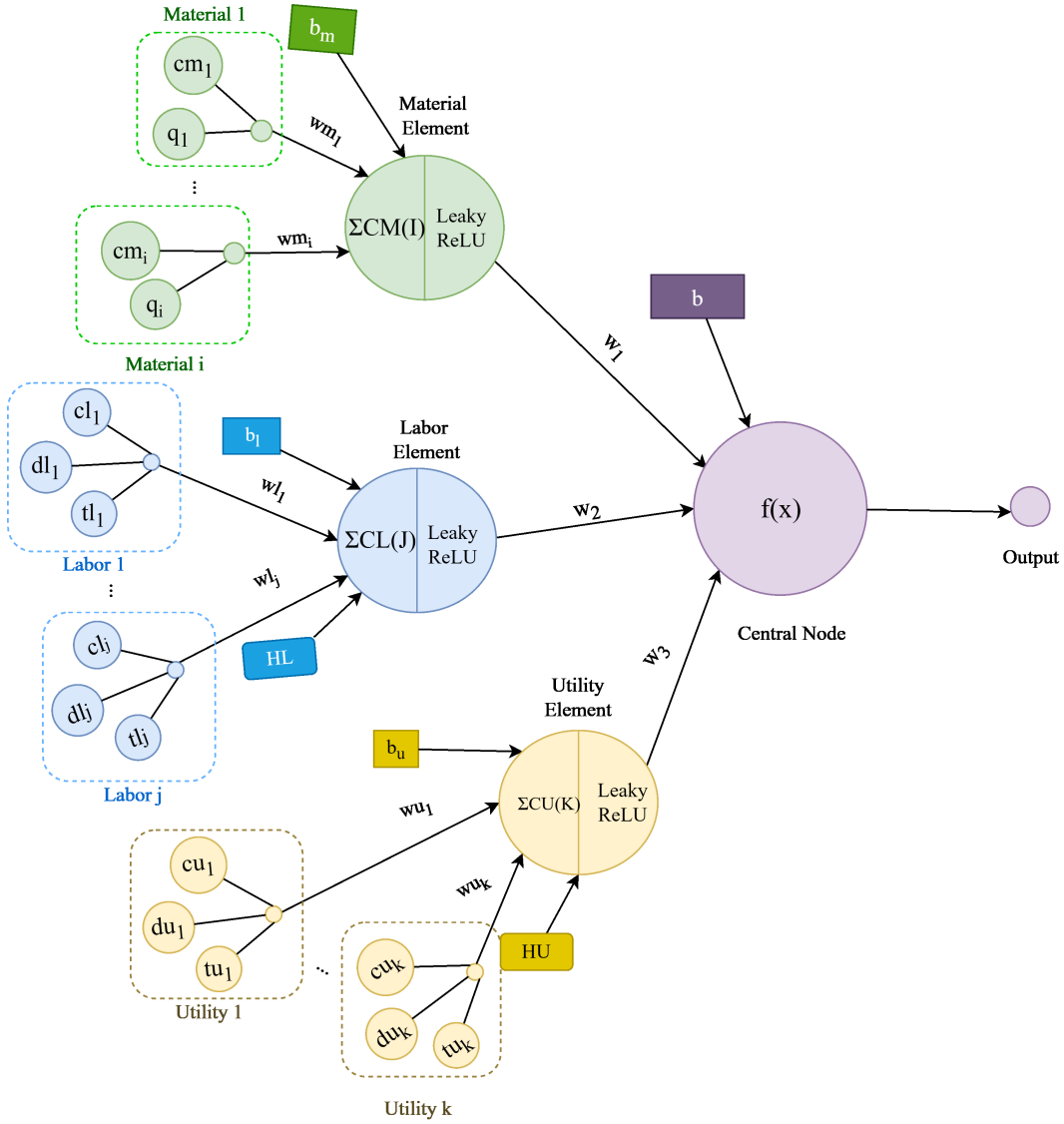


FIGURE 3. Model architecture of the extended version TDCE model

level does not produce a negative input.

$$w_{x,i,new} = \max(w_{x,i} - \alpha \nabla f(w_{x,i}), 0) \tag{14}$$

$$\nabla f(w_{x,i}) = w_{associated} \cdot MSE' \cdot LeakyReLU'(Input) \cdot w_x coefficient \tag{15}$$

$$b_{x,new} = b_x - \alpha \nabla f(b_x) \tag{16}$$

$$\nabla f(b_x) = w_{associated} \cdot MSE' \cdot LeakyReLU'(Input_x) \tag{17}$$

On the other hand, the model-level updated function can be displayed in Equations (18) and (19), in which $\beta$ refers to the learning rate of the model level, which can be different from the model-element level.

$$w_{x,new} = w_x - \beta(MSE' \cdot LeakyReLU(Input_x)) \tag{18}$$

$$b_{new} = b - \beta(MSE') \tag{19}$$

4. **Model Evaluation.** The performance evaluation of this model is performed by testing in four scenarios with different preprocessing pipelines: baseline experiment with the

original dataset, outlier removal, data augmentation, and the combined approach of these two techniques.

4.1. **Data preprocessing.** In all the scenarios, input data were normalized using Min-Max normalization with a modification to extend the margin of training data on both the max and min sides. The formula of normalization can be displayed as Equation (20).

$$X_{norm} = \frac{X_{old} - X_{\min} + 0.15(X_{\max} - X_{\min})}{X_{\max} - X_{\min} + 0.3(X_{\max} - X_{\min})} \tag{20}$$

For the outlier removal, the Interquartile Range (IQR) is selected to be a method to annotate outliers instead of the general Z-score method because the input data does not follow a normal distribution. 1.5 times the IQR is used as a threshold. Data augmentation is employed to balance the representation of the four material types. This is done by randomly selecting the input records.

4.2. **Datasets.** This experiment uses both simulated and real data, though the real dataset does not cover all of the experiment's goals. A controlled variation of the actual-inspired simulation is used in three of the simulated datasets.

According to a report on the average price per kilogram of raw blue swimming crab for fishermen from May 2015 to October 2016, 43 THB is the highest difference between the highest and lowest prices in each category [31]. The first simulated dataset is the simple dataset, where the material cost fluctuates randomly from the actual data record within the range of 0 to 30 THB, whereas in the second one, the high-variation simple dataset, the cost swings randomly between 0 and 60 THB. In addition, the other data in the high-variation simple dataset fluctuates between 30 and 60%, whereas the other data in the simple dataset fluctuates randomly between 0 and 10% from the real data record. Last but not least, the complicated dataset is a simulation in which the output of the data is linked to the outcome of the simple dataset; however, the input data oscillates to produce high-dimensional data.

Three of the four datasets in each set are inputs (material, labor, and utility cost usage), while the last set is the output data, also known as the process dataset.

Material usage contains the cost and quantity. The cost of each employee is included in the labor dataset along with the contract working quantity, the performed duration, and the contractual quantity (e.g., per day or month). However, this experiment cannot record the exact duration of our experiment; instead, the duration is calculated by the experimenters using the average time spent on a unit product activity, and the utility cost is estimated based on the interviews with the factory production experts. Lastly, the summarized cost of the product determined by the original factory method is contained in the process dataset.

The summarized variation of each dataset is presented in Tables 1 and 2. The amount represents the number of records in each dataset, the type refers to the unique class of data in the dataset, the Coefficient of Variation (CV) represents the ratio of its standard deviation to its mean, and the IQR represents the range between the first and the third quartile of each dataset.

Figure 4 illustrates the distribution of the output costs for each dataset; the high-variation simple dataset displays the highest range of spreading, and only a small portion of the data is removed when the outlier removal is applied.

4.3. **Experiment setup.** The dataset is split into 70% for training and 30% for validation. The experiment is on the model-element-level learning rate ($\alpha$) of 0.005, 0.01, 0.05, 0.1, and 0.5, and model-level learning rate ($\beta$) of $1 \times 10^{-7}$ and $1 \times 10^{-8}$. The use

TABLE 1. Data variation of each dataset in the simple and high-variation simple datasets

| Data | Simple dataset | | | | High-variation simple dataset | | | |
|---|---|---|---|---|---|---|---|---|
| | Amount | Type | CV | IQR | Amount | Type | CV | IQR |
| Total cost | 194 | 194 | 0.90 | 6,070.71 | 155 | 155 | 1.36 | 15,637.40 |
| Material unit cost | 194 | 4 | 0.13 | 35.00 | 283 | 4 | 0.15 | 43.00 |
| Material amount | 194 | 4 | 0.95 | 10.11 | 283 | 4 | 1.34 | 28.29 |
| Labor unit cost | 370 | 2 | 0.23 | 360.00 | 247 | 2 | 0.24 | 360.00 |
| Labor duration | 370 | 2 | 1.07 | 174.51 | 247 | 2 | 1.43 | 527.12 |
| Labor day amount | 370 | 2 | – | – | 247 | 2 | – | – |
| Utility cost | 740 | 2 | 0.94 | 97,884.00 | 494 | 2 | 0.84 | 12,959.00 |
| Utility cost day amount | 740 | 2 | 0.60 | 18.00 | 494 | 2 | 0.93 | 25.00 |
| Utility cost duration | 740 | 2 | 1.07 | 176.30 | 494 | 2 | 1.43 | 529.27 |

TABLE 2. Data variation of actual and complicated datasets

| Data | Actual dataset | | | | Complicated dataset | | | |
|---|---|---|---|---|---|---|---|---|
| | Amount | Type | CV | IQR | Amount | Type | CV | IQR |
| Total cost | 108 | 108 | 1.26 | 4,018.16 | 194 | 194 | 0.90 | 6,070.71 |
| Material unit cost | 105 | 3 | 0.37 | 120.00 | 194 | 4 | 0.21 | 62.24 |
| Material amount | 105 | 3 | 1.17 | 3.60 | 194 | 4 | 0.95 | 10.23 |
| Labor unit cost | 225 | 2 | 0.23 | 360.00 | 370 | 24 | 1.11 | 23,593.67 |
| Labor duration | 225 | 2 | 1.41 | 101.40 | 370 | 24 | 1.11 | 173.75 |
| Labor day amount | 225 | 2 | – | – | 370 | 24 | 1.13 | 6.00 |
| Utility cost | 450 | 2 | 0.99 | 397,000.00 | 740 | 2 | 0.94 | 85,387.92 |
| Utility cost day amount | 450 | 2 | – | – | 740 | 2 | 0.60 | 18.00 |
| Utility cost duration | 450 | 2 | 1.41 | 101.40 | 740 | 2 | 1.08 | 176.63 |

of fine-tuning the learning rate at the model level is necessary because this level receives unnormalized data from the element level, whereas the input level provides normalized data to the model-element level.

5. **Results.** Table 3 represents the validation error (in RMSPE) for different preprocessing scenarios and learning rates to display the model's performance under the different configurations. The results that exhibit overfitting and underfitting are removed from the table. Learning rate format is denoted as "$\beta/\alpha$", referring to the model-level learning

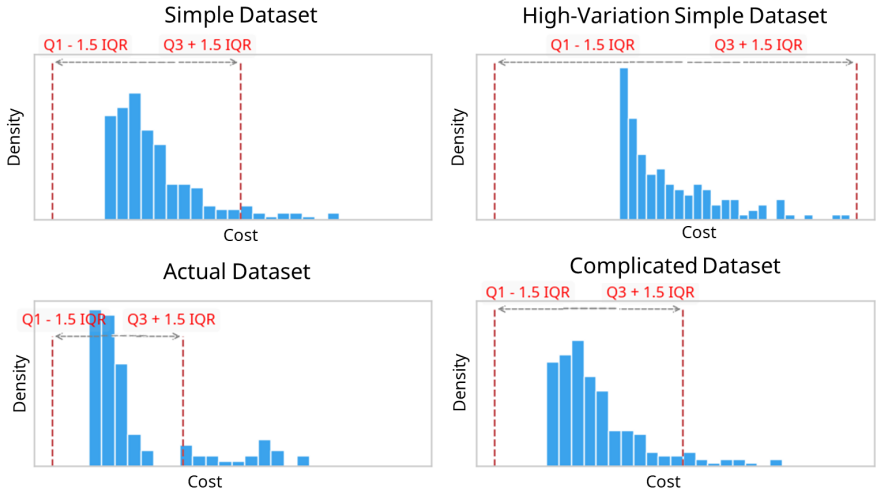


FIGURE 4. Histogram of total output cost dispersion in each dataset

TABLE 3. Validation error in RMSPE on each experiment scenario

| **Dataset** | **Scenario** | $\mathbf{1 \times 10^{-7}}$ | | | | | $\mathbf{1 \times 10^{-8}}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.005 | 0.01 | 0.05 | 0.1 | 0.5 | 0.005 | 0.01 | 0.05 | 0.1 | 0.5 |
| Simple | Original | 11.32 | 13.63 | 44.12 | 32.36 | 99.57 | 10.27 | 12.25 | 9.38 | 10.92 | 26.24 |
| | OR | 10.77 | 10.58 | 14.51 | 14.64 | – | 10.61 | 10.01 | 9.61 | 9.33 | 14.18 |
| | AU | 10.72 | 12.05 | – | 27.18 | – | 10.72 | 11.01 | 12.33 | 14.63 | – |
| | OR+AU | 11.78 | 10.44 | 12.64 | 21.78 | – | 10.77 | 9.82 | 11.02 | 11.54 | 15.07 |
| High-Variation Simple | Original | 171.31 | 131.27 | – | – | – | 47.76 | 43.41 | – | – | – |
| | OR | 56.98 | 64.94 | – | – | – | 39.89 | 45.89 | 47.22 | – | – |
| | AU | 1041.85 | 477.17 | – | – | – | 44.87 | 45.05 | – | – | – |
| | OR+AU | 46.31 | 48.57 | – | – | – | 47.70 | 42.39 | 52.60 | – | – |
| Actual | Original | 31.70 | 31.56 | 97.83 | 53.72 | – | 23.34 | 22.29 | 23.15 | 34.35 | 92.36 |
| | OR | 18.09 | 18.23 | 19.85 | 28.26 | – | 20.21 | 18.12 | 16.57 | 19.99 | 21.64 |
| | AU | 24.54 | 30.07 | 47.61 | – | – | 21.62 | 21.72 | 26.79 | 32.73 | 31.89 |
| | OR+AU | 18.53 | 15.76 | 29.40 | 39.60 | 29.86 | 16.15 | 16.59 | 16.65 | 21.08 | 23.44 |
| Complicated | Original | 30.63 | 29.12 | 73.49 | – | – | 25.90 | 23.40 | 27.50 | 31.79 | 59.48 |
| | OR | 25.40 | 29.39 | 32.09 | – | – | 23.80 | 25.91 | 26.48 | 30.34 | 39.74 |
| | AU | 29.35 | 32.81 | – | 29.78 | 30.10 | 23.46 | 26.15 | 25.62 | 30.59 | 32.10 |
| | OR+AU | 30.89 | 27.93 | – | 24.63 | 30.50 | 26.76 | 26.78 | 30.12 | 29.35 | 32.70 |

OR: Outlier Removal; AU: Data Augmentation
OR+AU: Combined Approach of Outlier Removal + Data Augmentation

rate $\beta$ and element-level learning rate $\alpha$. The outlier removal is represented by $OR$, data augmentation by $AU$, and the combination of them by $OR{+}AU$.

6. **Discussions.** The results indicate that using the model-level learning rate of $1 \times 10^{-8}$ allows the model to converge more effectively to the minimum point compared to using the model-level learning rate of $1 \times 10^{-7}$. On the same model learning rate, the optimal point that creates the accurate value is located around 0.01-0.1. On the simple dataset, the lowest achievable error is 9.33% with a learning rate of $1 \times 10^{-8}/0.1$, the high-variation simple dataset achieves 39.89% on a learning rate of $1 \times 10^{-8}/0.005$, the actual dataset yields 15.79% error with a learning rate of $1 \times 10^{-7}/0.01$, and finally, the complicated dataset delivers 23.40% error on $1 \times 10^{-8}/0.01$.

6.1. **Training behavior.** During the training, error adjustments are recorded for each iteration; an example of a learning curve on some element-level learning rate at the model-level learning rate $1 \times 10^{-8}$ can be displayed as Figure 5.

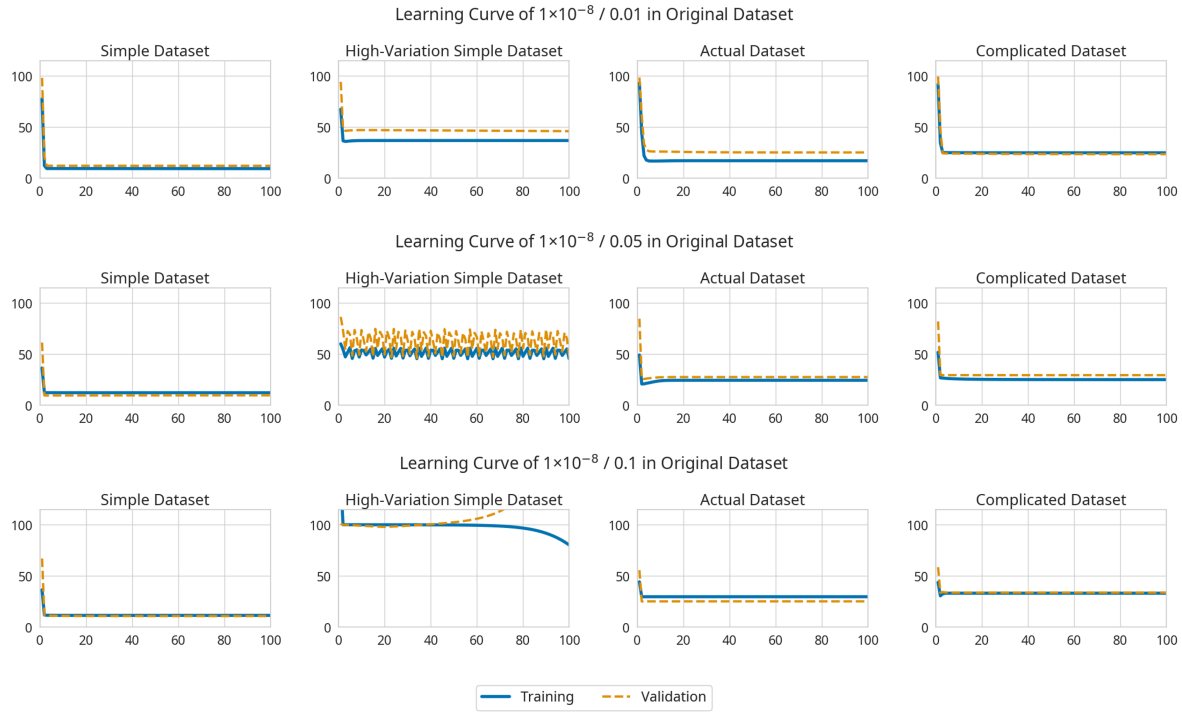


FIGURE 5. Learning curve of the model on the model-level learning rate $1 \times 10^{-8}$ on the original dataset

In the original dataset scenario, with the optimal element-level learning rate, the model can quickly find the optimal point using a small number of iterations with best-fit behavior. However, the model is unable to produce the desired result when applied to more oscillated datasets, such as the actual dataset or the high-variation simple dataset. For instance, the learning curve fluctuates when an element-level learning rate of 0.05 is used, overfitting happens while a learning rate of 0.1 is used, and the model is unable to converge to a small error value when a learning rate of 0.01 is used. Without the extra preprocessing, only the complicated can reach the maximum ability of the model. To improve accuracy, the preprocessing pipeline must be used for the remaining datasets.

Outlier removal effectively reduces the dataset variation and errors. Individual data augmentation can yield little benefit or even have the opposite effect, but it becomes effective when the data is high-dimensional. Lastly, preprocessing with the combined approach can improve the point at which data augmentation takes effect, but it usually has the same effect as removing outliers.

6.2. **Weight adjustment.** The model hyperparameters can be straightforwardly interpreted based on the model's design. The model-level weight can be interpreted as the intensity of importance of that category of cost to the overall cost, and the element-level weight can represent how influential that cost object is to the total cost. For example, the weight adjustment behavior of a model with a learning rate of $1 \times 10^{-8}/0.01$ can be displayed in Figure 6.

When compared to other model-level weights that are almost identical, the material element's weight is the highest and most variable, according to the model-level weight graph. It means that material is the most influential factor in the overall cost and the


FIGURE 6. Weight adjustment behavior using the learning rate of $1 \times 10^{-8}/0.01$

most important factor to monitor for changes in suppliers or other issues related to its cost. On the element-level weight, focus on the actual dataset; the A-size crab is ranked at the highest both in value and fluctuation. It refers to the real-world situation that A-size crab is the largest crab size; it can produce the most variety of product types, leading to a variation in cost across different products.

Even if evaluation with the actual-inspired simulated data might not capture all distinctions as in the real data, it can still prove that with our model's fundamental design, it can leverage its interpretability and be able to work with a small amount of data.

6.3. **Performance comparison.** Corresponding to the existing technique of ANN, the proposed model overcomes the black-box limitation of general-purpose ANN, with the ability to directly interpret the meaning of weights on the influence of the cost factor. The proposed model offers a clear and less complicated structure compared to the rich architecture of deep learning for the complex tasks in the ANN, and it is also suitable for a small dataset, unlike a large dataset required for a general-purpose ANN. However, it also contains a weakness in the requirement of model initialization, and the data should be complete in its features for an accurate prediction result.

In comparison between the proposed TDCE model and the traditional version of TDABC, both models offer transparent calculation, but the traditional one might not fully represent all the situations of the production process due to the hidden cost or undefined variables. TDCE solves this problem by introducing the ANN style of weights and biases, but it requires pre-training to acknowledge this value.

7. **Conclusions.** In order to create an effective and transparent model specifically for cost estimation, our paper improves the structure and expands the scope of our TDCE model, which is a hybrid combination of a machine learning model represented by an ANN and the curve fitting represented by the parametric cost estimation, and the TDABC. Then, we evaluate the model using the crab pasteurized manufacturing dataset.

Based on the experiment, the proposed model produces accurate results, which is 15.79% error with the actual dataset. For the simulated dataset, with low-complexity

data, it can achieve 9.33%, the complicated dataset with 23.40% error, and 39.89% error on the high-variation simple dataset. In low-dimensional datasets like simple and high-variation simple datasets, the outlier removal scenario offers the best accuracy. For the actual dataset to be as accurate as possible, both preprocessing steps must be used. In closing, the high-dimensional, complicated dataset performed best when none of the extra preprocessing pipelines were used.

Although the model does not automatically learn as a general-purpose ANN, its specific design allows it to leverage its interpretability, enabling analysis of the influence of each cost driver through model weights. The model can be trained with a small amount of data and can be used generally in cost estimation under the framework of TDABC. In the future, it can be implemented in more detail at each step of manufacturing. On that, it can acknowledge the finer relation of the cost elements in each step of production.

**Acknowledgment.** This research was funded by the Faculty of Engineering, Prince of Songkla University, under the scholarship for a former student. And the authors would like to thank Viyacrab Product Company Limited and Prince of Songkla University Science Park for the opportunity that allowed us to use their data and for their support in software implementation. The open-source part of the project source code and the simulated example dataset can be accessed at https://huggingface.co/iaecpsu-1/tdce-basic.

## REFERENCES

[1] A. Niazi, J. S. Dai, S. Balabani and L. Seneviratne, Product cost estimation: Technique classification and methodology review, *J. Manuf. Sci. Eng.*, vol.128, pp.563-575, 2006.

[2] R. Abbate, M. A. Turino, L. Morse, M. Fera, V. Mallardo and R. Macchiaroli, A cost estimation approach for aircraft design enhancement, *Int. J. Interact. Des. Manuf.*, vol.18, pp.83-96, 2023.

[3] S. B. Z. Adnan, A. A. M. Ariffin and M. Y. Misro, Curve fitting using quintic trigonometric Bézier curve, *AIP Conf. Proc.*, Bangi, vol.2266, no.1, 2020.

[4] M. Gunduz, L. O. Ugur and E. Ozturk, Parametric cost estimation system for light rail transit and metro trackworks, *Expert Syst. Appl.*, vol.38, pp.2873-2877, 2011.

[5] S. W. Yang, S.-W. Moon, H. Jang, S. Choo and S.-A. Kim, Parametric method and building information modeling-based cost estimation model for construction cost prediction in architectural planning, *Appl. Sci.*, vol.12, 2022.

[6] M. Mandolini, L. Manuguerra, M. Sartini, G. M. Lo Presti and F. Pescatori, A cost modelling methodology based on machine learning for engineered-to-order products, *Eng. Appl. Artif. Intell.*, vol.136, 2024.

[7] H. Zhang, W. Wang, S. Zhang, B. Huang, Y. Zhang, M. Wang, J. Liang et al., A novel method based on a convolutional graph neural network for manufacturing cost estimation, *J. Manuf. Syst.*, vol.65, pp.837-852, 2022.

[8] S. T. Hashemi, O. M. Ebadati and H. Kaur, Cost estimation and prediction in construction projects: A systematic review on machine learning techniques, *SN Appl. Sci.*, vol.2, no.10, 2020.

[9] M. M. I. Shamim, A. B. b. A. Hamid, T. E. Nyamasvisva and N. S. B. Rafi, Advancement of artificial intelligence in cost estimation for project management success: A systematic review of machine learning, deep learning, regression, and hybrid models, *Modelling*, vol.6, 2025.

[10] H. H. Elmousalami, Artificial intelligence and parametric construction cost estimate modeling: State-of-the-art review, *J. Constr. Eng. Manag.*, vol.146, 2019.

[11] J. Zhang, J. Yuan, A. Mahmoudi, W. Ji and Q. Fang, A data-driven framework for conceptual cost estimation of infrastructure projects using XGBoost and Bayesian optimization, *J. Asian Archit. Build. Eng.*, vol.24, no.2, pp.751-774, 2023.

[12] A. Z. A. Kadir, Y. Yusof and M. S. Wahab, Additive manufacturing cost estimation models – A classification review, *Int. J. Adv. Manuf. Technol.*, vol.107, nos.9-10, pp.4033-4053, 2020.

[13] S. Yoo and N. Kang, Explainable artificial intelligence for manufacturing cost estimation and machining feature visualization, *Expert Syst. Appl.*, vol.183, 2021.

[14] L. Chen, C. Xu, W. H. Lim, A. Sharma, S. S. Tiang, K. S. Chong, E.-S. M. El-kenawy et al., Transparent and reliable construction cost prediction using advanced machine learning and explainable AI, *Eng. Sci. Technol. Int. J.*, vol.70, 2025.

[15] F. Bodendorf and J. Franke, Synthesis of activity-based costing and deep learning to support cost management: A case study in the automotive industry, *Comput. Ind. Eng.*, vol.196, 2024.
[16] X. Huang and J. Marques-Silva, On the failings of Shapley values for explainability, *Int. J. Approx. Reason.*, vol.171, 2024.
[17] O. Al-Shamma and R. Ali, Parametric cost estimation models of civil aircraft for the preliminary aircraft design phase, *Aeronaut. J.*, vol.127, no.1308, pp.268-288, 2023.
[18] S. N. Areena and M. Y. Abu, A review on time-driven activity-based costing system in various sectors, *Journal of Modern Manufacturing Systems and Technology*, vol.2, pp.15-22, 2019.
[19] L. Siguenza-Guzman, A. Van den Abbeele, J. Vandewalle, H. Verhaaren and D. Cattrysse, Recent evolutions in costing systems: A literature review of time-driven activity-based costing, *Glob. Bus. Econ. Rev.*, vol.58, no.1, pp.34-64, 2013.
[20] O. Vedernikova, L. Siguenza-Guzman, J. Pesantez and R. Arcentales-Carrion, Time-driven activity-based costing in the assembly industry, *Australas. Account. Bus. Finance J.*, vol.14, no.4, pp.3-23, 2020.
[21] V. Vyas, P. Afonso, S. Silva and B. Boris, A stochastic costing model for manufacturing management and control, *IFAC-Pap.*, vol.55, pp.1116-1121, 2022.
[22] F. Koster, M. R. Kok, J. van der Kooij, G. Waverijn, A. Weel-Koenders and D. L. Barreto, Dealing with time estimates in hospital cost accounting: Integrating fuzzy logic into time-driven activity-based costing, *PharmacoEconomics – Open*, vol.7, no.4, pp.593-603, 2023.
[23] T. Savastham and N. Suvonvorn, Time-driven cost estimation learning model, in *Genetic and Evolutionary Computing, ICGEC 2024, Lecture Notes in Electrical Engineering*, J. S. Pan, T. T. Zin, T. W. Sung and J. C. W. Lin (eds.), Singapore, Springer, 2025.
[24] R. S. Kaplan and S. R. Anderson, Time-driven activity-based costing, *Harvard Business Review*, 2003.
[25] M. Namazi, Time-driven activity-based costing: Theory, applications and limitations, *Iran. J. Manag. Stud.*, vol.9, pp.457-482, 2016.
[26] N. F. Zamrud and M. Y. Abu, Comparative study: Activity-based costing and time-driven activity-based costing in electronic industry, *Journal of Modern Manufacturing Systems and Technology*, vol.2, pp.15-22, 2020.
[27] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach (Global Edition)*, 3rd Edition, Pearson, 2009.
[28] J. Xu, Z. Li, B. Du, M. Zhang and J. Liu, Reluplex made more practical: Leaky ReLU, *2020 IEEE Symp. Comput. Commun.*, Rennes, pp.1-7, 2020.
[29] L. Bi, A. G. Feleke and C. Guan, A review on EMG-based motor intention prediction of continuous human upper limb motion for human-robot collaboration, *Biomed. Signal Process. Control*, vol.51, pp.113-127, 2019.
[30] J. Zell, B. Bösch and G. Kändler, Estimating above-ground biomass of trees: Comparing Bayesian calibration with regression technique, *Eur. J. For. Res.*, vol.133, pp.649-660, 2014.
[31] T. Nitiratsuwan, K. Panwanitdamrong and V. Suksumjit, *Strategy for Decrease Berried Female Blue Swimming Crab (Portunus Pelagicus Linnaeus, 1758) Fishery from Small Scale Fisher: A Case Study in Trang Province*, Research Report (Technical Report), Faculty of Science and Fisheries Technology, Trang Campus, Rajamangala University of Technology Srivijaya, 2016.

## Author Biography

**Theethawat Savastham** received a Bachelor's degree in Computer Engineering from Prince of Songkla University, Thailand, in 2020.

He is currently a Master's degree student at the same university. His research interests include data-driven intelligent systems, neural networks, artificial intelligence, and software development.

**Nikom Suvonvorn** received a Ph.D. degree in Computer Science from l'Université de Paris Sud (XI), Orsay, France, in 2006. In 2003, He obtained a DEA (Diplôme d'Etudes Approfondies) on Electronic System and Information Processing (SETI) from l'Institute d'Electronique Fondamentale (IEF) at the same university. In that year, he also got another Master's degree in Computer Engineering from École Supérieure de Mécanique et d'Electricité (ESME)-Sudria Engineering School, Paris, France.

He is now an assistant professor in the Department of Computer Engineering, Prince of Songkla University, Thailand. His research interests include image processing, computer vision, machine vision, automation, and video surveillance.