Article

Hybrid Artificial Intelligence HFS-RF-PSO Model for Construction Labor Productivity Prediction and Optimization

Sara Ebrahimi, Aminah Robinson Fayek * and Vuppuluri Sumati

Department of Civil and Environmental Engineering, Hole School of Construction, University of Alberta, Edmonton, AB T6G 1H9, Canada; eb4@ualberta.ca (S.E.); sumati.vuppuluri@ualberta.ca (V.S.)
* Correspondence: aminah.robinson@ualberta.ca

Abstract: This paper presents a novel approach, using hybrid feature selection (HFS), machine learning (ML), and particle swarm optimization (PSO) to predict and optimize construction labor productivity (CLP). HFS selects factors that are most predictive of CLP to reduce the complexity of CLP data. Selected factors are used as inputs for four ML models for CLP prediction. The study results showed that random forest (RF) obtains better performance in mapping the relationship between CLP and selected factors affecting CLP, compared with the other three models. Finally, the integration of RF and PSO is developed to identify the maximum CLP value and the optimum value of each selected factor. This paper introduces a new hybrid model named HFS-RF-PSO that addresses the main limitation of existing CLP prediction studies, which is the lack of capacity to optimize CLP and its most predictive factors with respect to a construction company's preferences, such as a targeted CLP. The major contribution of this paper is the development of the hybrid HFS-RF-PSO model as a novel approach for optimizing factors that influence CLP and identifying the maximum CLP value.

1. Introduction

The construction industry is a key sector of the national economy for countries all around the world [1]. Since construction is a labor-intensive industry, poor construction labor productivity (CLP) usually causes cost and time overruns in projects [2,3]. To overcome this issue, the construction industry is constantly trying to identify CLP improvement strategies [4]. However, project managers first require a CLP model that helps them identify which factors lead to positively changing CLP and by how much [5,6]. Furthermore, the accurate prediction of CLP is essential for effective scheduling and planning prior to and during project execution [7]. CLP is a form of efficiency measure that is mainly defined as a ratio of units of output (i.e., project components) to units of input (e.g., labor work hours, and labor cost) or vice versa [7–9]. Many factors can potentially affect CLP, reducing the accuracy of a predictive model and imposing the risk of data overfitting [10]. The power of any prediction method relies on choosing the proper factors that affect the model output [11]. Since the identification of factors that influence construction productivity is essential for productivity performance improvement, different studies identified numerous factors that affect CLP [12]. Several studies used questionnaire surveys to identify factors with the greatest influence on CLP [13–19]. However, many studies notably focused on finding CLP factors, and few works are found in the literature on labor productivity prediction [20].

The most reliable estimate of productivity can be achieved using past project data because the important predictive productivity information can be extracted for future project management and planning [12].

Studies on productivity prediction can be classified into three types: statistical, simulation, and artificial intelligence (AI) techniques. Regression analysis is one of the most common statistical techniques [21]. Thomas and Sudhakumar [22] presented multiple regression models to quantify the effect of factors influencing masonry labor productivity. Hai and Tam [23] presented a multiple linear regression model as a statistical method for output prediction to evaluate the impact of 10 factor groups on CLP. Nevertheless, regression models are limited by the number of factors that can be considered as inputs [24]. System dynamics is one of the most applicable simulation techniques and is able to model a dynamic system [25]. Although system dynamics models are able to capture the probabilistic uncertainties of real-world systems, they cannot capture non-probabilistic uncertainties, that is, subjective or linguistically expressed information [26]. On the other hand, AI techniques, such as artificial neural network (ANN), are able to learn from experience to enhance their performance, adapt themselves to changes, and find patterns among datasets, which makes them effective prediction methods [27]. Accordingly, several AI techniques have been successfully applied to modeling and predicting construction productivity, which is discussed in the literature review section. CLP is affected by numerous factors that reduce the accuracy of the predictive model and impose the risk of data overfitting [10]. Feature selection (FS) is one of the important preprocessing procedures for data mining. Therefore, it is necessary to apply effective FS methods that are able to select key features affecting CLP and reject the nonessential features in order to achieve high prediction accuracy and reduce model complexity [28,29]. Filter and wrapper methods are two types of FS. Filter methods offer less computational time and work without a learning algorithm. However, they cannot consider model prediction, and their classification accuracy is lower, compared to wrapper methods [30]. Wrapper methods use a learning algorithm to evaluate the set of most suitable features. However, their applications are limited by high computational complexity. HFS methods are a combination of filter and wrapper methods and therefore, reduce the deficiencies of both methods [30,31]. Although comprehensive studies have identified CLP factors, few works have focused on applying different FS methods to CLP factors to reduce the risk of overfitting. In other words, a research gap exists regarding the development of HFS methods as an essential data cleaning process prior to CLP modeling.

Despite the wide application of predictive CLP models for project planning and control, a predictive model as a sole application cannot offer construction companies the optimum value of influencing factors for improving CLP [7]. Limsawasd and Athigakunagorn [32] developed a new discrete-event simulation for estimating the duration of activities under uncertainty, optimizing resource allocation in building construction projects, and improving productivity. However, they only considered resource allocation and activity duration as factors that influence productivity. Concretely, no study has presented a hybrid model for finding the maximum value of CLP and optimum value of each influential factor, using optimization techniques.

This study aims to fill the gap in the literature by developing a hybrid model that can identify the factors that most influence CLP as well as predict and optimize CLP and the factors influencing it. The proposed hybrid model will help project managers have more confidence in predicted CLP and be able to plan for improving each CLP factor.

The major contribution of this paper is developing a model for both predicting and optimizing CLP, using a combination of FS, AI, and evolutionary optimization techniques. To achieve this goal, this study had the following objectives: (1) identify factors that are the most predictive of CLP, using a combination of filter and wrapper methods as a hybrid feature selection (HFS) method; (2) predict CLP by developing and comparing four different predictive models, using the factors that most influence CLP; and (3) develop a novel hybrid evolutionary optimization model for finding the maximum CLP value and the optimum value of each selected factor.

The remainder of this paper is organized as follows. Section 2 presents a brief review of past research on modeling CLP. In Section 3, the proposed methodology for predicting and optimizing CLP is presented. Section 4 provides the experimental results from using the proposed methodology to predict and optimize CLP, using a data set. Finally, Section 5 presents conclusions and recommendations for future work.

2. Literature Review on Construction Productivity Modeling

Modeling CLP is challenging because it requires evaluating the impact of numerous factors simultaneously. To deal with this challenge, AI techniques, such as fuzzy logic, ANN, classifiers, learning algorithms, and hybrid techniques, are widely used in the construction management domain. Golnaraghi et al. [33] developed a CLP prediction model using ANN and compared it with other techniques, including adaptive neuro-fuzzy system (ANFIS) and radial basis function neural network. El-Gohary et al. [12] introduced the engineering approach, using ANN to map the relationship between CLP and factors influencing it. Nasirzadeh et al. [34] developed ANN-based prediction intervals to predict CLP, using historical data. Their model identified various sources of uncertainty affecting prediction. Momade et al. [35] proposed a data-driven approach, using support vector machine (SVM) and random forest (RF) to model and predict CLP. Their results showed that SVM achieved a higher rate of accuracy, compared to RF. Recently, Sarihi et al. [36] developed a comparative analysis of CLP models, using ANN, ANFIS, and logical fuzzy inference system (FIS). They found that ANFIS showed better accuracy, compared to the two other models.

However, in recent years, hybrid systems-based machine learning (ML), optimization algorithms, and simulation techniques have been applied in several construction problems because they are superior to sole AI techniques [7,37]. Gerami Seresht and Fayek [26] developed fuzzy system dynamics technique by integrating system dynamics and fuzzy logic to model the multifactor productivity of equipment-intensive activities. Tsehayae and Fayek [9] demonstrated the application of data-driven fuzzy clustering in the development of FIS. They then used genetic algorithm (GA)-based optimization to address the FIS limitation, which is the inability to learn from data. Khanzadi et al. [8] developed a hybrid simulation model by combining system dynamics and agent-based modeling to predict labor productivity by evaluating various influencing factors in a concrete placing project. Raoufi and Fayek [38] proposed the integration of fuzzy logic and agent-based modeling to predict the performance of construction crews, according to crew motivation and situational input variables. Gerami Seresht et al. [39] introduced a new fuzzy clustering algorithm, using Gustafson–Kessel's algorithm and Adam optimization to determine the number of clusters automatically and assign weights to the FIS rules to improve accuracy; they then used the proposed algorithm to predict CLP for concrete placing activities, and the results showed that the new approach improved accuracy and efficiency, compared to previous research. Although the aforementioned papers developed hybrid methods to model and predict construction productivity, very few studies have applied HFS methods as the combination of filter and wrapper FS methods in CLP prediction to find the most predictive factors. Ebrahimi et al. [10] proposed the integration of ANN and GA as a wrapper method for selecting the most influential CLP factors and then predicting CLP, using ANN. The results showed accuracy improvement, compared to previous work using filter methods. Recently, Cheng et al. [7] introduced a hybrid model, including least square SVM, symbiotic organisms search, and wrapper-based FS methods to predict construction productivity. Goodarzizad et al. [40] proposed the integration of ANN and the grasshopper optimization algorithm to identify the factors with the greatest influence on CLP. They then applied an ANN to measure CLP, using the identified factors.

One wrapper FS method that has been developed in other disciplines is the integration of SVM and GA, which shows appropriate efficiency in selecting the optimal feature subset. Fei and Min [41] developed the integration of SVM and GA to select a feature subset and optimize SVM parameters for solving binary classification problems. Furthermore,

Tao et al. [42] presented a novel approach based on GA for feature selection parameter optimization of SVM in hospitalization expense modeling. Due to the aforementioned superiority of HFS methods in the introduction, this study integrated the ReliefF algorithm as a filter method with SVM-GA as a wrapper method to develop the proposed HFS model.

Modeling the optimization process is another important challenge in CLP studies. Optimization techniques have been used in various construction domains. Jin et al. [43] proposed a workspace-based multi-objective optimization model to produce optimal solutions for scaffolding resource allocation and space planning. Lin and Lai [44] proposed a time-cost trade-off model to reduce project duration that used GA to evaluate variable productivity. Shahbazi et al. [45] presented a model, using mixed-integer nonlinear programming to allocate tasks to employees with different skill levels. However, the previous studies did not explore hybrid optimization in the area of construction labor productivity optimization by using evolutionary optimization techniques to evaluate the optimum value of factors influencing CLP.

3. Research Methodology

This paper presents an HFS method for identifying the factors that are most predictive of CLP and utilizing them as inputs of the developed CLP models, and develops a novel hybrid evolutionary optimization technique by integrating HFS, a predictive model, and particle swarm optimization (PSO) to optimize CLP and the factors that influence it. This proposed methodology for modeling, predicting, and optimizing CLP is accomplished in the following four main steps: (1) preparing CLP data, using different techniques (Section 3.1); (2) developing an HFS to reduce dimensionality and identify the factors most predictive of CLP (Section 3.2); (3) developing four algorithms widely used in predictive models, namely ANFIS, ANFIS-GA, ANN, and RF (Section 3.3); and (4) developing a hybrid optimization model, using PSO to search for the maximum CLP value and optimum value of each selected factor (Section 3.4). These four steps are presented in Figure 1.
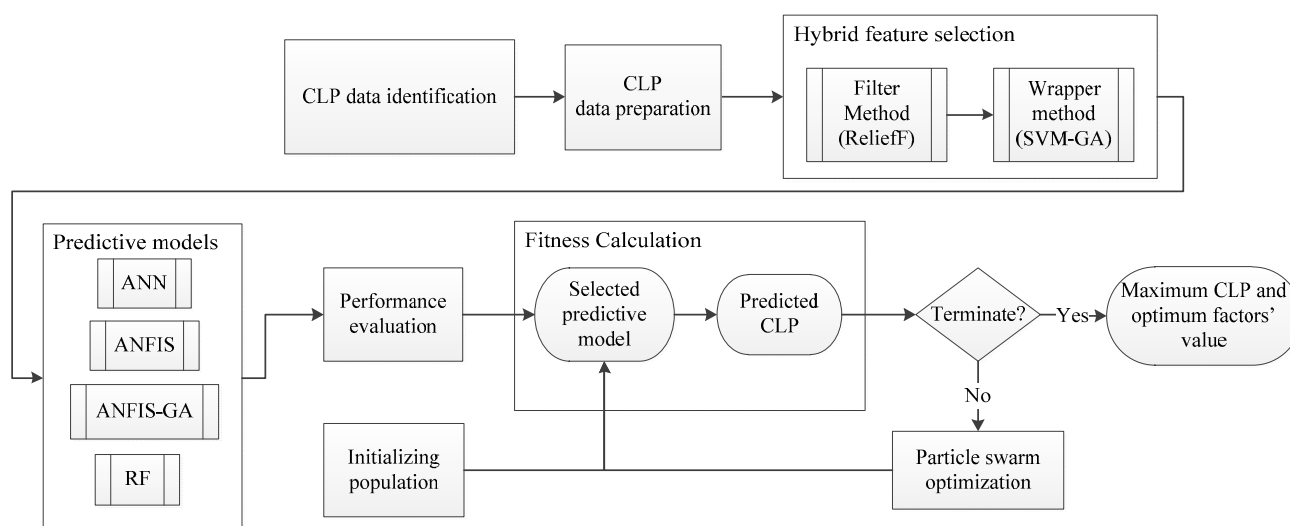

Figure 1. A general view of the proposed model for CLP prediction and optimization.

3.1. CLP Data Identification

In this study, the proposed methodology was used to predict and optimize CLP of concrete placing activities, using the data collected by Tsehayae and Fayek [9] in a previous study. Data were collected in Alberta, Canada, in four construction project contexts, including residential and commercial warehouse buildings, residential and commercial high-rise buildings, industrial buildings, and institutional buildings. A literature review conducted by Tsehayae and Fayek [13] initially identified 169 factors that influence CLP. They collected 112 factors influencing CLP for concrete placing activities over 92 days of

data collection. In this study, per Equation (1), CLP is defined as a ratio of output, which is installed quantity, to input, which is labor work hours; CLP has positive real values.

$$CLP = \frac{Installed\ quantity\ (output)}{Labor\ work\ hours\ (input)} \tag{1}$$

In the existing data set, some CLP factors are objective, such as crew size, which has a numerical measure (in terms of number of workers), while other factors are subjective, such as complexity of task, which does not have a well-defined measurement. Subjective factors were measured using a predetermined rating scale of 1–5, according to Tsehayae and Fayek [5]. CLP factors can be grouped into six levels: (1) activity, (2) project, (3) organizational, (4) provincial, (5) national, and (6) global.

3.2. CLP Data Preparation

The CLP data preparation process consists of normalization, imputing missing values, removing factors with zero variance, and eliminating outliers.

Mostly, CLP data have varying scales that lead to increased training time and biases in predictive models and affect convergence in prediction [33]. Hence, the experimental data are normalized using Equation (2) in a process called "max–min normalization", where x_{ij} is the value of instance i from factor j; x_{jmin} and x_{jmax} are the minimum and maximum values of factor j, respectively; and r_{ij} is the normalized value of instance i from factor j. Max–min normalization guarantees that all features have the exact same scale.

$$r_{ij} = \frac{x_{ij} - x_{jmin}}{x_{jmax} - x_{jmin}} \tag{2}$$

Data sets often have some missing values, due to human error or non-availability of real data. Imputation methods use ML algorithms to help estimate missing values. Based on Choudhury and Pal [46], the neural network-based imputation method is able to train a data set containing incomplete samples and identify instances similar to instances with missing values. Based on the results of several studies [46–48], neural network imputation was applied in the present study in order to impute missing values of CLP.

Standard deviation is a measure of the variance of each factor in a data set. Removing factors with no variation in data instances is a pre-processing step for data sets [49]. In this study, CLP factors with zero standard deviation were removed from the data set.

Detecting and eliminating outliers is another essential step in data preparation. Although outliers are part of a data set, they are significantly different from other observations. In this study, Tukey's method, which utilizes the median, upper, and lower quartiles of a data set, was applied as an outlier detection method [50]. Since quartiles are resistant to farthest data of the data set, Tukey's method is less sensitive, compared to methods using mean and standard variance [50].

3.3. Hybrid Feature Selection (HFS)

The developed HFS is a combination of the ReliefF algorithm as a filter method and the integration of SVM and GA as a wrapper method [51] and is utilized to identify the factors that are most predictive of CLP. The structure of three algorithms, namely ReliefF, SVM, and GA, are briefly discussed in the following sections.

3.3.1. ReliefF

Relief is a widely used filter-based FS method that identifies the best subset of features by measuring features' weights. Proposed by Kira and Rendell [52], this algorithm assigns weights to features based on correlation between features and categories, and then selects all features with weight greater than an artificial threshold. Notably, the Relief algorithm is limited to binary classification problems. To address this problem, Kononeko [53] proposed the ReliefF algorithm, which has the ability to work with multiclass problems. ReliefF is

a distance-based feature selector that uses Manhattan distance to measure weights. The evaluation criterion of ReliefF is presented in Equation (3), where: $W(f_{0,i})$ stands for the weight of the ith feature before updating; $W(f_i)$ is the updated weight of the ith feature; A is the vector of features; k is the number of nearest neighbors; m is the number of cycles; $f_{h(x_i)}$ and $f_{r(C)}$ are the value of k nearest neighbors of x_i in the same and different class, respectively; $P(C)$ is the ratio of the target samples C to the total sample; $P(class(x_i))$ is the ratio of the samples in the same class including x_i to the total samples; and $diff$ () denotes the distance of two samples on each feature in A.

$$W(f_i) = W(f_{0,i}) \quad - \frac{\sum_{j=1}^{k} diff(A, x_i, f_{h(x_i)})}{m \times k} \\ + \sum_{C \neq class(x_i)} \frac{P(C)}{1 - P(class(x_i))} \times \frac{\sum_{j=1}^{k} diff(A, x_i, f_{r(C)})}{m \times k} \tag{3}$$

This study uses Manhattan distance to measure the distance between two samples as shown in Equation (4), where $diff(A, R1, R2)$ is the difference between the sample $R1$ and $R2$ in the vector of feature A.

$$diff(A, R1, R2) = \frac{|R1 - R2|}{A_{max} - A_{min}} \tag{4}$$

In this study, ReliefF selected the most correlated CLP factors as its output. The factors selected by ReliefF were then applied as inputs to the combination of SVM and GA.

3.3.2. Support Vector Machine (SVM)

SVMs can solve linear and nonlinear problems and provide power classification results [54]. The most important advantage of SVM is that it can control overlearning and high dimensionality and decrease computational complexity and local extremum [42]. Introducing a kernel function can facilitate the solving of nonlinear problems. Types of kernel function include linear, polynomial, and sigmoid functions. The radial basis function (RBF), presented in Equation (5), is one of the most popular kernel functions because it requires only one parameter, δ, which is a free parameter with a significant effect on classification accuracy and has a lower complexity, compared with other functions [41,55]. Another essential parameter in SVM problems is C, which is the penalty factor and shows the cost of misclassification. According to the significance of C and δ on the result of SVM, they need to be optimized to obtain the desired accuracy, which can be accomplished using GA optimization.

$$K(x, y) = \frac{exp(-|x - y|^2)}{\delta^2} \tag{5}$$

3.3.3. Genetic Algorithm (GA)

GA is an adaptive heuristic search algorithm, the goal of which is finding an optimal solution [41]. GA uses a fitness function to estimate the significance of results from the evaluation step. Two GA operators, mutation and crossover functions, randomly transfer chromosomes and affect the fitness value [56]. Crossover specifies two chromosomes that will generate a new offspring chromosome. However, mutation is the process used to change genes in chromosomes from their initial state [10,57]. This study selected the four best chromosomes to be part of the next generation and used a single-point crossover and binary mutation.

GA minimizes the fitness function value, which is shown as FF and calculated for each chromosome, using Equation (6). SVM_RMSE is the root mean square error (RMSE) of an SVM model, w is a weight of the specified number of factors (n_f), s_i is '1' if the factor i is selected or '0' if the factor i is not selected, and c_i is the cost of factor i.

$$FF = SVM\_RMSE \times (1 + w \times (\sum_{i=1}^{n_f} c_i \times s_i)) \tag{6}$$

In the first step of HFS, ReliefF measures the weight of all CLP factors based on their correlations. Then, it selects all weights greater than or equal to a defined threshold as inputs of the wrapper method. In the second step, GA randomly generates the initial population, where each chromosome is an available feature subset for the problem. In the third step, using the selected factors from ReliefF as inputs, training of the SVM model begins and the RMSE of SVM is measured. Then, the *FF* calculation for each chromosome is completed. In the fourth step, if the result meets the termination criteria, the process stops; otherwise, the process goes back to the GA operation to find a better solution. Once the termination criteria are satisfied by the final generation, the iteration stops, and the final generation contains the factors that most influence CLP.

3.4. CLP Predictive Modeling

According to the literature review on past CLP modeling techniques, ANN and ANFIS have been found to perform well and thus, were chosen for this study. ANN is a suitable model for complex relationships between CLP and the factors that influence it, as these relationships cannot be obtained in a precise manner [12,24]. ANFIS models have been widely used in past CLP studies because of their superiority in being less reliant on expert knowledge and having a systematic data-driven process [36]. In order to optimize ANFIS parameters, the integration of ANFIS and GA was also developed. Another algorithm that shows accurate performance in a number of studies in other disciplines is RF, which was developed and compared with the other techniques in this study. Results from past studies show that RF is highly capable of solving non-linear classification problems, compared to other ML models [35]. As most crucial factors related to CLP do not follow a normal distribution, RF is a common ML technique in modeling construction productivity [58]. The following sections discuss the structure and components of these four widely used ML modeling techniques developed in this study.

3.4.1. Artificial Neural Network (ANN)

In the past few decades, ANN has become a popular and helpful technique for classification, clustering, pattern recognition, and prediction in many disciplines [59]. ANNs are able to deal with noisy or incomplete data and can be very effective, mostly in modeling problems where the relationships between inputs and outputs are not sufficiently known [60]. So, based on their abilities, ANN-based models can be ideal alternatives for modeling CLP. ANN consists of three types of layers: an input layer, hidden layers, and an output layer. In this study, a multilayer feedforward back-propagation network with one hidden layer was developed as an ANN model to predict CLP.

3.4.2. Adaptive Neuro Fuzzy Systems (ANFIS)

ANFIS is a hybrid technique that integrates the linguistic interpretability and fuzzy reasoning of FIS and learning capability of ANN in order to map inputs to an output [61]. In an ANFIS structure, fuzzy rules are extracted from ANN and the parameters of fuzzy membership functions are adaptively utilized during the hybrid learning process [62].

3.4.3. ANFIS-GA

The combination of ANFIS and GA is used to improve the performance of the ANFIS-based model and optimize its parameters. GA is utilized to find the optimum parameters of ANFIS.

3.4.4. Random Forest (RF)

RF could be considered an ensemble of classification and regression tree (CART) since multiple CART models are generated and used as base models [35,58]. In this approach, RF first generates several training data sets by sampling randomly from the original training data set. After generating new training data sets and before the tree splitting process, RF implements variable randomization to boost the diversity of trees. As both training data

and variable sets are generated randomly, trees in RF are different from each other and also independent [63]. Then, RF combines all trees by averaging their predictions. This joint prediction process increases accuracy and decreases large errors [64].

3.5. CLP Optimization

In the last step of the methodology, the PSO algorithm searches for the optimum values of CLP and the factors influencing it, using the predictive model proposed in the previous section.

PSO is one of the swarm intelligence-based algorithms first proposed by Kennedy and Eberhart [65]. PSO is simple to implement and is able to find solutions with acceptable accuracy, which makes it popular [66]. Each particle maintains three D-dimension vectors: position vector, velocity vector, and personal best vector. A particle retains its current position in position vector $X_i = (x_i^1, x_i^2, \ldots, x_i^D)$, for $i = 1$ to N (N = number of particles). Particles obtain their initial positions randomly in the search space. Velocity vector $V_i = (v_i^1, v_i^2, \ldots, v_i^D)$ of the ith particle is utilized to update its position, and the particle also obtains its initial value randomly. The best position attained by the ith particle is preserved in personal best vector and denoted as $Pbest_i = (Pbest_i^1, Pbest_i^2, \ldots, Pbest_i^D)$. Therefore, the swarm best position is evaluated as $Gbest = (Gbest^1, Gbest^2, \ldots, Gbest^D)$. The movement of a particle is related to updating its velocity and position attributes in the tth iteration ($t = 2, 3 \ldots$), based on Equations (7) and (8).

$$v_i^d(t+1) = wV_i^d(t) + c_1 r_1 (Pbest_i^d(t) - x_i^d(t)) + c_2 r_2 (Gbest_i^d(t) - x_i^d(t) \qquad (7)$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \qquad (8)$$

where w is the inertia weight, c_1 is the cognitive acceleration coefficient, c_2 is the social acceleration coefficient, and r_1 and r_2 are random values between 0 and 1. Figure 2 presents a flowchart of the PSO algorithm.

The objective of the optimization phase of this study contained the following goals:

- Goal 1: Predicted CLP (CLP_{Pred}) has minimum deviation from "targeted CLP" (CLP_{tgt}), as shown in Equation (9), where ω is the relative importance of Goal 1, compared to Goal 2.

$$\text{Goal 1} = \omega \times (CLP_{tgt} - CLP_{Pred})^2 \qquad (9)$$

- Goal 2: Predicted CLP factors (F_{Predi}) have minimum deviation from "average value of factors" (F_{Avgi}) in the data set, among all the possible combinations of improvement scenarios, as shown in Equation (10).

$$\text{Goal 2} = (1 - \omega) \times \sum_{i=1}^{n} (F_{Predi} - F_{Avgi})^2 \qquad (10)$$

In Goal 1, "targeted CLP" is the preferable CLP that a company tries to achieve. In this study, the value of CLP is between 0 and 1 after the normalization process, and greater CLP indicates better productivity in a project. Goal 1 tries to predict CLP considering the minimum distance from the targeted CLP.

Goal 2 tries to minimize changes in factors that most influence CLP. Companies mostly prefer minimum changes and corrective measures to achieve the preferable CLP because of the cost of implementing new strategies and corrective measures. In Goal 2, the average value of each factor is achieved from the existing CLP data set, which is discussed in Section 4. Since obtaining a value near the average value of each factor in the data set is feasible, the goal is to have a minimum distance between the average and optimum values for each factor. Therefore, the objective function is defined as in Equation (11):

$$\textit{Minimize } (Z = \omega \times (CLP_{tgt} - CLP_{Pred})^2 + \omega \times \sum_{i=1}^{n} (F_{Predi} - F_{Avgi})^2) \qquad (11)$$

where CLP_{tgt} and CLP_{Pred} are the targeted CLP and predicted CLP, respectively, n is the number of selected factors affecting CLP, F_{Predi} is the predicted value of the ith CLP factor, F_{Avgi} is the average value of the ith CLP factor in the data set, ω is the relative importance of Goal 1 compared to Goal 2, and Z is the minimum value of the objective function. Objective function ranges from 0 to 1, where 0.5 means that Goals 1 and 2 have equal importance. The outputs of this model are CLP_{Pred}, which is the optimized and predicted CLP value, and F_{Predi}, which is the predicted value of factors influencing CLP.
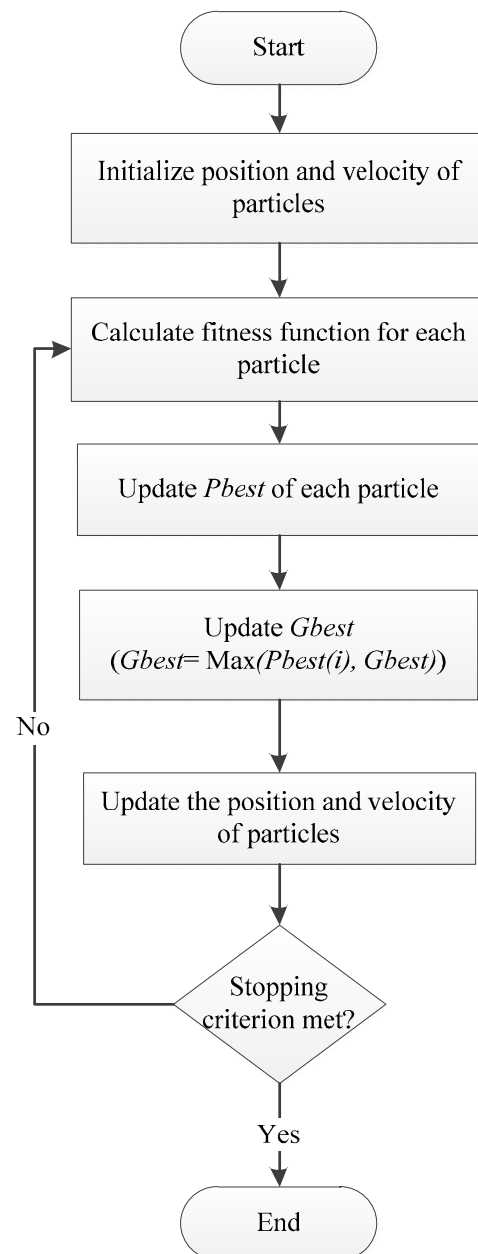

Figure 2. Overview of particle swarm optimization (PSO).

4. Experimental Results and Discussion

4.1. CLP Data Preparation and Feature Selection

Based on the data preparation process, the CLP value was normalized, using Equation (2). As a result of normalization, the CLP value was between 0 and 1, and greater CLP indicates better labor productivity for the project. After imputing missing values and removing factors with zero standard deviation, the number of factors was reduced to 108. By eliminating

outliers from the CLP data set, 7 data points were removed as outliers, and the total number of data points became 85. Therefore, the CLP data set after the preparation process had 85 data points, 108 CLP factors, and a CLP value.

Next, the number of features was reduced by the proposed HFS method. For this study, the threshold of 0.25 was defined for ReliefF. All features with weights greater than or equal to 0.25 were selected as essential features in the next HFS stage. From 108 factors in the final CLP data set, ReliefF selected 43 as essential features. In the next stage of HFS, which is the integration of SVM and GA as a wrapper method, the GA parameter settings were a population size of 50, GA maximum iteration of 60, crossover rate of 0.83, and mutation rate of 0.2. The SVM penalty factor *C* was 10, the kernel type was RBF, and the kernel cache was 200. These parameters were obtained by trial and error and are the optimum values for this case. The termination criteria were a maximum of 60 generations or no improvement of performance over 5 generations. The proposed wrapper method was developed considering these parameters, and it selected 14 of the 43 factors identified by ReliefF. The set of 14 factors was selected when the RMSE of the run of the wrapper model was lowest. Table 1 presents the selected CLP factors resulting from HFS. As shown in Table 1, the first 11 factors are all from the activity level, and the next three factors belong to the project level, which shows the significant impact of activity-level factors on predicting CLP. From the selected factors, "Level of interruption and disruption", "Complexity of task", "Working condition (dust and fumes)", "Location of work scope (elevation)" and "Congestion of work area" are factors that negatively influence CLP. In other words, after normalization, when negatively influencing factors have values close to zero, they result in greater CLP, compared to when their values are close to 1. The other selected factors are positively influencing factors, and when their values are close to 1, they result in greater CLP.

Table 1. Input factors for CLP modeling.

Selected Factor	Scale of Measure
(1) Crew size	Integer (Total number of crew members)
(2) Crew composition	Proportion (Ratio journeyman to apprentice to helper)
(3) Treatment of craftsperson by foreman	1–5 Predetermined rating
(4) Craftsperson trust in foreman	1–5 Predetermined rating
(5) Level of interruption and disruption	Integer (Number of interruptions and disruptions per day)
(6) Complexity of task	1–5 Predetermined rating
(7) Working condition (dust and fumes)	1–5 Predetermined rating
(8) Location of work scope (elevation)	Real number (elevation, m)
(9) Congestion of work area	Real number (ratio of actual peak manpower to actual average manpower)
(10) Fairness in performance review of crew by foreman	1–5 Predetermined rating
(11) Ground conditions	1–5 Predetermined rating
(12) Quality audits	Real number (Number of inspections per month)
(13) Risk monitoring and control	1–5 Predetermined rating
(14) Crisis management	1–5 Predetermined rating

4.2. CLP Modeling Comparison and Results

To develop the predictive CLP model, four different AI models were developed using the selected factors from HFS as input variables and CLP as the output. The accuracy of the four models was measured by comparing their predictions to the actual field data and calculating two commonly used error measures, mean absolute error (MAE) and RMSE, which are shown in Equations (12) and (13), where t_i and y_i are the actual and predicted CLP values for the ith instance, respectively, and m is the number of instances. For this purpose, data were divided into training and testing data sets, in which 70% of data were used for training and 30% for testing.

$$RMSE = \sqrt{\sum_i (t_i - y_i)^2 / m} \tag{12}$$

$$MAE = \left(\sum_i |y_i - t_i|\right) / m \tag{13}$$

For development of the ANN model, using MATLAB NN Toolbox, a multilayer feedforward back-propagation network with two hidden layers was considered, and the hidden layer sizes were 5 and 6. The learning rate was set to 0.33, and 200 training cycles were performed. The ANN model resulted in an RMSE of 0.164 and MAE of 0.130 for the training data set, and an RMSE of 0.165 and MAE of 0.135 for the testing data set.

The ANFIS model was generated using the ANFIS function of MATLAB Fuzzy Logic Toolbox. The basic learning rules for optimizing membership functions in ANFIS are either hybrid learning or back-propagation gradient descent. Hybrid learning combines the gradient descent and least square methods, and it overcomes the major limitation of the back-propagation method, which is that the learning process gets trapped in the local minima. Therefore, this study used the hybrid learning method. The training data set was grouped using subtractive clustering with an influence range of 0.4, squash factor of 1.15, and accept and reject ratios set at 0.5 and 1.15, respectively. The selected CLP factors were used as input variables and CLP as the output of ANFIS. The ANFIS model resulted in an RMSE of 0.042 and MAE of 0.034 for the training data set and an RMSE = 0.176 and MAE = 0.138 for the testing data set.

The ANFIS-GA model, developed using MATLAB, tries to optimize ANFIS parameters, and it showed better performance than ANFIS alone. In this study, the values of 0.2, 0.83, and 60 were assigned for the mutation rate, crossover percentage, and maximum iteration of GA, respectively. These parameters were obtained by trial and error and are the optimum values for this case. Different sizes were tested to find the appropriate population size and based on the results as shown in Table 2, the ANFIS-GA model with a population size of 25 had the best testing performance, which included an MAE of 0.096 for the training data set and MAE of 0.129 for the testing data set. Therefore, a population size of 25 was used in this study.

Table 2. Selecting the population size in ANFIS-GA modeling.

ANFIS-GA Model No.	Population Size	RMSE	
		Training	Testing
1	12	0.159	0.185
2	18	0.165	0.191
3	25	0.162	0.172
4	30	0.163	0.19

The RF model was developed using Python language programming and required three parameters, namely the minimum number of terminal nodes for each tree, the number of trees, and the number of randomly selected variables to grow the trees [63]. In this study, these three parameters were set to 5, 145, and 6, respectively. The results of the RF prediction model are listed in Table 3 along with results of the ANN, ANFIS, and ANFIS-GA models for comparison.

Table 3. Comparing the performance of the four developed models for predicting CLP.

Model	Training Dataset		Testing Dataset	
	RMSE	MAE	RMSE	MAE
ANN	0.164	0.130	0.165	0.135
ANFIS	0.042	0.034	0.176	0.138
ANFIS-GA	0.162	0.096	0.172	0.129
RF	0.074	0.051	0.137	0.112

The results presented in Table 3 indicate that the RF model had the highest accuracy among the four predictive models, with an RMSE of 0.137 and MAE of 0.112 in the testing data set. The second most accurate algorithm was the ANN model, with a testing data set RMSE of 0.165 and MAE of 0.135. The third most accurate algorithm was the combination of ANFIS and GA,

with an RMSE of 0.172 and MAE of 0.129 in the testing data set. Finally, testing data set RMSE of 0.176 and MAE of 0.138 indicate the ANFIS model was the least accurate.

According to the RMSE value of 0.137 for the RF testing data set, CLP predicted by RF was closer to the actual CLP values than for the other three developed models. In other words, RF was found to be better than ANN, ANFIS, and ANFIS-GA in mapping the relationship between the selected CLP factors and CLP. Moreover, the closeness of the RMSE values for the training and testing data sets indicate that ANN and RF were more stable than ANFIS and ANFIS-GA. Therefore, the RF model was selected to predict CLP in the optimization process for this study. Comparing the results of this study with past studies indicate that the RF predictive model has better performance. For example, Gerami Seresht et al. [39] obtained an RMSE value of 0.22 for their proposed CLP predictive model, while in this study, using the same data set, the RMSE value of the RF model was 0.137. Therefore, the proposed CLP predictive model achieved better performance accuracy in CLP prediction, compared with Gerami Seresht et al. [39].

4.3. CLP Optimization Results

Next, the integration of RF and PSO was developed to achieve the optimum value of the selected factors and maximum CLP value, according to the objective function in Equation (11). For this case study, the average value of each factor (F_{Avgi}) and CLP after normalization are shown in Table 4, and the average CLP value for the data set is 0.259.

Table 4. Average values of selected factors and CLP of the data set.

Selected Factor and CLP	Average Value in Normalized Data Set (F_{Avgi})
(1) Crew size	0.302
(2) Crew composition	0.289
(3) Treatment of craftsperson by foreman	0.569
(4) Craftsperson trust in foreman	0.518
(5) Level of interruption and disruption	0.162
(6) Complexity of task	0.500
(7) Working condition (dust and fumes)	0.218
(8) Location of work scope (elevation)	0.132
(9) Congestion of work area	0.438
(10) Fairness in performance review of crew by foreman	0.694
(11) Ground conditions	0.368
(12) Quality audits	0.832
(13) Risk monitoring and control	0.264
(14) Crisis management	0.634
Construction labor productivity	0.259

For the purpose of illustrating a CLP improvement trend, a sensitivity analysis was carried out to show the influence of different values of input parameters (namely ω and CLP_{tgt}) on output variables (CLP_{Pred} and Z) for understanding the impact of input parameters on model output. Table 5 shows the results of the sensitivity analysis, which indicates the value of Z and predicted CLP as outputs based on different values of ω and CLP_{tgt} as inputs of the RF-PSO model. The value of ω was changed to between 0.27 and 1; $\omega = 1$ is the largest possible value for ω and indicates that Goal 2 has no impact on the model. The CLP_{tgt} is in the range of 0.45 to 1, and $CLP_{tgt} = 1$ is the largest possible value for CLP resulting from the normalization process. Figure 3 is based on the results in Table 5, which shows the value of CLP_{Pred} for different values of ω and CLP_{tgt}. For a specific CLP_{tgt}, by increasing ω, CLP_{Pred} increases, which shows the model sensitivity to ω, which is the relative importance of Goal 1. For $CLP_{tgt} = 0.45$ and $CLP_{tgt} = 0.6$, the changes in CLP_{Pred} are much less, given ω greater than 0.4. So, it can be concluded that when CLP_{tgt} is less than or equal to 0.6, the most appropriate value of ω is less than or equal to 0.4. This means the minimum deviation of F_{Predi} (predicted value of CLP factors) from F_{Avgi} (average value of CLP factors in the dataset) as a Goal 2 in Equation (10) has more weight, compared to the minimum deviation of CLP_{Pred} from CLP_{tgt} as a Goal 1 in Equation (9).

Table 5. The results of sensitivity analysis.

CLP_{tgt}	ω	0.27	0.40	0.50	0.60	0.73	1.00
0.45	Z	0.041	0.045	0.038	0.056	0.033	1.15×10^{-5}
	CLP_{Pred}	0.374	0.430	0.441	0.439	0.448	0.449
0.60	Z	0.042	0.129	0.049	0.078	0.055	1.19×10^{-6}
	CLP_{Pred}	0.386	0.565	0.586	0.599	0.596	0.599
0.75	Z	0.057	0.049	0.079	0.124	0.116	0.0005
	CLP_{Pred}	0.522	0.561	0.616	0.649	0.671	0.721
0.90	Z	0.071	0.190	0.184	0.186	0.157	0.032
	CLP_{Pred}	0.555	0.558	0.664	0.678	0.685	0.728
1.00	Z	0.152	0.146	0.205	0.189	0.162	0.054
	CLP_{Pred}	0.713	0.697	0.714	0.728	0.737	0.769
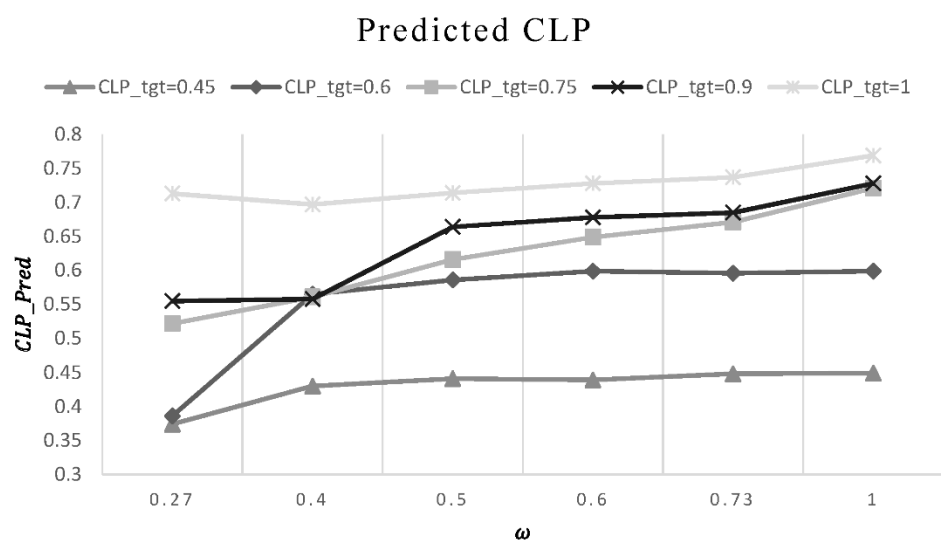
Figure 3. Predicted CLP from sensitivity analysis results.

For selecting the most appropriate weight and targeted CLP, a company's preference is important. Most companies prefer minimum deviation from "average value of factors", which is feasible to reach, helps them decrease the number of corrective measures that are required, and thus reduces the cost of implementing corrective measures. Based on this, Goal 2 needs to have more weight compared to Goal 1, which leads to selecting a value of ω less than or equal to 0.5 as a weight of Goal 1.

For this case study, the targeted CLP (CLP_{tgt}) of 0.75 and ω of 0.27 were selected. Equation (14) indicates the objective function of the HFS-RF-PSO algorithm, according to the selected factors. In the presented algorithm, the settings were the number of particles = 50, maximum number of iterations = 30, and maximum velocity = 2, and the value of learning factors c_1 and c_2 were both set to 2.05. The initial values of the parameters were established on the basis of the relevant literature [67]. A large number of trials were performed to obtain the optimum values for this case.

$$Minimize \left(Z = 0.27 \times (0.75 - CLP_{Pred})^2 + 0.73 \times \sum_{i=1}^{16} \left(F_{Predi} - F_{Avgi} \right)^2 \right) \qquad (14)$$

Based on the selected inputs, the result of the RF-PSO model indicated 0.057 as a minimum value of Z, which is the minimum value of objective function (Equation (14)), and 0.522 was achieved as a maximum value of predicted CLP (CLP_{Pred}). The optimum value of each factor is shown in Table 6.

Table 6. Result of the RF-PSO algorithm for selected factors and CLP.

Selected Factor and CLP	Optimum Value (F_{Predi})	Deviation ($F_{Predi} - F_{Avgi}$)
(1) Crew size	0.326	0.024
(2) Crew composition	0.364	0.075
(3) Treatment of craftsperson by foreman	0.587	0.018
(4) Craftsperson trust in foreman	0.535	0.017
(5) Level of interruption and disruption	0.043	−0.119
(6) Complexity of task	0.549	0.0490
(7) Working condition (dust and fumes)	0.108	−0.110
(8) Location of work scope (elevation)	0.176	0.044
(9) Congestion of work area	0.452	0.014
(10) Fairness in performance review of crew by foreman	0.808	0.114
(11) Ground conditions	0.372	0.004
(12) Quality audits	0.733	−0.099
(13) Risk monitoring and control	0.271	0.007
(14) Crisis management	0.629	−0.005
Construction labor productivity	0.522	0.263

The optimum value of each factor was obtained from the RF-PSO model as the predicted values for CLP factors (F_{Predi}) and the deviation of the optimum value from the average value for each factor. In other words, deviation from average value was achieved using Equation (15):

$$Deviation = F_{Predi} - F_{Avgi} \qquad (15)$$

As shown in Table 6, the optimum value of the factors "Ground condition," "Crisis management," and "Risk monitoring and control" have the least deviation from the average value of selected factors from dataset (F_{Avgi}) with values of 0.004, −0.005, and 0.007, respectively. Therefore, these factors do not need major changes to achieve the optimum CLP value, which is 0.522. It is notable in Table 6 that the optimum values of "Level of interruption and disruption," "Working condition (dust and fumes)," and "Fairness in performance review of crew by foreman" have the largest deviation from the average value of the factors, which are −0.119, −0.110, and 0.114, respectively. In other words, "Level of interruption and disruption" needs to be reduced to 0.043, "Working condition (dust and fumes)" needs to be reduced to 0.108, and "Fairness in performance review of crew by foreman" needs to increase to 0.808 in order to obtain the optimum CLP value. Improving factors with high deviation helps companies reach optimum predicted CLP. In order to improve factors that have a high deviation from their average value, a number of improvement strategies and corrective measures can be implemented. For example, for reducing dust and fumes in the working area, preventative maintenance for the air-conditioning system can be conducted.

The proposed HFS-RF-PSO model has the potential to benefit construction companies in achieving their preferred labor productivity by applying the minimum changes to factors influencing CLP. Another capability of the proposed model is that companies can define their targeted value for each factor influencing CLP instead of the average value of factors. The results of the model will give them the values of predicted CLP and predicted factors in regard to having the minimum deviation from their targeted values for CLP as well as each factor. This novel approach can help companies identify factors that need the most changes for achieving their targeted CLP and, consequently, to prioritize the management practices that focus on factors with the greatest deviation from average value in the HFS-RF-PSO model.

The proposed HFS-RF-PSO model presented in this paper has a few limitations that need to be addressed in future research. First, the hybrid model was developed, using field data collected for concrete placing activities. In order to develop a generic model of CLP for different types of labor-dependent activities, new field data need to be collected. Second, although the PSO algorithm is computationally efficient compared to other optimization techniques and is robust with respect to control parameters, it can fall into a local optimum

in high-dimensional space. In future research, an adaptive PSO algorithm can be developed and added to the hybrid model to improve diversity of the algorithm and avoid falling into local optimum.

5. Conclusions and Future Work

Developing models for predicting and improving a project's labor productivity is challenging because of the complexity of construction projects. Hence, accurate CLP prediction is required for effective decision making before and during project execution. The fact that numerous factors affect CLP is the main challenge in modeling labor productivity. This study deals with this challenge by developing an HFS method prior to CLP prediction. The main aim of this study was to develop a novel approach for predicting and optimizing CLP. After developing the HFS method, this novel methodology developed and compared four different predictive models, using ANN, ANFIS, ANFIS-GA, and RF and the 14 factors selected by HFS as inputs and CLP as an output. The comparative analysis of four predictive models showed that the RF model obtained better accuracy, compared with the three other models. Then, RF as the most accurate predictive model was selected to integrate with PSO for identifying the optimum value of influential factors and the maximum CLP value. The proposed HFS-RF-PSO model has the capability of obtaining the optimized CLP close to a company's preferred value and minimizing deviation of the predicted CLP factors from the average value of factors in a data set. Based on the results of the HFS-RF-PSO model using the mentioned data set, among 14 selected factors, "Level of interruption and disruption," "Working condition (dust and fumes)," and "Fairness in performance review of crew by foreman" have the largest deviation from their average value, which means major improvements regarding these factors are needed in order to obtain optimum CLP. Furthermore, comparing the four most common ML models highlighted some critical modeling features of the presented models, which can assist researchers in future studies.

The contributions of this study include (1) identifying the most predictive factors for CLP by developing an HFS model that contains the integration of ReliefF and SVM-GA, (2) developing and comparing four different predictive models for CLP and identifying the most accurate model, and (3) developing a novel approach—the HFS-RF-PSO algorithm—for optimizing the factors that influence CLP and identifying the maximum CLP value considering the minimum deviation from targeted CLP value and also finding the optimum value of the selected factors based on minimizing their deviation from their average value in the data set. The proposed HFS-RF-PSO model helps project managers predict, optimize, and improve the CLP value, taking into account the factors that are the most predictive of CLP. Although construction projects are unique and the factors affecting CLP may differ from project to project, the proposed model is flexible, and new influencing factors can be added to the existing model structure to predict and optimize CLP and its factors for a given project. The results of this study and implementation of the HFS-RF-PSO model will help project managers identify causes of low labor productivity and select and prioritize corrective measures based on the deviation of factors in the model, to improve CLP. The model also enables project managers to improve the reliability of predictions.

Future research can focus on using the proposed methodology to model and optimize multifactor construction productivity, which includes labor, equipment, and materials. Furthermore, future studies can present corrective measures to improve CLP, according to the HFS-RF-PSO results that show which factors need the most changes for reaching the targeted CLP. In addition, using new field data from other labor-dependent activities will help researchers overcome one of the mentioned limitations of this study and develop a generic hybrid model for predicting and optimizing CLP.

Author Contributions: S.E.: conceptualization, methodology, formal analysis, software, investigation, writing—original draft preparation, writing—review and editing. A.R.F.: conceptualization, writing—review and editing, supervision, project administration, funding acquisition. V.S.: conceptualization, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

References

1. Hafez, S.M. Critical factors affecting construction labor productivity in Egypt. *Am. J. Civ. Eng.* **2014**, *2*, 35–40. [CrossRef]
2. Doloi, H. Application of AHP in improving construction productivity from a management perspective. *Constr. Manag. Econ.* **2008**, *26*, 841–854. [CrossRef]
3. Abdel-Hamid, M.; Mohamed Abdelhaleem, H. Impact of poor labor productivity on construction project cost. *Int. J. Constr. Manag.* **2020**. [CrossRef]
4. Gurmu, A.T. Tools for measuring construction materials management practices and predicting labor productivity in multistory building projects. *J. Constr. Eng. Manag.* **2019**, *145*. [CrossRef]
5. Tsehayae, A.A.; Fayek, A.R. System model for analysing construction labour productivity. *Constr. Innov.* **2016**, *16*, 203–228. [CrossRef]
6. Moselhi, O.; Khan, Z. Significance ranking of parameters impacting construction labour productivity. *Constr. Innov.* **2012**, *12*, 272–296. [CrossRef]
7. Cheng, M.Y.; Cao, M.T.; Jaya Mendrofa, A.Y. Dynamic feature selection for accurately predicting construction productivity using symbiotic organisms search-optimized least square support vector machine. *J. Build. Eng.* **2021**, *35*, 101973. [CrossRef]
8. Khanzadi, M.; Nasirzadeh, F.; Mir, M.; Nojedehi, P. Prediction and improvement of labor productivity using hybrid system dynamics and agent-based modeling approach. *Constr. Innov.* **2017**, *18*, 2–19. [CrossRef]
9. Tsehayae, A.A.; Fayek, A.R. Developing and Optimizing Context-Specific Fuzzy Inference System-Based Construction Labor Productivity Models. *J. Constr. Eng. Manag.* **2016**, *142*. [CrossRef]
10. Ebrahimi, S.; Raoufi, M.; Fayek, A.R. Framework for Integrating an Artificial Neural Network and a Genetic Algorithm to Develop a Predictive Model for Construction Labor Productivity. In Proceedings of the Construction Research Congress 2020, Tempe, Arizona, 8–10 March 2020; American Society of Civil Engineers: Reston, VA, USA, 2020; pp. 58–66.
11. Durdyev, S.; Mbachu, J. On-site labour productivity of New Zealand construction industry: Key constraints and improvement measures. *Australas. J. Constr. Econ. Build.* **2011**, *11*, 18–33. [CrossRef]
12. El-Gohary, K.M.; Aziz, R.F.; Abdel-Khalek, H.A. Engineering Approach Using ANN to Improve and Predict Construction Labor Productivity under Different Influences. *J. Constr. Eng. Manag.* **2017**, *143*. [CrossRef]
13. Tsehayae, A.A.; Fayek, A.R. Identification and comparative analysis of key parameters influencing construction labour productivity in building and industrial projects. *Can. J. Civ. Eng.* **2014**, *41*, 878–891. [CrossRef]
14. Jarkas, A.M. Factors influencing labour productivity in Bahrain's construction industry. *Int. J. Constr. Manag.* **2015**, *15*, 94–108. [CrossRef]
15. Montaser, N.M.; Mahdi, I.M.; Mahdi, H.A.; Rashid, I.A. Factors Affecting Construction Labor Productivity for Construction of Pre-Stressed Concrete Bridges. *Int. J. Constr. Eng. Manag.* **2018**, *7*, 193–206. [CrossRef]
16. Alaghbari, W.; Al-Sakkaf, A.A.; Sultan, B. Factors affecting construction labour productivity in Yemen. *Int. J. Constr. Manag.* **2019**, *19*, 79–91. [CrossRef]
17. Kazerooni, M.; Raoufi, M.; Fayek, A.R. Framework to Analyze Construction Labor Productivity Using Fuzzy Data Clustering and Multi-Criteria Decision-Making. In Proceedings of the Construction Research Congress 2020, Tempe, Arizona, 8–10 March 2020; American Society of Civil Engineers: Reston, VA, USA, 2020; pp. 48–57.
18. Durdyev, S.; Ismail, S.; Kandymov, N. Structural equation model of the factors affecting construction labor productivity. *J. Constr. Eng. Manag.* **2018**, *144*. [CrossRef]
19. Irfan, M.; Zahoor, H.; Abbas, M.; Ali, Y. Determinants of labor productivity for building projects in Pakistan. *J. Constr. Eng. Manag. Innov.* **2020**, *3*, 85–100. [CrossRef]
20. Agrawal, A.; Halder, S. Identifying factors affecting construction labour productivity in India and measures to improve productivity. *Asian J. Civ. Eng.* **2020**, *21*, 569–579. [CrossRef]
21. Smith, S.D. Earthmoving Productivity Estimation Using Linear Regression Techniques. *J. Constr. Eng. Manag.* **1999**, *125*, 133–141. [CrossRef]

22. Thomas, A.V.; Sudhakumar, J. Modelling masonry labour productivity using multiple regression. In Proceedings of the 30th Annual Association of Researchers in Construction Management Conference, ARCOM 2014, Portsmouth, UK, 1–3 September 2014; Association of Researchers in Construction Management: Portsmouth, UK, 2014; pp. 1345–1354.

23. Hai, D.T.; Van Tam, N. Application of the Regression Model for Evaluating Factors Affecting Construction Workers' Labor Productivity in Vietnam. *Open Constr. Build. Technol. J.* **2020**, *13*, 353–362. [CrossRef]

24. Song, L.; AbouRizk, S.M. Measuring and modeling labor productivity using historical data. *J. Constr. Eng. Manag.* **2008**, *134*, 786–794. [CrossRef]

25. Al-Kofahi, Z.G.; Mahdavian, A.; Oloufa, A. System dynamics modeling approach to quantify change orders impact on labor productivity 1: Principles and model development comparative study. *Int. J. Constr. Manag.* **2020**. [CrossRef]

26. Gerami Seresht, N.; Fayek, A.R. Dynamic modeling of multifactor construction productivity for equipment-intensive activities. *J. Constr. Eng. Manag.* **2018**, *144*. [CrossRef]

27. Mirahadi, F.; Zayed, T. Simulation-based construction productivity forecast using neural-network-driven fuzzy reasoning. *Autom. Constr.* **2016**, *65*, 102–115. [CrossRef]

28. Topuz, K.; Zengul, F.D.; Dag, A.; Almehmi, A.; Yildirim, M.B. Predicting graft survival among kidney transplant recipients: A Bayesian decision support model. *Decis. Support. Syst.* **2018**, *106*, 97–109. [CrossRef]

29. Atallah, D.M.; Badawy, M.; El-Sayed, A.; Ghoneim, M.A. Predicting kidney transplantation outcome based on hybrid feature selection and KNN classifier. *Multimed. Tools Appl.* **2019**, *78*, 20383–20407. [CrossRef]

30. Venkatesh, B.; Anuradha, J. A hybrid feature selection approach for handling a high-dimensional data. In *Innovations in Computer Science and Engineering*; Saini, H., Sayal, R., Govardhan, A., Buyya, R., Eds.; Springer: Singapore, 2019; pp. 365–373. [CrossRef]

31. Piao, Y.; Ryu, K.H. A Hybrid Feature Selection Method Based on Symmetrical Uncertainty and Support Vector Machine for High-Dimensional Data Classification. In Proceedings of the Asian Conference on Intelligent Information and Database Systems, Kanazawa, Japan, 3–5 April 2017; Springer: Cham, Switzerland, 2017; pp. 721–727.

32. Limsawasd, C.; Athigakunagorn, N. Optimizing Construction productivity and resources in building projects under uncertainty. In Proceedings of the 6th CSCE-CRC International Construction Specialty Conference 2017-Held as Part of the Canadian Society for Civil Engineering Annual Conference and General Meeting, Vancouver, BC, Canada, 31 May–3 June 2017; pp. 1120–1129.

33. Golnaraghi, S.; Moselhi, O.; Alkass, S.; Zangenehmadar, Z. Predicting construction labor productivity using lower upper decomposition radial base function neural network. *Eng. Rep.* **2020**, *2*. [CrossRef]

34. Nasirzadeh, F.; Kabir, H.M.D.; Akbari, M.; Khosravi, A.; Nahavandi, S.; Carmichael, D.G. ANN-based prediction intervals to forecast labour productivity. *Eng. Constr. Archit. Manag.* **2020**, *27*, 2335–2351. [CrossRef]

35. Momade, M.H.; Shahid, S.; bin Hainin, M.R.; Nashwan, M.S.; Tahir Umar, A. Modelling labour productivity using SVM and RF: A comparative study on classifiers performance. *Int. J. Constr. Manag.* **2020**. [CrossRef]

36. Sarihi, M.; Shahhosseini, V.; Banki, M.T. Development and comparative analysis of the fuzzy inference system-based construction labor productivity models. *Int. J. Constr. Manag.* **2021**. [CrossRef]

37. Zhang, J.; Li, D.; Wang, Y. Predicting uniaxial compressive strength of oil palm shell concrete using a hybrid artificial intelligence model. *J. Build. Eng.* **2020**, *30*, 101282. [CrossRef]

38. Raoufi, M.; Fayek, A.R. Framework for Identification of Factors Affecting Construction Crew Motivation and Performance. *J. Constr. Eng. Manag.* **2018**, *144*. [CrossRef]

39. Gerami Seresht, N.; Lourenzutti, R.; Fayek, A.R. A fuzzy clustering algorithm for developing predictive models in construction applications. *Appl. Soft Comput. J.* **2020**, *96*, 106679. [CrossRef]

40. Goodarzizad, P.; Mohammadi Golafshani, E.; Arashpour, M. Predicting the construction labour productivity using artificial neural network and grasshopper optimisation algorithm. *Int. J. Constr. Manag.* **2021**. [CrossRef]

41. Fei, Y.; Min, H. Simultaneous feature with support vector selection and parameters optimization using GA-based SVM solve the binary classification. In Proceedings of the 2016 First IEEE International Conference on Computer Communication and the Internet, Wuhan, China, 13–15 October 2016; ICCCI: Piscataway, NJ, USA, 2016; pp. 426–433. [CrossRef]

42. Tao, Z.; Huiling, L.; Wenwen, W.; Xia, Y. GA-SVM based feature selection and parameter optimization in hospitalization expense modeling. *Appl. Soft Comput. J.* **2019**, *75*, 323–332. [CrossRef]

43. Jin, H.; Nahangi, M.; Goodrum, P.M.; Yuan, Y. Multiobjective Optimization for Scaffolding Space Planning in Industrial Piping Construction Using Model-Based Simulation Programming. *J. Comput. Civ. Eng.* **2020**, *34*. [CrossRef]

44. Lin, C.L.; Lai, Y.C. An improved time-cost trade-off model with optimal labor productivity. *J. Civ. Eng. Manag.* **2020**, *26*, 113–130. [CrossRef]

45. Shahbazi, B.; Akbarnezhad, A.; Rey, D.; Ahmadian Fard Fini, A.; Loosemore, M. Optimization of Job Allocation in Construction Organizations to Maximize Workers' Career Development Opportunities. *J. Constr. Eng. Manag.* **2019**, *145*. [CrossRef]

46. Choudhury, S.J.; Pal, N.R. Imputation of missing data with neural networks for classification. *Knowl. Based Syst.* **2019**, *182*. [CrossRef]

47. Nelwamondo, F.V.; Golding, D.; Marwala, T. A dynamic programming approach to missing data estimation using neural networks. *Inf. Sci.* **2013**, *237*, 49–58. [CrossRef]

48. Yuan, H.; Xu, G.; Yao, Z.; Jia, J.; Zhang, Y. Imputation of Missing Data in Time Series for Air Pollutants Using Long Short-Term Memory Recurrent Neural Networks. In Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, Singapore, 8–12 October 2018; pp. 1293–1300. [CrossRef]

49. Xu, W.; Jiang, L.; Yu, L. An attribute value frequency-based instance weighting filter for naive Bayes. *J. Exp. Theor. Artif. Intell.* **2019**, *31*, 225–236. [CrossRef]

50. Sandbhor, S.; Chaphalkar, N.B. Impact of outlier detection on neural networks based property value prediction. In *Information Systems Design and Intelligent Applications*; Satapathy, S., Bhateja, V., Somanah, R., Yang, X.S., Senkerik, R., Eds.; Springer: Singapore, Singapore, 2019; Volume 862, pp. 481–495. [CrossRef]

51. Ebrahimi, S.; Kazerooni, M.; Sumati, V.; Fayek, A.R. A Predictive Model for Construction Labour Productivity Using the Integration of Hybrid Feature Selection and PCA Methods. *Can. J. Civ. Eng.*. under review.

52. Kira, K.; Rendell, L.A. A Practical Approach to Feature Selection. In Proceedings of the Ninth International Workshop (ML92) at the Ninth International Machine Learning Conference, Aberdeen, Scotland, UK, 1–3 July 1992; Elsevier: Amsterdam, The Netherlands; pp. 249–256. [CrossRef]

53. Kononenko, I. Estimating attributes: Analysis and extensions of RELIEF. In *Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), Proceedings of the Machine Learning: ECML-94, Catania, Italy, 6–8 April 1994*; Bergadano, F., De Raedt, L., Eds.; Spring: Berlin/Heidelberg, Germany, 1994; Volume 784, pp. 171–182. [CrossRef]

54. Mathur, A.; Foody, G.M. Multiclass and binary SVM classification: Implications for training and classification users. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 241–245. [CrossRef]

55. Bao, Y.; Wang, T.; Qiu, G. Research on applicability of SVM kernel functions used in binary classification. *Adv. Intell. Syst. Comput.* **2014**, *255*, 833–844. [CrossRef]

56. Huang, C.L.; Wang, C.J. A GA-based feature selection and parameters optimization for support vector machines. *Expert Syst. Appl.* **2006**, *31*, 231–240. [CrossRef]

57. Aličković, E.; Subasi, A. Breast cancer diagnosis using GA feature selection and rotation forest. *Neural Comput. Appl.* **2017**, *28*, 753–763. [CrossRef]

58. Liu, X.; Song, Y.; Yi, W.; Wang, X.; Zhu, J. Comparing the random forest with the generalized additive model to evaluate the impacts of outdoor ambient environmental factors on scaffolding construction productivity. *J. Constr. Eng. Manag.* **2018**, *144*. [CrossRef]

59. Taheri, K.; Hasanipanah, M.; Golzar, S.B.; Majid, M.Z.A. A hybrid artificial bee colony algorithm-artificial neural network for forecasting the blast-produced ground vibration. *Eng. Comput.* **2017**, *33*, 689–700. [CrossRef]

60. Almási, A.D.; Woźniak, S.; Cristea, V.; Leblebici, Y.; Engbersen, T. Review of advances in neural networks: Neural design technology stack. *Neurocomputing* **2016**, *174*, 31–41. [CrossRef]

61. Siraj, N.B.; Fayek, A.R.; Tsehayae, A.A. Development and optimization of artificial intelligence-based concrete compressive strength predictive models. *Int. J. Struct. Civ. Eng. Res.* **2016**, *5*, 156–167. [CrossRef]

62. Moayedi, H.; Raftari, M.; Sharifi, A.; Jusoh, W.A.W.; Rashid, A.S.A. Optimization of ANFIS with GA and PSO estimating α ratio in driven piles. *Eng. Comput.* **2020**, *36*, 227–238. [CrossRef]

63. Wang, Z.; Wang, Y.; Zeng, R.; Srinivasan, R.S.; Ahrentzen, S. Random forest based hourly building energy prediction. *Energy Build.* **2018**, *171*, 11–25. [CrossRef]

64. Grandvalet, Y. Bagging equalizes influence. *Mach. Learn.* **2004**, *55*, 251–270. [CrossRef]

65. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the ICNN'95-International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995; IEEE: Piscataway, NJ, USA, 1995; Volume 4, pp. 1942–1948. [CrossRef]

66. Sengupta, S.; Basak, S.; Peters, R.A. Particle Swarm Optimization: A survey of historical and recent developments with hybridization perspectives. *Mach. Learn. Knowl. Extr.* **2018**, *1*, 157–191. [CrossRef]

67. El-Ghandour, H.A.; Elbeltagi, E. Comparison of five evolutionary algorithms for optimization of water distribution networks. *J. Comput. Civ. Eng.* **2018**, *32*. [CrossRef]