

Received October 19, 2021, accepted November 28, 2021, date of publication November 30, 2021, date of current version December 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3131702

Using Slightly Imbalanced Binary Classification to Predict the Efficiency of Winter Road Maintenance

MAHSHID HATAMZAD¹, GEANETTE POLANCO PINERES¹, AND JOHAN CASSELGREN²

¹Department of Industrial Engineering, UiT/The Arctic University of Norway, 8514 Narvik, Nordland, Norway

²Department of Engineering Sciences and Mathematics, Luleå University of Technology, 971 87 Luleå, Sweden

Corresponding author: Mahshid Hatamzad (mahshid.hatamzad@uit.no)

This work was supported by the UiT/The Arctic University of Norway.

ABSTRACT The prediction of efficiency scores for winter road maintenance (WRM) is a challenging and serious issue in countries with cold climates. While effective and efficient WRM is a key contributor to maximizing road transportation safety and minimizing costs and environmental impacts, it has not yet been included in intelligent prediction methods. Therefore, this study aims to design a WRM efficiency classification prediction model that combines data envelopment analysis and machine learning techniques to improve decision support systems for decision-making units. The proposed methodology consists of six stages and starts with road selection. Real data are obtained by observing road conditions in equal time intervals via road weather information systems, optical sensors, and road-mounted sensors. Then, data preprocessing is performed, and efficiency scores are calculated with the data envelopment analysis method to classify the decision-making units into efficient and inefficient classes. Next, the WRM efficiency classes are considered targets for machine learning classification algorithms, and the dataset is split into training and test datasets. A slightly imbalanced binary classification case is encountered since the distributions of inefficient and efficient classes in the training dataset are unequal, with a low ratio between classes. The proposed methodology includes a comparison of different machine learning classification techniques. The graphical and numerical results indicate that the combination of a support vector machine and genetic algorithm yields the best generalization performance. The results include analyzing the variables that affect the WRM and using efficiency classes to drive future insights to improve the process of decision-making.

INDEX TERMS Classification, data envelopment analysis, machine learning, prediction methods, road transportation safety, winter road maintenance.

I. INTRODUCTION

A. MOTIVATION

In today's modern world, the number of vehicles on roads has rapidly increased due to a surge in the population [1], resulting in an increased number of traffic accidents, especially in winter, when conditions making driving more difficult than in normal situations. Traffic accidents on roads are among the most dangerous and serious problems in society and can lead to detrimental impacts on both individuals and communities, resulting in fatalities, health issues, and economic losses [2]. Driving conditions during the winter in countries with cold climates, such as Norway, Sweden, Finland, and Canada, can

be hazardous due to ice formation on the ground, snowfall, and poor visibility. Slippery roads can reduce the friction between tires and road surfaces and can consequently lead to decreased road safety [3]. Therefore, winter road maintenance (WRM) is important for improving traffic safety. There are two main types of WRM: anti-icing and deicing. Anti-icing maintenance prevents ice from forming on the road surface (such as by using salt), and deicing maintenance contributes to removing ice from the road surface (such as by plowing). WRM needs to be well planned, including preparing a sufficient number of trucks and truck drivers and using suitable salt quantities to provide drivers with safe roads on demand. Additionally, costs and damage to the environment must be minimized. WRM should make roads usable, meaning that the speed of vehicles on the road should be the

The associate editor coordinating the review of this manuscript and approving it for publication was Jinjia Zhou¹.

same as that in the summer; however, in some cases, such a demand is unrealistic, and lower travel speeds are required. Therefore, roads must be effectively and efficiently maintained in winter, when the weather and driving conditions can be detrimental to driving. Effectiveness and efficiency are essential dimensions of WRM methods. Effective WRM means that road users can drive on a road with sufficient quality and safety. The term safe means that the friction between the tire and the road surface (traction) is suitable and that the road conditions allow drivers to maintain the appropriate velocity for driving. The term quality is related to minimizing bumps in the road (an uneven road surface) as a result of improper WRM. Additionally, too many bumps in the road could annoy drivers and harm vehicles. Efficient WRM involves achieving effective WRM by minimizing the time spent on maintenance, numbers of trucks and truck drivers, and amount of salt (chemicals) applied, which consequently equate to minimizing expenses. Considering only effectiveness without efficiency is an incomplete approach for assessing WRM performance, and in such cases, the result might be high expenses in the WRM budget [4].

B. STATE-OF-THE-ART METHOD

There have been many studies of WRM that aimed to maximize safety while minimizing cost and environmental impacts using different methods. For instance, a review was performed by Nixon and Foster [5]. on technological developments that can enhance WRM effectiveness and efficiency, such as road weather information systems (RWIS), mechanical methods, and materials for anti-icing and deicing. Minsk [6] published a book and summarized the available technology for collecting real-time meteorological information to predict road conditions. Hallmark and Jing [7] proposed visualization and analysis methods to examine the impacts of WRM on traffic safety using traffic, crash, weather, and snowplow data. Lorentzen [8] used a multivariate regression method and Monte Carlo simulations to analyze the effect of climate variations on WRM costs. Xu and Kwon [9] applied the tabu search algorithm to optimize snowplow routing problems based on different fleet sizes and scenarios. Ahabchane *et al.* [10] presented a methodology that uses machine learning (ML) regression algorithms to predict the quantity of abrasive and salt in a street segment; they used street segment geomatic information, weather data recorded by the Canadian Meteorological Service, and telemetry data measured by truck sensors. Sedivy *et al.* [11] conducted a study considering technological innovation to increase the economic sustainability of WRM. In their research, they noted that the WRM process should reflect the safe mobility of road users and goods, global climate change, and novel approaches for cost effectiveness. Although previous studies made remarkable contributions to the overall WRM knowledgebase, a framework for predicting WRM efficiency has not been established. The prediction of WRM efficiency involves making projections of future WRM performance. A highly accurate prediction of WRM efficiency can

strengthen decision support systems (DSSs) and planning in advance, as well as enhance WRM performance according to experience and accessible historical data.

One of the leading approaches for measuring efficiency scores is DEA. DEA is a good method for organizing and analyzing data, as it permits efficiency to vary over time and does not require prior assumptions regarding the fittest practice frontier. Nevertheless, DEA has been rarely used in performance predictions for decision-making units (DMUs) [12]. In this case, DEA can be combined with ML methods to fulfill certain prediction objectives. The idea of combining DEA and ML methods was first proposed in 1996, and DEA and artificial neural networks (ANNs) were used to assess the performance of bank branches [13]. Later, different studies were performed on the combination of DEA and ML. For instance, in 1999, a methodology was proposed by Hong *et al.* [14] to overcome DEA limitations through unsupervised ML techniques for clustering in system integration projects. The results provided decision-makers with a step-wise path for improving the efficiency scores of inefficient DMUs. Yeh *et al.* [15] applied a hybrid approach to integrate DEA, rough set theory, and a support vector machine (SVM) by treating efficiency scores as predictive features (inputs) to enhance the accuracy of business failure prediction; moreover, they used a DEA model to calculate the efficiency of input/output operations. In addition, rough set theory was employed to reduce the number of independent variables, which increased the classification ability of the SVM. Mirzozaffari *et al.* [16] combined DEA with an ML clustering method to determine the most efficient DMUs and the best clustering algorithm.

C. CONTRIBUTIONS

To model the complex WRM problem (due to the presence of different factors, such as variations in weather, temperature and road surface conditions), the main contribution of this study is to design a framework that combines the DEA method and an ML classification algorithm to predict slightly imbalanced binary classes of WRM efficiency using comprehensive measurements collected from RWIS stations, optical sensors, and road mounted sensors. This framework can help decision-makers improve DSSs for DMUs and achieve effective and efficient WRM.

The proposed methodology uses real measured data from three different sources and utilizes the DEA model to calculate WRM efficiency scores and divide DMUs into efficient and inefficient classes. Then, the binary classes are considered targets for the ML classification algorithm to predict the WRM efficiency classes for a section of European road E18 in Sweden. Different ML classification algorithms are compared to select the most accurate algorithm with the best generalization ability to predict the WRM efficiency classes. The algorithms include the k-nearest neighbor (KNN), decision tree (DT), multilayer perceptron (MLP), SVM, logistic regression (LR), and combined SVM and genetic algorithm (GA) methods.

D. OUTLINE OF THE PAPER

The rest of this paper is organized as follows. Section II explains the theory. The evaluation metrics are defined in Section III. The methodology is formulated in Section IV. In Section V, results are explained. Finally, in Section VI, we present conclusions.

II. THEORY

A. DATA ENVELOPMENT ANALYSIS

The DEA model calculates the weight of each input and output based on an optimization problem. According to the corresponding weights, DMUs are classified into two classes: i) efficient and ii) inefficient [17]. DEA can also identify inefficiency and possible improvements for inefficient DMUs (i.e., how changes in the inputs and outputs can lead to an efficient DMU [12]). There are two main formulations for the DEA model: i) Charnes, Cooper and Rhodes proposed the CCR model, which operates with constant returns to scale (CRS) [18], and ii) Banker, Charnes, and Cooper proposed the BCC model, which operates with variable returns to scale (VRS) [19]. The BCC model includes a convexity constraint and an extra free variable compared to the CCR model. Therefore, feasible region in the CCR model is larger than that in the BCC model; this difference can influence the efficiency of each model and ultimately lead to a reduction in the number of efficient DMUs in the CCR model [12]. In the DEA model, n DMUs are considered in the evaluation, and each DMU has m various inputs ($i=1, 2, \dots, m$) for producing s different outputs ($r=1, 2, \dots, s$). Moreover, x_{ij} and y_{rj} indicate the input and output values of DMU _{j} ($j=1, 2, \dots, n$). u_r and v_i represent input and output weights. The nature of DEA is based on maximizing the efficiency of DMU _{p} , subject to a few constraints: i) the efficiency of other DMUs needs to be lower than 1, ii) the input and output variables must be nonzero and nonnegative, and iii) the input and output weights must not be less than zero. The original DEA-CCR model that computes the highest efficiency Z_p for DMU _{p} based on the CCR approach is a linear programming model that can be defined as follows [18]:

$$\begin{aligned} \text{Max } Z_p &= \sum_{r=1}^s u_r y_{rp} \\ \text{st } : \quad &\sum_{i=1}^m v_i x_{ip} = 1 \\ &\sum_{r=1}^s u_r y_{rj} \leq \sum_{i=1}^m v_i x_{ij} \\ &u_r \geq 0, \quad v_i \geq 0 \end{aligned} \quad (1)$$

To analyze the outputs of the DEA model, n LP problems need to be solved [12]. The outputs are the values of decision variables or the weights of inputs and outputs (u_r and v_i), as well as the maximum efficiency of each DMU.

B. MACHINE LEARNING METHODS

ML in artificial intelligence enables machines to learn and improve without new computer programming. ML tech-

niques are flexible in processing large amounts of data, and the availability of advanced computational power plays an important role in analyzing data and discovering knowledge from data [20]. There are two main types of ML techniques: i) supervised ML, which is based on labeled data, and ii) unsupervised ML, which is based on unlabeled data. Classification is a supervised ML method, whereas clustering is an unsupervised ML method [21]. In this study, five different supervised ML techniques are used for classification prediction and modeling: KNN, DT, MLP, SVM, and LR. The details of these algorithms are explained as follows.

1) K-NEAREST NEIGHBORS

The KNN classification algorithm is a type of learning method known as an instance-based learning method. These learning methods are based on the principle that the observations within a dataset are usually located in close vicinity to other instances with the same or similar characteristics [22]. The KNN classifier tries to predict output classes by finding the K nearest neighbors based on the associated inputs [23]. The KNN approach typically uses a distance metric, generally the Euclidean distance, for classification [24].

2) DECISION TREE

A DT is a nonparameteric technique with high flexibility [25], and DTs are the most commonly used technique for classification [26]. A DT consists of a root node, branches and leaf nodes [27]. The nodes correspond to input variables, and the leaves are associated with decision outcomes [28]. The main objective of the DT approach is to build a model that calculates an output according to input variables [26].

3) MULTILAYER PERCEPTRON

Among the different types of ANNs, MLPs have been widely used for various classification purposes [28]. An MLP network is able to solve nonlinear problems by learning complicated patterns involving output and input variables that can be difficult to model with a single-layer network method. MLPs can be applied to different problems in which decision-making plays a significant role [29].

4) LOGISTIC REGRESSION

LR is a classification technique in statistics used to training models and find the most accurate, best fitting, and most sensible model to explore the relationship between inputs and an output [30]. Moreover, LR is a method used to obtain predictions for dichotomous dependent variables (i.e., predicting the absence or presence of an outcome according to predictor variables) [25].

5) SUPPORT VECTOR MACHINE

SVMs are among the most promising methods for classification [31] and have become popular due to several attractive features, such as their excellent generalization performance, avoidance of underfitting and overfitting and lack of sensitivity to input dimensions. SVM theory involves first mapping

the input data from a low-dimensional feature space to a high-dimensional feature space by using a kernel function; then, an SVM transforms the nonlinear separable problem into a linearly separable problem in the sample space [32]. The mathematical theory of SVM is defined based on [33].

It is assumed that $(x_i, y_i)_{i=1}^l$ is the training data set, where x_i indicates the input data, y_i is the corresponding output value and l is the number of values in the training set. The optimization problem is defined as:

$$\begin{aligned} \text{Min} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\varepsilon_i + \varepsilon_i^*) \\ \text{st} \quad & y_i - f(x_i) \leq \varepsilon + \varepsilon_i^* \\ & f(x_i) - y_i \leq \varepsilon + \varepsilon_i \\ & \varepsilon_i, \quad \varepsilon_i^* \geq 0 \end{aligned} \quad (2)$$

where C is a positive hyperparameter, often called a regularization parameter.

The general form of SVM theory after solving the dual problem can be written as:

$$f(x) = \sum_{i=1}^l (\alpha_i^* - \alpha_i) k(x_i, x) + b \quad (3)$$

where α_i and α_i^* are Lagrange multipliers. $k(x_i, x)$ is the kernel function. The form of the RBF kernel function is expressed as follows:

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (4)$$

where γ represents an RBF parameter and is always nonzero and nonnegative.

C. OPTIMIZATION METHOD

Optimization methods search for optimal values of parameters under different constraints, such as in a given search space. In this study, we use a GA optimization method to find the optimal values of SVM parameters.

1) GENETIC ALGORITHM

GAs have been successfully used in many fields because of their robustness, reliability, high speed in solving complex problems, and simplicity, which allows GA methods to be combined with other models. The GA algorithm is based on nonlinear global optimization, which is inspired by the mechanism of biological evolution [32]. The rule of probability transform and a fitness function are used to guide the search direction and obtain optimized values through iterative computations. The basic GA process is as follows [34].

Step 1: Randomly generate candidate solutions (initial population).

Step 2: Calculate the fitness values of candidate solutions (every chromosome in the population).

Step 3: Perform selection, crossover, and mutation operations.

Step 4: Stop if the termination condition is met (an approximate optimal solution is discovered); otherwise, return to the second step.

D. COMBINATION OF SUPPORT VECTOR MACHINE AND GENETIC ALGORITHM

The generalization performance of an SVM is highly dependent on the appropriate setting of parameters. The regularization parameter (C) specifies the tradeoff cost between minimizing model complexity and minimizing the training error. The kernel parameter γ determines the width of the kernel and is related to the range of inputs in the training data set [35]. After selecting a suitable kernel function, we must choose appropriate values for SVM parameters. To find optimal values of SVM parameters, we can utilize optimization methods such as the GA, which includes the following steps [36]:

- i. First, define the size of the initial population and the ranges of C and γ .
- ii. The error of the SVM model is selected to establish an objective function and calculate the fitness value.
- iii. A new population is generated through genetic operators (selection, crossover, and mutation).
- iv. The current population is replaced with the newly generated population.
- v. If the termination rules are met, the model can be used to optimize the values of SVM parameters.

III. EVALUATION METRICS

Generally, three types of binary classification prediction problems are encountered [37]: i) the distribution of classes is equal in the training dataset (balanced classification); ii) the distribution of classes is not equal in the training dataset, and the ratio between classes is moderately low (e.g., 4:6) (slightly imbalanced classification); iii) the distribution of classes is not equal in the training dataset, and the ratio between classes is high (e.g., 1:4) (severely imbalanced classification).

In most cases, slightly imbalanced classification is often not a challenge and can be modeled in the same way as a normal classification prediction problem [38]. In this study, we deal with slightly imbalanced classes; however, we provide the results for both classes, as they can provide us with valuable information about model performance.

To understand whether the samples are correctly classified by the models, evaluation metrics need to be used. Different evaluation metrics can be used to estimate the classification performance of ML methods in imbalanced binary classification. Most of these factors can be calculated from a confusion matrix. This matrix contains performance information for a given classifier based on the test dataset. The confusion matrix for binary classes is a matrix with two dimensions: one dimension includes the actual values, and the other dimension includes the predicted values [39]. There are two types of classes in problems that deal with slightly imbalanced binary classification: (i) the majority class, called the negative

		Predicted Values	
		Inefficient WRM	Efficient WRM
Actual Values	Inefficient WRM	True Negative (TN)	False Positive (FP)
	Efficient WRM	False Negative (FN)	True Positive (TP)

FIGURE 1. Confusion matrix.

TABLE 1. Explanation of four possible outputs for the confusion matrix.

Term	Explanation
TN	Appropriately classified samples into the negative class
FP	Falsely classified samples into the positive class
TP	Appropriately classified samples into the positive class
FN	Falsely classified samples into the negative class

outcome, and (ii) the minority class, called the positive outcome [40]. In this study, the majority class is associated with inefficient WRM, which constitutes a negative test result, and the minority class is associated with efficient WRM, which constitutes a positive test result (see Figure 1 and Table 1).

The performance of classification algorithms can be measured by the precision, recall, F1 score, accuracy, and area under the receiver operator characteristic curve (ROC), called the AUC. Table 2 illustrates the precision, recall, F1 score, and accuracy formulas for imbalanced binary classification. For the majority class, precision quantifies the appropriately predicted instances for the negative class (5), and precision for the minority class quantifies the appropriately predicted instances for the positive class (6). Therefore, precision measures the accuracy achieved for the negative and positive classes separately [40]. Recall for the negative class (also referred to as specificity [41]) is the proportion of true negative (TN) samples that are correctly observed (7), and recall for the positive class (also referred to as sensitivity [41]) reflects the classifier’s ability to correctly identify the proportion of true positive (TP) samples (8) [28], [30]. The difference between precision and recall for the positive class is that precision reflects the number of true positive predictions compared to the total number of positive predictions, whereas recall considers the number of missed positive predictions. The same relationships hold for the negative class. The F1 score captures the properties of precision and recall by combining the ratios of these two metrics into a single number (9) [40]. Accuracy is the number of correctly predicted samples divided by the total number of predicted observations (10). The AUC can be used to evaluate the overall performance of classification models. The ROC curve (probability curve) illustrates the tradeoff between the TP rate and false positive (FP) rate [28]. The AUC represents the capability of the model to distinguish different classes and varies between

TABLE 2. Precision, recall, F1-score, and accuracy formulas.

Evaluation metrics	Negative class	Positive class
Precision	$\frac{TN}{TN+FN}$ (5)	$\frac{TP}{TP+FP}$ (6)
Recall	$\frac{TN}{TN+FP}$ (7)	$\frac{TP}{TP+FN}$ (8)
F1-score	$\frac{2 \times Precision \times Recall}{Precision + Recall}$ (9)	
Accuracy	$\frac{TP+TN}{TP+FP+TN+FN}$ (10)	

0.5 and 1, where 0.5 is indicative of a poor classifier and 1 is indicative of an excellent (perfect) classifier.

IV. METHODOLOGY

The proposed methodology works based on six different stages used to predict WRM efficiency classes. In this study, a section of road E18 in Sweden was selected to obtain observations every 10 minutes in February 2019. The data were collected by two types of sensors (optical and road-mounted sensors) and the RWIS station. This methodology is implemented in Python software using different libraries: NumPy [42], Pandas [43], Matplotlib [44], Openpyxl, Collections, MLxtend [45], SciPy [46], and Scikit-learn [47]. Every stage is explained in detail as follows.

A. STAGE 1: ROAD SELECTION

In this paper, a section of a European road, E18 (Sweden), is selected as a case study. The data are from the Swedish Transport Administration’s RWIS station at Test site E18. This road section was chosen to use the observations obtained at the RWIS station, and optical and road-mounted sensors were also used to record data every 10 minutes in February 2019. Each 10-minute observation is considered one DMU (DMU_{ij}), and every 96 observations are considered one part (part_i). In total, 40 different parts are considered in this study, of which 39 parts have 96 different DMUs and one part (the last part) has 99 DMUs.

B. STAGE 1: DATA COLLECTION AND PREPROCESSING

There is an obvious relationship between weather conditions and road surface conditions. The important variables that influence WRM include the surface temperature (°C), air temperature (°C), base temperature (°C), ground temperature (°C), precipitation (mm), dew point (°C), snow height (mm), grip, conductivity (mS/cm), concentration of chemicals (g/l) and amount of chemicals (g/m²) used for WRM. The conductivity and concentration provide information about the salt on the road. Air temperature, precipitation, and dew point are measured by the RWIS station. Moreover, optical sensors were used to measure surface temperature and grip. Furthermore, the sensor mounted in the wheel track of the road was

TABLE 3. Number of missing values for each variable.

Variables	Number of missing values
Surface temperature	2
Concentration	2
Air temperature	0
Dew point	0
Grip	2
Precipitation	0
Conductivity	2
Chemical	2
Snow height	2
Base temperature	0
Ground temperature	2

used to measure the base temperature, ground temperature, snow height, conductivity, concentration of chemicals, and amount of chemicals.

1) DATA PREPARATION

In this research, data preparation efforts concentrated on three steps:

(i) Removing missing values: due to the low number of missing values, we removed missing values because they did not change the distributions or characteristics of the variables. Table 3 shows the number of missing values for each feature.

(ii) Scaling the data points: as data points are in different ranges, to obtain the best results, the data points were scaled. Table 4 gives an example statistical description of the raw data. Feature scaling is performed using the MinMax scaler method to transform the data range to (0.001, 1). Table 5 presents the statistical information after data transformation.

(iii) Removing highly correlated features: the Pearson correlation coefficient was calculated, and highly correlated features (absolute value of correlation coefficient greater than 0.8) were removed (Figure 2), because they do not provide extra information and increase model complexity. Therefore, air temperature, dew point, and ground temperature were removed from the dataset.

2) DIAGRAMMATIC DATA PRESENTATION

Diagrammatic data presentation is a process of employing graphical charts to visually depict the relationship between two or more features in a dataset. To understand the relationship between two variables, we employed a two-dimensional (bivariate) kernel density estimation (KDE) plot, which is a nonparametric method that visualizes the distribution of data points. We utilized this plot to extract the density of data points for the input variables and the amount of chemical (salt) used for WRM. In addition, we can see where no observation is presented. As shown in Figure 3, the density of observations for the salt used for WRM is typically lower than 1 gr when: (i) the density of data points for conductivity is between 0 and 2.5 gr, (ii) base temperature is between -2°C and +2°C, (iii) level of grip is almost 0.8, (iv) concentration

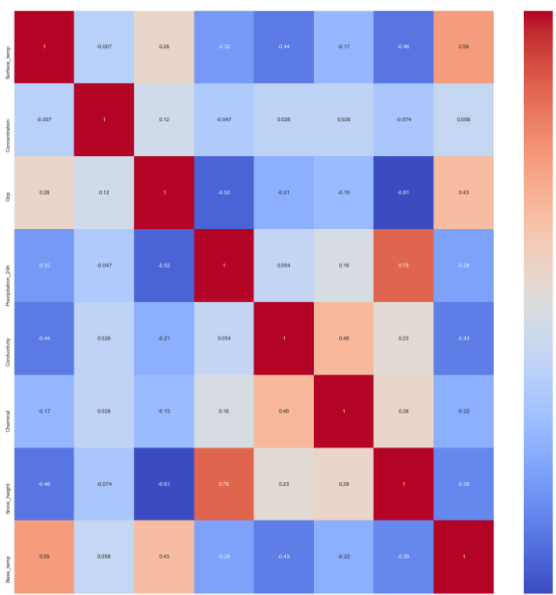


FIGURE 2. Pearson correlation coefficients between different features.

is between 0 and 20 gr, (v) surface temperature is between -3°C and +3°C, and (vi) snow height and precipitation is almost 0.

C. STAGE 3: DATA ENVELOPMENT ANALYSIS MODEL

We must determine the inputs and output of the DEA model from the 8 remaining variables (surface temperature, base temperature, precipitation, snow height, grip, conductivity, concentration of chemicals, and amount of chemicals).

1) APPLICATION OF DATA ENVELOPMENT ANALYSIS IN WINTER ROAD MAINTENANCE

In the DEA model, input variables can be defined as conditions that influence the performance of DMUs. On the other hand, the output variable can be defined as a result drawn by the conditions of the DMUs [4]. One of the main advantages of DEA is that no parametric function is required to solve the model. Hence, any type of input and output variables can be considered in the model without specifying parametric relationships. For instance, environmental attributes are uncontrollable factors because they are not under the control of decision makers. These attributes can be simply included as input variables in the DEA model. In addition, all the variables included in the DEA model have a similar opportunity to affect the efficiency scores [48]. Therefore, relevant and appropriate variables must be selected because they are firmly related to the goals of the DMU and influence the efficiency of the DMU. Thus, the DEA model is a suitable method for calculating the efficiency in our research. In addition, surface temperature, base temperature, precipitation, snow height, level of grip, conductivity, and concentration of chemicals are the input variables that can lead to different road surface conditions and can have negative impacts on the performance of

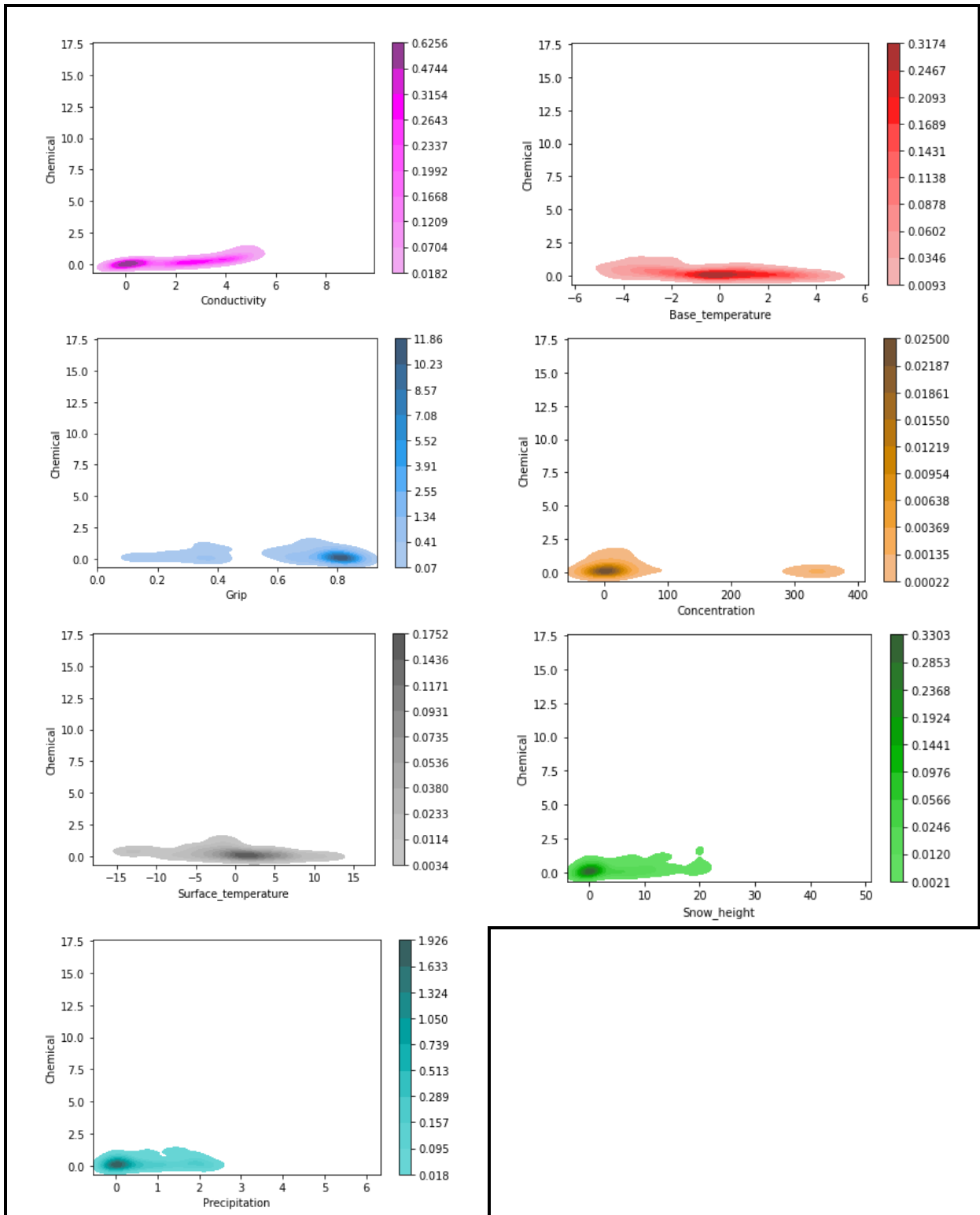


FIGURE 3. Kernel density estimation for each input variable based on the amount of chemical used for WRM.

DMUs(road section) in severe weather conditions. Moreover, road surface conditions determine the type of WRM, such as salting (and the salt quantity), that must be performed on the road section.

2) IMPLEMENTATION OF THE DATA ENVELOPMENT ANALYSIS MODEL

Based on the DEA principals, when we have a DMU with its associated input and output data for k periods, then each

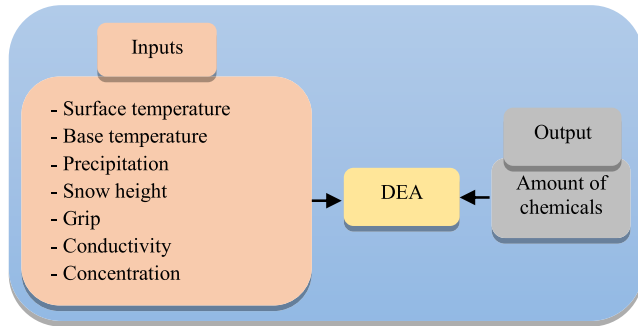


FIGURE 4. Schematic view of the inputs and outputs for DEA-CCR model to calculate the WRM efficiency of DMU_{ij} .

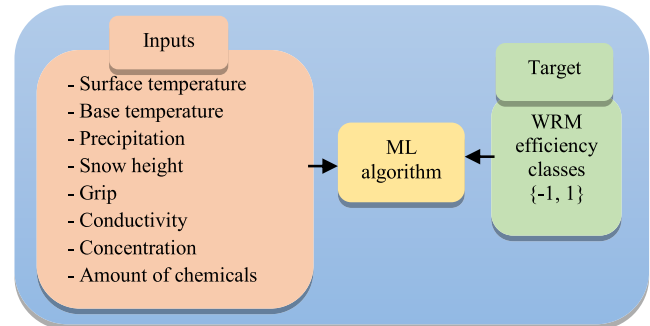


FIGURE 5. Schematic view of the inputs and targets of the ML classification model.

period can be considered a different DMU [49]. In our case, the observations are measured every 10 minutes in a section of road E18 in Sweden in February 2019. Therefore, in total, we have 3843 DMUs based on each 10-minute period; however, the efficiency score of each DMU is calculated for each part separately. In total, we have 40 parts, and each part includes three different time intervals (Table 6). The first observation was recorded on 01.02.2019 at 00:00, and the last observation was recorded on 28.02.2019 at 00:00.

Therefore, in this stage, the amount of chemicals is selected as the output, and the rest of the variables are considered inputs (Figure 4). Then, the DEA-CCR model is applied to calculate the efficiency score of DMU_{ij} for part_i.

D. STAGE 4: EFFICIENCY CLASSES

In this stage, if the efficiency score of DMU_{ij} is greater than or equal to 0.90, DMU_{ij} is assumed to be an efficient DMU and classified into class 1; otherwise, DMU_{ij} is inefficient and classified into class -1 . The results show that 2058 DMUs are efficient and 1785 DMUs are inefficient. Efficiency scores are considered targets for the ML binary classification method, and both the inputs and output of the DEA model are considered inputs for the ML classification method (Figure 5). Next, the dataset is split into training and testing sets, with 70% of instances selected for the training set and 30% of instances chosen for the testing set; thus, 2690 observations are considered for training the model, and 1153 observations are considered for testing the model. In the training set, 1262 observations are inefficient, and 1428 samples are efficient; in the test sets, 523 observations are inefficient, and 630 instances are efficient. Therefore, a slightly imbalanced binary classification problem is encountered since the distributions of inefficient and efficient classes are unequal in the training dataset, with a low ratio between classes.

E. STAGE 5: BUILDING THE ML CLASSIFICATION MODEL AND PERFORMING CROSS-VALIDATION

The KNN, DT, MLP, SVM, and LR algorithms are built based on the default Python software. However, for the SVM-GA model, the parameters of the SVM (c and γ) optimized by the GA algorithm are as follows.

1) SVM-GA

Since the generalization performance of SVMs is highly dependent on their parameters, these parameters must be set appropriately. In this study, SVM parameters are optimized simultaneously instead of separately. The first step here is cross-validation, and the reason behind this decision is enhancing SVM generalization performance without using the test set [33]. The generalization performance is later calculated based on the test set. The accuracy of generalization is considered a termination condition (accuracy > 0.9); if the SVM cannot meet this condition, the SVM parameters are optimized by the GA algorithm based on the following steps [36].

Step 1- The ranges of c and γ are specified, and the initial population is determined.

These ranges are $c \in [1, 10^4]$ and $\gamma \in [10^{-3}, 10]$ in this case, and the initial size of the population for the first generation is 20, which is usually selected randomly.

Step 2- c and γ are encoded

The SVM parameters are encoded directly based on the appropriate ranges to generate random chromosomes.

Step 3- Fitness function

Here, the classification model error is the basis of the fitness function for the training set, and the objective is to minimize the errors. Therefore, solutions that yield the minimum error have the highest fitness, and they are generally retained in the next generation.

Step 4- Selection, crossover, and mutation

The GA operators are specified in this step. A stochastic uniform selection operator is used. A scattered operator with a rate of 0.8 is selected as the crossover operator. A constraint dependence operator is selected for mutation based on the default setting in MATLAB software, and a new population is used here instead of the old population.

Step 5- Stopping criteria

The maximum number of generations is 100, and the function tolerance criterion is zero. If the stopping criteria are met, the algorithm is ready to generate optimal values of c and γ ; otherwise, the algorithm process returns to step 3. The optimal values of c and γ determined by the GA are shown in Table 7.

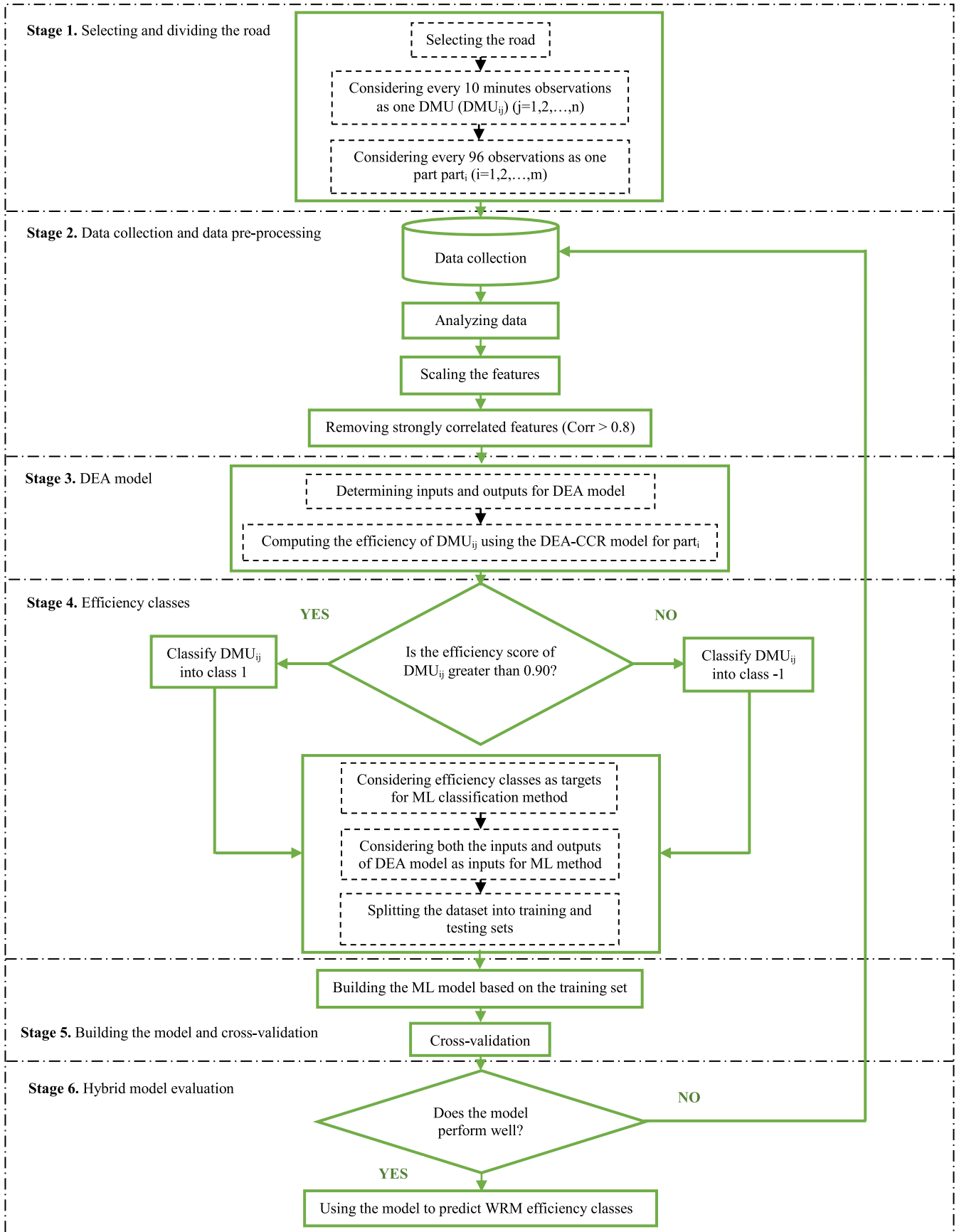
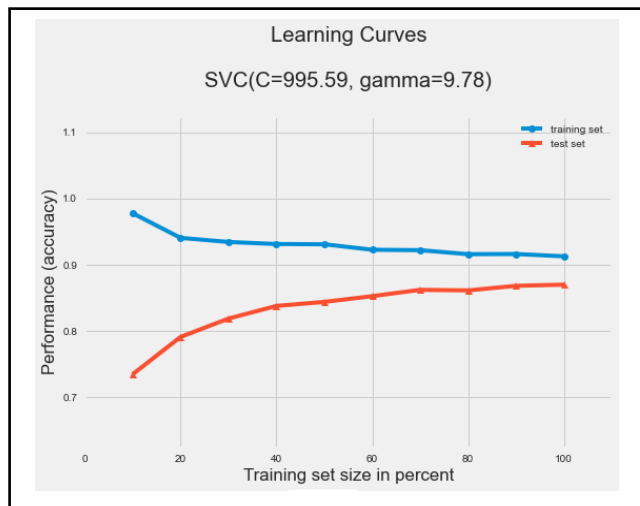
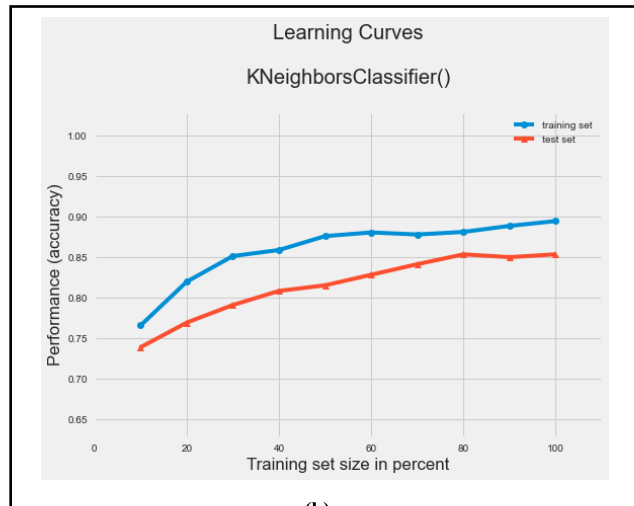


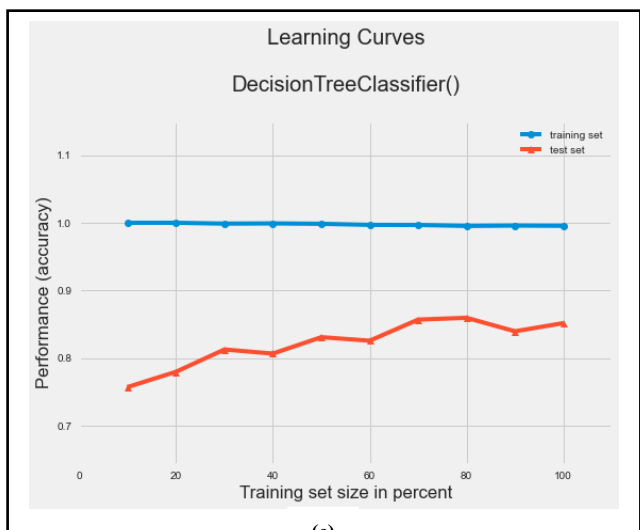
FIGURE 6. Methodology architecture.



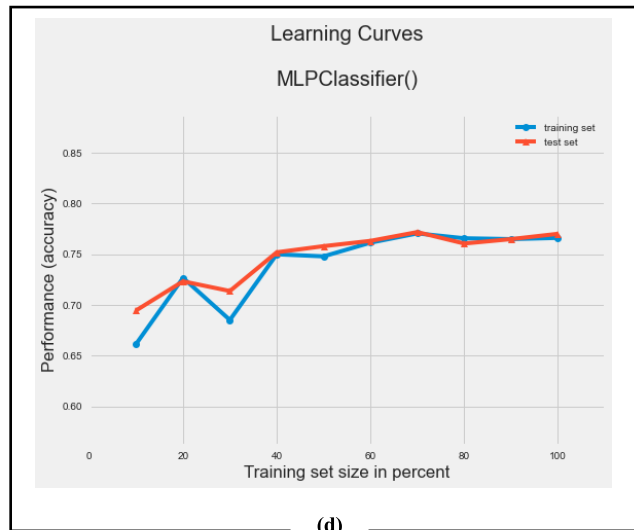
(a)



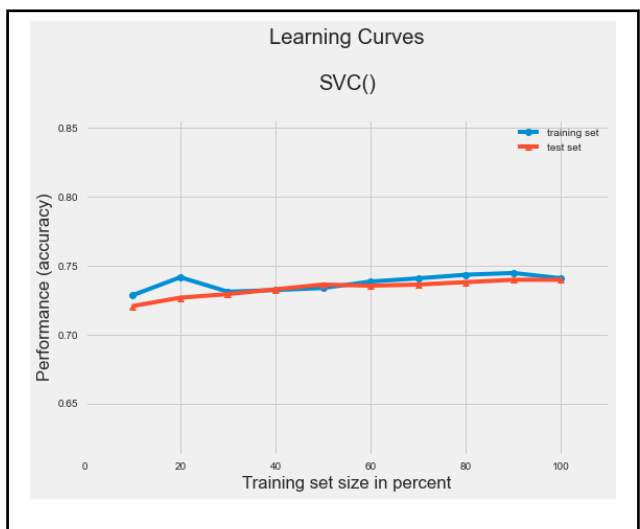
(b)



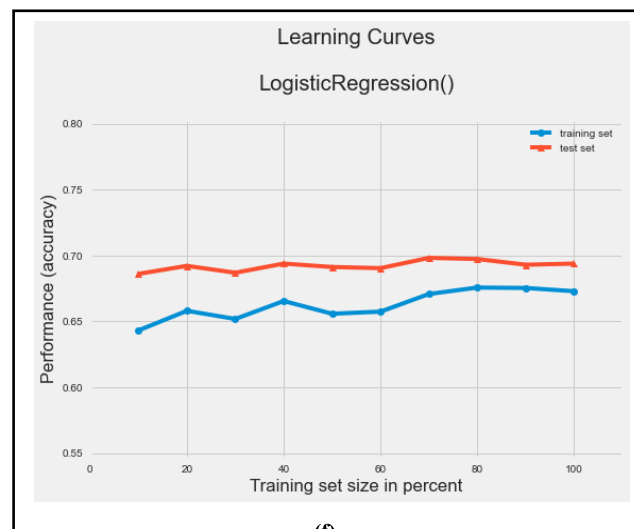
(c)



(d)



(e)



(f)

FIGURE 7. Learning curves obtained with the (a) SVM-GA model, (b) KNN model, (c) DT model, (d) MLP model, (e) SVM model, and (f) LR model.

TABLE 4. Statistical description of the raw data.

Variables	Count	Mean	Std	Min	25%	50%	75%	Max
Surface temperature	3843	0.610	4.622	-14.600	-1.500	1.100	3.300	14.200
Air temperature	3843	0.823	4.961	-20.000	-0.900	1.900	3.800	10.400
Base temperature	3843	-0.039	2.076	-4.600	-1.500	-0.100	1.400	4.600
Ground temperature	3843	0.599	3.755	-11.300	-1.200	0.800	3.000	10.300
Precipitation	3843	0.449	0.732	0.000	0.000	0.000	0.600	5.800
Dew point	3843	-2.393	4.860	-21.900	-4.200	-0.900	1.200	3.700
Snow height	3843	2.549	5.327	0.000	0.000	0.000	2.000	47.000
Grip	3843	0.750	0.149	0.110	0.780	0.810	0.820	0.820
Conductivity	3843	1.945	1.758	0.000	0.000	1.900	3.500	8.600
Concentration	3843	25.393	77.335	0.000	0.000	1.500	7.300	352.700
Chemical	3843	0.238	0.900	0.000	0.000	0.100	0.200	16.900

TABLE 5. Statistical description of the data after transformation.

Variables	Count	Mean	Std	Min	25%	50%	75%	Max
Surface temperature	3843	0.529	0.160	0.001	0.455	0.546	0.622	1.000
Air temperature	3843	0.685	0.163	0.001	0.629	0.721	0.783	1.000
Base temperature	3843	0.496	0.225	0.001	0.338	0.490	0.652	1.000
Ground temperature	3843	0.551	0.174	0.001	0.468	0.561	0.662	1.000
Precipitation	3843	0.078	0.126	0.001	0.001	0.001	0.104	1.000
Dew point	3843	0.762	0.190	0.001	0.692	0.820	0.902	1.000
Snow height	3843	0.055	0.113	0.001	0.001	0.001	0.044	1.000
Grip	3843	0.901	0.209	0.001	0.944	0.986	1.000	1.000
Conductivity	3843	0.227	0.204	0.001	0.001	0.222	0.408	1.000
Concentration	3843	0.073	0.219	0.001	0.001	0.005	0.022	1.000
Chemical	3843	0.015	0.053	0.001	0.001	0.007	0.013	1.000

TABLE 6. Information about DMUs and parts for WRM.

Time intervals	Number of intervals for 39 parts	Number of intervals in the last (40 th) part	Summation of number of intervals
00:00-16:00	32	33	(39*32)+33=1281
16:00-08:00	32	33	(39*32)+33=1281
08:00-00:00	32	33	(39*32)+33=1281

After the parameters of the SVM are optimized by the GA algorithm, the SVM model is run in Python software based on the optimized values of c and γ . The rest of the parameters are set according to the defaults in Python.

When the models are built for training, fivefold cross-validation is selected to validate the algorithms. Table 8 shows the mean accuracy and standard deviation of cross-validation and the accuracy score of the test set for the classification algorithms. The SVM-GA and KNN models exhibit the best mean cross-validation accuracy and the same standard deviation values; however, SVM-GA achieves better accuracy for the test set. (The all stages of the methodology is graphically shown in Figure 6.)

F. STAGE 6: MODEL EVALUATION

The numerical results from the previous stage provide a general overview of model performance, but we cannot deter-

TABLE 7. Optimal values of SVM parameters.

SVM parameter	Optimal value
c	995.59
γ	9.78

TABLE 8. Cross-validation results and accuracy of each model based on the test set.

ML supervised model	Mean accuracy of 5-fold cross-validation	Std of 5-fold cross-validation	Accuracy of the test set
SVM-GA	0.84	0.02	0.87
KNN	0.84	0.02	0.85
DT	0.83	0.01	0.85
MLP	0.76	0.02	0.78
SVM	0.74	0.01	0.74
LR	0.67	0.02	0.69

mine which model is best because we have to evaluate the models to determine if overfitting, underfitting, or fitting occurs. Learning curves allow us to monitor the learning process for the training and test sets and determine which model is the best fit for a given dataset. Learning curves are graphs that include training and test curves based on the training set size. The training curve reflects the learning procedure,

TABLE 9. Classification reports for each algorithm.

ML Algorithm	Precision		Recall		F1-score	
	-1	1	-1	1	-1	1
SVM-GA	0.86	0.88	0.85	0.89	0.86	0.88
KNN	0.83	0.88	0.85	0.85	0.84	0.86
DT	0.85	0.86	0.83	0.86	0.84	0.87
MLP	0.72	0.84	0.83	0.73	0.77	0.78
SVM	0.66	0.85	0.86	0.64	0.75	0.73
LR	0.66	0.72	0.67	0.71	0.67	0.72

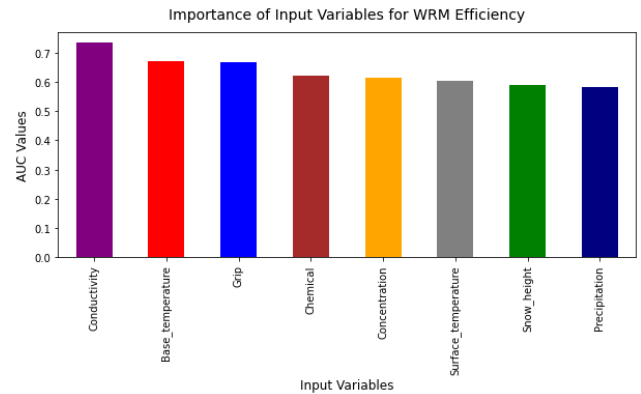


FIGURE 9. Importance of input variables affecting WRM efficiency.

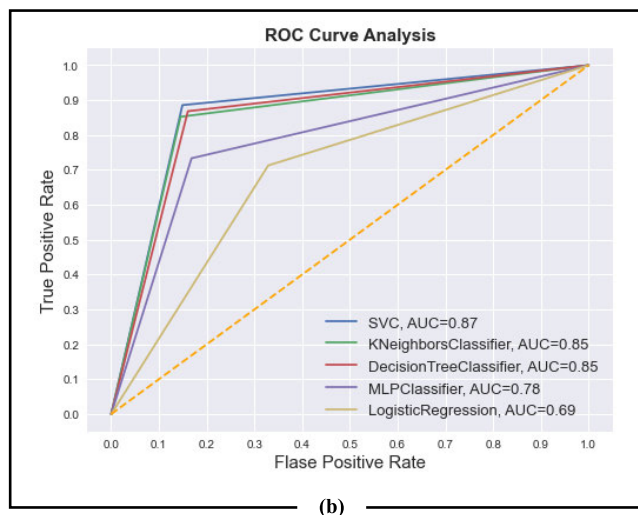
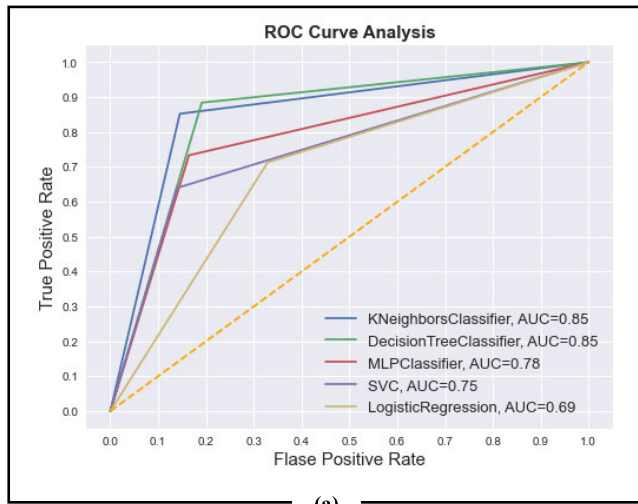


FIGURE 8. Comparing the AUC graphs and values of the KNN, DT, MLP and LR models with the results of the SVM model (a) without optimized parameters and (b) with optimized parameters.

whereas the test curve illustrates the generalization trend. Figure 7 presents the learning curves for the algorithms used in this article. Here, we briefly describe each of the curves: (a) SVM-GA: the gap between the learning and test curves reflects model variance, which is generally high when the size of the training set is small. Initially, the training set includes 10% of the samples, and the accuracy of the learning curve is 100%. This trend is typical because the algorithm does

not have any problem fitting so few data points. However, the accuracy of the test curve is approximately 74% because the algorithm is trained with only 10% of the training data points and is not capable of achieving an appropriate level of generalization for unseen observations. As the number of data points increases in the training set, the accuracy of the testing curve also increases, but the accuracy of the training curve decreases, which indicates that the algorithm is unable to predict all the data points in the training set. This trend continues until the two curves converge to the point called the Bayes error point (low variance and low bias); this result indicates that the SVM-GA is a good fit for our dataset. (b) KNN: in this figure, when the training set includes 10% of the samples, the accuracies of the training and test curves are approximately 76% and 74%, respectively; i.e., the gap between the training and test curves is low. As the training set increases in size from 10% to 80% of the samples, the accuracy curves converge; however, as training data points are added, the model performance varies (the accuracy for the training set increases, and that for the test set decreases). This trend suggests that by adding data points, the gap between the two curves becomes wider, and the curves no longer converge. (c) DT: the accuracy for the training set is 100% regardless of the size of the training set, which is a sign of severe overfitting. In addition, the accuracy of the test set increases as the size of the training set increases; however, there is a large gap between the two curves, indicating a high variance. Perhaps adding data or reducing model complexity would be helpful for overcoming these issues. (d) MLP: according to the training and test curves, the accuracy values are low, which reflects the poor performance of the model for both curves. In this situation, underfitting occurs for the model. (e) SVM: the accuracy for the test set remains generally stable, and there is no increase in accuracy as the size of the training set increases. This result indicates that the model is not able to learn from the data. In addition, the accuracy for the training set slightly increases until a training set size of 20% is reached. Then, the accuracy decreases until the training set size is 30%; consequently, an accuracy plateau

is observed. This behavior for the training set indicates that the model is underfitting the output, and the model may need to be adjusted (e.g., increase the complexity of the model). (f) LR: there are slight changes in both the training and test curves. Notably, the curves are almost stagnant throughout the learning and generalization processes, suggesting that the model cannot learn from the data. This strange result and the low accuracy scores demonstrate that the model is underfitting the output and performs poorly.

Overall, the proposed learning curves indicate that SVM-GA is a good fit for the considered dataset and can accurately predict efficiency classes. Additionally, AUC is another metric used to assess the ability of a prediction model. Figure 8 evaluates the AUC metric for the different algorithms considered in this study. Chart (a) compares the SVM model with other models before optimizing the SVM parameters, and chart (b) compares the SVM model with other models after optimizing the SVM parameters using the GA algorithm. As shown, the AUC value is enhanced after optimization, and the SVM-GA yields the highest AUC score among the considered models.

Table 9 illustrates the classification reports for the algorithms. We only analyze the results of classification reports for the SVM-GA algorithm, as it displays good performance based on the cross-validation results, the accuracy score for the test set, and the learning curves. In fact, the SVM-GA algorithm is able to accurately classify 86% of the samples in the inefficient WRM class and 88% of the samples in the efficient WRM class. In addition, the model correctly classifies 85% of inefficient WRM observations and 89% of efficient WRM observations. Moreover, the F1 score captures the properties of precision and recall, and the corresponding scores are 86% and 88% for the inefficient WRM class and efficient WRM class, respectively. Thus, the proposed algorithm achieves good performance in predicting efficient and inefficient observations, with slight superiority in predicting efficient WRMsamples.

V. RESULTS

In fact, analysis of the variables affecting the WRM and use of the WRM efficiency prediction model make it possible to drive future insights to reliably predict conditions at a specific time. WRM prediction model enables us to discover the relationship between input and output variables and ultimately leads to an improved decision-making process.

We started exploring what the model learned by plotting the importance of each input variable by building a machine-learning model that utilizes a single input variable individually to predict the target and calculate the performance metric. Next, these input variables were ranked according to the values of the performance metric. Figure 9 shows the importance of input variables in predicting WRM efficiency classes using the SVM-GA model and the AUC values. Next, we organized the features in descending order of AUC values (the higher the AUC value is, the better performance). As shown in the chart, conductivity, base tem-

perature, and grip are the top three factors affecting WRM efficiency prediction.

Furthermore, we used a scatter plot that colored the data points based on efficient and inefficient classes (Figure 10): the green points indicate inefficient WRM, and the orange points denote efficient WRM. The scatter plots produced some valuable findings:

(i) If the amount of salt used on the road and the conductivity of the chemical (salt) are low, the DMUs are more efficient. This trend makes sense as using more salt when the conductivity is also high leads to more inefficient DMUs because it is a waste of money, time, and material, in addition to the greater negative impact on the environment.

(ii) The number of inefficient DMUs increases when the base temperature drops because there is a need to use more salt on the road surface to increase safety. However, the number of efficient WRM cases increases when the base temperature increases. The same rule applies for precipitation.

(iii) It is not possible to draw a helpful conclusion from the grip scatter plot.

(iv) The presence of salt on the road surface demonstrates that there is no need to apply more salt on the ground to obtain efficient DMUs.

(v) The number of inefficient DMUs increases by using a high amount of salt when the surface temperature is approximately 0°C. In addition, the DMUs can be efficient when the surface temperature is drastically low by using low amount of salt on the road surface. This can be due to the presence of salt on the road surface.

(vi) When the snow height is almost 15 mm, applying chemicals (salt) leads to more inefficient DMUs. Under this condition, applying chemicals is clearly useless because there is a need to plow snow. However, when the snow height is between 0 and 9 mm, applying no salt or a low amount of salt can be helpful to increase the efficiency of the DMUs.

VI. CONCLUSION

In this research, a methodology was proposed to model slightly imbalanced binary WRM efficiency classes using a data envelopment analysis model and machine learning classification algorithms. In addition, several ML algorithms were implemented in Python software and compared based on real data collected at a road weather station and by optical and road-mounted sensors on European road E18 in Sweden. The results obtained in this study verify that the combined SVM-GA approach yields the best performance among the considered algorithms in discriminating to between efficient and inefficient WRM classes. Therefore, this algorithm can be used to predict WRM efficiency classes based on the output learning curves, AUC graphs, and precision, recall, and F1 score metrics. Furthermore, the results extracted from the methodology includes the analyzing factors affecting WRM to drive future insights in order to enhance decision support systems. This methodology is not limited to the prediction of WRM efficiency classes, and it can be used in various applications in which the prediction of efficiency classes is

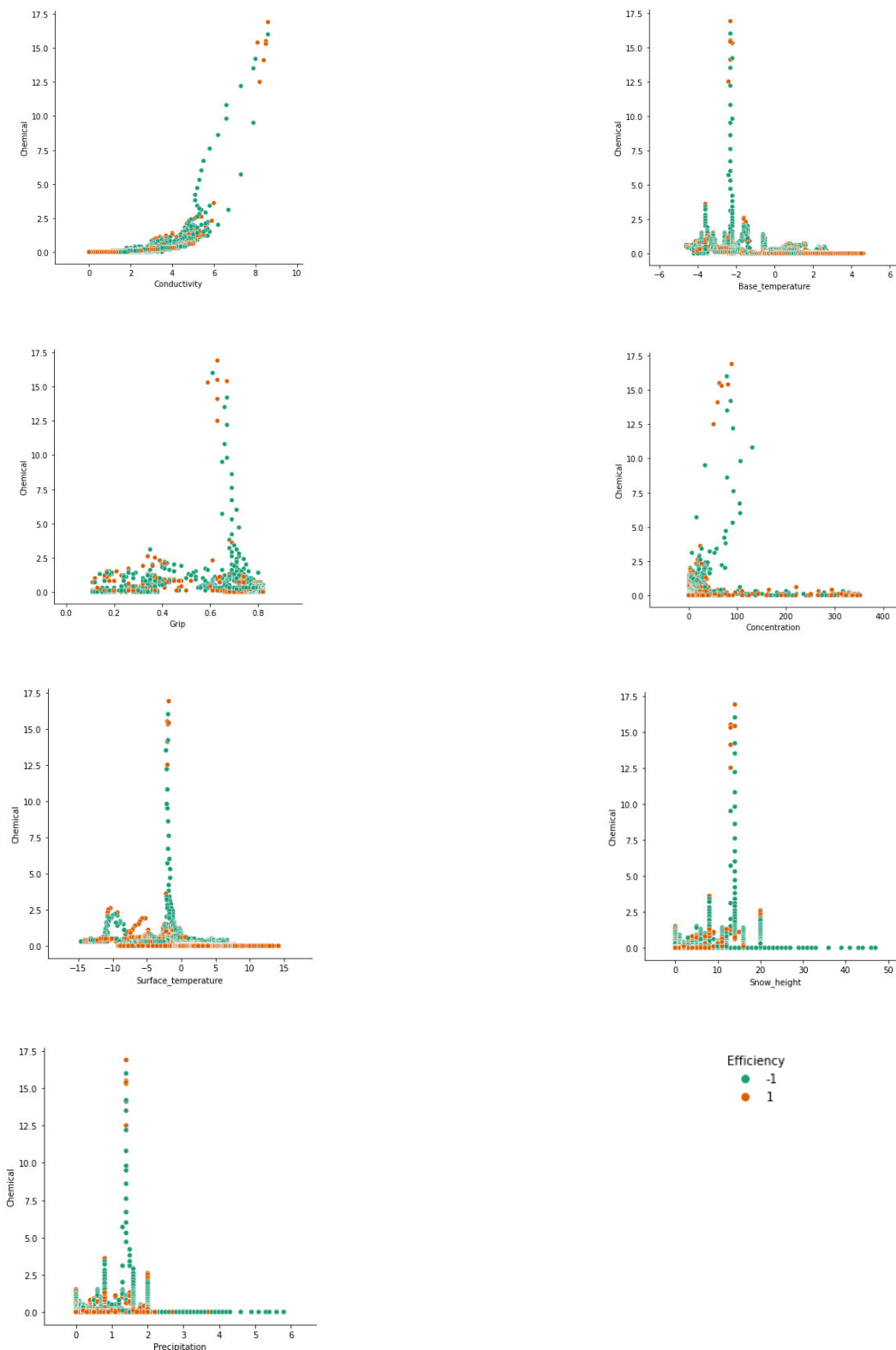


FIGURE 10. Scatter plots of input variables based on efficient and inefficient WRM.

important for developing management strategies and decision support systems. Generally, the performance of data-driven methods is mainly dependent on the quantity of data, qual-

ity of data, and adjustment of different parameters in the algorithms. In future work, connected (combined) physics-based modeling and data-driven modeling approaches will be

TABLE 10. List of abbreviations.

Abbreviation	Word
ANN	Artificial Neural Network
AUC	Area Under the (ROC) Curve
BCC	Banker, Charnes, and Cooper
CCR	Charnes, Cooper, and Rhodes
CRS	Constant Returns to Scale
Corr	Correlation
DEA	Data Envelopment Analysis
DMUs	Decision-Making Units
DSS	Decision Support Systems
DT	Decision Tree
FN	False Negative
FP	False Positive
GA	Genetic Algorithm
KNN	K-Nearest Neighborhood
LR	Logistic Regression
ML	Machine Learning
MLP	Multilayer Perceptron
ROC	Receiver Operator Characteristic
RWIS	Road Weather Information System
Std	Standard deviation
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
VRS	Variables Returns to Scale
WRM	Winter Road Maintenance

considered as alternatives to overcome the limitations of the current approach in terms of accuracy, system complexity, and speed (e.g., lower dependence on the number of data points).

LIST OF ABBREVIATIONS

Table 10 shows the abbreviations used in this study.

REFERENCES

- [1] J. Xiao, Z. Xiao, D. Wang, J. Bai, V. Havyarimana, and F. Zeng, "Short-term traffic volume prediction by ensemble learning in concept drifting environments," *Knowl.-Based Syst.*, vol. 164, pp. 213–225, Jan. 2019, doi: 10.1016/j.knsys.2018.10.037.
- [2] Z. E. A. Ellassad, H. Mousannif, and H. Al Moatassime, "A proactive decision support system for predicting traffic crash events: A critical analysis of imbalanced class distribution," *Knowl.-Based Syst.*, vol. 205, pp. 1–14, Oct. 2020, doi: 10.1016/j.knsys.2020.106314.
- [3] T. Usman, L. Fu, and L. F. Miranda-Moreno, "Quantifying safety benefit of winter road maintenance: Accident frequency modeling," *Accident Anal. Prevention*, vol. 42, no. 6, pp. 1878–1887, Nov. 2010, doi: 10.1016/j.aap.2010.05.008.
- [4] M. E. Ozbek, "Development of a comprehensive framework for the efficiency measurement of road maintenance strategies using data envelopment analysis," Ph.D. dissertation, Dept. Comput. Sci., Virginia Tech., Blacksburg, VI, USA, 2007. [Online]. Available: <http://hdl.handle.net/10919/29117>
- [5] W. A. Nixon and N. S. Foster, "Strategies for winter highway maintenance," *Transp. Vehicle Saf. Policy*, Univ. Iowa, Iowa City, IA, USA, Tech. Rep., 1996. [Online]. Available: https://ir.uiowa.edu/ppc_transportation/20
- [6] L. D. Minsk, *Snow and Ice Control Manual for Transportation Facilities*. New York, NY, USA: McGraw-Hill, 1998. [Online]. Available: <http://worldcat.org/isbn/0070428093>
- [7] B. Hallmark and J. Dong, "Examining the effects of winter road maintenance operations on traffic safety through visual analytics," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2020, pp. 1–6, doi: 10.1109/ITSC45102.2020.9294266.
- [8] T. Lorentzen, "Climate change and winter road maintenance," *Climatic Change*, vol. 161, no. 1, pp. 225–242, Jul. 2020, doi: 10.1007/s10584-020-02662-0.
- [9] S. Xu and T. J. Kwon, "Optimizing snowplow routes using all-new perspectives: Road users and winter road maintenance operators," *Can. J. Civil Eng.*, vol. 99, no. 999, pp. 1–10, 2021, doi: 10.1139/cjce-2019-0768.
- [10] C. Ahabchane, M. Trépanier, and A. Langevin, "Street-segment-based salt and abrasive prediction for winter maintenance using machine learning and GIS," *Trans. GIS*, vol. 23, no. 1, pp. 48–69, Feb. 2019, doi: 10.1111/tgis.12495.
- [11] S. Sedivy, L. Mikulová, Z. Florkova, and G. Gaspar, "Increasing the economic sustainability of winter road maintenance processes through the use of technological innovation," in *Proc. IOP Conf. Mater. Sci. Eng.*, 2021, vol. 1015, no. 1, Art. no. 012098.
- [12] D. D. Wu, Z. Yang, and L. Liang, "Using DEA-neural network approach to evaluate branch efficiency of a large Canadian bank," *Expert Syst. Appl.*, vol. 31, no. 1, pp. 108–115, 2006, doi: 10.1016/j.eswa.2005.09.034.
- [13] A. D. Athanassopoulos and S. P. Curram, "A comparison of data envelopment analysis and artificial neural networks as tools for assessing the efficiency of decision making units," *J. Oper. Res. Soc.*, vol. 47, no. 8, pp. 1000–1016, Aug. 1996, doi: 10.1057/jors.1996.127.
- [14] H. K. Hong, S. H. Ha, C. K. Shin, S. C. Park, and S. H. Kim, "Evaluating the efficiency of system integration projects using data envelopment analysis (DEA) and machine learning," *Expert Syst. Appl.*, vol. 16, no. 3, pp. 283–296, 1999, doi: 10.1016/S0957-4174(98)00077-3.
- [15] C.-C. Yeh, D.-J. Chi, and M.-F. Hsu, "A hybrid approach of DEA, rough set and support vector machines for business failure prediction," *Expert Syst. Appl.*, vol. 37, no. 2, pp. 1535–1541, Mar. 2010, doi: 10.1016/j.eswa.2009.06.088.
- [16] M. Mirmozaffari, A. Boskabadi, G. Azeem, R. Massah, E. Boskabadi, H. A. Dolatsara, and A. Liravian, "Machine learning clustering algorithms based on the DEA optimization approach for banking system in developing countries," *Eur. J. Eng. Res. Sci.*, vol. 5, no. 6, pp. 651–658, Jun. 2020, doi: 10.24018/ejers.2020.5.6.1924.
- [17] K. Vincova, "Using DEA models to measure efficiency," *Biaterc*, vol. 13, no. 8, pp. 24–28, 2005.
- [18] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *Eur. J. Oper. Res.*, vol. 2, no. 6, pp. 429–444, 1978, doi: 10.1016/0377-2217(78)90138-8.
- [19] A. Charnes, C. T. Clark, W. W. Cooper, and B. Golany, "A developmental study of data envelopment analysis in measuring the efficiency of maintenance units in the U.S. Air forces," *Ann. Oper. Res.*, vol. 2, no. 1, pp. 95–112, Dec. 1984, doi: 10.1007/BF01874734.
- [20] X.-W. Chen and X. Lin, "Big data deep learning: Challenges and perspectives," *IEEE Access*, vol. 2, pp. 514–525, 2014, doi: 10.1109/ACCESS.2014.2325029.
- [21] A. L'Heureux, K. Grolinger, H. F. Elyamany, and M. A. M. Capretz, "Machine learning with big data: Challenges and approaches," *IEEE Access*, vol. 5, pp. 7776–7797, 2017, doi: 10.1109/ACCESS.2017.2696365.
- [22] P. J. García-Laencina, J.-L. Sancho-Gómez, A. R. Figueiras-Vidal, and M. Verleysen, "K nearest neighbours with mutual information for simultaneous classification and missing data imputation," *Neurocomputing*, vol. 72, nos. 7–9, pp. 1483–1493, Mar. 2009, doi: 10.1016/j.neucom.2008.11.026.
- [23] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: A review of classification and combining techniques," *Artif. Intell. Rev.*, vol. 26, no. 3, pp. 159–190, Nov. 2006, doi: 10.1007/s10462-007-9052-3.
- [24] Y. Song, J. Huang, D. Zhou, H. Zha, and C. L. Giles, "IKNN: Informative K-nearest neighbor pattern classification," in *Proc. Eur. Conf. Princ. Data Mining Knowl. Discovery Berlin, Germany*: Springer, 2017, pp. 248–264, doi: 10.1007/978-3-540-74976-9_25.
- [25] A. Shen, R. Tong, and Y. Deng, "Application of classification models on credit card fraud detection," in *Proc. Int. Conf. Service Syst. Service Manage.*, Jun. 2007, pp. 1–4, doi: 10.1109/ICSSSM.2007.4280163.
- [26] A. Soofi and A. Awan, "Classification techniques in machine learning: Applications and issues," *J. Basic Appl. Sci.*, vol. 13, pp. 459–465, Aug. 2017, doi: 10.6000/1927-5129.2017.13.76.
- [27] H. H. Patel and P. Prajapati, "Study and analysis of decision tree based classification algorithms," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 10, pp. 74–78, Oct. 2018, doi: 10.26438/ijcse/v6i10.7478.

- [28] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Comput. Struct. Biotechnol. J.*, vol. 13, no. 1, pp. 8–17, 2015, doi: [10.1016/j.csbj.2014.11.005](https://doi.org/10.1016/j.csbj.2014.11.005).
- [29] R. Ramezani, A. Peymanfar, and S. B. Ebrahimi, "An integrated framework of genetic network programming and multi-layer perceptron neural network for prediction of daily stock return: An application in Tehran stock exchange market," *Appl. Soft Comput.*, vol. 82, Sep. 2019, Art. no. 105551, doi: [10.1016/j.asoc.2019.105551](https://doi.org/10.1016/j.asoc.2019.105551).
- [30] G. Panahandeh, E. Ek, and N. Mohammadiha, "Road friction estimation for connected vehicles using supervised machine learning," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2017, pp. 1262–1267, doi: [10.1109/IVS.2017.7995885](https://doi.org/10.1109/IVS.2017.7995885).
- [31] A. H. Nizar, Z. Y. Dong, and Y. Wang, "Power utility nontechnical loss analysis with extreme learning machine method," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 946–955, Aug. 2008, doi: [10.1109/TPWRS.2008.926431](https://doi.org/10.1109/TPWRS.2008.926431).
- [32] D. Liu, D. Niu, H. Wang, and L. Fan, "Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm," *Renew. Energy*, vol. 62, pp. 592–597, Feb. 2014, doi: [10.1016/j.renene.2013.08.011](https://doi.org/10.1016/j.renene.2013.08.011).
- [33] Y. Ren and G. Bai, "Determination of optimal SVM parameters by using GA/PSO," *J. Comput.*, vol. 5, no. 8, pp. 1160–1168, Aug. 2010, doi: [10.4304/jcp.5.8.1160-1168](https://doi.org/10.4304/jcp.5.8.1160-1168).
- [34] M. Zhao, C. Fu, L. Ji, K. Tang, and M. Zhou, "Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5197–5204, May 2011, doi: [10.1016/j.eswa.2010.10.041](https://doi.org/10.1016/j.eswa.2010.10.041).
- [35] C.-H. Wu, G.-H. Tzeng, and R.-H. Lin, "A novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 4725–4735, Apr. 2009, doi: [10.1016/j.eswa.2008.06.046](https://doi.org/10.1016/j.eswa.2008.06.046).
- [36] W. Chen and H.-M. Yuan, "An improved GA-SVM algorithm," in *Proc. 9th IEEE Conf. Ind. Electron. Appl.*, Jun. 2014, pp. 2137–2141, doi: [10.1109/ICIEA.2014.6931525](https://doi.org/10.1109/ICIEA.2014.6931525).
- [37] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, *Learning From Imbalanced Data Sets*. Cham, Switzerland: Springer, 2018, doi: [10.1007/978-3-319-98074-4](https://doi.org/10.1007/978-3-319-98074-4).
- [38] J. Brownlee. (2020). *A Gentle Introduction to Imbalanced Classification*. [Online]. Available: <https://machinelearningmastery.com/what-is-imbalanced-classification/>
- [39] C. Sammut and G. I. Webb, Eds., *Encyclopedia of Machine Learning and Data Mining*, 2nd ed. New York, NY, USA: Springer, 2017, doi: [10.1007/978-1-4899-7687-1](https://doi.org/10.1007/978-1-4899-7687-1).
- [40] J. Brownlee. (2020). *How to Calculate Precision, Recall, and F-Measure for Imbalanced Classification*. [Online]. Available: <https://machinelearningmastery.com/precision-recall-and-f-measure-for-imbalanced-classification>
- [41] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and J. Vanderplas, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011. [Online]. Available: https://scikitlearn.org/stable/modules/generated/sklearn.metrics.classification_report.html
- [42] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, and R. Kern, "Array programming with numpy," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020, doi: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2).
- [43] W. McKinney, "Data structures for statistical computing in Python," in *Proc. 9th Python Sci. Conf.*, 2010, pp. 51–56.
- [44] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, May/Jun. 2007, doi: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55).
- [45] S. Raschka, "MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack," *J. Open Source Softw.*, vol. 3, no. 24, p. 638, Apr. 2018, doi: [10.21105/joss.00638](https://doi.org/10.21105/joss.00638).
- [46] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, and S. J. Van Der Walt, "SciPy 1.0: Fundamental algorithms for scientific computing in Python," *Nature Methods*, vol. 17, pp. 261–272, Feb. 2020, doi: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).
- [47] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and J. Vanderplas, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.
- [48] M. K. Epstein and J. C. Henderson, "Data envelopment analysis for managerial control and diagnosis," *Decis. Sci.*, vol. 20, no. 1, pp. 90–119, Mar. 1989, doi: [10.1111/j.1540-5915.1989.tb01399.x](https://doi.org/10.1111/j.1540-5915.1989.tb01399.x).
- [49] M. E. Ozbek, J. M. de la Garza, and K. Triantis, "Efficiency measurement of bridge maintenance using data envelopment analysis," *J. Infrastruct. Syst.*, vol. 16, no. 1, pp. 31–39, 2010, doi: [10.1061/\(ASCE\)1076-0342\(2010\)16:1\(31\)](https://doi.org/10.1061/(ASCE)1076-0342(2010)16:1(31)).



MAHSHID HATAMZAD received the B.S. and M.S. degrees in industrial engineering, in 2010 and 2013, respectively, and the M.S. degree in risk management from the University of Stavanger, in 2018. She is currently pursuing the Ph.D. degree in industrial engineering with the UiT/The Arctic University of Norway. Her main research interests include data analysis, machine learning, and uncertainty assessment.



GEANETTE POLANCO PINERIZ received the B.S. and M.S. degrees in mechanical engineering, in 1997 and 2000, respectively, and the Ph.D. degree in mechanical engineering from Coventry University, in 2009. She is currently an Associate Professor at the UiT/The Arctic University of Norway. Her research interests include fluid mechanics, optical measurement techniques, and winter road maintenance.



JOHAN CASSELGREN received the M.S. degree in applied physics and the Ph.D. degree from the Luleå University of Technology, in 2005 and 2011, respectively. In 2010, he returned to the Luleå University of Technology as a Researcher. In 2019, he became an Associate Professor. His research interests include optical measurement techniques, focusing on winter road conditions and winter road maintenance.