https://doi.org/10.1016/j.autcon.2020.103476

Automated Rebar Diameter Classification using Point Cloud Data based Machine Learning

Abstract

Inspecting the diameter and spacing of rebar is an important task conducted by fabricators and site engineers during the manufacturing and construction stages. This is because the bearing capacity of reinforced concrete structures is affected by the size and position of the rebar, so installing rebar of the correct size and position should be ensured to safeguard the structural integrity of the structure. This study presents a new terrestrial laser scanning (TLS)-based method using machine learning to automatically classify rebar diameters and accurately estimate rebar spacing. To this end, a new methodology, named density based machine model, is proposed to improve classification accuracy. To validate the proposed method, experimental tests on laboratory specimens with rebars of seven different diameters are conducted. The results show that the prediction accuracy for large rebar diameters measuring D25-D40 are up to 97.2%, demonstrating great potential for the application of the proposed technique on manufacturing and construction sites. The key findings of the study are: (1) the proposed DBM method for rebar diameter prediction is superior to the traditional machine learning approach; (2) scan density is one of the most important factors in the prediction results, especially in the small rebar diameter group; and (3) a scan density value of at least 80 pts/mm^2 computed on the cross section plane for a rebar instance with 100 mm length is necessary for successful rebar diameter prediction. It is expected that the proposed rebar diameter and rebar spacing technique will be useful in providing autonomous and accurate rebar inspection in manufacturing factories and on construction sites.

Key words: Rebar diameter, classification, machine learning, point cloud data, laser scanning

1. Introduction

Inspecting rebar diameters and rebar spacing is important for fabricators and site engineers to check dimensional compliance with the design model during the manufacturing and construction stages. This is because the bearing capacity of the reinforced concrete structures is dictated by the size and position of the rebar. Therefore, rebar of the correct size should be installed in the correction position, as determined by the blueprints, to ensure the structural integrity of the reinforced concrete structure. In this regard, dimensional inspection of rebars is conducted primarily by qualified workers to detect any abnormalities related to rebar diameter and rebar spacing using measurement tapes. However, this is a time-consuming and labor-intensive task, so there is an urgent need for automated rebar diameter and rebar spacing measurement that can save time and increase the reliability of the inspection.

Thanks to improvements in 3D sensing technology, several studies have been conducted into dimensional inspection of prefabricated RC components such as precast slabs [1–5] and precast girders [6] over the past decade. However, there have been relatively few studies [7-9] into rebar inspection. Recently, the authors' group proposed a technique that estimates the dimensions of rebar and formwork using the RANdom SAmple Consensus (RANSAC) [10] based on terrestrial laser scanning (TLS) approach. However, the prior study assumes rebar diameters are known prior to use as the input parameters of the RANSAC algorithm, making the algorithms unreliable due to the manual input. To fully automate the process, rebar diameters will need to be predicted from raw rebar scan data without any manual input. However, determining rebar diameters is a challenging task due to the small size of the rebar and irregular shapes of the rebar surfaces. In addition, diameter prediction in the current circle fitting methods remains a major challenge in the presence of noise, outliers, distortion, and missing boundaries in the unregistered scan data [11]. In order to tackle these technical issues, this paper aims to develop a TLS-based rebar diameter classification technique that automatically classifies rebar diameters using a machine learning approach. In this study, a new concept of density-based machine model is proposed and validation tests are conducted to demonstrate the applicability of the proposed rebar diameter prediction technique. The uniqueness of the study are (1) the development of a rebar diameter classification technique for the first time; and (2) successful applicability validation of the proposed technique through various tests including comparison tests with traditional methods.

This paper is organized as follows. Related background of the study and state-of-the-art studies is detailed in Section 2, followed by explanation of the proposed method and its procedure in Section 3. Next, validation tests and results are presented and interpreted comprehensively in Section 4. The primary factors affecting the result of the proposed method are discussed in Section 5. Finally, this study is concluded with a summary and future work.

2. Research Background

2.1 Overview of circle fitting methods

There are three main approaches in classical circle fitting, which are 1) geometric, 2) algebraic, and 3) robust fitting. Geometric fitting minimizes the sum of the squared geometric (orthogonal) distance from the estimated circle to the given data points. According to [12], geometric fit is commonly regarded as the most accurate, but it is implemented by iterative schemes that are computationally intensive and subject to occasional divergence. Another limitation of geometric fitting is that its accuracy depends on the choice of the initial parameters of center circle and diameter of the circle [12-13]. Algebraic fitting is a form of non-iterative fitting that minimizes the approximate (algebraic) distances by determining the constraint of the algebraic equations based on mathematical law and theory [12]. Algebraic fits are faster than geometric fitting, which is an advantage for large size point cloud data. Kasa [14] developed a simple and fast algebraic fitting method that has been used as a basis of many circle fitting methods. However, Al-Sharadqah and Chernov [15] found that the method introduced by Kasa [14] is heavily biased toward smaller circular arcs. This limitation was studied by Pratt [16] and Taubin [17] who each developed a popular method by changing the parameter constraints of the Kasa fit. Al-Sharadqah and Chernov [15] subsequently developed a method, called 'Hyper', which works by minimizing the algebraic functions associated with the two constraints used by Pratt [16] and Taubin [17]. On the other hand, Gander-Golub-Strebel (GGS) fit [18] developed the concept of the least squares for estimating the circle center points and radius. Nievergelt [19] enhanced the GGS fit by using translation shifts of the center coordinates and the 'Constrained Total Least-Squares' concept for approximation in the algebraic objective function. However, the geometric and algebraic algorithms has limitations – namely, the presence of outliers.

To counter the effect of outliers, several researchers have proposed robust statistical methods for circular fitting. Wang and Suter, for example [20], developed the Least Trimmed Symmetry Distance (LTSD) approach for robust model fitting using Symmetry Distance, coupled with a Least Trimmed Square (LTS) regression. However, their method is limited to spatially symmetric points [21]. The RANSAC, one of the most widely used methods, was developed in order to cope with high presence of outliers. The RANSAC iteration process starts by randomly fitting a model with pre-determined threshold and computes number of points located inside the area of the RANSAC model. At the end of iteration, an optimal RANSAC model with maximum number of points is selected. However, the RANSAC algorithm is sensitive to thresholding parameters [22-23], which needs to be selected manually. In order to overcome this problem, a method to determine the RANSAC parameters is required. In this study, a machine learning approach that enables automatic determination of RANSAC input parameters is presented.

2.2 Circle fitting applications in the construction industry

In recent years, studies on circle fitting have been conducted using 3D measurement sensors such as Kinect and TLS. Nahangi et al. [24], for example, put forward a curvature estimation method for estimating the cylinder radius of pipes using Kinect. The study proposed a radius estimation algorithm involving normal vector estimation and curvature computation and shows an accuracy ranging from ±1.1 cm to ±2 cm. Separately, Díaz-Vilariño et al. [25] used the Hough Transform algorithm [26] to detect and estimate the radius of columns in a residential building. The study found that the Hough Transform algorithm is sensitive to the circumference of columns. In addition, Bueno et al. [27] examined the Hough Transform algorithm reliability using simulated column data. The study shows that completeness of edges is the most sensitive variable affecting the performance of the Hough transform algorithm and it was found that at least 30% to 40% edge completeness is required to robustly estimate the radius of columns. In addition, to deal with incomplete circular edges and noisy data, Nurmunabi et al [21] introduced a circle fitting method named Repeated Least Trimmed Squares (RLTS) that combines the concept of the least trimmed square [28] and the hyper-algebraic fitting [15]. The RLTS was validated with 1000 quarter-circle datasets with noise, and the results showed an average mean square error of 0.42. In summary of the circle fitting applications in the construction industry, the existing literature are focused on estimating the circular radius of target objects having a circular surface. However, considering the focus of this study that is rebar diameter classification, circle fitting and rebar diameter estimation is a challenging task because construction rebar has deformed shapes and its diameter is relatively small. In addition, the rebar diameter gaps between different sizes of rebar is relatively low, e.g. D10 and D12.

2.3 Rebar inspection studies in the construction industry

Thanks to the fast and accurate nature of laser scanning, there have been multiple studies conducted on sensing-based rebar and formwork inspection. Han et al. [8] used a density histogram of scan points generated based on the Structure-from-Motion (SfM) and multi-view stereo algorithms for rebar layout inspection. In the study, a validation test was performed using fifteen targets placed near the rebar, and 850 images were used for the implementation of SfM to generate a set of point cloud data. However, the study focused on the generation of a dense point cloud data from images and numerous high-resolution photos were required from different angles to generate the formwork and rebar scan points. Subsequently, Akula et al. [7] used a drilling monitoring framework that maps the locations of rebars using the SfM and laser scanning to provide real-time feedback for for drill operator based on the position and orientation of the drill. A comparison analysis in the study showed that the accuracy of the vision-based technique was 28.9% lower than the laser scanning-based method. Nishio et al. [29] conducted a study with a rebar core wire extraction algorithm designed to perform well in noisy scan data. A density distribution function was proposed in the study to filter out unwanted scan points near the rebars. However, the study was focused on the extraction of rebar scan points, and there was no further analysis of rebar spacing and rebar diameter estimation. Recently, Kim et al [11] proposed a new technique that estimates the dimensions of rebar and formwork, concrete covers and rebar spacing. The study used a RANSAC algorithm to automatically detect and estimate rebar layers. Validation results showed an estimation accuracy of around 2.5 mm. However, the study assumed that the rebar diameters used for input parameters of the RANSAC algorithm were known, making the method unreliable. In order to achieve full automation and make the method reliable, rebar diameters needs to be classified accurately from raw rebar scan data. For this, a rebar classification technique that predicts correct rebar diameters is needed to ensure accurate and automated rebar spacing estimation during the fabrication stages or on construction sites.

2.4 Point cloud data based machine learning application in the construction industry

There have been multiple studies that have looked into adopting machine learning methodologies with point cloud data in the construction industry. Wang et al. [30] developed an automated technique for position estimation of rebars in reinforced precast slabs. A one class SVM, which is an unsupervised outlier detection method, was used to extract rebars based on the geometric and color features of scan points. Experiments on two reinforced precast concrete decks were conducted, and the results showed a rebar position estimation error of 0.9 mm. Valero et al. [31] proposed a strategy for the automatic detection and classification of defects on masonry walls. They achieved this using geometric and color features including roughness, circularity and RGB values, for their machine learning approach. Lastly, Bassier et al. [32] proposed an automated classification method for building elements for Scan-to-BIM. A

Random Forests classifier was used for the classification of the floors, ceilings, roofs, walls and beams. Both contextual and geometric features were used, culminating in an average precision of 85% for structural element classification. In summary, a few machine learning methods have been implemented using point cloud data previously. However, these previous studies are limited in that they mainly focus on the extraction of planar-type elements including roofs, walls and slabs.

2.5 Research gaps and objectives

Based on the review of related literature, there has been no study conducted into rebar diameter classification using point cloud data. In addition, the direct use of circle fitting methods, including geometrical and algebraic fits, may not be accurate and robust due to the small size and potentially deformed shape of rebars. Moreover, the RANSAC, which is a robust circle fitting method, requires initial manual parameters including the circle radius and tolerance threshold. To overcome these gaps in research the objectives of this study are to 1) develop a technique that classifies rebar diameters using machine learning approach for accurate and automated estimation of rebar spacings; and 2) validate the feasibility of the proposed rebar classification technique on rebars with various diameters.

3. Methodology

Figure 1 shows the workflow of the proposed rebar diameter classification, which consists of three stages: 1) training, 2) prediction and 3) estimation stages. The training stage is composed of five sub-steps, including data collection, data pre-processing, feature extraction, feature selection and machine learning model selection. In the prediction stage, new data sets are prepared and predicted using the learning model chosen in the training stage. Note that there is no step involving feature selection, as optimal features are already determined in the feature selection step of the training stage. In the last stage, rebar spacings between adjacent individual rebars are estimated based on the predicted rebar diameters in the rebar estimation stage. Details of each step are presented in the following sections.

3.1 Data collection

In the first step of the training process, deformed rebars with various rebar diameters are scanned by a TLS in the laboratory to collect rebar point cloud data. In this study, seven rebar diameters of 10, 12, 16, 20, 25, 32 and 40 mm, named as D10, D12, D16, D20, D25, D32 and D40, which are widely used in the Hong Kong construction industry, are selected and used for the data collection. Three scan variables including the incident angle, distance and angular resolution, are controlled in the data collection to collect rebar surface point cloud data in as many different conditions as possible to ensure the generation of a high-accuracy machine learning model.

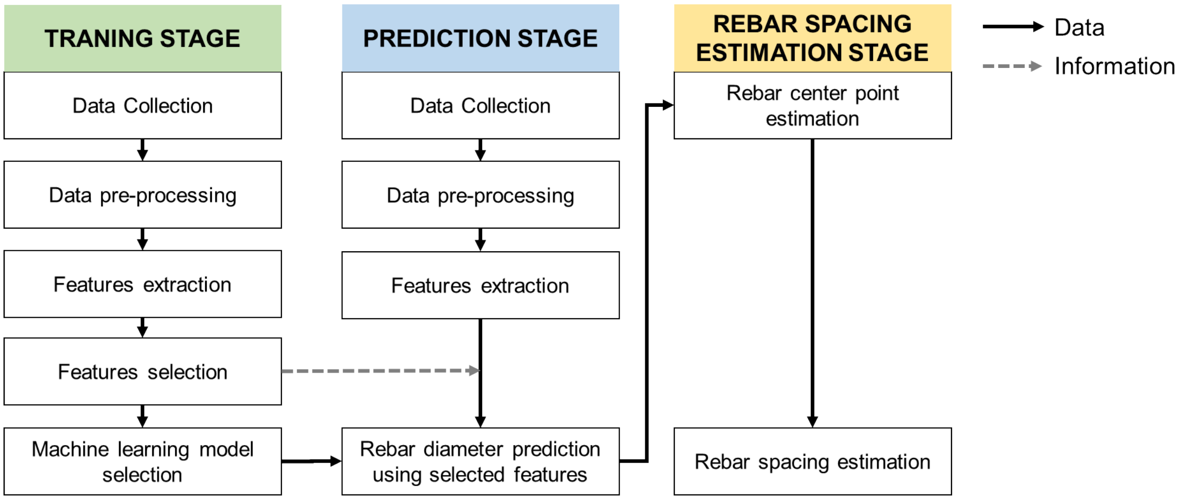

Figure 1. Overall procedure of the proposed rebar diameter classification technique for autonomous rebar spacing estimation

3.2 Data pre-processing

It is important to eliminate unwanted scan points near the rebar scan data including background noise and mixed pixels prior to the feature extraction of each rebar instance. Thus, the data pre-processing step aims to 1) remove background noise and mixed pixels in raw scan points and 2) slice the individual rebars into multiple small instances to be used for the next step, i.e. feature extraction. Figure 2 shows the data pre-processing steps consisting of three steps: 1) the elimination of the background noise and mixed pixels; 2) rebar rotation; and 3) rebar slicing and angle adjustment. Firstly, the DBSCAN (Density-based spatial clustering of applications with noise) [34] is used to segment the rebar scan points with respect to scan density and to remove the scan points of the background noises and mixed pixels [35]. Secondly, the segmented scan points of rebars are rotated by angle (α) between the 1st principal axis of the rebar scan points and the X axis as shown in Figure 2(b) for the purpose of easy further data processing. Third, the rotated rebars are sliced into multiple instances and the scan points of the sliced rebar instances are then projected into the YZ plane. Note that a minor rotation by the angle β (defined as the angle discrepancies between X axis and 1st principal axes of the sliced rebar part) is performed before the projection to ensure that the rebar part is parallel to the X axis as illustrated in Figure 2(c). Here, the projected scan points of each sliced on to the YZ plane is used to implement the feature extraction.

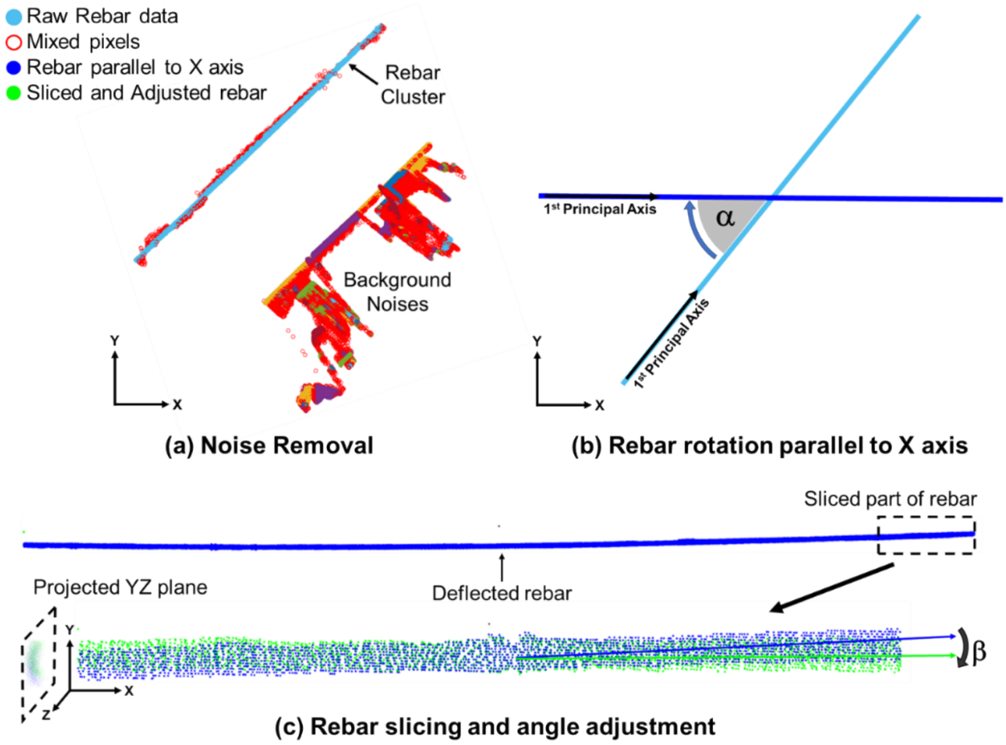

Figure 2. Data pre-processing steps: (a) Noise removal - mixed pixels and background noises are eliminated as sets of rebar scan data clusters using the DBSCAN; (b) Rebar rotation - the 1st rebar principal axis is computed then rotated by α degree to make it parallel to the X axis; (c) Rebar slicing and angle adjustment - the individual rebar is sliced, rotated by β degree and projected onto the YZ plane.

3.3 Features extractions

The purpose of this step is to extract the key features of rebar instance for machine learning. For this study, several features of algebraic fitting and principal axes of the instance's scan points were used due to the fact that they are sensitive to geometrical properties of rebar diameter. Unlike previous studies [33, 36-38], which used non-geometrical features such as material intensity [37], surface color [33] and surface roughness [36, 38], this study used seven geometrical features that can be extracted from the algebraic fittings and PCA [39] because the material property of rebar was assumed to be same.

Firstly, six algebraic fittings method were chosen to extract the six geometrical features: Kasa [14], Pratt [16], Taubin [17], Hyper [15], Nievergelt [19], GGS [18]. This is because no parameters are required for the algebraic fittings. Note that the outputs of the algebraic fitting that uses point cloud data of rebar instance as input are 1) estimated rebar diameter and center point of the circle fit. Figure 3 shows the feature values of the six algebraic fittings. The estimated rebar diameter values vary largely depending on fitting method. It is for this reason that the 6 diameter features are used in the study for machine learning instead of the directly use of estimated rebar diameters based on algebraic fitting methods. Secondly, the 2nd eigenvalue of point cloud of the rebar instance is used as the 7th feature. Figure 3 shows that the 2nd eigenvalue feature indicates the length of the point cloud data in the cross section and is closely associated with rebar diameter. In other words, the larger rebar diameter is, the larger the 2nd eigenvalue obtained, illustrating that the 2nd eigenvalue of rebar instance can be used as a unique feature that represents the rebar diameter. Note that only 2nd eigenvalue feature is used among the three eigenvalues because 1) the 1st eigenvalue (λ_1) indicating the length of sliced rebar instance has the same value of 100 mm; and 2) the 3rd eigenvalue (λ_3) is unreliable because of its small range of variation, as seen in Figure 3.

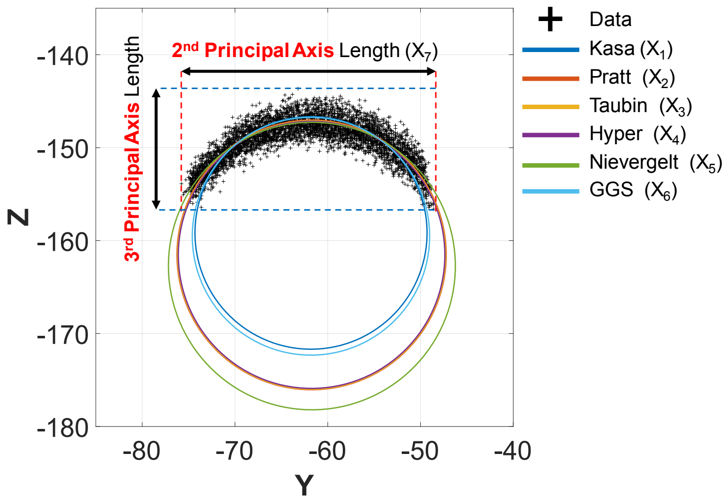

Figure 3. Features extraction of rebar instance in the YZ plane

3.4 Features selection

Feature selection aims to identify the most relevant or key features that increase machine learning prediction accuracy and reduces computational time [40]. In this study, correlation-based selection (CFS) [41] , which is a robust method in selecting features when there are linear relationships between feature values and predicted values [42], was utilized because the extracted feature values of rebar instance tend to have a linear relationship with the predicted rebar diameters. The CFS selects the best subsets of features using the correlation coefficients of features which can be computed using Pearson's coefficient [43]:

$$r_{xy} = \frac{\sum_{i} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i} (x_i - \bar{x})^2} \sqrt{\sum_{i} (y_i - \bar{y})^2}} \qquad \text{Eq. (1)}$$

where x and y are the series of features with the i^th number, $\bar{x}$ and $\bar{y}$ are the average value of the x and y array. r_{xy} is the Pearson's correlation value ranging from -1 to 1. The larger the absolute value of r_{xy}, the more strongly x and y are correlated. If the value is zero, the x and y variables are independent each other. In the CFS, a scoring method called Merit [41] is used to determine the best sets of features using the correlation coefficients of the features. Here, $Merit_{Sk}$ is the Merit score of a feature subset S consisting of k features.

$$Merit_{Sk} = \frac{k\,\overline{r_{fc}}}{\sqrt{k + k\,(k-1)\,\overline{r_{ff}}}} \qquad \text{Eq. (2)}$$

where $\overline{r_{fc}}$ is the average value of all feature-classification correlations and $\overline{r_{ff}}$ is the average value of all feature-feature correlations. Finally, the feature subset with the highest Merit score is determined as the features to be used for machine learning.

3.5 Machine learning classification test and selection

To classify different rebar diameters, machine learning classifiers are used to learn unique data patterns based on the selected features. Although machine learning is able to find key patterns from a set of training data [44-45], as well as accurately predicting classes based on the patterns found, the performance of rebar prediction varies with different classifiers. Thus, it is necessary to test diverse classifiers to see which classifier fits best the rebar scan data. To this end, five supervised machine learning algorithms, which are the Naïve Bayes (NB) [46-47], Discriminant Analysis (DA) [48-49], Classification Tree (CT) [50], Nearest Neighbor (NN) [51-52], and Support Vector Machines (SVM) [53] were employed in this study to find the optimal classifier. For the evaluation of the performance, the 10-fold cross-validation [54] method was used to decrease the bias and variance of the classifiers. Finally, the optimal model with the best features and classifier was used for predicting rebar diameters in the prediction stage.

3.6 Density-based modeling

In this study, a new methodology, named Density based Modeling (DBM), is proposed for the purpose of enhancing the classification accuracy. The DBM was developed based on observations from this study as well as previous related studies [1-5, 11, 31-33] that a higher scan density results in a higher accuracy. For example, rebar instances with high scan density form much clearer circular shape compared to those with low scan density. In this regard, density-based modeling is important in increasing the accuracy of machine learning classifiers. Figure 4 shows the concept of the proposed DBM method. The training data sets can be divided into a certain number of density groups so each density group has an equal or a similar percentage of training data. For instance, if the training data is divided into 4 density groups, every density group will have 25% of the training data. Once the data has been divided, 7 features are extracted for every instance and feature selection and machine learning selection are performed independently for each group.

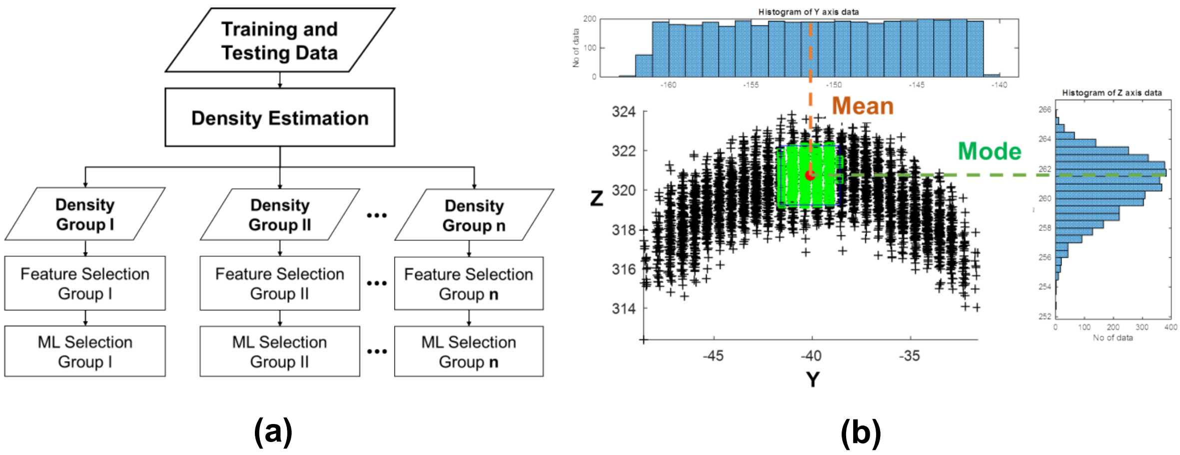

Figure 4. Density-based modeling (DBM): (a) The schematic of the DBM. The training data sets are divided into a number of density groups for machine learning; and (b) The estimation of scan density for a rebar instance

In order to estimate the scan density of each instance, a small area in the center of the scan points in the YZ plane is used. Figure 4(b) shows the scan density calculation in a rebar instance. Note that the Y and Z coordinates of the center point of each rebar instance in the YZ plane are calculated as the mean of the Y coordinates and the mode of Z coordinates of the scan points, respectively. Also note that the area with the dimensions of 9 mm^2 (3 mm × 3 mm) was used for the scan density calculation.

3.7 Rebar diameter prediction and rebar spacing estimation

Once a machine learning model was determined for each density group using the DBM method, the prediction of rebar diameters was performed on new rebar scan data with different rebar diameters using the optimized machine learning model. After the rebar diameter prediction, the rebar spacings between adjacent rebars was estimated using the predicted rebar diameters. The details of rebar spacing estimation, including the process of rebar center point estimation, are presented in [11]. It is important to note that compared to the previous study [11], the proposed method provides automatic rebar spacing estimation based on predicted rebar diameters, whereas two manual inputs, i.e, circle radius and tolerance threshold, are necessary for the implementation of the circular RANSAC in the previous literature.

4. Experimental Validation

4.1 Data collection for training

Figure 5 shows the two different data collection set-ups. A phase-shift TLS, FARO M70, with a measurement accuracy of ±3 mm in a range of 0.6 m to 70 m and a measurement rate of up to 488,000 points/sec., was used for data acquisition. For the first set-up as shown in Figure 5(a), 14 individual rebars with 7 different diameters (D10-D40), as shown in Figure 5(c), and a standard length of 3 m, were scanned. Note that the rebar layers were scanned at 14 random scan positions with two different scan angular resolutions of 0.036° and 0.072° to collect extensive training data with different scan densities, resulting in a total of 11,099 rebar slicing instances from the first data collection. The second data collection, configured as shown in Fig. 5(b), was focused on collecting scan data of rebar diameters of 12 mm and 16 mm at scan angles of 45° and 90° with respect to the rebar cage. For this configuration, the TLS was positioned on a desk to scan the entire rebar cage and a total of 2,888 instances of rebar diameter 12 mm and 16 mm were collected. In total, 13,987 rebar slicing instances were collected.

4.2 Feature selection and machine learning classifier selection using the DBM

Table 1 shows the results of the feature set selection based on the CFS in three different scan density groups. Note that each density group has its own feature set selected based on the Merit Score. It was found that the features of 1, 6 and 7 corresponding to Kasa fit, GGS fit and the 2nd principal axis length respectively were selected in most cases. As can be seen in the table, the primary finding was that the best performance on the feature selection was obtained in the highest density group, i.e. density group 3. For example, the selected feature sets in the density group 3 provided higher Merit scores when the number of features in the feature set were same.

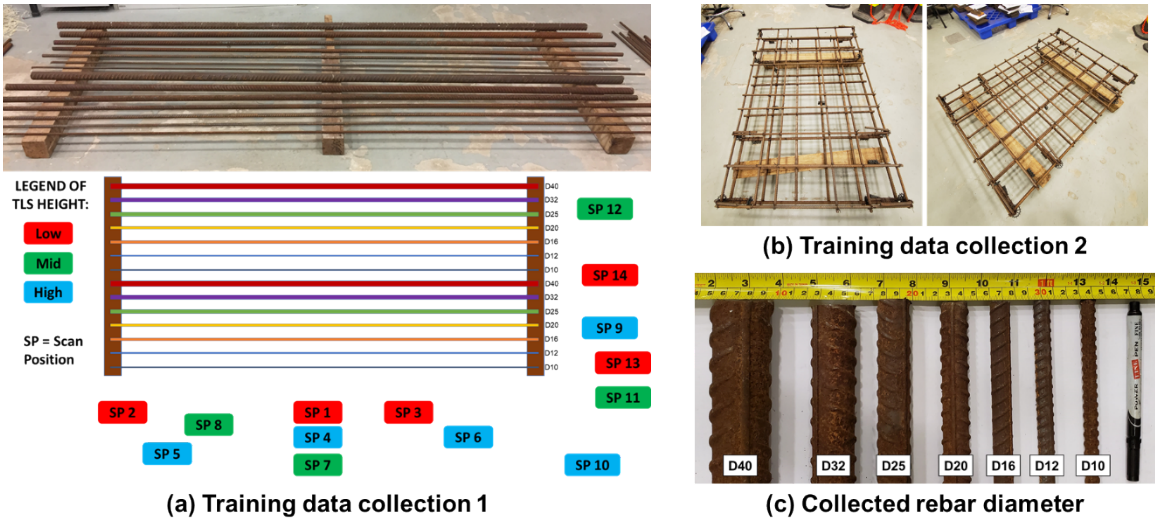

Figure 5. Data collection set-ups: (a) The first data collection configuration - the training data collection on 14 individual rebars with 7 different rebar diameters from 14 different positions with three different TLS heights; (b) The second data collection configuration – the training data collection on the rebar cage specimen having rebar diameters of 12 mm and 16 mm with two different scan angles of 45° and 90° with respect to the rebar cage; and (c) used rebars with seven different diameters (D10-D40)

Table 1. Feature selection results in three scan density groups

Density group 1		**Density group 2**		**Density group 3**	
Selected features	Merit Score	Selected features	Merit Score	Selected features	Merit Score
6, 7	**0.9680**	6, 7	**0.9724**	**6**	**0.9838**
6	0.9624	6	0.9704	6, 1	0.9837
6, 7, 1	0.9676	6, 7, 1	0.9723	6, 1, 7	0.9821
6, 7, 1, 5	0.9160	6, 7, 1, 4	0.9311	6, 1, 7, 4	0.9532
6, 7, 1, 5, 4	0.7985	6, 7, 1, 4, 5	0.8706	6, 1, 7, 4, 5	0.9111
6, 7, 1, 5, 4, 2	0.6776	6, 7, 1, 4, 5, 2	0.7897	6, 1, 7, 4, 5, 2	0.8726
6, 7, 1, 5, 4, 2, 3	0.6776	6, 7, 1, 4, 5, 2, 3	0.7897	6, 1, 7, 4, 5, 2, 3	0.8726

Table 2 shows the rebar diameter classification accuracy of the training data with five different machine learning algorithms. Note that the rebar diameter classification accuracy sets are divided into two groups: small diameters (D10-D20) and large diameters (D25-D40). Here, the number of 3 was selected as the optimal number of scan density group. The reasoning behind the selection is discussed in section 5.2. There are three primary observations. Firstly, in most cases, the classification accuracy increased as the density increases. Secondly, the average accuracy of the small rebar diameter cases (61.4%) was much lower than that (95.9%) of the large rebar diameter cases. This is attributed to the lack of scan points on the rebar surface in the cases of small diameters. Thirdly, the SVM provides the highest average classification accuracy of 78.7% and 96.0% for the total cases (D10-40) and the large diameters (D25-D40), respectively. Therefore, the SVM was chosen as the optimal classifier for the implementation of new rebar scan data prediction.

Table 2. Comparison of the training accuracy among the five machine learning classifiers in the three density groups

Classifier	**Classification accuracy**								
	Small diameter (D10-D20)				Large diameter (D25-D40)				**Total avg.**
	Density group			**Avg.**	Density group			**Avg.**	
	1	2	3		1	2	3		
	Selected features				Selected features				
	6, 7	6, 7	6		6, 7	6, 7	6		
Naïve Bayes (NB)	52.9%	44.4%	64.5%	53.9%	92.7%	95.8%	98.7%	95.7%	74.8%
Discriminant Analysis (DA)	52.3%	45.5%	64.0%	53.9%	92.0%	94.6%	99.2%	95.3%	74.6%
Classification Tree (CT)	59.0%	65.6%	61.5%	62.0%	90.6%	96.0%	96.5%	94.4%	78.2%
Nearest Neighbours (NN)	58.5%	65.0%	59.1%	60.9%	90.3%	96.2%	96.8%	94.4%	77.6%
Support Vector Machine (SVM)	59.3%	56.5%	68.6%	61.5%	93.1%	96.9%	98.0%	**96.0%**	**78.7%**

In order to investigate the details of the classification result, the confusion matrix of the SVM in three different density groups as shows in Figure 6. Three key pieces of information obtained through this are: (1) A large number of the false predictions occur in the small rebar diameter group (D10-D20) due to the lack of scan points on the rebar surface for small diameters; (2) the false prediction occurs in the adjacent to the diagonal cells, indicating fasely predicted rebar diameters have similar diameters. For example, most falsely predicted rebar diameters for the cases of rebars with D12 are D10 or D16; and (3) classification accuracy is largely affected by scan density. In the confusion matrices, as the density increases, the number of non-diagonal cases is reduced. The causes are described in detail in Section 5.1, which compares the performance of the traditional machine model and the DBM.

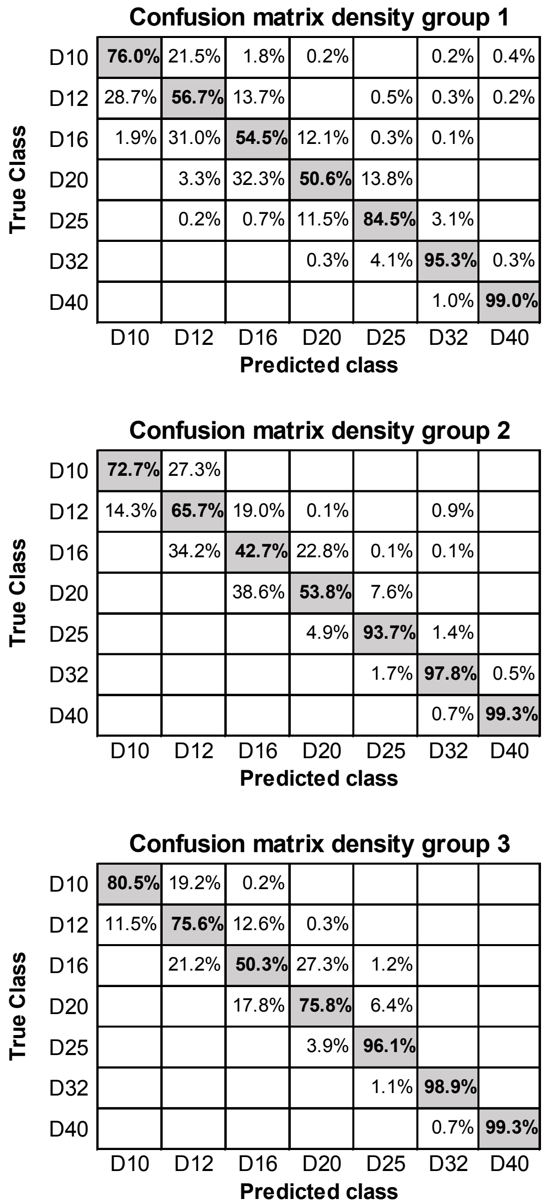

Figure 6. Confusion matrix result in the three different density groups

4.3 Prediction result of Rebar grid with varies diameter

In order to validate the proposed rebar diameter classification method, prediction tests were conducted on a specimen with 3 m length and 1 m width. Figure 7 shows the experimental set up for the prediction test. The specimen is composed of 10 longitudinal rebars and 18 transversal rebars. The longitudinal rebars also have 7 different diameters from D10 to D40 and the transversal rebars have 2 different diameters of D12 and D16.

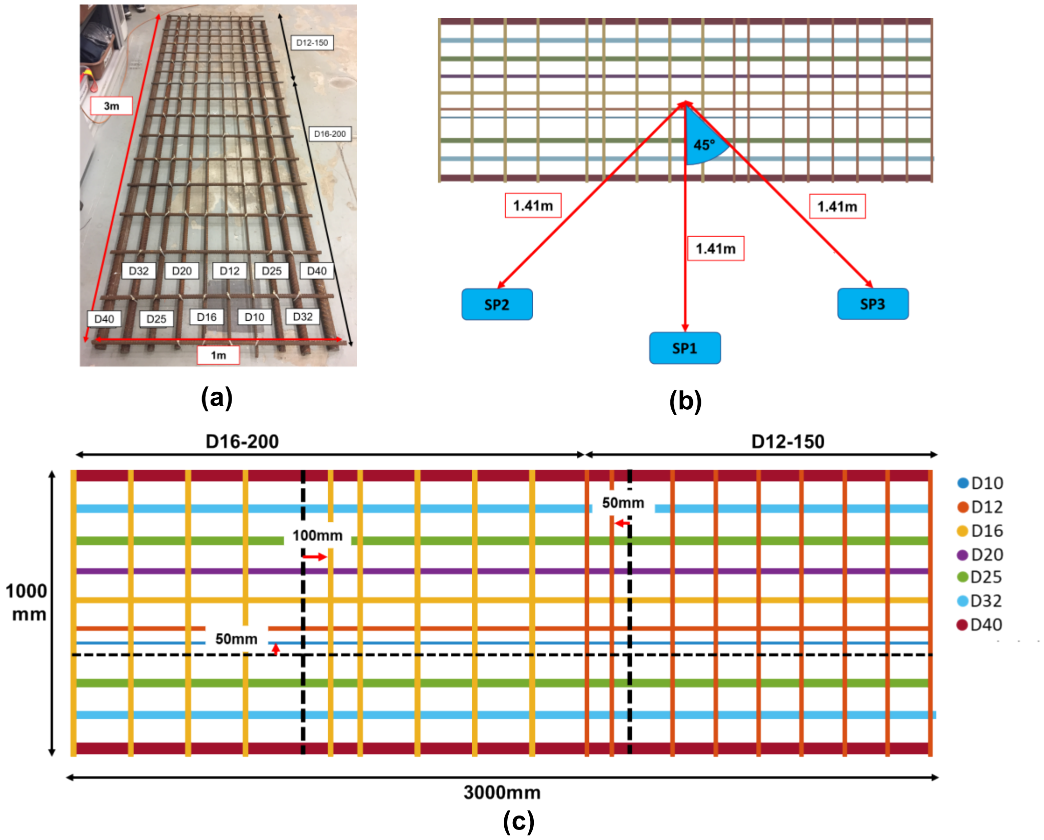

Figure 7. Experimental set up for the prediction test: (a) Specimen with 3 m long and 1 m width; (b) the TLS at three positions (SP1-SP3) for the prediction test acquisition; and (c) artificial dimensional changes on the 3 individual rebars (1 longitudinal and 2 transversal) were introduced for rebar diameter prediction and rebar spacing estimation.

Figure 8 shows the prediction results of the longitudinal and transversal rebars. Note that the results were generated from the scan location of SP2 with the high scan density. Also note that the instances with different colors indicate the predicted diameters. Here, there are two types of prediction results, which are 1) instance-level prediction and 2) rebar-level prediction in the figure. The instance-level prediction gives the prediction value for each instance. Meanwhile, the rebar-level prediction provides one prediction value for each individual rebar based on the assumption that the diameter of each individual rebar is predicted as the diameter having the largest number of predicted diameter at the instance-level. For example, taking the example of the D40 bottom longitudinal rebar in Figure 8, the diameter with the largest number of predicted diameter at the instance-level is D40, although there are 5 instances of false predictions, as seen in D32, in blue. Note that the rebar-level prediction results are presented in the middle of each individual rebar in Figure 8.

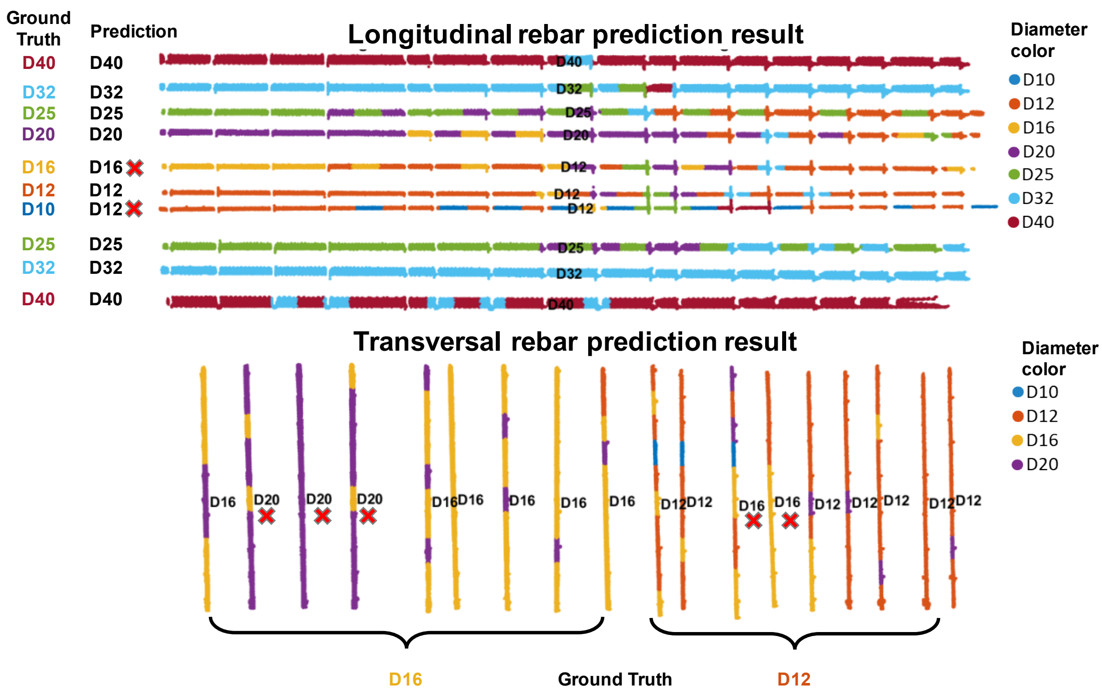

Figure 8. Rebar diameter prediction results. The upper and bottom parts indicate the longitudinal and transversal rebars prediction results respectively. (The results come from the scan location of SP2 with the high scan density)

Table 3 shows the prediction results obtained from the 3 scan positions of SP1 to SP3. There are three primary observations. First, a higher prediction accuracy is obtained in high scan density cases on both the instance-level and the rebar-level from the three scan positions of SP1 to SP3. As can be seen in the bottom part of the table, the average accuracy improvement is 17% in the high scan density among SP1 to SP3. Further, as expected, higher prediction performance was obtained at the rebar-level prediction compared to the instance-level prediction in both diameter groups. For instance, the average prediction accuracy at the instance-level is 71.1% whereas that at the rebar-level is 97.2% for the large diameter group. Finally, similar to the trend observed in the training data classification results, the small rebar diameters were hard to accurately classify (56.0%) at the rebar-level compared to of the large rebar diameters (97.2%).

Table 3. The prediction test results with three scan positions (SP1 to SP3)

	Prediction accuracy								
	Scan position						Avg. at Inst. level	Num. of correct rebar out of total rebar	Avg. at rebar level
	SP1		SP2		SP3				
	Scan density								
	Low	High	Low	High	Low	High			
D10	0.0%	33.3%	38.1%	30.0%	30.0%	46.6%	**35.1%**	4/6	**56.0%**
D12 (L&T)	58.5%	53.6%	60.0%	61.4%	36.3%	41.3%		47/60	
D16 (L&T)	30.2%	50.0%	22.5%	52.5%	12.4%	37.5%		20/60	
D20	0.0%	29.6%	13.3%	57.5%	3.3%	45.0%		3/6	
D25	50.0%	43.3%	70.0%	61.6%	63.3%	50.0%	**71.1%**	12/12	**97.2%**
D32	76.6%	66.6%	85.4%	78.0%	61.6%	75.0%		12/12	
D40	90.0%	76.6%	88.3%	90.0%	73.3%	80.8%		11/12	
Avg. at Inst. level	43.6%	50.4%	53.9%	61.6%	40.0%	53.7%			
Avg. at rebar level	**64%**	**82%**	**57%**	**75%**	**46%**	**61%**			

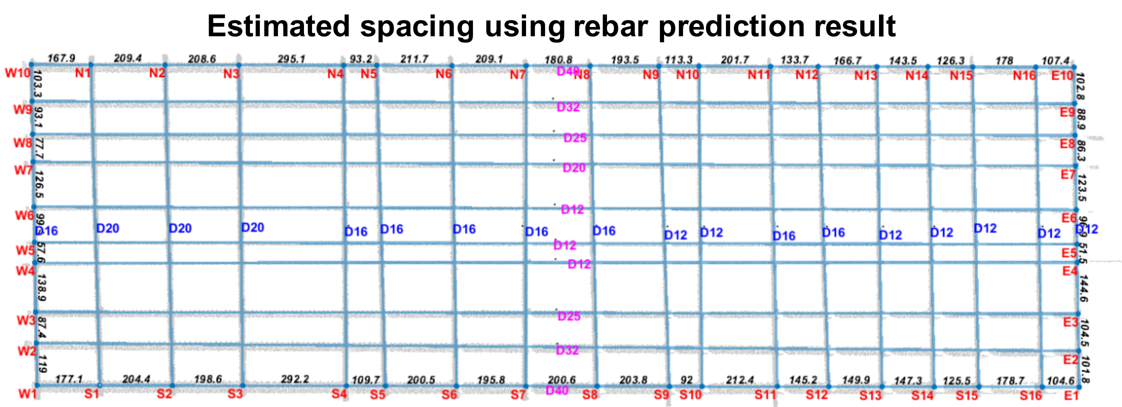

Figure 9. Estimation result using predicted rebar under the left scan position with a high scan density (The results come from the scan location of SP2 with the high scan density)

4.1 Rebar spacing estimation result

Figure 9 shows the rebar spacing estimation results based on the results of rebar prediction in Section 4.3. Note that four directions of the rebars are denoted as 'N', 'S', 'E', and 'W', and the value presented between two points at the outer boundaries refers to the estimated rebar spacing. Also, note that the diameter presented in the middle of rebars represents the predicted rebar diameters in Section 4.3. The rebar spacing estimation was made at 52 locations (17 locations in each 'N' and 'S' directions, and 9 locations in each 'E' and 'W' directions). Table 3 shows the rebar spacing estimation discrepancies for the longitudinal and transversal rebars. Note that the rebar spacing estimation discrepancy is defined as the difference between the estimated rebar spacing using the proposed method and a manual measurement using a measurement tape. The results show the rebar spacing achieves estimation accuracy of 2.2 mm with the proposed automatic rebar diameter prediction - a comparable accuracy to that (2.1 mm) of on actual rebar diameter, indicating that the proposed method has potential for automated rebar diameter prediction as well as rebar spacing estimation in an accurate manner.

Table 4. Comparison of rebar spacing estimation results between the proposed automatic method and the manual method. Here, the discrepancy between the estimated rebar spacing and the manually measured (ground-truth) rebar spacing

		Discrepancy between the estimated rebar spacing and the measured rebar spacing (mm)						
		Scan position						Total avg.
		SP1		SP2		SP3		
		Scan density						
		Low	High	Low	High	Low	High	
Proposed method (automatic rebar diameter estimation)	N. Trans.	2.3	2.4	3.2	2.7	2.9	1.8	2.5
	S. Trans.	1.6	1.7	2.4	1.5	2.0	1.8	1.8
	W. Long.	2.5	1.7	3.1	1.3	3.2	2.9	2.4
	E. Long.	3.3	1.7	2.8	2.2	3.0	1.7	2.4
	Average	**2.3**	**1.9**	**2.8**	**2.0**	**2.6**	**2.0**	**2.2**
Manual method (rebar diameter are given manually)	N. Trans.	2.3	2.0	3.0	2.8	2.0	2.0	2.3
	S. Trans.	1.5	1.8	2.0	1.7	1.9	1.7	1.7
	W. Long.	2.3	2.2	2.8	1.7	2.6	2.5	2.3
	E. Long.	2.4	1.6	3.3	2.4	2.6	1.5	2.3
	Average	**2.1**	**1.9**	**2.7**	**2.2**	**2.2**	**1.9**	**2.1**

5. Discussion

To further identify the effectiveness of the proposed method, further investigation into the three aspects, which are 1) accuracy comparison between the traditional machine learning approach and the proposed DBM approach; 2) optimal number of density group; and 3) recommendation of scan density for performing rebar diameter classification, was conducted.

5.1 Accuracy comparison with traditional machine learning approach

A comparison test was performed using the SVM which was as selected as the optimal machine learning classifier. Note that the traditional machine learning approach uses training data sets to generate one model, while the DBM generates a number of models according to the number of scan density groups. Table 4 shows the comparison results. Overall, the DBM approach offers better accuracy compared to the traditional model in both the cases of small diameters (D10-D20) and large diameters (D25-D40). Particularly, a significant improvement of 17.9% and 27.8% was observed in the large diameter group (D25 to D40) at the instance level and the rebar level respectively.

Table 4. Accuracy comparison between the traditional method and the proposed DBM method

		Classification Accuracy					
		Small diameter (D10 to D20)			**Large diameter** (D25 to D40)		
		Traditional	DBM	Improvement	Traditional	BDM	Improvement
Training	Instance level	57.6%	61.7%	**4.1%**	90.5%	95.8%	**5.3%**
Prediction	Instance level	33.9%	35.1%	**1.2%**	53.2%	71.1%	**17.9%**
	Rebar level	57.6%	56.1%	**-1.5%**	69.4%	97.2%	**27.8%**

The reasons for the improvement can be found in Figure 10, which illustrates the effect of the density-based modeling on the feature extraction results. Figure 10(a) shows the 13,987 rebar instances comprising the training data plotted in the feature space. Note that the chosen features in the plot are GGS (Feature #6) and the 2nd principal axis (Feature #7), which are the primary selected features by the CFS. In traditional modeling, the instances' feature values largely overlap with the adjacent rebar diameters. However, a clear separation between the different rebar diameters can be observed in the DBM, particularly in the highest-density (Density group 3). For example, in the highest scan density group of the large rebar diameter group, the separation level between the rebar diameter D25, D32 and D40 in the feature space can be seen in Figure 10. This indicates that high scan density needs to be assured for performing high accurate rebar diameter prediction. Meanwhile, Figure 11 illustrates the reason why prediction accuracy in the small rebar diameter group is relatively low compare to the large rebar diameter group. Taking the example of D20, the variation level of feature value among the 6 fitting features is relatively large compared to that of the large diameter cases, which results in the large overlap between adjacent small rebar diameters and reduce prediction accuracy. On the other hand, the variation level of feature values among the 6 features is much lower in the large rebar diameter group and the feature values are closer to the actual rebar diameters, resulting in a clear separation in the feature space and achieving a high prediction accuracy.

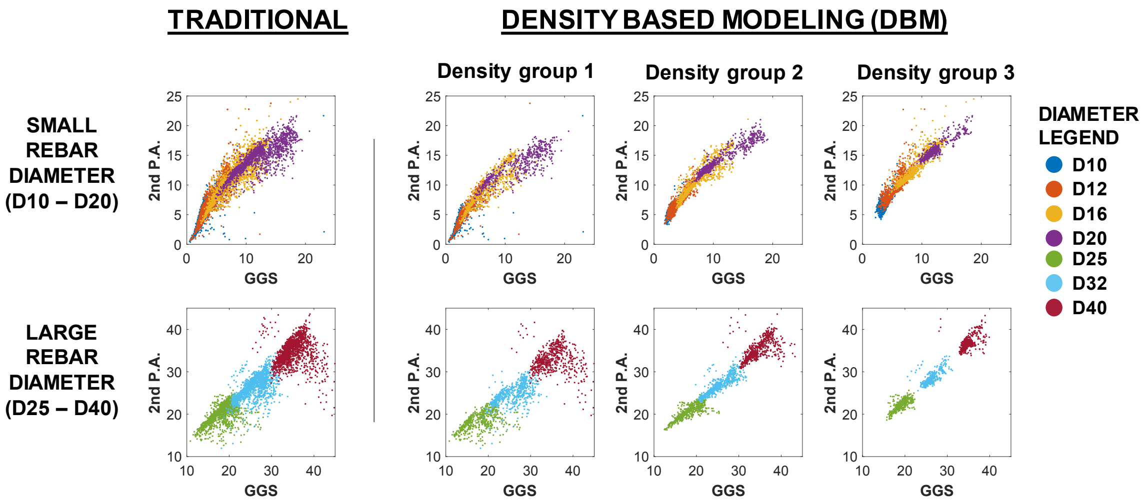

Figure 10. Performance comparison between the traditional method and the proposed DBM method on the feature extraction results. Compared to the traditional method, a clear separation among the different rebar diameters can be observed using the DBM method, particularly in the highest-density (Density group 3) in the large rebar diameter group.

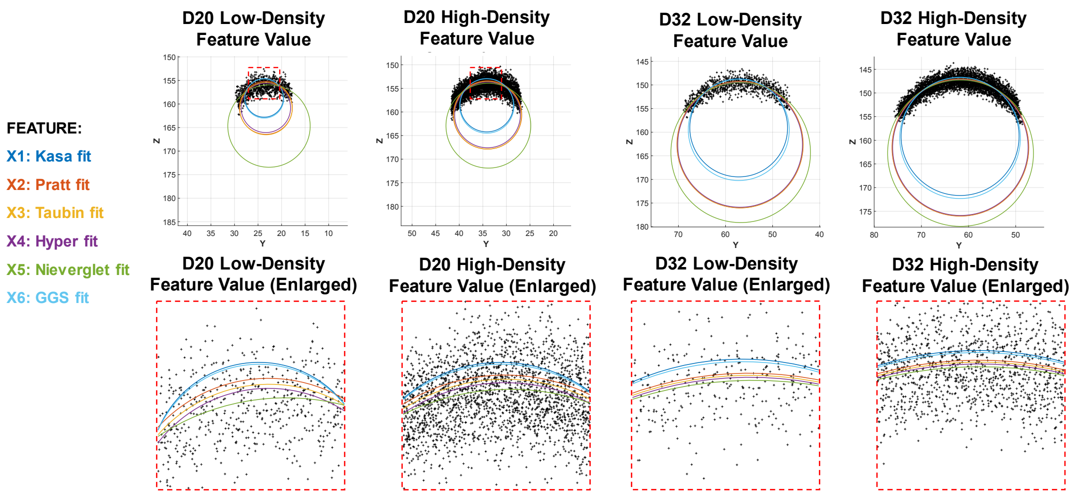

Figure 11. Density effect on the feature values in both the small and large rebar diameter groups. Compared to the large variation level of feature values among the 6 features in the small rebar diameter group (e.g. D20), the variation level of feature values in the large rebar

diameter group (e.g. D32) is much lower and the feature values are close to the actual rebar diameters.

5.2 Selection of the number of density group

The selection of the optimal number of density group is another important issue in the proposed DBM method. Figure 12 shows the effect of number of density group on instance-level classification accuracy in training data using the SVM. It can be seen that although a 4.1 % increase is obtained when increasing the number from 2 to 3, there is no significant increase in accuracy when the number of density group is equal to or larger than 3. Based on the results, the selected number of density group was 3 in this study.

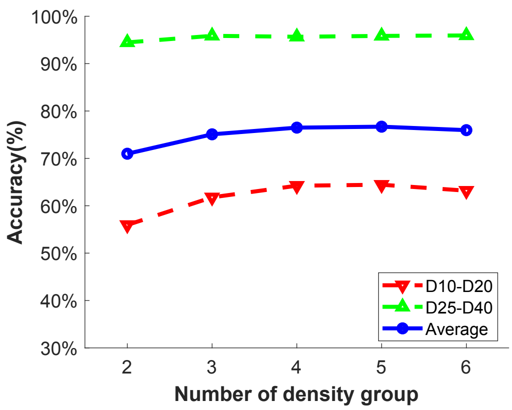

Figure 12. The effect of number of density group on instance-level classification accuracy in training data using the SVM

5.3 Recommended scan density for rebar diameter prediction

Throughout the investigation in Section 4 and Section 5.1, it was found that the scan density is one of the most important factors for rebar diameter classification. This section investigates the recommended scan density level for accurate rebar diameter prediction. Figure 13 shows the effect of scan density on rebar diameter prediction accuracy. Figures 13(a) and (b) show the percentages of correctly predicted instances out of the total instances with respect to the scan density in the small and large rebar diameter groups, respectively. Note that each bin has a width of 10 pts/mm^2. In the small rebar diameter group, it was found that in order to achieve an accuracy over 75%, a scan density value of at least 80 pts/mm^2 is necessary. In contrast, for the large rebar diameter group, all the density bins from 0-150 pts/mm^2 had a prediction accuracy over 80%, even in the case of the lowest scan density ranging between 0-10 pts/ mm^2, which had an accuracy of 84.8%. This is because the cross section shape of each rebar section instance formed in the large rebar diameter groups is always larger and clearer than that of small rebar diameter groups, resulting in a robust and consistent prediction accuracy. Figure 13(c) shows an exemplary case that represents the impact of scan density on prediction accuracy in both the small and large diameter rebars. The TLS was positioned on the left and as a result, the area of small rebar diameter groups in the middle of the horizontal rebar cage has many cases of false predictions. This is because the area has small rebar diameters, resulting in a low scan density ranging from 14.3 pts/ mm^2 to 82.4 pts/ mm^2. Based on the findings, a scan density value of at least 80 pts/mm^2 on small rebar diameters from D10-D20 is necessary for successful rebar diameter prediction. Note that for each instance, the length of the instance is 100 mm, so much larger scan points are contained in the unit area (mm^2). In summary, it can be concluded that it is essential to conduct scan planning before actual scanning in order to achieve a high rebar diameter prediction, as well as rebar spacing estimation.

6. Conclusion and future work

This study presents a new TLS-based approach that automates the classification of rebar diameters using machine learning in order to enable accurate rebar spacing inspection. In this study, a new methodology named Density based Modeling (DBM) is proposed to improve classification accuracy. Experimental tests on laboratory specimens with rebars of seven different diameters (D10-D40) were conducted and the results show that the prediction accuracy for large rebar diameter group (D25-D40) was up to 97.2%. However, it was found that its performance in predicting small rebar diameter group (D10-D20) is much lower - around 56.0%. The lessons learned from the results are (1) the proposed DBM method for rebar diameter prediction is superior to the traditional machine learning approach; and (2) scan density is one of the most important factors affecting the prediction results, especially in the small rebar diameter group (D10-D20); and (3) based on the findings of the study, a scan density value of at least 80 pts/mm^2 on the cross section plane to rebar instance with 100 mm length is necessary in small rebar diameters from D10-D20 for successful rebar diameter prediction. In addition, in practice, it is essential to conduct scan planning before actual scan in order to achieve a high rebar diameter prediction as well as rebar spacing estimation.

However, there are some limitations of the proposed technique, which are avenues for future research. First, accurate prediction of small-diameter rebars is still a challenging task using the proposed method. This issue may be further investigated by finding more robust features for accurate machine learning performance. In addition, a scan planning method that computes the scan density and determines an optimal scan position may address the issue of low accuracy in small-diameter rebars.

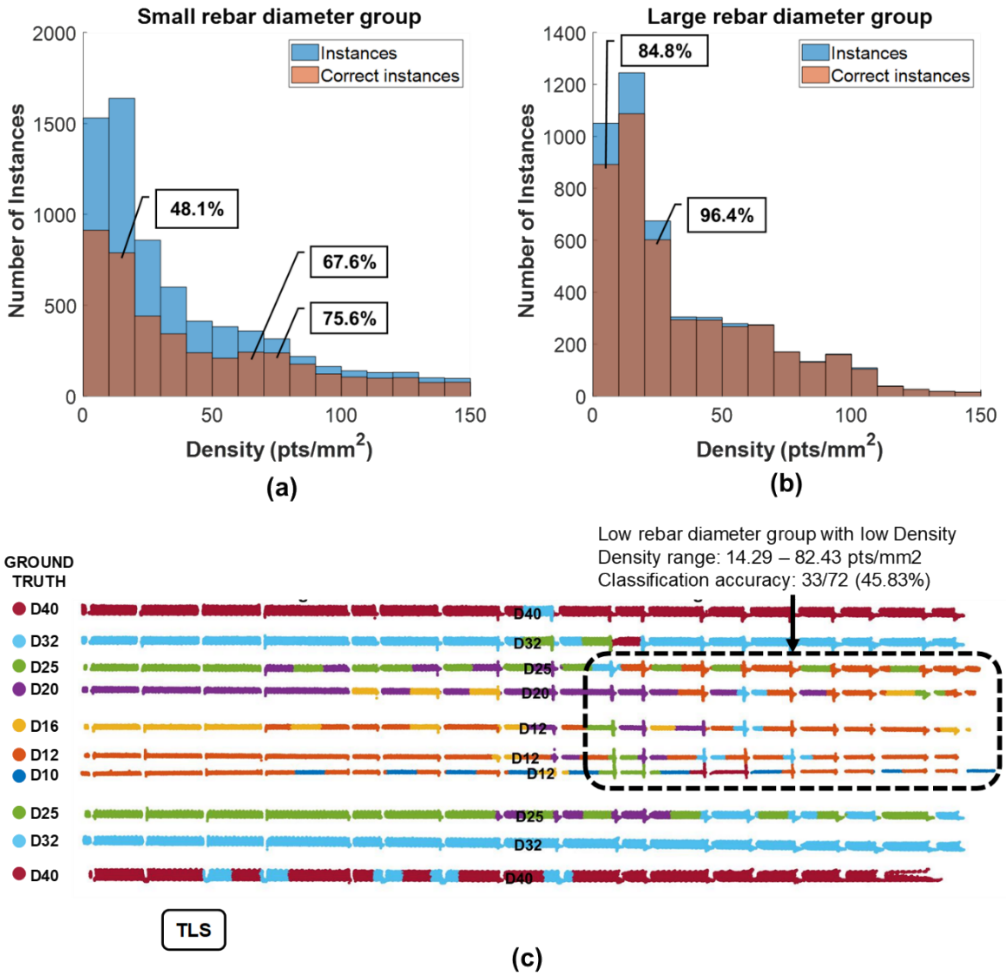

Figure 13. Effect of scan density on the rebar diameter prediction accuracy at the instance level: (a) In the small rebar diameter group; (b) in the large rebar diameter group; and (c) on the left scan (SP1) with the scan high density

Acknowledgement

The first author would like to acknowledge that this research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIP) (No. 2018R1A5A1025137) and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1A6A3A03010355).

References

1. Kim, M.-K., et al., *A framework for dimensional and surface quality assessment of precast concrete elements using BIM and 3D laser scanning.* Automation in Construction, 2015. **49**: p. 225-238.

2. Kim, M.-K., H. Sohn, and C.-C. Chang, *Automated dimensional quality assessment of precast concrete panels using terrestrial laser scanning.* Automation in construction, 2014. **45**: p. 163-177.

3. Kim, M.-K., H. Sohn, and C.-C. Chang, *Localization and quantification of concrete spalling defects using terrestrial laser scanning.* Journal of Computing in Civil Engineering, 2015. **29**(6): p. 04014086.

4. Kim, M.-K., et al., *Automated dimensional quality assurance of full-scale precast concrete elements using laser scanning and BIM.* Automation in Construction, 2016. **72**: p. 102-114.

5. Wang, Q., et al., *Automated quality assessment of precast concrete elements with geometry irregularities using terrestrial laser scanning.* Automation in construction, 2016. **68**: p. 170-182.

6. Yoon, S., Q. Wang, and H. Sohn, *Optimal placement of precast bridge deck slabs with respect to precast girders using 3D laser scanning.* Automation in construction, 2018. **86**: p. 81-98.

7. Akula, M., et al., *Real-time drill monitoring and control using building information models augmented with 3D imaging data.* Automation in Construction, 2013. **36**: p. 1-15.

8. Han, K., et al. *Vision-based field inspection of concrete reinforcing bars*. in *13th International Conference on Construction Applications of Virtual Reality, London, UK. Oct. 30-31*. 2013.

9. Nishio, K., et al. *A method of core wire extraction from point cloud data of rebar*. in *The 25th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*. 2017. University of West Bohemia, Plzen, Czech Republic: Václav Skala - UNION Agency.

10. Fischler, M.A. and R.C. Bolles, *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography.* Communications of the ACM 1981. **24**(6): p. 381-395.

11. Kim, M.-K., J.P.P. Thedja, and Q. Wang, *Automated dimensional quality assessment for formwork and rebar of reinforced concrete components using 3D point cloud data.* Automation in Construction, 2020. **112**: p. 103077.

12. Chernov, N., *Circular and linear regression: Fitting circles and lines by least squares*. 2010: CRC Press.

13. Chernov, N. and C. Lesort, *Least squares fitting of circles.* Journal of Mathematical Imaging and Vision, 2005. **23**(3): p. 239-252.

14. Kåsa, I., *A circle fitting procedure and its error analysis.* IEEE Transactions on instrumentation and measurement, 1976(1): p. 8-14.

15. Al-Sharadqah, A. and N. Chernov, *Error analysis for circle fitting algorithms.* Electronic Journal of Statistics, 2009. **3**: p. 886-911.

16. Pratt, V. *Direct least-squares fitting of algebraic surfaces*. in *ACM SIGGRAPH computer graphics*. 1987. ACM.

17. Taubin, G., *Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation.* IEEE Transactions on Pattern Analysis & Machine Intelligence, 1991(11): p. 1115-1138.

18. Gander, W., G.H. Golub, and R. Strebel, *Least-squares fitting of circles and ellipses.* BIT Numerical Mathematics, 1994. **34**(4): p. 558-578.

19. Nievergelt, Y., *Hyperspheres and hyperplanes fitted seamlessly by algebraic constrained total least-squares.* Linear Algebra and its Applications, 2001. **331**(1-3): p. 43-59.

20. Wang, H. and D. Suter, *Using symmetry in robust model fitting.* Pattern Recognition Letters, 2003. **24**(16): p. 2953-2966.

21. Nurunnabi, A., Y. Sadahiro, and D.F. Laefer, *Robust statistical approaches for circle fitting in laser scanning three-dimensional point cloud data.* Pattern Recognition, 2018. **81**: p. 417-431.

22. Nurunnabi, A., D. Belton, and G. West, *Robust statistical approaches for local planar surface fitting in 3D laser scanning data.* ISPRS journal of photogrammetry and Remote Sensing, 2014. **96**: p. 106-122.

23. Nurunnabi, A., G. West, and D. Belton, *Outlier detection and robust normal-curvature estimation in mobile laser scanning 3D point cloud data.* Pattern Recognition, 2015. **48**(4): p. 1404-1419.

24. Nahangi, M., et al., *Pipe radius estimation using Kinect range cameras.* 2019. **99**: p. 197-205.

25. Díaz-Vilariño, L., et al., *Automatic detection and segmentation of columns in as-built buildings from point clouds.* 2015. **7**(11): p. 15651-15667.

26. Ballard, D.H., *Generalizing the Hough transform to detect arbitrary shapes.* Pattern Recognition, 1981. **12**(2): p. 111-122.

27. Bueno, M., et al., *Quantitative Evaluation of CHT and GHT for Column Detection under Different Conditions of Data Quality.* 2017. **31**(5).

28. Rousseeuw, P.J., *Least median of squares regression.* Journal of the American statistical association, 1984. **79**(388): p. 871-880.

29. Nishio, K., et al., *A method of core wire extraction from point cloud data of rebar.* 2017.

30. Wang, Q., J.C. Cheng, and H. Sohn, *Automated estimation of reinforced precast concrete rebar positions using colored laser scan data.* Computer-Aided Civil and Infrastructure Engineering, 2017. **32**(9): p. 787-802.

31. Valero, E., et al., *Automated defect detection and classification in ashlar masonry walls using machine learning.* Automation in Construction, 2019. **106**: p. 102846.

32. Bassier, M., B. Van Genechten, and M. Vergauwen, *Classification of sensor independent point cloud data of building objects using random forests.* Journal of Building Engineering, 2019. **21**: p. 468-477.

33. Wang, Q., J.C.P. Cheng, and H. Sohn, *Automated estimation of reinforced precast concrete rebar positions using colored laser scan data.* Computer-Aided Civil and Infrastructure Engineering, 2017. **32**(9): p. 787-802.

34. Ester, M., et al. *A density-based algorithm for discovering clusters in large spatial databases with noise*. in *Kdd*. 1996.

35. Tang, P., D. Huber, and B. Akinci. *A comparative analysis of depth-discontinuity and mixed-pixel detection algorithms*. in *Sixth International Conference on 3-D Digital Imaging and Modeling (3DIM 2007)*. 2007. IEEE.

36. Fardin, N., Q. Feng, and O. Stephansson, *Application of a new in situ 3D laser scanner to study the scale effect on the rock joint surface roughness.* International Journal of Rock Mechanics and Mining Sciences, 2004. **2**(41): p. 329-335.

37. Franceschi, M., et al., *Discrimination between marls and limestones using intensity data from*

terrestrial laser scanner. ISPRS journal of photogrammetry and remote sensing, 2009. **64**(6): p. 522-528.

38. Pollyea, R.M. and J.P. Fairley, *Estimating surface roughness of terrestrial laser scan data using orthogonal distance regression.* Geology, 2011. **39**(7): p. 623-626.

39. Wold, S., K. Esbensen, and P. Geladi, *Principal component analysis.* Chemometrics and intelligent laboratory systems, 1987. **2**(1-3): p. 37-52.

40. Yu, L. and H. Liu. *Feature selection for high-dimensional data: A fast correlation-based filter solution*. in *Proceedings of the 20th international conference on machine learning (ICML-03)*. 2003.

41. Hall, M.A., *Correlation-based feature selection for machine learning.* 1999.

42. Wang, F., et al. *Feature selection using feature ranking, correlation analysis and chaotic binary particle swarm optimization*. in *2014 IEEE 5th International Conference on Software Engineering and Service Science*. 2014. IEEE.

43. Pearson, K., *VII. Note on regression and inheritance in the case of two parents.* proceedings of the royal society of London, 1895. **58**(347-352): p. 240-242.

44. Holzinger, A. and I. Jurisica, *Knowledge discovery and data mining in biomedical informatics: The future is in integrative, interactive machine learning solutions*, in *Interactive knowledge discovery and data mining in biomedical informatics*. 2014, Springer. p. 1-18.

45. Wang, Q., J.C.P. Cheng, and H. Sohn, *Automated Estimation of Reinforced Precast Concrete Rebar Positions Using Colored Laser Scan Data.* Computer-Aided Civil and Infrastructure Engineering, 2017. **32**(9): p. 787-802.

46. Bickel, P.J. and E. Levina, *Some theory for Fisher's linear discriminant function,naive Bayes', and some alternatives when there are many more variables than observations.* Bernoulli, 2004. **10**(6): p. 989-1010.

47. Hastie, T., et al., *The elements of statistical learning: data mining, inference and prediction*. 2009: Springer, New York, NY.

48. Fisher, R.A., *The use of multiple measurements in taxonomic problems.* Annals of eugenics, 1936. **7**(2): p. 179-188.

49. McLachlan, G., *Discriminant analysis and statistical pattern recognition*. 1992: John Wiley & Sons.

50. Breiman, L., et al., *Classification And Regression Trees*. 2017: Routledge.

51. Cover, T.M. and P. Hart, *Nearest neighbor pattern classification.* IEEE transactions on information theory, 1967. **13**(1): p. 21-27.

52. Dasarathy, B.V., *Nearest neighbor (NN) norms: NN pattern classification techniques*. 1991, Los Alamitos, California: IEEE Computer Society Press. 447.

53. Cortes, C. and V. Vapnik, *Support-vector networks.* Machine learning, 1995. **20**(3): p. 273-297.

54. Kohavi, R. *A study of cross-validation and bootstrap for accuracy estimation and model selection*. in *Ijcai*. 1995. Montreal, Canada.