Probabilistic model for rebar-concrete bond failure mode prediction considering corrosion

Ahmad Soraghi, Graduate Student,[1] Qindan Huang, Associate Professor,[2] and Derek A.J. Hauff, Graduate Student[3]

[1]University of Akron, Department of Civil and Environmental Engineering, The University of Akron, 44325-3905; e-mail: as481@zips.uakron.edu
[2]University of Akron, Department of Civil and Environmental Engineering, The University of Akron, 44325-3905; e-mail: qhuang@uakron.edu
[3]University of Akron, Department of Civil and Environmental Engineering, The University of Akron, 44325-3905; e-mail: adh@zips.uakron.edu

ABSTRACT

Adequate bonding between rebar and concrete is critical to ensuring the reliable performance of RC structures. It is found empirically that bond behavior is affected by many factors, including concrete cover, transverse reinforcement, rebar geometry, concrete properties, and and etc. While many past studies have focused on the prediction of bond strength, how those factors influence the bond failure mode is not well investigated.

The goal of this research is to develop a bond failure mode prediction model considering corrosion. The model development is based on bond testing results of 44 beam-end specimens with various rebar size, corrosion levels, covers, and stirrup confinement. This study adopts logistic and lasso logistic regression, where the failure mode is the categorical dependent variable and the aforementioned factors that could influence the bond behavior are the independent variables. The developed model can be can be further used for corroded RC structure performance evaluation.

Key words:

Bond, reinforced concrete, failure mode, logistic regression, lasso logistic regression

INTRODUCTION

Reinforced concrete (RC) has been widely used for civil infrastructures such as bridges, buildings, dams, and etc. The performance of the bonding between concrete and rebar (i.e., rebar-concrete interaction) is a critical factor to determine the performance of RC structure, as the bonding is to ensure the force transformation between concrete and rebar. The bond behavior directly affects the load carrying capacity and failure mode of the structure. Previous studies have shown that this bond is influenced by many factors such as concrete cover, concrete cover bar size ratio, transverse reinforcement, concrete properties, rebar corrosion, loading type, and etc.

Most of the past research has studied the bond strength and how the aforementioned factors influence the ultimate bond load carrying capacity (e.g., Kivell, 2012; Almusallam et al., 1996). For example, Wang (2009) studied bond strength in unconfined concrete and found that the ratio of concrete cover to the diameter of main rebar contributes to the bond strength. Harajli et al.

(2004) recognized that the amount of transverse rebars influences both the bond strength and increases the ductility of bond failure in non-corroded structure. Torre-Casanova et al. (2013) and Castel (2016) both developed and incorporated the effects of stirrups confinement on rebar-concrete bond strength in a numerical model. In addition, studies have showed that the bond strength is a function of compressive strength or tensile strength of concrete (Sajedi & Huang 2015; Torre-Casanova et al., 2013; Castel et al., 2016). In addition, many studies have found corrosion of rebar has a significant impact on the bond strength prediction, and thus on the RC structure performance (e.g., Hussain et al., 1995; Kivell, 2012; Wang, 2009; Fu & Chun, 1997; Stanish, 1999; Sajedi & Huang 2015).

Recognizing the failure mode of the concrete and rebar bond (i.e., either pull-out or splitting failure) is also critical to determine the performance of RC structures. Prediction of failure mode is also needed. Pull-out failure happens due to shearing of concrete between ribs, when there is adequate concrete cover that prevents splitting and presence of transvers reinforcement that maintains the small cracks (ACI, 2012). Splitting failure occurs under circumstances of lack of cover or confinement to attain full pull-out strength, and it is due to radial forces caused by deformation bearing forces that spreads to the sides of the member, leading in loss of concrete cover and bonding (ACI, 2012).

However, the failure mode of the concrete and rebar bond in RC structures have not been well studied especially when corrosion is presented or under cyclic loading or the combination of the two. ACI (2012) uses transverse stirrups cross section and spacing to determine if the bond failure will be pull-out or splitting; while CEB-FIB (2010) determines pull-out failure when both concrete cover bar size ratio amount are larger than certain values. Zandi Hanjari et al. (2011) and Cucchiara et al. (2004) also studied the stirrup contribution to failure mode. Kivell (2012) showed that specimens with high corrosion levels (e.g. 12%) can change the mechanism of failure tending to see pull-out behavior accompanied by reduced stiffness at rupture point and that specimens subjected to cyclic loading showed a tendency of failing in pull-out.

The goal of this paper is to develop a prediction model for bond failure mode considering various factors including cover to bar size ratio, transverse stirrups, corrosion, and loading type. In particular, this paper adopts logistic and lasso logistic regression for the model development. These two methods are suitable for categorical responses and compared with other approaches, they also yield to an explicit formulation, allowing for practical application from an engineering perspective.

The outline of this paper is as follows: a review of logistic and lasso logistic regression methodology; brief explanation about the test data for the model development is summarized; model development and model selection are elaborated, and the performance of the two developed models are presented; and final statements and remarks are made in the Summary and Conclusions section.

LOGISTIC AND LASSO LOGISTIC REGRESSION

Supervised machine learning techniques have been widely used in engineering for response prediction, and it is generally categorized into regression and classification algorithms. While regression is suitable for numerical responses, classification method is appropriate for categorical responses (e.g., failure modes) (Mangalathu Sivasubramanian Pillai, 2017). In particular, logistic regression and lasso logistic regression are adopted in this study, as they provide an explicit formulation, which provide greater assistance from an engineering stand-point.

Logistic regression

Logistic regression is a supervised machine learning technique used for categorical responses. This method evaluates the relation between independent and dependent variables (i.e., categorical response) through a logistic function. When the response is binary (e.g., let $Y = 1$ refer to pull-out failure and $Y = 0$ refer to splitting failure), the logistic regression formulation to predict the probability of Y being 1, $p(\mathbf{x})$, is shown as:

$$p(\mathbf{x}) = \Pr(Y = 1 \mid \mathbf{X} = \mathbf{x}) = \frac{\exp\left(\beta_0 + \sum \beta_i x_i\right)}{1 + \exp\left(\beta_0 + \sum \beta_i x_i\right)} \quad (1)$$

where $\mathbf{X} = \{X_i\}$, $\mathbf{x} = \{x_i\}$, X_i = independent variables selected, β_0 and β_i are logistic regression coefficients that can be estimated by maximum likelihood approach (Mangalathu Sivasubramanian Pillai, 2017), through likelihood function as shown below:

$$l(\boldsymbol{\beta}) = \sum_{j=1}^{N} \left(y_j \boldsymbol{\beta}^T \tilde{\mathbf{x}}_j - \log\left[1 + \exp\left(\tilde{\mathbf{x}}_j \boldsymbol{\beta}\right)\right] \right) \quad (2)$$

where subscript j refers to the j^{th} observation data, $\tilde{\mathbf{x}}_j = \{1\ \mathbf{x}\}^{\mathrm{T}}$ and $\boldsymbol{\beta} = \{\beta_0\ \beta_1\ \beta_2\ \ldots\ \beta_p\}^{\mathrm{T}}$. As Y is binary variable, $\Pr(Y = 0|\mathbf{x}) = 1 - p(\mathbf{x})$. Note that deviance is proportional to $-\log[l(\boldsymbol{\beta})]$; thus when maximizing $l(\boldsymbol{\beta})$ to estimate, it is minimizing the deviance.

Lasso logistic regression

Lasso logistic regression uses the same prediction formulation as logistic regression (shown in Eq. (1)); the difference is that lasso logistic regression uses a different approach to assess the model parameters, $\boldsymbol{\beta}$. lasso logistic is preferred when the independent variables are correlated and the size of data set used for the model development is small (Tibshirani, 1996). In particular, lasso logistic imposes a constraint on the coefficients in the maximum likelihood evaluation, which can be shown as follows:

$$l(\boldsymbol{\beta}) = \sum_{j=1}^{N} \left(y_j \boldsymbol{\beta}^T \mathbf{X}_j - \ln\left[(1 + \exp\left(\mathbf{X}_j \boldsymbol{\beta}\right)\right] - \lambda \sum_{i=0}^{p} |\beta_i| \right) \quad (3)$$

where λ refers to penalty factor. Instead of minimizing the deviance, lasso logistic departs from optimality to stabilize a system (which is called sparse regularization to avoid over-fitting) by adding a cost of the sum of absolute values of the coefficients. Lasso regression is preferred mostly when a number of independent variables is large and the sample size is small; it reduces deviance of the fitted model without a substantial increase in the prediction bias.

Model accuracy

In this study, two different quantities are used to measure the model accuracy. One measure quantity is the sum of squared errors of prediction, *SSE*, which can be shown as below:

$$SSE = \sqrt{\sum \left(y_j - p(\mathbf{x}_j)\right)^2} \quad (4)$$

The other measure quantity is based on hit or miss method. With the prediction probability by Eq. (1), one could select a threshold, α, to determine the failure mode. For example, by setting α = 50%, then $p(\mathbf{x}) \geq 50\%$ indicates that the predicted failure will be pull-out failure; and $p(\mathbf{x}) < 50\%$ indicates the failure is predicted to be splitting failure. Thus, there are four possible outcomes (shown in Table 1) depending on if a failure mode is correctly predicted or not. As shown in Table 1, TS and TN are the correct detections. To measure the model accuracy, the probability of correct detection, P_{CD}, can be applied. In this study, it is calculated based on the number of TP (n_{TP}), the number of TN tests (n_{TN}), and the total number of tests, shown as following:

$$P_{CD} \approx \frac{n_{TS} + n_{TN}}{n_{TS} + n_{TN} + n_{FP}} \tag{5}$$

where n_{FP} refers to the number of FP tests.

Table 1. Four possible prediction outcomes

Failure mode	Predicted to be pull-out	Predicted to be splitting
Pull-out ($Y = 1$)	True positive (TP)	False positive (FP)
Splitting ($Y = 0$)	False positive (FP)	True negative (TN)

EXPERIMENTAL DATA

To study the bond behavior of intact and corroded rebar, a group of beam-end specimens were designed. Based on previous study (Sajedi & Huang, 2015), the specimens were designed based on four factors: concrete compressive strength (f'_c), the ratio of cover size to diameter of rebar (C/d), corrosion level (Q), and confinement provided by transverse reinforcement that can be described by index K_{tr} specified by (Orangun et al., 1977), as shown below

$$K_{tr} = \frac{A_{tr} \cdot f_{yt}}{600 s \cdot d_b} \tag{6}$$

Where A_{tr} is the area of transverse reinforcement in inches squared (in^2), f_{yt} is the yield strength of transverse reinforcement in pounds per square inch (psi), s is the spacing of transverse reinforcement in inches (in), and d_b is the bar diameter in inches. The specimens are grouped in three based on designated compressive strength levels: 30MPa, 40MPa, and 50MPa. At this stage, only the group with 40MPa and a total of 44 specimens has been cast and tested. Table 2 shows the design parameters of the 44 specimens in this group.

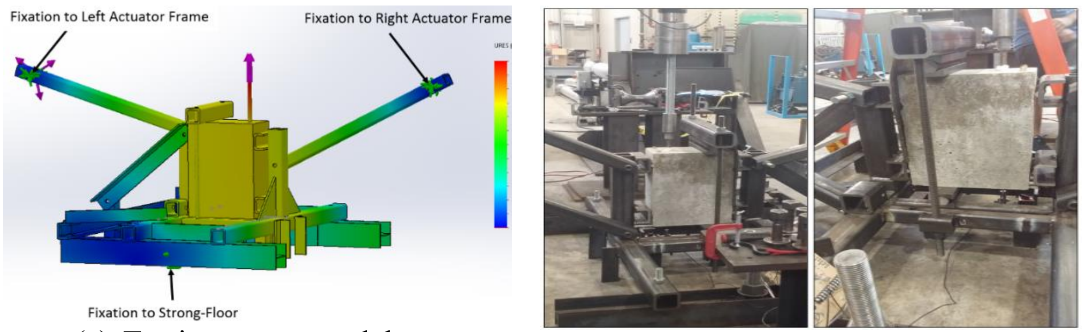
In summary, as shown in Table 2, there are 12 intact (0% corrosion) and 22 corroded specimens ranging from 5% to 20% as designed corrosion levels (corresponding to 3.2% to 15.6% actual corrosion levels measured after the load testing). Among the 44 specimens, 22 had transverse stirrups with K_{tr} ranging from 3.68 to 5.89, and the remaining 22 specimens were without transverse stirrups (i.e., K_{tr} = 0). The specimens were split into two groups, with each group being subjected to different loading behaviors. 18 specimens are tested under monotonic loading, while the rest were tested under cyclic loading. Moreover, as can be seen in Table 2, there were 10 specimens with the failure mode denoted as NA, indicating the failure mode is not recognizable or the test was not completed; thus, they have been removed from the collected data.

The lab's 55-kip actuator was mounted in a vertical position on a rigid frame secured to the base slab of the testing center. A special testing frame was designed and constructed to allow for the beam specimens to be subjected to monotonic or cyclic loading in a vertical position. This varied from the suggested horizontal setup shown in ASTM Standard A944-10. The testing frame allowed the beam-end specimen's rebar to connect to the actuator through a threaded rod welded to the rebar and attached to a special connection specifically design for this test, as shown in Figure 1.

Table 2. Experimental database for beam-end specimens

Specimen	*d* (mm)	*C* (mm)	*C*/*d*	*Q* (%) designed	*Q* (%) actual	*MC*	K_{tr}	Failure mode
1	15.875	50.8	3.2	0	0	0	0.00	0
2	15.875	63.5	4.0	10	4.9	0	0.00	1
3	15.875	76.2	4.8	20	7.6	0	0.00	0
4	15.875	50.8	3.2	0	0	0	5.89	1
5	15.875	63.5	4.0	10	5.3	0	5.89	1
6	15.875	76.2	4.8	20	9.9	0	5.89	1
7	15.875	50.8	3.2	20	0	1	0.00	1
8	15.875	25.4	1.6	10	10.3	1	0.00	1
9	15.875	63.5	4.0	5	11.0	1	0.00	1
10	15.875	38.1	2.4	15	10.1	1	0.00	1
11	15.875	76.2	4.8	0	12.0	1	0.00	1
12	15.875	50.8	3.2	20	0	1	5.89	1
13	15.875	25.4	1.6	10	7.9	1	5.89	N/A
14	15.875	63.5	4.0	5	4.3	1	5.89	1
15	15.875	38.1	2.4	15	8.2	1	5.89	1
16	15.875	76.2	4.8	0	11.3	1	5.89	1
17	19.05	38.1	2.0	0	0	0	0.00	0
18	19.05	25.4	1.3	10	3.6	0	0.00	0
19	19.05	50.8	2.7	20	15.6	0	0.00	N/A
20	19.05	38.1	2.0	0	0	0	4.91	0
21	19.05	25.4	1.3	10	3.2	0	4.91	0
22	19.05	50.8	2.7	20	7.1	0	4.91	N/A
23	19.05	38.1	2.0	5	0.0	1	0.00	0
24	19.05	63.5	3.3	0	8.5	1	0.00	N/A
25	19.05	25.4	1.3	10	7.6	1	0.00	1
26	19.05	76.2	4.0	20	9.9	1	0.00	1
27	19.05	50.8	2.7	15	13.4	1	0.00	1
28	19.05	38.1	2.0	5	0	1	4.91	N/A
29	19.05	63.5	3.3	0	8.6	1	4.91	0
30	19.05	25.4	1.3	10	6.9	1	4.91	1
31	19.05	76.2	4.0	20	7.7	1	4.91	1
32	19.05	50.8	2.7	15	11.0	1	4.91	1
33	25.4	63.5	2.5	0	0	0	0.00	0
34	25.4	50.8	2.0	10	4.3	0	0.00	0
35	25.4	38.1	1.5	20	10.2	0	0.00	0
36	25.4	63.5	2.5	0	0	0	3.68	0
37	25.4	50.8	2.0	10	7.7	0	3.68	0
38	25.4	38.1	1.5	20	11.9	0	3.68	0
39	25.4	63.5	2.5	0	0	1	3.68	0
40	25.4	50.8	2.0	10	5.2	1	0.00	N/A
41	25.4	38.1	1.5	20	13.1	1	0.00	N/A
42	25.4	63.5	2.5	0	0	1	3.68	N/A
43	25.4	50.8	2.0	10	5.7	1	3.68	N/A
44	25.4	38.1	1.5	20	13.7	1	3.68	N/A

During the testing, the applied force and slippage at the force end and the slippage at the free end were recorded. From the results of testing, the two distinct failure modes, pull-out and splitting failure, could be determined. As mentioned before, tests that produced unusual results or could not be completed were dismissed and label as NA in Table 2. The failure modes are shown in Table 2, where the failure mode “1” refers to pull-out failure and failure mode “0” refers to splitting failure.



(a) Testing set-up model

(b) Real test set-up

Figure 1. Testing frame set-up

MODEL DEVELOPEMENT

Independent variables selected for models

To develop prediction models for the failure mode using Eq. (1), a preliminary study is first conducted to select potential x_i. The variables that show the potential influences on the failure mode (Y) are: C/d, K_{tr}, Q, and MC, where MC is a dummy variable and defined as

$$MC = \begin{cases} 0 \; for\, specimens\, under\, monotonic\, loading \\ 1 \qquad for\, specimens\, under\, cyclic\, loading \end{cases} \tag{7}$$

In addition, the interactions among the four variables and higher orders of these variables are also examined using scatter plots. Figure 2 shows a scatter plot of an interaction term, $C/d{\cdot}K_{tr}$ vs. response, y, with a fitted logistic curve, as an example. As for results, the potential terms for x_i are listed in Table 3.

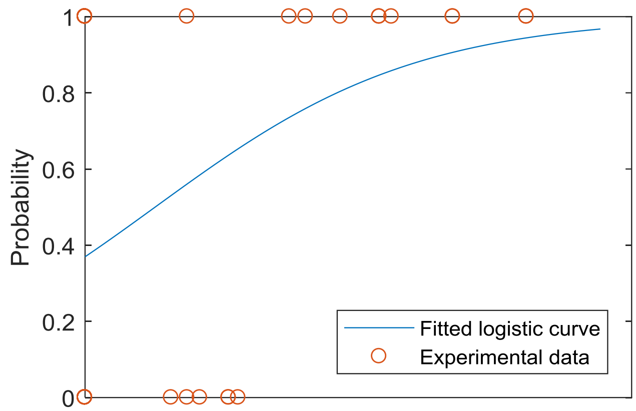


Figure 2. Example of a logistic curve for an interaction term (K_{tr}·*C/d*)

Table 3. Independent variables used in the model

Term types	x_i
Single variable	K_{tr}, *C/d, Q, MC*
Interaction with 2 variables	K_{tr}·*C/d, C/d·Q, Q·MC,* K_{tr}·*MC, C/d·MC,* K_{tr}·*Q*
Interaction with 3 variables	K_{tr}·*C/d·Q, C/d·Q·MC,* K_{tr}·*Q·MC,* K_{tr}·*C/d·MC*
Interaction with 4 variables	K_{tr}·*C/d·Q·MC*

Model selection

Inserting the 15 terms shown in Table 3 into Eq. (1), the model is a full model with a model size of 15. A model selection is then applied to the full model to eliminate the variables that do not have a statistically significant contribution to the model prediction. For logistic regression, an all possible subset approach (Lindsey & Sheather, 2010) is adopted. In all possible subset, all possible combination of x_i is first found for each reduced model size (varying from 1 to 14), resulting in a total of 32,717 models. The model parameters are then assessed using maximum likelihood approach. The models that have any model parameters with *p*-value larger than 10% or Variance Inflation Factors (*VIF*s) larger than 10 are considered to be invalid and they are not considered further in the model selection. Then for each model size, the model quality is measured by several statistical measurements: R-squared (*R-sq*), adjusted R-squared (*Adj-R-sq*), Akaike Information Criterion (*AIC*), and Bayesian Information Criterion (*BIC*). The model with the largest *Adj-R-sq* and lowest *AIC* and *BIC* are the most desirable model. For the same model size, the model with the largest *Adj-R-sq* also has the lowest *AIC* or *BIC*; however, when comparing the models with different model sizes, those statistical measurements can result in the different most desirable model.

Table 4 shows the top three best models for different model sizes, and the best model for each model size is denoted in bold. Note that when the model size is larger than 3, none of the models are valid due to large *p-values* and *VIF*s. When comparing the models of model size 1 or 2 with the models of model size 3, the model size 3 models have better accuracy in terms of *R-sq* and *Adj-R-sq*, even though their *AIC* and *BIC* are very close for model size 2 and model size 3 models. Table 4 also compares the *SSE* and P_{CD} of the models. As shown in Table 4, model size 3 models has much better accuracy in terms of *SSE*, while the difference in P_{CD} for different model size models is not significant.

As shown in Table 4, all the statistical measures point to the model with three terms (*MC*, $K_{tr}{\cdot}C/d$, $C/d{\cdot}Q$) to be the best model that has the largest *Adj-R-sq* and lowest *AIC* and *BIC*. The statistics of the model coefficients for this model are listed in Table 5.

Table 4. Statistics summary of top three logistic regression models for each model size

Model size	Independent variables	*R-sq* (%)	*Adj-R-sq* (%)	*AIC*	*BIC*	*SSE*	P_{CD}
1	$C/d{\cdot}MC$	27.4	25.1	40.0	43.1	6.0	0.76
	MC	28.4	26.2	40.4	43.5	6.0	0.50
	$\boldsymbol{C/d{\cdot}Q}$	30.5	28.4	39.1	42.2	5.8	0.73
2	$K_{tr}{\cdot}C/d$, $C/d{\cdot}Q$	40.8	37.1	35.8	40.4	4.95	0.73
	MC, $K_{tr}{\cdot}C/d$	44.9	40.3	36.2	40.7	4.69	0.85
	$\boldsymbol{MC, C/d{\cdot}Q}$	44.4	40.8	34.7	39.3	4.65	0.85
3	*Q*, $K_{tr}{\cdot}C/d$, $C/d{\cdot}MC$	50.7	45.8	34.5	40.6	2.03	0.79
	Q, *MC*, $K_{tr}{\cdot}C/d$	54.9	50.4	33.6	39.7	2.11	0.85
	$\boldsymbol{MC, K_{tr}{\cdot}C/d, C/d{\cdot}Q}$	56.7	52.3	31.8	37.9	1.91	0.88

Table 5. Model coefficients for the logistic regression model

Model coefficients	β_0 (Intercept)	β_1 (*MC*)	β_2 ($K_{tr}{\cdot}C/d$)	β_3 ($C/d{\cdot}Q$)
Mean	-3.06	2.41	0.15	9.43
Standard deviation	1.13	1.07	0.08	4.59
Coefficient of variance	-0.37	0.44	0.53	0.49

In lasso regression, $\boldsymbol{\beta}$ is estimated for a sequence of penalty factor values. For a given penalty factor value, the independent variables are auto-selected (i.e., the variables with the corresponding coefficient being zero are deleted) when applying the sparse regularization through the penalization. Thus, corresponding to the sequence of the penalty factor values, there is a sequence of models with various model sizes. It is suggested that the model with average deviance that is one standard deviation away from the minimum average deviance should be selected as the final model, which can balance prediction (measured by deviance) against false discovery. As results, the selected model based on the lasso regression has three variables: C/d, $K_{tr}{\cdot}C/d$, and $MC{\cdot}Q$. The estimated model coefficients are provided in Table 6.

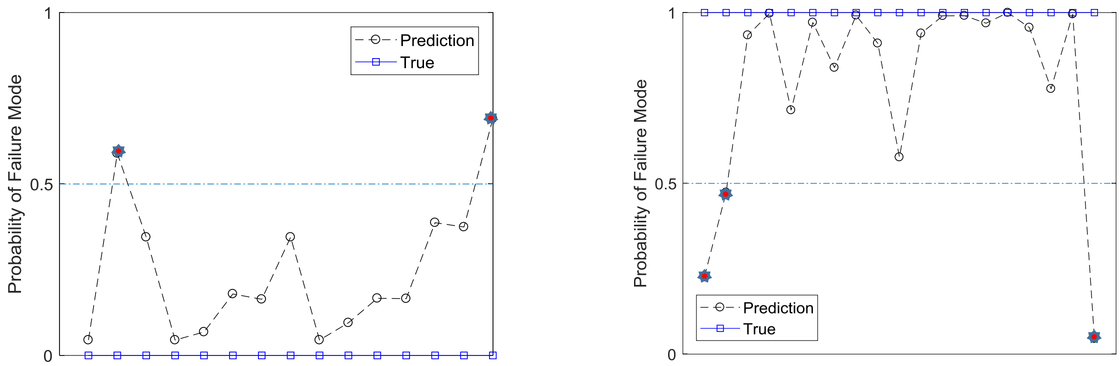
Table 4. Model coefficients for lasso logistic regression

Model coefficients	β_0 (Intercept)	β_1 (C/d)	β_2 ($K_{tr}\cdot C/d$)	β_3 ($MC\cdot Q$)
Mean	-3.56	0.70	0.10	53.35
Standard deviation	0.41	0.10	0.01	6.18
Coefficient of variance	-0.12	0.14	0.10	0.12

Logistic and lasso logistic regression comparison

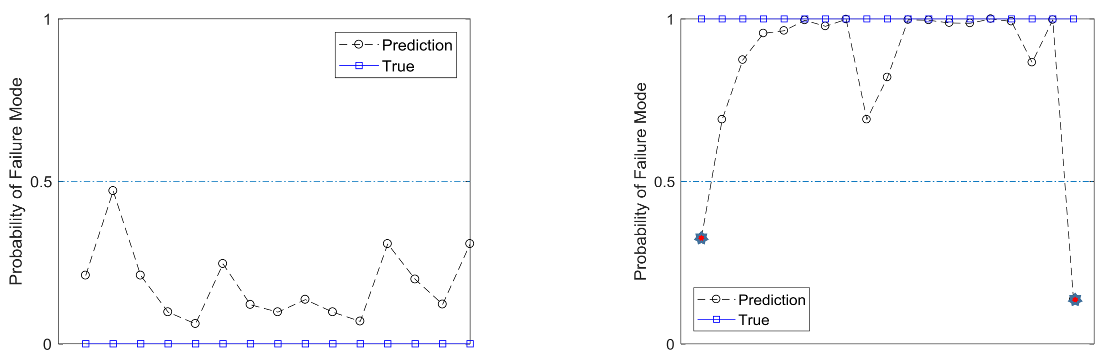
As shown in Table 5 and Table 6, all the mean values of β_1, β_2, and β_3 are positive for both models. The positive coefficients show that the likelihood of being pull-out failure mode increases with the increase of the associated term. Even though the variable terms selected in both models are different, except $K_{tr}\cdot C/d$, both models indicate that the failure mode leans to pull-out failure when more confinement (e.g., the ratio of cover to the diameter of rebar, C/d, and transverse stirrups) is provided, or corrosion level increases, or structure subjected to cyclic loading. In Kivell (2012), they have also found that specimens corroded specimens subjected to cyclic loading have more tendency of failing in pull-out. In addition, as expected, the coefficients of variance of $\boldsymbol{\beta}$ are much smaller for the lasso logistic regression than the ones for the logistic regression.

Figures 3 and 4 show the predicted probabilities of the specimen based on the developed regression models compared to the actual value (e.g., 1 for pull-out failure and 0 for splitting failure). The predictions of a failure mode using lasso logistic regression are closer to actual failure mode compared to the failure modes found using logistic regression. In particular, when using logistic regression model, there are two prediction probabilities larger than 50%, but the actual failure modes are splitting failure (as shown Figure 3(a) marked as red circles); and there are three prediction probabilities less than 50%, but the actual failures modes are pull-out failure (as shown Figure 3(b) marked as red circles). When using lasso logistic regression model, all the prediction probabilities are less than 50% when the specimens fail in splitting (as shown in Figure 4(a) marked as red circles); and there are only two prediction probabilities less than 50% when the specimens fail in pull-out (as shown in Figure 4(b) marked as red circles). The better accuracy in lasso logistic regression are also shown in the terms of SSE and P_{CD} defined earlier. Lasso logistic regression model has a smaller of SSE values and its probability of correct detection to be 94%, when using $\alpha = 50\%$ as the threshold.



(a) Splitting failure (b) Pull-out failure

Figure 3. Prediction plot based on logistic regression model



(a) Splitting failure (b) Pull-out failure

Figure 4. Prediction plot based on lasso logistic regression model

Table 5. Logistic and lasso logistic comparison

Model	Independent variables used	*SSE*	P_{CD}
Logistic regression	MC, $K_{tr}\cdot C/d$, $C/d\cdot Q$	2.03	88%
Lasso logistic regression	C/d, $K_{tr}\cdot C/d$, $MC\cdot Q$	1.5	94%

SUMMARY AND CONCLUSIONS

Adequate bonding between rebar and concrete is the key to ensuring the reliable performance of RC structures. This rebar-concrete bond behavior directly influences the structural load carrying capacity and structure failure mode. While many past studies have focused on the prediction of bond strength, predictions on the bond failure mode (i.e., pullout failure or splitting failure) is not well investigated, particularly when the concrete is not well confined and/or corrosion is present.

In this study, engineering machine learning techniques, logistic and lasso logistic regression, are implemented to develop bond failure mode prediction models. Compared with other machine learning techniques, logistic and lasso logistic regression are able to provide an explicit prediction formulation. To developed the prediction models, this study uses the bond tests of 44 beam-end specimens. For logistic regression, an all possible subset approach is adopted to simplify the model formulation by removing the variables that do not contribute to the model predictions. For lasso logistic, sparse regularization used in estimating the model parameters auto-selects variables for the model.

As results, both logistic and lasso logistic regression select 3 independent variables in their final model formulations. Even though the variables are not completely the same, they all suggest that the failure mode tends more to be pull-out failure when more confinement (e.g., the ratio of cover to the diameter of rebar, C/d, and transverse stirrups) is provided, corrosion level increases, or structure subjected to cyclic loading. In addition, the results show that the predictions of the failure mode using lasso logistic regression are closer to the true failure mode than the ones using logistic regression. When using a threshold, α = 50%, to determine the failure mode, the lasso logistic has a probability of correct detection of 94%. Since lasso logistic regression is suitable for large number of potential independent variables and small sample size, which is verified in this study. The results shows that the performance of lasso logistic regression is better than logistic regression. The prediction model developed in this study will be further updated when new experiment data become available. These models can be further used for corroded RC structure performance evaluation.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. CMMI 1635507. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

ACI (American Concrete Institute). "Report on bond of steel reinforcing bars under cyclic loads." (2012).

Almusallam, Abdullah A., Ahmad S. Al-Gahtani, and Abdur Rauf Aziz. "Effect of reinforcement corrosion on bond strength." *Construction and building materials* 10.2 (1996): 123-129.

ASTM A944. "Standard test method for comparing bond strength of steel reinforcing bars to concrete using beam-end specimens." *Annual Book of ASTM Standards* (2014).

Böhni, Hans, ed. *Corrosion in reinforced concrete structures*. Elsevier, 2005.

Castel, Arnaud, et al. "Modeling steel concrete bond strength reduction due to corrosion." *ACI Structural Journal* 113.5 (2016): 973.

Code, CEB-FIB Model. "CEB-FIB Model Code 2010–Final draft." *Thomas Thelford, Lausanne, Switzerland* (2010).

Cucchiara, Calogero, Lidia La Mendola, and Maurizio Papia. "Effectiveness of stirrups and steel fibres as shear reinforcement." *Cement and concrete composites* 26.7 (2004): 777-786.

Hadje-Ghaffari, Hossain, et al. "Bond of epoxy-coated reinforcement: cover, casting position, slump, and consolidation." American Concrete Institute, 1994.

Harajli, Mohamed H., Bilal S. Hamad, and Ahmad A. Rteil. "Effect of Confinement of Bond Strength between Steel." *ACI Structural Journal* 101.5 (2004): 595-603.

Hussain, S. E., A. Al-Musallam, and A. S. Al-Gahtani. "Factors affecting threshold chloride for reinforcement corrosion in concrete." *Cement and Concrete Research* 25.7 (1995): 1543-1555.

Kivell, Anton Richard Lean. "Effects of bond deterioration due to corrosion on seismic performance of reinforced concrete structures." (2012).

Lindsey, Charles, and Simon Sheather. "Variable selection in linear regression." *Stata Journal* 10.4 (2010): 650.

Mangalathu Sivasubramanian Pillai, Sujith. *Performance based grouping and fragility analysis of box-girder bridges in California*. Diss. Georgia Institute of Technology, 2017.

Orangun, C. O., J. O. Jirsa, and J. E. Breen. "A reevaulation of test data on development length and splices." *Journal Proceedings*. Vol. 74. No. 3. 1977.

Sajedi, Siavash, and Qindan Huang. "Probabilistic prediction model for average bond strength at steel–concrete interface considering corrosion effect." *Engineering Structures* 99 (2015): 120-131.

Standard, An ACI. "Building Code Requirements for Structural Concrete (ACI 318-11)." *American Concrete Institute*. 2011.

Stanish, Kyle David. *Corrosion effects on bond strength in reinforced concrete*. Diss. National Library of Canada= Bibliothèque nationale du Canada, 1999.

Tibshirani, Robert. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society. Series B (Methodological)* (1996): 267-288.

Torre-Casanova, Anaëlle, et al. "Confinement effects on the steel–concrete bond strength and pull-out failure." *Engineering Fracture Mechanics* 97 (2013): 92-104.

Turk, Kazim, and M. Sukru Yildirim. "Bond strength of reinforcement in splices in beams." *Structural Engineering and Mechanics* 16.4 (2003): 469-478.

Wang, Huanzi. "An analytical study of bond strength associated with splitting of concrete cover." *Engineering Structures* 31.4 (2009): 968-975.

Zandi Hanjari, Kamyab, Dario Coronelli, and Karin Lundgren. "Bond capacity of severely corroded bars with corroded stirrups." *Magazine of concrete Research* 63.12 (2011): 953-968.