ARTICLE IN PRESS

Engineering Applications of Artificial Intelligence xx (xxxx) xxxx–xxxx

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

# The use of artificial intelligence combiners for modeling steel pitting risk and corrosion rate

Jui-Sheng Chou[a,*], Ngoc-Tri Ngo[a,b], Wai K. Chong[c]

[a] Department of Civil and Construction Engineering, National Taiwan University of Science and Technology, 43, Sec. 4, Keelung Rd., Taipei 106, Taiwan
[b] Faculty of Project Management, The University of Danang – University of Science and Technology, 54 Nguyen Luong Bang, Danang, Vietnam
[c] Del E. Webb School of Construction, School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ 85287, USA

ARTICLE INFO

*Keywords:*
Artificial intelligence
Machine learning
Meta ensemble
Metaheuristic regression
Pitting risk
Corrosion rate
Engineering application

ABSTRACT

Corrosion is a common deterioration that reduces the service life of concrete structures and steels. Particularly, corrosion behavior is a highly nonlinear problem influenced by complex characteristics. This study used advanced artificial intelligence (AI) techniques to predict pitting corrosion risk of steel reinforced concrete and marine corrosion rate of carbon steel. The AI-based models used for prediction included single and ensemble models constructed from four well-known machine learners including artificial neural networks (ANNs), support vector regression/machines (SVR/SVMs), classification and regression tree (CART), and linear regression (LR). Notably, a hybrid metaheuristic regression model was implemented by integrating a smart nature-inspired metaheuristic optimization algorithm (*i.e.*, smart firefly algorithm) with a least squares SVR. Prediction accuracy was evaluated using two real-world datasets. According to the comparison results, the hybrid metaheuristic regression model was better than the single and ensemble models in predicting the pitting corrosion risk (mean absolute percentage error=5.6%) and the marine corrosion rate (mean absolute percentage error=1.26%). The hybrid metaheuristic regression model is a promising and practical methodology for real-time tracking of corrosion in steel rebar. Civil engineers can use the hybrid model to schedule maintenance process that leads to risk reduction of structure failure and maintenance cost.

## 1. Introduction

Civil infrastructures are often subject to aggressive conditions and consequently serious deterioration may occur. In recent years, the maintenance of concrete structures has become important with respect to the sustainability of infrastructure. Corrosion of reinforcing steel-bar (rebar) is a typical deterioration process that reduces the service lives of reinforced concrete structures and steel structures. To evaluate the performance of concrete members and structures, the corrosion of steel rebar must be predicted as early as possible.

For the industrial sectors in the USA, the repairing cost of corrosion damage is equal annually to 3.1% of the gross national product, among which the utilities, transportation and infrastructures contribute the most (Shaw and Kelly, 2006). Approximately 70% of corrosion occurs in a localized area (Nimmo and Hinds, 2003) and localized corrosion is more dangerous than uniform corrosion because it is more difficult to detect. The most common localized corrosion type is pitting corrosion, which may decrease steel strength as a result of a loss in rebar thickness and initiate early fatigue crack (Nimmo and Hinds, 2003). Worsening fatigue cracks can eventually cause a catastrophic failure of a structure. Risks are endemic in practically every aspect of our lives, and they always cause a great deal of potential damage (Wu and Birge, 2016). Thus, pitting corrosion must be identified before it reaches a dangerous level.

The rapid growth in the number of offshore structures urgently demand a model for predicting corrosion rate in order to reduce potential structural failures (Caines et al., 2013). However, corrosion is a highly nonlinear problem influenced by complex characteristics and models for predicting the corrosion rate of steel currently lack a theoretical basis. Researchers have yet to reach a consensus on the best model to predict corrosion rate or pitting risk due to the lack of good understanding of factors that affect the corrosion process. Many of the better known factors that affect corrosion of steel structure, like the electrochemical measurement of pitting corrosion, are not included in existing models. Therefore, an accurate model for predicting corrosion based on electrochemical data is needed.

Machine learning (ML) and artificial intelligence (AI)-based approaches have attracted a great deal of scientific attention (Chou and Ngo, 2016a; Wu et al., 2015) and have successfully used in civil engineering (Chou et al., 2015; Chou et al., 2016; Goel and Pal, 2009;

* Corresponding author.
*E-mail addresses:* jschou@mail.ntust.edu.tw (J.-S. Chou), D10205804@mail.ntust.edu.tw, trinn@dut.udn.vn (N.-T. Ngo), oswald.chong@asu.edu (W.K. Chong).

http://dx.doi.org/10.1016/j.engappai.2016.09.008
Received 17 May 2016; Received in revised form 30 August 2016; Accepted 17 September 2016
Available online xxxx
0952-1976/

Wen et al., 2009) such as modeling of pier scour (Pal et al., 2011). For instance, Wen et al. (2009) applied a single support vector regression (SVR) model for predicting the corrosion rate of 3C steel in five different seawater environments (Wen et al., 2009). Single artificial neural networks (ANNs) has been applied to predict pitting corrosion in steel reinforced concrete (Shi et al., 2011; Wen et al., 2009). To the best knowledge of the authors, only single AI models are proposed in literature to modeling the corrosion related issues.

Although the single AI-based models have proven moderately effective for solving prediction problems, one of the critical problems is how to select an appropriate model and fine-tune the model parameters, which plays an important role in good generalization performance and prediction accuracy for future use (Chou and Ngo, 2016b; Jiménez-Come et al., 2013; Yang et al., 2011). To overcome the drawback and enhance the accuracy, ensemble and metaheuristic approaches appear to be promising solutions for the above situations. Hybrid approaches that combine multiple AI models and optimization algorithms have been proposed to enhance the prediction accuracy of single AI models (Kazem et al., 2013).

This study investigated the applicability of four single AI models, four meta ensembles (*i.e.*, voting, bagging, stacking, and tiering), and a hybrid metaheuristic regression in solving the non-linear problem of estimating pitting risk or corrosion rate of steel rebar. The single AI models are ANNs, support vector machines and regression (SVMs/SVR), classification and regression tree (CART), and linear regression (LR). These single AI models are the most commonly used techniques in related works (Chou and Pham, 2013). Moreover, four ensemble AI models are combined from the above single AI models. The ensemble AI models consists of voting, bagging, stacking, and tiering methods. The hybrid metaheuristic regression model integrates the nature-inspired metaheuristic optimization (*i.e.*, smart firefly algorithm (SFA)) and least squares support vector regression (LSSVR).

The first originality of this study was a comprehensive investigation of AI models for modeling pitting corrosion risk and marine corrosion rate. The combination of those AI models has not been proposed yet in modeling steel corrosion. Secondly, the hybrid metaheuristic regression was presented to predict steel corrosion, in which the SFA was used to automatically optimize the hyperparameters of the LSSVR for improving prediction accuracy. Lastly, the findings of this study provide civil engineers with a promising and practical methodology for tracking of steel corrosion. Two real-world datasets were used to evaluate the applicability of various AI models. The performance of the investigated models is compared in terms of mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and synthesis index (SI).

The rest of this paper is organized as follows. Section 2 briefly reviews the relevant literature on corrosion behavior of steel rebar in reinforced concrete structures as well as steel structures. The machine learners, meta ensembles, metaheuristic regression, and the method used for model performance evaluation are introduced in Section 3. Section 4 presents the results of numerical experiments using two real-world datasets. Finally, Section 5 concludes with remarks and discussions.

## 2. Literature review

Risk management has become a vital topic both in academia and practice during the past decades and most intelligence tools have been used to enhance risk management (Wu, 2016; Wu et al., 2014). In civil infrastructure domain, corrosion behavior has been identified as a critical issue of risk by many researchers. Han et al. investigated the initial corrosion behavior of carbon steel exposed to cyclic wet-dry conditions in an outdoor environment (Han et al., 2014). They found that the stages of localized corrosion are pitting corrosion followed by fusiform corrosion. The corrosion rate of carbon steel exposed to outdoor wet-dry cycles is three times faster than that of carbon steel exposed to a natural environment, which suggests a strong correlation between outdoor wet-dry cycling and exposure to the atmospheric environment.

Mendili et al. (2014) investigated corrosion of carbon steel in a carbon dioxide clay-rich environment to understand its behavior under different geological conditions. They found that magnetite formation was the main corrosion product in the first step of the corrosion process, followed by the formation of different corrosion products with complex mixtures of iron-oxide, hydroxycarbonate, hydroxychloride, and sulfide phases. Their study demonstrated the need to consider all components of a disposal site, including the atmosphere composition to intent to explain the corrosion mechanism.

Shi et al. (2011) performed a quantitative study of the main factors in the chloride threshold of pitting corrosion of steel in concrete. To identify these factors, they performed a series of laboratory tests to assess the corrosion potential ($E_{\mathrm{corr}}$) and pitting potential ($E_{\mathrm{pit}}$) of steel coupons in simulated concrete pore solutions. They proposed a phenomenological model with the aid of an ANN to correlate the influential factors (total chloride concentration, chloride binding, solution pH, and dissolved oxygen (*DO*) concentration) with the pitting corrosion risk (characterized by $E_{\mathrm{corr}}$–$E_{\mathrm{pit}}$). Performance tests of this model obtained a coefficient of determination (*R-square*) of 0.8741.

The 3C steel is carbon steel which has a good operational performance and widely used in offshore engineering, pipelines, mining and construction, and so forth. Seawater is recognized to be one of the most corrosive natural electrolytes under natural environment. Carbon steel immersed in seawater is gradually eroded by chemical or electrochemical reactions. Wen et al. (2009) established an optimal support vector regression (SVR) model for predicting the corrosion rate of 3C steel in five different seawater environments. Their analytical results showed that the prediction accuracy of the SVR model was consistently higher than that of the back-propagation neural network (BPNN) under similar training and test conditions.

Most studies have focused on the mechanical aspects of corrosion behavior rather than on predicting the pitting risk or the corrosion rate. Since predicting uncertainties is beneficial in practices (Wu and Wu, 2016), a major advantage of early prediction of corrosion is reduced failure risk and maintenance costs. Besides, no study in literature has identified optimal models for corrosion prediction. This study thus compared the above models in terms of accuracy in forecasting pitting risk and corrosion rate.

## 3. Methodology

Researchers in many fields now use ML techniques to simulate material behavior. The ANNs, CART, LR, and SVMs are the most commonly used techniques in related works and are also considered the best data mining algorithms (Wu et al., 2007). Therefore, these four techniques were adopted in this study to develop single (baseline) models as well as their ensembles, and metaheuristic regression model. The meta ensembles are voting, bagging, stacking, and tiering models. The single and ensemble models were developed in the WEKA environment which is an open source software for data mining (Hall et al., 2009a). The hybrid metaheuristic regression model was coded and developed in the MATLAB environment. The following subsections elaborate theories and rationales of machine learners, meta ensembles, and metaheuristic regression model.

### 3.1. Machine learners

#### 3.1.1. Artificial neural networks

The information-processing units in artificial neural networks are artificial neurons similar to the neurons in the human brain (Haykin, 1998). Neural networks learn by experience; that is, they generalize from previous experiences to new ones, and make decisions using those experiences. A neural network consists of a group of neural nodes that