Computational Materials Science 168 (2019) 180–187

Contents lists available at ScienceDirect

Computational Materials Science

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

Towards machine learning approaches for predicting the self-healing efficiency of materials

Wenjun Wang[a,1], Nicolette G. Moreau[b], Yingfang Yuan[a], Paul R. Race[b,c,2], Wei Pang[a,*]

[a] Department of Computing Science, University of Aberdeen, UK
[b] Department of Biochemistry, University of Bristol, UK
[c] BrisSynBio Synthetic Biology Research Centre, University of Bristol, UK

ARTICLE INFO

Keywords:
Self-healing efficiency
Predictive model
Regression and classification
Artificial neural network
Online ensemble learning framework

ABSTRACT

Self-healing materials with an inherent repair mechanism have been widely studied. However, the self-healing efficiencies of most materials can only be measured by laboratory-based experiments, which can be time consuming and expensive. Inspired by modern machine learning approaches, we are interested in predicting the self-healing efficiency of new bio-hybrid materials, as part of our ongoing EPSRC funded "Manufacturing Immortality" project. By modelling existing experimental data, predictive models can be built to forecast self-healing efficiency. This has the potential to reduce the time input required by laboratory experiments, guide material and component selection, and inform hypotheses, thereby facilitating the design of novel self-healing materials. In this position paper, we first present preliminary knowledge and quantitative definitions of the self-healing efficiency of materials. We then demonstrate several widely used machine learning approaches and review an experimental case of predictive modelling based on neural networks. Furthermore, and aiming to expedite self-healing material development, we propose an on-line ensemble learning framework as the whole system model for the optimization of predictive computational models. Finally, the rationality of our on-line ensemble learning framework is experimentally studied and validated.

1. Introduction

The ability to autonomously sense and repair damage is an inherent property of many biological systems. Bio-mimetic approaches that seek to reproduce natural mechanisms of self-repair using both bio-derived and synthetic components, have opened up new opportunities in the development of materials with latent self-healing abilities [1,2]. Such materials have the potential to benefit human sustainable manufacturing systems by increasing product lifetime, reducing repair costs, and will ultimately contribute to the goal of manufacturing self-sustaining materials [3,4]. Generally, prospective self-healing materials are those which can repair damage and restore lost or degraded properties using inherently accessible resources [5] without any requirement for external diagnosis or extrinsic intervention [2]. This is currently measured quantitatively by comparing the extent of restoration and recovery following damage, versus the original state, and can be defined as self-healing efficiency (SHE) [1]. Recently, many self-healing materials, including plastics, rubbers, ceramics, concrete, glass, and metal based systems have been explored, encompassing both coatings and bulk materials [1,2,6].

Despite successes, significant challenges in the design and manufacture of self-healing materials remain. The reason for this is two-fold: i) There are a vast array of potential design choices for new materials, meaning that the underlying search space for new self-healing systems is so large as to be insurmountable; and ii) As environmental conditions for self-healing are varied, the self-healing efficiencies may vary as well. It is therefore necessary to accurately assess the self-healing behavior in all possible environmental conditions. As a result, the selection of appropriate test configurations in combination with phenomenologically different experiments and healing conditions is required, and thus many tens, or even hundreds, of laboratory experiments are required to thoroughly measure SHE [7]. Undertaking such extensive testing is often intractable in 'extreme' conditions, including those that pose significant health and safety risks to humans, and where self-healing materials would have the most significant impact.

* Corresponding author. ORCID: 0000-0002-1761-6659.
E-mail address: pang.wei@abdn.ac.uk (W. Pang).
[1] ORCID: 0000-0003-3579-1474.
[2] ORCID: 0000-0003-0184-5630.

https://doi.org/10.1016/j.commatsci.2019.05.050
Received 21 February 2019; Received in revised form 23 May 2019; Accepted 25 May 2019

Usually, a new material prototype is designed according to the knowledge of the material itself, or closely related materials, with this knowledge derived from previous experimentation. This is followed by the in-laboratory production of several specimens and associated testing of their properties (or performance), with regards to self-healing behaviour. The SHEs of these new material prototypes are then statistically analysed [8]. According to the experimental results, the design of new materials is further refined and adjusted, for instance, increasing or decreasing the amount of specific components, the addition of new components or the removal of trivial ones. Subsequently laboratory testing will be repeated to monitor SHE, and then a better design solution of self-healing materials may be selected, and the experiments will be carried out accordingly. This design-build-test-learn cycle can be repeated multiple times to produce the best prototype, making the design of a new material an iterative and time-consuming process. From the perspective of computational optimisation, material design may be considered as a search problem to discover the underlying design space of a novel self-healing material. We are aiming to design and discover better materials with high SHE under various environmental conditions, based on existing knowledge and experiences, historical data, and intuition [9].

Machine learning approaches enable more knowledge to be extracted from historical data, and furthermore, this knowledge can be used to predict properties of new materials [10], which could accelerate discovery and design processes for novel materials. For example, chemists have previously trained support vector machines (SVM) and neural networks (NN) to categorise a spectrum of reagents as reactive or non-reactive in order to discover new chemical reactions [11]. Consequently, new reactions were found and then verified by laboratory experimentation [11]. In a further example a general procedure of applying machine learning for lithium-ion battery development is presented as three key steps: i) sample construction, ii) model building, and iii) model evaluation [12]. Based on these steps, several commonly used machine learning algorithms and their value in predicting the properties, structures and components of novel materials are reviewed [13]. Moreover, the fundamental interplay between the predictive capability of machine learning models and the availability of data were analysed, and a method of using small datasets was introduced by Zhang et al. in [14].

In our ongoing EPSRC funded project “New Industrial Systems: Manufacturing Immortality”, one of the key ambitions is to use machine learning approaches, such as regression and classification, to predict the SHE of new material prototypes. We aim to construct predictive models which are trained and validated by historical data, and these predictions will then be verified by laboratory experimentation. By using these predictive models, significant theoretical results can be drawn to support the intuitional inferences from experienced materials specialists, and it is expected that this will lead to new insights and inspirations. However, it should be acknowledged that predicted results may not be as accurate as those measured by lab experiments. Generally, a larger quantity and higher quality of historical data will improve the performance of predictive models, and therefore, our project also aims to create an experimentally verified database of parameter effects on SHE, that can be used in this way.

The contributions of this work are summarized as follows. Section 1 contains a brief introduction on self-healing materials and the related mathematical definitions regarding damage and SHE. Section 2 presents some primary knowledge on self-healing materials and extracts the definition of SHE. Then, machine learning approaches such as regression and classification are reviewed in Section 3, leading to a description of a neural network model for predicting SHE in Section 4. In Section 5, to enable the modelling of SHE, we propose an on-line ensemble learning framework including multiple learning approaches, which is experimentally studied and validated. Finally, in Section 6, conclusions and future suggestions are presented.

2. Primary definitions and types of self-healing materials

In this section, the background of self-healing materials will be presented in terms of the following three parts: (1) an explanation of intrinsic and extrinsic self-healing, (2) quantitative definitions of damage and SHE, and (3) a classification of three types of bio-hybrid self-healing materials. Whilst biological and bio-hybrid self-healing materials are only a subset of the self-healing materials available, this project has chosen to concentrate on these for their relatively untapped potential in this area. Biology has evolved over millions of years to both sense and repair damage rapidly and effectively and this is a property that is highly desirable in the manufacture of new materials. Furthermore, recent advances in the field of synthetic biology have made the incorporation of biological components tractable.

2.1. Intrinsic and extrinsic self-healing for material

Our extensive literature review on self-healing materials reveals that basic self-healing behaviour can be classified into two types: intrinsic and extrinsic self-healing [6]. Intrinsic self-healing means that the material can be healed by changing only the environmental conditions to stimulate intrinsic activities, such as Diels-Alder (DA) reactions [15], sol-gel transitions (SG) [16], or other spontaneous reactions [17]. In contrast, extrinsic self-healing is implemented by adding external healing components, such as those contained in capsules [5,18], or vascular based [19,20] healing systems. When damage occurs the vessel housing is broken, releasing the healing components to participate in or activate the healing process. The vascular based self-healing materials keep the healing components in capillaries (or vascular nets) [21]. The latter can be particularly useful or components that need to be continuously circulated or that need to be accessible at all locations within the material.

2.2. The quantitative definitions for damage and self-healing efficiency

There are many kinds of material damage, which can be visible or invisible, such as cracks or functional failures. This can include materials such as electrical wiring or fibre optic damage, as well as the more commonly considered bulk materials e.g. metal sheeting. To define it quantitatively, the property (or performance) of the virgin material and damaged material are denoted as F_0 and F_d, respectively [2]. It should be noted that the properties or performance of materials can include stress, stiffness, conductivity, bulk, and so on. Thus, the measurement of damage Δf_0 (i.e. loss of property or performance) can simply be defined as follows [1,7]:

$$\Delta f_0 = F_0 - F_d \tag{1}$$

As we can see, a lager value of Δf_0 indicates a bigger damage. If the damage material can be self-healed, then the measurement of recovery can be denoted as follows:

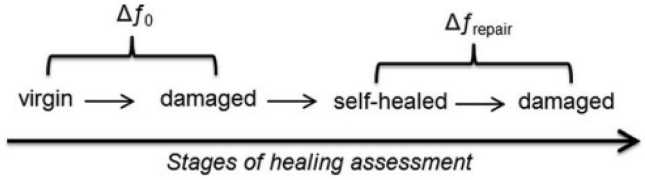


Fig. 1. The definition of self-healing efficiency. (Figure credits: [22]).

$$\Delta f_{repair} = F_{repair} - F_d \quad (2)$$

where the recovered property from self-healing is F_{repair}. Based on the definitions of damage and recovery, SHE can be defined as the percentage of performance that is recovered compared to that of the damaged performance, as shown in Fig. 1 [22]. Thus SHE can be calculated by the following equation [1,7]:

$$\eta = \Delta f_{repair} / \Delta f_0 \quad (3)$$

For some material M, a larger number of experiments on the functional property are carried out between the virgin status and damaged status to get a good estimation of Δf_0. Also, many experimental repeats are used to obtain an accurate, averaged damaged status for calculating Δf_{repair}. Supposing that all the damage occurs in the same (or very similar) manner, then the percentage calculated by Eq. (3) can stand for the self-healing efficiency of material M under the healing process H, termed η^{HM}. However, for any self-healing material, variation to the healing process H (including different environmental configurations for the self-healing process) may lead to different results. Thus the optimization approaches shown here could be beneficial in new material design, incorporating a wide range of parameters that cannot all be experimentally tested for.

2.3. The types of bio-hybrid self-healing materials

For the future research in our project, we will focus on the design of bio-hybrid self-healing material. Here, bio-hybrid materials and non-biological materials can be distinguished according to whether they contain biological components, such as nucleic acid, peptide, protein, or whole microorganisms [21,22], or not. If a material does not contain any biological components, e.g., some types of alloy metals, it should be considered as a non-biological material. For ease of clarification in this process, we further classify the bio-hybrid materials into three types (see Table 1), and the non-biological material as one type because we are currently focused on investigating those that include biological components. Furthermore, as shown in this table, we consider that the added components may either directly react with the material or indirectly activate as catalysts to accelerate the healing process.

Table 2
Regression models for prediction. (Table credits: [24]).

$F_a = ct^k + \varepsilon$	$F_a = c(T)t^k + \varepsilon$
$F_a = ct^{k(T)} + \varepsilon$	$F_a = c(T)t^{k(T)} + \varepsilon$
$F_a = ct^k + F_i + \varepsilon$	$F_a = c(T)t^k + F_i + \varepsilon$
$F_a = ct^{k(T)} + F_i + \varepsilon$	$F_a = c(T)t^{k(T)} + F_i + \varepsilon$
$F_a = ct^k + F_i(T) + \varepsilon$	$F_a = c(T)t^k + F_i(T) + \varepsilon$
$F_a = ct^{k(T)} + F_i(T) + \varepsilon$	$F_a = c(T)t^{k(T)} + F_i(T) + \varepsilon$
$F_a = ce^{tk} + \varepsilon$	$F_a = c(T)e^{tk} + \varepsilon$
$F_a = ce^{tk(T)} + \varepsilon$	$F_a = c(T)e^{tk(T)} + \varepsilon$
$F_a = ce^{tk} + F_i + \varepsilon$	$F_a = c(T)e^{tk} + F_i + \varepsilon$
$F_a = ce^{tk(T)} + F_i + \varepsilon$	$F_a = c(T)e^{tk(T)} + F_i + \varepsilon$
$F_a = ce^{tk} + F_i(T) + \varepsilon$	$F_a = c(T)e^{tk} + F_i(T) + \varepsilon$
$F_a = ce^{tk(T)} + F_i(T) + \varepsilon$	$F_a = c(T)e^{tk(T)} + F_i(T) + \varepsilon$

3. Machine learning approaches for predicting self-healing efficiency

Regression and classification are well-known machine learning approaches [27]. In this section we will briefly introduce our ideas to use regression and classification for the prediction of SHE.

3.1. Regression for prediction

Regression has been widely used for analyzing the relationship between dependent and independent variables [27]. For example, ridge regression, support vector regression and back-propagation neural networks have been used to predict the onset temperature of As_xSe_{1-x} glasses transition with respect to the input variable x (x indicates the Molar ratio of glasses components) [28]. In the case of SHE prediction, suppose the input of experiments (including the configuration of material and environmental conditions) are the independent variables, and the outputs (properties of material) are dependent variable(s). Regression models can be built to reveal the relationship between the input and output. An example of a bio-inspired self-healing material design, which benefitted from regression modelling was demonstrated by Sariola et al. [24]. In that work the authors used a structural protein with alternating soft and rigid domains to act as the self-healing mechanism and through regression modelling they learnt the relationships between the adhesion force F_a with relation to the duration time of the experiment t, residual ε, exponent parameter k, temperature-dependent parameter c, and temperature-independent constant F_0. This is shown in Table 2.

All the above-mentioned regression models in Table 2 are trained by historical data, i.e., all parameters are estimated. According to the criteria of minimum residual sum of squares (RSS), $F_a = c(T)t^k + F_0 + \varepsilon$ is the best fitted model, and its fitting curve is plotted in Fig. 2. As a result, if the adhesion forces F_a under other duration time is required, we could use this model $F_a = c(T)t^k + F_0 + \varepsilon$ to predict it without lab experiments.

Table 1
Bio-hybrid and non-bio-hybrid materials.

Component + *Material*	Non-bio- *Material*	Bio- *Material*
Non-bio Component	Epoxy + *Glass* [18]	Furan/amine + *polymer* [25]
Bio Component	Bacteria + *Concrete* [23]	Chitosan + *Hydrogel* [26]

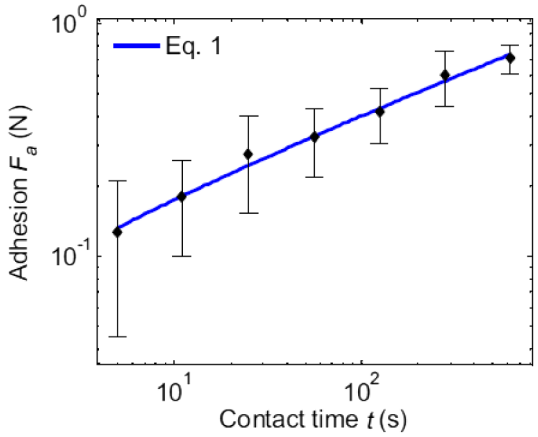


Fig. 2. The fitting curve for adhesion force F_a. (Figure credits: [24]).

3.2. Classification for prediction

Sometimes an exact fitting curve is desired but there is less historical data available, which means the problem is underfitted. In this case, the aim of predicting a real number for SHE should be reduced to categorizing a binary conclusion on whether the material can be healed or not. As presented above, there may be many factors that affect SHE, such as healing temperature, duration time, and PH value, which can be written as a variable vector X. Consider the following logistic regression model [27,29]:

$$p(y|X;\theta) \tag{4}$$

where θ is the parameter vector, y is a binary dependent variable with values 1 or 0 (which mean true or false, respectively). For ease of illustration, we can model this classifier as $h_\theta(X) = \theta^T X$. Then we would predict 1 on an input X if and only if $\theta^T X \geq 0$, as shown below:

$$p(y|X;\theta) = \begin{cases} 1, & \theta^T X \geq 0 \\ 0, & \theta^T X < 0 \end{cases} \tag{5}$$

It should be noted that a larger $\theta^T X$ indicates a larger $h_\theta(X) \Rightarrow p(y=1|X;\theta)$, thus a higher degree of confidence that the label being 1. On the other hand, if $\theta^T X \ll 0$, we consider a very high confident prediction of $y = 0$. So, an optimal classifier should be obtained by maximizing the confidences either $y = 1$ or $y = 0$. Usually the margins are intuitionally used to describe the confidences, so the optimal classifier can be achieved by maximizing the margins, as shown in Fig. 3.

In Fig. 3, the nearest distance from data points to the linear classifier is defined as the margin, and the red and blue data points (which have the same margin value in this figure) are reasonably defined as support

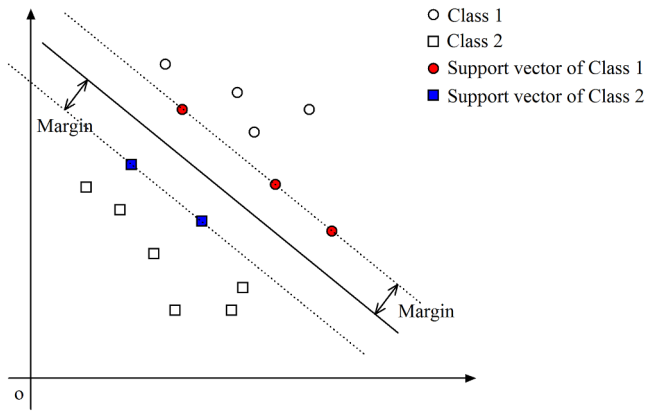


Fig. 3. Linear SVM and margins.

Table 3
Conversion percentage (Yields %) in the ROMP reaction of ENB and DCPD with different catalysts. (Table credits: [30]. ROMP denotes ring-opening metathesis polymerization; ENB denotes 5-ethylidene-2-norbornene; DCPD denotes dicyclopentadiene.)

Catalyst	Monomer	Temperature	Time	Yields
G1	DCPD	25	15	34
G1	DCPD	10	30	0
G1	ENB	25	0.5	100
G1	ENB	0	13	100
G1	ENB	−30	1440	74
G1	ENB	−40	1440	52
HG1	ENB	25	0.5	96
HG1	ENB	−50	450	99

vectors (with respect to the original point o). It is proved that the formula of classifier can be deduced based on its support vectors, so this method is named as support vector machine (SVM) [27]. SVMs are widely used in classification problems, and recently it was employed by an organic synthesis robot to search for new reactions [11]. In that study, SVM was trained by the spectrum of reactive and non-reactive reagents, and then it was used to predict a new reagent.

There are many material design problems that can use SVM to predict the properties of materials. For example, if we have already obtained some test data as shown in Table 3 [30], we can specify that a yield of more than 80% is considered as healed. Otherwise, a yield of below 80% is labelled as unhealed. An SVM model (classifier) can then be trained by these labelled data in order to predict the healing status under new configurations of reagents. In Table 3, the Ring-Opening Metathesis Polymerization was carried out using ENB and DCPD with different catalysts, and the reaction yield (conversion percentage) under different conditions are recorded, where G1 and HG1 denote Grubbs catalyst and Hoveyda-Grubbs 1, DCPD and ENB stand for dicyclopentadiene and 5-ethylene-2 norbornene, respectively.

4. Machine learning examples by using artificial neural networks

In this section we will first introduce the fundamental models of artificial neural networks (ANNs), and then present a classification experiment realized ANN to predict new reactions [11], and a simulated regression experiment for learning the curve of tensile stress (TS) with respect to the number-average molecular weight (Mn) [31].

4.1. Neuron model and neural networks

A single neuron model (also called a perception model) contains several inputs and one output as shown in Fig. 4 [32], where x_i ($i = 0, 1, \ldots, n-1$) are n input variables, where $x_0 = 1$ is a constant number. w_i ($i = 0, 1, \ldots, n-1$) are the n unknown weight parameters to be estimated, and the value of function f is the output. Usually, a neuron is a computational unit that takes n input $(1, x_1, x_2, \ldots x_{n-1})$, and the output as $f(X) = \sum w_i x_i$, (i = 0, 1, …, n − 1) , where f: $R \rightarrow R$ is named as the activation function. There are many choices for activation functions, such as linear summation function, sigmoid function, tanh function, that can be used for building a neural network [32].

An artificial neural network (ANN) is constructed by hooking together many of the simple neuron models, and thus the input of a neuron can be the output of another neuron [27]. It is noted that NN can contain many layers as shown in Fig. 5, where the leftmost layer is called the input layer, and the rightmost layer is called output layer (which, in this example, only has one node), and the middle layers are named as hidden layers. An ANN model can be constructed by using many neurons, layers, and different activation functions, and it is widely used in tackling those highly complex and large-scale problems [27].

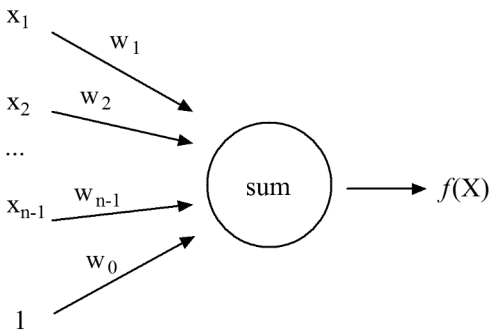


Fig. 4. Neuron model.

4.2. ANN for classification

In order to further explore the chemical reaction space, an ANN model was applied to the Suzuki-Miyaura reaction space dataset [33], in which the dimensions of independent variable increase to 37 [11], as shown in Fig. 6. In the first layer of ANN, the input is a 37-bits variable vector, which code reactant 1 with 7 bits, reactant 2 with 4 bits, ligand with 12 bits, chemical base with 8 bits, and chemical solvent with 6 bits, thus the final chemical reaction space is 37 dimensions. While in the second layer, 50 neurons with sigmoid activation functions are constructed. The third layer comprises 7 neurons in the fully connected layer (also with sigmoid activation functions). The final prediction of yield was obtained as the output from the third layer. Mean squared errors (MSE) between the predicted and experimental yield were used as the loss function to train this ANN. As reported, this ANN model was trained by 3,456 reactions and validated by 576 reactions with only 0.01208 MSE, when predicting the Suzuki-Miyaura dataset [33].

4.3. ANN for regression

To further demonstrate the value of ANN in regression, the relationship curve between the simulated tensile stress (TS) and the number-average molecular weight (M_n) is learned. An ANN with two hidden layers was constructed; each layer contains 4 neuron models with sigmoid activation functions. Fig. 7 shows the computational results for linear, nonlinear regression (with order 2 and 3), and ANN regression together with ANN predictions. It should be noted that the ANN regression only fits the TS values of the data points that have exact Mn values with the experimental data points, while the ANN prediction gives the whole prediction of TS with respect of a uniformly sampled Mn values from 1.0×10^5 to 1.2×10^6.

As shown in Fig. 7, compared to the traditional linear and nonlinear regressions, ANN regression has the minimum mean square errors

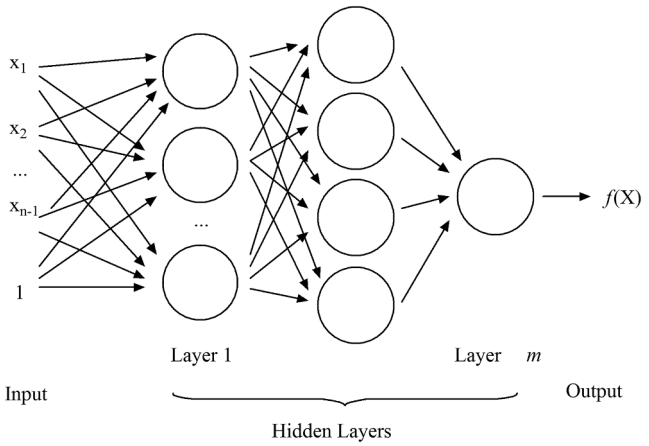


Fig. 5. The artificial neural network model.

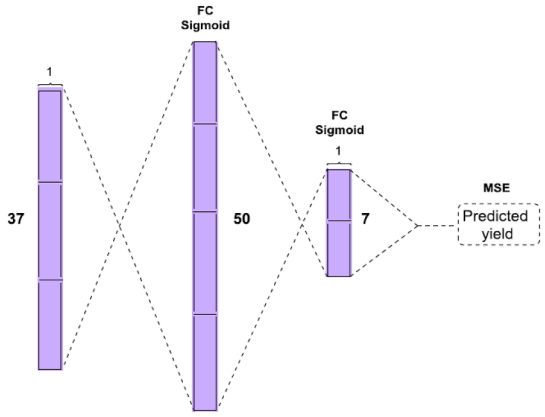


Fig. 6. The structure of ANN used for chemical reaction prediction (Figure credit: the supplementary file of [11]).

(MSE), and its predictive curve is more promising than other methods. To improve the performance of nonlinear regression, one approach is to increase the order of the nonlinear function, e.g., using 4 or even higher order of polynomial model to fit the data, however, this may lead to overfitting [34].

5. An on-line ensemble learning framework and experimental study

As lab experiments are undertaken we acquire an increasing number of specimens and experimental results. These can be used further to refine predictive models. In this section we present a novel on-line ensemble learning (OEL) framework with three open issues: i) a termination condition; ii) a refining method; and iii) an adviser of new material designs and environmental settings. Finally, the rationality of our OEL framework is experimentally studied.

5.1. An on-line ensemble learning framework

In order to refine the predictive model, we propose an OEL framework based on considering the SHE prediction error as feedback or reward. By analyzing this feedback, we can adjust the predictive models, or re-train the model based on both historical data and new experimental data, thus forming an on-line learning framework [35,36]. Here we present this on-line learning framework for SHE prediction with details shown in Fig. 8.

In this framework, there are two processes indicated by black and

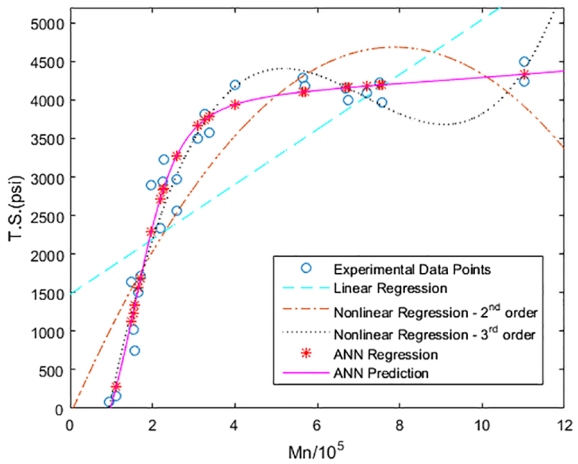


Fig. 7. Experimental data points and regression results by linear, nonlinear (order 2 and 3) and ANN models, and their prediction.

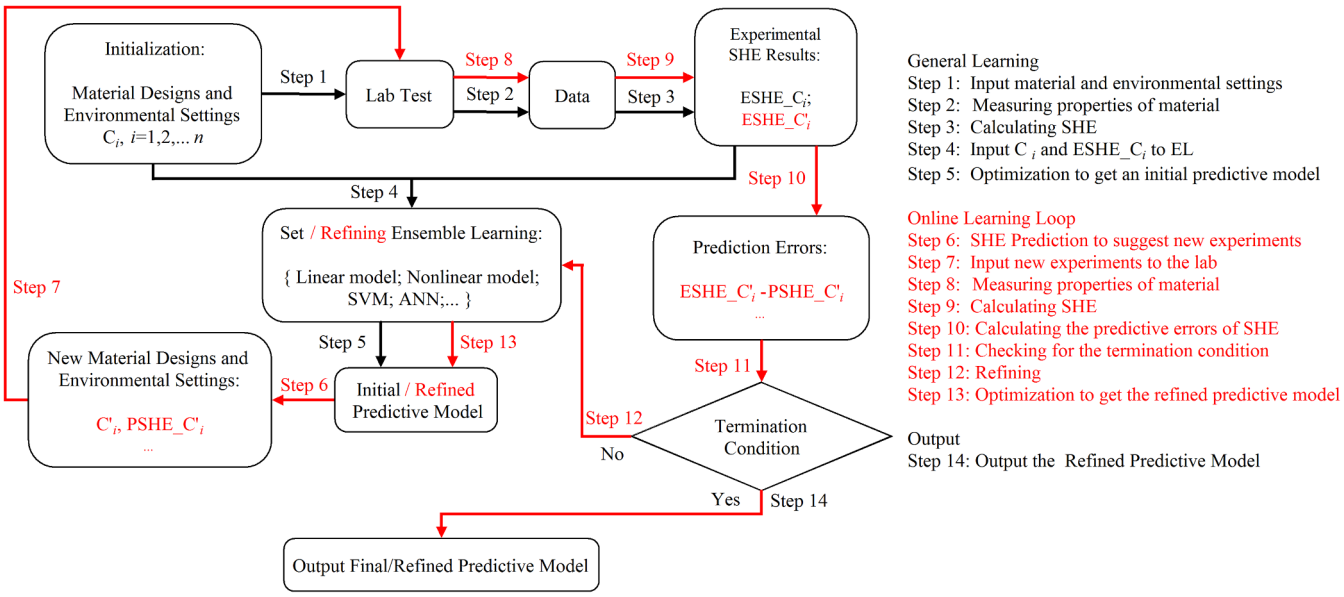


Fig. 8. An on-line ensemble learning framework for SHE prediction.

red lines, respectively. The black loop is the general learning process, and its starting point is the module of 'Initialization', while the end point is 'Initial predictive model'. The general learning process contain 5 steps as follow:

Step 1: Input material and environmental settings
Step 2: Measuring properties of material
Step 3: Calculating SHE
Step 4: Input C_i and ESHE_C_i to ensemble learning (EL)
Step 5: Optimization to get an initial predictive model

Firstly, some material design and environmental settings termed as C_i (i = 1, 2, …, n) are carried out based on historical experimental data. In Step 1, $\{C_i\}$ will be sent to lab for test. In Step 2, testers measure/record the property values concerned with SHE, then we can get the experimental SHE termed as ESHE_C_i through calculation in Step 3. Step 4 denotes the modelling process, and in this step, we construct an ensemble of machine learning models for prediction, including such as linear model, nonlinear model, SVM, and ANN, but we are not restricted to these models. According to the MSE criterion, in Step 5 the optimal machine learning model is selected as the initial predictive model.

The red loop stands for the on-line learning progress, which contains 7 steps as follows:

Step 6: SHE prediction to suggest new experiments
Step 7: Request new experiments to the lab
Step 8: Measuring properties of material
Step 9: Calculating SHE
Step 10: Calculating the predictive errors of SHE
Step 11: Checking for the termination condition
Step 12: Refining the ensemble of predictive models
Step 13: Optimization to get the refined predictive model

In Step 6, based on the presented initial predictive model, several new material designs with high SHE will be carried out in feasible environmental settings, which are termed as C_i' (i = 1, 2, …, n'), and their SHEs are predicted and termed as PSHE_C_i'. Step 7, Step 8, and Step 9 are similar to Step 1, Step 2 and Step 3, which are used to get new experimental SHEs, termed as ESHE_C_i'. Step 10 calculates the prediction error of the proposed predictive model, and through Step 11 these prediction errors are sent for termination condition checking. If the termination condition is satisfied, then in Step 14 we will have the final predictive model, otherwise it will go to Step 12 for refining, e.g, adding a new selective model, deleting a poorly performed one, or simply tuning the parameters of existing models. Through Step 13 the predictive models will be adjusted by using the information from SHE prediction errors, material configurations and environmental settings as feedbacks. Then we will go to Step 6 to re-start a new on-line learning iteration, until the termination condition is satisfied.

There are three open issues in this framework that should be addressed further, and they are listed as follows.

(1) Termination condition

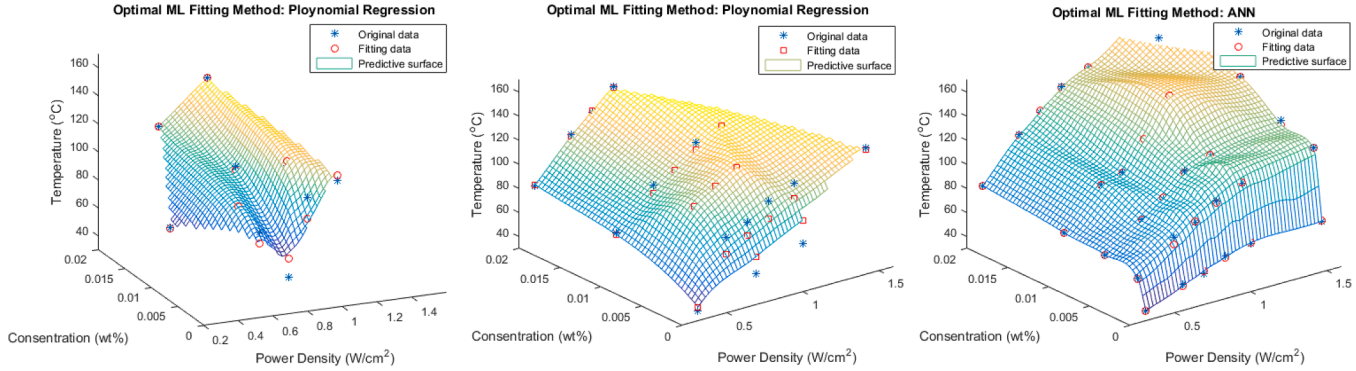
Intuitively, we can set a threshold for the predictive error of SHE. However, as we know there might be over-fitting issues which would lead to poor predictions. By combining the MSE threshold with other criterion e.g., smoothness constraints [34] or regularization [37], this OEL approach would give more robust results.

(2) Refining ensemble and its predictive models

With the prediction errors as feedback, we can retrain the initial predictive model to adjust its parameters. Alternatively, we can also add more predictive models with new typical configurations and settings into the ensemble that we constructed. For instance, if we choose ANN as the optimal initial predictive model, we can not only refine its parameters (usually referring to weight vectors), but can also change its layer and node numbers, activation functions and training methods (the so-called hyperparameters). As different configurations of ANN have different performance, in this case, adding several different ANNs with characteristic configurations into the ensemble would show the advantage of ensemble learning. However, characteristic configuration of ANN is a challenging issue, which will depend on specific problems in material design.

(3) Suggesting new material design solutions and environmental settings

With the initial/refined predictive model, any SHE of new material design solutions within varying environmental settings could be



(a) Results by sub-dataset 1; (b) Results by sub-datasets 1 & 2; (c) Results by sub-datasets 1, 2 & 3

Fig. 9. Experimental results of the OEL framework.

calculated. However, we cannot predict all the cases in both chemical and environmental setting spaces. According to the predictive model, a fast adviser that can quickly guide material design should be further investigated.

5.2. Experimental study

To validate the rationality of the proposed OEL framework, a simulated experiment using incremental data has been studied. We choose self-healing electrical conductivity polyurethane nanocomposites as target material (termed as TM). As the SHE of TM depend on its surface temperature heavily, we suppose the surface temperature can stand for SHE of TM. The surface temperature property varies with respect to two variables: the concentration of the structure graphene oxide materials (rmGO) and laser density [38]. The former belongs to material design choices, and the latter belongs to environmental settings.

The whole experimental results are randomly divided into three subsets, and each subset contains 10 samples. We first use subset 1 as $\{C_i\}$ which is presented in Section 5.1, and then suppose that subset 2 and subset 3 respectively stand for the on-line dataset $\{C_i'\}$ and $\{C_i''\}$. Here we do not consider how to suggest new material design solutions and environmental settings in this context. Also for simplicity, the refining method now we used is just re-training the model, while the termination condition is set to be a threshold of prediction errors in terms of MSE.

The ensemble of predictive models contains two models: 1) nonlinear polynomial regression model; 2) an ANN with 9 hidden layers, and with both sigmoid and radial basis functions for its activation functions. Fig. 9 shows one of the experimental results of the proposed OEL framework. It is indicated that the optimal model changed with respect to the size of the dataset: at the beginning when only $\{C_i\}$ is available, the optimal predictive model is the polynomial regression model (Fig. 9(a)). However, later on when new experimental data are available, the optimal predictive model is changed to ANN (Fig. 9(c)). This demonstrates the rationality of using ensemble learning. Moreover, as the on-line dataset scales up, the nonlinear polynomial regression model seems not suitable for describing complicated surface any more, while the ANN model demonstrates its advantage. However, according to our experiments ANN will not be always better than nonlinear polynomial regression model, due to the randomly generated initial parameters used in training ANN.

6. Conclusions and future works

This position paper aims to give a brief introduction to self-healing materials and explore possible predictive models to elucidate their SHEs. By reviewing the fundamental knowledge of self-healing materials the typical machine learning methods in related studies, as well as performing several simulated experiments, three main conclusions can be summarized as follow:

1) Regression methods can be used for discovering the relationship of SHE and other properties of material. Suppose that there is a new design solution for a self-healing material, but the relationship between its properties affects its SHE value, in this case regression approaches to approximate the functional curve of SHE can be used. The desired SHE value estimation under new properties and environmental configurations can then be predicted by machine learning models (predictive models).
2) Classification has been successfully used to predict new chemical reactions. Inspired by this, both the SVM and ANN models trained by historical data can be used to predict whether a new material would possess the self-healing ability.
3) By using prediction errors of predictive models and lab experiments as feedback, an on-line ensemble learning framework is proposed to improve the accuracy and reliability of predictive models. Also, the effectiveness of ensemble and on-line learning are experimentally studied.

As for our future work, we would like to focus on further exploring the proposed OEL framework, especially to tackle the three open issues mentioned in Section 5.1. With respect to the immediate next step of our EPSRC project, the presented OEL framework and predictive models would be adapted and employed to facilitate the advanced bio-hybrid self-healing material design tasks, and this will also further verify the effectiveness of OEL on these tasks.

CRediT authorship contribution statement

Wenjun Wang: Conceptualization, Formal analysis, Software, Validation, Writing - original draft. **Nicolette G. Moreau:** Resources, Investigation, Writing - original draft. **Yingfang Yuan:** Investigation, Writing - review & editing. **Paul R. Race:** Methodology, Project administration, Supervision, Writing - review & editing, Funding acquisition. **Wei Pang:** Methodology, Project administration, Supervision, Writing - review & editing, Funding acquisition.

Acknowledgements

This research is supported by the Engineering and Physical Sciences Research Council (EPSRC) funded Project on New Industrial Systems: Manufacturing Immortality (EP/R020957/1). The authors are also grateful to the Manufacturing Immortality consortium.

References

[1] B.J. Blaiszik, S.L.B. Kramer, S.C. Olugebefola, J.S. Moore, N.R. Sottos, S.R. White, MR40CH08-white self-healing polymers and composites, Annu. Rev. Mater. Res. (2010), https://doi.org/10.1146/annurev-matsci-070909-104532.

[2] S.K. Ghosh, Self-Healing Materials: Fundamentals, Design Strategies, and Applications, 2009, https://doi.org/10.1002/9783527625376.

[3] M. Sarkar, D. Adak, A. Tamang, B. Chattopadhyay, S. Mandal, Genetically-enriched microbe-facilitated self-healing concrete-a sustainable material for a new generation of construction technology, RSC Adv. 5 (2015) 105363–105371, https://doi.org/10.1039/c5ra20858k.

[4] D. Sun, H. Zhang, X.Z. Tang, J. Yang, Water resistant reactive microcapsules for self-healing coatings in harsh environments, Polym. (United Kingdom) (2016), https://doi.org/10.1016/j.polymer.2016.03.044.

[5] H. Jin, C.L. Mangun, A.S. Griffin, J.S. Moore, N.R. Sottos, S.R. White, Thermally stable autonomic healing in epoxy using a dual-microcapsule system, Adv. Mater. (2014), https://doi.org/10.1002/adma.201303179.

[6] D.G. Bekas, K. Tsirka, D. Baltzis, A.S. Paipetis, Self-healing materials: a review of advances in materials, evaluation, characterization and monitoring techniques, Compos. B Eng. (2016), https://doi.org/10.1016/j.compositesb.2015.09.057.

[7] M.D. Hager, P. Greil, C. Leyens, S. Van Der Zwaag, U.S. Schubert, Self-healing Materials, 2016, https://doi.org/10.1002/adma.201003036.

[8] B.M.D. Hager, P. Greil, C. Leyens, S. Van Der Zwaag, U.S. Schubert, Self-healing materials, Adv. Mater. 22 (2010) 5424–5430, https://doi.org/10.1002/adma.201003036.

[9] A. Mannodi-Kanakkithodi, G. Pilania, T.D. Huan, T. Lookman, R. Ramprasad, Machine learning strategy for accelerated design of polymer dielectrics, Sci. Rep. 6 (2016), https://doi.org/10.1038/srep20952.

[10] M. Wang, N. Pan, Predictions of effective physical properties of complex multiphase materials, Mater. Sci. Eng. R Rep. (2008), https://doi.org/10.1016/j.mser.2008.07.001.

[11] J.M. Granda, L. Donina, V. Dragone, D.L. Long, L. Cronin, Controlling an organic synthesis robot with machine learning to search for new reactivity, Nature 559 (2018) 377–381, https://doi.org/10.1038/s41586-018-0307-8.

[12] S. Shi, J. Gao, Y. Liu, Y. Zhao, Q. Wu, W. Ju, C. Ouyang, R. Xiao, Multi-scale computation methods: their applications in lithium-ion battery research and development, Chin. Phys. B 25 (2016), https://doi.org/10.1088/1674-1056/25/1/018212.

[13] Y. Liu, T. Zhao, W. Ju, S. Shi, S. Shi, S. Shi, Materials discovery and design using machine learning, J. Mater. 3 (2017) 159–177, https://doi.org/10.1016/j.jmat.2017.08.002.

[14] Y. Zhang, C. Ling, A strategy to apply machine learning to small datasets in materials science, Npj Comput. Mater. (2018), https://doi.org/10.1038/s41524-018-0081-z.

[15] G. Rivero, L.T.T. Nguyen, X.K.D. Hillewaere, F.E. Du Prez, One-pot thermo-remendable shape memory polyurethanes, Macromolecules (2014), https://doi.org/10.1021/ma402471c.

[16] X. Hao, H. Liu, Y. Xie, C. Fang, H. Yang, Thermal-responsive self-healing hydrogel based on hydrophobically modified chitosan and vesicle, Colloid Polym. Sci. (2013), https://doi.org/10.1007/s00396-013-2910-4.

[17] C.C. DuFort, B. Dragnea, Bio-enabled synthesis of metamaterials, Annu. Rev. Phys. Chem. (2010), https://doi.org/10.1146/annurev.physchem.012809.103300.

[18] A.R. Jones, B.J. Blaiszik, S.R. White, N.R. Sottos, Full recovery of fiber/matrix interfacial bond strength using a microencapsulated solvent-based healing system, Compos. Sci. Technol. (2013), https://doi.org/10.1016/j.compscitech.2013.02.007.

[19] J.F. Patrick, K.R. Hart, B.P. Krull, C.E. Diesendruck, J.S. Moore, S.R. White, N.R. Sottos, Continuous self-healing life cycle in vascularized structural composites, Adv. Mater. (2014), https://doi.org/10.1002/adma.201400248.

[20] A.M. Coppola, P.R. Thakre, N.R. Sottos, S.R. White, Tensile properties and damage evolution in vascular 3D woven glass/epoxy composites, Compos. Part A Appl. Sci. Manuf. (2014), https://doi.org/10.1016/j.compositesa.2013.12.006.

[21] K.S. Toohey, C.J. Hansen, J.A. Lewis, S.R. White, N.R. Sottos, Delivery of two-part self-healing chemistry via microvascular networks, Adv. Funct. Mater. 19 (2009) 1399–1405, https://doi.org/10.1002/adfm.200801824.

[22] C.E. Diesendruck, N.R. Sottos, J.S. Moore, S.R. White, Biomimetic self-healing, Angew. Chemie – Int. Ed. (2015), https://doi.org/10.1002/anie.201500484.

[23] K. Vijay, M. Murmu, S.V. Deo, Bacteria based self healing concrete – a review, Constr. Build. Mater. 152 (2017) 1008–1014, https://doi.org/10.1016/j.conbuildmat.2017.07.040.

[24] V. Sariola, A. Pena-Francesch, H. Jung, M. Çetinkaya, C. Pacheco, M. Sitti, M.C. Demirel, Segmented molecular design of self-healing proteinaceous materials, Sci. Rep. (2015), https://doi.org/10.1038/srep13482.

[25] R. Araya-Hermosilla, A.A. Broekhuis, F. Picchioni, Reversible polymer networks containing covalent and hydrogen bonding interactions, Eur. Polym. J. (2014), https://doi.org/10.1016/j.eurpolymj.2013.10.014.

[26] M.A. Darabi, A. Khosrozadeh, R. Mbeleck, Y. Liu, Q. Chang, J. Jiang, J. Cai, Q. Wang, G. Luo, M. Xing, Skin-inspired multifunctional autonomic-intrinsic conductive self-healing hydrogels with pressure sensitivity, stretchability, and 3D printability, Adv. Mater. 29 (2017) 1–8, https://doi.org/10.1002/adma.201700533.

[27] C.M. Bishop, Patterns Recognition and Machine Learning, 2006, https://doi.org/10.1016/B978-044452701-1.00059-4.

[28] Y. Liu, T. Zhao, G. Yang, W. Ju, S. Shi, The onset temperature (T g) of $As_x Se_{1-x}$ glasses transition prediction: a comparison of topological and regression analysis methods, Comput. Mater. Sci. 140 (2017) 315–321, https://doi.org/10.1016/j.commatsci.2017.09.008.

[29] Ng, CS229 Lecture notes Margins : Intuition, Http://Cs229.Stanford.Edu/Notes/Cs229-Notes3.Pdf. (2017) 1–25, http://cs229.stanford.edu/notes/cs229-notes3.pdf.

[30] L. Guadagno, M. Raimondo, C. Naddeo, P. Longo, A. Mariconda, Self-healing materials for structural applications, Polym. Eng. Sci. (2014), https://doi.org/10.1002/pen.23621.

[31] P.J. Flory, Effects of molecular structure on physical properties of butyl rubber, Ind. Eng. Chem. 38 (1946) 417–436, https://doi.org/10.5254/1.3543214.

[32] K. Gurney, An Introduction to Neural Networks, 1st ed., UCL Press Limited, London and New York, 1997, , https://doi.org/10.1016/S0140-6736(95)91746-2.

[33] D. Perera, J.W. Tucker, S. Brahmbhatt, C.J. Helal, A. Chong, W. Farrell, P. Richardson, N.W. Sach, A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow, Science 359 (6734) (2018) 429–434, https://doi.org/10.1126/science.aap9112.

[34] G.C. Cawley, N.L.C. Talbot, I. Guyon, A. Saffari, Preventing over-fitting during model selection via bayesian regularisation of the hyper-parameters, J. Mach. Learn. Res. 8 (2007) 841–861 http://www.jmlr.org/papers/volume8/cawley07a/cawley07a.pdf.

[35] M.V. Butz, Rule-Based Evolutionary Online Learning Systems: Learning Bounds, Classification, and Prediction, 2004, p. 294, , https://doi.org/10.1007/b104669.

[36] M. Opper, A bayesian approach to online learning, On-Line Learn. Neural Networks, Cambridge University Press, New York, 1997, pp. 363–378 https://dl.acm.org/citation.cfm?id=304756.

[37] H. Ohlsson, Regularization for Sparseness and Smoothness and Signal Processing, Linköping University, 2010 http://users.isy.liu.se/en/rt/ohlsson/HOthesis.pdf.

[38] C. Lin, D. Sheng, X. Liu, S. Xu, F. Ji, L. Dong, Y. Zhou, Y. Yang, NIR induced self-healing electrical conductivity polyurethane/graphene nanocomposites based on Diels−Alder reaction, Polymer (Guildf) 140 (2018) 150–157, https://doi.org/10.1016/j.polymer.2018.02.036.