REVIEW ARTICLE OPEN

A review of deep learning in the study of materials degradation

Will Nash[1,2], Tom Drummond[3] and Nick Birbilis[1,2]

Deep learning is revolutionising the way that many industries operate, providing a powerful method to interpret large quantities of data automatically and relatively quickly. Deterioration is often multi-factorial and difficult to model deterministically due to limits in measurability, or unknown variables. Deploying deep learning tools to the field of materials degradation should be a natural fit. In this paper, we review the current research into deep learning for detection, modelling and planning for material deterioration. Driving such research are factors such as budget reductions, increasing safety and increasing detection reliability. Based on the available literature, researchers are making headway, but several challenges remain, not least of which is the development of large training data sets and the computational intensity of many of these deep learning models.

INTRODUCTION

The degradation of engineered materials presents significant environmental, safety and economic risks. Modern society depends on the ongoing integrity of materials—from the reliability of aircraft to the efficacy of sanitary systems. Designers impose ever increasing demands on man-made materials that are thermodynamically driven to deteriorate.[1] For all the novel materials created in laboratories around the world, their potential degradation in service is a significant barrier to adoption. Magnesium alloys provide a salient example, promising light-weight and strong parts, but suffering from rapid corrosion rates.

Aside from the mechanistic research regarding materials degradation, research is nowadays underway that seeks to employ deep learning to understand how to detect defects, improve durability and manage the associated risks associated with materials degradation.

Background on Deep Learning

Recently, advances in Artificial Intelligence (A.I.) seem to be broadcast weekly, even daily. To a large extent the burgeoning A.I. revolution has been supported by silicon transistor technology, arguably the material technology that defines our current age. Alongside the development of cheaper more powerful Graphical Processing Units (GPUs), A.I. improvement in recent years has been driven by the collection of massive data sets via the Internet, novel learning architectures and programming languages.[2,3] A recent review by Dimiduk et al.[4] reveals that materials design and development is benefiting from Deep Learning; and quantum matter researchers using artificial neural nets have revealed previously hidden patterns in cuprate superconductor psuedogap images,[5] providing insight to fundamental questions that have gone unanswered for decades. The critical review herein intends to explore how Deep Learning methods are being used to automate the detection of degradation, improve modelling of materials durability and assist decision making by analysis of large sets of degradation data. The true power of Deep Learning arises when the computer is able to discover its own interpretation of the data, often leading to faster and more accurate predictive power than hand-crafted algorithms.

Common taxonomy

The field of A.I. is awash with apparently complicated terminology, in many cases with different descriptors having an identical meaning, (i.e. due to the pace of research there can be various names given to the same concepts); understandably this can create confusion, even to researchers in the field. The common taxonomy of terms are defined below, for a more detailed description of Deep Learning A.I. systems, readers should refer to 'Deep Learning'.[6]

Artificial Intelligence

Within this review we define A.I. to refer to machine learning models that can process data to make meaningful decisions. This definition is narrower than the traditional one, and excludes A.I. that is hard coded such as expert systems.

Artificial Neural Network

The Artificial Neural Network (ANN) was first proposed in 1958 by Rosenblatt[7] as a computer 'Perceptron' that mimics the brain. As the name suggests an ANN is made up of artificial neurons represented by an activation function, each neuron is fed inputs that are weighted and summed, once the activation threshold is exceeded the neuron 'fires', producing an output signal. The neurons are arranged in a layered network with neurons taking inputs from preceding layers, thus transforming an input signal to an output. The weights of the neurons can be tuned to adjust how they react to the inputs. Rosenblatt's original diagram of an ANN has been reproduced in Fig. 1.

Backpropagation

The measured error of a network can be passed backwards, using the chain rule of derivatives to determine the contribution of each

[1]Department of Materials Science and Engineering, Monash University, Clayton 3800 VIC, Australia; [2]Woodside Innovation Centre, Monash University, Clayton 3800 VIC, Australia and [3]Department of Electrical and Computer Systems Engineering, Monash University, Clayton 3800 VIC, Australia
Correspondence: Will Nash (Will.Nash@monash.edu)

npj

A review of deep learning in the study of materials degradation
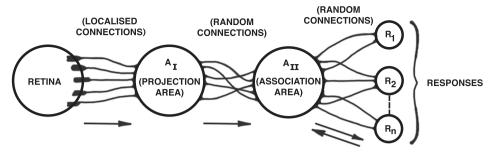W Nash et al.

2

Fig. 1 The original perceptron concept from Rosenblatt (ref. 7) [public domain]; artificial neurons mimic the function of the brain, transforming inputs at the retina into responses

weight to that error—this is termed 'backpropagation'. This method was developed independently by a number of researchers in the 1970s and 1980s. Its use in machine learning was first popularised in 1986.[8]

Data sets
The development of accurate deep learning models relies heavily on 'good quality' data sets. Any underlying biases and systemic errors that are present in data sets utilised for training can compromise the accuracy and effectiveness of deep learning. For this reason, design of data sets is a major concern for A.I. researchers and takes considerable effort. Ideally the distribution of information contained in data sets will match the distribution encountered in deployment. During training researchers typically break the data set into the following subsets: training set, used for training the model; validation set, used during training to check the accuracy of the model on 'unseen' samples; and a testing set, reserved to evaluate performance after training. Thankfully, sites like www.kaggle.com provide benchmark data sets for researchers to develop their models and compete for prizes removing the data set bottleneck and helping drive research.

Deep Learning
Until the 1990s ANNs were largely limited to three layers, comprising one input, one hidden, and one output layer. In 2009 parallelisation of ANN training using Graphical Processing Units (GPUs) was demonstrated.[9] Subsequently ANNs have been successfully extended to so-called Deep Learning models, extending to 100 s of hidden layers. It is useful to consider that each neuron in the network transforms the incoming data to a distinct output signal. As the depth of the ANN is increased the network can transform the data in more complex manners, effectively adding variables to the learned relationship between inputs and outputs.

Convolutional layers
There is a special class of neural network layer called a 'convolutional layer' that was first proposed in 1982.[10] The convolutional layers consist of neurons grouped into filters that convolve the input data to produce activated outputs. For example, if the input is an image made up of an array of red, green and blue channels, the filters scan across the image and produce an output map where the filter neurons are activated. Extending the explanation of Deep Learning above, the lower layers of convolutional neural networks close to the input have been found to detect simple features such as edges or colours, whereas the higher layers are able to use these lower level representations to interpret more complex features such as faces and text.[11] Networks that use convolutional layers are commonly called Convolutional Neural Networks (CNN) or ConvNets.

Recurrent neural network and long short-term memory
Recurrent neural nets (RNNs) are designed to process sequential data, using a connection from the output to the input of the next sequence. This network architecture is particularly suited to processing temporal data. Simple RNNs suffer from gradient instability, when the sequence of inputs grows, the gradient vanishes or 'explodes'. To overcome this issue Long Short-Term Memory (LSTM) networks were introduced,[12] and later refined.[13] LSTMs incorporate a memory cell into RNNs to store the state of the neuron, preventing the gradient instability problem.

Semantic segmentation
Within object detection researchers typically use semantic segmentation, this is a term that refers to segmenting an image into its semantic components. Semantic segmentation models produce a label for each pixel, are trained using data sets that are themselves segmented into their different objects, Fig. 2 provides an example image alongside its semantic segmentation for illustration from the Pascal VOC Dataset[14]. Accuracy of semantic segmentation is typically reported using the F1-score, the harmonic average of the precision (how many positives predicted were true) and recall (how many true positives were predicted out of labelled positives). We can examine human performance on labelling data sets to formulate a benchmark for performance— e.g. the Microsoft Common Objects in Context semantic segmentation data set expert labellers achieve an average F1-score of 0.81.[15]

Training Deep Learning models
Deep learning models are trained to be able to interpret the input data in a useful way. Simply put, models are initialised with random weights, and example inputs are fed through the network. The difference between the target labels and the model outputs is then measured as the error. The contribution of each neuron to the error is determined using backpropagation, and the weights are updated to reduce the error. This process is repeated until a set number of iterations are completed, or the error is reduced to an acceptable level, and the model adequately interprets the input data into the desired output. The whole process is termed Stochastic Gradient Descent (SGD), although there are several variants in use that employ different methods to increase the speed of converging on a solution. To set up the training phase there are several so-called hyper-parameters that affect the speed of convergence, including the number of iterations to train with, the learning rate (i.e. how large of a step to take with each iteration), and the specific calculation of the error signal. Selecting an appropriate measurement of error is important and depends on the problem space, within the literature the error is also referred to as the 'cost' and the 'loss'.
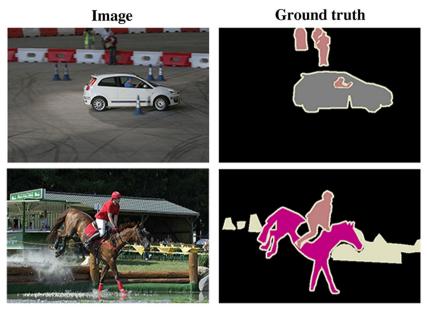
A review of deep learning in the study of materials degradation
W Nash et al.

npj

3

Fig. 2 Example Images and Ground Truth Maps illustrating semantic segmentation. Adapted by permission from Springer Customer Service Centre Gmbh (ref. [14])
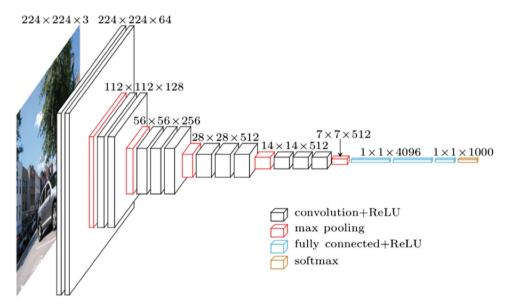

Fig. 3 An overview of the VGG-16 model architecture, this model uses simple convolutional blocks to transform the input image to a 1000 class vector representing the classes of the ILSVRC, figure reproduced from ref. [20]

Fine-tuning models

Deep Learning models can be trained on one task, and then fine-tuned on another task, otherwise known as transfer learning.[16,17] Typically, fine-tuning involves locking all the previously learned weights bar those on the output layer. Commonly this approach is used by training on a large and freely available data set, and then fine-tuning on a specific task with a smaller data set. This works for tasks in similar domains where the weights learned at lower levels are similar.

ImageNet large scale visual recognition challenge

In 2010 the annual ImageNet Large Scale Visual Recognition Challenge (ILSVRC) was launched and has become the benchmark for object detection and classification computer vision models.[18] The ILSVRC provides a data set of 1.4 M labelled images in 1000 classes for competitors to develop and train models for ~4 months. Models are assessed on a reserved data set where labels are only known to the organisers, and scored based on the number of accurate predictions.

VGG-16

In 2014 the Visual Geometry Group from Oxford University placed second in the ILSVRC for classification using a very deep but simple convolutional neural network architecture that has come to be known as VGG-16.[19] This model has become very popular in the research community due to its simple approach and because the pre-trained weights were made freely available online, facilitating the fine-tuning of this powerful model on new tasks. Several of the papers reviewed make use of this model, and so its network architecture is provided in Fig. 3.[20]

npj

A review of deep learning in the study of materials degradation
W Nash et al.

4

Fig. 4 Results from deep learning semantic segmentation of railway ties—pink segmentation indicates crumbled concrete, red segmentation indicates chipped concrete. © 2017 IEEE. Reprinted, with permission, from ref. 24

DEEP LEARNING FOR DETECTION OF DEGRADATION

Detection of degradation is necessary to allow intervention prior to failure; undetected deterioration can lead to catastrophic failure in extreme cases. Direct detection involves measuring change in materials that are detectable in ambient conditions, for example visual presence of corrosion products, cracks and changes in dimensions. Indirect detection requires the application of an excitation signal for which the response of the material can be measured to indicate deterioration, for example ultrasonic thickness testing may reveal loss of wall thickness in pipes. Both direct and indirect detection methods are used throughout industry, and provide complementary functions. Typically, direct detection is used to focus indirect detection efforts to areas of distress.

Direct detection of degradation

Research by the European project MINOAS (Marine INspection rObotic Assistant System) has demonstrated the effectiveness of a simple Artificial Neural Network (ANN) for corrosion and crack detection using a micro-aerial vehicle in ships ballast tanks.[21] Using traditional computer vision techniques to produce inputs related to colour and texture to various ANNs comprising one hidden layer. The analysis determined that the optimum configuration consisted of 34 inputs and 37 neurons, achieving accuracies of 74 to 87%. This hybrid computer vision + ANN approach may be necessary with shallow networks that are unable to learn to discern higher order features, such as texture. Colour information was provided to the network by filtering hue and saturation values; and texture information by processing the distribution of neighbouring pixel intensity. Thus, the approach does not exploit the true power of deep learning, i.e. allowing the computer to determine the best representation of the input data to achieve the task. It is likely that these models overfit the limited training data of ship ballasts, although this is appropriate for the task at hand, transferring this approach to other environments and subjects may require significant rework.

Deep learning models and traditional computer vision systems for corrosion detection were compared in 2016.[22] The deep learning architecture utilised transfer learning of the AlexNet model architecture that won the ImageNet competition in 2012[3]—thus the model was pre-trained to identify low level features like edges. The AlexNet model incorporates five convolutional layers, and consists of ~650,000 neurons. Even with a small data set of 3,500 images it was demonstrated that Deep Learning outperforms computer vision with total accuracies of 78% and 69%, respectively. Unfortunately, neither accuracy would be considered equivalent to human performance, measured as 88–95% when tested on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC).[17] The authors posit that a computer vision system could augment Deep Learning to improve classification accuracy further. The model also requires images to be downsized to 256 × 256 pixels, discarding some (perhaps significant) available information. Image classification for the presence of corrosion at the accuracy achieved would however still require humans to review nearly all the data captured.

Deep Learning Fully Convolutional Networks (DLFCNs) have been trained to detect degradation of railway ties[23] and fasteners[24] from greyscale images. In such work, a four-layer material classification network with 493,226 trainable weights and biases was able to discern crumbled and chipped concrete from good concrete, as well as other materials such as ballast, rail and fasteners, with an accuracy of 95.02%. A classifier uses the output of the material detector as its input and has been trained to identify five types of fasteners and whether they are broken or not. Example detections from the model are presented in Fig. 4. Learning to detect defects in railway ties and fasteners benefits greatly from the nature of data capture—the position of the camera is fixed with respect to the subject, and therefore the images are well controlled. Even so, the authors were required to manipulate their data set to enable good training: applying a global gain normalisation, preferentially training on good quality images, and resampling data to balance the data set to include difficult images. Furthermore, the alignment of images was strictly controlled to avoid intra-class variation, necessitating additional annotation by an individual researcher to frame the region of interest inside 'bounding boxes'; this extra constraint complicates and effectively prevents outsourcing of data set creation.

A CNN has been successfully used for classification of cracked and un-cracked pavement regions,[25] the approach was able to

A review of deep learning in the study of materials degradation
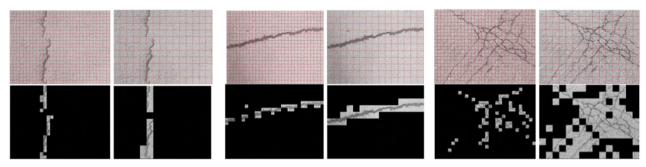W Nash et al.

npj

5

Fig. 5 Crack detection and classification using grid tiling coupled with a Convolutional Neural Network, from Wang and Hu, 2017. © 2017 IEEE. Reprinted, with permission, from ref. 25

achieve greater than 90% classification accuracy. This model of Wang and Hu[25] took varying image sizes and divided them into the input grid size, thus providing a quasi-localisation function, an example output from the model is presented in Fig. 5. However, it is unclear how well regulated the feature size to grid size needs to be for the model to operate accurately. The pre-processing required introduces more complexity than the more straightforward semantic segmentation techniques. Images are also downsized and made greyscale, reducing the information available for the model.

CrackNet was developed using a CNN to detect cracks in 3D images of asphalt,[26] with a reported precision of over 90%. By foregoing max-pooling layers (i.e. layers that down-size resolution) the CrackNet model was able to preserve the dimensionality of the input and produce pixel level segmentation. A specialised hand-coded feature extractor was used to feed data into the model, which somewhat limited robustness, and was attributed to false negatives detected for the case of hairline cracks. Although the technique required a specialised PaveVison3D scanning camera, it is expected that similar results can be achieved with standard 2D scanning cameras.

Two deep neural nets were trained to detect corrosion,[27] based on ZF Net, VGG-16, and two smaller CNNs of 5 and 7 layers; where ZF Net and VGG-16 are freely available models that ranked highly in the ILSVRC. A so-called sliding window was used to scan across the images and provide localisation of features. Different window size and input image colour formats were investigated, as well as the impact of fine-tuning after training on the ILSVRC data set against training end to end on a corrosion data set. Increasing the window size improved the model's success rate, at the cost of decreased fineness of detection. Accuracy was similar for the RGB and YCbCr colour formats, which is anticipated as they contain simple transforms of the same data. Overall however, and a recurring outcome in the literature, is that the CNNs demonstrate superior accuracy to traditional computer vision filtering techniques in feature detection. An important finding from this work is that the networks trained end to end on corrosion learned colour and texture based filters, but suffered from overfitting; whereas the fine-tuned models provide more general representations at a significant computational cost. It is not clear that the scope of training images included features that may confuse a CNN, and the example images demonstrate false positives when the model is faced with gravel, presumably due to the similar texture to corrosion. A robust A.I. detection of corrosion is likely to require contextual clues that are provided by detecting other features such as structural steel frames versus foliage that may be confused when relying on colour and texture alone. The drawbacks of fineness of segmentation, overfitting and computational demands remain to be overcome for an automated A.I. corrosion detector.

The 'Faster-R-CNN' model, recently presented in the work of Cha and co-workers,[28] fuses a region of interest detector with an object detector, and has been trained to detect a variety of infrastructure defects, including cracked concrete, bolt corrosion, steel delamination, and general corrosion. This method produces region of interest bounding boxes around the detected defect in real time on a video resolution of 500×375 pixels. Fusing a region of interest with the object detector is intuitively similar to how human vision operates, focusing on the important features. Although reporting an impressive average precision of 87.8% the data set was limited to two bridges and a building at the University of Manitoba. Unfortunately, the subjects used for performance evaluation were not fully described, and it appears to be applied to a restricted domain—therefore, it is not clear whether the model will perform reliably in other environments. Additional fineness of detection is likely to be improved by replacing the object detector with a semantic segmentation method. Unequivocally however, such a model shows both the value of CNN models, and the possibilities of automated defect detection removing the need for difficult site access and personnel.

Returning to the domain of ship ballast tanks, another 'Faster-RCNN' based on VGG19 has been trained to segment Coating Breakdown and Corrosion (CBC) in natural colour images.[29] The model detects four classes of defects: CBC on edges, CBC on welds, surface corrosion (termed 'hard rust') and pitting; example output from the model is presented in Fig. 6. Accuracy is reported to vary from 45 to 95%, although this performance is distorted by the data set bias toward background class (no CBC), representing 40% of pixel labels in the ground truths. Excluding the background class the F1-score is calculated to be 0.69, approaching individual human performance of 0.81 in the MS-COCO semantic segmentation task.[15] As the training data set increases in size we should expect the F1-score to reach and even exceed individual human level performance.

Aircraft fuselage inspection is required frequently, typically a visual inspection is undertaken between each flight. An automated system has been developed that utilised a deep learning convolutional neural network, based on the VGG-16 model pre-trained on ILSVRC and fine-tuned on fuselage images.[30] This fuselage inspection system coupled the CNN with a traditional computer vision feature detector (called SURF), which locates areas of change in the image; increasing inspection speed up to 6 times. A Gaussian filter was also used to smooth the image, reducing false positives occurring from dirt on an unwashed plane. Images were divided into 64×64 pixel patches, which the model classified as defect/no defect. Accuracy of more than 96% was measured for new unseen images, with an average run time of 15.78 sec. This system could be improved by providing pixel level segmentation, and reducing the run time, which may be achieved if pre-processing is reduced.

Recently researchers have demonstrated an ability to perform semantic segmentation of video at 12 frames-per-second on a resolution of 1024×1280 pixels.[31] The model was trained to segment coating, water, rivet, wet or corroded surfaces of

npj

A review of deep learning in the study of materials degradation
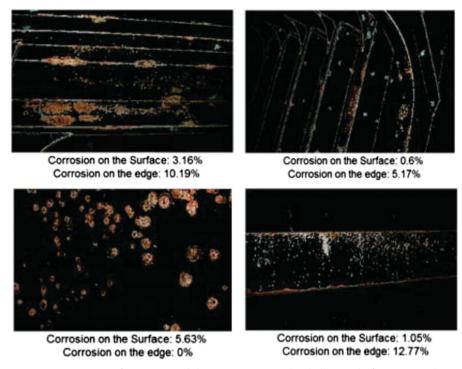W Nash et al.

6

Fig. 6 Model output Semantic Segmentation of Coating Breakdown Corrosion in ship ballast tanks from Liu et. al. © 2018 IEEE. Reprinted, with permission, from ref. 29

penstocks. In order to compensate for the imbalance of relatively small data set of 40 images the authors weight the loss function to focus on the less common classes. This technique achieved an F-score of 52.5%, and although the performance achieved falls short of human level accuracy, with an increased data set size this can be expected to improve, particularly within this restricted domain. The work was extended to produce a 3D volume rendering of the penstock, which is useful for automating inspection of inaccessible assets.

Combining visual and infrared imaging was shown to permit a CNN to detect concrete cracks smaller than 0.5 mm.[32] The additional information from the infrared camera improved the F1-score from 0.45 to 0.99, approaching 'near perfect' crack detection. This method relies on a laser based excitation unit to provide a signal for the infrared detector. If this system can be successfully transferred from the laboratory to site it would replace the tedious task of crack mapping. This approach of fusing different detection methods shows great potential to be extended to other domains.

Indirect detection of degradation
Indirect detection of degradation aims to identify signals of change that arise as a result of material deterioration. Typically, these methods measure the response to an energy source, either imparted by inspectors as in the case of microwave thermal imaging of composites, or from operating conditions as commonly used for vibration analysis. Deep Learning is well suited to seek out the signs of deterioration hidden in the enormous amount of data generated by these indirect methods.

Cracks in welds
Deep neural networks have been trained to detect flaws in welds from radiographic scans[33] in an automated process presented by Hou et. al. that achieved a maximum of 91.84% classification accuracy. The training process uses unlabelled data to pre-train Stacked Sparse Autoencoders, before fine-tuning on labelled data;

reducing the need for a very large data set. This approach and similar should be considered by researchers with difficult to produce data sets. A sliding window is used to provide location of the defects, a convolutional architecture could reduce the computational complexity and provide pixel level labelling. Extensive data pre-processing was implemented to train on a limited data set of 88 scans, that would then be required to be undertaken on new scans; placing a (minor) bottleneck on the process.

Carbon fibre reinforced polymer composites
The astounding performance of carbon fibre reinforced polymer (CFRP) composites can be undermined by subsurface flaws that grow in service hidden from view. In-service integrity checks are commonly performed using ultrasonic non-destructive testing. Interpretation of the ultrasonic signal requires expert knowledge from experienced inspectors. In order to provide faster and reliable inspection a deep learning Convolutional Neural Network was trained on the ultrasonic wavelet packet decomposition signal to detect flaws deliberately introduced to the composite samples.[34] The two layer CNN was able to detect flaws with 95% classification accuracy. Post processing of the ultrasonic mapping denoised the detection by removing defects with different class neighbouring areas, thus introducing a lower limit for detectable defect size. Just ten defective CFRP composite samples were used to produce the data set, increasing the likelihood that the CNN presented is overfit. The authors also do not provide a measure of the speed of detection, a significant factor in deployment.

Aircraft fuselage composites
An early attempt by the US National Aeronautics and Space Administration (NASA) to automate detection of corrosion in aluminium composites used simple neural networks to analyse thermal data.[35] Two ANNs were trained to detect flaws and extent of corrosion from the thermal response of composite panels subjected to quartz lamp heating. The flaw detector model was

A review of deep learning in the study of materials degradation
W Nash et al.

npj

7

binary, while the corrosion detector was binned into 10 percentile ranges. By averaging the training data over several frames of the imaging the signal to noise ratio was effectively increased. The research went on to compare the performance of the individual models against a combined architecture and demonstrated that this combined architecture provided superior performance. Unfortunately, the authors did not provide details of the accuracy of this approach according to any metrics.

Aluminium plate
Giant Magnetoresistive (GMR) sensing data has been used as an input to a simple neural network to detect defects in aluminium plate.[36] The method successfully identified cracks, holes and deformation using Eddy current testing, with the aim of producing a low-cost, fast and robust defect detection sensor array. The network architecture utilised is described as a multilayer perceptron (1 input, 1 hidden and 1 output layer) followed by a competitive neural network of one layer. This competitive neural network effectively performs the softmax function on the output. The authors did not provide details on the size of their data set, although data setit appears to be small. The classification accuracy for holes and cracks is reported as 83 and 95%, respectively. The GMR sensor ANN has a heavy reliance on filtering and feature extraction prior to input to the neural network, thus the method does not leverage the ability of deep neural networks to extract relevant features, and makes it difficult to retrain the network on other sensor geometries or material types. The small number of machined defects used to generate the data set raises issues of overfitting. No field performance evaluation was undertaken, and demonstrating the efficacy of the method on aluminium plate with unknown dimensions would be necessary before the technique could be deployed.

Stainless steel coupons
The onset of pitting and crevice corrosion in stainless steels was shown to be able to be predicted from electrochemical data using a simple ANN in 1993.[37] The ANN was trained for 30,000 iterations on a data set of 50 files (based on potentiodynamic scans of 304 stainless steel, presumably in chloride containing electrolyte), after which it achieved a 90% accuracy at identifying the initiation localised corrosion. This approach shows promise to detect insidious forms of corrosion using potential monitoring that are not easily observed otherwise. Although at first the method of assigning a pitting or crevice corrosion initiation event to the electrochemical data is straightforward enough that an algorithm may be a better choice than a neural network, where it was stated that '*The start of the corrosion event was arbitrarily assigned as the first of at least three data points above the mean baseline plus 2 standard deviations calculated from baseline noise*'. The authors contend that the neural network identified corrosion earlier than simple current limit monitoring, and can distinguish between pitting and crevice corrosion. A sensor based on this technology may be able to detect the onset of pitting through measuring potential changes of a coupon due to chemical excursion events in processing industries. This work could be revisited using Deep Neural Networks to improve the prediction capability.

Steel pipelines for subsea oil transmission
Simple neural networks were effective at processing multiple NDT sensor inputs to predict the degree of oil pipeline corrosion under laboratory conditions.[38] Ultrasonic and magnetic flux leakage sensors were used to collect a data set from machined defects in steel pipes. This work showed great promise to automate the detection of corrosion on subsea pipelines, an ongoing concern in the oil and gas industries. It's unclear if field testing has validated this approach, and there would be questions of overfitting from the small data set. The ANN training methods used in this work have since fallen out of favour, this is another example where modern deep neural networks could yield accuracy improvements, if the data set were made available and ideally enlarged.

Steel transmission tower footings
Prediction of corrosion of electrical transmission tower footings[39] was achieved using a basic neural network. The network consisted of 6 input neurons, 5 hidden neurons and one output neuron that estimates the degree of corrosion on a 0–50 scale. Input data consisted of close and remote soil resistivity, corrosion potential, polarization and noise resistance. The reported accuracy of 0.999 shows that the ANN method is able to learn very well when clear correlations exist even with small, shallow networks. Sensitivity analysis of the model to the inputs revealed that the corrosion potential is the most influential in determining the degree of corrosion.

Concrete reinforcement steel
A machine learning approach was used to predict the linear polarization resistance (nominally determined by electrochemical testing) of reinforcing steel in concrete without requiring destructive breakout.[40] NDT measurements of concrete resistivity, galvanostatic resistivity and air temperature were provided as inputs to the simple ANNs. The author investigated the network architecture to find the best arrangement neurons, which provided R-squared accuracies above 95% on the testing data. A tool to measure reinforcing steel corrosion without breakout would prove very useful, while no information on in-field performance was reported the work is extremely promising.

ANNs have also been trained to interpret ElectroMagnetic Anomaly Detection (EMAD) of reinforcing steel in concrete.[41] EMAD is a non-destructive technique developed in 2009[42] that magnetizes the reinforcement via electromagnetic induction, and can detect defects from Magnetic Flux Leakage (MFL) sensors. In real world performance testing, recurrent network architectures were shown to provide the best predictive accuracy due to the time-dependence of the EMAD signal. Recently developed 'attention-based' neural networks may be able to further improve accuracy for EMAD.

For reinforced concrete bridge monitoring, research using vibration sensors as inputs to CNNs has shown the capability to learn features that correspond with vibration mode.[43] The networks were trained on simulated data for simple beams, but extending the technique to real bridges with complex mixed modes appears straightforward. Automating the deployment and tuning of these sensors should increase their use thanks to savings in time and costs. The accuracy and speed achieved by researchers is suitable for field deployment on simple bridges, although small defects are undetected in the presence of noise. The authors recognise that obtaining unbiased data is difficult because bridges are generally maintained in good condition, thus the data set needs to be augmented by robust numerical simulations.

Machine health monitoring
Machine condition monitoring typically involves measuring vibration to detect faults in bearings and rotating parts. Once again, deep learning methods are well suited to identify fault signals from copious amounts of data. A comprehensive survey of DL research into machine health monitoring was undertaken by Zhao et. al in 2016 [44]; a handful of salient examples are reviewed below.

CNNs were trained to identify bearing faults with 93.61% accuracy.[45] Fifty minutes of vibration data from eight different bearing fault conditions was used for training. Feature extraction

npj

A review of deep learning in the study of materials degradation
W Nash et al.

8

segmented the data into one minute windows, which were fed into two neural nets, the first classified the machine state as balanced/unbalanced, and the second classified the bearing fault type. Machine learning improved accuracy by ~6.4% compared to hand-crafted features. Similar CNNs trained for classification of gearbox faults outperformed manual feature extraction by roughly 10%.[46]

Using CNNs to analyse temporal data requires pre-processing into discrete time-windows, this dictates fineness of detection. To overcome this limitation, researchers have turned to Recurrent Neural Network (RNN) methods, in particular using Long Short-Term Memory (LSTM) networks. LSTM models incorporate a hidden state that acts as a memory of previous inputs, providing an advantage when interpreting time series data such as machine health.

LSTMs have been trained to predict CNC machine tool wear from vibration and cutting force data.[47] This research indicated that Deep LSTMs outperform basic LSTMs, RNN, MLP and traditional regression models. Follow-up research trained a novel Convolutional Bi-Directional LSTM (CBLSTM) network to monitor machine health.[48] The CBLSTM extracted features from the data using a CNN, these features were then analysed by two bidirectional LSTMs both forward and backward in time. The CBLSTM accuracy outperformed the compared state-of-the-art methods across all data sets, achieving a root mean square error of ~10 compared to offline tool wear measurement. The work presented is promising because it works on raw data, and is able to analyse time series data continuously. However, the test set up is problematic, and it is anticipated that health monitoring would need tuning for each individual machine. It appears that the deep LSTM models sometimes produce significant error excursions, which may trigger improper maintenance decisions. Furthermore, the models produce errors that show a reverse of wear, which is not logically consistent.

An alternative approach for machine health monitoring was proposed by Jia et. al.[49] The authors designed a Normalised Sparse Auto-Encoder network coupled with a Local Connection Network (NSAE-LCN) to predict planetary gearbox health from vibration data. A data set of 4,000 samples over ten-classes was used to train the network to detect machine faults from vibration accelerometers. The authors posited that the NSAE component enabled them to automatically find relevant features from the inputs, while the LCN component ensured that the features identified are independent and shift-invariant. Although it seems that this method is a complex implementation of a straightforward deep network, the advantage is that the features learned are able to be directly extracted from the NSAE. Impressively, the NSAE-LCN achieved greater than 99.9% accuracy of classification on the testing set.

DEEP LEARNING FOR DEGRADATION MODELLING AND FORECASTING

Moving from detection of existing defects to prediction of deterioration is important for managing critical assets and forecasting budgets. Deep learning methods have the potential to improve prediction of materials deterioration, especially where the interaction of variables is not empirically understood, and there is significant uncertainty of variables to the extent that many variables may remain unknown.

Remaining useful life of aero-engines

The US National Aeronautics and Space Administration (NASA) has developed the C-MAPSS aero-engine simulator that has been used to produce a Remaining Useful Life (RUL) data set. The RUL data set provides time or cycles to failure labels based on 21 input channels from temperature and pressure sensors. This data set has

Table 1. A comparison of the accuracy (root mean squared accuracy) of various Deep Learning methods to predict remaining useful life from the NASA C-MAPSS data sets

Model architecture	Reference	Mean RMSE for C-MAPSS
Deep Convolutional Neural Network	50	17.73
Vanilla LSTM	53	19.76
Deep Belief Network (MODBNE)	54	20.32
Deep Convolutional Neural Network	51	24.43

been used to train and evaluate predictive models based on Deep Convolutional Neural Network, DCNN,[50,51] Long Short-Term Memory, LTSM,[52,53] and Deep Belief Networks.[54] The mean Root Mean Squared Accuracy (RMSE) across the four C-MAPSS data sets is presented in Table 1.

The best accuracy achieved on the RUL data set to date used a DCNN model.[50] In order to input the time series data into a convolutional network the multi-sensor vectors were concatenated into 2D arrays. Using the DCNN in this way reduced the computational demands compared to an LSTM, however, it required the model to observe a limited time window. Example outputs of the DCNN are presented in Fig. 7.

Inspection of the data in Fig. 7 indicates that the DCNN fails to model the underlying phenomenon driving deterioration of the aero-engine, the prediction is not similar to the expected behaviour, and occasionally the RUL prediction increases with increasing cycles. This behaviour may present frustration to engineers entrusting the prediction to plan maintenance— although the authors posit that accuracy at the critical end of life stage is adequate for decision making.

A major drawback of the DCNN method is that a limited time window has to be selected for analysis, this discards historical information that may be indicative of premature failures. Although LSTM achieved a lower accuracy, the memory gate allows it to make decision on all the historical information, at the cost of increased computational demand. Presumably with more training and tuning of hyper-parameters an LSTM could match or exceed the accuracy of the DCNN. An output from the Vanilla LSTM is presented in Fig. 8 where the general form appears to more closely match the underlying degradation of the engine, although again there are instances of increasing RUL prediction with increasing cycles. Interestingly, the bulk of the error occurs prior to the decline in RUL, and the model detects an anomaly that presumably causes the subsequent deterioration. Although these methods show promise for predicting failures from NASA's C-MAPSS data set, new applications would require obtaining a data set by running the subjects to failure.

Lithium-ion battery remaining cycles

The remaining useful life of lithium-ion batteries[55] has been predicted using a deep LSTM network with two hidden layers. This network is able to provide early failure warning—enabling users to switch batteries prior to insufficient charge availability. Only one battery lifetime running at 25 °C was used to train the network, and it is likely that this prediction model is overfit. To develop a more generalized model more training data on battery lifetimes over various temperature and other operating conditions needs to be captured and utilised.
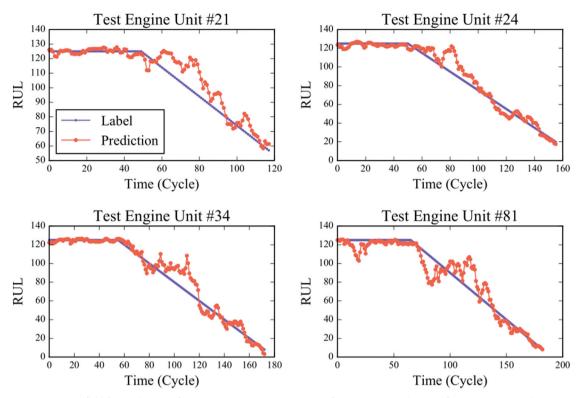
A review of deep learning in the study of materials degradation
W Nash et al.

np|

9

Fig. 7 DCNN remaining useful life predictions from the NASA C-MAPPS RUL of aero-engines data set from Li, Ding, and Sun. Reprinted from ref. 50 Copyright (2018), with permission from Elsevier
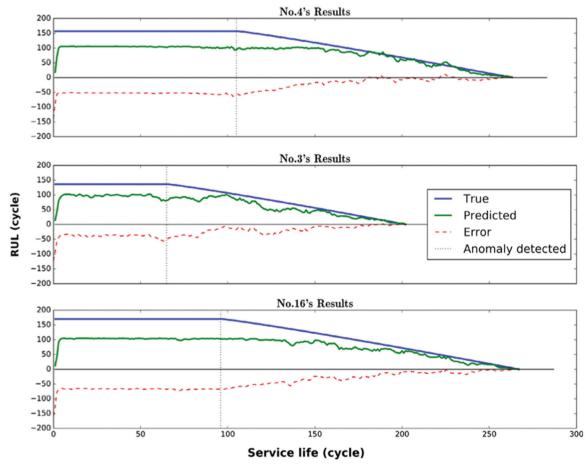

Fig. 8 RUL prediction from NASA C-MAPSS data set using Vanilla LSTM. Reprinted from ref. 53 Copyright (2018), with permission from Elsevier

npj

A review of deep learning in the study of materials degradation
W Nash et al.

10

DEEP LEARNING FOR DECISION MAKING

Beyond individual asset deterioration we can envision deep learning systems providing decision support for managing large infrastructure portfolios with complex interdependencies. Decision support systems based on so-called 'big data' rely on collecting vast amounts of disparate sensor measurements to monitor and forecast system health. Deep learning A.I. has the capability to assess the quantity and complexity of this heterogeneous and unstructured data in real time. Although we are not aware of any successful implementations, the potential of deep learning for interpreting big data has been explored.[56] Multiple deep learning architectures were evaluated for suitability to handle the challenges of big data analytics, including volume of data, speed of processing and low data quality.

While not strictly materials degradation modelling, deep learning has been used to model risk management of the San Jose-Mountain View transportation network in the event of the extreme natural disaster of an earthquake.[57] The simulation data set illustrates the capability of deep learning to process interdependencies of assets, and could feasibly be coupled with health monitoring sensors to provide city planners with a forecast risk profile.

Several challenges remain for deploying deep learning decision support systems. Not least amongst these is the task specific nature of deep learning models, that require training or tuning, as well as the increasing computing complexity with increasing inputs. One final issue is the case when unknown factors are driving deterioration, if these aren't captured by sensors then the computer is as blind to their influence as human operators.

DISCUSSION OF DEEP LEARNING METHODS CHALLENGES AND FUTURE DIRECTION

A summary of the methods reviewed herein is presented in Table 2.

The challenge for deploying deep learning to materials degradation has less to do with computing power and model architecture, and more to do with lack of 'good quality' training data. This latter point relates to everything from a lack of useful (or available) collected data, to appropriately (or expertly) labelled data. It is instructive that the areas where machine learning is making great strides are supported by freely available large data sets: KITTI for self-driving cars,[58] ImageNet[18] and MS-COCO[15] for object detection and BRATS for brain tumours[59] to name but a few. Recognising the social and economic costs of materials degradation, the creation of these data sets to drive innovation in the field may be an appropriate undertaking for public institutions. A discrete effort at addressing this point has been recently made via a web based corrosion detection resource called corrosiondetector.com, although, significant data sets that are free and readily available are in need.

Nonetheless, the transformative nature of deep learning in many related fields is illuminating for researchers in materials degradation. Just as A.I. is becoming adept at detecting 'deterioration' of the human body within the medical imaging field, we are beginning to see these advances for our built infrastructure. In particular vibration analysis,[49] detection of railway defects,[24] and corrosion of ship ballast tanks[29] have been successfully demonstrated using deep learning.

Materials degradation researchers that are interested to deploy deep learning would benefit from standardising data collection

Table 2.	Deep Learning methods for degradation that have been reviewed and their applications		
Reference	Network	Application	Input
21	Hybrid + ANN	Corrosion segmentation of ship ballast tanks	Natural colour images
22	CNN (AlexNet)	Corrosion classification	Natural colour images
23,24	FCN	Degradation segmentation of Railway ties and fasteners	Natural colour images
25	CNN	Pavement crack classification	Greyscale images
26	CNN (CrackNet)	Asphalt crack segmentation	3D scans (PaveVision3D)
27	CNN	Corrosion detection and localization	Natural colour images
28	Faster-RCNN	Defect detection and localization on Infrastructure	Natural colour video
29	Faster-RCNN	Coating defects and corrosion on steel	Natural colour images
30	CNN (VGG-16)	Aircraft fuselage defects	Natural colour images
31	U-Net	Defect segmentation of penstocks	Natural colour images
32	CNN	Concrete crack detection	Natural colour and infrared images
33	Deep NN	Flaw detection in welds	Radiographic scans
34	CNN	Flaw detection in carbon fibre reinforced polymer composites	Ultrasonic scans
35	ANN	Flaw detection in aluminium composites	Thermal response
36	ANN	Hole and crack detector in aluminium plate	Eddy current (giant magnetoresistive sensing)
37	ANN	Pitting and crevice corrosion in stainless steels	Electrochemical data
38	ANN	Corrosion of steel pipes	Magnetic flux leakage
39	ANN	Corrosion of transmission tower footings	Soil resistivity, corrosion potential, polarization and noise resistance
40	ANN	Corrosion of concrete reinforcement	Concrete resistivity, galvanostatic resistivity and air temperature
41	ANN	Defects of concrete reinforcement	Magnetic flux leakage
43	CNN	Damage detection of bridges	Vibration sensors (simulation)
45	CNN	Bearing faults	Vibration sensors
47	LSTM	Tool wear	Vibration and cutting force
50,51	CNN	Aero-engine remaining useful life	Pressure, temperature and vibration sensors
52,53	LSTM	Aero-engine remaining useful life	Pressure, temperature and vibration sensors
54	Deep Belief Network	Aero-engine remaining useful life	Pressure, temperature and vibration sensors
55	LSTM	Lithium-ion batteries	Battery capacity (Ah)

A review of deep learning in the study of materials degradation
W Nash et al.

npj

11

where possible, so that data sets can be effectively built up from multiple published sources. Providing as much information as possible about the set-up of experiments will also aid machine learning, as well as publishing results in a digital format. As much as possible, deep learning will benefit from having raw data available, that represents the expected distribution of data that will be seen in the field. To leverage the power of deep learning, models should be free to find the representation that best fits the data—which is a mental barrier for some researchers, where allowing a computer to determine mechanistic trends is (or was) considered anathema to basic science. Researchers using Deep Learning methods should adhere to established data practices, most importantly splitting data sets into training, validation and test sets; and reporting standard metrics on the reserved test set that the model has not seen prior.

Finally, within this review we have not reported on the speed of the models, largely due to the difficulty in comparing models developed on different hardware and with different objectives. It must be noted that where models are to be deployed into real world environments researchers need to strive to reduce the speed of prediction to run on available hardware—in many cases the hardware for training of models is required to be vastly more powerful than that available in the field.

CONCLUSIONS

- Deep learning has produced some very impressive results for detecting materials degradation to date, in particular the use of DLFCN for railway tie defect detection,[24] achieved greater than 95% accuracy. Furthermore, examples that employed CNNs indicated the possibility of automated defect detection of infrastructure[28] and aircraft fuselage,[30] making it clear that autonomous A.I. has merit in detecting materials degradation, with ramifications in the future role of personnel and access.
- In the case of deep learning tools applied to indirect detection of degradation; it was shown that the most promise to date was in machine health monitoring via vibration sensors— where an accuracy of 99% was achieved on the testing set.[49]
- When forecasting degradation, preliminary results summarised herein are promising, but are based on small data sets, and exhibit errors that indicate that the models do not satisfactorily reflect the underlying degradation phenomenon. The latter is anticipated in the case of incomplete learning, and highlights that deep learning is a method that relies on 'learning' as opposed to mechanistic 'hard coding'. The critical review herein has identified that models with large training data sets are likely to outperform those with small data sets, and that large data sets are to date, not widely available.
- Many of the deep learning models investigated for materials degradation have incorporated traditional hard-coded algorithms to filter and transform input data—this approach does not fully leverage the power of deep learning to learn its own representations, and limits the ability to deploy models to different problems. This conclusion also highlights a reluctance of researchers to 'let go' of mechanistic rules, which will allow the deep learning models to learn their own weightings of relevance.
- There are limited data sets publicly available for deep learning of degradation; forcing most researchers to develop their own data sets. Producing large and high-quality data sets is resource intensive, especially if it requires running multiple assets to failure. The relatively small data sets produced for training tend to show evidence of overfitting based on the works reviewed herein.
- Several of the examples presented utilising simple ANNs or outdated methods could be revisited using modern deep

learning models to yield improvements in accuracy; in particular for indirect detection.[36–38,41]

Where industry is interested in driving research into degradation it is suggested that the 'competition' formula is followed, such as the ImageNet Large Scale Visual Recognition Challenge,[18] where data sets are created and made publicly available, and subsequently some incentive is awarded for the best performing models.

REFERENCES
1. Gin, S., Dillmann, P. & Birbilis, N. Material degradation foreseen in the very long term: The case of glasses and ferrous metals. *npj Mater. Degrad.* **1**, 10 (2017).
2. Schmidhuber, J. Deep Learning in neural networks: An overview. *Neural Netw.* **61**, 85–117 (2015).
3. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. in *Proceedings of the 25th International Conference on Neural Information Processing Systems* 1–9 (IEEE, New Jersey, 2012).
4. Dimiduk, D. M., Holm, E. A. & Niezgoda, S. R. Perspectives on the impact of machine learning, deep learning, and artificial intelligence on materials, processes, and structures engineering. *Integr. Mater. Manuf. Innov.* **7**, 157–172 (2018).
5. Zhang, Y. et al. Using machine learning for scientific discovery in electronic quantum matter visualization experiments. arXiv pre-print at http://arxiv.org/abs/1808.00479 (2018).
6. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
7. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **65**, 386–408 (1958).
8. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagation errors. *Nature* **323**, 533–536 (1986).
9. Raina, R., Madhavan, A. & Ng, A. Y. Large-scale Deep Unsupervised Learning using Graphics Processors. in *Proc of the 26th International Conference on Machine Learning* (IEEE, New Jersey, 2009).
10. Fukushima, K. & Miyake, S. Neocognitron: a new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern Recognit.* **15**, 455–469 (1982).
11. Yosinski, J., Clune, J., Nguyen, A., Fuchs, T. & Lipson, H. Understanding Neural Networks Through Deep Visualization. In *Proc. Deep Learning Workshop, 31st International Conference on Machine Learning* 12 (2015).
12. Hochreiter, S. & Urgen Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
13. Gers, F. A., Schmidhuber, J. & Cummins, F. Learning to forget: continual prediction with LSTM. *Neural Comput.* **12**, 2451–2471 (2000).
14. Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J. & Zisserman, A. The Pascal Visual Object Classes (VOC) challenge. *Int. J. Comput. Vis.* **88**, 303–338 (2010).
15. Lin TY. et al. Microsoft COCO: Common Objects in Context. (eds Fleet D., Pajdla T., Schiele B., Tuytelaars T.) In *Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science* Vol 8693. (Springer, New York, 2014).
16. Rich Caruana (School of Computer Science Carnegie Mellon University). Learning Many Related Tasks at the Same Time With Backpropagation. in *NIPS'94 Proc of the 7th International Conference on Neural Information Processing Systems* 657–664 (MIT Press Cambridge, MA, 1994).
17. Bengio, Y. Deep learning of representations for unsupervised and transfer learning. *JMLR Work. Conf. Proc.* **7**, 1–20 (2011).
18. Russakovsky, O. et al. ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**, 211–252 (2015).

npj

A review of deep learning in the study of materials degradation
W Nash et al.

12

19. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. in ICLR 2015 1–14 (2014). https://doi.org/10.1016/j.infsof.2008.09.005

20. Blier, L. A brief report of the heuritech deep learning meetup#5. (2016). Available at: https://blog.heuritech.com/2016/02/29/a-brief-report-of-the-heuritech-deep-learning-meetup-5/. (Accessed 21 June 2018).

21. Ortiz, A., Bonnin-Pascual, F., Garcia-Fidalgo, E. & Company, J. P. Visual Inspection of Vessels by Means of a Micro-Aerial Vehicle: An Artificial Neural Network Approach for Corrosion Detection. in Proc. Robot 2015: Second Iberian Robotics Conference 418, 543–555 (Springer International Publishing, New York, 2016).

22. Petricca, L., Moss, T., Figueroa, G. & Broen, S. Corrosion Detection using A.I.: A Comparison of Standard Computer Vision Techniques and Deep Learning Model. In Proc. The Sixth International Conference on Computer Science, Engineering and Information Technology 91–99 (IEEE, New Jersey, 2016).

23. Gibert, X., Patel, V. M. & Chellappa, R. Material Classification and Semantic Segmentation of Railway Track Images with Deep Convolutional Neural Networks. in Proc. IEEE International Conference on Image Processing (IEEE, New Jersey, 2015). 10.1109

24. Gibert, X., Patel, V. M. & Chellappa, R. Deep multitask learning for railway track inspection. IEEE Trans. Intell. Transp. Syst. 18, 153–164 (2017).

25. Wang, X. & Hu, Z. Grid-based pavement crack analysis using deep learning. In Proc. The 4th International Conference on Transportation Information and Safety (ICTIS) 917–924 (IEEE, New Jersey, 2017). https://doi.org/10.1109/ICTIS.2017.8047878

26. Zhang, A. et al. Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network. Comput. Civ. Infrastruct. Eng. 32, 805–819 (2017).

27. Atha, D. J. & Jahanshahi, M. R. Evaluation of deep learning approaches based on convolutional neural networks for corrosion detection. Struct. Heal. Monit. An Int. J. 1–19 (2017). https://doi.org/10.1177/1475921717737051

28. Cha, Y. J., Choi, W., Suh, G., Mahmoudkhani, S. & Büyüköztürk, O. Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. Comput. Civ. Infrastruct. Eng. 00, 1–17 (2017).

29. Liu, L., Tan, E., Zhen, Y. & Yin, X. J. AI-facilitated Coating Corrosion Assessment System for Productivity Enhancement. In Proc. 2018 13th IEEE Conf. Ind. Electron. Appl. 606–610 (IEEE, New Jersey, 2018).

30. Malekzadeh, T., Abdollahzadeh, M., Nejati, H. & Cheung, N.-M. Aircraft Fuselage Defect Detection using Deep Neural Networks. arXiv pre-print at http://arxiv.org/abs/1712.09213 (2017).

31. Nguyen, T. et al. U-Net for MAV-based Penstock Inspection: an Investigation of Focal Loss in Multi-class Segmentation for Corrosion Identification. arXiv pre-print at http://arxiv.org/abs/1809.06576 (2018).

32. Jang, K., Kim, B., Cho, S. & An, Y. Deep learning-based concrete crack detection using hybrid images. in Proc. Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018 (ed Sohn, H.). 1059812, 36 (SPIE, Washington, USA, 2018).

33. Hou, W., Wei, Y., Guo, J., Jin, Y. & Zhu, C. Automatic detection of welding defects using deep neural network. J. Phys. Conf. Ser. 933, 012006 (2018).

34. Meng, M., Chua, Y. J., Wouterson, E. & Ong, C. P. K. Ultrasonic signal classification and imaging system for composite materials via deep convolutional neural networks. Neurocomputing 257, 128–135 (2017).

35. Prabhu, D. R. & Winfree, W. P. Neural network based processing of thermal NDE data for corrosion detection. Rev. Prog. Quant. Nondestruct. Eval. 12, 775–782 (1993).

36. Postolache, O., Ramos, H. G. & Ribeiro, A. L. Detection and characterization of defects using GMR probes and artificial neural networks. Comput. Stand. Interfaces 33, 191–200 (2011).

37. Barton, T. F., Tuck, D. I. & Wells, D. B. The identification of pitting and crevice corrosion using a neural/network. In Proc. 1993 The First New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems. 0–1 (IEEE, New Jersey, 1993). 10.1109/ANNES.1993.323012

38. Jingwen, T., Meijuan, G. & Jin, L. Corrosion detection system for oil pipelines based on multi-sensor data fusion by improved simulated annealing neural network. In 2006 International Conference on Communication Technology (IEEE, New Jersey, 2006). https://doi.org/10.1109/ICCT.2006.341978

39. Uruchurtu-Chavarin, J., M. Malo-Tamayo, J. & Hernandez-Perez, A. J. Artificial intelligence for the assessment on the corrosion conditions diagnosis of transmission line tower foundations. Recent Pat. Corros. Sci. 2, 98–111 (2012).

40. Sadowski, L. Non-destructive investigation of corrosion current density in steel reinforced concrete by artificial neural networks. Arch. Civ. Mech. Eng. 13, 104–111 (2013).

41. Butcher, J. B. et al. Defect detection in reinforced concrete using random neural architectures. Comput. Civ. Infrastruct. Eng. 29, 191–207 (2014).

42. Butcher, J. B. et al. in Concrete Solutions (eds. Grantham, M., Majorana, C. & Valentina, S.). 417–424 (CRC Press, USA, 2009).

43. Lin, Y. Z., Nie, Z. H. & Ma, H. W. Structural damage detection with automatic feature-extraction through deep learning. Comput. Civ. Infrastruct. Eng. 32, 1025–1046 (2017).

44. Zhao, R. Deep learning and its applications to machine health monitoring. Mech. Syst. Signal Process. 115, 213–237 (2019).

45. Janssens, O. et al. Convolutional neural network based fault detection for rotating machinery. J. Sound Vib. 377, 331–345 (2016).

46. Jing, L., Zhao, M., Li, P. & Xu, X. A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. Measurement 111, 1–10 (2017).

47. Zhao, R., Wang, J., Yan, R. & Mao, K. Machine health monitoring with LSTM networks. In Proc. 2016 10th International Conference on Sensing Technology (ICST) (2016). https://doi.org/10.1109/ICSensT.2016.7796266

48. Zhao, R., Yan, R., Wang, J. & Mao, K. Learning to monitor machine health with convolutional Bi-directional LSTM networks. Sens. (Switz.) 17, 1–18 (2017).

49. Jia, F., Lei, Y., Guo, L., Lin, J. & Xing, S. A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines. Neurocomputing 272, 619–628 (2017).

50. Li, X., Ding, Q. & Sun, J. Q. Remaining useful life estimation in prognostics using deep convolution neural networks. Reliab. Eng. Syst. Saf. 172, 1–11 (2018).

51. Sateesh Babu, G., Zhao P., Li XL. Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life. (eds Navathe S., Wu W., Shekhar S., Du X., Wang X., Xiong H.) Database Systems for Advanced Applications. DASFAA 2016. Lecture Notes in Computer Science, vol 9642. (Springer International Publishing AG, Cham, Switzerland, 2016).

52. Yuan, M., Wu, Y. & Lin, L. Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network. Int. Conf. Aircr. Util. Syst. 135–140 (2016). https://doi.org/10.1109/AUS.2016.7748035

53. Wu, Y., Yuan, M., Dong, S., Lin, L. & Liu, Y. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. Neurocomputing 275, 167–179 (2017).

54. Zhang, C., Lim, P., Qin, A. K. & Tan, K. C. Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics. IEEE Trans. Neural Netw. Learn. Syst 28, 2306–2318 (2017).

55. Zhang, Y., Xiong, R., He, H. & Liu, Z. A LSTM-RNN method for the lithuim-ion battery remaining useful life prediction. 2017 Progn. Syst. Heal. Manag. Conf. PHM-Harbin 2017 - Proc. (2017). https://doi.org/10.1109/PHM.2017.8079316

56. Zhang, Q., Yang, L. T., Chen, Z. & Li, P. A survey on deep learning for big data. Inf. Fusion 42, 146–157 (2018).

57. Nabian, M. A. & Meidani, H. Deep learning for accelerated seismic reliability analysis of transportation networks. Comput. Civ. Infrastruct. Eng 33, 443–458 (2018).

58. Fritsch, J., Kuhnl, T. & Geiger, A. A new performance measure and evaluation benchmark for road detection algorithms. In Proc. 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013) 1693–1700 (IEEE, New Jersey, 2013). https://doi.org/10.1109/ITSC.2013.6728473

59. Menze, B. H. et al. The multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Trans. Med Imaging 34, 1993–2024 (2016).