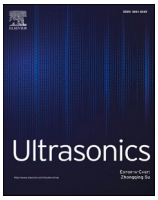
Ultrasonics 124 (2022) 106776

Contents lists available at ScienceDirect

Ultrasonics

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



A review of ultrasonic sensing and machine learning methods to monitor industrial processes

Alexander L. Bowler [a], Michael P. Pound [b], Nicholas J. Watson [a,*]

[a] Food, Water, Waste Research Group, Faculty of Engineering, University of Nottingham, University Park, Nottingham NG7 2RD, UK
[b] School of Computer Science, Jubilee Campus, University of Nottingham, Nottingham NG8 1BB, UK

ARTICLE INFO

Keywords:
Ultrasonic measurements
Machine learning
Deep learning
Industrial digital technologies
Transfer learning
Domain adaptation

ABSTRACT

Supervised machine learning techniques are increasingly being combined with ultrasonic sensor measurements owing to their strong performance. These techniques also offer advantages over calibration procedures of more complex fitting, improved generalisation, reduced development time, ability for continuous retraining, and the correlation of sensor data to important process information. However, their implementation requires expertise to extract and select appropriate features from the sensor measurements as model inputs, select the type of machine learning algorithm to use, and find a suitable set of model hyperparameters. The aim of this article is to facilitate implementation of machine learning techniques in combination with ultrasonic measurements for in-line and on-line monitoring of industrial processes and other similar applications. The article first reviews the use of ultrasonic sensors for monitoring processes, before reviewing the combination of ultrasonic measurements and machine learning. We include literature from other sectors such as structural health monitoring. This review covers feature extraction, feature selection, algorithm choice, hyperparameter selection, data augmentation, domain adaptation, semi-supervised learning and machine learning interpretability. Finally, recommendations for applying machine learning to the reviewed processes are made.

1. Introduction

The manufacturing sector is increasingly using the collection and interpretation of data to inform decision making and improve productivity, sustainability, and product quality [1]. This is part of the fourth industrial revolution, which is projected to culminate in Industry 4.0 and consist of fully interconnected supply chains, processes, and markets where intelligent, automatic decision-making adjusts to demands in real-time [2]. This transformation will be realised through the deployment of industrial digital technologies (IDTs) such as smart sensors, edge computing, cloud computing, the internet of things (IoT), and machine learning (ML). Sensors underpin this transition by acquiring the real-time data required to inform the decision-making process. This necessitates in-line and on-line sensors which do not require human operators, where in-line techniques directly measure the process stream and on-line measurements use automatic sampling methods [3]. Sensors can be adapted into smart sensors through additional functionalities such as wireless IoT connection or by providing some processing of the acquired data to reduce the complexity of the data being transferred [4]. Hardware solutions are required for process interconnectedness such as edge computing, where compute nodes are located close the end devices, or cloud computing, where data is transferred to a centralized cloud location [5]. ML can be used at all levels, from the individual sensors to the centralised data in the cloud, to analyse data and provide automatic decisions [6].

Discrete manufacturing is leading process manufacturing in IDT implementation owing to the much simpler processes to be monitored [7]. A wider range of sensor options is needed for process manufacturing to monitor more complex and often highly variable operations. The process analytical technology (PAT) initiative, first introduced to the pharmaceutical industry in 2004 and since spread to other sectors such as food, demonstrates the desire for greater process understanding [8,9]. PAT focuses on real-time sensor measurements, preferably in-line or on-line, which monitor critical process parameters that effect critical quality attributes of the products. There are many sensor techniques in development, each with different advantages and disadvantages making them suitable for specific applications [10]. Low power (intensities below 1 Wcm^2), high frequency (higher than 100 kHz) ultrasonic (US) sensors monitor the interaction of materials with mechanical sound waves. They benefit from being low cost, small in size, able to monitor

* Corresponding author.
E-mail address: nicholas.watson@nottingham.ac.uk (N.J. Watson).

https://doi.org/10.1016/j.ultras.2022.106776
Received 18 January 2022; Received in revised form 29 April 2022; Accepted 26 May 2022
Available online 28 May 2022

opaque materials, low in power consumption, able to operate non-invasively, non-destructive, real-time, in-line, and do not cause changes to the structure of the material through which they pass [11]. These attributes make US sensors the optimal sensor for certain applications and there use has been demonstrated for monitoring industrially relevant processes as reviewed in Section 3. Therefore, their appeal to industry can be expected to continue to grow.

The most commonly used US measurements include velocity, attenuation, and acoustic impedance [11]. The US velocity is calculated by measuring the time of flight and distance the sound wave has travelled. Attenuation is measured as a loss in sound wave energy as it passes through a material. Attenuation may be caused by absorption in homogeneous materials due to effects such as fluid viscosity, or by scattering due to encountering discontinuities in heterogeneous materials [12]. Acoustic impedance, the product of the sound velocity and material density, is typically monitored by measuring the proportion of a sound wave reflected from a boundary between two materials [13]. Pulse-echo sensing techniques utilise a single sensor to both transmit and receive a sound wave after reflection from an interface. Pitch-catch techniques use one sensor to produce the sound wave and another to receive it [13].

Traditionally, physical inversion models were developed from first principles to determine material properties from US measurements [14]. However, their development becomes challenging in real-life applications where the paths of the sound wave are often complex or the sound wave travels through multiple material interfaces. Furthermore, US properties are highly dependent on temperature and the presence of gas bubbles causes strong reflection of the sound wave, both of which must be accounted for [15,16]. As such, calibration procedures are commonplace that correlate ultrasonic measurements (such as the speed of sound, attenuation, or acoustic impedance) to desired material properties across a range of process parameters without defining the underlying paths of the sound wave. Calibration procedures also become complicated in industrial processes when many parameter ranges must be investigated, such as temperature, gas content, and the content of other heterogeneities [17]. ML uses algorithms to learn solutions to tasks without requiring explicit instructions. Supervised ML is a type of ML that maps inputs (or features) to outputs (or target variables) during training with the aim of producing a model that accurately predicts the outputs of previously unseen input data [18]. Supervised ML offers some distinct advantages over calibration methods: (1) The time investment for calibration procedures can be eliminated simply by monitoring the desired process across its natural parameter variations, so long as these are recorded and a reference measurement is available to label the sensor data with target variables. (2) ML typically uses a greater number of more complex US waveform features compared with calibration procedures allowing more US waveform information to be used in determining material properties. (3) ML models typically employ more complex fitting procedures to map input data to outputs. This allows more accurate predictions while minimising over-fitting to the training data through model regularisation and validation procedures. (4) Validation procedures can encourage development of ML models that accurately predict on new data from outside the range of process parameters that they were trained on. (5) ML models can be continuously retrained as more data becomes available to increase prediction accuracy. (6) Lastly, ML models can correlate sensor data directly to useful process information (such as classifying the state of a process or predicting the processing time remaining) rather than to material properties.

However, a lack of knowledge and experience in applying ML is a barrier to its deployment for US measurement analysis. To develop an adequate ML model, features must be extracted and selected from the US waveform, suitable ML algorithms must be identified and investigated, and a satisfactory set of hyperparameters must be chosen or found. Hyperparameters are any variables that may be selected by the ML model developer. The aim of this article is to facilitate the use of ML in combination with US sensors for in-line and on-line industrial process monitoring. This article first reviews the ability of US measurements to monitor processes before reviewing the combination of US measurements and ML including other areas such as structural health monitoring (SHM). This review covers feature extraction, feature selection and unsupervised learning, algorithm choice, hyperparameter selection, data augmentation, domain adaptation, semi-supervised learning and ML interpretability. Finally, recommendations are provided for combining ML and US measurements for the reviewed processes.

2. Machine learning background

This review includes supervised, unsupervised, and semi-supervised ML methods. Supervised learning uses features as inputs along with corresponding target variables as outputs (also known as labelled data) [18]. The ML algorithms then map the inputs to these outputs with the aim of accurately predicting the target variables for previously unseen data. This may be classification tasks, in which the target variables are discrete categories, or regression tasks where the targets are continuous variables. Unsupervised learning only uses input data for tasks such as finding patterns within the data or reducing its dimensionality. Semi-supervised learning is typically employed when a large volume of unlabelled data and a small volume of labelled data is available [19]. This may be due to the time and expense required to label each data point. Semi-supervised techniques may be used to pseudo-label previously unlabelled data points using knowledge from the labelled data. Then, a more accurate ML model can be constructed using the labelled and pseudo-labelled data compared with using the labelled data alone [18]. Fig. 1 displays a pipeline for supervised ML model development.

Labelled data is required to create the set of model outputs, or target variables, for the model inputs to be correlated to by the ML algorithm. Labelled data may be collected using: an alternative in-line or on-line sensing technique as a reference measurement such as imaging, density measurement, or particle sizer; off-line techniques where material samples are periodically collected during the process; stopping or sampling the process at stages, collecting data using a reference measurement technique, and using semi-supervised learning to pseudo-label the unlabelled data; using the US sensor measurements combined with prior process knowledge to infer process stages; or by transferring ML models between similar processes after domain adaptation [20]. Data labelling may be challenging in factory environments and its consideration should be taken into account throughout the different development stages of sensor and ML combinations.

During the supervised ML pipeline, the input data is usually divided in training, validation, and test sets. The training data is used for model training. Multiple trained models can be evaluated on the validation data to compare between algorithm choice, model architecture, hyperparameter and feature selection. Finally, the best performing models on the validation set are evaluated (or retrained and evaluated using the combined training and validation sets) on the test set to provide an assessment of the full ML pipeline. Several validation techniques are available such as holdout, k-fold cross validation, stratified k-fold cross validation, leave-one-out cross validation, leave-p-out cross-validation, and nested cross-validation. For a detailed comparison of these methods, readers are encouraged to visit [21]. Choice of the training, validation, and test data can also be used to evaluate the ML model's extrapolation capability beyond the process parameter bounds it was trained on [22]. This is a useful approach for applications with limited training data available or highly variable processes.

The success of ML tasks is in part dependent on the features used. A feature is any measurable property of the process being monitored that is inputted into an ML model [23]. Features can originate from the US signal or from other process parameters such as temperature or flowrate. Features may be extracted from the time domain US waveform, the frequency domain (for example after Fourier transformation [24,25]), or time–frequency domain following wavelet decomposition [26]. Wavelet

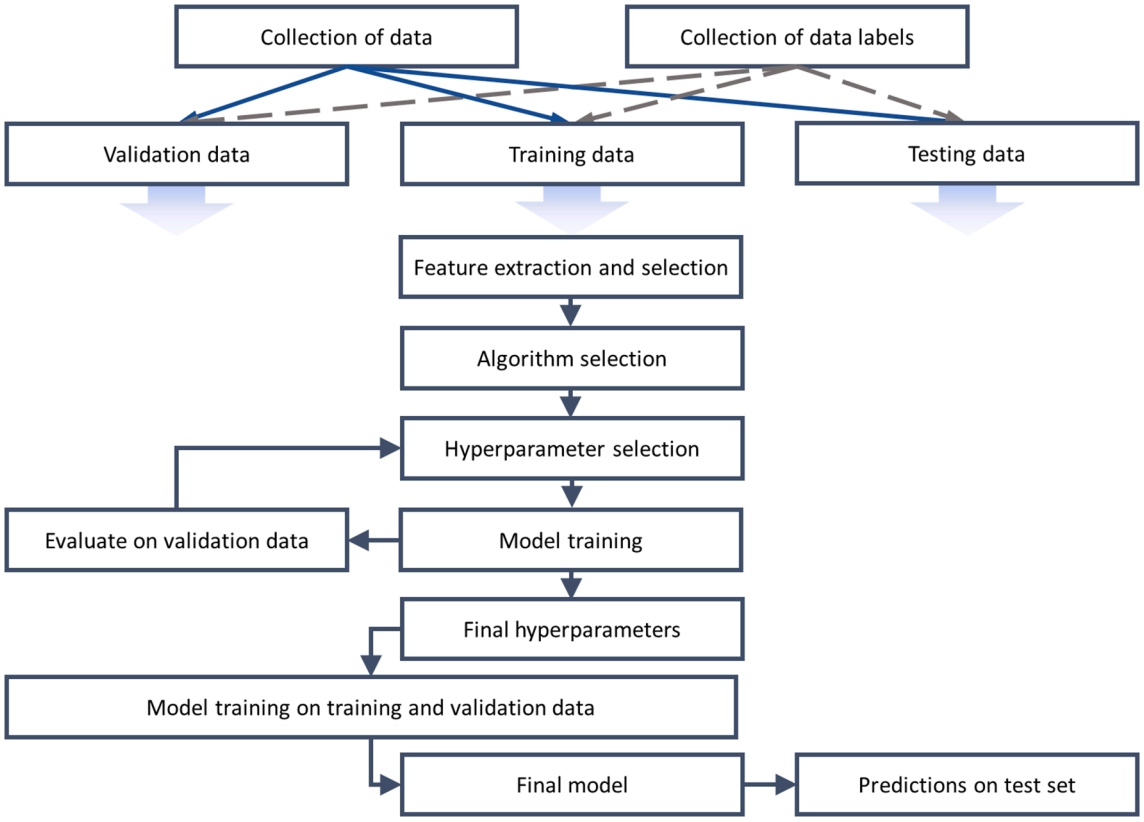


Fig. 1. An exemplar pipeline for developing supervised machine learning models.

analysis uses decaying waveforms as the transform function compared with the non-decaying sine or cosine waves used in the Fourier transform [27]. The continuous wavelet analysis uses a continuous range of frequencies to decompose the US signal whereas the discrete wavelet transform and wavelet packet transform use discrete frequencies at each decomposition. Wavelet packet decomposition performs successive decompositions on each branch of the original signal whereas the discrete wavelet transform only applies successive decompositions to the higher frequency signal content [15,27].

Feature selection encompasses methods of choosing which features to use in ML models or reducing the number of features by using algorithms. A common method is Principal Component Analysis (PCA) as used in [27–29] which is an unsupervised ML method that linearly transforms input variables into new, uncorrelated features called principal components (PCs) [30]. Feature selection can be used to improve ML model fitting by removing redundant information, reducing the likelihood of a model overfitting to its training data, providing simpler optimisation problems, and reducing the computational requirement to train the model [30,31]. However, some information from the input features could be lost leading to a reduction in ML model accuracy.

According to the No Free Lunch theorem, all optimisation techniques are equally as accurate when averaged over all possible problems [32]. Therefore, the optimal algorithm to use is dependent on the application. However, some knowledge of the procedure of each algorithm can help in identifying which to try. For classification tasks, support vector machines (SVMs) find a hyperplane that separates two classes of data by maximizing its distance from the closest data points from each category. In regression tasks, support vector regressors fit lines to continuous data by only accounting for the error from data points outside a set distance from the fitted line. SVMs generalise well to new data and, as they are effective with high dimensional feature spaces, make use of the kernel trick for non-linear fitting [33]. Decision and regression trees (DT) use conditions which are successively applied to the input data until an output decision is reached. They are simple, interpretable, have low computational cost, can be graphically represented, but typically have lower accuracy compared to other algorithms [34]. Random forests (RF) are an ensemble method that combines the predictive performance of multiple DTs by, for example, selecting the most common class predicted in classification tasks or the mean output for regression tasks [16]. K-nearest neighbours (KNN) uses the distance between data points in the feature space and a voting procedure of the K nearest training instances to determine the class or regression value of the queried data point [16]. Artificial neural networks (ANNs) can create new features in their hidden layers from combinations of input features to non-linearly fit model inputs to the outputs. In ANNs, information flows by feed-forward propagation from the input layer, through hidden layers, to the output layer. Weight and bias terms connect all units in the previous layer to all the units in the following layer. During training, the weight and bias terms are iteratively altered through backpropagation of the prediction error and gradient descent steps [34]. A deep neural network (DNN) is an ANN with more than one hidden layer. Convolutional neural networks (CNN) have convolutional layers as well as fully connected layers and are widely used in image recognition tasks [35]. The convolutional layers consist of filters of weights which perform cross-correlation on the input data. This enables CNNs to learn their own features from the input data at lower computational expense and with simpler optimisation than fully connected neural networks of similar size. Furthermore, CNNs are spatially invariant meaning that they are robust to changes in feature locations, unlike fully connected neural networks. The accuracy of ML methods is limited by the choice of features inputted into the models. Therefore, CNNs offer the advantage of negating the need for feature extraction or selection by automatically learning features important to the task (Fig. 2). Long short-term memory neural networks (LSTMs) are able to learn sequences of time series data and are widely used in natural language processing applications. LSTMs are a development of recurrent neural networks (RNNs) that reduce the likelihood of exploding or

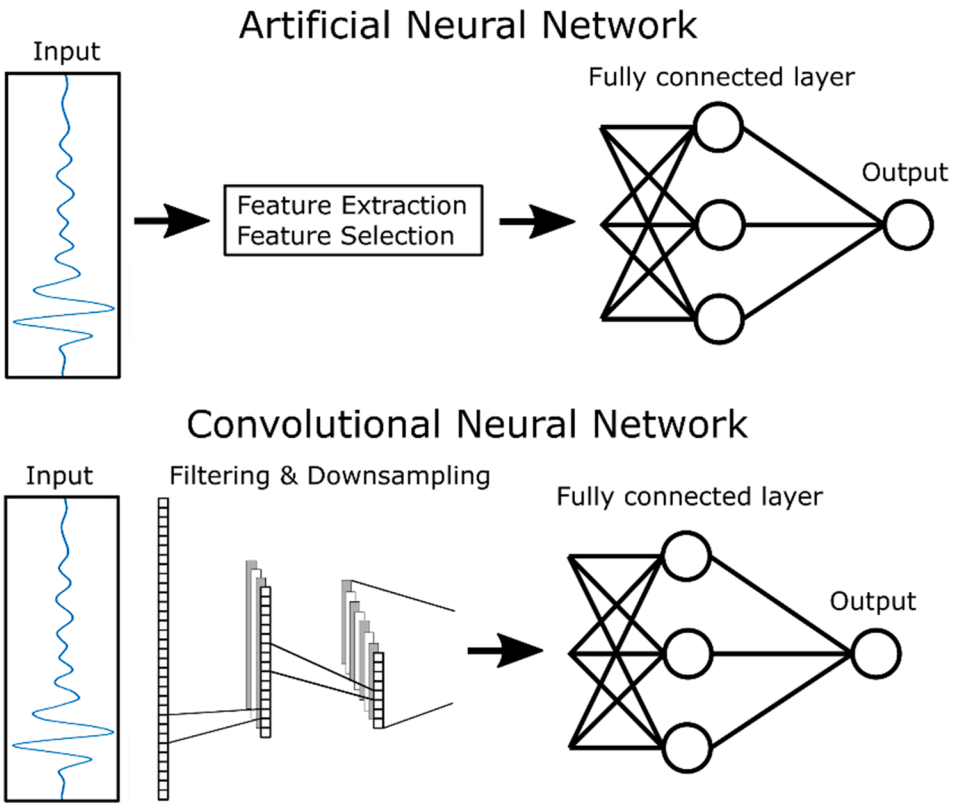


Fig. 2. A comparison between Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs). ANNs require feature extraction and, if necessary, feature selection before using the model. In contrast, CNNs use convolutional filters, and pooling layers to downsample the data, to automatically extract features. ANNs may be used with feature extraction or selection, i.e. the raw data is used as inputs. However, this network would not be robust against spatial variance of the features, unlike CNNs.

vanishing gradients, enabling the learning of long-term dependencies [36]. LSTMs can store representations of sequences by using gate units to update their internal network state. At each time step, LSTMs use the input features as well as information passed from the previous time step to make their prediction.

Hyperparameters are variables that may be selected by the developer. This may be through trial and error, using values previously employed in other works, through grid-searches of possible hyperparameter combinations, or through other procedures such as Bayesian optimisation. Hyperparameter selection can be evaluated using part of the data as a validation set. Hyperparameters may define the structure of the algorithm or how it trains. For example, neural networks (encompassing ANNs, DNNs, CNNs, and LSTMs) often require regularisation to prevent them from over-fitting to the training data and limiting their ability to generalise to new data. Common regularisation techniques include L1 and L2 penalties, early stopping, and dropout layers. L1 regularisation adds a penalty term to the error value that is the sum of all the weight magnitudes, whereas L2 sums the square of the weights [37]. Therefore, a model is penalised for having many or large weight values depending on the chosen magnitude of these regularisation penalties. Early stopping evaluates the current model on a validation set during training. After the prediction accuracy on this validation set decreases for a specified number of iterations, training is stopped [37]. Dropout layers randomly drop network nodes during training according to a specified probability. This effectively "thins" the network during training, allowing multiple input data propagation paths through the network and reducing co-adaptation of the hidden nodes [38].

3. Process monitoring using ultrasonic sensors

The section reviews the use of in-line and on-line ultrasonic measurement techniques to monitor processes including cleaning, fermentation, crystallisation, mixing, extrusion, injection moulding, curing, reactions, tabletting, and membrane fouling.

3.1. Cleaning

Cleaning is a process used to remove material from the internal surfaces of processing equipment in sectors such as food and drink, pharmaceutical and Fast-Moving Consumer Goods (FMCG) [39]. However, cleaning is usually carried out for a predetermined length of time which is designed to over-clean the equipment. With real-time monitoring of fouling removal, time and cleaning resource (e.g. water, energy, and chemicals) use can be minimised. This not only improves process economics but sustainability as well [40]. Furthermore, the build-up of fouling decreases the efficiency of heat exchangers and so a method to detect the presence of this fouling would allow for improved scheduling of heat transfer equipment cleaning [41].

Wallhäußer et al. [41] combined US measurements and an ANN to classify whether a model heat exchanger was fouled by dairy protein deposits. A single US sensor monitored waveforms reflected from the plate-fouling interface and from the far wall of the heat exchanger. The ANN achieved an accuracy of 98.6%. Wallhäußer et al. [42] used ANNs and SVMs to classify the presence of protein or mineral fouling. The SVMs achieved higher accuracies compared with the ANNs. It is suggested that this is due to the ability of SVMs to find global minima, opposed to local minima found by ANNs. In actuality, this may be due to the ability of SVMs to generalise well to new data and that not enough regularisation was applied to the ANNs. This is likely, as no validation set was used to evaluate hyperparameters for the ANN. Only features from the reflection from the plate-fouling interface was used. The SVMs achieved 97.6 % accuracy when the mineral and protein fouling datasets were combined, and 100 % and 98.2 % for the protein and mineral fouling, respectively, when each dataset was used individually. Wallhäußer et al. [43] monitored the cleaning process of protein fouling using a sodium hydroxide solution. Seven US waveform features were used along with the process temperature and cleaning fluid mass flow rate. The gradients of the US features were also monitored to identify a plateau that could indicate the end of cleaning. SVMs achieved greater than 94 % accuracy when classifying whether the heat exchanger was

fouled or cleaned. Úbeda et al. [40] used a single US transducer to monitor cleaning of milk fouling in a model plate heat exchanger. Milk protein deposits build up on heat transfer surfaces during thermal treatment, reducing their efficiency and necessitating cleaning. Cleaning was carried out for 95 min using sodium hydroxide and, subsequently, nitric acid solutions. Two sections of the received US waveforms were analysed: the start of the waveform corresponding to sound wave reflections from the plate-fouling interface, and the third echo reflecting from the far wall of the plate heat exchanger (after passing through the fouling and cleaning fluid or cleaning fluid alone). For the sound wave reflections from the plate-fouling interface, three US waveform features were monitored: the temporal spread, temporal roll-off, and temporal inertia. These features were used in an ANN to classify whether the plate was fouled or cleaned. For the third echo, four features were used: the maximum amplitude, spectral crest factor, spectral centroid, and temporal entropy. Classification accuracies of up to 98 % for the plate-interface features and 96% for the 3rd echo features were achieved.

Escrig et al. [39] monitored the removal of three types of food fouling (gravy, tomato paste, and malt) at two cleaning fluid temperatures (12 °C and 45 °C) using a single US sensor. The US sensing method monitored the US waveform reflected from the interface between the pipe wall and the fouling material. The experiments were conducted in a lab-scale pipe section. The pipe section was cuboid with a flat, stainless steel bottom plate where the US sensor was attached externally. Three features from the received US waveform were monitored: the waveform energy, peak-to-peak amplitude, and root mean square error of the amplitudes at every sample point in the waveform compared to that of a clean pipe. The US sensor could identify differences in the cleaning mechanisms between the mechanical removal of the tomato paste and gravy compared with the dissolution of the malt. The US technique was only sensitive to the area of fouling coverage not the fouling thickness. Escrig et al. [16] expanded this study by training classification machine learning models to predict whether the pipe section was fouled or cleaned. The highest model accuracy was attained when the amplitudes at each sample point in the waveforms were used directly instead of using any further feature extraction methodologies. A K-best feature selection methodology was used to select the number of amplitudes to be used. The classifiers used were KNN, SVM, RF and adaboost RF. It was found that combining multiple datasets from different fouling materials resulted in improved model accuracy. Accuracies up to 99 % were achieved. Simeone et al. [15] used the same pipe test section and materials to monitor cleaning by combining US and optical sensors. A three-level wavelet package transform using the 3 Daubechies mother wavelet was performed on the US waveforms. Afterwards, the mean, standard deviation, minimum, maximum, skewness, kurtosis, and energy were extracted from the decomposed signals. These features were then input into an ANN for a regression task to predict the surface area or volume of fouling remaining. Escrig et al. [44] used classification methods to monitor tomato paste and malt fouling in plastic (PMMA) and metal (stainless steel) cylindrical pipe sections. Accuracies up to 100% were achieved for both pipes. Finally, Chen et al. [45] used a single US sensor to monitor cleaning of wax deposits from a flat duct section over the course of 3 h. A decorrelation coefficient of coda waves compared with a clean plate was monitored and cleaning was completed after 2.2 h.

3.2. Fermentation

Fermentation processes are conventionally monitored through sampling and off-line analysis [46]. However, this has issues of requiring manual operation, risking contamination, and lacking timely results [17]. There are several types of fermentation that have been monitored using US sensors, such as alcoholic fermentation where yeast converts sugar into ethanol and carbon dioxide [46], lactic acid fermentation where lactose is converted to lactic acid through bacteria metabolism [47], and malolactic fermentation in red wines where malic acid is converted into lactic acid which is an important process for developing sensory characteristics [48].

Becker, Mitzscherling, and Delgado [49] used a single externally mounted sensor to monitor beer fermentation in a 300 m^3 tank under industrial conditions for 90 h. The US wave was transmitted across the 4.5 m diameter vessel and was reflected at the far wall before returning to the transducer. The US velocity was monitored and an ANN was used to compensate for the effects of temperature. Resa, Elvira, and De Espinosa [46] mounted two non-invasive US sensors to a square glass bottle (64 mm ID) to monitor the US velocity using a pitch-catch transmission method. A water bath was used to keep a constant temperature of 30 °C. The US velocity decreased with the decreasing density of the fermenting medium. Resa et al. [47,50] used similar experimental methods to monitor lactic acid fermentation. The US velocity decreased throughout fermentation despite no significant change in density. Resa et al. [17] monitored wine and beer fermentations and reported a decreasing US velocity with the decreasing density. The US amplitude was also used to monitor the beer wort fermentation and an increase in attenuation was obtained during the start of ethanol production due to the production of CO_2 bubbles. Similarly, Lamberti et al. [51] used two transducers for transmission across a 35 mm diameter square bottle to monitor wine fermentation. A decreasing US velocity during ethanol production was found.

Hussein, Hussein, and Becker [24] implemented a single US sensor on a circulation line for in-line monitoring of a 60 L (working volume) yeast fermentation process. The sensor used a reflector plate to transmit through the fermenting liquid. Frequency domain analysis, after Fourier transformation, and phase shift correction were used for the time-of-flight measurements. The US velocity was combined with nine waveform features and the temperature and was inputted into an ANN to predict the liquid density. The ANN produced a maximum error of 0.95%. The US velocity increased throughout the fermentation, contradicting the previously reported results presented above. Hoche et al. [52] also found the US velocity to increase during fermentation and the reflection coefficient to decrease. The reason for these results may be due to the larger scale process and industrial conditions monitored in [24,52] compared with [17,46,47,50,51]. At these specific combinations of temperature, along with the content of sugar, ethanol, yeast, and CO_2, the US velocity may increase during fermentation. [52] used invasive sensor probes with a 50 mm reflector plate transmission distance to monitor alcoholic fermentation in vessels up to 2140 L (maximum volume) in size. A sound velocity–density–temperature calibration model achieved an average root mean square error of 0.53% g/g sugar and 0.26% g/g ethanol during the fermentations. However, this does require a secondary measurement of the density of the fermenting wort. Bowler et al. [53] used LSTMs to predict the alcohol concentration during beer fermentation in a 30 L vessel. An invasive probe with a reflector plate was used. Accurate monitoring of the alcohol content was achieved without using the sound wave reflection that had passed through the wort or using the process temperature. This indicates that a non-invasive, reflection-mode US sensing technique could be possible. The energy of the reflection from the probe-wort interface increased throughout fermentation although no trend in the speed of sound was identified owing to variations in the process temperature.

Ogasawara et al. [54] used two invasive probes in pitch-catch mode with a 15 mm transmission distance to monitor yogurt fermentation in a beaker. The US velocity was used to detect phase changes of the yogurt along with variations in the internal temperature caused by the exothermic fermentation reaction. Meng et al. [55] used a single non-invasive sensor to monitor yogurt fermentation through the wall of a 250 ml stainless-steel reactor. The acoustic impedance was monitored using multiple reflections from the wall-yogurt interface. The acoustic impedance of the yogurt increased between pH 5.6 and 5.3.

Novoa-Díaz et al. [48] used an invasive US sensor probe with a reflector plate to monitor malolactic fermentation in red wine. The US velocity increased during the fermentation until reaching a stationary

phase after approximately six days. It was proposed that identification of the stationary phase could be used to determine the fermentation end point. However, temperature compensation would be required to uncover its masking effect on the velocity. Amer et al. [56] presented a temperature compensation strategy by determining the effect of temperature at different concentrations of alcohol and Amer et al. [57] presented two further temperature compensation methods. Çelik et al. [58] presented the design of an invasive sensor probe with reflector plate to be installed in the side of industrial vessels for malolactic acid fermentation.

Lastly, Keskinoğlu and Aydln [59] used two non-invasive transducers in pitch-catch mode to monitor cell growth in a small-scale vessel. From the US velocity, growth curves were obtained that could identify the lag, growth, and stationary phases.

3.3. Crystallisation

Crystallisation is a process predominantly used in the chemical and pharmaceutical industries [60]. The most important parameters to control are the mean crystal size and the crystal size distribution as these determine the properties of the product and effect downstream processing [61]. Mougin et al. [62] used an invasive probe consisting of two pairs of broadband US transducers to monitor the crystallisation of two organic compounds. US attenuation spectroscopy using the Epstein and Carhart and Allegra and Hawley (ECAH) scattering model was utilised. The frequency range of the probe spanned from 1 MHz to over 150 MHz and was capable of measuring particle sizes in the range from 0.01 to 1000 μm. The minimum crystal concentration for size characterisation was 0.1% vol. The technique had limited effectiveness for monitoring the crystallisation of urea owing to the formation of high aspect ratio needle crystals whose long axial length was beyond the sensor measurement range. Mougin, Wilkinson, and Roberts [63] went on to monitor the particle size of crystals in two different polymorphic forms. In this work, the US attenuation spectroscopy method was less sensitive than turbidometric measurements for the determination of the onset of crystallization. Mougin et al. [61] used a 470 ml stainless steel flow-through cell connected to the side-ports of a double-jacketted 2.6 L glass reactor to determine the crystal size and solid concentration during crystallisation. The method could identify secondary nucleation, growth and crystal breakage and the results were used to determine kinetic parameters such as the secondary nucleation rate and growth rate. Li et al. [64] inputted US attenuation spectroscopy measurements into two neural networks. The first ANN was used to predict the mean crystal size and crystal size standard deviation. The five inputs were the attenuation at four different frequencies along with the process temperature. The second neural network was used to predict the US attenuation at a reference frequency to determine solids concentration. Its inputs were the temperature and the mean crystal diameter and standard deviation predictions from the first ANN. Although the authors acknowledge that a single ANN could have been used to make both sets of predictions, two separate models were used to simplify the ANN training. ANNs were used as they did not need knowledge of the solid and liquid physical parameters required for the ECAH model. Furthermore, the ANNs could be used in-line where as the ECAH model must be completed off-line due to the long iterative process required. Lyall et al. [65] was able to monitor crystal breakage and the mass and linear crystal growth rates were determined from the US measurements. Shukla, Prakash, and Rohani [66] employed US attenuation spectroscopy using a single invasive sensor with a reflector plate in a jacketed glass reactor of 0.115 m diameter and 0.25 m height.

Pertig et al. [67] used an invasive probe and measurements of the US velocity and attenuation at a single frequency to determine the mean particle size and suspension density. Experiments were performed isothermally in a jacketed vessel with a diameter of 115 mm and a height of 200 mm. The method could measure particle sizes between 200 and 800 μm with solids content up to 40 wt%. This method was presented as faster, less expensive, and simpler than US attenuation spectroscopy which must be conducted across multiple frequencies. Stelzer, Pertig, and Ulrich [60] used two invasive probes in a 1 L jacketed glass crystallizer to monitor the suspension density, mean crystal size and liquid concentration. One sensor was surrounded by a mesh to prevent crystals from entering the measurement line so the liquid properties could be monitored. Frohberg and Ulrich [68] showed that the same two-sensor technique could be used for the determination of the metastable zone width, nucleation and growth kinetics, seeding events, and detection of phase transitions. Helmdach, Feth, and Ulrich [69] showed that calibration transfer using the same sensing technique was possible between lab and pilot scale processes so long as the influence of gas was minimal in the pilot-plant setup. Morris et al. [70] used a single US sensor with reflector plate in a 250 ml reactor. US attenuation spectroscopy and monitoring of multiple reflections was utilised.

3.4. Mixing

Mixing is a ubiquitous process across manufacturing, such as in the food, chemical, and pharmaceutical industries [10]. In many industries, mixing is typically carried out for a predetermined length of time without monitoring the product quality. Classification of whether the materials were mixed or a prediction of when the mixing process will finish would enable more consistent product quality, more efficient resource use, and better equipment scheduling. Bamberger and Greenwood [71] used an invasive probe to monitor slurry suspension in a 1.91 m diameter tank. The US attenuation was measured across three transmitter–receiver transducer pairs located at different heights along the probe and separated by a 10.2 cm distance. Fox, Smith, and Sahi [72] and Salazar et al. [73] used single, invasive sensors to monitor air incorporation into batters during mixing. Both sensors monitored the changing acoustic impedance of the batter by measuring the reflected sound wave. Tourbin and Frances [74] used a flow-through cell and US attenuation spectroscopy to monitor the suspension and aggregation of nanoparticles (mean diameter of 80 nm) in a 1 L capacity stirred tank. Liu et al. [75] also used a flow-through cell and US attenuation spectroscopy in combination with Electrical Resistance Tomography (ERT) to monitor crossflow membrane emulsification. The size distribution and concentration of droplets was determined using the ECAH inversion model. Hunter et al. [76] employed an in-situ, multi-frequency acoustic backscatter system to monitor high concentration particle dispersion. Homogeneous glass powder dispersions were monitored at small and large (2 m^3 mixing tank) scale. Transducers with central frequencies of 1, 2, 4, and 5 MHz were used in pulse-echo mode and the attenuation decay with time of the returning signal was dependent on the particle concentrations. Bowler, Bakalis, and Watson [27] used single, non-invasive US sensors in reflection-mode to monitor two model mixing systems: honey-water blending and flour–water batter mixing. Classification ML models were developed to predict if the materials were mixed or not, and regression models were trained to predict the time remaining until (or time passed since) the materials were fully mixed. ANNs, SVMs, LSTMs, and CNNs were all tested with extensive feature extraction in both the time and time–frequency domains (after applying the discrete or continuous wavelet transform). Multi-sensor fusion between two sensors was also investigated. Classification accuracies of up to 96.3% for the honey-water blending and 92.5% for the flour–water batter mixing were achieved, as well as R^2 values for the regression models of up to 0.977 for the honey-water blending and 0.968 for the flour–water batter mixing. Each prediction task achieved optimal accuracy using different ML algorithms and feature extraction methods.

3.5. Extrusion

Using ultrasonic sensors to monitor extrusion processes has applications in industries such as polymer processing and food production [77]. Coates et al. [78] used US and spectroscopic sensors to monitor the

blending of polyethylene and polypropylene pellets in a 38 mm single screw extruder. Two US transducers were used in pitch-catch mode across a path distance of 15 mm. The US velocity was more sensitive than the spectroscopic measurements in detecting a change of blend composition. Barnes et al. [79] monitored the blending of ethylene vinyl acetate (EVA) random co-polymers with varying vinyl acetate (VA) content. US probes were implemented into dies attached to the end of the extruder barrel and the time of flight was measured. Although, pressure fluctuations that effected the US velocity measurements did not affect the NIR probe measurements. Alig et al. [80] combined US, spectroscopic, and rheometry measurements, as well as US and dielectric measurements, in slit die extruders. US attenuation was used to monitor polymer filler blending. Sun et al. [81] monitored filler dispersion during extrusion using two US sensors. An ANN was trained using the US velocity and attenuation along with the pressure, temperature, filler type and feed rate for prediction of the dispersion index. An error of less than 5% was achieved.

Fischer et al. [82] combined US and spectroscopic measurements to monitor additive blending during extrusion. Fischer et al. [83] used NIR spectroscopy and US attenuation spectroscopy to monitor the extrusion of polymer nanocomposite blends. The measurements were used to determine the dispersion extent and the impact strength of the polymer product. Schober et al. [84] also used US attenuation spectroscopy to monitor dispersion and particle size during polymer melt extrusion. Wöckel et al. [85] monitored the reflected sound wave using a single US sensor. The standard deviation of consecutive signals was used to determine the filler concentration. Halmen et al. [86] used US tomography to determine filler distribution in polymer melts. The US velocity and attenuation from a 60 mm ID sensor ring consisting of 40 transducers were used.

3.6. Injection moulding

US sensors have been widely applied to monitor injection moulding of polymer materials. For a full review of this area, the interested reader is directed to [87]. Recently, Wu et al. [88] presented a T-shaped extension nozzle with two integrated high temperature US transducers in transmission mode for non-invasive monitoring of injection moulding. The US velocity and attenuation could follow the process stages and was also correlated to the polymer flow speed. Altmann, Praher, and Steinbichler [89] used three US transducers (10 MHz) in pulse-echo to monitor melting behaviour in injection moulding. Zhao et al. [90] used a single US transducer for in-line monitoring of microcellular injection moulding. The duration of the US signals and the change in US velocity could be used to monitor variations in cell size and thickness of the skin layer. Cheng and Wu [91] used two high temperature US sensors in transmission mode to non-invasively monitor injection moulding of two types of plastic. Each stage of the injection moulding process could be identified from the US measurements, and the effect of injection speed on the quality of the final product could be monitored using the US velocity. Zhao et al. [92] used the US velocity and pressure measurements for in-line temperature measurement during injection moulding. Finally, outside of polymer processing, Grob et al. [93] used four US transducers (two to transmit US waves and two to receive) to monitor the crystallisation, solidification, contraction, and mould wall detachment of chocolate products. Detachment from the mould wall produced a reduction in the US amplitude.

3.7. Curing

Rath et al. [94] used US sensors in a compression mould and measured the velocity and attenuation during curing to evaluate the effects of different mouldings compounds, elevated temperatures, and filler, moisture, and hardener content. Lionetto, Tarzia, and Maffezzoli [95] used two air-coupled US sensors to monitor the curing of resin. The US measurements were corrected for the variations in air temperature caused by the exothermic reaction by periodically switching the US sensors from pitch-catch to pulse-echo mode. Lionetto and Maffezzoli [96] used the US velocity and attenuation to monitor the curing processes of thermosetting resins. Both contact and air-coupled US techniques were used. Koissin, Demčenko, and Korneev [97] used a noncollinear US wave mixing approach to monitor curing of resin. This technique uses the interaction of two US waves to produce scattered waves with mixed frequencies. Ghodhbani, Maréchal, and Duflo [98] used a single US sensor to monitor the liquid viscous, glassy transition, and saturation solid stages during curing of an epoxy resin. Dominguez-Macaya et al. [99] used an air-coupled US sensor to monitor longitudinal and, after the gel point was reached, shear waves during ultraviolet curing of a vinyl ester resin. The US system was also used to monitor the change in thickness of the resin due to shrinkage. Finally, [100–103] measured the velocity and attenuation of US waves to monitor curing of carbon fibre-reinforced plastics.

3.8. Reaction monitoring

Pawelzyk, Toledo, and Willenbacher [104] monitored US velocity and attenuation during styrene emulsion polymerization. However, this was conducted using a through-transmission method across a small-scale sample volume (15 ml). Buckin and Atlas [105] and Buckin [106] demonstrate how non-invasive, through-transmission measurement of the US velocity and attenuation can be used to determine many phenomena of reactions at small scale. For example, [106] reviews the monitoring of substrate and product concentrations, degree of polymerisation, polymer molar mass, reaction rates, catalyst inhibition, reversible and irreversible thermal deactivation, and particle size changes in dispersions. Figueiredo et al. [107] used a single transmission-based US sensor to monitor the transesterification process of biodiesel at small scale (70 mm diameter vessel). Stabilisation of the US velocity and amplitude indicated the achievement of the maximum yield and that the process should be stopped at this point. Baêsso et al. [108] used the US velocity and attenuation to determine the content of contaminants or by-products in biodiesel samples. In this way, US sensors could eventually be used for in-line monitoring of transesterification final products. Schmachtl et al. [109] monitored the synthesis of zeolite A and zeolite X using a transmission US sensing method a small scale. Decreases in the US velocity and attenuation were correlated with gel formation at the start of the process. An increase in attenuation and corresponding peak in US velocity indicated zeolite crystallisation. Hums, Baser, and Schwieger [110] used an invasive US transducer and reflector plate to monitor nucleation and crystal growth during the hydrothermal synthesis of zeolite A and X from coal fly ash. Van Groenestijn et al. [111] used a US nanoparticle sizer probe to monitor the synthesis of spherical silica nanoparticles. The ECAH method was used to obtain the size and concentration of particles.

3.9. Tabletting

Tabletting involves the compaction of powders into tablet forms using punches. Stephens et al. [112] used a single, non-invasive, embedded US sensor to monitor the mechanical properties of tablets during compaction. The time of flight and reflection coefficient of the sound wave reflecting from the interface between the upper punch and the powder was monitored. Leskinen et al. [113] used two transducers implemented inside the upper and lower punches of a tableting machine. Through-transmission was used to measure the US velocity and frequency spectra to monitor the mechanical properties of tablets.

3.10. Membrane fouling

US measurements have been widely applied as a non-invasive technique to monitor membrane fouling during micro-, ultra- and nano-filtration separation processes [114,115]. The amplitudes of reflected

sound waves from the fouling and membrane layers are measured to monitor fouling formation. For example, Li et al. [116] used an in-situ US technique to monitor organic and colloidal fouling during nano-filtration. Differences in the fouling process were observed in the US measurements when using different mixtures of foulants.

4. Ultrasonic sensors and machine learning

This section reviews the use of ML with US sensor measurements applicable to industrial process monitoring and other similar applications. Feature extraction, feature selection, algorithm choice, hyper-parameter selection, data augmentation, and domain adaptation is reviewed. For reviews on similar areas, the interested reader is directed to [117–128].

4.1. Feature extraction

The choice of features to investigate may be decided through either understanding of US sensors and the process being monitored, plotting features over the course of the process and monitoring their trends, or preliminary experiments to evaluate the accuracy of ML models using different feature combinations. As explained in Section 2, features can be extracted from the US waveform in the time domain, frequency domain following the Fourier transform, time–frequency domain following wavelet transform, or after other transformations such as cosine [25], chirplet [129], or short-time Fourier [130]. Empirical mode decomposition has also been widely applied in the SHM community and a review of this area can be found in [131].

The choice of Mother wavelet and level of decomposition are important decisions when applying wavelet transformation to US waveforms [25]. For this reason, a review of wavelet methods to analyse US signals is provided in Table 1. Typically, the choice of mother wavelet is selected as that most visually similar to the received US waveform [27]. The number of decompositions and vanishing moments can be decided by evaluating ML models using different values [27].

Table 2 reviews features extracted from US waveforms and inputted into ML models. A US waveform in the time, frequency, or time–frequency domain is a function composed of amplitudes. Specific amplitudes in these functions can be monitored, such as the maximum, minimum, or peak-to-peak amplitudes. Other features use many of these function amplitudes in a single measure. For example: the standard deviation or variance monitor amplitude dispersion relative to an average magnitude, features such as the temporal slope monitor the rise or decrement of amplitudes in a function, features similar to the energy provide a measure of the overall magnitude of the function, the crest factor measures the dominance of the maximum amplitudes, and skewness and kurtosis provide measures of the shape of the function. Another set of features can provide measurements of the position of the function in its respective domain, such as the time centre, average frequency, temporal duration, or bandwidth. Also, all, or a subset of, function amplitudes may be used as features directly rather than incorporated into other measures [16,27,35,148]. The time of flight monitors the distance in the time domain between two waveforms and is used to measure the speed of sound in the process material. The variance between consecutive waveforms of any of the previously listed features may also be monitored for example to monitor the production of CO_2 during fermentation [53] or identify flow regimes [149,150]. Time-lagged features or feature gradients can be used to incorporate information from past time-steps into the ML models for processes that progress over time. Finally, additional features, such as the process temperature or material mass flowrate can be used to provide extra information to the ML models about the process being monitored.

Table 1
A review of US waveform decomposition through wavelet analysis as a feature extraction methodology for ML models.

Reference	Wavelet decomposition transform	Number of decompositions	Mother wavelet and number of vanishing moments
[15]	Wavelet packet	3	3 Daubechies
[27]	Discrete wavelet	3, 5, 7	Symlet
[27]	Continuous wavelet	N/A	Morlet
[132]	Wavelet packet	5	5 Daubechies
[133]	Discrete wavelet	4	Daubechies
[134]	Wavelet packet	4	5 Daubechies
[135]	Discrete wavelet	4	5 Coiflet
[136]	Continuous wavelet	N/A	Various investigated
[137]	Continuous wavelet	N/A	3 Morse
[138]	Discrete wavelet	4	1 Debuchet
[139]	Wavelet packet	3	4 Daubechies
[140]	Discrete wavelet	7	Created own mother wavelet
[141]	Wavelet packet	3	8 Symlet
[142]	Discrete wavelet	3	5 Coiflet
[143]	Discrete wavelet		10 Daubechies
[144]	Discrete wavelet	5	10 Daubechies
[145]	Discrete wavelet	5	8 Daubechies
[146]	Discrete wavelet	5	8 Symlet
[147]	Continuous wavelet		Gaus

4.2. Feature selection

This section reviews feature selection methods used with US measurements and ML. Feature selection encompasses methods to reduce the number of features inputted into ML models, however, it is not a mandatory step in the ML pipeline. Feature selection can improve ML model training by removing redundant information, reducing the likelihood of overfitting, providing an easier optimisation problem for the algorithm, and reducing the computational requirement for model training [30,31]. However, some information from the input features could be lost leading to a reduction in accuracy. As explained in Section 2.2, PCA is a common method of feature selection and was employed in [27–29,156–159]. Ref. [16] used a K-best predictors method to select the sample point amplitudes from waveforms. This involved using a grid search of the amplitudes used and an F-test to determine their importance. Ref. [28] used a Garson's method which calculated feature importance from weights of a previously trained ANN. The feature importance was scored between 0 and 1 and features scoring below a threshold value of 0.35 were discarded. Ref. [25] used the Wilcoxon-Mann-Whitney rank test to find class discriminant features. This method is usable in binary classification tasks and does not determine if features are redundant or not.

Autoencoders can also be used as unsupervised ML methods for feature selection. Autoencoders are a type of neural network that aims to reconstruct its inputs after having passed the data through a bottleneck, or latent space, in the network. For example, [160] used a convolutional autoencoder as a feature extraction methodology. During training, the information about the input signal contained in the latent space is maximised so that it may be reconstructed. After training, the encoder part of the network (from the inputs to the latent space) may be applied to new data as a feature extractor. Similarly, [161] used autoencoders as a feature extraction method for detecting fatigue damage in structures. Autoencoders have also been used for other applications when used with US sensor data. For example, [162] used a convolutional autoencoder to reconstruct noiseless US signals after artificial noise has been added. In this way, the trained autoencoder could then be used to denoise new US signals. The input signals consisted of 2048 datapoints and the latent space was 256 neurons. Ref. [163] used convolutional denoising autoencoders to remove the effects of temperature on US guided waves for structural health monitoring applications. Refs. [164,165] applied denoising autoencoders for US waveforms to improve the signal to noise ratio. As noise is random fluctuations overlaying a US waveform, autoencoders are unable to learn a relationship between the noise and

Table 2
A review of features extracted from US measurements in the time, frequency, or time–frequency domain for ML.

Category of features	Features	Description	References
Specific amplitudes in function	Maximum amplitude	The largest amplitude in the investigated function interval	[28,40,53,133,140]
	Minimum amplitude	The smallest (or largest negative) amplitude in the investigated function interval	[15,53,133,140]
	Peak-to-peak amplitude	The difference between the largest and smallest amplitude in the investigated function interval	[28,52,140]
Dispersion of amplitudes along the function	Standard deviation	A measure of the dispersion of amplitude values with respect to the mean	[15,140]
	Variance	A measure of the dispersion of amplitude values with respect to the mean	[28,29,133]
	Temporal spread	A measure of the dispersion of amplitude values with respect to the mean	[40]
	Temporal entropy	A measure of amplitude variability along the function	[24,40]
	Spectral standard deviation	A measure of the dispersion of frequency amplitudes with respect to the mean	[28]
	Spectral smoothness	The variability of frequency amplitudes with respect to their neighbouring amplitudes	[42,43]
	Spectral spread	The variance of frequency amplitudes with respect to the average	[24]
	Spectral entropy	The amplitude variability along a frequency domain function	[24]
Measures of the rise or descent of function amplitudes	Temporal roll-off	The time value at which 90 % of the signal energy is concentrated	[40]
	Logarithmic decrement	The logarithmic decrease of amplitudes in a function	[41]
	Temporal slope	A measure of the rate of decrease in function amplitudes	[28,43]
	Descent time	A time value. For example, the time at which the slope of amplitude descent crosses zero.	[28,29,43]
	Lower 25 % of power spectrum	Fraction of total energy between lower 25 % level and peak frequency amplitude	[28]
	Upper 25 % of power spectrum	Fraction of total energy between peak and upper 25 % level frequency amplitude	[28]
	Rising time	The time value for function increase from 25% level amplitude to peak	[29]
Energy	Temporal energy	The sum squared amplitude of the waveform interval investigated	[15,16,24,27,29,41-43,53,151,152]
	Spectral energy	The sum squared amplitude of the frequency domain interval investigated	[24]
	Temporal inertia	Weighted average of the signal amplitude in time domain	[40]
	Mean	The mean amplitude in a function	[15,28,133,140]
	Sum absolute amplitude	A measure that gives lesser weight to large amplitudes compared with the energy	[27]
	Median	The median amplitude in a function	[140]
Crest factor	Temporal crest factor	The magnitude of the maximum signal amplitude in the time domain compared to the average	[24,42,43]
	Spectral crest factor	The magnitude of the dominant frequency compared with the average	[24,40,42,43]
Average frequency	Mean frequency	The mean frequency value	[28,133]
	Spectral centroid	The frequency value where half of the waveform energy is contained	[24,28,29,40,133]
Temporal position	Time centre	The centre of the function in the time domain	[29]
Temporal duration	Pulse duration	The length of time between the start and end of the waveform	[29,151]
Bandwidth	Measured bandwidth	The range of the measured frequency values	[28,29,151]
Skewness	Temporal skewness	A measure of the lack of symmetry in the waveform	[15,151]
	Spectral skewness	A measure of the lack of symmetry in the frequency domain	[24,29]
Kurtosis	Temporal kurtosis	A measure of the tailedness of the waveform	[15]
	Spectral kurtosis	A measure of the tailedness of the frequency domain function	[29]
Amplitudes at sample points in the waveform		Using the amplitude at each sample point in a waveform as individual features	[16,27,35,148,154]
Variations in features between consecutive waveforms	Standard deviation of the energy	A measure of the dispersion of consecutive waveform energy values	[53]
	Frequency analysis of consecutive amplitudes	A measure of the dispersion of consecutive amplitude values	[149]
	Features extracted from velocity variations	Measures of the dispersion of consecutive velocity values	[150]
Time of flight		The length of time for a sound wave to travel through a material	[53,81,152,153,155]
Feature gradients		A measure of the current time step feature with respect to previous time steps	[27,53]
Other features	Temperature	The process temperature	[24,43,81]
	Mass flow rate	The process mass flow rate	[43,81]
	Pressure	The pressure of the process	[81]
	Material type	Information about the material being processed	[81]

Table 3
Hyperparameters used for ANNs with features from US measurements.

Reference	Number of input features	Number of neurons in each hidden layers	Training algorithm	Additional information
[172]	Varied between 128 and 512	128		Root mean square error goal of 0.01 during training Learning rate and momentum term varied
[173]		5	Levenberg–Marquardt	
[153]	2	10		Training stopped after mean squared error of 1×10^{-6} or 6,000 iterations reached Learning rate of 0.6 Momentum rate of 0.4
[152]	2	10		Training stopped after 20,000 iterations or 1×10^{-6} mean squared error reached Learning rate of 0.6 Momentum rate of 0.4
[64]	5	50	Levenberg–Marquardt	Training stopped after a maximum of 500 iterations or desired minimum mean square error of 5.0 $\times 10^{-4}$ reached
[64]	4	20	Levenberg–Marquardt	Training stopped after a maximum of 500 iterations or mean square error of 5 $\times 10^{-4}$ reached
[28]	24	40		Stopping criteria were: maximum epochs 500, minimum error gradient equal to 1×10^{-5}, minimum mean square error equal to 1×10^{-5}
[144]	Varied between 1 and 7	Varied between 6 and 22		Learning rate of 0.2 Additional momentum 0f 0.5
[145]		12		Learning rate of 0.05
[174]	2	3, 3, 1	Evolutionary optimisation	Learning rate of 0.5, Mutation rate of 0.04, population size of 50, and cross over rate of 0.2
[175]	10	10, 2	Levenberg–Marquardt	
[132]	32	100		Training stopped when accuracy of 1×10^{-3} reached
[143]		32, 12	Scaled conjugate gradient	
[176]	4	3 hidden layers	Evolutionary algorithm	Population of 50 individuals Crossover probability of 0.95 Mutation probability of 0.01 200 generations
[133]	8	8, 25	Scaled Conjugate Gradient	Training continued until error goal of 1×10^{-2} achieved
[41]	5	2		
[146]		10	Scaled Conjugate Gradient	
[24]	6	11		
[42]	5	2		
[43]	9	14		
[177]		10	Scaled Conjugate Gradient	
[40]	3	9		Training continued until error was below 0.1 % accuracy
[178]	4	5, 2	Bayesian regularization Levenberg–Marquardt	Network trained 100 times and weights with lowest score on the validation set were used Training was stopped once the Summation of Squared Errors reached below 10^2 or 10^1 depending on the prediction task, or 1000 epochs were reached
[139]	3	4	Levenberg–Marquardt	
[148]	502	980, 270		3 dropout layers with 0.5 probability Trained for 3000 epochs
[179]	151	10	Levenberg–Marquardt	
[180]		50	Levenberg–Marquardt	Trained for 5 epochs
[35]		3 hidden layers		Dropout layers with 0.5 probability Trained for 500 epochs
[154]	11,501	502		Trained for 500 epochs
[154]	11,501	4 hidden layers		
[181]	5	5		
[182]	3300	1300, 660, 330, 165	Scaled Conjugate Gradient	ReLu activation function Learning rate of 0.001 Trained for 1000 epochs
[15]	7	7	Bayesian Regularization	
[27]	Various	Determined through grid-search	Levenberg–Marquardt for regression Scaled Conjugate Gradient for classification	Early stopping applied with a validation patience of six iterations Ten networks were trained and their scores averaged A grid search determined L2 regularisation magnitude
[183]		1024, 512, 265, 128	Adam	ReLu activation function used Dropout layers with probabilities 0.2, 0.3, 0.4, 0.5 Learning rate of 0.0001 Batch size of 8 Trained for 10,000 epochs
[183]		2048, 2048, 1024, 1024	Adam	ReLu activation function Dropout layers with probabilities 0.2, 0.2, 0.2, 0.2 Learning rate of 0.0001 Batch size of 8 Trained for 10,000 epochs

(*continued on next page*)

Table 3 (*continued*)

Reference	Number of input features	Number of neurons in each hidden layers	Training algorithm	Additional information
[184]	5000	1000	Learning rate of 0.001	Dropout layers with 0.7 probability before and after each fully connected layer ReLu activation function
[184]	5000	1000, 1000	Learning rate of 0.001	Dropout layers with 0.7 probability before and after each fully connected layer ReLu activation function
[185]	24	6, 10, 2	Levenberg–Marquardt Algorithm	1000 epochs

the signal and so fail to reconstruct noise during training. Ref. [166] used simulations and a small number of experimentally collected samples to train an autoencoder to reconstruct full wavefield data from sparsely sampled US measurements. This allows US data to be collected at lower sampling frequencies and artificially reconstructed as higher sampling frequency waveforms, thus reducing measurement acquisition time. Ref. [167] trained autoencoders to reconstruct flawless US signals so that when a flaw is detected the autoencoder fails to reconstruct the waveform. This allows flaws to be identified even if they overlap the initial transducer pulse. Ref. [168] used stacked autoencoders to localise and classify acoustic emission sources in riveted panels.

As comparatively few feature selection methods have been used with US sensor measurements, the interested reader is directed to [169–171] for further information on techniques available such as wrapper, filter, hybrid, and embedded methods.

4.3. Algorithms

This section provides a review of the hyperparameters used in ML models with US measurements. Table 3 reviews the hyperparameters used for ANNs with US measurements. These can determine the structure of the network (e.g. the number of hidden layers and the number of neurons in each hidden layer) or the training of the network (e.g. training algorithm used and the learning rate).

Similar to ANNs, CNNs also require hyperparameter selection to decide the structure of the network and how it trains. However, the hyperparameters in the convolutional layers, which detect features, and pooling layers, which downsample the data, must also be chosen. Table 4 reviews the hyperparameters used for CNNs combined with US measurements. LSTMs also require selection of similar hyperparameters to ANNs in addition to the number of LSTM units to employ. Ref. [27] used 50 LSTM units, a fully connected layer of 50 neurons, a dropout layer with a probability of 0.5. The network was trained for 60 epochs using the Adam optimisation algorithm, a learning rate of 0.01, a batch size of 2 and a gradient threshold of 1 to prevent exploding gradients. Ref. [182] used the Scaled Conjugate Gradient optimisation algorithm and a learning rate of 0.01 for 400 epochs. Two dropout layers with probabilities of 0.1 and 0.2 were used. Ref. [186] used 32 LSTM units, two fully connected layers (with 512 and 128 neurons), and two dropout layers with probabilities of 0.25 and 0.2. Training was carried out for 500 epochs with a learning rate of 5×10^{-5} and a batch size of 8. Ref. [138] used 6 LSTM units and [187] used 7. Ref. [188] combined CNNs with two layers of 8 Gated Recurrent Units (GRU) along with dropout probabilities of 0.5 to extract temporal features from US waveforms. GRUs are similar mechanisms to LSTMs only simpler with two gates rather than three and therefore, generally, lower performance when learning long sequences. Ref. [189] trained a ConvLSTM encoder-decoder DNN on finite element simulations of 2-D US wave propagation. ConvLSTMs allow the learning of spatio-temporal dependence in input sequences by employing convolutional structures within the LSTM units. The trained model was comparable in accuracy to finite element simulations but faster to solve by approximately an order of magnitude through negating the computation of numerical calculations. Ref. [190] used an LSTM layer following a CNN for damage detection of copper pipelines using laser ultrasonic scanning.

CNNs have also been used for other applications such as for B-scan US images [208], combining multiple B-scan images [209], C-scan images [210], and guided waves [211]. Ref. [199] used a CNN to deconvolve overlapping US signals and extract the time of flight and amplitude. Ref. [212] used 3D CNN for defect detection by using US images of wave propagation from multiple time steps. Refs. [213,147] presented a 22-layer GFresNet and GFresNET-2D for guided-wave focusing defect signal classification, respectively. Ref. [214] employed a U-net style CNN for predicting the material thickness of plate-like structures using acoustic steady-state excitation spatial spectroscopy. Ref. [215] used a CNN for corrosion inspection on an aluminium plate using broadband Lamb waves.

For models using support vectors, the most commonly tuned hyperparameters include C, the penalty factor, γ or σ, the influence a single training example has, and the type of kernel used, e.g. linear, polynomial, or radial. In support vector regression, epsilon defines the distance from the fitted function where the error cost of datapoints is not counted. Hyperparameters used for support vector models and US measurements are reviewed in Table 5. Decision trees can require a choice in the number of trees used (500 [16], maximum tree depth (1 [16], 3 [44], 4 [16], the minimum number of leaf instances (10 [16]), the learning rate (1 [16]), or the maximum number of splits (4, 20, 100 [216]). The type of ensemble method can also be chosen, such as Adaboost used in [16]. Furthermore, K-nearest neighbour algorithms require a selection of the number of neighbours to use, such as 5 [216], 11 [217], 25 [44], 50 [44], or 105 [16].

Gaussian Processes (GP) are an algorithm for classification or regression which interpolates datapoints with normal distributions and thereby provides confidence intervals for its predictions. GPs have recently been applied to measure oil film thickness in journal bearings using US measurements [230] and to localise acoustic emission sources in SHM by measuring the time of flight with multiple sensor pairs [231]. Ref. [232] used data-driven GPs to model guided waves in composite materials. Physical knowledge of the system was inputted into the model by specifying constraints of the GP's kernels such as rotational symmetry, exponential decay for viscoelastic damping, and attenuation due to geometric spreading using a polynomial kernel. Ref. [233] used Bayesian linear regression to decompose guided wave signals into individual modes to enable damage sizing and localisation following two-dimensional Fourier transformation. Ref. [234] used a GP to predict thermal barrier coating porosity. Input features to the GP were first selected by evaluating a neural network using different feature combinations.

4.4. Out-of-distribution detection

Out-of-distribution (OOD) detection methods are used to identify datapoints that fall outside the range of normally expected values. For example, [235] used OOD detection to identify damage in wind turbine blades. During feature extraction, DWT, Fast Fourier Transform and PCA were used. One-class classification algorithms were used including support vector machine data description, K-means, and Euclidean distance measures. Autoencoders are commonly used for OOD detection. If anomalous data is passed through as encoder, the distribution of the latent space variables will be different to the training data and the

autoencoder fails to reconstruct the input. Ref. [167] used autoencoders to allows flaws to be identified even if they overlap the initial transducer pulse. Ref. [236] used autoencoders in a similar network structure for a Ganomaly approach. Furthermore, they used another generative model, normalizing flows, to learn transformations between normal and anomalous samples for OOD detection. Ref. [237] used variational AEs as an OOD detection system to detect defects from ultrasonic B-scans. In this study, a second encoder was added after the decoder and was found to provide increased accuracy in detecting defects. Ref. [238] compared autoencoders to one-class SVMs, isolation forests, and hidden Markov models.

4.5. Data augmentation

Data augmentation can be used to artificially increase the size of a dataset. This can be particularly useful when training deep learning models (e.g. DNNs, CNNs, and LSTMs) which can require many training instances to tune all the model parameters. Ref. [35] time-shifted US signals forwards and backwards by 5 and 10 μs forward to increase the dataset size by four times. Similarly, [162] time-shifted US signals forwards and backwards by 6×10^{-3}, 10×10^{-3}, 14×10^{-3}, and 20×10^{-3} μs. Ref. [239] laterally translated signals and used magnification to increase the dataset size by five times. Ref. [205] added six different levels of Gaussian white noise to US signals ranging from 20 to 30 dB. Ref. [240] rotated samples by 90°, 180°, and 270° for damage localisation in plate-like structures. Ref. [204] extracted parts of US waveforms received from measuring flawed samples and inserted them into flawless signals. In this way, virtual data could be created by implanting flaws into different locations of a test section. The choice of data augmentation techniques, such as lateral translation, magnification, or noise addition, must be decided based on the application it is being used for. For example, lateral translation could not be used if the time of flight is an important parameter to measure during the process being monitored. Refs. [35,162] were classifying weldment defects. As such, as shift in the time domain would only represent a change in depth of the flaw rather than its presence or type. Ref. [239] used data augmentation for US waveform feature learning in a CNN. The CNN was trained on an auxiliary task to classify the dataset membership of previously collected US measurements. The pretrained CNN weights were then used as a feature extractor on new US measurement datasets. Therefore, the lateral translation and magnification did not represent a change in any physical parameters of a system. Ref. [241] doubled the size of an experimental dataset by reversing US images of defects around the vertical axis. Ref. [207] used time shifting and the addition of white Gaussian noise for US flaw classification in weldments. Ref. [202] added white Gaussian noise to create three datasets with signal-to-noise ratios of 5, 10, and 20.

4.6. Semi-supervised learning

[242] used a hierarchical clustering algorithm to detect whether pipe sections were damaged or undamaged. This is traditionally an unsupervised learning method which divides the input data into the number of clusters specified (in this case, two). To label each of the clusters, only one labelled instance of an undamaged pipe was required to perform the classification. Similarly, [243] used a k-means clustering technique to monitor the growth of simulated cracks in pipes. An alarm threshold was developed to trigger when the size of the defect becomes critical based on the distance of the US measurement from each cluster. Ref. [244] presented a semi-supervised Gaussian mixture model which was updated through the expected maximisation algorithm over both the labelled and unlabelled data.

4.7. Active learning

Active learning uses methods to select unlabelled datapoints that would have the most benefit to model performance if labelled thereby minimising the total number of datapoints to be labelled. Ref. [245] used active learning to improve a probabilistic mixture model initially trained on a small sample of labelled data. This approach was evaluated on three datasets: Z24 Bridge data, a machining acoustic emission dataset, and data from ground vibration aircraft tests. Ref. [246] used the expected value of perfect information for SHM on a numerical case study and the Z24 Bridge benchmark.

4.8. Generative models

Generative Adversarial Networks (GAN) and Variational Autoencoders (VAE) are methods to produce realistic data from random inputs. The generator component of a GAN is trained by aiming to fool the discriminator component into determining whether its input data are real or synthetic. During training, a VAE learns a probability distribution of the input data in the latent space from which new samples can be drawn. Ref. [247] used a GAN for generating defects in US B-scan images and successfully increased defect detection from 70 % to 76 % when combining real and synthetic images. Ref. [248] compared two GAN structures to produce images of defects in US signals and confirmed that the generated images could not be identified by human experts. Ref. [249] used GANs to increase the size of Finite Element simulation datasets for welding defect detection. The highest defect detection accuracy was achieved by supplementing the generated data with noise derived from experiments and extracted using the sliding kernel approach. Ref. [250] also used GANs to generate B-scans from simulated US data for non-destructive evaluation applications. Ref. [251] used GANs to create synthetic acoustic emission spectrograms in the US range for detecting cavitation in hydraulic turbines. This increased cavitation detection accuracy from 94.2% using CNNs alone to 95.1%.

4.9. Transfer learning

Transfer learning encompasses methods which transfer knowledge learned from one task to different, but similar, tasks. Ref. [241] trained a Faster-CNN to detect defects in US images on simulated datasets first before training on a small set of experimental data. This greatly reduced the loss function compared with training on the experimental datasets alone. Ref. [252] presented an experimental dataset of 7004 ultrasonic images collected from 18 stainless steel plates and evaluate the performance of many pre-trained CNNs. They conclude that their dataset may be used by others for pre-training their own CNN models. Ref. [253] transferred fixed layers of a neural network trained on easier-to-classify tasks to more difficult cases for damage localisation on an aircraft wing. Ref. [254,255] used a pre-trained VGG16 to classify acoustic emission sources following CWT by training the last convolutional layer, two fully connected layers, and output layer, and the output layer only, respectively. Similarly, [256] updated the last layer of a pre-trained ResNet34 model for vibration data for SHM. Ref. [257] pre-trained a CNN on compressed vibration data in the form of a histogram of response thresholds for an SHM application. This CNN was then fine-tuned using extremely compressed, smoothed histogram data in the form of a mean, variance, and scale factor.

4.9.1. Domain adaptation

An ML model trained on one task (source domain) will predict poorly on a second task (target domain) if the feature distributions between the domains change. US waveform features may be different across domains due to differences in the path of the sound wave or the materials being monitored. Even for similar processes, differences in the sensor used, attachment procedure, or contact pressure may alter the feature distributions [20]. Domain adaptation is a subcategory of transfer learning which alters how a ML model trains so that it predicts accurately across both domains. Ref. [20] used unlabelled domain adaptation of a single-feature waveform feature to apply a trained ML model to new, similar

Table 4
A review of hyperparameters used for CNNs with US measurements.

Reference	Convolutional layers, number of filters, size of filters	Pooling layers	Fully connected layers	Training	Additional information
[134]	2 16 7 × 5 32 5 × 3	2 × 2 max pooling	128 neurons and an SVM final layer		
[135]	3 5 5 × 1 8 8 × 1 16 7 × 1	Max pooling: 2 × 1, stride 1 × 1	2 150 and 75 neurons	Adam optimisation algorithm Learning rate varied between 10^{-5} and 1 Trained for 100 epochs	ReLu activation function
[135]	2 6 10 × 2 16 10 × 2	2 × 2 max pooling	2 500 and 250 neurons	Adam optimisation algorithm Learning rate varied between 10^{-5} and 1 100 epochs	
[191]	2 16 10 × 1 32 10 × 1		2 200, 200		ReLu activation function
[192]	3 32 3 × 3 32 3 × 3 64 3 × 3	2 × 2 max pooling	64 neurons	Batch size of 8 3 epochs	Three dropout layers with 0.5 probability ReLu activation function
[188]	6 layers Combined with GRUs			Batch size of 512 300 epochs Adam optimisation function Initial learning rate of 0.001 Learning rate reduced every 50 epochs Final learning rate of 0.0001	ReLu activation function Batch normalisation used Dropout rate of 0.5
[35]	2 32 16 × 1 filter size, 8 × 1 stride 64 3 × 1 filter size, 2 × 1 stride	No pooling layer between 1st and 2nd convolutional layers Max pooling, 2 × 1 size with 2 × 1 stride	300 neurons	Trained for 500 epochs	Padding: "Same" for convolutional layers, "valid" for max pooling layer Activation function: Elu in convolutional layers, ReLu in fully connected layers Two dropout layers, 0.25 and 0.5 probability
[27]	2 8 5 × 5 16 5 × 5	2 × 2 max pooling		Adam optimisation algorithm Learning rate 0.01, drop factor of 0.33 after 4 epochs Trained for 8 epochs in total Batch size of 256, shuffled every epoch	"Same" padding Batch normalisation ReLu activation function Single dropout layer, probability varied between 0.1 and 0.5
[162]	2 32 25 × 1, 8 × 1 stride 64 3 × 1, 2 × 1 stride	No max pooling between 1st and 2nd convolutional layers Max pooling: 2 × 1 size, 2 × 1 stride			Three dropout layers with 0.7, 0.5 and 0.5 probability, respectively
[136]	2 16 3 × 3 512 3 × 3	2 × 2 max pooling	SVM final layer	Learning rate of 0.1	Batch normalisation ReLu activation function Single dropout layer with 0.5 probability
[193]	7 Structure based on VGGNet	3 max pooling layers			
[194]	3 32 5 × 1 64 5 × 1 96 5 × 1	Max pooling: 2 × 1, stride of 2	1000		
[195]	2 30 3 × 3 in both	Max pooling: 3 × 3, stride of 2	56, 28	Learning rate of $1e^{-6}$ Trained for 5000 iterations	ReLu activation function
[196]	21 convolutional layers across three channels Filters ranging in size from 1 × 1 to 12 × 20			Adam optimisation algorithm Trained for 6 epochs Learning rate of 0.001 Batch size of 32	Batch normalisation Leaky ReLu activation function
[197]	4 32 3 × 3 32 3 × 3 64 3 × 3 1 1 × 1	1 × 2 max pooling			Batch normalisation ReLu activation function
[198]	3 48 3 × 3 96 3 × 3 192 3 × 3	2 × 2 max pooling	64	Adam optimisation algorithm Learning rate of 0.001 Batch size of 128	10 % dropout rate before and after fully connected layer ReLu activation function

(*continued on next page*)

Table 4 (*continued*)

Reference	Convolutional layers, number of filters, size of filters	Pooling layers	Fully connected layers	Training	Additional information
				400 epochs Validation patience of 150 epochs	
[199]	1 64 filters			Stochastic gradient descent optimisation algorithm used Initial learning rate of 0.005 and decreased logarithmically 30,000 epochs	Dropout rate of 0.5
[200]	1 8 5 × 5	3 × 3 max pooling	2 1024 and 64 neurons	Adam optimisation algorithm	
[201]	5 Inception blocks with 1 × 1, 3 × 3, and 5x5 filters	3 × 3 max pooling Stride of 2	3	Momentum of 0.9 Weight decay of 0.0002 Stochastic gradient descent optimisation algorithm Initial learning rate of 0.001 160 epochs	Batch normalisation and ReLu activation function before each convolutional layer Dropout ratios of 0.5
[147]	4 ResBlocks			Adam optimisation algorithm Learning rate of 0.0001 20 epochs Batch size of 30	
[202]	3 convolutional blocks			Adam optimisation algorithm Learning rate of 0.01 Batch size of 4 500 epochs	
[137]	3 16 3 × 1 32 3 × 1 64 3 × 1	2x1 max pooling	16 neurons	Learning rate of $1e^{-4}$ Adam optimisation algorithm Trained for 250 epochs Batch size of 64	ReLu activation function
[137]	5 16 3 × 1 32 3 × 1 64 3 × 1 128 3 × 1 256 3 × 1	2x1 max pooling	2 512 and 128 neurons	Learning rate of $1e^{-5}$ Adam optimisation algorithm Trained for 500 epochs Batch size of 32	ReLu activation function Two dropout layers with probabilities of 0.25 and 0.2, respectively
[203]	5 16 3 × 1 32 3 × 1 64 3 × 1 128 3 × 1 256 3 × 1	2x1 max pooling	128 neurons	Adam optimisation algorithm Learning rate of $1e^{-5}$ Trained for 2500 epochs Batch size of 128	ReLu activation function Single dropout layer with 0.25 probability
[203]	5 16 3 × 1 32 3 × 1 64 3 × 1 128 3 × 1 256 3 × 1	2 × 1 max pooling	128 neurons	Adam optimisation algorithm Learning rate of $1e^{-5}$ 500 epochs Batch size of 32	ReLu activation function Single dropout layer with 0.25 probability
[204]	4 96 3 × 3 64 3 × 3 48 3 × 3 32 3 × 3	Max pooling with varying sizes of 7 × 1, 2x8, 3 × 4	14 neurons		ReLu activation function
[184]	32 50 × 1, stride of 5 × 1 64 4 × 1, stride of 2 × 1	2 × 1 max pooling, stride of 2 × 1	1000 neurons	Learning rate of 0.001	2 dropout layers, before and after full-connected layer, probabilities of 0.7 ReLu activation function
[205]	Dual-headed convolutional neural network 4 convolutional layers in each head 576 11 × 1 484 11 × 1 500 5 × 1 324 5 × 1	4 × 1 max pooling	1 fully connected layer in each head, 256 neurons in each Final fully connected layer with 196 neurons	Batch size of 24 Trained for 70 epochs	Two dropout layers in each head, 0.2 and 0.3 probabilities ReLu activation function L2 regularisation value of 0.01 Early stopping with patience of 10 epochs
[130]	1 64 3 × 3	2 × 2 max pooling	512 neurons	Adam optimisation algorithm Learning rate of 0.003 Trained for 30 epochs Batch size of 32 Momentum of 0.9	ReLu activation function
[206]	CNN base on WaveNet Dilation rate of 3 × 1 filters increased to 3, 9, 27, 81 through			Trained for 900 epochs	"Casual", "valid", and "same" padding used

(*continued on next page*)

Table 4 (*continued*)

Reference	Convolutional layers, number of filters, size of filters	Pooling layers	Fully connected layers	Training	Additional information
	four residual blocks 16 filters in each convolutional layer				
[207]	19-layer ResNet	1 × 2 max pooling	30o neurons	Adam optimisation algorithm 500 epochs	ReLu activation function Dropout probability of 0.5

Table 5
Hyperparameters used with support vector models and US measurements.

References	Information
[218]	Radial basis function (RBF) kernel C varied between 1 and 32 γ varied between 0.00049 and 0.5
[29]	C = 1000 γ = 10 Gaussian kernel Kernel Fisher discriminant used to optimise parameters
[219]	Linear kernel with C = 10 2nd order polynomial kernel with C = 0.1 RBF kernel with C = 10 and γ = 1
[220]	RBF and linear kernels γ = 20
[221]	Linear kernel, polynomial kernel, RBF kernel, and sigmoid kernel Range of C values tested: 1, 10^1, 10^2, 10^3, 10^4, 10^5
[42]	Third order polynomial kernels used
[43]	RBF kernel γ = 0.7
[222]	Epsilon, C, and γ determined through particle swarm analysis
[223]	Linear, polynomial, and RBF kernels used C varied between 0.001 and 100 in increasing powers of 10 Polynomial degree evaluated between 2 and 5 γ for RBF kernel evaluated between 0.01 and 100 in increasing powers of 10
[224]	Linear, quadratic, RBF, and polynomial kernels tested γ values tested: 4, 5, 6, 7 Polynomial orders tested: 2, 3, 4
[177]	Linear, quadratic, polynomial, multilayer perceptron and RBF kernels tested A 3rd degree polynomial kernel function was used
[225]	C = 2 γ = 10,000 RBF kernel
[129]	Binary tree SVM C and γ varied
[180]	Bias = 42.57 Box constraint = 9.4885 Epsilon = 0.9489 Number of iterations = 64 RBF kernel
[141]	RBF kernel C and γ parameters determined using particle swarm optimisation
[216]	Linear, polynomial and RBF kernel tested. Linear kernel performed best. The best box constraint was found to be 1.
[181]	Linear and RBF kernels investigated
[226]	RBF kernel C and γ optimised using Particle Swarm Optimisation
[227]	RBF kernel C = 0.1 γ = 0.3
[16]	C = 0.0001
[44]	C = 0.001
[27]	Bayesian optimisation for 60 evaluations to select box constraint value, kernel scale, kernel function, polynomial order, and whether the inputs were standardised The expected improvement acquisition function was used
[228]	C and γ optimised using Bat Optimisation Algorithm
[229]	Cross-validation to determine C and γ RBF kernel

processes (investigated for two mixing and three cleaning processes) without needing to label data in the new domain. This is therefore a method of eliminating the data labelling burden in a factory environment when applying a US sensor to a new, similar process. Combining multiple source datasets or using datasets collected from a wider range of process parameters (e.g. temperature variations) enabled the models to better adapt to changes in feature distributions. Ref. [258] used labelled domain adaptation, where a reference measurement is available in the target domain, to use previously collected data from a laboratory fermentation process to reduce the development time of accurate ML models for an industrial process. Three methods were investigated to train DNNs with LSTM layers: simultaneous training over both datasets, federated learning using both datasets, and fine-tuning of the previously trained models on the target domain dataset. Federated learning was investigated as a method for potentially using datasets from multiple companies while maintaining data privacy. Federated learning shares network weights between a local models from each dataset and a global model rather than sharing the real data. All methodologies provided an increase in prediction accuracy over solely using the industrial fermentation dataset. Federated learning provided the highest increase in accuracy by allowing further convergence to minima during training. Ref. [259] transferred knowledge from microseismic data of earthquake studies to acoustic logging tools for collecting borehole information in oil fields. The purpose of this study was to overcome the effects of noise on accurate determination of the time-of-flight using pulse-echo transducers. The maximum mean discrepancy was used to align feature spaces and a convolutional autoencoder was used to ensure class discriminant features were extracted. Ref. [260] used a dictionary learning method to use simulated US wavefields to isolate damage wavefields in experimental data from plate structures. Firstly, a dictionary is learned to optimally reconstruct the simulated wavefields and then transferred to the experimental data. Secondly, the reconstructed experimental data is aligned with the experimental data to account for changes in domain between the simulated and experimental datasets.

Ref. [261] used Transfer Component Analysis (TCA) to transfer damage detectors between experimental datasets from three aircraft tailplanes. The data from both domains were matched in a latent space and two Transfer Components were extracted. Ref. [262] used metric-informed joint domain adaptation to overcome the problem of pre- and post-repair changes in data distribution in SHM. Joint domain adaptation aligns both the marginal and conditional distributions in a latent space using pseudo-labels from the target domain. In this work, the Mahalanobis squared distance was used to select the data for pseudo-labelling. Ref. [263] used a domain-adapted Gaussian mixture model (DA-GMM) to transfer labelled information between two bridge datasets for SHM. A linear mapping was used to transform the target domain data and the model was optimised using an expectation maximisation technique. Ref. [264] presented a kernelised Bayesian transfer learning (KBTL) approach for SHM applications. The approach maps data from each domain onto a shared latent space where labelled data from the source domain is used to classify the data from the target domain. This method may be used to transfer labelled data from uncommon damage types from similar structures or simulations to reduce the burden of labelling these rare states in the target domain. Ref. [265] used Balanced

Table 6
An explanation of the recommendations presented in Table 7.

Recommendations	Explanation
Other sensors	Other in-line or on-line sensors as a reference measurement
Semi-supervised	Semi-supervised learning to pseudo-label the unlabelled data
Sampling	Sampling and off-line sensors as a reference measurement
US measurements	US measurements to infer process state
Unlabelled domain adaptation	Unlabelled domain adaptation from similar processes
Labelled domain adaptation	Labelled domain adaptation from similar processes
Coarse time dome features	E.g., specific amplitudes in a function, dispersion of amplitudes along the function, measures of the rise or descent of function amplitudes, energy, crest factor, kurtosis, skewness, temporal duration
Convolutional feature extraction	Convolutional feature extraction methodology as presented in [239]
Feature variations	Variations in features between consecutive waveforms

Distribution Adaptation to transfer damage localisation models between different types of aircraft wings. Three metrics were evaluated to determine the structural and data similarities between domains before domain adaptation. This was to minimise the risk of negative transfer where transfer learning leads to a decrease in model performance.

4.10. Object detection

Object detection is a type of image analysis where a CNN identifies the presence, location, and class of objects and has been used for defect detection in ultrasonic images. Ref. [241] used a Faster R-CNN to identify, locate, and size defects using simulated data and a small experimental dataset. Ref. [266] presented EfficientDet and a method to select anchor hyperparameters for the high aspect ratios expected. They achieved a 9 % increase in accuracy (89.6% accuracy overall) compared to the YOLOv3 architecture used in [267]. Ref. [268] presented DefectDet which uses a lightweight encoder-decoder feature extractor and a detection head with custom anchor box aspect ratio and stride to detect high aspect ratio defects. Ref. [269] presented two methods of using multiple ultrasonic B-scans for defect detection by merging feature maps using convolutional and convolutional LSTM layers.

5. Future directions - interpretability in machine learning

A particular barrier to implementation of sensor and ML combinations is the perceived lack of explainability and interpretability of ML models and therefore a lack of trust it their predictions. These issues must be addressed to increase buy-in from companies and operators, and, furthermore, to meet any potential regulation criteria that requires adequate transparency in the ML prediction process. There are three points in the supervised ML pipeline where these problems can be considered. The first is during feature extraction and selection. Inherently explainable features such as the waveform energy (typically a measure of the reflection coefficient, and therefore acoustic impedance, or attenuation in a system) or the time of flight (a measure of the speed of sound through a material) may be preferred over other, more abstract features acquired from a US waveform. Feature selection methods may be used to reduce the number of features to make models simpler or calculate the importance of each feature and thereby make the models more interpretable. Secondly, transparent algorithms could be used such as linear or logistic regression, decision trees, or k-nearest neighbour models [270]. Predictions from linear and logistic regression could be accompanied by the weights applied to each feature used to make the prediction. In this way, the prediction process can be made fully explainable. Similarly, decision trees could produce the hierarchical decision process used and k-nearest neighbour models could present the k nearest training points used to inform the prediction made. Finally, post-hoc explanation of individual predictions can be used to understand the decision-making process [271]. Local interpretable model-agnostic explanations (LIME) perturb training data around a particular query point and build a transparent model (e.g. decision tree) correlating the new synthetic training data with model predictions to understand the decision-making process around this particular point. Shapley Additive explanations (SHAP) calculate the change in predictions by varying feature values at a particular data point to understand the impact of each feature on the prediction being made. For CNNs, additional techniques such as, gradients, class activation mappings, saliency maps, or occlusions can be used to indicate the datapoints contributing to a particular prediction.

6. Recommendations

Tables 6 and 7 contain recommendations for combining ML and US sensors for the reviewed process applications. To obtain labelled data in industrial environments, other in-line or on-line sensors can be used as a reference measurement for all the reviewed processes. Periodic sampling combined with off-line analysis could be employed for all processes other than cleaning, curing, or membrane fouling. Instead, for cleaning, the process would need to be ended at different stages, the equipment dismantled, and the sensor data labelled. During curing, sampling would not be possible due to the toughening or hardening of the material. For membrane fouling, no non-disruptive sampling of the fouling material could be performed. As such, semi-supervised learning could be used in all processes to pseudo-label the unlabelled data except for curing or membrane fouling processes where single data points cannot be collected. For all the processes reviewed, the US measurements could be used to infer process stages (such as the attainment of homogeneity during mixing, or the start of ethanol production during fermentation), apart from for tabletting. This is because during tabletting the final product is monitored opposed to the compaction process. For every process, unlabelled and labelled domain adaptation from similar processes could be used to reduce the data labelling burden in a factory environment. The recommended features include the coarse time domain features and convolutional feature extraction methods as compared in [239]. Fig. 3 displays the convolutional feature extraction method as an unrolled DNN with an LSTM presented in [239]. The coarse time domain features include specific amplitudes in the function, dispersion of amplitudes along the function, measures of the rise or descent of function amplitudes, energy, crest factor, kurtosis, skewness, and temporal duration. These features do not misattribute waveform variations of narrow frequency band US sensors to changes in frequency content as do the Fourier or wavelet transformations [239]. These features also overcome the problem of lateral sample point shifting of waveforms due to temperature changes as suffered by using the amplitudes at sample points as features directly [239]. The time of flight can be used for all processes to monitor the speed of sound throughout the material. The variations between consecutively acquired US waveforms can be used to monitor product quality variation in extrusion and injection moulding, the degree of homogeneity in mixing processes, or the presence of CO_2 bubbles during fermentation. Feature gradients along with LSTMs can be used for all processes that progress over time to

Table 7
Recommendations for combining ML and US measurements with the reviewed processes.

	Cleaning	Fermentation	Crystallisation	Mixing	Extrusion	Injection moulding	Curing	Reaction monitoring	Tabletting	Membrane fouling
Obtaining labelled data	Other sensors Ending the process at different stages and using off-line sensors as a reference measurement Semi-supervised learning	Other sensors Sampling Semi-supervised learning	Other sensors Sampling Semi-supervised learning	Other sensors Sampling Semi-supervised learning	Other sensors Sampling Semi-supervised learning	Other sensors Sampling Semi-supervised learning	Other sensors	Other sensors Sampling Semi-supervised learning	Other sensors Sampling Semi-supervised learning	Other sensors
Reducing data labelling burden	US measurements Unlabelled domain adaptation Labelled domain adaptation	US measurements Unlabelled domain adaptation Labelled domain adaptation	US measurements Unlabelled domain adaptation Labelled domain adaptation	US measurements Unlabelled domain adaptation Labelled domain adaptation	US measurements Unlabelled domain adaptation Labelled domain adaptation	US measurements Unlabelled domain adaptation Labelled domain adaptation	US measurements Unlabelled domain adaptation Labelled domain adaptation	US measurements Unlabelled domain adaptation Labelled domain adaptation	Unlabelled domain adaptation Labelled domain adaptation	US measurements Unlabelled domain adaptation Labelled domain adaptation
Feature extraction	Time of flight Coarse time domain features Feature gradients Other features available (e.g. temperature, material type, mass flow rate) Convolutional feature extraction	Time of flight Coarse time domain features Feature gradients Other features available (e.g. temperature, material type) Convolutional feature extraction Feature variations	Time of flight Coarse time domain features Feature gradients Other features available (e.g. temperature, material type) Convolutional feature extraction	Time of flight Coarse time domain features Feature gradients Other features available (e.g. temperature, material type) Convolutional feature extraction Feature variations	Time of flight Coarse time domain features Other features available (e.g. temperature, material type, pressure, mass flow rate) Convolutional feature extraction Feature variations	Time of flight Coarse time domain features Other features available (e.g. temperature, material type, pressure, mass flow rate) Convolutional feature extraction Feature variations	Time of flight Coarse time domain features Feature gradients Other features available (e.g. temperature, material type) Convolutional feature extraction	Time of flight Coarse time domain features Feature gradients Other features available (e.g. temperature, material type) Convolutional feature extraction	Time of flight Coarse time domain features Other features available (e.g. temperature, material type) Convolutional feature extraction	Time of flight Coarse time domain features Feature gradients Other features available (e.g. temperature, mass flow rate, material type) Convolutional feature extraction
Algorithms	LSTMs	LSTMs	LSTMs	LSTMs	LSTMs for short time sequences of features ANNs or DNNs, using US waveforms	LSTMs for short time sequences of features ANNs or DNNs, using US waveforms	LSTMs	LSTMs	ANNs or DNNs	LSTMs

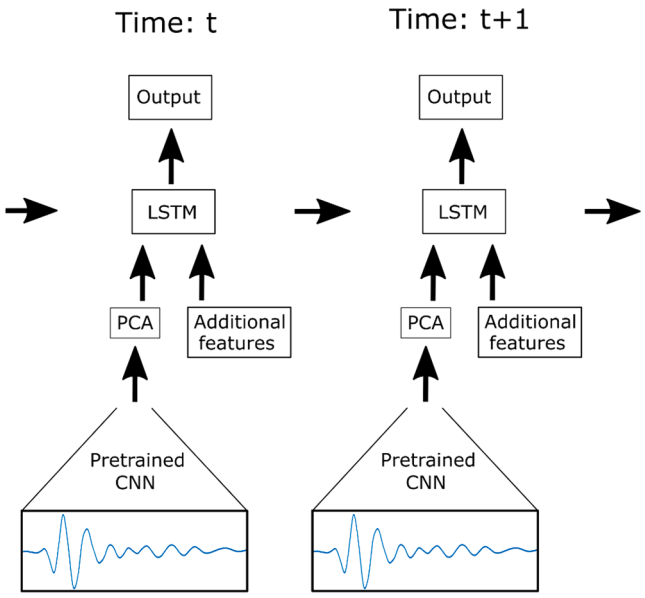


Fig. 3. A diagram of the convolutional feature extraction method presented in [239]. It is depicted as an unrolled DNN with an LSTM layer. A pretrained CNN extracts features from the US waveform which are then combined with additional features, such as the time of flight or deviations between consecutively acquired waveforms and inputted into the DNN.

incorporate knowledge from previous time steps. LSTMs can be used on short time sequences of features for extrusion and injection moulding to monitor variations in product qualities, or ANNs and DNNs can be used on single time-step features. ANNs or DNNs should be used for data that is not part of a time sequence due to their ability to construct new features from input data that correlate to output variables. This ability for complex fitting requires adequate regularisation to prevent over-fitting. ANNs and DNNs should also be used over CNNs as they allow for incorporation of other features such as the time of flight, mass flow rate, temperature, pressure, material type, or variations between consecutively acquired waveform features.

7. Summary

The manufacturing sector is increasingly using data to inform decision making. In-line and on-line sensors underpin this transition by automatically acquiring real-time data. Supervised ML techniques can be combined with US measurements and provide advantages over calibration procedures. However, their implementation is lagging due to expertise required to extract and select appropriate features from the sensor measurements, select the ML algorithm to use, and find a suitable set of model hyperparameters. The aim of this article is to facilitate the combination of ML and US measurements for in-line and on-line process monitoring or other similar applications. The article first reviews the use of US sensors for monitoring processes before reviewing the combination of US measurements and ML including literature from other sectors. This review covers feature extraction, feature selection, algorithm choice, hyperparameter selection, data augmentation, domain adaptation, semi-supervised learning and ML interpretability. Recommendations for applying ML methods for monitoring of the reviewed processes are also provided.

Funding: This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) standard research studentship (EP/R513283/1).

CRediT authorship contribution statement

Alexander L. Bowler: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Michael P. Pound:** Conceptualization, Methodology, Software, Supervision, Visualization, Writing – review & editing. **Nicholas J. Watson:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Visualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Ghobakhloo, Industry 4.0, digitization, and opportunities for sustainability, J. Clean. Prod. 252 (2020) 119869, https://doi.org/10.1016/j.jclepro.2019.119869.
[2] H.S. Kang, J.Y. Lee, SangSu Choi, H. Kim, J.H. Park, J.Y. Son, B.H. Kim, S.D. Noh, Smart manufacturing: Past research, present findings, and future directions, Int. J. Pr. Eng. Man.-GT. 3 (1) (2016) 111–128, https://doi.org/10.1007/s40684-016-0015-5.
[3] T. De Beer, A. Burggraeve, M. Fonteyne, L. Saerens, J.P. Remon, C. Vervaet, Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes, Int. J. Pharm. 417 (1-2) (2011) 32–47, https://doi.org/10.1016/j.ijpharm.2010.12.012.
[4] D.R. Sjödin, V. Parida, M. Leksell, A. Petrovic, Smart factory implementation and process innovation: a preliminary maturity model for leveraging digitalization in manufacturing moving to smart factories presents specific challenges that can be addressed through a structured approach focused on people, processes, and technologies, Res. Technol. Manag. 61 (5) (2018) 22–31, https://doi.org/10.1080/08956308.2018.1471277.
[5] J. Chen, X. Ran, Deep learning with edge computing: a review, Proc. IEEE 107 (8) (2019) 1655–1674.
[6] R.Y. Zhong, X. Xu, E. Klotz, S.T. Newman, Intelligent manufacturing in the context of industry 4.0: a review, Engineering 3 (5) (2017) 616–630, https://doi.org/10.1016/J.ENG.2017.05.015.
[7] O. Fisher, N. Watson, L. Porcu, D. Bacon, M. Rigley, R.L. Gomes, Cloud manufacturing as a sustainable process manufacturing route, J. Manuf. Syst. 47 (2018) 53–68, https://doi.org/10.1016/j.jmsy.2018.03.005.
[8] L.L. Simon, H. Pataki, G. Marosi, F. Meemken, K. Hungerbühler, A. Baiker, S. Tummala, B. Glennon, M. Kuentz, G. Steele, H.J.M. Kramer, J.W. Rydzak, Z. Chen, J. Morris, F. Kjell, R. Singh, R. Gani, K.V. Gernaey, M. Louhi-Kultanen, J. O'Reilly, N. Sandler, O. Antikainen, J. Yliruusi, P. Frohberg, J. Ulrich, R. D. Braatz, T. Leyssens, M. von Stosch, R. Oliveira, R.B.H. Tan, H. Wu, M. Khan, D. O'Grady, A. Pandey, R. Westra, E. Delle-Case, D. Pape, D. Angelosante, Y. Maret, O. Steiger, M. Lenner, K. Abbou-Oucherif, Z.K. Nagy, J.D. Litster, V. K. Kamaraju, M.-S. Chiu, Assessment of recent process analytical technology (PAT) trends: a multiauthor review, Org. Process Res. Dev. 19 (1) (2015) 3–62, https://doi.org/10.1021/op500261y.
[9] A. Gowen, C. Odonnell, P. Cullen, G. Downey, J. Frias, Hyperspectral imaging – an emerging process analytical tool for food quality and safety control, Trends Food Sci. Tech. 18 (12) (2007) 590–598, https://doi.org/10.1016/j.tifs.2007.06.001.
[10] A.L. Bowler, S. Bakalis, N.J. Watson, A review of in-line and on-line measurement techniques to monitor industrial mixing processes, Chem. Eng. Res. Des. 153 (2020) 463–495, https://doi.org/10.1016/j.cherd.2019.10.045.
[11] B. Henning, J. Rautenberg, Process monitoring using ultrasonic sensor systems, Ultrasonics 44 (2006) e1395–e1399, https://doi.org/10.1016/j.ultras.2006.05.048.
[12] D.J. McClements, Advances in the application of ultrasound in food analysis and processing, Trends Food Sci. Tech. 6 (9) (1995) 293–299, https://doi.org/10.1016/S0924-2244(00)89139-6.
[13] T.S. Awad, H.A. Moharram, O.E. Shaltout, D. Asker, M.M. Youssef, Applications of ultrasound in analysis, processing and quality control of food: a review, Food Res. Int. 48 (2) (2012) 410–427, https://doi.org/10.1016/j.foodres.2012.05.004.
[14] R.E. Challis, M.J.W. Povey, M.L. Mather, A.K. Holmes, Ultrasound techniques for characterizing colloidal dispersions, Rep. Prog. Phys. 68 (7) (2005) 1541–1637.
[15] A. Simeone, E. Woolley, J. Escrig, N.J. Watson, Intelligent industrial cleaning: a multi-sensor approach utilising machine learning-based regression, Sensors 20 (2020) 1–22, https://doi.org/10.3390/s20133642.
[16] J.E. Escrig, A. Simeone, E. Woolley, S. Rangappa, A. Rady, N.J. Watson, Ultrasonic measurements and machine learning for monitoring the removal of surface fouling during clean-in-place processes, Food Bioprod. Process. 123 (2020) 1–13, https://doi.org/10.1016/j.fbp.2020.05.003.
[17] P. Resa, L. Elvira, F. Montero de Espinosa, R. González, J. Barcenilla, On-line ultrasonic velocity monitoring of alcoholic fermentation kinetics, Bioproc. Biosyst. Eng. 32 (3) (2009) 321–331, https://doi.org/10.1007/s00449-008-0251-3.
[18] Z. Ge, Z. Song, S.X. Ding, B. Huang, Data mining and analytics in the process industry: the role of machine learning, IEEE Access 5 (2017) 20590–20616, https://doi.org/10.1109/ACCESS.2017.2756872.

[19] J.E. van Engelen, H.H. Hoos, A survey on semi-supervised learning, Mach. Learn. 109 (2) (2020) 373–440, https://doi.org/10.1007/s10994-019-05855-6.
[20] A.L. Bowler, N.J. Watson, Transfer learning for process monitoring using reflection-mode ultrasonic sensing, Ultrasonics 115 (2021) 106468, https://doi.org/10.1016/j.ultras.2021.106468.
[21] A. Joby, What Is Cross-Validation? Comparing Machine Learning Models, 2021. <https://learn.g2.com/cross-validation> (Accessed 17/11/2021).
[22] O.J. Fisher, N.J. Watson, J.E. Escrig, R. Witt, L. Porcu, D. Bacon, M. Rigley, R. L. Gomes, Considerations, challenges and opportunities when developing data-driven models for process manufacturing systems, Comput. Chem. Eng. 140 (2020) 106881, https://doi.org/10.1016/j.compchemeng.2020.106881.
[23] G. Chandrashekar, F. Sahin, A survey on feature selection methods, Comput. Electr. Eng. 40 (1) (2014) 16–28, https://doi.org/10.1016/j.compeleceng.2013.11.024.
[24] W.B. Hussein, M.A. Hussein, T. Becker, Robust spectral estimation for speed of sound with phase shift correction applied online in yeast fermentation processes, Eng. Life Sci. 12 (6) (2012) 603–614, https://doi.org/10.1002/elsc.201100183.
[25] F.C. Cruz, E.F. Simas Filho, M.C.S. Albuquerque, I.C. Silva, C.T.T. Farias, L. L. Gouvêa, Efficient feature selection for neural network based detection of flaws in steel welded joints using ultrasound testing, Ultrasonics 73 (2017) 1–8, https://doi.org/10.1016/j.ultras.2016.08.017.
[26] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE T. Pattern Anal. 11 (1989) 674–693, https://doi.org/10.1109/34.192463.
[27] A.L. Bowler, S. Bakalis, N.J. Watson, Monitoring mixing processes using ultrasonic sensors and machine learning, Sensors 20 (2020) 1813, https://doi.org/10.3390/s20071813.
[28] F. Cau, A. Fanni, A. Montisci, P. Testoni, M. Usai, Artificial neural networks for non-destructive evaluation with ultrasonic waves in not accessible, IEEE Ind. Applic. Soc. 1 (2005) 685–692, https://doi.org/10.1109/IAS.2005.1518382.
[29] C. Miao, Y. Wang, Y. Zhang, J. Qn, M.J. Zuo, X. Wang, A SVM classifier combined with PCA for ultrasonic crack size classification, Can. Conf. Electr. Comput. Eng. (2008) 1627–1630, https://doi.org/10.1109/CCECE.2008.4564817.
[30] S. Khalid, T. Khalil, S. Nasreen, A survey of feature selection and feature extraction techniques in machine learning, in: Proceedings of 2014 Science and Information Conference, 2014, pp. 372–378, https://doi.org/10.1109/SAI.2014.6918213.
[31] R. Jenke, A. Peer, M. Buss, Feature extraction and selection for emotion recognition from EEG, IEEE T Affect. Comput. 5 (3) (2014) 327–339, https://doi.org/10.1109/TAFFC.2014.2339834.
[32] D. Gómez, A. Rojas, An empirical overview of the no free lunch theorem and its effect on real-world machine learning classification, Neural Comput. 28 (1) (2016) 216–228, https://doi.org/10.1162/NECO_a_00793.
[33] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (3) (2004) 199–222, https://doi.org/10.1023/B:STCO.0000035301.49549.88.
[34] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, M. Chica-Rivas, Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines, Ore Geol. Rev. 71 (2015) 804–818, https://doi.org/10.1016/j.oregeorev.2015.01.001.
[35] N. Munir, H.-J. Kim, J. Park, S.-J. Song, S.-S. Kang, Convolutional neural network for ultrasonic weldment flaw classification in noisy conditions, Ultrasonics 94 (2019) 74–81, https://doi.org/10.1016/j.ultras.2018.12.001.
[36] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput 9 (8) (1997) 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.
[37] J. Vaněk, J. Zelinka, D. Soutner, J. Psutka, A regularization post layer: an additional way how to make deep neural networks robust, Lect. Notes Comp. Sc. 10583 (2017) 204–214, https://doi.org/10.1007/978-3-319-68456-7_17.
[38] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutsever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (2014) 1929–1958.
[39] J.E. Escrig, E. Woolley, S. Rangappa, A. Simeone, N.J. Watson, Clean-in-place monitoring of different food fouling materials using ultrasonic measurements, Food Control 104 (2019) 358–366, https://doi.org/10.1016/j.foodcont.2019.05.013.
[40] M.A. Úbeda, W.B. Hussein, M.A. Hussein, J. Hinrichs, T.M. Becker, Acoustic sensing and signal processing techniques for monitoring milk fouling cleaning operations, Eng. Life Sci. 16 (1) (2016) 67–77, https://doi.org/10.1002/elsc.201400235.
[41] E. Wallhäußer, W.B. Hussein, M.A. Hussein, J. Hinrichs, T.M. Becker, On the usage of acoustic properties combined with an artificial neural network – a new approach of determining presence of dairy fouling, J. Food Eng. 103 (4) (2011) 449–456, https://doi.org/10.1016/j.jfoodeng.2010.11.015.
[42] E. Wallhäußer, W.B. Hussein, M.A. Hussein, J. Hinrichs, T. Becker, Detection of dairy fouling: combining ultrasonic measurements and classification methods, Eng. Life Sci. 13 (3) (2013) 292–301, https://doi.org/10.1002/elsc.201200081.
[43] E. Wallhäußer, A. Sayed, S. Nöbel, M.A. Hussein, J. Hinrichs, T. Becker, Determination of cleaning end of dairy protein fouling using an online system combining ultrasonic and classification methods, Food Bioprocess Tech. 7 (2) (2014) 506–515, https://doi.org/10.1007/s11947-012-1041-0.
[44] J. Escrig, E. Woolley, A. Simeone, N.J. Watson, Monitoring the cleaning of food fouling in pipes using ultrasonic measurements and machine learning, Food Control 116 (2020) 107309, https://doi.org/10.1016/j.foodcont.2020.107309.
[45] B. Chen, D. Callens, P. Campistron, E. Moulin, P. Debreyne, G. Delaplace, Monitoring cleaning cycles of fouled ducts using ultrasonic coda wave interferometry (CWI), Ultrasonics 96 (2019) 252–260, https://doi.org/10.1016/j.ultras.2018.12.011.
[46] P. Resa, L. Elvira, F. Montero de Espinosa, Concentration control in alcoholic fermentation processes from ultrasonic velocity measurements, Food Res. Int. 37 (6) (2004) 587–594, https://doi.org/10.1016/j.foodres.2003.12.012.
[47] P. Resa, L. Elvira, F.M. De Espinosa, T. Bolumar, G. Pérez, Ultrasonic velocity measurements in the ternary mixtures water-lactose-lactate, for the purpose of monitoring the lactic acid fermentation of lactose, Proc. – IEEE Ultrason. Symp. 3 (2004) P3U-Y-1, https://doi.org/10.1109/ULTSYM.2004.1418306.
[48] D. Novoa-Díaz, J.M. Rodríguez-Nogales, E. Fernández-Fernández, J. Vila-Crespo, J. García-Álvarez, M.A. Amer, J.A. Chávez, A. Turó, M.J. García-Hernández, J. Salazar, Ultrasonic monitoring of malolactic fermentation in red wines, Ultrasonics 54 (6) (2014) 1575–1580, https://doi.org/10.1016/j.ultras.2014.04.004.
[49] T. Becker, M. Mitzscherling, A. Delgado, Hybrid data model for the improvement of an ultrasonic-based gravity measurement system, Food Control 13 (4-5) (2002) 223–233, https://doi.org/10.1016/S0956-7135(01)00104-9.
[50] P. Resa, T. Bolumar, L. Elvira, G. Pérez, F.M. de Espinosa, Monitoring of lactic acid fermentation in culture broth using ultrasonic velocity, J. Food Eng. 78 (3) (2007) 1083–1091, https://doi.org/10.1016/j.jfoodeng.2005.12.021.
[51] N. Lamberti, L. Ardia, D. Albanese, M. Di Matteo, An ultrasound technique for monitoring the alcoholic wine fermentation, Ultrasonics 49 (1) (2009) 94–97, https://doi.org/10.1016/j.ultras.2008.06.003.
[52] S. Hoche, D. Krause, M.A. Hussein, T. Becker, Ultrasound-based, in-line monitoring of anaerobe yeast fermentation: Model, sensor design and process application, 51 (2016) 710–719. doi:10.1111/ijfs.13027.
[53] A. Bowler, J. Escrig, M. Pound, N. Watson, Predicting alcohol concentration during beer fermentation using ultrasonic measurements and machine learning, Fermentation 7 (2021) 34, https://doi.org/10.3390/fermentation7010034.
[54] H. Ogasawara, K. Mizutani, T. Ohbuchi, T. Nakamura, Acoustical experiment of yogurt fermentation process, Ultrasonics 44 (2006) e727–e730, https://doi.org/10.1016/j.ultras.2006.05.084.
[55] R. Meng, J. Zhou, X. Ye, D. Liu, On-line monitoring of yogurt fermentation using acoustic impedance method, Appl. Mech. Mater. 101–102 (2012) 737–742, https://doi.org/10.4028/www.scientific.net/AMM.101-102.737.
[56] M.A. Amer, D. Novoa-Díaz, A. Puig-Pujol, J. Capdevila, J.A. Chávez, A. Turó, M. J. García-Hernández, J. Salazar, Ultrasonic velocity of water-ethanol-malic acid-lactic acid mixtures during the malolactic fermentation process, J. Food Eng. 149 (2015) 61–69, https://doi.org/10.1016/j.jfoodeng.2014.09.042.
[57] M.A. Amer, D. Novoa-Díaz, J.A. Chávez, A. Turó, M.J. García-Hernández, J. Salazar, Temperature compensation of ultrasonic velocity during the malolactic fermentation process, Meas. Sci. Technol. 26 (12) (2015) 125602, https://doi.org/10.1088/0957-0233/26/12/125602.
[58] D.A. Çelik, M.A. Amer, D. Novoa-Díaz, J.A. Chávez, A. Turó, M.J. García-Hernández, J. Salazar, Design and implementation of an ultrasonic sensor for rapid monitoring of industrial malolactic fermentation of wines 46 (2018) 387–407. DOI:10.1080/10739149.2017.1394878.
[59] C. Keskinoğlu, A. Aydın, Ultrasound based noninvasive real-time cell proliferation process monitoring, J. Acoust. Soc. Am. 149 (5) (2021) 3345–3351, https://doi.org/10.1121/10.0004993.
[60] T. Stelzer, D. Pertig, J. Ulrich, Ultrasonic crystallization monitoring technique for simultaneous in-line measurement of liquid and solid phase, J. Cryst. Growth 362 (2013) 71–76, https://doi.org/10.1016/j.jcrysgro.2011.11.027.
[61] P. Mougin, A. Thomas, D. Wilkinson, G. White, K.J. Roberts, N. Herrmann, R. Jack, R. Tweedie, On-line monitoring of a crystallization process, AIChE J. 49 (2) (2003) 373–378, https://doi.org/10.1002/aic.690490209.
[62] P. Mougin, D. Wilkinson, K.J. Roberts, R. Tweedie, Characterization of particle size and its distribution during the crystallization of organic fine chemical products as measured in situ using ultrasonic attenuation spectroscopy, J. Acoust. Soc. Am. 109 (2001) 274–282, https://doi.org/10.1121/1.1331113.
[63] P. Mougin, D. Wilkinson, K.J. Roberts, In situ measurement of particle size during the crystallization of L-glutamic acid under two polymorphic forms: influence of crystal habit on ultrasonic attenuation measurements, Crystal Growth Des. 2 (2002) 227–234, https://doi.org/10.1021/cg0155752.
[64] M. Li, D. Wilkinson, K. Patchigolla, P. Mougin, K.J. Roberts, R. Tweedie, On-line crystallization process parameter measurements using ultrasonic attenuation spectroscopy, Cryst. Growth Des. 4 (5) (2004) 955–963, https://doi.org/10.1021/cg030041h.
[65] E. Lyall, P. Mougin, D. Wilkinson, K.J. Roberts, In situ ultrasonic spectroscopy study of the nucleation and growth of copper sulfate pentahydrate batch crystallized from supersaturated aqueous solutions, Ind. Eng. Chem. Res. 43 (16) (2004) 4947–4956, https://doi.org/10.1021/ie0342560.
[66] A. Shukla, A. Prakash, S. Rohani, Online measurement of particle size distribution during crystallization using ultrasonic spectroscopy, Chem. Eng. Sci. 65 (10) (2010) 3072–3079, https://doi.org/10.1016/j.ces.2010.01.034.
[67] D. Pertig, R. Buchfink, S. Petersen, T. Stelzer, J. Ulrich, Inline analyzing of industrial crystallization processes by an innovative ultrasonic probe technique, Chem. Eng. Technol. 34 (2011) 639–646, https://doi.org/10.1002/ceat.201000558.
[68] P. Frohberg, J. Ulrich, Single-frequency ultrasonic crystallization monitoring (UCM): Innovative technique for in-line analyzing of industrial crystallization processes, Org. Process Res. Dev. 19 (1) (2015) 84–88, https://doi.org/10.1021/op400362f.
[69] L. Helmdach, M.P. Feth, J. Ulrich, Application of ultrasound measurements as pat tools for industrial crystallization process development of pharmaceutical

compounds, Org. Process Res. Dev. 19 (1) (2015) 110–121, https://doi.org/10.1021/op4001803.
[70] L. Morris, E. Simeone, Z.J. Glover, H. Powell, S. Marty-Terrade, M. Francis, M. J. Povey, Dynamic monitoring of glycine crystallisation with low power ultrasound reflection spectroscopy, Chem. Eng. Res. Des. 170 (2021) 213–223, https://doi.org/10.1016/j.cherd.2021.04.003.
[71] J.A. Bamberger, M.S. Greenwood, Using ultrasonic attenuation to monitor slurry mixing in real time, Ultrasonics 42 (1-9) (2004) 145–148, https://doi.org/10.1016/j.ultras.2004.02.016.
[72] P. Fox, P.P. Smith, S. Sahi, Ultrasound measurements to monitor the specific gravity of food batters, J. Food Eng. 65 (3) (2004) 317–324, https://doi.org/10.1016/j.jfoodeng.2004.01.028.
[73] J. Salazar, A. Turó, J.A. Chávez, M.J. García, Ultrasonic inspection of batters for on-line process monitoring, Ultrasonics 42 (1-9) (2004) 155–159, https://doi.org/10.1016/j.ultras.2004.02.017.
[74] M. Tourbin, C. Frances, Monitoring of the aggregation process of dense colloidal silica suspensions in a stirred tank by acoustic spectroscopy, Powder Technol. 190 (1-2) (2009) 25–30, https://doi.org/10.1016/j.powtec.2008.04.067.
[75] L. Liu, R.F. Li, S. Collins, X.Z. Wang, R. Tweedie, K. Primrose, Ultrasound spectroscopy and electrical resistance tomography for online characterisation of concentrated emulsions in crossflow membrane emulsifications, Powder Technol. 213 (1-3) (2011) 123–131, https://doi.org/10.1016/j.powtec.2011.07.018.
[76] T.N. Hunter, L. Darlison, J. Peakall, S. Biggs, Using a multi-frequency acoustic backscatter system as an in situ high concentration dispersion monitor, Chem. Eng. Sci. 80 (2012) 409–418, https://doi.org/10.1016/j.ces.2012.06.038.
[77] G.M. Owolabi, M.N. Bassim, J.H. Page, M.G. Scanlon, The influence of specific mechanical energy on the ultrasonic characteristics of extruded dough, J. Food Eng. 86 (2) (2008) 202–206, https://doi.org/10.1016/j.jfoodeng.2007.09.029.
[78] P.D. Coates, S.E. Barnes, M.G. Sibley, E.C. Brown, H.G.M. Edwards, I.J. Scowen, In-process vibrational spectroscopy and ultrasound measurements in polymer melt extrusion, Polymer 44 (19) (2003) 5937–5949, https://doi.org/10.1016/S0032-3861(03)00544-5.
[79] S.E. Barnes, E.C. Brown, M.G. Sibley, H.G.M. Edwards, P.D. Coates, Vibrational spectroscopic and ultrasound analysis for the in-process monitoring of poly (ethylene vinyl acetate) copolymer composition during melt extrusion, Analyst 130 (2005) 286–292, https://doi.org/10.1039/b416244g.
[80] I. Alig, D. Fischer, D. Lellinger, B. Steinhoff, Combination of NIR, Raman, ultrasonic and dielectric spectroscopy for in-line monitoring of the extrusion process 230 (2005) 51–58. doi:10.1002/masy.200551141.
[81] Z. Sun, C.-K. Jen, J. Yan, M.-Y. Chen, Application of ultrasound and neural networks in the determination of filler dispersion during polymer extrusion processes, Polym. Eng. Sci. 45 (6) (2005) 764–772, https://doi.org/10.1002/pen.20328.
[82] D. Fischer, K. Sahre, M. Abdelrhim, B. Voit, V.B. Sadhu, J. Pionteck, H. Komber, J. Hutschenreuter, Process monitoring of polymers by in-line ATR-IR, NIR and Raman spectroscopy and ultrasonic measurements, CR Chim. 9 (2006) 1419–1424, https://doi.org/10.1016/j.crci.2006.06.006.
[83] D. Fischer, J. Müller, S. Kummer, B. Kretzschmar, Real time monitoring of morphologic and mechanical properties of polymer nanocomposites during extrusion by near infrared and ultrasonic spectroscopy, Macromol. Symp. 305 (2011) 10–17, https://doi.org/10.1002/masy.201000113.
[84] G. Schober, P. Heidemeyer, K. Kretschmer, M. Bastian, T. Hochrein, Degree of dispersion monitoring by ultrasonic transmission technique and excitation of the transducer's harmonics, AIP Conf. Proc. 1593 (2014) 24–28, https://doi.org/10.1063/1.4873727.
[85] S. Wöckel, h. Arndt, U. Steinmann, J. Auge, K. Dietl, G. Schober, C. Kugler, T. Hochrein, Statistical ultrasonic characterization of particulate filler in polymer compounds, IEEE Int. Ultra. Sym. 2016-November (2016) 7728747. doi:10.1109/ULTSYM.2016.7728747.
[86] N. Halmen, C. Kugler, T. Hochrein, P. Heidemeyer, M. Bastian, Ultrasound tomography for inline monitoring of plastic melts, J. Sens. Sens. Syst. 6 (1) (2017) 9–18, https://doi.org/10.5194/jsss-6-9-201710.5194/jsss-6-9-2017-supplement.
[87] M. Kariminejad, D. Tormey, S. Huq, J. Morrison, M. McAfee, Ultrasound sensors for process monitoring in injection moulding, Sensors 21 (2021) 5193, https://doi.org/10.3390/s21155193.
[88] Y.-L. Wu, C.-C. Cheng, M. Kobayashi, C.-H. Yang, Novel design of extension nozzle and its application on real-time injection molding process diagnosed by ultrasound, Sensor. Actuat. A-Phys. 263 (2017) 430–438, https://doi.org/10.1016/j.sna.2017.06.023.
[89] D. Altmann, B. Praher, G. Steinbichler, Simulation of the melting behavior in an injection molding plasticizing unit as measured by pressure and ultrasound measurement technology, AIP Conf. Proc. 2055 (2019) 040003, https://doi.org/10.1063/1.5084818.
[90] P. Zhao, Y. Zhao, H. Kharbas, J. Zhang, T. Wu, W. Yang, J. Fu, L.-S. Turng, In-situ ultrasonic characterization of microcellular injection molding, J. Mater. Process. Tech. 270 (2019) 254–264, https://doi.org/10.1016/j.jmatprotec.2019.03.012.
[91] C.-C. Cheng, Y.-L. Wu, Diagnosis of multi-stage injection molding process by ultrasonic technology at a T-shape extension nozzle, J. Mater. Process. Tech. 282 (2020) 116650, https://doi.org/10.1016/j.jmatprotec.2020.116650.
[92] P. Zhao, K. Ji, J. Zhang, Y. Chen, Z. Dong, J. Zheng, J. Fu, In-situ ultrasonic measurement of molten polymers during injection molding, J. Mater. Process. Tech. 293 (2021) 117081, https://doi.org/10.1016/j.jmatprotec.2021.117081.
[93] L. Grob, K. Papadea, P. Braun, E.J. Windhab, In-line detection method for crystallization, contraction and mold detachment during cooling of confectionery products, J Food Eng 292 (2021) 110322, https://doi.org/10.1016/j.jfoodeng.2020.110322.
[94] M. Rath, J. Döring, W. Stark, G. Hinrichsen, Process monitoring of moulding compounds by ultrasonic measurements in a compression mould, NDT&E Int. 33 (2) (2000) 123–130, https://doi.org/10.1016/S0963-8695(99)00029-8.
[95] F. Lionetto, A. Tarzia, A. Maffezzoli, Air-coupled ultrasound: a novel technique for monitoring the curing of thermosetting matrices, IEEE T. Ultrason. Ferr. 54 (7) (2007) 1437–1444, https://doi.org/10.1109/TUFFC.2007.404.
[96] F. Lionetto, A. Maffezzoli, Monitoring the cure state of thermosetting resins by ultrasound, Materials 6 (2013) 3783–3804, https://doi.org/10.3390/ma6093783.
[97] V. Koissin, A. Demčenko, V.A. Korneev, Isothermal epoxy-cure monitoring using nonlinear ultrasonics, Int. J. Ahes. Adhes. 52 (2014) 11–18, https://doi.org/10.1016/j.ijadhadh.2014.01.003.
[98] N. Ghodhbani, P. Maréchal, H. Duflo, Ultrasound monitoring of the cure kinetics of an epoxy resin: Identification, frequency and temperature dependence, Polym. Test. 56 (2016) 156–166, https://doi.org/10.1016/j.polymertesting.2016.10.009.
[99] A. Dominguez-Macaya, T.E.G. Álvarez-Arenas, I. Saenz-Dominguez, I. Tena, J. Aurrekoetxea, A. Iturrospe, Monitoring the evolution of stiffness during ultraviolet curing of a vinyl ester resin with quasi-normal air-coupled ultrasonic spectroscopy, Polym. Test. 80 (2019) 106112, https://doi.org/10.1016/j.polymertesting.2019.106112.
[100] J.S. Chilles, A.E. Koutsomitopoulou, A.J. Croxford, I.P. Bond, Monitoring cure and detecting damage in composites with inductively coupled embedded sensors, Compos. Sci. Technol. 134 (2016) 81–88, https://doi.org/10.1016/j.compscitech.2016.07.028.
[101] K. Mizukami, S. Yoshimoto, K. Ogi, In-process acquisition of cure-dependent viscoelastic properties of carbon fiber reinforced composites using micromechanics-based guided wave analysis, Polym. Test. 65 (2018) 459–467, https://doi.org/10.1016/j.polymertesting.2017.12.032.
[102] K. Mizukami, T. Ikeda, K. Ogi, Measurement of velocity and attenuation of ultrasonic guided wave for real-time estimation of cure-dependent anisotropic viscoelastic properties of carbon fiber-reinforced plastics, Ultrasonics 99 (2019) 105952, https://doi.org/10.1016/j.ultras.2019.105952.
[103] X. Liu, J. Li, J. Zhu, Y. Wang, X. Qing, Cure monitoring and damage identification of CFRP using embedded piezoelectric sensors network, Ultrasonics 115 (2021) 106470, https://doi.org/10.1016/j.ultras.2021.106470.
[104] P. Pawelzyk, M.L. Toledo, N. Willenbacher, Ultrasonic in-line monitoring of styrene miniemulsion polymerization, Chem. Eng. J. 219 (2013) 303–310, https://doi.org/10.1016/j.cej.2013.01.014.
[105] V. Buckin, M.C. Atlas, Ultrasonic monitoring of biocatalysis in solutions and complex dispersions, Catalysts 7 (2017) 336, https://doi.org/10.3390/catal7110336.
[106] V. Buckin, High-resolution ultrasonic spectroscopy, J. Sens. Sens. Syst. 7 (2018) 207–217, https://doi.org/10.5194/jsss-7-207-2018.
[107] M.K.K. Figueiredo, C.E.R. Silva, A.V. Alvarenga, R.P.B. Costa-Félix, Relating speed of sound and echo amplitude with biodiesel manufacture, Chem. Eng. Res. Des. 136 (2018) 825–833, https://doi.org/10.1016/j.cherd.2018.06.038.
[108] R.M. Baêsso, R.P.B. Costa-Felix, P. Miloro, B. Zeqiri, Ultrasonic parameter measurement as a means of assessing the quality of biodiesel production, Fuel 241 (2019) 155–163, https://doi.org/10.1016/j.fuel.2018.12.032.
[109] M. Schmachtl, T.J. Kim, W. Grill, R. Herrmann, O. Scharf, W. Schwieger, R. Schertlen, C. Stenzel, Ultrasonic monitoring of zeolite synthesis in real time, Ultrasonics 38 (2000) 809–812, https://doi.org/10.1016/S0041-624X(99)00201-2.
[110] E. Hums, H. Baser, W. Schwieger, In situ ultrasonic measurements: a powerful tool to control the synthesis of zeolites from coal fly ash, Res. Chem. Intermediat. 42 (10) (2016) 7513–7532, https://doi.org/10.1007/s11164-016-2550-7.
[111] G.J.A. Van Groenestijn, N.M.M. Meulendiiks, A.W.F. Volker, P.L.M.J. Van Neer, P. J.P. Buskens, Real-time monitoring of size and concentration of nanoparticles inside a reactor using ultrasound, IEEE Int. Ultra. Sym. (2018) 143804, https://doi.org/10.1109/ULTSYM.2018.8580198.
[112] J.D. Stephens, B.R. Kowalczyk, B.C. Hancock, G. Kaul, I. Akseli, C. Cetinkaya, In-die ultrasonic and off-line air-coupled monitoring and characterization techniques for drug tablets, AIP Conf. Proc. 1430 (2012) 1691–1698, https://doi.org/10.1063/1.4716416.
[113] J.T.T. Leskinen, S.-P. Simonaho, M. Hakulinen, J. Ketolainen, Real-time tablet formation monitoring with ultrasound measurements in eccentric single station tablet press, Int. J. Pharm. 442 (1-2) (2013) 27–34, https://doi.org/10.1016/j.ijpharm.2012.09.004.
[114] X. Li, Y. Mo, J. Li, W. Guo, H.H. Ngo, In-situ monitoring techniques for membrane fouling and local filtration characteristics in hollow fiber membrane processes: a critical review, J. Membrane Sci. 528 (2017) 187–200, https://doi.org/10.1016/j.memsci.2017.01.030.
[115] G. Rudolph, T. Virtanen, M. Ferrando, C. Güell, F. Lipnizki, M. Kallioinen, A review of in situ real-time monitoring techniques for membrane fouling in the biotechnology, biorefinery and food sectors, J. Membrane Sci. 588 (2019) 117221, https://doi.org/10.1016/j.memsci.2019.117221.
[116] X. Li, H. Zhang, Y. Hou, Y. Gao, J. Li, W. Guo, H.H. Ngo, In situ investigation of combined organic and colloidal fouling for nanofiltration membrane using ultrasonic time domain reflectometry, Desalination 362 (2015) 43–51, https://doi.org/10.1016/j.desal.2015.02.005.
[117] M. Mitra, S. Gopalakrishnan, Guided wave based structural health monitoring: a review, Smart Mater. Struct. 25 (5) (2016) 053001, https://doi.org/10.1088/0964-1726/25/5/053001.
[118] P. Gardner, R. Fuentes, N. Dervilis, C. Mineo, S.G. Pierce, E.J. Cross, K. Worden, Machine learning at the interface of structural health monitoring and non-

destructive evaluation: Machine Learning in SHM and NDE, Philos. Trans. Royal Soc. A 378 (2182) (2020) 20190581, https://doi.org/10.1098/rsta.2019.0581.
[119] M. Azimi, A.D. Eslamlou, G. Pekcan, Data-driven structural health monitoring and damage detection through deep learning: State-ofthe- art review, Sensors (Basel) 20 (10) (2020). doi:10.3390/s20102778.
[120] X.W. Ye, T. Jin, C.B. Yun, A review on deep learning-based structural health monitoring of civil infrastructures, Smart Struct. Syst. 24 (5) (2019) 567–585, https://doi.org/10.12989/sss.2019.24.5.567.
[121] S. Sony, K. Dunphy, A. Sadhu, M. Capretz, A systematic review of convolutional neural network-based structural condition assessment techniques, Eng. Struct. 226 (2021) 111347, https://doi.org/10.1016/j.engstruct.2020.111347.
[122] Y. Bao, Z. Chen, S. Wei, Y. Xu, Z. Tang, H. Li, The state of the art of data science and engineering in structural health monitoring, J. Eng. 5 (2) (2019) 234–242, https://doi.org/10.1016/j.eng.2018.11.027.
[123] M. Flah, I. Nunez, W. Ben Chaabene, M.L. Nehdi, Machine learning algorithms in civil structural health monitoring: a systematic review, Arch Comput. Methods Eng 28 (4) (2021) 2621–2643.
[124] J.B. Harley, D. Sparkman, Machine learning and NDE: Past, present, and future, in: AIP Conf. Proc. doi:10.1063/1.5099819.
[125] H. Ahmed, H.M. La, N. Gucunski, Review of non-destructive civil infrastructure evaluation for bridges: state-of-the-art robotic platforms, sensors and algorithms, Sensors (Basel) 20 (14) (2020) 1–38, https://doi.org/10.3390/s20143954.
[126] M. Gordan, S.-R. Sabbagh-Yazdi, Z. Ismail, K. Ghaedi, P. Carroll, D. McCrum, B. Samali, State-of-the-art review on advancements of data mining in structural health monitoring, Meas.: J. Int. Meas. Confed. 193 (2022) 110939, https://doi.org/10.1016/j.measurement.2022.110939.
[127] G. Toh, J. Park, Review of vibration-based structural health monitoring using deep learning, Appl. Sci. 10 (5) (2020) 1680, https://doi.org/10.3390/app10051680.
[128] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, Deep learning and its applications to machine health monitoring, Mech. Syst. Signal. Pr. 115 (2018) 213–237, https://doi.org/10.1016/j.ymssp.2018.05.050.
[129] S. Lu, L.i. Zhou, Fatigue crack monitoring of aerospace structure based on lamb waves and binary tree support vector machines, J. Vibroeng. 19 (5) (2017) 3271–3282, https://doi.org/10.21595/jve.2017.17528.
[130] A.K.U. Malikov, Y. Cho, Y.H. Kim, J. Kim, J. Park, J.-H. Yi, Ultrasonic assessment of thickness and bonding quality of coating layer based on short-time fourier transform and convolutional neural networks, Coatings 11 (2021) 909, https://doi.org/10.3390/coatings11080909.
[131] M. Barbosh, P. Singh, A. Sadhu, Empirical mode decomposition and its variants: a review with applications in structural health monitoring, Smart Mater. Struct. 29 (9) (2020) 093001, https://doi.org/10.1088/1361-665X/aba539.
[132] L. Zhang, H. Li, B. Gao, Combination of wavelet packet analysis with BPNN flaw type identification in concrete ultrasonic testing, in: 3rd International Conference on Innovative Computing Information and Control, 2008, p. 4603704, https://doi.org/10.1109/ICICIC.2008.195.
[133] S. Sambath, P. Nagaraj, N. Selvakumar, Automatic defect classification in ultrasonic NDT using artificial intelligence, J. Nondestruct. Eval. 30 (1) (2011) 20–28, https://doi.org/10.1007/s10921-010-0086-0.
[134] M. Meng, Y.J. Chua, E. Wouterson, C.P.K. Ong, Ultrasonic signal classification and imaging system for composite materials via deep convolutional neural networks, Neurocomputing 257 (2017) 128–135, https://doi.org/10.1016/j.neucom.2016.11.066.
[135] K. Virupakshappa, M. Marino, E. Oruklu, A multi-resolution convolutional neural network architecture for ultrasonic flaw detection, IEEE Int. Ultra. Sym. (2018) 8579888, https://doi.org/10.1109/ULTSYM.2018.8579888.
[136] Y. Yan, D. Liu, B. Gao, G.Y. Tian, Z.C. Cai, A deep learning-based ultrasonic pattern recognition method for inspecting girth weld cracking of gas pipeline, IEEE Sens. J. 20 (14) (2020) 7997–8006, https://doi.org/10.1109/JSEN.2020.2982680.
[137] M. Rautela, S. Gopalakrishnan, Ultrasonic guided wave based structural damage detection and localization using model assisted convolutional and recurrent neural networks, Exp. Syst. Appl. 167 (2021) 114189, https://doi.org/10.1016/j.eswa.2020.114189.
[138] S. Bosse, D. Weiss, D. Schmidt, Supervised distributed multi-instance and unsupervised single-instance autoencoder machine learning for damage diagnostics with high-dimensional data—a hybrid approach and comparison study, Computers 10 (2021) 34, https://doi.org/10.3390/computers10030034.
[139] J. Liu, G. Xu, L. Ren, Z. Qian, L. Ren, Defect intelligent identification in resistance spot welding ultrasonic detection based on wavelet packet and neural network, Int. J. Adv. Manuf. Tech. 90 (9–12) (2017) 2581–2588, https://doi.org/10.1007/s00170-016-9588-y.
[140] K. Sudheera, N.M. Nandhitha, V.B.V. Sai, N.V. Kumar, Deep learning techniques for flaw characterization in weld pieces from ultrasonic signals, Russ. J. Nondestruct+ 56 (10) (2020) 820–830, https://doi.org/10.1134/S1061830920100083.
[141] X. Wang, S. Guan, L. Hua, B. Wand, X. He, Classification of spot-welded joint strength using ultrasonic signal time-frequency features and PSO-SVM method, Ultrasonics 91 (2019) 161–169, https://doi.org/10.1016/j.ultras.2018.08.014.
[142] K. Virupakshappa, E. Oruklu, Ultrasonic flaw detection using hidden markov model with wavelet features, IEEE Int. Ultra. Sym. (2016) 7728491, https://doi.org/10.1109/ULTSYM.2016.7728491.
[143] Y. Lu, L. Ye, Z. Su, L. Zhou, L. Cheng, Artificial Neural Network (ANN)-based crack identification in aluminum plates with lamb wave signals, J. Intel. Mat. Syst. Str. 20 (2009) 39–49, https://doi.org/10.1177/1045389X07088782.
[144] F. Lanza di Scalea, P. Rizzo, S. Coccia, I. Bartoli, M. Fateh, E. Viola, G. Pascale, Non-contact ultrasonic inspection of rails and signal processing for automatic defect detection and classification, Insight: Non-Destruct. Test. Cond. Monitor. 47 (6) (2005) 346–353, https://doi.org/10.1784/insi.47.6.346.66449.
[145] P. Rizzo, I. Bartoli, A. Marzani, F.L. di Scalea, Defect classification in pipes by neural networks using multiple guided ultrasonic wave features extracted after wavelet processing, J. Press. Vess.-T. ASME 127 (2005) 294–303, https://doi.org/10.1115/1.1990213.
[146] S. Iyer, S.K. Sinha, B.R. Tittmann, M.K. Pedrick, Ultrasonic signal processing methods for detection of defects in concrete pipes, Autom. Constr. 22 (2012) 135–148, https://doi.org/10.1016/j.autcon.2011.06.012.
[147] H. Sun, S. Huang, S. Wang, W. Zhao, L. Peng, Quantification of defects with point-focusing shear horizontal guided wave EMAT using deep residual network, IEEE Intl. Conf. Ind. I (2021), https://doi.org/10.1109/INDIN45523.2021.9557567.
[148] N. Munir, H.-J. Kim, S.-J. Song, S.-S. Kang, Investigation of deep neural network with drop out for ultrasonic flaw classification in weldments, J. Mech. Sci. Technol. 32 (7) (2018) 3073–3080, https://doi.org/10.1007/s12206-018-0610-1.
[149] B.M. Abbagoni, H. Yeung, Non-invasive classification of gas-liquid two-phase horizontal flow regimes using an ultrasonic Doppler sensor and a neural network, Meas. Sci. Technol. 27 (8) (2016) 084002, https://doi.org/10.1088/0957-0233/27/8/084002.
[150] Y. Zhang, A.N. Azman, K.-W. Xu, C. Kang, H.-B. Kim, Two-phase flow regime identification based on the liquid-phase velocity information and machine learning, Exp. Fluids 61 (2020) 212, https://doi.org/10.1007/s00348-020-03046-x.
[151] S.-J. Song, H.-J. Kim, H. Cho, Development of an intelligent system for ultrasonic flaw classification in weldments, Nucl. Eng. Des. 212 (1-3) (2002) 307–320, https://doi.org/10.1016/S0029-5493(01)00495-2.
[152] M.B. Utomo, T. Sakai, S. Uchida, Use of neural network-ultrasonic technique for measuring gas and solid hold-ups in a slurry bubble column, Chem. Eng. Technol. 25 (2002) 293–299, https://doi.org/10.1002/1521-4125(200203)25:3<293::AID-CEAT293>3.0.CO;2-X/.
[153] M.B. Utomo, T. Sakai, S. Uchida, A. Maezawa, Simultaneous measurement of mean bubble diameter and local gas holdup using ultrasonic method with neural network, Chem. Eng. Technol. 24 (2001) 493–500, https://doi.org/10.1002/1521-4125(200105)24:5<493::AID-CEAT493>3.0.CO;2-L.
[154] Z. Tang, N. Munir, T.-G. Lee, Y.-T. Yeom, S.-J. Song, Lamb wave flaw classification in al plates using time reversal and deep neural networks, J. Korean Phys. Soc. 75 (12) (2019) 978–984, https://doi.org/10.3938/jkps.75.978.
[155] L.C. Silva, E.F. Simas Filho, M.C.S. Albuquerque, I.C. Silva, C.T.T. Farias, Segmented analysis of time-of-flight diffraction ultrasound for flaw detection in welded steel plates using extreme learning machines, Ultrasonics 102 (2020) 106057, https://doi.org/10.1016/j.ultras.2019.106057.
[156] D. Sen, K. Erazo, W. Zhang, S. Nagarajaiah, L. Sun, On the effectiveness of principal component analysis for decoupling structural damage and environmental effects in bridge structures, J. Sound Vib. 457 (2019) 280–298, https://doi.org/10.1016/j.jsv.2019.06.003.
[157] M.R. Azim, M. Gül, Data-driven damage identification technique for steel truss railroad bridges utilizing principal component analysis of strain response, Struct. Infrastruct. Eng. 17 (8) (2021) 1019–1035, https://doi.org/10.1080/15732479.2020.1785512.
[158] Z. Nie, Z. Shen, J. Li, H. Hao, Y. Lin, H. Ma, H. Jiang, Using a single sensor for bridge condition monitoring via moving embedded principal component analysis, Struct. Health Monit. 20 (6) (2021) 3123–3149, https://doi.org/10.1177/1475921720980516.
[159] Y. Yang, S. Nagarajaiah, Blind denoising of structural vibration responses with outliers via principal component pursuit, Struct. Control Health Monit. 21 (6) (2014) 962–978, https://doi.org/10.1002/stc.1624.
[160] X. Hong, B. Zhang, Y. Liu, H. Qi, W. Li, Deep-learning-based guided wave detection for liquid-level state in porcelain bushing type terminal, Struct. Control Health Monitor. 28 (1) (2021), https://doi.org/10.1002/stc.v28.110.1002/stc.2651.
[161] H. Lee, H.J. Lim, T. Skinner, A. Chattopadhyay, A. Hall, Automated fatigue damage detection and classification technique for composite structures using Lamb waves and deep autoencoder, Mech. Syst. Signal Pr. 163 (2022) 108148, https://doi.org/10.1016/j.ymssp.2021.108148.
[162] N. Munir, J. Park, H.-J. Kim, S.-J. Song, S.-S. Kang, Performance enhancement of convolutional neural network for ultrasonic flaw classification by adopting autoencoder, NDT E Int. 111 (2020) 102218, https://doi.org/10.1016/j.ndteint.2020.102218.
[163] M. Rautela, S. Jayavelu, J. Moll, S. Gopalakrishnan, Temperature compensation for guided waves using convolutional denoising autoencoders, P. Soc. Photo.-Opt. Ins. 11593 (2021) 1159319, https://doi.org/10.1117/12.2582986.
[164] F. Gao, B. Li, L. Chen, X. Wei, Z. Shang, C. He, Ultrasonic signal denoising based on autoencoder, Rev. Sci. Instrum. 91 (4) (2020) 045104, https://doi.org/10.1063/1.5136269.
[165] W. Xu, X. Li, J. Zhang, Z. Xue, J. Cao, Ultrasonic signal enhancement for coarse grain materials by machine learning analysis, Ultrasonics 117 (2021) 106550, https://doi.org/10.1016/j.ultras.2021.106550.
[166] K.S. Alguri, J.B. Harley, Transfer learning of ultrasonic guided waves using autoencoders: a preliminary study, AIP Conf. Proc. 2102 (2019) 050013, https://doi.org/10.1063/1.5099779.
[167] J.M. Ha, H.M. Seung, W. Choi, Autoencoder-based detection of near-surface defects in ultrasonic testing, Ultrasonics 119 (2022) 106637. doi:10.1016/j.ultras.2021.106637.

[168] A. Ebrahimkhanlou, B. Dubuc, S. Salamone, A generalizable deep learning framework for localizing and characterizing acoustic emission sources in riveted metallic panels, Mech. Syst. Signal. Pr. 30 (2019) 248–272, https://doi.org/10.1016/j.ymssp.2019.04.050.
[169] Y. Saeys, I. Inza, P. Larranaga, A review of feature selection techniques in bioinformatics, Bioinformatics 23 (19) (2007) 2507–2517, https://doi.org/10.1093/bioinformatics/btm344.
[170] Z.M. Hira, D.F. Gillies, A review of feature selection and feature extraction methods applied on microarray data, Adv. Bioinform. 2015 (2015) 1–13, https://doi.org/10.1155/2015/198363.
[171] V. Bolón-Canedo, N. Sánchez-Maroño, A. Alonso-Betanzos, A review of feature selection methods on synthetic data, Knowl. Inf. Syst. 34 (3) (2013) 483–519, https://doi.org/10.1007/s10115-012-0487-8.
[172] F.W. Margrave, K. Rigas, D.A. Bradley, P. Barrowcliffe, The use of neural networks in ultrasonic flaw detection, Meas.: J. Int. Meas. Confed. 25 (2) (1999) 143–154, https://doi.org/10.1016/S0263-2241(98)00075-X.
[173] S. Legendre, D. Massicote, J. Goyette, T.K. Bose, Neural classification of lamb wave ultrasonic weld testing signals using wavelet coefficients, IEEE T. Instrum. Meas. 50 (2001) 672–678, https://doi.org/10.1109/19.930439.
[174] M.A. Kewalramani, R. Gupta, Concrete compressive strength prediction using ultrasonic pulse velocity through artificial neural networks, Automat. Constr. 15 (3) (2006) 374–379, https://doi.org/10.1016/j.autcon.2005.07.003.
[175] Ó. Martín, M. López, F. Martín, Artificial neural networks for quality control by ultrasonic testing in resistance spot welding, J. Mater. Process. Tech. 183 (2-3) (2007) 226–233, https://doi.org/10.1016/j.jmatprotec.2006.10.011.
[176] R. Madandoust, R. Ghavidel, N. Nariman-zadeh, Evolutionary design of generalized GMDH-type neural network for prediction of concrete compressive strength using UPV, Comp. Mater. Sci. 49 (3) (2010) 556–567, https://doi.org/10.1016/j.commatsci.2010.05.050.
[177] S. Agarwal, M. Mitra, Lamb wave based automatic damage detection using matching pursuit and machine learning, Smart Mater. Struct. 23 (8) (2014) 085012, https://doi.org/10.1088/0964-1726/23/8/085012.
[178] M.M.F. Figueredo, J.L. Goncalves, A.M.V. Nakashima, A.M.F. Fileti, R.D. M. Carvalho, The use of an ultrasonic technique and neural networks for identification of the flow pattern and measurement of the gas volume fraction in multiphase flows, Exp. Therm. Fluid. Sci. 70 (2016) 29–50, https://doi.org/10.1016/j.expthermflusci.2015.08.010.
[179] Y. Nagatani, S. Okumura, S. Wu, Neural network based bone density estimation from the ultrasound waveforms inside cancellous bone derived by FDTD simulations, IEEE Int. Ultra. Sym. (2018) 8580010, https://doi.org/10.1109/ULTSYM.2018.8580010.
[180] J.Y. Park, Y.G. Yoon, T.K. Oh, Prediction of concrete strength with P-, S-, R-wave velocities by support vector machine (SVM) and artificial neural network (ANN), Appl. Sci.-Basel 9 9 (19) (2019) 4053, https://doi.org/10.3390/app9194053.
[181] K. Virupakshappa, E. Oruklu, Investigation of feature inputs for binary classification of ultrasonic NDT signals using SVM and neural networks, Midwest Symp. Circuit. (2019) 638–641, https://doi.org/10.1109/MWSCAS.2019.8884852.
[182] T.H.L. Nguyen, S. Park, Intelligent ultrasonic flow measurement using linear array transducer with recurrent neural networks, IEEE Access 8 (2020) 137564–137573, https://doi.org/10.1109/ACCESS.2020.3012037.
[183] S. Godfrey Nnabuife, B. Kuang, J.F. Whidborne, Z. Rana, Non-intrusive classification of gas-liquid flow regimes in an S-shaped pipeline riser using a Doppler ultrasonic sensor and deep neural networks, Chem. Eng. J. 403 (2021) 126401, https://doi.org/10.1016/j.cej.2020.126401.
[184] S.-H. Park, J.-Y. Hong, T. Ha, S. Choi, K.-Y. Jhang, Deep learning-based ultrasonic testing to evaluate the porosity of additively manufactured parts with rough surfaces, Metals-Basel 11 (2021) 290, https://doi.org/10.3390/met11020290.
[185] S. Lari, Y. Qian, H.-J. Kwon, Assessment of geometrical features of internal flaws with artificial neural network, Int. J. Precis. Eng. Man. 22 (5) (2021) 777–789, https://doi.org/10.1007/s12541-021-00515-z.
[186] W. Ren, N. Jin, L. OuYang, L. Zhai, Y. Ren, Gas volume fraction measurement of oil-gas-water three-phase flows in vertical pipe by combining ultrasonic sensor and deep attention network, IEEE T. Instrum. Meas. 70 (2021) 1–9, https://doi.org/10.1109/TIM.2020.3031186.
[187] Y. Qin, M. Ma, E. Zhu, Z. Mao, M. Haile, M. Shiao, T.-K. Chen, Temperature compensation of ultrasonic guided waves via recurrent neural network, Proceedings of the 7th Asia-Pacific Workshop on Structural Health Monitoring, APWSHM 2018 (2018) 402–409.
[188] Y. Guo, Z. Xiao, L. Geng, J. Wu, F. Zhang, Y. Liu, W. Wang, Fully convolutional neural network with GRU for 3D braided composite material flaw detection, IEEE Access 7 (2019) 151180–151188, https://doi.org/10.1109/ACCESS.2019.2946447.
[189] T. Gantala, K. Balasubramaniam, DPAI: A Data-driven simulation-assisted-Physics learned AI model for transient ultrasonic wave propagation, Ultrasonics 121 (2022) 106671, https://doi.org/10.1016/j.ultras.2021.106671.
[190] L. Huang, X. Hong, Z. Yang, Y. Liu, B. Zhang, CNN-LSTM network-based damage detection approach for copper pipeline using laser ultrasonic scanning, Ultrasonics 121 (2022) 106685, https://doi.org/10.1016/j.ultras.2022.106685.
[191] J. Melville, K.S. Alguri, C. Deemer, J.B. Harley, Structural damage detection using deep learning of ultrasonic guided waves, AIP Conf. Proc. 1949 (2018) 230004, https://doi.org/10.1063/1.5031651.
[192] M. de Oliveira, A. Monteiro, J. Vieira Filho, A new structural health monitoring strategy based on PZT sensors and convolutional neural network, Sensors (Basel) 18 (9) (2018) 2955, https://doi.org/10.3390/s18092955.
[193] C. Hu, B. Yang, J. Yan, Y. Xiang, S. Zhou, F.-Z. Xuan, Damage localization in pressure vessel by guided waves based on convolution neural network approach, J. Press. Vessel Technol. 142 (6) (2020), https://doi.org/10.1115/1.4047213.
[194] T. Zhang, Flow measurement of natural gas in pipeline based on 1d-convolutional neural network, Int. J. Comput. Int. Sys. 13 (2020) 1198–1206, https://doi.org/10.2991/ijcis.d.200803.002.
[195] H.J. Lim, H. Sohn, Online stress monitoring technique based on lamb-wave measurements and a convolutional neural network under static and dynamic loadings, Exp. Mech. 60 (2) (2020) 171–179, https://doi.org/10.1007/s11340-019-00546-8.
[196] Y. Li, K. Xu, Y. Li, F. Xu, D. Ta, W. Wang, Deep learning analysis of ultrasonic guided waves for cortical bone characterization, IEEE T Ultrason. Ferr. 68 (4) (2021) 935–951, https://doi.org/10.1109/TUFFC.2020.3025546.
[197] B. Filipovic, F. Milkovic, M. Subasic, S. Loncaric, T. Petkovic, M. Budimir, Automated ultrasonic testing of materials based on C-scan flaw classification, Int. Symp. Image. Sig. (2021) 230–234, https://doi.org/10.1109/ISPA52656.2021.9552056.
[198] R.J. Pyle, R.L.T. Bevan, R.R. Hughes, R.K. Rachev, A.A.S. Ali, P.D. Wilcox, Deep learning for ultrasonic crack characterization in NDE, IEEE T. Ultrason. Ferr. 68 (5) (2021) 1854–1865, https://doi.org/10.1109/TUFFC.2020.3045847.
[199] A. Chapon, D. Pereira, M. Toews, P. Belanger, Deconvolution of ultrasonic signals using a convolutional neural network, Ultrasonics 111 (2021) 106312, https://doi.org/10.1016/j.ultras.2020.106312.
[200] F. Gao, B. Li, L. Chen, Z. Shang, X. Wei, C. He, A softmax classifier for high-precision classification of ultrasonic similar signals, Ultrasonics 112 (2021) 106344, https://doi.org/10.1016/j.ultras.2020.106344.
[201] Z. Xiao, Q. Guo, Y. Guo, Y. Huang, Ultrasonic A-scan image detection for 3D braided composites based on convolutional neural network, ICCAI 2021 (2021) 8–14, https://doi.org/10.1145/3467707.3467709.
[202] K. Gopalakrishnan, M. Rautela, Y. Deng, Deep learning based identification of elastic properties using ultrasonic guided waves, Lect. Notes Civ. Eng 128 (2021) 77–90, https://doi.org/10.1007/978-3-030-64908-1_8.
[203] M. Rautela, J. Senthilnath, J. Moll, S. Gopalakrishnan, Combined two-level damage identification strategy using ultrasonic guided waves and physical knowledge assisted machine learning, Ultrasonics 115 (2021) 106451, https://doi.org/10.1016/j.ultras.2021.106451.
[204] I. Virkkunen, T. Koskinen, O. Jessen-Juhler, J. Rinta-aho, Augmented ultrasonic data for machine learning, J. Nondestruct. Eval. 40 (2021) 4, https://doi.org/10.1007/s10921-020-00739-5.
[205] A. Rai, M. Mitra, Lamb wave based damage detection in metallic plates using multi-headed 1-dimensional convolutional neural network, Smart Mater. Struct. 30 (3) (2021) 035010, https://doi.org/10.1088/1361-665X/abdd00.
[206] S. Mariani, Q. Rendu, M. Urbani, C. Sbarufatti, Causal dilated convolutional neural networks for automatic inspection of ultrasonic signals in non-destructive evaluation and structural health monitoring, Mech. Syst. Signal Pr. 157 (2021) 107748, https://doi.org/10.1016/j.ymssp.2021.107748.
[207] J. Park, S.-E. Lee, H.-J. Kim, S.-J. Song, S.-S. Kang, System invariant method for ultrasonic flaw classification in weldments using residual neural network, Appl. Sci.-Basel 12 (3) (2022) 1477, https://doi.org/10.3390/app12031477.
[208] T. Koskinen, I. Virkkunen, O. Siljama, O. Jessen-Juhler, The effect of different flaw data to machine learning powered ultrasonic inspection, J. Nondestruct. Eval. 40 (1) (2021), https://doi.org/10.1007/s10921-021-00757-x.
[209] O. Siljama, T. Koskinen, O. Jessen-Juhler, I. Virkkunen, Automated flaw detection in multi-channel phased array ultrasonic data using machine learning, J. Nondestruct. Eval. 40 (3) (2021), https://doi.org/10.1007/s10921-021-00796-4.
[210] Y. Cai, Y. Song, P. Ni, X. Liu, X. Li, Subwavelength ultrasonic imaging using a deep convolutional neural network trained on structural noise, Ultrasonics 117 (2021) 106552, https://doi.org/10.1016/j.ultras.2021.106552.
[211] R. Miorelli, C. Fisher, A. Kulakovskyi, B. Chapuis, O. Mesnil, O. D'Almeida, Defect sizing in guided wave imaging structural health monitoring using convolutional neural networks, NDT&E Int. 122 (2021) 102480, https://doi.org/10.1016/j.ndteint.2021.102480.
[212] J. Ye, N. Toyama, Automatic defect detection for ultrasonic wave propagation imaging method using spatio-temporal convolution neural networks, Struct. Health Monit. (2022), https://doi.org/10.1177/14759217211073503.
[213] H. Sun, L. Peng, S. Wang, S. Huang, K. Qu, Development of frequency-mixed point-focusing shear horizontal guided-wave EMAT for defect inspection using deep neural network, IEEE T. Instrum. Meas. 70 (2021) 1–14, https://doi.org/10.1109/TIM.2020.3033941.
[214] J.D. Eckels, E.M. Jacobson, I.T. Cummings, I.F. Fernandez, K. Ho, N. Dervilis, E. B. Flynn, A.J. Wachtor, Predicting local material thickness from steady-state ultrasonic wavefield measurements using a convolutional neural network, Ultrasonics 123 (2022) 106661, https://doi.org/10.1016/j.ultras.2021.106661.
[215] F. Gao, J. Hua, Damage characterization using CNN and SAE of broadband Lamb waves, Ultrasonics 119 (2022) 106592, https://doi.org/10.1016/j.ultras.2021.106592.
[216] A. Arcos Jiménez, C.Q. Gómez Muñoz, F.P. García Márquez, Dirt and mud detection and diagnosis on a wind turbine blade employing guided waves and supervised learning classifiers, Reliab. Eng. Syst. Safe 184 184 (2019) 2–12, https://doi.org/10.1016/j.ress.2018.02.013.
[217] L.F.M. Rodrigues, F.C. Cruz, M.A. Oliveira, E.F. Simas Filho, M.C.S. Albuquerque, I.C. Silva, C.T.T. Farias, Carburization level identification in industrial HP pipes using ultrasonic evaluation and machine learning, Ultrasonics 94 (2019) 145–151, https://doi.org/10.1016/j.ultras.2018.10.005.

[218] K. Lee, V. Estivill-Castro, Feature extraction and gating techniques for ultrasonic shaft signal classification, Appl. Soft Comput. J. 7 (1) (2007) 156–165, https://doi.org/10.1016/j.asoc.2005.05.003.
[219] M. Cacciola, S. Calcagno, F.C. Morabito, M. Versaci, Computational intelligence aspects for defect classification in aeronautic composites by using ultrasonic pulses, IEEE T. Ultrason. Ferr. 55 (4) (2008) 870–878, https://doi.org/10.1109/TUFFC.2008.722.
[220] S. Saechi, W. Kongprawechnon, R. Sahamitmongkol, Test system for defect detection in construction materials with ultrasonic waves by support vector machine and neural network, in: 6th International Conference on Soft Computing and Intelligent Systems, and 13th International Symposium on Advanced Intelligence Systems, 2012, pp. 1034–1039, https://doi.org/10.1109/SCIS-ISIS.2012.6505090.
[221] L.H. Lee, R. Rajkumar, L.H. Lo, C.H. Wan, D. Isa, Oil and gas pipeline failure prediction system using long range ultrasonic transducers and euclidean-support vector machines classification approach, Exp. Syst. Appl. 40 (6) (2013) 1925–1934, https://doi.org/10.1016/j.eswa.2012.10.006.
[222] H. Zhao, L. Peng, T. Takahashi, T. Hayashi, K. Shimizu, T. Yamamoto, Support vector regression-based data integration method for multipath ultrasonic flowmeter, IEEE T. Instrum. Meas. 63 (12) (2014) 2717–2725, https://doi.org/10.1109/TIM.2014.2326276.
[223] P. Yang, Q. Li, Wavelet transform-based feature extraction for ultrasonic flaw signal classification, Neural. Comput. Appl. 24 (3-4) (2014) 817–826, https://doi.org/10.1007/s00521-012-1305-7.
[224] H. Zamani Hosseinabadi, R. Amirfattahi, B. Nazari, H.R. Mirdamadi, S. A. Atashipour, GUW-based structural damage detection using WPT statistical features and multiclass SVM, Appl. Acoust. 86 (2014) 59–70, https://doi.org/10.1016/j.apacoust.2014.05.002.
[225] Y.-F. Shih, Y.-R. Wang, K.-L. Lin, C.-W. Chen, Improving non-destructive concrete strength tests using support vector machines, Materials 8 (2015) 7169–7178, https://doi.org/10.3390/ma8105368.
[226] Z. Yang, Q.u. Zhou, X. Wu, Z. Zhao, A novel measuring method of interfacial tension of transformer oil combined PSO optimized SVM and multi frequency ultrasonic technology, IEEE Access 7 (2019) 182624–182631, https://doi.org/10.1109/ACCESS.2019.2954899.
[227] H. Xiao, D. Chen, J. Xu, S. Guo, Defects identification using the improved ultrasonic measurement model and support vector machines, NDT&E Int. 111 (2020) 102223, https://doi.org/10.1016/j.ndteint.2020.102223.
[228] J.-G. Minonzio, B. Cataldo, R. Olivares, D. Ramiandrisoa, R. Soto, B. Crawford, V. H.C. De Albuquerque, R. Munoz, Automatic classifying of patients with non-traumatic fractures based on ultrasonic guided wave spectrum image using a dynamic support vector machine, IEEE Access 8 (2020) 194752–194764, https://doi.org/10.1109/ACCESS.2020.3033480.
[229] R. Miorelli, A. Kulakovskyi, B. Chapuis, O. D'Almeida, O. Mesnil, Supervised learning strategy for classification and regression tasks applied to aeronautical structural health monitoring problems, Ultrasonics 113 (2021) 106372, https://doi.org/10.1016/j.ultras.2021.106372.
[230] C.A. Lindley, S. Beamish, R.S. Dwyer-Joyce, N. Dervilis, K. Worden, A Bayesian approach for shaft centre localisation in journal bearings, Mech. Syst. Signal. Pr. 174 (2022) 109021, https://doi.org/10.1016/j.ymssp.2022.109021.
[231] M.R. Jones, T.J. Rogers, K. Worden, E.J. Cross, A Bayesian methodology for localising acoustic emission sources in complex structures mech, Syst. Signal. Pr. 163 (2022) 108143, https://doi.org/10.1016/j.ymssp.2021.108143.
[232] M. Haywood-Alexander, N. Dervilis, K. Worden, E.J. Cross, R.S. Mills, T.J. Rogers, Structured machine learning tools for modelling characteristics of guided waves, Mech. Syst. Signal. Pr. 156 (2021) 107628, https://doi.org/10.1016/j.ymssp.2021.107628.
[233] M. Haywood-Alexander, N. Dervilis, K. Worden, G.B. Dobie, T.J. Rogers, Decomposition of multi-mode signals using dispersion curves and Bayesian linear regression, Proc. SPIE (2021), https://doi.org/10.1117/12.2582967.
[234] Z. Ma, W. Zhang, Z. Luo, X.u. Sun, Z. Li, L.i. Lin, Ultrasonic characterization of thermal barrier coatings porosity through BP neural network optimizing Gaussian process regression algorithm, Ultrasonics 100 (2020) 105981, https://doi.org/10.1016/j.ultras.2019.105981.
[235] M.A. Oliveira, E.F. Simas Filho, M.C.S. Albuquerque, Y.T.B. Santos, I.C. da Silva, C.T.T. Farias, Ultrasound-based identification of damage in wind turbine blades using novelty detection, Ultrasonics 108 (2020) 106166, https://doi.org/10.1016/j.ultras.2020.106166.
[236] L. Posilović, D. Medak, F. Milković, M. Subašić, M. Budimir, S. Lončarić, Deep learning-based anomaly detection from ultrasonic images, Ultrasonics 124 (2022) 106737, https://doi.org/10.1016/j.ultras.2022.106737.
[237] F. Milkovic, B. Filipovic, M. Subasic, T. Petkovic, S. Loncaric, M. Budimir, Ultrasound anomaly detection based on variational autoencoders, Int. Symp. Image. Sig. (2021) 225–229, https://doi.org/10.1109/ISPA52656.2021.9552041.
[238] I. Kraljevski, F. Duckhorn, C. Tschope, M. Wolff, machine learning for anomaly assessment in sensor networks for NDT in aerospace, IEEE Sens. J. 21 (9) (2021) 11000–11008, https://doi.org/10.1109/JSEN.2021.3062941.
[239] A. Bowler, M. Pound, N. Watson, Convolutional feature extraction for process monitoring using ultrasonic sensors, Comput. Chem. Eng. 155 (2021) 107508, https://doi.org/10.1016/j.compchemeng.2021.107508.
[240] S. Zhang, C.M. Li, W. Ye, Damage localization in plate-like structures using time-varying feature and one-dimensional convolutional neural network, Mech. Syst. Signal Pr. 147 (2021) 107107, https://doi.org/10.1016/j.ymssp.2020.107107.
[241] T. Latête, B. Gauthier, P. Belanger, Towards using convolutional neural network to locate, identify and size defects in phased array ultrasonic testing, Ultrasonics 115 (2021) 106436, https://doi.org/10.1016/j.ultras.2021.106436.
[242] D. Sen, A. Aghazadeh, A. Mousavi, S. Nagarajaiah, R. Baraniuk, A. Dabak, Data-driven semi-supervised and supervised learning algorithms for health monitoring of pipes, Mech. Syst. Signal. Pr. 131 (2019) 524–537, https://doi.org/10.1016/j.ymssp.2019.06.003.
[243] A.E. Bouzenad, M.E. Mountassir, S. Yaacoubi, F. Dahmene, M. Koabaz, L. Buchheit, W. Ke, A semi-supervised based k-means algorithm for optimal guided waves structural health monitoring: a case study, Inventions 4 (2019) 17, https://doi.org/10.3390/inventions4010017.
[244] L.A. Bull, K. Worden, N. Dervilis, Towards semi-supervised and probabilistic classification in structural health monitoring, Mech. Syst. Signal. Pr. 140 (2020) 106653, https://doi.org/10.1016/j.ymssp.2020.106653.
[245] L.A. Bull, T.J. Rogers, C. Wickramarachchi, E.J. Cross, K. Worden, N. Dervilis, Probabilistic active learning: an online framework for structural health monitoring, Mech. Syst. Signal. Pr. 134 (2019) 106294, https://doi.org/10.1016/j.ymssp.2019.106294.
[246] A.J. Hughes, L.A. Bull, P. Gardner, R.J. Barthorpe, N. Dervilis, K. Worden, On risk-based active learning for structural health monitoring, Mech. Syst. Signal. Pr. 167 (2022) 108569, https://doi.org/10.1016/j.ymssp.2021.108569.
[247] L. Posilović, D. Medak, M. Subašić, M. Budimir, S. Lončarić, Generative adversarial network with object detector discriminator for enhanced defect detection on ultrasonic B-scans, Neurocomputing 459 (2021) 361–369, https://doi.org/10.1016/j.neucom.2021.06.094.
[248] L. Posilović, D. Medak, M. Subašić, M. Budimir, S. Lončarić, S., Generating ultrasonic images indistinguishable from real images using Generative Adversarial Networks. Ultrasonics 119 (2022). doi:10.1016/j.ultras.2021.106610.
[249] T. Gantala, K. Balasubramaniam, Automated defect recognition for welds using simulation assisted TFM imaging with artificial intelligence, J. Nondestruct. Eval. 40 (1) (2021), https://doi.org/10.1007/s10921-021-00761-1.
[250] K. Virupakshappa, E. Oruklu, Using generative adversarial networks to generate ultrasonic signals, IEEE Int. Ultra. Sym. (2020), https://doi.org/10.1109/IUS46767.2020.9251382.
[251] A. Look, O. Kirschner, S. Riedelbauch, Building robust classifiers with generative adversarial networks for detecting cavitation in hydraulic turbines, ICPRAM 2018 (2018) 456–462, https://doi.org/10.5220/0006636304560462.
[252] J. Ye, N. Toyama, Benchmarking deep learning models for automatic ultrasonic imaging inspection, IEEE Access 9 (2021) 36986–36994, https://doi.org/10.1109/ACCESS.2021.3062860.
[253] G.P. Tsialiamanis, D.J. Wagg, P.A. Gardner, N. Dervilis, K. Worden, On partitioning of an SHM problem and parallels with transfer learning, Conf. Proc. Soc. Exp. Mech. Ser. (2021) 41–50, https://doi.org/10.1007/978-3-030-47717-2_5.
[254] D.F. Hesser, S. Mostafavi, G.K. Kocur, B. Markert, Identification of acoustic emission sources for structural health monitoring applications based on convolutional neural networks and deep transfer learning, Neurocomputing 453 (2021) 1–12, https://doi.org/10.1016/j.neucom.2021.04.108.
[255] B. Ren, J. Chen, Fracture acoustic emission signals identification of broken wire using deep transfer learning and wavelet analysis, ACM Int. Conf. Proc. Ser. (2021), https://doi.org/10.1145/3469213.3470230.
[256] G. Reyes-Carmenaty, M.A. Pérez, Use of transfer learning for detection of structural alterations, Proc. Comput. Sci. 200 (2022) 1368–1377, https://doi.org/10.1016/j.procs.2022.01.338.
[257] M. Azimi, G. Pekcan, Structural health monitoring using extremely compressed data through deep learning, Comput.-Aid. Civil Infrastruct. Eng. 35 (6) (2020) 597–614.
[258] A.L. Bowler, M.P. Pound, N.J. Watson, Domain adaptation and federated learning for ultrasonic monitoring of beer fermentation, Fermentation 7 (2021) 253, https://doi.org/10.3390/fermentation7040253.
[259] X. Gao, Y. Shi, Q.i. Zhu, Z. Li, H.u. Sun, Z. Yao, W. Zhang, Domain adaptation in intelligent ultrasonic logging tool: from microseismic to pulse-echo, IEEE T. Instrum. Meas. 70 (2021) 1–14, https://doi.org/10.1109/TIM.2021.3050154.
[260] K.S. Alguri, C.C. Chia, J.B. Harley, Sim-to-Real: Employing ultrasonic guided wave digital surrogates and transfer learning for damage visualization, Ultrasonics 111 (2021) 106338, https://doi.org/10.1016/j.ultras.2020.106338.
[261] L.A. Bull, P.A. Gardner, N. Dervilis, E. Papatheou, M. Haywood-Alexander, R. S. Mills, K. Worden, On the transfer of damage detectors between structures: an experimental case study, J. Sound Vib. 501 (2021) 116072, https://doi.org/10.1016/j.jsv.2021.116072.
[262] P. Gardner, L.A. Bull, N. Dervilis, K. Worden, Overcoming the problem of repair in structural health monitoring: metric-informed transfer learning, J. Sound Vib. 510 (2021) 116245, https://doi.org/10.1016/j.jsv.2021.116245.
[263] P. Gardner, L.A. Bull, N. Dervilis, K. Worden, Domain-adapted Gaussian mixture models for population-based structural health monitoring, J. Civ. Struct. Health Monit. (2022), https://doi.org/10.1007/s13349-022-00565-5.
[264] P. Gardner, L.A. Bull, N. Dervilis, K. Worden, On the application of kernelised Bayesian transfer learning to population-based structural health monitoring, Mech. Syst. Signal Process. 167 (2022) 108519, https://doi.org/10.1016/j.ymssp.2021.108519.
[265] P. Gardner, L.A. Bull, J. Gosliga, J. Poole, N. Dervilis, K. Worden, A population-based SHM methodology for heterogeneous structures: Transferring damage localisation knowledge between different aircraft wings, Mech. Syst. Signal. Pr. 172 (2022) 108918, https://doi.org/10.1016/j.ymssp.2022.108918.

[266] D. Medak, L. Posilovic, M. Subasic, M. Budimir, S. Loncaric, Automated defect detection from ultrasonic images using deep learning, IEEE T. Ultrason. Ferr. 68 (10) (2021) 3126–3134, https://doi.org/10.1109/TUFFC.2021.3081750.

[267] L. Posilovic, D. Medak, M. Subasic, T. Petkovic, M. Budimir, S. Loncaric, Flaw detection from ultrasonic images using YOLO and SSD, Int. Symp. Image. Sig. (2019) 163–168, https://doi.org/10.1109/ISPA.2019.8868929.

[268] D. Medak, L. Posilović, M. Subašić, M. Budimir, S. Lončarić, DefectDet: a deep learning architecture for detection of defects with extreme aspect ratios in ultrasonic images, Neurocomputing 473 (2022) 107–115, https://doi.org/10.1016/j.neucom.2021.12.008.

[269] D. Medak, L. Posilovic, M. Subasic, M. Budimir, S. Loncaric, Deep learning-based defect detection from sequences of ultrasonic B-scans, IEEE Sens. J. 22 (3) (2022) 2456–2463, https://doi.org/10.1109/JSEN.2021.3134452.

[270] A.B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, F. Herrera, Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Inform. Fusion 58 (2020) 82–115, https://doi.org/10.1016/j.inffus.2019.12.012.

[271] P. Linardatos, V. Papastefanopoulos, S. Kotsiantis, Explainable AI: a review of machine learning interpretability methods, Entropy 23 (2021) 18, https://doi.org/10.3390/e23010018.