Convolutional feature extraction for process monitoring using ultrasonic sensors

Alexander Bowler[a], Michael Pound[b], Nicholas 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

E-mail addresses: alexander.bowler@nottingham.ac.uk (A.B.), michael.pound@nottingham.ac.uk (M.P.), nicholas.watson@nottingham.ac.uk (N.W.)

* Corresponding author

Abstract

Ultrasonic sensors are a low-cost and in-line technique and can be combined with machine learning for industrial process monitoring. However, training accurate machine learning models for process monitoring using sensor data is dependent on the feature selection methodology. This paper compares a convolutional feature extraction method to a traditional, coarse feature engineering approach. The convolutional method uses filter weights pre-trained on an auxiliary task to classify ultrasonic waveform dataset membership using previously obtained sensor data. The filter weights are used to extract features from the ultrasonic waveform. Principal component analysis is then applied to produce five principal components to be input into long short-term memory neural networks. The two approaches are compared on fermentation, mixing and cleaning datasets monitored using ultrasonic sensors. Overall, the convolutional feature method produced more informative waveform features than the coarse feature engineering approach, achieving higher model accuracy for datasets requiring substantial waveform information and for 65% of tasks overall. Multi-task learning also improved feature trajectory learning but led to reduced model accuracy for data points far from the classification decision boundaries. This can be overcome by further optimisation of neural network hyperparameters, though at increased model development time. Once trained, the convolutional feature extraction approach is a fast and convenient way of producing high quality features from ultrasonic waveforms using convolutional neural networks with little training data.

Keywords

Convolutional neural networks; Ultrasonic sensors; Transfer learning; Process monitoring; Industrial digital technologies, Multi-task learning

1 Introduction

The fourth industrial revolution, also termed Industry 4.0, has the potential to improve the productivity, efficiency, and sustainability of process manufacturing (Sjödin et al., 2018). This will be via the implementation of industrial digital technologies which include: The Internet of Things to enable connectivity between devices; Cloud, Fog, and Edge Computing to process large data stream (Chen and Ran, 2019; Wu et al., 2017); and Machine Learning (ML) to provide automatic data analysis and decision making. Industry 4.0 requires continuous data streams to enable real-time communication across processes, markets, and supply chains. Therefore, in-line and on-line sensors are a key technology in this transformation as they provide process data with no human intervention. In-line sensors directly measure the process stream while on-line sensors use automatic sampling systems (De Beer et al., 2011).

Ultrasonic (US) sensors have the benefits of being: low-cost, in-line, real-time, able to be non-invasive, small in size, low energy consuming, non-destructive, and able to characterise opaque materials. US sensors have been widely applied across manufacturing, such as fermentation (Ojha et al., 2017), polymerisation, crystallisation (Henning and Rautenberg, 2006), and food product analysis (Awad et al., 2012, Mohd Khairi et al., 2015). US sensors consist of a piezoelectric transducer which converts electrical pulses into sound waves and vice versa. Single sensors may be used in pulse-echo mode, where the sound wave is reflected back to the transducer from an interface between two neighbouring materials, or in pitch-catch mode where a second sensor receives the sound wave after it has been transmitted through a material (Awad et al., 2012). High frequency (>1 MHz), low power (<1 Wcm^{-2}) sound waves are used which do not affect the structure of the material that they pass through (Ojha et al., 2017). However, US properties are highly dependent on temperature and large changes in the acoustic impedance at a material interface (e.g. if gas bubbles are present in a liquid) causes strong reflection of the sound waves making transmission techniques difficult to use for many industrial applications (Henning and Rautenberg, 2006).

Traditionally, either first principle or empirical correlations are used to determine material properties from US sensor data, or waveforms. However, first principle models soon become complex under industrial conditions, where the sound wave travels through multiple interfaces and process parameters (e.g. temperature) are changing. Similarly, empirical models require extensive calibration to account for all process parameter variations. In contrast, ML can be used to predict material properties without extensive calibration procedures by learning the relationships between these variations and the US waveform. ML also provides automatic interpretation of the sensor data. For example: training an ML model to predict the processing time remaining would enable improved batch scheduling; classifying the end of processing would reduce resource consumption; and anomaly detection methods would provide early warning of problems with batches and ensure product quality.

During training, ML models fit input data, or features, to the desired prediction outputs. The success of the ML models is partly dependent on the choice of features used for the model. For ultrasonic techniques, the speed of sound is commonly used as a feature as it is dependent on the density and compressibility of the material it passes through and is calculated by measuring the sound wave time of flight and distance travelled (Utomo et al., 2001, Utomo et al., 2002, Supardan et al., 2003, Sun et al., 2005). The changing amplitude between consecutively acquired waveforms can be used as a feature to identify process states and has been applied to determine flow regimes (Ren et al., 2021; Abbagoni and Yeung, 2016). Other process information can also be used to aid the prediction accuracy of the ML model, such as the temperature, material composition and concentration (Sun et al., 2005), or mass flow rate (Wallhäußer et al., 2014). Along with these features, measurements that describe the oscillations of the waveform are also required. The energy of the waveform (the sum of the squared amplitudes at each point in the waveform) may be used to monitor attenuation of the sound wave as it passes through a material (Utomo et al., 2001, Utomo et al., 2002, Supardan et al., 2003, Sun et al., 2005) or to monitor a change in acoustic impedance by measuring the proportion of the sound wave reflected from a material boundary (Wallhäußer et al., 2013, Wallhäußer et al., 2014, Figueiredo et al. 2016). However, the energy may not account for all the changes to the waveform, as some peaks may increase in amplitude while others decrease, or the waveform could be composed of multiple overlapping sound waves. Further features can be extracted which describe the shape of the waveform by monitoring information such as maximum amplitudes, variance in the amplitudes, the rising and falling slopes of the waveform, the duration of the waveform, and the relationship between all of these (Wallhäußer et al., 2013, Wallhäußer et al., 2014, Cau et al., 2005). Nevertheless, this is still a coarse method of monitoring waveform changes, which are indirectly measured rather than directly identified. Signal features similar to those previously listed can also be extracted in the frequency domain, commonly after using the discrete wavelet transform (Cau et al., 2005, Simeone et al., 2020). However, US transducers used for material characterisation typically have narrow frequency bands. Therefore, areas where the waveform changes or overlaps may be mis-identified as frequency changes. The amplitudes at each sample point in the time domain waveform can also be used as individual features (Escrig et al., 2020a, Escrig et al., 2020b, Munir et al., 2018). Though, should a peak translate along sample points, whether due to changes to the monitored materials or a change in temperature, the information regarding this part of the waveform is lost.

Convolution Neural Networks (CNNs) overcome these issues by using convolutional filters to measure spatial relationships in the waveform. CNNs use representation learning to automatically extract features by transforming the data into higher, more abstract levels (Lecun et al., 2015). CNNs have been used previously with US signals (Virupakshappa et al., 2018, Meng et al., 2017, Munir et al., 2019, Munir et al., 2020, Bowler et al., 2020). However, previous work has also shown that Long Short-Term Memory (LSTMs) neural network layers are required to accurately monitor time-evolving processes (Bowler et al., 2020 and 2021). LSTMs are able to retain process information from previous timesteps and are a type of recurrent neural network which uses gate units to reduce the likelihood of vanishing or exploding gradients. This enables them to be used over much longer sequences (Hochreiter and Schmidhuber, 1997). Previous time-step information could also be included in CNN inputs or even fully-connected neural networks; however, LSTMs are more memory efficient than fully connected structures and are better equipped to handle long sequences and sequences of varying length. In this work a pre-trained CNN is used to extract features from the waveform. The CNN is pre-trained on an auxiliary task using previously collected US data. The auxiliary task is to classify which dataset each US waveform belongs to. This is a transfer learning task, in which the CNN learns features of a US waveform in the auxiliary task which are then used to aid prediction on the main tasks. Augmentation of the waveforms for the auxiliary task is used to improve CNN feature learning. Furthermore, Principal Component Analysis (PCA) is applied to these extracted features to enable the use of additional features (such as the speed of sound, changes between consecutively acquired waveforms, the process temperature, feature gradients, and time-lagged representations of waveform features) and to reduce the dimensionality of the extracted features to improve LSTM unit training accuracy and stability. The extracted Principal Components (PCs) and additional features are used as input features to the LSTM models. The novelty of this work can be summarised as: the use of CNN extracted features from US waveforms used as inputs to LSTM models, the pre-training of a CNN on an auxiliary task to identify features in US waveforms, using previously collected US datasets to improve ML model prediction through transfer learning, and applying PCA to CNN extracted US features to enable the use of additional features. The convolutional feature extraction method is compared to traditional, coarse features extracted from the time-domain waveform, such as the waveform energy, peak-to-peak amplitude or sample point position of the maximum peak. The benefits of another type of transfer learning, multi-task learning, to tasks which require multiple outputs is also evaluated throughout. The feature extraction and ML methods are compared on previously collected fermentation, cleaning, and mixing process monitoring tasks to provide a comprehensive evaluation of their advantages.

2 Method

2.1 Ultrasonic data collection

For all experiments, a US box (Lecoeur Electronique) was used to excite the transducers and digitise the received sound waves. The temperature sensors were connected to a PT-104 Data Logger (Pico Technology). The US box and temperature data logger were connected to a laptop and a bespoke MATLAB software controlled the hardware components and acquired the data.

2.1.1 Beer fermentation

Full experimental details are provided in Bowler et al. (2021). The fermentation batches were conducted in a 30 l cylindrical plastic vessel. A US probe consisting of a US transducer (Sonatest, 2 MHz central frequency) and a temperature sensor (RTD, PT1000) was installed into the vessel wall. A Tilt hydrometer provided real-time density measurements of the wort. 1.5 kg of malt (Coopers Real Ale), 1 kg of brewing sugar (The Home Brew Shop) and yeast (Coopers Real Ale) were used. In total, 13 batches were completed with the fermentation lasting between 4 to 7 days. The US waveform consisted of two sound wave reflections: the first from the interface between the probe material and the wort, and the second being transmitted through the wort and reflecting from the far probe interface (Fig. 1). The US and temperature data were collected periodically. Each set of collected data consisted of 36 US waveforms and temperature readings. The US waveforms were averaged for each set to minimise noise disturbance. Between the collection of each set of data, 200 s elapsed.

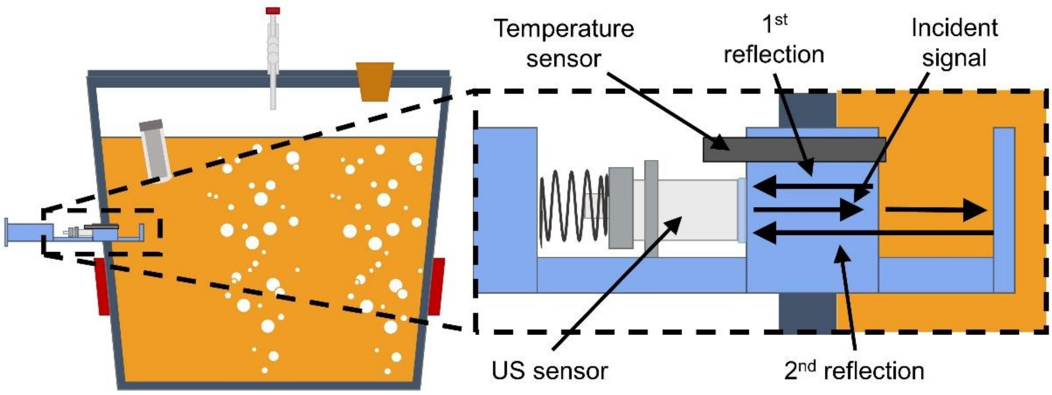
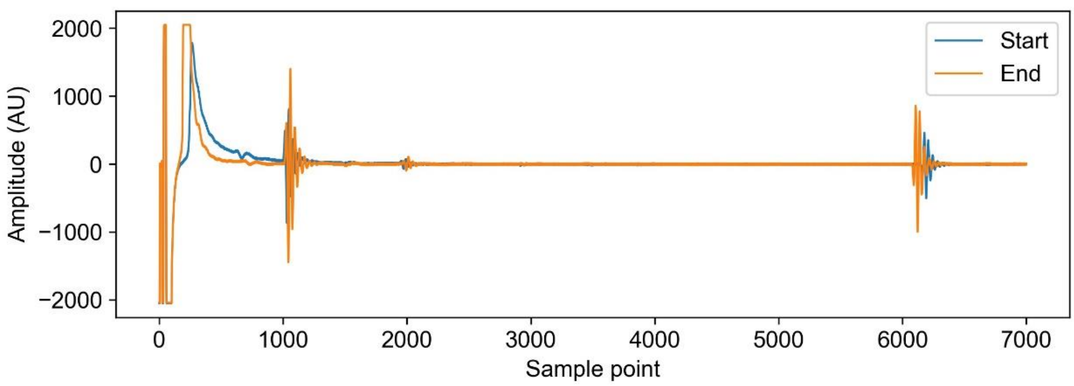
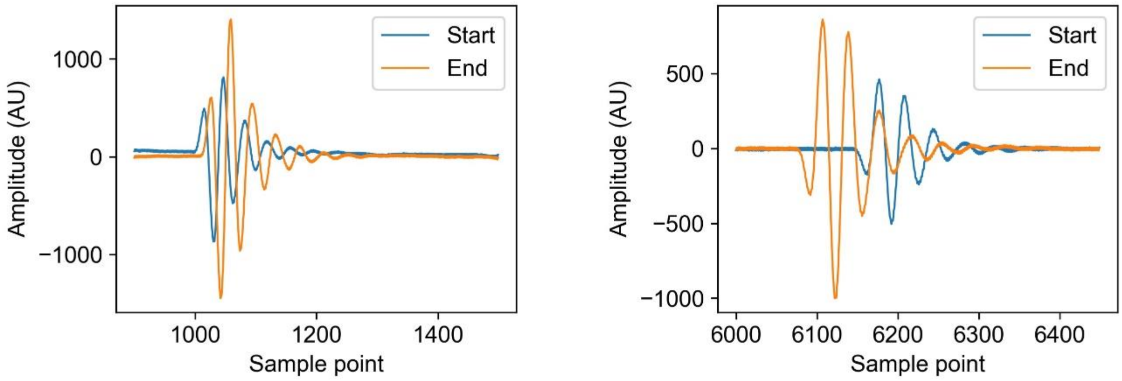


Fig. 1. The experimental apparatus and path of the received US sound wave reflections. Adapted from Bowler et al. (2021).



(a)

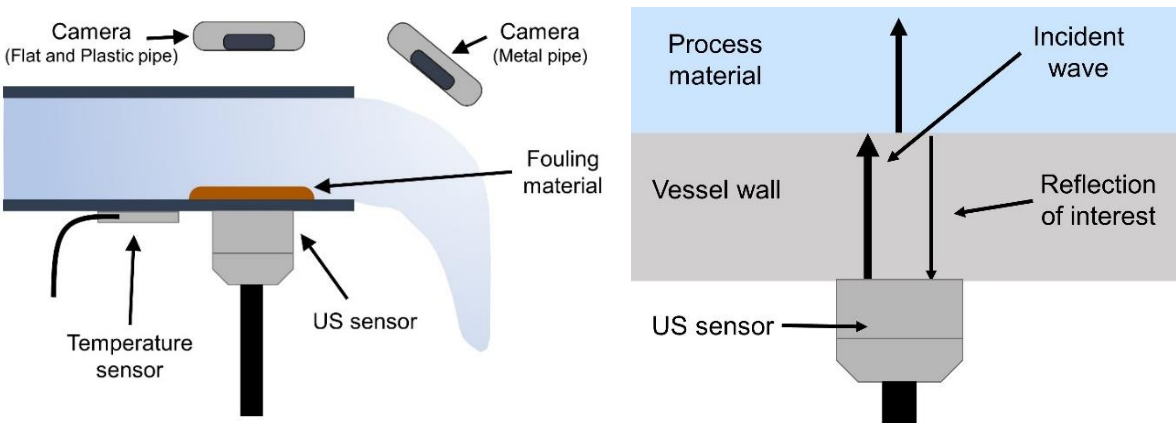


(b) (c)

Fig. 2.(a) Example US waveforms obtained for the start and end of a fermentation batch. (b) The first reflection, located between sample points 900 and 1400. (c) The second reflection, located between sample points 6000 and 6500.

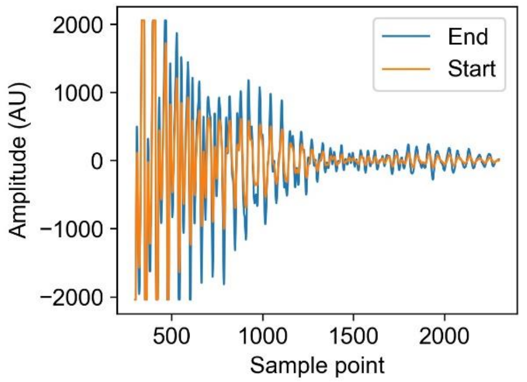
2.1.2 Cleaning of pipe fouling

Full experimental details are provided in Escrig et al. (2019, 2020a, and 2020b). Three pipe test sections were used: A rectangular rig with a SS340 base plate and clear, PMMA sides; a circular pipe section constructed from clear PMMA; and an opaque, circular pipe section constructed from SS316. Three different food materials were used to foul the pipe test sections: tomato paste, concentrated malt, and gravy. The fouling material was spread onto the pipes and allowed to dry. It was placed in the centre of the base plate for the rectangular rig and 30 mm from the exit for the circular pipes. The temperature of the water used for cleaning was set at either 12 °C or 45 °C and a flowrate of 6 l/s was used. For the rectangular test section, a magnetic sensor (5 MHz resonance, M1057, Olympus) was externally attached to the base plate. For the circular pipe sections, the US transducers (2 MHz, Yushi, 2P10N) were glued externally to the bottom of the pipes in the location where the fouling material would be placed. The temperature sensors were attached at the same locations. A camera was used to determine the time at which all the fouling material was removed. The position of the camera was moved depending on whether the pipe section was clear or opaque. The US and temperature data was recorded every 4 seconds producing 4 waveforms which were averaged. A reflection-mode, pulse-echo sensing technique was used to monitor the waveform reflected from the interface between the pipe wall and the fouling material. The camera images were recorded every 20 seconds. A minimum of 7 repeats were conducted for every permutation of pipe test section, fouling material and fluid temperature, producing 93 runs in total.

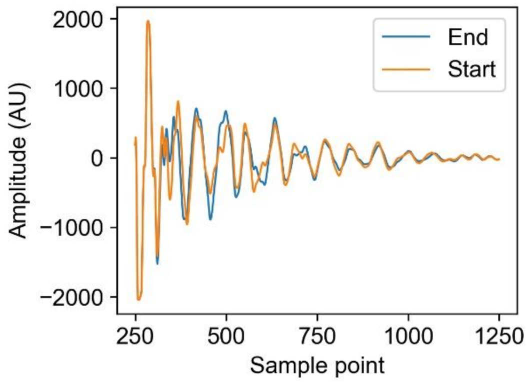


(a) (b)

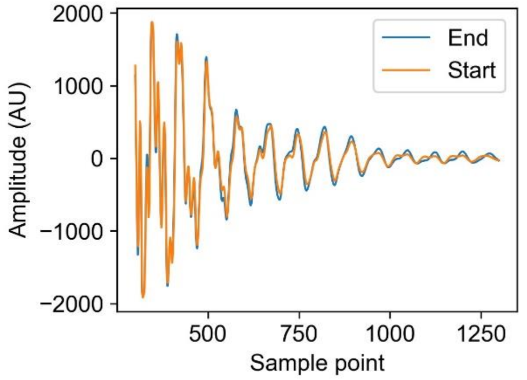
Fig. 3. (a) The experimental apparatus including the positions of the pipe section, US sensor, temperature sensor, and fouling material. (b) The paths of the received US reflections.



(a)



(b)



(c)

Fig. 4. The received US waveform at the start and end of the cleaning process for the (a) Flat, (b) Plastic, and (c) Metal pipe sections.

2.1.3 Honey-water mixing

Full experimental details are provided in Bowler et al. (2020). Two US sensors (5 MHz resonance, M1057, Olympus) were externally attached to the base of a 250 ml glass mixing vessel. An overhead stirrer was used to stir the mixture. One sensor (the central sensor) was attached in the centre of the vessel base. Another sensor (the non-central sensor) was attached approximately 2 cm offset from the centre. The temperature sensor was also attached to the base of the vessel. A reflection-mode, pulse-echo sensing technique was used to monitor the sound wave reflected from the interface between the vessel wall and the mixture. US signals were acquired continuously for 1 s for each probe consecutively. On average, this acquired two US waveforms which were then averaged to minimise noise disturbance. Two different volumes of pure clear honey (Wm Morrison Supermarkets plc) were used: 20 and 30 ml. 200 ml of tap water was used for all runs. The impeller speed was set to either 200 or 250 rpm. These four parameter permutations were repeated three times whilst varying the environmental temperature, producing a set of 12 runs. This methodology was repeated across two days, producing two datasets. Between, the US sensors were removed and reattached. The ground truth was obtained using a video camera to determine the time for complete mixing.

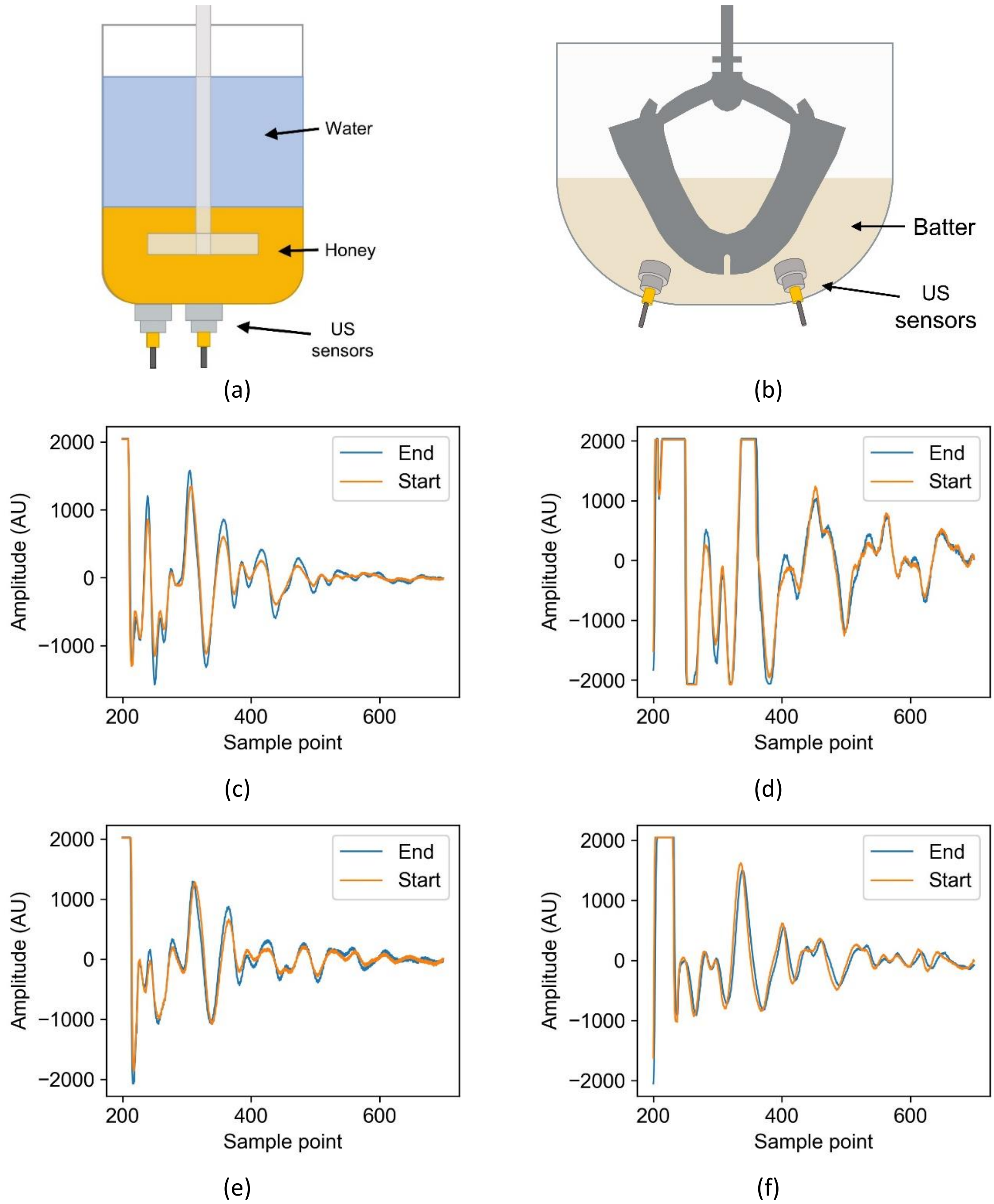


Fig. 5. The experimental apparatus for (a) the honey-water mixing experiments and (b) the flour-water batter mixing experiments. The received US waveforms reflections for (c) honey-water mixing probe 1, (d) batter mixing probe 1, (e) honey-water mixing probe 2, and (f) batter mixing probe 2.

2.1.4 Batter mixing dataset

Full experimental details are provided in Bowler et al. (2020). Two US sensors (5 MHz resonance, M1057, Olympus) were externally attached to a stand mixer glass mixing bowl (1000 W Kenwood kmix kmx754). The temperature sensor was also attached to the outside of the mixing bowl. A reflection-mode, pulse-echo sensing technique monitored the sound wave reflected from the interface between the mixing bowl and the mixture. US signals were continuously acquired for 1 s for each probe consecutively. On average, this produced 2 waveforms which were averaged to minimise disturbance from signal noise. The quantity of strong white flour (Wm Morrison Supermarkets plc) and tap water used was varied. A total of 9 runs were monitored. The optimal mixing time was obtained by determining the time of maximum power input to the impeller. This was measured using a YouThink plug socket power meter.

2.2 Feature extraction

Two feature extraction methodologies were compared: extracting coarse, time-domain signal features (Coarse method) and convolutional feature extraction using a CNN pre-trained on an auxiliary task (Convolutional method). The Coarse features method obtains coarse information about the changing waveform oscillations compared with the Convolutional method which can identify changing amplitudes at individual sample points in the waveform. The Coarse features method is designated as the next best approach (as justified in Section 1) and was the method used in Bowler et al. (2021). Therefore, a comparison between these two methods will evaluate the advantage of using convolutional feature extraction.

2.2.1 Coarse feature extraction

In total, 10 signal features were extracted from the waveform. The sum absolute amplitude (SAA), energy, sum root amplitude (SRA), standard deviation, skewness and kurtosis (equations 1 – 7) provide measurements of the distribution of amplitudes within the waveform. In addition, the amplitude and position of the maximum and minimum peaks were used as features to monitor the largest peaks in the waveform.

$$SAA = \sum_{i=1}^{i=SP} |A_i| \tag{1}$$

Where *SAA* is the sum absolute amplitude, *SP* is the number of sample points in the waveform, *A* is the waveform amplitude at sample point *i* (Zhan et al., 2015).

$$E = \sum_{i=1}^{i=SP} {A_i}^2 \tag{2}$$

Where *E* is the waveform energy (Zhan et al., 2015).

$$SRA = \sum_{i=1}^{i=SP} \sqrt{|A_i|} \tag{3}$$

Where *SRA* is the sum root amplitude (Zhan et al., 2015).

$$\mu = \frac{\sum_{i=1}^{i=SP} A_i}{SP} \tag{4}$$

$$SD = \sqrt{\frac{1}{SP} \sum_{i=1}^{i=SP} (A_i - \mu)^2} \tag{5}$$

Where μ is the mean amplitude of the waveform, and SD is the standard deviation (Zhan et al., 2015).

$$S = \frac{\sum_{i=1}^{i=SP}(A_i - \mu)^3}{SP \times STD^3} \tag{6}$$

Where S is the waveform skewness (Caesarendra and Tjahjowidodo, 2017).

$$K = \frac{\sum_{i=1}^{i=SP}(A_i - \mu)^4}{SP \times STD^4} \tag{7}$$

Where K is the waveform kurtosis (Caesarendra and Tjahjowidodo, 2017).

2.2.2 Convolutional feature extraction

Previous work has determined LSTM layers are required for accurate time-series process monitoring. Training a convolutional neural network to the target data without an LSTM layer to obtain pre-trained convolutional filter weights would be a sub-optimal task due to the LSTM layer's ability to learn the important process feature trajectories. Therefore, the input waveforms would not be able to fit to the target data optimally without an LSTM layer and informative waveform features would not be learned (Bowler et al., 2020 and 2021). Training convolutional and LSTM layers simultaneously would also be a difficult task especially with long time sequences and limited training data used in the present case studies. This is because the many weights present in the convolutional filters and LSTM units would compete during the training process and likely fail to fit to the target data or make the training unstable. As such, to easily train convolutional layers that extract informative ultrasonic waveform features, this work trained a 1D CNN on an auxiliary task to predict waveform dataset membership. Table 1 summarises the 11 waveform datasets used. Segments of 1000 samples points in length were taken from each waveform. The position of the 1000 sample point length window was chosen for each waveform by investigating the difference between the start and end of the corresponding process. The areas with the largest visual change throughout the process were used. To increase the training set size for the network, and to improve meaningful feature extraction in the convolutional layers, a 600 x 1 input to the CNN was used. Data augmentation using a sliding window, laterally translated by 100 sample points each time, produced five waveform segments of 600 sample points in length. Further data augmentation through separate normalisation of each waveform segment was used to differentially magnify the waveform. This ensures that the network learns features specific to each waveform, rather than the position or magnitude of features.

Table 1. A summary of the datasets used to train the convolutional feature extractor on the auxiliary task and also evaluate the performance of the proposed feature extraction methodology.

Experimental dataset	**ML task**	**Waveforms for CNN auxiliary task**	**Total number of runs (train/ validation/ test split)**	**Maximum sequence length**
Fermentation	• Regression to predict alcohol concentration	Reflection 1 Reflection 2	13 (9/2/2)	3112
Cleaning of food fouling from pipe sections	• Classify the end of cleaning • Regression to predict cleaning time remaining	Flat rig Circular, plastic Circular, metal	35 (25/5/5) 30 (20/5/5) 28 (20/4/4)	400 300 200

Honey-water mixing 1	• Classify the end of mixing • Regression to predict mixing time remaining	Central sensor Non-central sensor	12 (8/2/2)	165
Honey-water mixing 2	• Classify the end of mixing • Regression to predict mixing time remaining	Central sensor Non-central sensor	12 (8/2/2)	123
Batter mixing	• Classify the end of mixing • Regression to predict mixing time remaining	Sensor 1 Sensor 2	9 (5/2/2)	153

A summary of the 1D CNN trained is presented in Table 2 which also presents CNN structures used in other previous works as a comparison. It should be noted that optimal CNN architectures are task-specific and should be chosen through validation procedures. CNN architectures for US sensor signals are included as a literature review for the interested reader. A grid search was used to select the learning rate, batch size and number of neurons in the fully connected layer. No padding was used. Training was performed with the Adam optimiser. The minimum number of neurons in the fully connected layer to achieve 100% accuracy for the dataset membership prediction was used to ensure feature identification in the convolutional layers rather than the fully connected layer. The designated training and validation sets for all datasets were used. A training accuracy of 100% was achievable after only 3 epochs, highlighting the rapidity in developing our proposed convolutional feature extraction methodology. The pre-trained convolutional weights were then used to extract features on the full-size waveform for each dataset.

Table 2. A summary of the feature extraction layers of the proposed convolutional neural network and a comparison with the other 1D CNN structure present in the literature for US sensor data.

Layer	**Proposed network**	**Virupakshappa et al., 2018**	**Meng et al., 2017**	**Munir et al., 2019**	**Munir et al., 2020**
1	1D Convolutional layer 7 x 1 filter size 16 filters	1D Convolutional layer 5 x 1 filter size 5 filters	2D Convolutional layer 7 x 5 filter size 16 filters	1D Convolutional layer 16 x 1 filter size 32 filters 8 x 1 stride	1D Convolutional layer 25 x 1 filter size 32 filters 8 x 1 stride
2	Max Pooling layer 2 x 1 pool size	Max Pooling layer 2 x 1 pool size	Max Pooling layer 2 x 2 pool size	1D Convolutional layer 3 x 1 filter size 64 filters 2 x 1 stride	1D Convolutional layer 3 x 1 filter size 64 filters 2 x 1 stride
3	1D Convolutional layer 5 x 1 filter size 32 filters	1D Convolutional layer 8 x 1 filter size 8 filters	2D Convolutional layer 5 x 3 filter size 32 filters	Max Pooling layer 2 x 1 pool size 2 x 1 stride	Max Pooling layer 2 x 1 pool size 2 x 1 stride
4	Max Pooling layer 2 x 1 pool size	Max Pooling layer 2 x 1 pool size	Max Pooling layer 2 x 2 pool size	-	-

5	-	1D Convolutional layer 7 x 1 filter size 16 filters	-	-	-
6	-	Max Pooling layer 2 x 1 pool size	-	-	-

To reduce the dimensionality of the data, minimise non-useful information input into the network, aid LSTM unit training accuracy and stability, and enable the use of additional features such as the US time of flight and standard deviation between consecutive waveforms, PCA was applied to the waveform features extracted using the pre-trained convolutional filter weights. PCA extracts a set of orthogonal principal components (PCs) which are a combination of the co-linear original features (Abdi and Williams, 2010). Alternatively, a CNN feature extractor structure with more downsampling or additional layers to reduce the number features extracted could have been used. However, preliminary investigations showed this method produced features too specific to the auxiliary training task. Furthermore, an autoencoder could have been used to learn non-linear feature relationships compared to the linear relationships assumed using PCA. However, as outlined in Section 1, the convolutional feature extraction methodology only needs to overcome a possible translation in waveform peaks by measuring spatial relationships between sample point amplitudes. Therefore, compared with autoencoders, owing to the sufficient feature extraction capability, elimination of hyperparameter optimisation, model training and convenient selection of the number of features extracted, PCA was identified as the optimal methodology. Table 3 includes the percentage variability explained by each PC for the US waveform datasets and the number of PCs required to explain 95 % of the variability. The first PC likely follows the common waveform changes across the full dataset caused by variations in the US properties of the materials being monitored (either due to changing composition or process temperature). Successive PCs will identify waveform changes more specific to each batch, most likely due to the different process temperatures. Therefore, it is anticipated that only a small number of PCs are required (i.e. greater than one) to monitor the changing material composition and account for changes in the monitoring US waveform at different temperatures. This is supported by Table 3 where the percentage variability explained drops off after the first two PCs. As shown in Table 3, the smallest number of PCs required to explain 95% of the variability in the dataset, a common method for selecting the number of PCs to use, is eight for the Plastic Cleaning dataset and nine for fermentation monitoring using only the first reflection. Therefore, using these two pieces of guidance (the primacy of the first and second PCs and the smallest number of PCs to explain 95% of dataset variability), five PCs were selected to obtain useful waveform information while minimising noise. The PCs were also combined with the standard deviation of the energy between consecutive waveforms in an acquisition block (where the number of waveforms was greater than two) to provide a measure of material differences between consecutive waveform acquisitions (Equation 8). In the case of the fermentation dataset using both the first and second waveform reflections, the sound wave time of flight was also added. The time of flight was calculated using a thresholding method, identifying the sample point where the waveform rises above the signal noise.

$$\text{ESD} = \sqrt{\frac{1}{W}\sum_{i=1}^{i=W}(E_i - \bar{E})^2} \qquad (8)$$

Where *ESD* is the standard deviation in the energy of the waveforms in the acquired block, and *W* is the number of waveforms in the acquired block.

Table 3. A summary of the distribution of the explained variance by each PC for the US waveform datasets after convolutional feature extraction.

Experimental dataset	Waveforms	Number of PCs to explain 95% of variability	Variability explained by 1st PC (%)	Variability explained by 2nd PC (%)	Variability explained by 3rd PC (%)	Variability explained by 4th PC (%)	Variability explained by 5th PC (%)
Fermentation	Reflection 1	9	56.4	23.1	9.2	2.1	1.5
	Reflection 2	18	30.4	21.6	14.9	9.0	6.1
Cleaning of food fouling from pipe sections	Flat rig	15	60.4	15.2	7.4	4.3	1.8
	Circular, plastic	8	56.7	14.3	12.4	6.3	1.9
	Circular, metal	32	50.9	12.3	8.5	4.6	3.7
Honey-water mixing 1	Central sensor	24	52.1	18.8	7.5	4.6	2.3
	Non-central sensor	41	51.4	17.0	4.7	3.8	2.8
Honey-water mixing 2	Central sensor	19	38.6	30.8	12.4	4.1	2.9
	Non-central sensor	25	41.6	36.8	4.6	2.7	2.4
Batter mixing	Sensor 1	42	49.1	15.1	14.3	4.5	2.6
	Sensor 2	16	60.5	16.3	7.5	3.1	1.5

2.3 Model training and testing

Neural networks consisting of an LSTM layer followed by a fully-connected layer were used for all ML tasks. A fully-connected layer allows for the creation of modified features which better match the prediction task output while the LSTM layer learns the trajectories of the input features. The input features were normalised and zero-padding at the start extended the sequence lengths to that of the maximum. A masking layer specified the LSTM to disregard the zero-padding. The Adam optimisation algorithm and a gradient norm clipping value of 1 was used. A single-fold validation procedure determined the learning rate, number of LSTM units, dropout probability, L2 regularisation penalty, number of neurons in the fully-connected layer, and batch size. As many tasks and hyperparameters were investigated, only a single validation set was used to reduce the training time required. The optimal set of hyperparameters were used to retrain a model using all of the training data. The LSTMs were trained using TensorFlow 2.3.0. The coefficient of determination (R^2), mean squared error (MSE), and mean absolute error (MAE) were used as performance metrics to evaluate the regression ML models. The accuracy, precision, and recall were used to evaluate the classification models. Evaluation of multiple performance metrics allow for improved comparison between models. Multi-task learning was also investigated to aid LSTM learning of the process trajectory (Fig. 6). By training on two correlated tasks (in this case, both the classification and regression tasks for the mixing and cleaning datasets), the shared LSTM layer may learn more effective feature trajectories while reducing redundant information being stored (Li et al., 2016).

This may have two benefits. The first being increased model accuracy through global learning of feature trajectories important to the process being monitored. The second being more stable model training by optimising for two combined losses. To reduce the model validation time, the number of neurons in the fully-connected layers and the dropout rate for the task-specific branches of the neural networks were fixed as the optimal hyperparameters determined from the single-task learning networks. Alternatively, a shared fully connected layer could have been used for the multi-task networks. However, to provide easier evaluation of multi-task learning utility compared with the single task learning networks, only the LSTM layer was shared. This allows for task specific feature combinations to be learned in the fully connected layers. A single-fold validation procedure optimised the number of LSTM units, dropout rate, L2 regularisation parameter, learning rate, batch size, and weighting of the individual classification and regression losses. A coarse grid search optimised the loss weighting by monitoring the unweighted classification and regression losses individually, followed by a fine grid search which optimised by monitoring the combined loss.

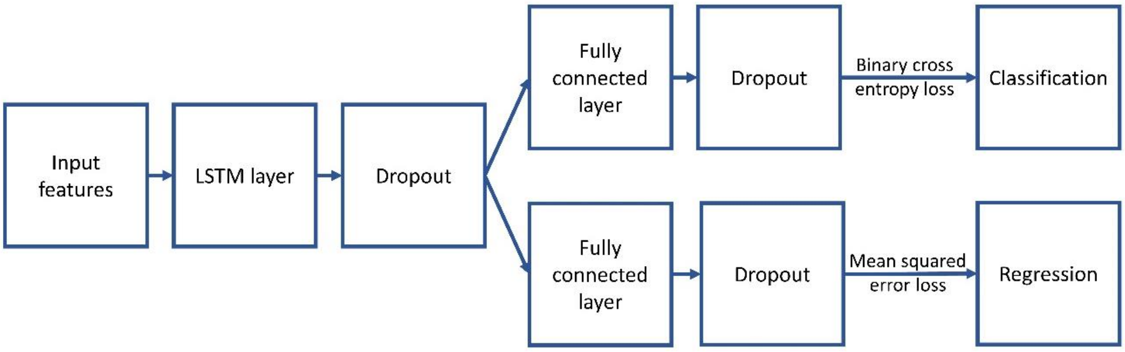
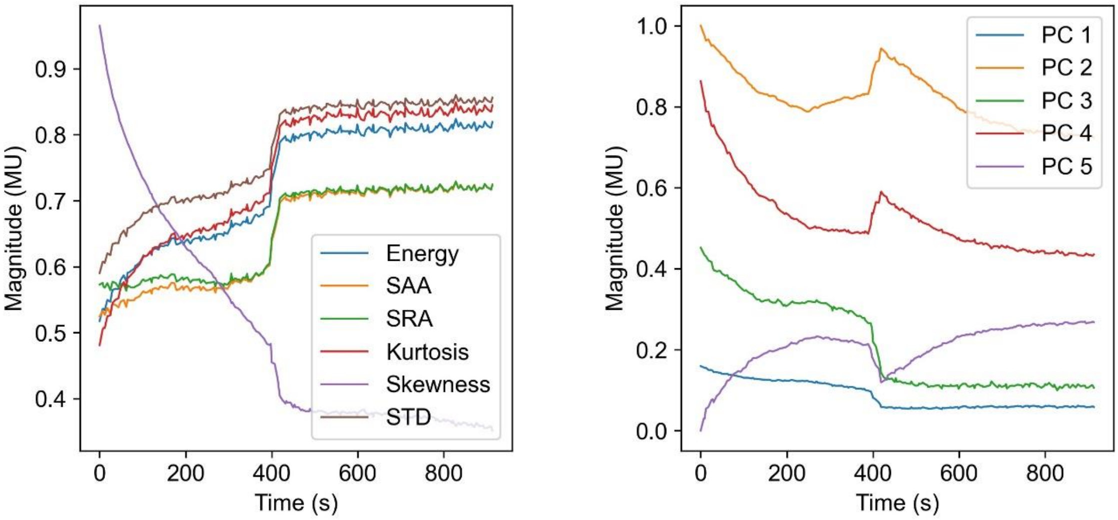


Fig. 6. The structure of the multi-task learning network evaluated using the cleaning and mixing datasets. The cleaning and mixing datasets were used as both entail classification and regression tasks.

3 Results and discussion

To highlight the differences between the features extracted by the two methodologies, Fig. 7 displays the Coarse features (Fig. 7a) and the Convolutional PCs (Fig. 7b) for the first batch of the Flat Cleaning experiments. In Fig. 7a it is shown that the Energy, Sum Absolute Amplitude (SAA), Sum Root Ampltiude (SRA), Kurtosis, and standard deviation (STD) all follow similar trends. In contrast, the convolutionally extracted PCs follow different trajectories, highlighting the additional waveform information presented to the ML models through use of the Convolutional approach.



(a) (b)

Fig. 7. A comparison between (a) the Coarse features and (b) the Convolutional extracted features for the first batch of the Flat Cleaning experiments. The end of cleaning was identified using the camera at 425 s. Note the similar process trajectories of the Energy, Sum Squared Amplitude (SAA), Sum Root Amplitude (SRA), Kurtosis, and Standard Deviation (STD). In contrast, the five convolutionally extracted principal components show differing trajectories, making additional US waveform information more accessible to the ML models.

Overall, the Convolutional method was more accurate for over half of the tasks evaluated. For the fermentation datasets (Fig. 8), the Convolutional approach achieved lower accuracies than the Coarse feature method. In contrast, the Convolutional method proved more accurate for all cleaning tasks (Fig. 9.) and flour-water batter mixing (Fig. 10). However, the results were mixed for the honey-water mixing datasets (Fig. 11) with a Convolutional based approach scoring the highest for three tasks compared with five using the Coarse features method. Table 4 compares the results from this work with previous published works using these datasets. It should be noted that training, validation, and test sets, along with validation and testing procedure, differ between the previous published results and the current work. As such, the accuracies are not directly comparable. In practice, optimising for the number of PCs, employing k-fold cross validation, and possibly using past process information (in the form of feature gradients, time-lagged feature representations, or the time since the beginning of the process) would improve model accuracy on the test set data. However, this is not necessary in the current work in which the aim is to present the superior feature extraction ability of the Convolutional method compared with the Coarse features. Interestingly, for the datasets where the Convolutional method was more accurate than the Coarse method, cleaning and flour-water batter mixing, high accuracy was achieved in previous works using complex feature extraction methodologies. For example, Bowler et al (2020) achieved 92.5 % accuracy in classifying the end of flour-water batter mixing through using a CNN training on the continuous wavelet decomposition of the waveform. Escrig et al. (2020a, 2020b) used a K-best predictors method to selected the 200 most informative sample points in the waveform to predict the end point of pipe section cleaning. Furthermore, no LSTM layers or past process information (e.g. features gradients or time-lagged features) were required for these tasks. Contrastingly, for the datasets where the Coarse method was more accurate than the Convolutional method, honey-water mixing and fermentation, previous work suggests that using past process information as features was vital for high model accuracy but not complex feature extraction methodologies.

The increased accuracy of the Convolutional feature method for tasks that require a lot of waveform information in the previous works, namely; cleaning and flour-water batter mixing, shows that this method is capable of extracting more usable information from the waveform. As such, this proves that the Convolutional method is a superior feature extractor to using Coarse features. Resultantly, the lower accuracy of the fermentation and honey-water mixing results indicates that the Convolutional feature method degraded the feature trajectory learning of the LSTM layer. There are several reasons why this may be the case. Firstly, the more complicated trajectories of the PCs could have been more difficult for the LSTM layer to learn. To overcome this, the results from previous works suggest the use of feature gradients aids in LSTM layer learning of feature trajectories. Secondly, due to the increased waveform information extracted, the Convolutional method may have overfitted to the range of the training data with the testing data falling outside of the training feature ranges. This can be overcome through using a k-fold cross-validation procedure instead of the single-fold validation used in this study. Single-fold validation was used to reduce model development time, owing to the large number of tasks and hyperparameters evaluated. K-fold cross-validation was not required in this study, where the aim was to showcase the superior feature extraction capability of the Convolutional method, as has been presented. Thirdly, the Coarse feature method may have benefitted from the similarity in the input features. These features will most strongly follow the changes in US properties of the monitored materials, similar to the first PC extracted using the Convolutional method. Therefore, the Coarse feature method allows the LSTM layer more opportunities to learn this strong feature trend. In contrast, the Convolutional method can only learn the trajectory of the first principal component through a single path in the network. It is anticipated that, again, k-fold cross-validation would strengthen the impact of the first PC relative to the subsequent, less informative PCs.

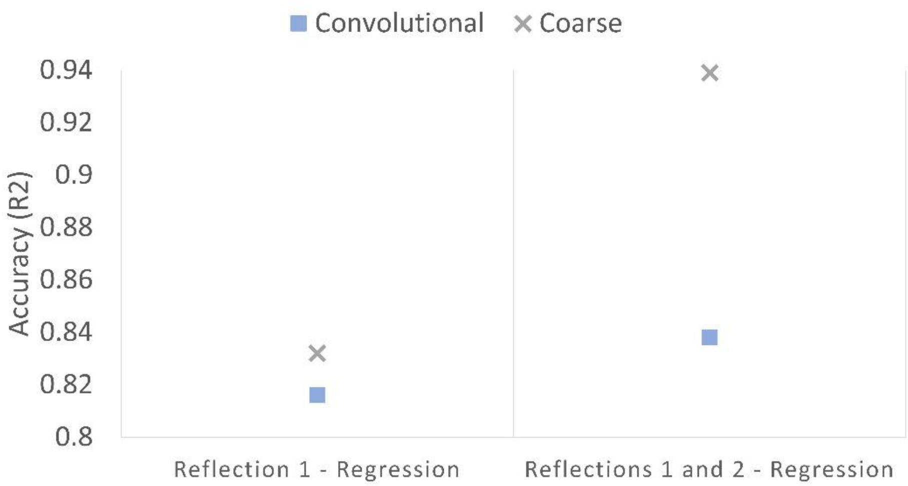


Fig. 8. The R^2 scores for the feature extraction methodologies applied to the fermentation dataset.

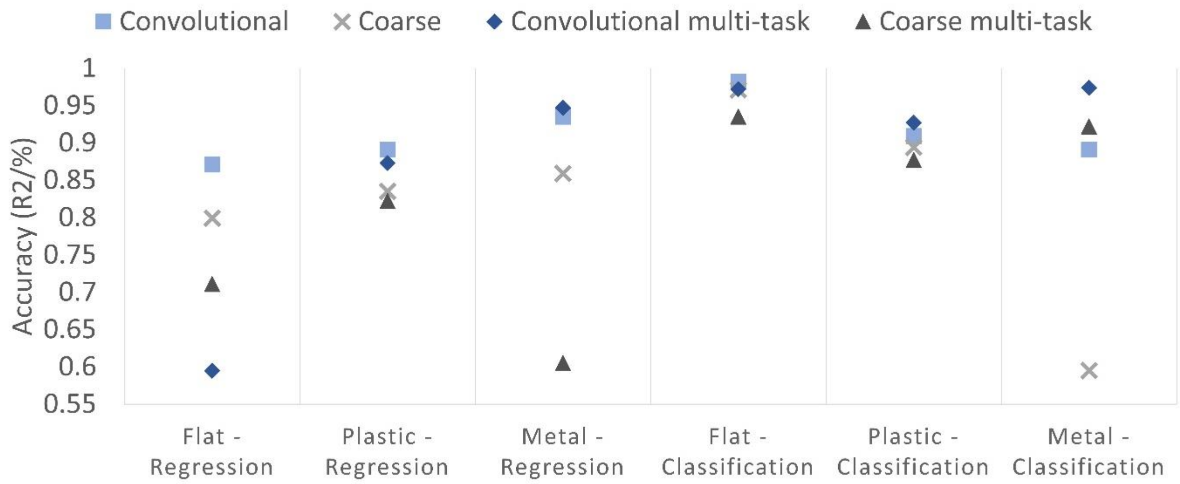


Fig. 9. The regression (R^2) and classification (% correct) accuracy for the feature extraction methodologies evaluated on the cleaning tasks. A CNN method was most accurate for every task.

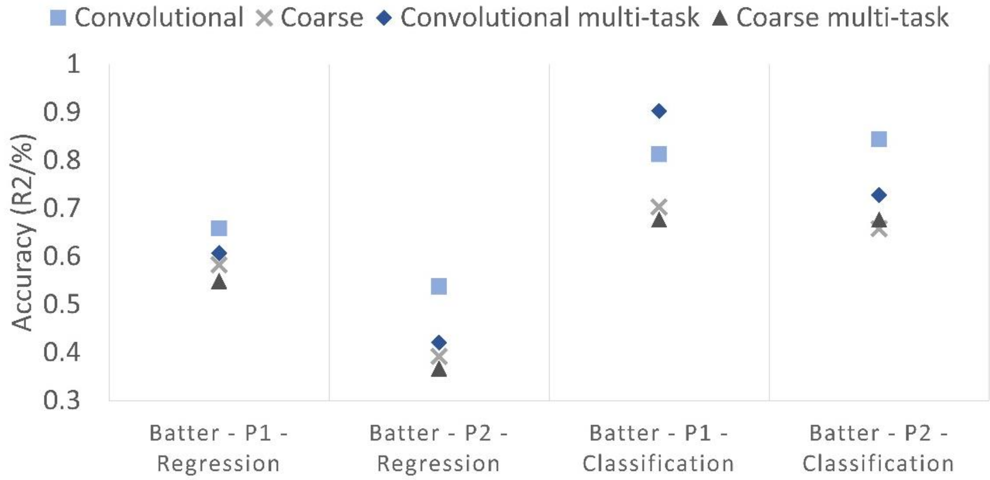


Fig. 10. The regression (R^2) and classification (% correct) accuracies for the feature extraction methodologies evaluated on the flour-water batter mixing dataset. A Convolutional method was most accurate for every task.

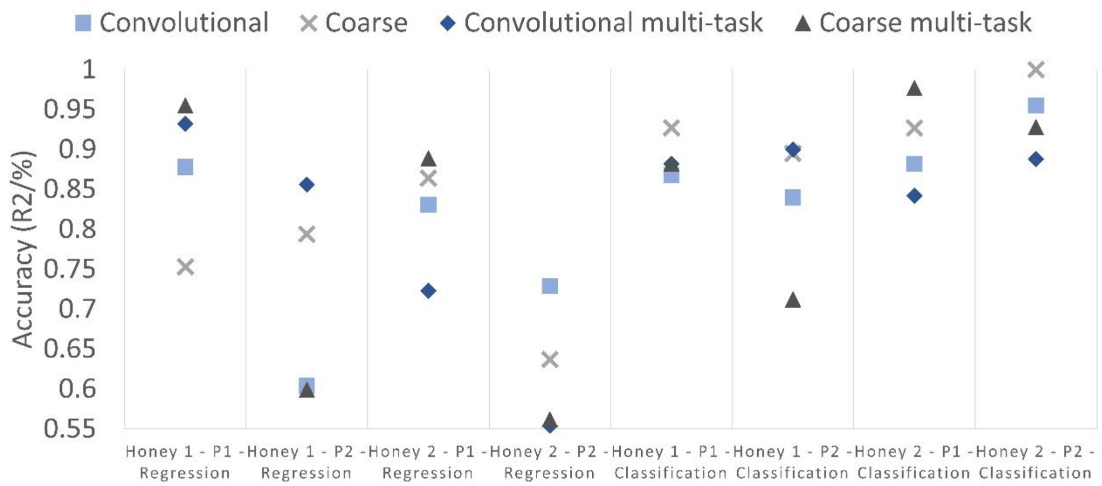


Fig. 11. The regression (R^2) and classification (% correct) accuracies for the feature extraction methodologies evaluated on the honey-water mixing datasets. P1 indicates the non-central sensors and P2 denotes the central sensors.

Table 4. A comparison of the presented convolutional feature extraction method to previously published ML results obtained using the same datasets.

Previous works	**Task**	**ML accuracy (R^2 / %)**	**Presented convolutional feature extraction method accuracy (R^2 / %)**	**Differences in previous works methodology**	**Conclusions**
Bowler et al. (2021)	Fermentation monitoring Regression	0.952 – 1^{st} and 2^{nd} reflections 0.948 – 1^{st} reflection	0.816 – 1^{st} and 2^{nd} reflections 0.838 – 1^{st} reflection	Feature gradients used as features	The results indicate that the addition of time-lagged feature representations improves LSTM model training
Bowler et al. (2020)	Honey-water mixing Classification	96.3 % central sensor 89.8 % non-central sensor	95.5 % central sensor 88.2 % non-central sensor		
	Honey-water mixing Regression	0.960 central sensor 0.965 non-central sensor	0.856 central sensor 0.932 non-central sensor		
	Flour-water batter mixing Regression	0.976	0.659		
	Flour-water batter mixing Classification	92.5 %	90.3 %	Wavelet analysis used	The results indicate a greater number of PCs may improve model accuracy
Escrig et al. (2020a)	Cleaning Flat pipe Classification	Up to 99 %	98.2 %	200 waveform sample points used as features, selected though K-best predictors	
Escrig et al. (2020b)	Cleaning Plastic and Metal pipes Classification	Up to 100 %	92.7 % - Plastic pipe 97.4 % - Metal pipe		
Simeone et al., 2020	Cleaning Flat pipe Regression	0.955	0.871	US sensor data combined with optical sensor data	The image analysis allowed for early monitoring of the cleaning process

The results for the multi-task learning neural networks were mixed. Overall, multi-task learning performed worse for 23 out of 36 tasks compared with the single task learning counterparts. The reason for this likely that the networks failed to optimise for both tasks but instead generalised across them. The hyperparameters for the task-specific branches of the multi-task neural networks were fixed as the optimal values from the respective optimised single-task networks. Optimising for these hyperparameters as well may improve multi-task learning accuracy though requires longer development time. As such, the decision to use multi-task learning should be made during the task validation stage. However, multi-task learning showed more benefits to the classification tasks compared with regression, providing improved accuracy for 8 out of 18 tasks. This is likely because the regression part of the network aids in identifying the approximate position of the classification decision boundary for the classification branch to optimise. This indicates that the regression results for the multi-task learning networks may be improved around the classification decision boundary but failed to learn feature trajectories far from this point. Multi-task learning showed more benefits in the regression tasks for the honey-water mixing experiments, achieving higher accuracy for half of the tasks. As the results from previous works show that learning feature trajectories is vital for these tasks, this indicates that multi-task learning may allow improved feature trend learning in the LSTM layer. This is further supported by multi-task learning proving more benefits for the Convolution method, achieving higher accuracies for 7 out of 18 tasks compared with 5 for the Coarse method. As feature trajectory learning is more difficult using the Convolutional method without feature gradients, this indicates that multi-task learning could alleviate this problem.

3.1 Advantages of the convolutional feature extraction method

The Convolutional feature extraction method presented in this work and evaluated on time-series data also has benefits for non-time series data. Firstly, it obtains informative convolutional filter weights from an easier task to be used for a more difficult desired task as either a feature extraction method or as starting points for weight fine-tuning. Data augmentation and the minimisation of the number of neurons in the fully connected layer of the auxiliary task CNN ensures useful convolutional layer feature learning. Secondly, by using the pre-trained filter weights as feature extractors rather than a starting point for fine-tuning time, model development time is saved. Thirdly, the use of PCA allows the incorporation of other features useful to process monitoring, such as the process temperature, speed of sound, standard deviation between consecutively acquired signals, feature gradients or time lagged feature representations, and other process parameters. Furthermore, the use of PCA reduces the dimensionality of the data to improve model training and amplifies the contribution the previously listed additional features.

4 Conclusion

The performance of ML models is partly dependent on the quality of features extracted from the data. This work compared two feature extraction methodologies for process monitoring using US sensor data. The Convolution feature extraction method produces more informative waveform features; however, presents a more difficult feature trajectory learning task. Multi-task learning improves process trajectory learning but regression accuracy is degraded far from the classification decision boundary. This may be overcome through more extensive hyperparameter selection though at increased model development time. Once trained, the convolutional method represents a fast and convenient way of extracting high quality US waveform features for future applications.

Author contributions

Alex Bowler: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. Michael Pound: Conceptualization, Supervision, Writing – review & editing. Nicholas Watson: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) standard research studentship (EP/R513283/1) and EPSRC network+ Connected Everything (EP/P001246/1).

Declaration of Competing Interest

The authors declare no conflict of interest.

References

Abbagoni, B.M., Yeung, H., 2016. Non-invasive classification of gas-liquid two-phase horizontal flow regimes using an ultrasonic Doppler sensor and a neural network. Meas. Sci. Technol. 27, 084002. 10.1088/0957-0233/27/8/084002

Abdi, H., Williams, L.J., 2010. Principal component analysis. Wiley Interdiscip. Rev. Comput. Stat. 2, 433–459. 10.1002/wics.101

Awad, T.S., Moharram, H.A., Shaltout, O.E., Asker, D., Youssef, M.M., 2012. Applications of ultrasound in analysis, processing and quality control of food: A review. Food Res. Int. 48, 410–427. 10.1016/j.foodres.2012.05.004.

Bowler, A.L., Bakalis, S., Watson, N.J., 2020. Monitoring mixing processes using ultrasonic sensors and machine learning. Sensors 20, 1813. 10.3390/s20071813

Bowler, A.L., Escrig, J., Pound, M., Watson, N. Predicting Alcohol Concentration during Beer Fermentation Using Ultrasonic Measurements and Machine Learning. Fermentation 7, 34. 10.3390/fermentation7010034

Caesarendra, W., T. Tjahjowidodo., 2017. A Review of Feature Extraction Methods in Vibration-Based Condition Monitoring and Its Application for Degradation Trend Estimation of Low-Speed Slew Bearing. Machines 5, 21. 10.3390/machines5040021

Cau, F., Fanni, A., Montisci, A., Testoni, P., Usai, M., 2005. Artificial neural networks for non-destructive evaluation with ultrasonic waves in not accessible. IEEE Ind. Applic. Soc. 1, 685-692. 10.1109/IAS.2005.1518382

Chen, J., Ran, X., 2019. Deep Learning With Edge Computing: A Review. P. IEEE. 10.1109/JPROC.2019.2921977

De Beer, T., Burggraeve, A., Fonteyne, M., Saerens, L., Remon, J.P., Vervaet, C, 2011. Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes. Int. J. Pharm. 417, 32–47. 10.1016/j.ijpharm.2010.12.012.

Escrig, Escrig, J., Woolley, E., Rangappa, S., Simeone, A., Watson, N.J., 2019. Clean-in-place monitoring of different food fouling materials using ultrasonic measurements. Food Control 104, 358-366. 10.1016/j.foodcont.2019.05.013

Escrig, J.E., Simeone, A., Woolley, E., Rangappa, S., Rady, A., Watson, N.J., 2020a. Ultrasonic measurements and machine learning for monitoring the removal of surface fouling during clean-in-place processes. Food Bioprod. Process 123, 1-13. 10.1016/j.fbp.2020.05.003

Escrig, J., Woolley, E., Simeone, A., Watson, N.J., 2020b. Monitoring the cleaning of food fouling in pipes using ultrasonic measurements and machine learning. Food Control 116, 107309. 10.1016/j.foodcont.2020.107309

Henning, B., Rautenberg, J., 2006. Process monitoring using ultrasonic sensor systems. Ultrasonics 44, 1395–1399. 10.1016/j.ultras.2006.05.048.

Figueiredo, M.M.F., Goncalves, J.L., Nakashima, A.M.V., Fileti, A.M.F., Carvalho, R.D.M., 2016. 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, 29-50. 10.1016/j.expthermflusci.2015.08.010

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Comput. 9, 1735-1780. 10.1162/neco.1997.9.8.1735.

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. 10.1038/nature14539

Li, X., Zhao, L., Wei, L., Yang, M.-H., Wu, F., Zhuang, Y., Ling, H., Wang, J., 2016. DeepSaliency: Multi-Task Deep Neural Network Model for Salient Object Detection. IEEE T. Image. Process. 25, 3919-3930. 10.1109/TIP.2016.2579306

Meng, M. , Chua, Y.J., Wouterson, E., Ong, C.P.K., 2017. Ultrasonic signal classification and imaging system for composite materials via deep convolutional neural networks. Neurocomputing 257, 128-135. 10.1016/j.neucom.2016.11.066

Mohd Khairi, M.T., Ibrahim, S., Md Yunus, M.A., and Faramarzi, M., 2015. Contact and non-contact ultrasonic measurement in the food industry: A review. Meas. Sci. Technol. 27, 012001. 10.1088/0957-0233/27/1/012001.

Munir, N., Kim, H.-J., Song, S.-J., Kang, S.-S., 2018. Investigation of deep neural network with drop out for ultrasonic flaw classification in weldments. J. Mech. Sci. Technol. 32, 3073-3080. 10.1007/s12206-018-0610-1

Munir, N., Kim, H.-J., Park, J., Song, S.-J., Kang, S.-S., 2019. Convolutional neural network for ultrasonic weldment flaw classification in noisy conditions. Ultrasonics 94, 74-81. 10.1016/j.ultras.2018.12.001

Munir, N., Park, J., Kim, H.-J., Song, S.-J., Kang, S.-S., 2020. Performance enhancement of convolutional neural network for ultrasonic flaw classification by adopting autoencoder. NDT&E Int. 111, 102218. 10.1016/j.ndteint.2020.102218

Ojha, K.S., Mason, T.J., O'Donnell, C.P., Kerry, J.P., and Tiwari, B.K., 2017. Ultrasound technology for food fermentation applications. Ultrason. Sonochem. 417, 32-47. 10.1016/j.ultsonch.2016.06.001.

Utomo, M.B., Sakai, T., Uchida, S., Maezawa, A., 2001. Simultaneous measurement of mean bubble diameter and local gas holdup using ultrasonic method with neural network. Chem. Eng. Technol. 24, 493-500. 10.1002/1521-4125(200105)24:5<493::AID-CEAT493>3.0.CO;2-L

Utomo, M.B., Sakai, T., Uchida, S., 2002. Use of neural network-ultrasonic technique for measuring gas and solid hold-ups in a slurry bubble column. Chem. Eng. Technol. 25, 293-299. 10.1002/1521-4125(200203)25:3<293::AID-CEAT293>3.0.CO;2-X

Ren, W., Jin, N., Ouyang, L., Zhai, L., Ren, Y., 2021. 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, 9244102. 10.1109/TIM.2020.3031186

Simeone, A., Woolley, E., Escrig, J., Watson, N.J., 2020. Intelligent industrial cleaning: A multi-sensor approach utilising machine learning-based regression. Sensors 20, 1-22. 10.3390/s20133642

Sjödin, D.R., Parida, V., Leksell, M., Petrovic, A., 2018. 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, 22-31. 10.1080/08956308.2018.1471277

Sun, Z., Jen, C.-K., Yan, J., Chen, M.-Y., 2005. Application of ultrasound and neural networks in the determination of filler dispersion during polymer extrusion processes. Polym. Eng. Sci. 45, 764-772. 10.1002/pen.20328

Supardan, M.D., Maezawa, A., Uchida, S., 2003. Determination of local gas holdup and volumetric mass transfer coefficient in a bubble column by means of an ultrasonic method and neural network. Chem. Eng. Technol. 26, 1080-1083. 10.1002/ceat.200301752

Virupakshappa, K., Marino, M., Oruklu, E., 2018. A Multi-Resolution Convolutional Neural Network Architecture for Ultrasonic Flaw Detection. IEEE Int. Ultra. Sym. 2018, 8579888. 10.1109/ULTSYM.2018.8579888

Wallhäußer, E., Hussein, W.B., Hussein, M.A., Hinrichs, J., Becker, T., 2013. Detection of dairy fouling: Combining ultrasonic measurements and classification methods. Eng. Life Sci. 13, 292-301. 10.1002/elsc.201200081

Wallhäußer, E., Sayed, A., Nöbel, S., Hussein, M.A., Hinrichs, J.b,, Becker, T., 2014. Determination of cleaning end of dairy protein fouling using an online system combining ultrasonic and classification methods. Food Bioprocess Tech. 7, 506-515. 10.1007/s11947-012-1041-0

Wu, D., Liu, S., Zhang, L., Terpenny, J., Gao, R.X., Kurfess, T., Guzzo, J.A., 2017. A fog computing-based framework for process monitoring and prognosis in cyber-manufacturing. J. Manuf. Syst. 43, 25-34. 10.1016/j.jmsy.2017.02.011

Zhan, X., Jiang, S., Yang, Y., Liang, J., Shi, T., Li, X., 2015. Inline Measurement of Particle Concentrations in Multicomponent Suspensions using Ultrasonic Sensor and Least Squares Support Vector Machines. Sensors 15, 24109–24124. 10.3390/s150924109