Food Control 147 (2023) 109622

Contents lists available at ScienceDirect

Food Control

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

# Machine learning and domain adaptation to monitor yoghurt fermentation using ultrasonic measurements

Alexander Bowler [a], Samet Ozturk [a,b], Vincenzo di Bari [c], Zachary J. Glover [d], Nicholas J. Watson [a,*]

[a] Food, Water, Waste Research Group, Faculty of Engineering, University of Nottingham, University Park, Nottingham, UK
[b] Department of Food Engineering, Gümüşhane University, Gümüşhane, Turkey
[c] Division of Food, Nutrition and Dietetics, School of Biosciences, University of Nottingham, Sutton Bonington Campus, Nottingham, UK
[d] Arla Foods Ltd, 4 Savannah Way, Leeds Valley Park, Leeds, LS10 1AB, UK

ARTICLE INFO

*Keywords:*
Yogurt
Fermentation
Ultrasonic sensors
Machine learning
Domain adaptation
Process monitoring

ABSTRACT

In manufacturing environments, real-time monitoring of yoghurt fermentation is required to maintain an optimal production schedule, ensure product quality, and prevent the growth of pathogenic bacteria. Ultrasonic sensors combined with machine learning models offer the potential for non-invasive process monitoring. However, methods are required to ensure the models are robust to changing ultrasonic measurement distributions as a result of changing process conditions. As it is unknown when these changes in distribution will occur, domain adaptation methods are needed that can be applied to newly acquired data in real-time. In this work, yoghurt fermentation processes are monitored using non-invasive ultrasonic sensors. Furthermore, a transmission based method is compared to an industrially-relevant non-transmission method which does not require the sound wave to travel through the fermenting yoghurt. Three machine learning algorithms were investigated including fully-connected neural networks, fully-connected neural networks with long short-term memory layers, and convolutional neural networks with long short-term memory layers. Three real-time domain adaptation strategies were also evaluated, namely; feature alignment, prediction alignment, and feature removal. The most accurate method (mean squared error of 0.008 to predict pH during fermentation) was non-transmission based and used convolutional neural networks with long short-term memory layers, and a combination of all three domain adaption methods.

## 1. Introduction

The Food and Drink industry is undergoing pressures of increasing competition, expanding product diversity, increased safety regulations, and demands from consumers for healthier, higher quality, and more sustainable products. To meet this, advanced sensing techniques combined with other digital technologies such as Machine Learning (ML) are required to supersede the current destructive or time-consuming analytical methods (Hassoun et al., 2023). One such area is yoghurt manufacturing. Yoghurt is a coagulated milk product created through fermentation by the lactic acid bacteria *Streptococcus thermophilus* and *Lactobacillus delbrueckii* (Food and Agriculture Organisation of the United Nation (FAO), 2018). During acidification, the casein micelles aggregate and a gel network is formed (Horne, 1999). In yoghurt manufacturing, the fermentation process is typically monitored through periodic sampling and off-line pH measurements. The purpose of this is to monitor that the process is progressing as expected and that the production schedule can be maintained. Yoghurt is a living product and changes in the raw milk composition can cause variations in the fermentation process and final product quality. A slow fermentation rate may cause accumulation of raw milk arriving on the manufacturing site which may exceed storage capacity. Furthermore, fermentation rates affect product quality, yield, texture, and flavour profile and so must be made as consistent as possible. Lastly, slow fermentation rates may increase the risk of pathogenic bacterial growth (Walstra et al., 2005).

For optimal monitoring of the fermentation rate, a real-time sensing

*Abbreviations:* CNN, Convolutional neural network; FCNN, Fully-connected neural network; LSTM, Long short-term memory; ML, Machine learning; MSE, Mean squared error; TOF, Time of flight; US, Ultrasonic.

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

https://doi.org/10.1016/j.foodcont.2023.109622
Received 28 October 2022; Received in revised form 31 December 2022; Accepted 9 January 2023
Available online 10 January 2023

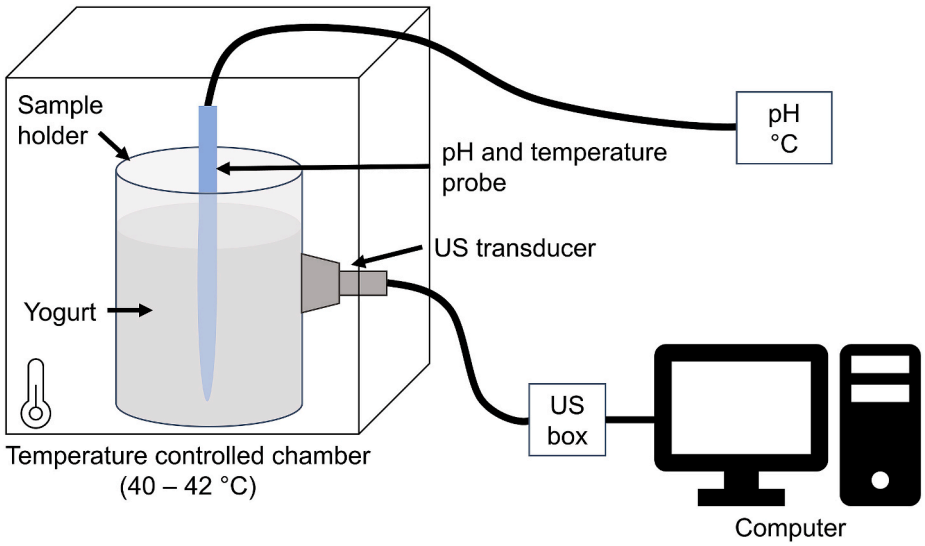


**Fig. 1.** A diagram of the yoghurt fermentation experimental set-up. The inoculated milk samples were placed into a sample holder connected to a US transducer. The pH and temperature probe was also used to monitor the fermentation. The samples were placed in a temperature controlled chamber set to either 40 or 42 °C. The US transducer was connected to a US Box.

solution is required. This would enable fast corrective action (e.g. adjusting the process temperature) or deciding to abandon the batch, reducing time and energy waste. Several sensor methods have been developed for this purpose but each have limitations. For example, near-infrared spectroscopy (Abildgaard et al., 2015; Muncan et al., 2021) and image analysis (Verdú et al., 2019) methods cannot monitor through optically opaque materials requiring them to be internally installed in the fermentation vessels. Furthermore, an invasive pH probe would require cleaning after each batch to prevent contamination. In contrast, low power (intensities below 1 $Wcm^2$), high frequency (higher than 100 kHz) Ultrasonic (US) sensors can monitor opaque materials and are therefore able to be applied non-invasively. US sensors monitor the interaction of materials with mechanical sound waves and have been widely used in the food industry to characterise materials, ensure safety, and monitor product quality (Mohd Khairi et al., 2015 and, 2018; Awad et al., 2012). They are low cost, small in size, low in power consumption, non-destructive, real-time, in-line, and do not cause changes to the structure of the material through which they pass (Henning & Rautenberg, 2006). Several studies have previously used US measurements to monitor yoghurt fermentation. In previous applications, an increase in US velocity, attenuation, and acoustic impedance have been found to follow the trend of decreasing pH (Aljaafreh & Lucklum, 2015; Alouache et al., 2015; Krasaekoopt et al., 2005; Masuzawa et al., 2003; Meng et al., 2012). However, in factories, the yoghurt fermentation process is conducted in large vessels which causes high attenuation of the sound wave. Therefore, transmission-based methods are not feasible and the speed of sound and attenuation cannot be obtained. Meng et al. (2012) used a reflection-mode sensing method to monitor fermentation which requires no transmission of the sound wave through the yoghurt or reflection from a far wall. As this negates obtaining the speed of sound or attenuation, the acoustic impedance was monitored.

To enable operators to observe the state of the fermentation process, the rate of fermentation, and inform decision-making, a method is required to convert the real-time sensor measurements into useful process information. Supervised ML trains algorithms on data to find correlations between inputs (e.g. sensor measurements), also known as features, and outputs (e.g. useful information such as pH value, classifying the state of the fermentation, or predicting the processing time remaining), also known as labels. The aim of ML is to produce a model that accurately predicts the outputs of previously unseen input data and is being increasingly combined with US measurements for process monitoring applications due to its ability to learn complex correlations that are accurate on unseen data (Bowler et al., 2022). However, if the new input data falls outside the range the model has previously encountered during training, erroneous predictions could be obtained. In manufacturing, different process conditions (e.g. temperatures, materials) may affect sensor measurements and change the input feature distribution. Furthermore, the desire to quickly deploy sensing and ML solutions with minimal disruption to the manufacturing process during the data collection stage may limit the number of batches that can be monitored and thereby reduce the variability encountered during model training. Domain adaptation is a category of ML that alters model training or prediction to ensure accuracy is maintained on data outside of the training set range (Bowler et al., 2022). Unlabelled domain adaptation is a further subcategory where the technique has no access to labelled output data for the new input feature range. Unlabelled domain adaptation has previously been used with US measurements to align feature distributions (Alguri et al., 2021; Gao et al., 2021), extract common features across both domains (Bowler & Watson, 2021; Bull et al., 2021), and to predict which features will provide the most benefit to the new domain (Gardner et al., 2022).

However, these previous works have all used domain adaptation to transfer ML models from a source to a target dataset and none have applied domain adaptation techniques to incoming, real-time sensor measurements. This real-time domain adaptation problem negates techniques applied during model training (as during in-line monitoring, the ML models have already been trained) or the use of computationally expensive methods such as deep learning (Gao et al., 2021) or kernel-based approaches (Bowler & Watson, 2021). Therefore, in this work, three domain adaptation methods are investigated that are applicable to this type of problem. Firstly, feature alignment by aligning the mean, or mean and standard deviation, of the training data and new sensor measurements are performed. Secondly, prediction alignment is utilised through discretising the regression output as a classification task and using multitask learning. The aim of this is to maintain a consistent output prediction distribution despite changing input feature distributions. Lastly, real-time feature removal is performed to discard features of the new data which are outliers of the training dataset. The aim of this method is to use common features across the training dataset and the new sensor measurements. The novelty of this work can be summarised as the following: 1) Combining non-invasive US measurements with ML to predict the pH value during yoghurt fermentation. 2) Comparing the

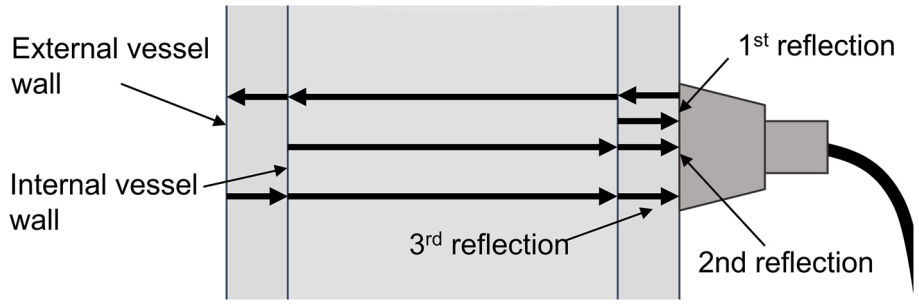


**Fig. 2.** A diagram of the three US waveform reflections monitored during the yoghurt fermentation. The 1st reflection was reflected from the near vessel-yoghurt interface, the 2nd reflection reflected from the far vessel-yoghurt interface after being transmitted through the fermenting yoghurt, and the 3rd reflection was reflected from the far vessel wall after transmission through the yoghurt. The vessel wall thickness and vessel diameter are not to scale.

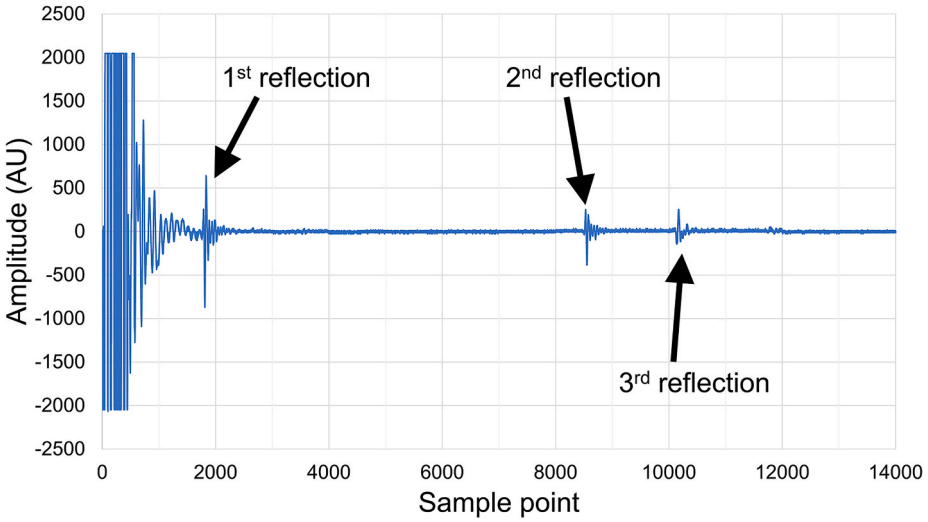


**Fig. 3.** An exemplar US waveform including the three reflected sound waves. The 1st reflection is located at approximately 2000 sample points, the 2nd reflection at approximately 9000 sample points, and the 3rd reflection at approximately 10,000 sample points.

accuracy of transmission and non-transmission US methods. 3) Evaluating the accuracy of ML models using US measurements on outlier fermentation batches to ensure robustness to changing process conditions. 4) Investigating several real-time unlabelled domain adaptation methods to enhance model robustness to the outlier data. These methods were feature alignment, prediction alignment, and feature removal.

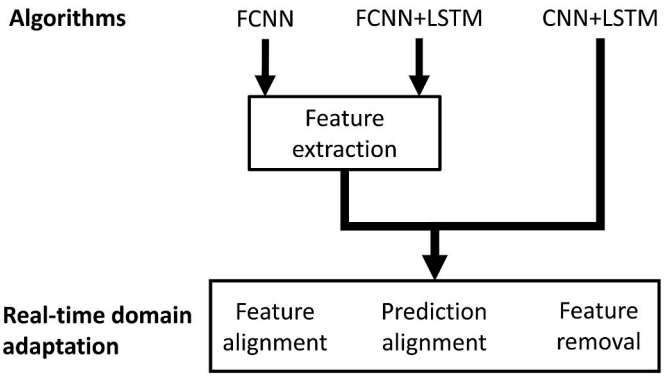


**Fig. 4.** A flow diagram of the ML investigations. Three ML algorithms were investigated: Fully-Connected Neural Networks (FCNN), Long-Short Term Memory (LSTM) neural networks, and Convolutional Neural Networks with LSTM layers (CNN + LSTM). Three real-time domain adaptation methodologies were investigated: 1) feature alignment, 2) prediction alignment, and 3) feature removal.

## 2. Methodology

### 2.1. Yoghurt fermentation and ultrasonic measurement

Samples of 100 ml whole fat UHT sterilised milk samples (Morrisons) were heated to 80 °C using a hot plate whilst stirring with a spatula before being cooled down to 45–50 °C using an ice container. Either 10 or 15 ml of whole fat yoghurt (Morrisons) was added to the milk samples and stirred for 30 s at 1000 rpm using a magnetic stirrer. The inoculated milk samples were then placed into a custom sample holder connected to 5 MHz transducer (Sonatest, SLIM5-10 t049502) (Fig. 1). A pH and temperature meter (HI-99164, Hanna Instruments) was placed into the sample holder as a reference measurement to provide the labelled data with which to train the ML models. The pH and temperature probe was placed against the wall of the sample holder to prevent blocking the path of the US wave. The optimal growth temperature of lactic acid bacteria is between 35 and 45 °C (Adamberg et al., 2003; Lee & Lucey, 2010; Schiraldi et al., 2006). Therefore, the samples were placed into a temperature controlled chamber which was set to either 40 or 42 °C to provide batch variability. Each fermentation lasted approximately 8 h. The pH and temperature values were recorded every 30 min. In total, twelve fermentation batches were monitored; three repeats for each of the yoghurt volume and temperature combinations to produce variability in the received US measurements. The US transducer was connected to a Lecouer Electronique US Box that provided the excitation pulses to the transducer and digitised the received US waveform. The US acquisition parameters were set to a frequency of 160 MHz and a gain of

50. Waveforms were continuously acquired in blocks of 60 s. Between each block of collected US waveforms, 600 s elapsed. There were approximately 17 US waveforms acquired in each block, each with 14, 000 sample points. Three US wave reflections were monitored: the 1st reflection was reflected from the near vessel-yoghurt interface, the 2nd reflection reflected from the far vessel-yoghurt interface after being transmitted through the fermenting yoghurt, and the 3rd reflection was reflected from the far vessel wall after transmission through the yoghurt (Fig. 2). An exemplar US waveform depicting these three reflections is presented in Fig. 3.

### 2.2. Machine learning

In this work, ML was used to predict the pH value during the yoghurt fermentation by using the US sensor measurements as input features. Real-time domain adaptation techniques were also used to ensure robustness of the ML models to outlier fermentation batches. A flow diagram is provided in Fig. 4 to show the algorithms and domain adaptation methodologies investigated. Three ML algorithms were investigated: Fully-Connected Neural Networks (FCNN), Long-Short Term Memory (LSTM) neural networks, and Convolutional Neural Networks with LSTM layers (CNN + LSTM). Three real-time domain adaptation methodologies were investigated: 1) feature alignment, 2) prediction alignment, and 3) feature removal. The first 8 h of each fermentation were used in the subsequent ML investigations to enable the same sequence length inputted into the neural networks with LSTM layers.

#### 2.2.1. Feature extraction

The following section describes the coarse features calculated from each waveform reflection to be used as inputs into the FCNNs and LSTMs. Coarse feature extraction was compared with convolutional feature extraction using the CNN + LSTMs as these are the two optimal methods for feature extraction when using US measurements for process monitoring (Bowler et al., 2021a, 2022).

*2.2.1.1. Energy.* The waveform Energy is the sum of all the squared amplitudes in a section of a waveform (Equation (1)). It is a measure of the magnitude of the sound wave and therefore may be used to monitor reflection coefficients at material boundaries or attenuation through a medium.

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

where $E$ is the Energy, $i$ is the sample point in a waveform, *start* is the first sample point, *end* is the last sample point, $A_i$ is the waveform amplitude at sample point $i$ (Zhan et al., 2015).

*2.2.1.2. Sum absolute amplitude.* The Sum Absolute Amplitude is the sum of all the absolute amplitudes in a section of a waveform (Equation (2)). It is also a measure of the magnitude of a sound wave, but it assigns larger weighting to smaller amplitudes than the Energy and therefore can be used in comparison with the Energy to identify changes in the shape of the US waveform.

$$SAA = \sum_{i=start}^{i=end} |A_i| \tag{2}$$

where *SAA* is the Sum Absolute Amplitude (Zhan et al., 2015).

*2.2.1.3. Peak-to-peak amplitude.* The peak-to-peak amplitude measures the range between the maximum and minimum amplitudes in a waveform section (Equation (3)).

$$PPA = max(A_{start:end}) - min(A_{start:end}) \tag{3}$$

Where *PPA* is the peak-to-peak amplitude, *max* and *min* indicate functions to find the maximum or minimum amplitudes in a waveform section, respectively.

*2.2.1.4. Maximum amplitude.* The maximum amplitude is the largest positive value in a waveform section (Equation (4)).

$$A_{max} = max(A_{start:end}) \tag{4}$$

Where $A_{max}$ is the maximum amplitude in a waveform section.

*2.2.1.5. Minimum amplitude.* The minimum amplitude is the largest negative value in a waveform section (Equation (5)).

$$A_{min} = min(A_{start:end}) \tag{5}$$

Where $A_{min}$ is the minimum amplitude in a waveform section.

*2.2.1.6. Standard deviation.* The standard deviation of sample point amplitudes along a waveform section is a measure of their dispersion relative to the mean value and its variation may be used to monitor changes in the waveform shape (Equations (6) and (7)).

$$\mu = \frac{\sum_{i=\text{start}}^{i=end} A_i}{start - end} \tag{6}$$

$$STD = \sqrt{\frac{1}{start - end} \sum_{i=start}^{i=end} (A_i - \mu)^2} \tag{7}$$

where $\mu$ is the mean amplitude in the waveform, *STD* is the standard deviation (Zhan et al., 2015).

*2.2.1.7. Skewness.* Skewness is a measure of the lack of symmetry in the waveform section (Equation (8)).

$$S = \frac{\sum_{i=start}^{i=end} (A_i - \mu)^3}{(end - start) \times STD^3} \tag{8}$$

where $S$ is the skewness (Caesarendra & Tjahjowidodo, 2017).

*2.2.1.8. Kurtosis.* The Kurtosis is a measure of the tailed-ness of the waveform section (Equation (9)).

$$K = \frac{\sum_{i=start}^{i=end} (A_i - \mu)^4}{(end - start) \times STD^4} \tag{9}$$

where $K$ is the kurtosis (Caesarendra & Tjahjowidodo, 2017).

*2.2.1.9. Time of flight.* The Time of Flight (TOF) was calculated for each of the three US wave reflections. The TOF of the 1st reflection can indicate changes in vessel temperature and the TOFs of the 2nd and 3rd reflections can identify changes to the yoghurt during fermentation. The TOF was calculated using a thresholding method followed by calculating the point of zero-crossing, i.e., the sample point were the waveform rises above the signal noise was identified and subsequently the next sample point were the waveform crosses the zero amplitude level was determined. This provides a consistent determination of the TOF that is not dependent on waveform shape.

*2.2.1.10. Energy standard deviation.* The Energy standard deviation is a measure of the dispersion in the Energy of consecutively acquired waveforms (Equation (10)). This can be used to identify fluctuations in the materials being measured by the US sensor.

$$ESTD = \sqrt{\frac{1}{W} \sum_{i=1}^{i=W} (E_i - \overline{E})^2} \tag{10}$$

where *ESTD* is the Energy standard deviation, $W$ is the number of

consecutively acquired waveforms investigated, and $\bar{E}$ is the average energy of the consecutively acquired waveforms.

*2.2.1.11. Time lagged features.* Time-lagged features were also included to provide the ML models with data from previous process time-steps and enable identification of the process trajectory. The US features from 1, 2, and 3 h previous were provided to the ML models. This increased the number of inputs features four-fold.

*2.2.1.12. Feature mean since process start.* The mean of each feature since the beginning of the process was also used as an ML model input to provide a measure of the current US features compared with the historic mean. To calculate this, at each time in the process, the mean of all previous values for that feature were calculated. This was then used as a new set of features alongside the features from which it was calculated. This further doubled the number of the input features.

### 2.2.2. Algorithms

*2.2.2.1. Fully-connected neural networks.* Fully-connected neural networks were investigated due to their ability to create new, complex features from combinations of original input features (Bowler et al., 2022). The FCNNs were used to determine whether the time-lagged features (see Section 2.2.1.11 Time lagged features) and historic feature means (see Section 2.2.1.12 Feature mean since process start) were sufficient to learn the process trajectory or if LSTM layers were also required. The Adam optimisation algorithm was used. A validation set was used to select the optimal number of fully-connected layers, the number of neurons in each layer, learning rate, batch size, and number of epochs for training. The selection of the validation and test sets are described in Section 3.1.1 Selection of test set. In total, 80 input features were used for the non-transmission methods and 240 for the transmission methods.

*2.2.2.2. Long short-term memory neural networks.* Long Short-Term Memory layers are able to learn sequences of time series data. LSTMs are a development of recurrent neural networks which reduces the likelihood of vanishing or exploding gradients (Hochreiter & Schmidhuber, 1997). All timesteps for each batch were used as a single sequence opposed to being truncated into multiple sequences of shorter length. While long sequences (500–1000 timesteps) are prone to producing vanishing or exploding gradients in LSTM layers when predicting a single output at the end of a sequence, this is less likely to occur when predicting an output at every timestep (Machine Learning Mastery, 2019). Using all timesteps enables the use of the full feature sequences for the ML models to make a prediction. A validation set was used to select the number of epochs for training, batch size, learning rate, number of neurons in the fully-connected layers, number of fully-connected layers, and number of units in the LSTM layer. The selection of the validation and test sets are described in Section 3.1.1 Selection of test set. The Adam optimisation algorithm was used with a gradient threshold of 1 to reduce the likelihood of exploding gradients. Similarly, 80 input features were used for the non-transmission methods and 240 for the transmission methods.

*2.2.2.3. Convolutional neural network with long short-term memory layers.* Convolutional layers in the CNN + LSTMs were also explored for their ability to directly monitor small changes to a US waveform compared with the coarse features outlines in Section 2.2.1 (Bowler, Pound, & Watson, 2021). LSTM layers were also included to enable learning of the process trajectory as the time-lagged features (2.2.1.11 Time lagged features) and historic feature means (2.2.1.12 Feature mean since process start) could not be applied to the full US waveform. The Adam optimisation algorithm was used. A validation set was used to select the optimal batch size, learning rate, number of convolutional layers, number of epochs for training, number of convolutional filters, and filter size. The selection of the validation and test sets are described in Section 3.1.1 Selection of test set.

**Table 1**
A summary of the features inputted into each of the ML algorithms.

| Algorithm | Fully-connected neural networks | Long Short-Term Memory neural networks | Convolutional neural network with Long Short-Term Memory layers |
|---|---|---|---|
| Input data type | Non-time series | Time series | Time series |
| Waveform reflections used | 1st reflection for non-transmission methods<br>1st, 2nd, and 3rd reflections for transmission methods | | |
| Features used | Energy, sum absolute amplitude, peak-to-peak amplitude, maximum amplitude, minimum amplitude, standard deviation, skewness, kurtosis, time of flight, and energy standard deviation calculated for each waveform reflection<br>Time lagged variations of the previously listed features<br>The feature mean since the start of the process for the previously listed features and time lagged feature variations | | Amplitudes at each sample point in the waveform reflections |
| Total number of features | 80 for non-transmission methods<br>240 for transmission methods | | 1000 for non-transmission methods<br>8000 for transmission methods |

**Table 2**
The hyperparameter ranges explored during model training.

| Hyperparameter | Range searched |
|---|---|
| Number of fully-connected layers | 1–5 |
| Number of epochs for training | 100–10,000 |
| Batch size | 1–9 |
| Learning rate | 0.00001–0.1 |
| Number of neurons in each layer | 1–64 |
| Number of units in the LSTM layer | 1–64 |
| Number of convolutional layers | 1–5 |
| Number of convolutional filters | 1–64 |
| Convolutional filter size | 3x1 – 50x1 |

*2.2.2.4. Summary of features used for each algorithm.* A summary of the features inputted into each algorithm is provided in Table 1. Time series data is input into the algorithms containing LSTM layers.

*2.2.2.5. Summary of hyperparameter search procedure.* A summary of the hyperparameter ranges search during this investigation is provided in Table 2. A combination of coarse-fine grid searches, random search, and prior knowledge were used to identify the optimal hyperparameters to predict the outputs of the validation set.

### 2.2.3. Real-time unlabelled domain adaptation

Three real-time unlabelled domain adaptation methodologies were investigated to improve ML model accuracy on data that falls outside the training dataset range: feature alignment, prediction alignment, and feature removal. This may occur due to changes in process conditions, such as operating temperature or materials being processed, affecting US sensor measurements. As it is not known when a change in feature distribution will occur or what the cause of these changes may be, a real-time domain adaptation method facilitates process monitoring in a factory environment by adjusting the US measurements as they are acquired.

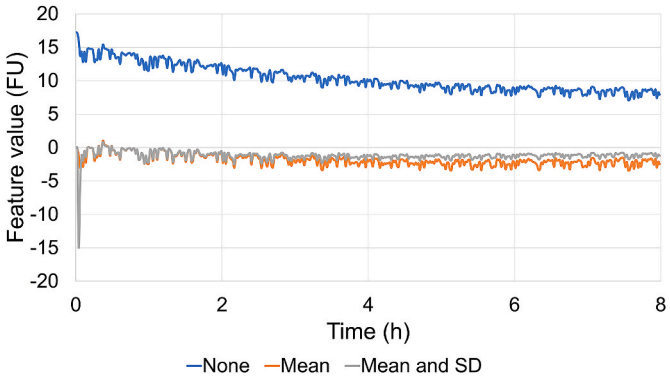


**Fig. 5.** An example of the feature alignment methods applied to a single US measurement feature taken from a yoghurt fermentation batch (None, blue line). Alignment of the historic mean (Mean, orange line) was either applied to the new data or to the training and new data as. Alignment of the historic mean and standard deviation (Mean and SD, grey line) was, similarly, applied to either the new data or to the training and new data.

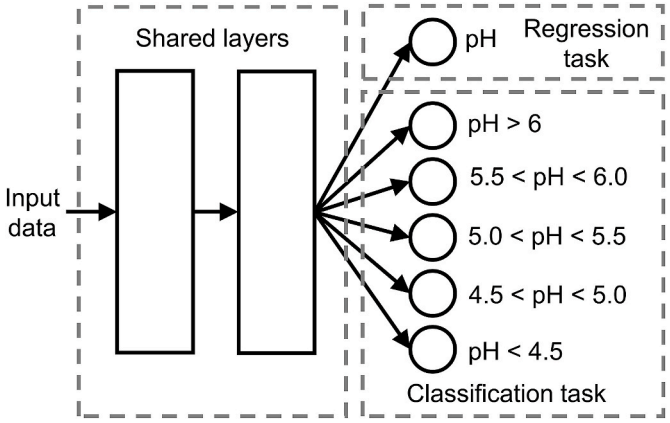


**Fig. 6.** A diagram of the multi-task network structure for alignment of the output predictions. To prevent offset of the predictions if the distribution of the input data were to change, a multi-task strategy with no task-specific feature extraction layers was used to constrain the outputs.

*2.2.3.1. Feature alignment.* Feature alignment aims to minimise the change in input data distribution between the training data and the new domain. Four methods of real-time feature alignment were investigated. These included: 1) aligning the historic mean of the new data to zero, 2) standardising the historic feature values of the new data, 3) aligning the historic mean of all training batches and the new data to zero, 4) standardising all historic feature values of the training batches and the new data. To achieve this, for each new waveform acquired, the means of all previous feature values were subtracted from the current feature values to align the mean of the new data to zero. If standardisation was also employed, then the feature values were subsequently divided by the standard deviation of all feature values from the beginning of the process as well. An example of these methodologies is provided in Fig. 5.

*2.2.3.2. Prediction alignment.* As ML models correlate input data to output data it is expected that if the distribution of the input data changes then the learned correlation will produce outputs with an offset from their true values. To counteract this, multi-task learning was employed to also predict a classification task by discretising the pH values. Therefore, even if the feature distribution were to change, the output predictions are more constrained to the output pH encountered during training. To achieve this, both regression and classification output neurons were connected to a shared final feature extraction layer (Fig. 6). Five classes were predicted including pH values of greater than 6.0, 5.5–6.0, 5.0–5.5, 4.5–5.0, and 4.0–4.5.

*2.2.3.3. Feature removal.* Finally, it is hypothesised that features from the new data that deviate far from the training values will be of less use when predicting the outputs of the new data. The reason for this is that the trained ML models will be less able to maintain accurate prediction when these features are included as their value ranges were not encountered during model training. To remove these features, following standardisation using the mean and standard deviation of the training set, any feature values more than three standard deviations from the mean of training set were set to zero, preventing interference with the prediction (Fig. 7). This therefore gives larger influence to features which have a lower deviation from the training set.

## 3. Results

Fig. 8 presents the ultrasonic measurements for each yoghurt fermentation batch. Fig. 8a displays the waveform energy of the 1st reflection for all fermentation batches. Notably, the 1st reflection energy for Batch 7 is far greater than all other batches. Resultantly, Batch 7 was chosen as the test set to evaluate the various domain adaptation methodologies. Fig. 8b also displays the 1st reflection energy, but with Batch 7 removed. Overall, a decrease in the 1st reflection energy was found during fermentation corresponding to an increasing acoustic impedance of the yoghurt and is in agreement with the results found in Meng et al. (2012). The acoustic impedance is the product of the speed of sound and material density (McClements, 1995) and as the density of the yoghurt does not change during fermentation (Meng et al., 2012) the change in acoustic impedance is caused by the changing speed of sound. The TOF through the fermenting yoghurt is shown in Fig. 8d. The TOF initially increases as the sample is placed into the temperature controlled chamber and then decreases throughout fermentation. A decreasing TOF corresponds to an increasing speed of sound and was also found in previous studies (Aljaafreh & Lucklum, 2015; Alouache et al., 2015; Krasaekoopt et al., 2005; Masuzawa et al., 2003; Meng et al., 2012). Furthermore, the TOF follows the same trend as the pH (Fig. 8e) by rapidly decreasing during approximately the first 4 h before the rate of decrease slows until the end of fermentation. The increasing speed of sound was found to follow the same trend of decreasing pH in (Aljaafreh & Lucklum, 2015) as well as increasing viscosity in (Krasaekoopt et al., 2005, Masuzawa et al., 2003). The increase in viscosity is due to the formation of a gel structure which increases the bulk modulus of the yoghurt and subsequently the speed of sound (Masuzawa et al., 2003). The attenuation (Fig. 8c) is shown to rapidly increase during the first 2 h of fermentation before reaching a plateau. This is similar to the results found in (Aljaafreh & Lucklum, 2015; Alouache et al., 2015; Masuzawa

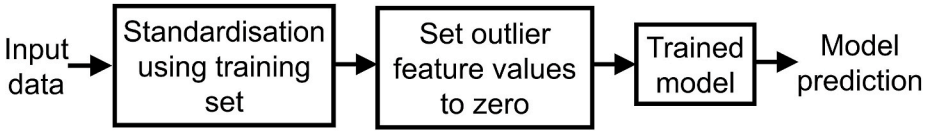


**Fig. 7.** The steps involved in the feature removal methodology. It is anticipated that features from the new data that were not encountered during model training would be disruptive to the model predictions. Therefore, these features were removed and more importance was given to the feature values located in the training set ranges. An outlier was defined as being three standard deviations from the mean value.

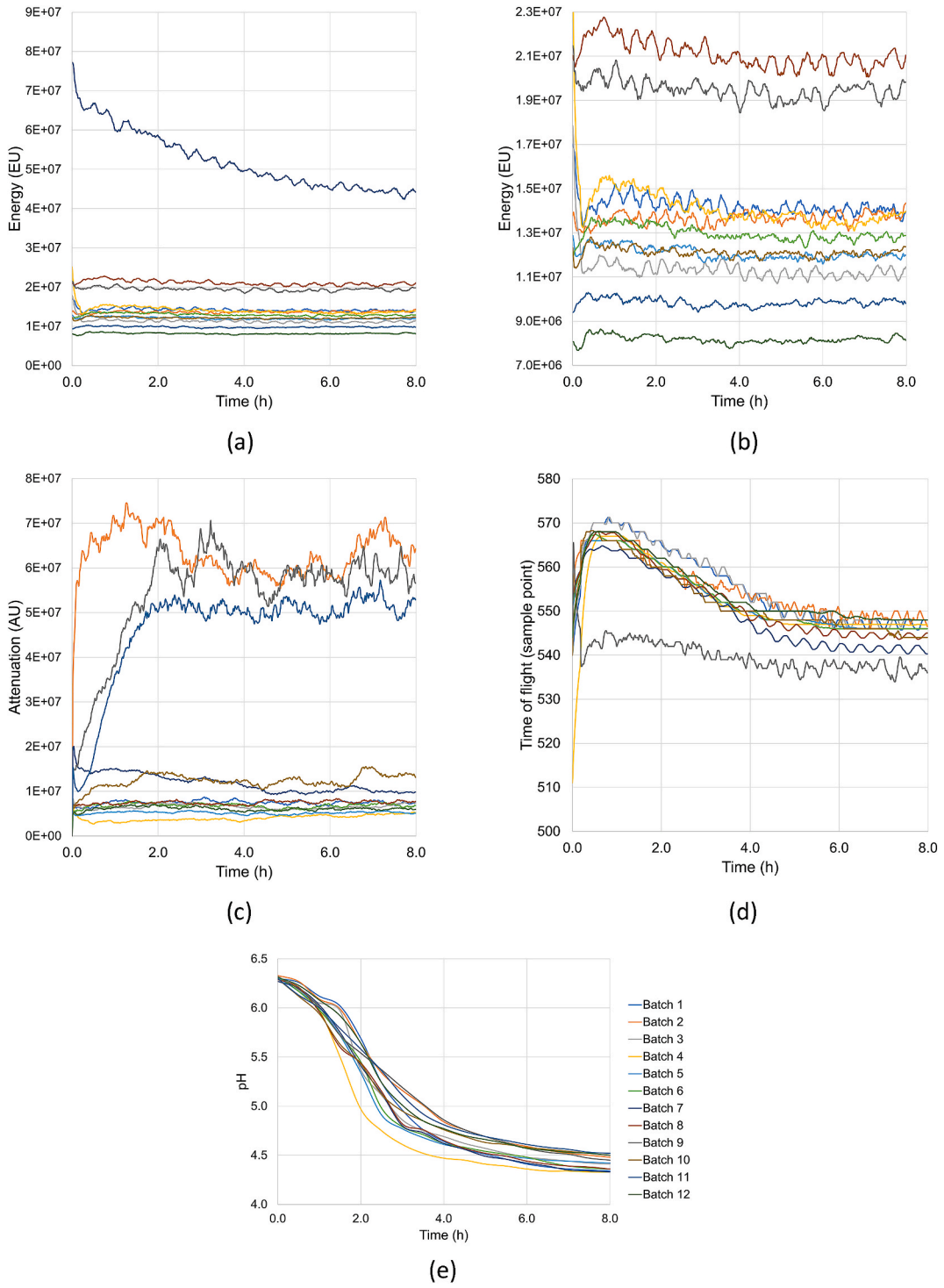


**Fig. 8.** The results of ultrasonic measurements for each yoghurt fermentation batch. (a) The waveform energy of the 1st reflection for all batches. A decrease in the 1st reflection energy corresponds to an increase in the yoghurt acoustic impedance. (b) The waveform energy of the 1st reflection. Batch 7 excluded. (c) The attenuation of the sound wave as it passes through the fermenting yoghurt for all batches. (d) The time of flight (TOF) of the sound wave for all batches. A decrease in the TOF corresponds to an increase in the speed of sound through the yoghurt. (e) The pH during fermentation for all batches.

et al., 2003) where the attenuation follows the same trend as the speed of sound due to the increasing viscosity causing increased energy absorption and the aggregation of micelles causing increased scattering of the sound wave.

### 3.1. Machine learning investigations

#### 3.1.1. Selection of test set

As previously discussed (Section 3), Batch 7 was chosen as the test set owing to it displaying the greatest variability compared to all other batches (Fig. 9). Therefore, it was the ideal fermentation batch to evaluate the real-time domain adaptation strategies to enable ML model robustness to changing input data distributions caused by changing US

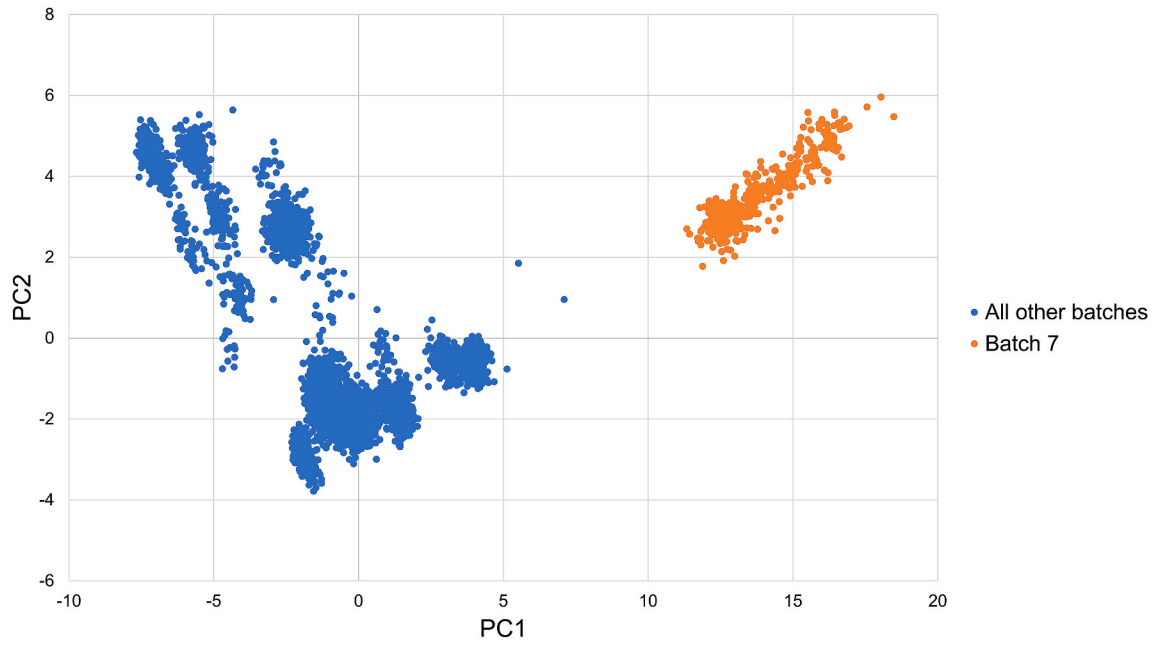


**Fig. 9.** A comparison of Batch 7 to all other batches. Principal component analysis was used to extract two principal components for all the coarse features (240 features) of the transmission based methodology. As can be seen, Batch 7 is an outlier compared to all other batches and therefore was selected as the test set to evaluate the real-time domain adaptation methodologies to changing input feature distributions. Batches 11 and 12 were selected as the validation set for hyperparameter evaluation in all investigations.

**Table 3**
The results of the domain adaptation investigations to predict the pH value during fermentation. The most accurate results discussed in the text are highlighted in bold. Overall, the most accurate method achieved an MSE of 0.008 using the non-transmission method, LSTM + CNN algorithm, and a combination of all three domain adaptation methodologies with feature alignment of the historic mean and standard deviation for all batches.

| Algorithm | | | FCNN | | LSTM | | LSTM + CNN | |
|---|---|---|---|---|---|---|---|---|
| Transmission | | | ✓ | × | ✓ | × | ✓ | × |
| Prediction alignment | Feature removal | Feature alignment | | | | | | |
| × | × | None | 8.881 | 0.588 | 0.903 | 0.608 | **0.049** | **0.024** |
| | | Test mean | 2.327 | 5.987 | 6.311 | 2.234 | 1.225 | 0.634 |
| | | Test mean and SD | 20.644 | 3.111 | 5.762 | 0.144 | 3.366 | 0.731 |
| | | All mean | 17.193 | 2.945 | 10.054 | 5.708 | 0.135 | 2.637 |
| | | All mean and SD | 0.344 | 1.466 | 0.509 | 0.559 | 0.010 | 0.148 |
| ✓ | × | None | 16.060 | 0.904 | 1.422 | 1.180 | 0.127 | 0.982 |
| | | Test mean | 0.600 | 6.624 | 0.574 | 0.437 | 7.691 | 0.447 |
| | | Test mean and SD | 26.452 | 3.402 | 1.888 | 0.224 | 0.903 | 0.638 |
| | | All mean | 14.676 | 2.891 | 0.484 | 0.437 | 0.372 | 0.108 |
| | | All mean and SD | 0.576 | 1.581 | 0.441 | 2.079 | **0.040** | 0.157 |
| × | ✓ | None | 10.590 | 0.643 | 0.979 | 0.424 | 0.488 | 6.974 |
| | | Test mean | 0.498 | 0.502 | 0.694 | 0.862 | 3.663 | 0.810 |
| | | Test mean and SD | 1.385 | 1.008 | 10.260 | 0.514 | 3.159 | 1.867 |
| | | All mean | 2.322 | 0.930 | 0.854 | 2.146 | 0.195 | 0.110 |
| | | All mean and SD | 0.420 | 1.808 | 1.198 | 0.738 | 0.491 | **0.013** |
| ✓ | ✓ | None | 14.148 | 0.739 | 0.566 | 0.763 | 0.452 | 1.585 |
| | | Test mean | 0.601 | 0.546 | 0.474 | 0.801 | 1.026 | 0.314 |
| | | Test mean and SD | 1.863 | 1.387 | 0.208 | 0.627 | 0.734 | 3.458 |
| | | All mean | 3.129 | 0.915 | 1.061 | 0.695 | 0.273 | 0.194 |
| | | All mean and SD | 0.400 | 1.683 | 0.156 | 1.067 | 0.173 | **0.008** |

features. The validation set for hyperparameter tuning in all investigations was selected as Batches 11 and 12. After hyperparameter selection, new models were trained using all of the training and validation sets. Fig. 8a displays the waveform energy during Batch 7 to be greater than the remaining batches, whereas Batch 7 is not displayed as an outlier in Fig. 8b or 8c displaying the waveform attenuation or TOF, respectively. However, the domain adaptation methods used in this investigation are agnostic to the change in feature distribution causing a batch to become an outlier. This is because during industrial application the change in feature distribution cannot be predicted. Therefore, these methods are still of utility for variations in features other than the waveform magnitude as present in this investigation.

### 3.1.2. Domain adaptation

Table 3 presents the results of all the domain adaptation investigations. Overall, the most accurate method used the non-transmission method, LSTM + CNN algorithm, and all three domain adaptation methodologies with feature alignment of the historic mean and standard deviation for all batches. This method achieved a Mean Squared Error (MSE) of 0.008 to predict the pH value during fermentation and is displayed in Fig. 10. As can be seen with the closeness of the predicted and true pH value, this method is suitable for in-line

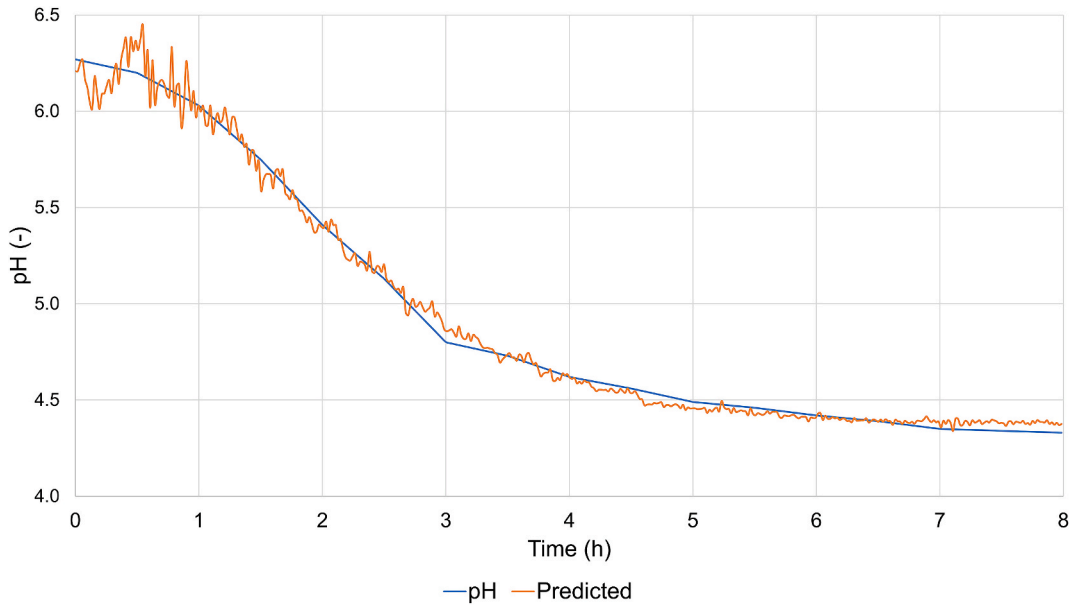


**Fig. 10.** The results of the most accurate ML model (MSE = 0.008). This method used the non-transmission method, LSTM + CNN algorithm, and a combination of all three domain adaptation methodologies with feature alignment of the historic mean and standard deviation for all batches. The closeness of the real and predicted pH evidences the model's applicability to real time fermentation monitoring.

monitoring of yoghurt fermentation processes to aid optimal monitoring of the fermentation rate. Overall, the LSTM + CNN algorithm was used for all the most accurate methods (bolded Table 3). As previously stated, the LSTM + CNN method achieved an MSE of 0.008 using the non-transmission method and 0.040 using the transmission method. This shows that the convolutional feature extraction method was able to learn features more informative to the task compared with the coarse features extracted using the FCNN and LSTM methods. This result was also found previously in (Bowler, Pound, & Watson, 2021) where convolutional and coarse US features were compared to monitor mixing, cleaning, and alcoholic fermentation processes. The LSTM method achieved lower MSE results, 0.144 versus 0.344, compared with FCNNs as also found in (Bowler, Escrig, et al., 2021) where these algorithms were compared to monitor alcoholic fermentation using US measurements. This demonstrates that despite the addition of time-lagged features and the historic process mean, the more complex trajectory learning capable by the LSTM layers were still required. Notably, the non-transmission method achieved higher accuracy compared with the transmission based method despite the greater information available from the 2nd and 3rd sound wave reflections in transmission based methods. The reason for this may be that by using fewer features, the models extracted more general information for the US waveform that transferred more effectively to the outlier batch test set. However, this is a promising result for industrial application of the US sensing and ML combination given that transmission methods are not possible due to the large size of the fermentation vessels.

Prediction alignment of the historic mean and standard deviation for all batches produced the most accurate models for both the transmission and non-transmission methods. This indicates that this method was able to align the features from Batch 7 as well as provide the model with similar examples by also aligning the features for the training batches. Utilising prediction alignment based on multi-task learning as well as feature removal produced the most accurate method overall. This indicates that each domain adaptation method could provide benefits to an ML model and enable robustness to outlier data. However, only using multitask learning to align the prediction produced the highest accuracy using the transmission method. This may be because the feature removal procedure removed features that despite being outliers to the training set still contained useful information to the ML model. Interestingly, using no domain adaptation strategies were the second and third most accurate for the transmission and non-transmission tasks, respectively. The optimal hyperparameters when using no domain adaptation methods were the same as the most accurate methods using domain adaptation (Table 4). Therefore, this shows that the small network size (two convolutional filters in each layer, two neurons in the fully connected layer, and two LSTM units) compressed the input waveforms into

**Table 4**
Optimal hyperparameters for best performing ML models (bolded in Table 3). The small network size was able to extract generalisable features to the outlier test batch.

| Hyperparameter | Value |
|---|---|
| Number of convolutional layers | 3 |
| Filter size | 3x1 |
| Number of filters in each layer | 2 |
| Max pooling size | 2x1 |
| Number of neurons in fully connected layer | 2 |
| Number of LSTM units | 2 |
| Learning rate | 0.01 |
| Batch size | 2 |

**Table 5**
Evaluation of the best performing model on various cross-sections of the data set. The best performing model was non-transmission based, used a convolutional neural network with long short-term memory layer and a combination of all three domain adaption methods.

| Number of batches included in the test set | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Number of batches included in the training set | 11 | 10 | 9 | 8 | 7 |
| Batches included in the test set | Batch 1 | Batches 1 and 2 | Batches 1, 2, and 3 | Batches 1, 2, 3, and 4 | Batches 1, 2, 3, 4, and 5 |
| MSE | 0.009 | 0.017 | 0.019 | 0.137 | 0.191 |

the simplest trends in the US measurements to maximise transfer across changing feature distributions.

### 3.1.3. Evaluation of best method

The best performing model was non-transmission based, used a convolutional neural network with long short-term memory layer and a combination of all three domain adaption methods. During the feature removal domain adaptation method, 1.2% of the input data from Batch 7 was removed to aid with model generalisation. To further test the method and to evaluate whether a greater number of fermentation batches should be monitored during application of the US and ML combination, varying test set sizes have been used to evaluate the optimal method. As Batches 11 and 12 were previously used as a validation set, Batches 1–5 have been used as a test set for this investigation. The results in Table 5 show that with up to 3 batches included in the test set, the MSE is still equivalent to the best performing models depicted in Table 3 (shown in bolded text) proving that the method is robust for test sets other than Batch 7. However, the MSE continues to decrease with decreasing test set size which indicates that collection of US sensor data from further batches would increase model accuracy.

## 4. Conclusions

In manufacturing environments, real-time monitoring of yoghurt fermentation batches is required to optimally maintain production schedule, product quality, and prevent the growth of pathogenic bacteria. In this work, non-invasive ultrasonic transmission and non-transmission based methods were compared. Overall, the non-transmission method achieved a higher accuracy (0.008 MSE of the pH value versus 0.0399) which is promising for industrial application where transmission based methods are not possible due to the large vessel sizes. Three machine learning algorithms were compared. It was found that using convolutional layers followed by long short-term memory units were most accurate suggesting that they were able to extract the most informative features from the waveform along with optimal learning of the process trajectory. Three real-time domain adaptation procedures were compared to ensure machine learning model robustness when encountering batch variability. These were feature alignment of the historic mean and standard deviation, prediction alignment using multitask learning, and removal of features that were outliers to the training set. A combination of all three methods produced the lowest MSE.

## Funding

This work was supported by the EPSRC grant for EATS - Enhancing Agri-Food Transparent Sustainability (EP/V041371/1) and the Higher Education Innovation Funding (HEIF) from the University of Nottingham.

## CRediT authorship contribution statement

**Alexander Bowler:** Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing. **Samet Ozturk:** Conceptualization, Methodology, Investigation. **Vincenzo di Bari:** Supervision, Writing – review & editing. **Zachary J. Glover:** Writing – review & editing. **Nicholas J. Watson:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

## Declaration of competing interest

None.

## Data availability

Data will be made available on request.

## References

Abildgaard, O. H. A., Kamran, F., Dahl, A. B., Skytte, J. L., Nielsen, F. D., Thomsen, C. L., Andersen, P. E., Larsen, R., & Frisvad, J. R. (2015). Non-invasive assessment of dairy products using spatially resolved diffuse reflectance spectroscopy. *Applied Spectroscopy, 69*(9), 1096–1105. https://doi.org/10.1366/14-07529

Adamberg, K., Kask, S., Laht, T. M., & Paalme, T. (2003). The effect of temperature and pH on the growth of lactic acid bacteria: A pH-auxostat study. *International Journal of Food Microbiology, 85*, 171–183. https://doi.org/10.1016/S0168-1605(02)00537-8

Alguri, K. S., Chia, C. C., & Harley, J. B. (2021). Sim-to-Real: Employing ultrasonic guided wave digital surrogates and transfer learning for damage visualization. *Ultrasonics, 111*, Article 106338. https://doi.org/10.1016/j.ultras.2020.106338

Aljaafreh, A., & Lucklum, R. (2015). On-line monitoring of yogurt fermentation using ultrasonic characteristics. *New Developments in Circuits, Systems, Signal Processing, Communications and Computers, 1*.

Alouache, B., Touat, A., Boutkedjirt, T., & Bennamane, A. (2015). Monitoring of lactic fermentation process by ultrasonic technique. *Physics Procedia, 70*, 1057–1060. https://doi.org/10.1016/j.phpro.2015.08.225

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 Research International, 48*(2), 410–427. https://doi.org/10.1016/j.foodres.2012.05.004

Bowler, A., Escrig, J., Pound, M., & Watson, N. (2021). Predicting alcohol concentration during beer fermentation using ultrasonic measurements and machine learning. *Fermentation, 7*(1), 34. https://doi.org/10.3390/fermentation7010034

Bowler, A., Pound, M., & Watson, N. (2021). Convolutional feature extraction for process monitoring using ultrasonic sensors. *Computers & Chemical Engineering, 155*, Article 107508. https://doi.org/10.1016/j.compchemeng.2021.107508

Bowler, A. L., Pound, M. P., & Watson, N. J. (2022). A review of ultrasonic sensing and machine learning methods to monitor industrial processes. *Ultrasonics, 124*, Article 106776. https://doi.org/10.1016/j.ultras.2022.106776

Bowler, A. L., & Watson, N. J. (2021). Transfer learning for process monitoring using reflection-mode ultrasonic sensing. *Ultrasonics, 115*, Article 106468. https://doi.org/10.1016/j.ultras.2021.106468

Bull, L. A., Gardner, P. A., Dervilis, N., Papatheou, E., Haywood-Alexander, M., Mills, R. S., & Worden, K. (2021). On the transfer of damage detectors between structures: An experimental case study. *Journal of Sound and Vibration, 501*, Article 116072. https://doi.org/10.1016/j.jsv.2021.116072

Caesarendra, W., & Tjahjowidodo, T. (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*(4), 21. https://doi.org/10.3390/machines5040021

Food and Agriculture Organisation of the United Nation (FAO). (2018). Standard for fermented milks. *CXS*, 243–2003.

Gao, X., Shi, Y., Zhu, Q. I., Li, Z., Sun, H. U., Yao, Z., & Zhang, W. (2021). Domain adaptation in intelligent ultrasonic logging tool: From microseismic to pulse-echo. *IEEE Transactions on Instrumentation and Measurement, 70*, 1–14. https://doi.org/10.1109/TIM.2021.3050154

Gardner, P., Bull, L. A., Gosliga, J., Poole, J., Dervilis, N., & Worden, K. (2022). A population based SHM methodology for heterogeneous structures: Transferring damage localisation knowledge between different aircraft wings. *Mechanical Systems and Signal Processing, 172*, Article 108918. https://doi.org/10.1016/j.ymssp.2022.108918

Hassoun, A., Jagtap, S., Garcia-Garcia, G., Trollman, H., Pateiro, M., Lorenzo, J. M., Trif, M., Rusu, A. V., Aadil, R. M., Šimat, V., Cropotova, J., & Câmara, J. S. (2023). Food quality 4.0: From traditional approaches to digitalized automated analysis. *Journal of Food Engineering, 337*, Article 111216. https://doi.org/10.1016/j.jfoodeng.2022.111216

Henning, B., & Rautenberg, J. (2006). Process monitoring using ultrasonic sensor systems. *Ultrasonics, 44*, e1395–e1399. https://doi.org/10.1016/j.ultras.2006.05.048

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Horne, D. S. (1999). Formation and structure of acidified milk gels. *International Dairy Journal, 9*(3–6), 261–268. https://doi.org/10.1016/S0958-6946(99)00072-2

Krasaekoopt, W., Bhandari, B., & Deeth, H. (2005). Comparison of gelation profile of yoghurts during fermentation measured by RVA and ultrasonic spectroscopy. *International Journal of Food Properties, 8*(2), 193–198. https://doi.org/10.1081/JFP-200059469

Lee, W. J., & Lucey, J. A. (2010). Formation and physical properties of yoghurt. *Food Research International, 30*(7), 529–542. https://doi.org/10.1016/S0963-9969(98)00015-5

Machine learning Mastery. Available online: https://machinelearningmastery.com/handle-long-sequences-long-short-term-memory-recurrent-neural-networks/, (2019)–. (Accessed 12 January 2022).

Masuzawa, N., Kimura, A., & Ohdaira, E. (2003). Ultrasonic monitoring of the progress of lactic acid fermentation. *Japanese Journal of Applied Physics. Part 1: Regular Papers and Short Notes and Review Papers, 42*(5 B), 2963–2964. https://doi.org/10.1143/jjap.42.2963

McClements, D. J. (1995). Advances in the application of ultrasound in food analysis and processing. *Trends in Food Science & Technology, 6*(9), 293–299. https://doi.org/10.1016/S0924-2244(00)89139-6

Meng, R., Zhou, J., Ye, X., & Liu, D. (2012). On-line monitoring of yogurt fermentation using acoustic impedance method. *Applied Mechanics and Materials, 101–102*, 737–742. https://doi.org/10.4028/www.scientific.net/AMM.101-102.737

Mohd Khairi, M. T., Ibrahim, S., Md Yunus, M. A., & Faramarzi, M. (2015). Contact and non-contact ultrasonic measurement in the food industry: A review. *Measurement Science and Technology, 27*(1), Article 012001. https://doi.org/10.1088/0957-0233/27/1/012001

Mohd Khairi, M. T., Ibrahim, S., Md Yunus, M. A., & Faramarzi, M. (2018). Noninvasive techniques for detection of foreign bodies in food: A review. *Journal of Food Process Engineering, 41*(6), Article e12808. https://doi.org/10.1111/jfpe.12808

Muncan, J., Tei, K., & Tsenkova, R. (2021). Real-time monitoring of yoghurt fermentation process by aquaphotomics near-infrared spectroscopy. *Sensors, 21*(1), 1–18. https://doi.org/10.3390/s21010177

Schiraldi, C., Valli, V., Molinaro, A., Cartení, M., & De Rosa, M. (2006). Exopolysaccharides production in Lactobacillus bulgaricus and Lactobacillus casei exploiting microfiltration. *Journal of Industrial Microbiology and Biotechnology, 33*, 384–390. https://doi.org/10.1007/s10295-005-0068-x

Verdú, S., Barat, J. M., & Grau, R. (2019). Non destructive monitoring of the yoghurt fermentation phase by an image analysis of laser-diffraction patterns: Characterization of cow's, goat's and sheep's milk. *Food Chemistry, 274*, 46–54. https://doi.org/10.1016/j.foodchem.2018.08.091

Walstra, P., Wouters, J. T. M., & Geurts, T. J. (2005). *Dairy science and technology* (2nd ed.) (Chapter 22)).

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. https://doi.org/10.3390/s150924109