Optimizing Terahertz Waveform Selection of a Pharmaceutical Film Coating Process Using Recurrent Network

Xiaoran Li, Bryan M. Williams, Robert K. May, Michael J. Evans, Shuncong Zhong, Lynn F. Gladden, Yaochun Shen, J. Axel Zeitler, and Hungyen Lin, *Senior Member, IEEE*

***Abstract*—In-line terahertz pulsed imaging has been utilized to measure the film coating thickness of individual tablets during the coating process in a production-scale pan coater. A criteria-based waveform selection algorithm (WSA) was developed to select terahertz signals reflected from the surface of coating tablets and determine the coating thickness. Since the WSA uses many criteria thresholds to select terahertz waveforms of sufficiently high quality, it could reject some potential candidate tablet waveforms that are close but do not reach the threshold boundary. On the premise of the availability of large datasets, we aim to improve the efficiency of WSA with machine learning. This article presents a recurrent neural network approach to optimize waveform selection. In comparison with the conventional method of WSA, our approach allows more than double the number of waveforms to be selected while maintaining excellent agreement with offline thickness measurements. Moreover, the processing time of waveform selection decreases so that it can be applied for real-time coating monitoring in the pharmaceutical industry, which leads to more advancements in quality control for pharmaceutical film coating.**

***Index Terms*—Coating uniformity, machine learning, neural network, pharmaceutical film coating thickness, terahertz pulsed imaging (TPI).**

Manuscript received September 13, 2021; revised December 16, 2021 and March 1, 2022; accepted March 28, 2022. Date of publication April 1, 2022; date of current version July 5, 2022. The work of Xiaoran Li and Hungyen Lin was supported in part by the EPSRC under Grant EP/R019460/1, in part by H2FC Supergen Flexible Grant EP/P024807/1, and in part by Australian Defence Science and Technology Group. Additional data sets related to this publication are available from the Lancaster University data repository https://doi.org/10.17635/lancaster/researchdata/523

Xiaoran Li and Hungyen Lin are with the Department of Engineering, Lancaster University, Lancaster LA1 4YW, U.K. (e-mail: x.li59@lancaster.ac.uk; h.lin2@lancaster.ac.uk).

Bryan M. Williams is with the Department of Computing and Communications, Lancaster University, Lancaster LA1 4YW, U.K. (e-mail: b.williams6@lancaster.ac.uk).

Robert K. May and Shuncong Zhong are with TeraView Ltd., Cambridge CB4 0DS, U.K. (e-mail: rob.may@teraview.com; sczhong@fzu.edu.cn).

Michael J. Evans is with the Laboratory of Optics, Terahertz and Nondestructive Testing, School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou 350025, China (e-mail: mike.evans@teraview.com).

Lynn F. Gladden is with the Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge CB2 1TN, U.K. (e-mail: lfg1@cam.ac.uk).

Yaochun Shen is with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3BX, U.K. (e-mail: y.c.shen@liverpool.ac.uk).

J. Axel Zeitler is with the Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge CB2 1TN, U.K. (e-mail: jaz22@cam.ac.uk).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TTHZ.2022.3164353.

Digital Object Identifier 10.1109/TTHZ.2022.3164353

I. Introduction

PHARMACEUTICAL film coating processes are widely used to ensure color uniformity, light protection and taste masking of the dosage forms. Functional coatings can be used to mask the taste or smell of a product, to protect the active pharmaceutical ingredient (API) against the acidic environment of the stomach or the gastric mucosa against an aggressive API, and to prolong API release. Coating thickness and integrity are particularly important for functional coatings where a minimum thickness is required to ensure gastro-resistance of a dosage form or to achieve a targeted release profile/ rate. Active coatings contain an API, the amount of which is directly correlated to coating thickness. Several non-destructive sensing techniques have been demonstrated to quantify film coating thickness such as Raman spectroscopy [1], near-infrared spectroscopy [2] and more recently, optical coherence tomography [3], [4]. Terahertz pulsed imaging (TPI) was introduced approximately 15 years ago and has attracted interest in the pharmaceutical industry as a fast, nondestructive modality for quantifying film coatings on pharmaceutical tablets. Extensive reviews on the subject matter have been conducted [5]–[8]. In short, incident terahertz pulses penetrate through the coatings where a portion of the terahertz pulse reflects back to the detector at each coating interface or abrupt change in refractive index. Film coating thickness is then determined from the separation of measured reflection peaks in the processed signal, given by $d = c\Delta t/(2n)$, where d is the coating thickness, c is the speed of light, Δt is the peak separation time and n is the refractive index of the coating material. In an effort to better understand the pharmaceutical coating process, in-line TPI has been demonstrated to measure tablet coating thickness directly during the coating operation with a fast acquisition rate (up to 120 Hz) thus yielding statistical information on the coating variability of the tablet population inside the coating unit. This has been demonstrated for production [9], [10] and lab-scale [11] processes and consolidated against numerical modeling [12], [13]. In processing the saved data stream of raw waveforms, a criteria-based waveform selection algorithm (WSA) was developed previously [9], which

can select waveforms that represent reflections from the coated tablets that are both in the focus of the terahertz beam and normally aligned to the terahertz detector. The WSA is dependent on a combination of thresholds determined using offline data including the amplitude, width, and position of the peaks that represent tablet surface and subsequent coating/core interface. Depending on the threshold values, WSA could reject potential candidate tablet waveforms resulting in limited thickness measurements taken i.e., hits. Opportunities therefore loom to increase the number of tablet waveforms selected by relaxing the parameters [10] without significantly compromising on the accuracy to offline thickness measurements. This would then allow for a more complete representation of the process allowing for systematic investigations to be performed leading to greater process understanding [10].

Significant progress have been made on analyzing terahertz measurements of samples with simple geometries using conventional terahertz time-domain spectroscopy in an ideal laboratory condition [14], [15]. Under such scenarios, an understanding of the underlying mechanisms is often needed e.g., Lorentzian profiles for water vapor removal [16] or a double Debye in resolving molecular water states [17]. For analyzing measurements that are convoluted by additional effects, such as shape irregularity, variations in thickness, attenuations, Fabry–Perot oscillations and the combinations thereof, experimental data-driven approaches using machine learning algorithms have found success in automatic signal recognition without manual and professional intervention. For example, convolutional neural networks (CNNs) [18]–[21] and various shallow classifiers, like support vector machines [22]–[27], k-nearest neighbors [23], [28], [29], random forest [30], and other single-layer neural networks [31]–[34] have been used for image/waveform classification. With the exception of CNN, these common classifiers are usually used in conjunction with feature extraction methods, e.g., principal component analysis. In the context of thickness measurements geometric algebra has also been proposed, though for classifying transmission measurements at thicknesses between 0.5–3 mm in an ideal laboratory setting [35]. Given the complexity of the pharmaceutical tablet pan coating process (spray atomization, thermodynamic drying and tablet mixing) [36] coupled terahertz in-line measurement through perforations of a rotating coating pan travelling at 0.41 m/s, developing an analytical solution is non-trivial. Consequently, mechanistic modeling using discrete element models are used yielding promising results [12], [13] As large volumes of experimental datasets can be generated, the challenge is to investigate whether a data-driven approach using machine learning could also be used to learn the features for real-life classification, representative of a production scale coating environment.

A. Long-Short Term Memory Network

Recurrent neural networks (RNN) form a widely-used class of artificial neural network, which is capable of handling sequential data. Since it was first described in 1982 [37], this class of network has been developed for application in many different fields, such as speech recognition [38] and machine translation [39]. It contains a learning structure where all inputs and outputs are inter-connected [40]. Hence, the input from each unit depends on the previous output. Long-short term memory (LSTM) was developed by [41] to overcome the limitations of RNNs with improved ability to capture long-term dependencies. Unlike standard RNN unit, the memory cell in LSTM replaces the traditional node in the hidden layer.

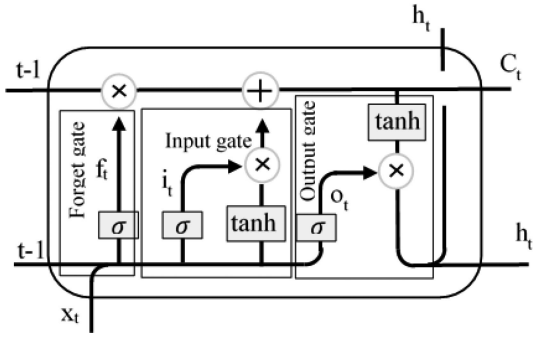

Fig. 1. Illustration of LSTM cell unit.

Fig. 1 shows the internal structure of a LSTM unit. It contains memory blocks and memory cells, along with the gate units they contain [42]. The input gate takes charge of the input stream to the memory cell and the output gate takes charge of the output flow to other LSTM units. The forget gate is controlled by a one-layer neural network. The output of the forget gate is a sigmoid activation function that is connected to the previous memory block. The activation of this gate is calculated as

$$f_t = \sigma\left(P\left[x_t, h_{t-1}, C_{t-1}\right] + V_f\right) \tag{1}$$

where x_t is the sequence input, C_{t-1} is the previous LSTM block memory, h_{t-1} is the previous block output and V_f is the bias vector. P is the weight vectors for the input and σ represents the sigmoid activation function. Since the output of the forget gate is applied to the memory block of previous LSTM unit, the previous memory block will affect the signal of the current LSTM. If the activation output vector values are close to zero, then the previous memory will be erased [40]. The input gate creates new memory determined by a simple neural network with the tan*h* activation function and the previous memory block. These signals can be expressed as

$$i_t = \sigma\left(P\left[x_t, h_{t-1}, C_{t-1}\right] + V_i\right) \tag{2}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tanh\left(P\left[x_t, h_{t-1}, C_{t-1}\right] + V_c\right). \tag{3}$$

The output gate generates the output signal of the current LSTM unit. These outputs can be expressed as

$$o_t = \sigma\left(P\left[x_t, h_{t-1}, C_{t-1}\right] + V_o\right) \tag{4}$$

$$h_t = \tanh\left(C_t\right) \cdot o_t. \tag{5}$$

As the most popular type of RNN, LSTM has also been widely applied for classifying sequential data such as electrocardiogram signals [40] and power load forecasts [43]. Since terahertz waveforms are time dependent, LSTM is used to analyze terahertz

waveforms. Additional justifications for the choice of LSTM over MLP or CNN can be found in supplementary information where higher sensitivity, precision, and accuracy are observed. Furthermore, CNN requires the waveforms to be converted to images prior to classification, which would require additional overheads thus may not be ideal for waveform selection in real-time [44]. Here, a particular type of LSTM network i.e., bidirectional LSTM (BLSTM) network, which divides regular RNN neuron states into forward and backward for higher selection accuracy [40].

II. Methodology

A. In-Line and Offline Dataset Acquisition

Details of the coating experiments used in this work have been covered elsewhere [9]. In short, a production-scale perforated pan tablet coater (Premier 200; Oystar Manesty, Merseyside, U.K.) was used for the coating trials. The diameter of this coater was 1.3 m and the diameter of each circular perforation was 3 mm. TPI (TeraView Ltd., Cambridge, U.K.) was mounted onto the side of the coater where terahertz signals were focused onto the surface of tablets inside the coating pan. Tubular baffles were additionally fitted into the coater pan to facilitate tablet bed mixing. Individual terahertz waveforms were acquired at a rate of 120 Hz throughout a 300-min coating process. The rotational speed of the coater was set as 6 r/min. The tablet geometry was biconvex (10 mm diameter, 370 mg). Two runs under the same process conditions (runs A and B) were performed. The tablet load for these two coating trials was 175 kg. In order to validate the in-line measurement data, several tablets were removed randomly from the coating pan at 60, 120, 180, 240, and 300 min for both runs, and were measured using the offline TPI. It should be noted that even though process conditions were identical, coating variations do exist between runs due to the inherent process complexities [10]. As a matter of fact, weight gain variations of coatings from runs A and B span from 16% to 20% throughout the process thus affecting product quality. In another separate coating trial (run C), tablets underwent a different coating process altogether compared to runs A and B. In particular, a batch of uncoated tablets of the same load was added into the coating process after 140 min of coating [10]. The dataset resulting from run C is used to assess network's ability to predict on a process completely different to what it was trained with.

B. Design of Neural Network

Fig. 2 shows the structure of the BLSTM network. The first layer is a sequence input layer of size 1x69, to accommodate our cropped waveform data. The BLSTM layers learn long-term dependencies between time steps with 50 hidden units. This is followed by a dropout layer with rate 0.5 and a fully connected layer. For the classification output layer, we used the cross entropy error function and the softmax activation function, which is widely used for 1 of K classification [45]. The softmax function normalizes the network outputs to between zero and one. To prevent the network from overfitting, a dropout layer has been added into the network, which prevents the units being too consistent with each other during the training process.

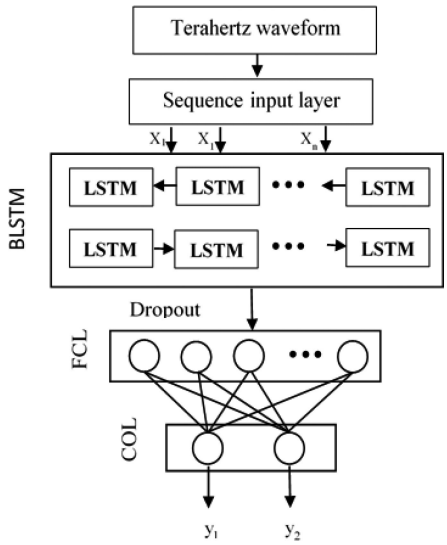

Fig. 2. LSTM network used for classification. FCL: fully connected layer; COL: classification output layer.

The BLSTM network is trained with cross-entropy loss for 20 epochs with the Adam solver, a learning rate of 0.0001 and a mini-batch size of 125.

C. Training Dataset Preparation

Because the in-line measurement dataset contains waveforms from the tablets, surface of the coating pan, perforation edge and noise waveforms, the training dataset is split into two classes: hit class (reflected from tablet) and miss class (reflected from other places) for network training. The primary training dataset is generated from the offline measurement of run A at the process endpoint. Offline waveforms from run A are used to represent the hit class, the training data of the miss class is generated from an empty coater.

Fig. 3 shows examples of measured offline terahertz time-domain waveforms of a coated tablet at different coating time points, which shows a main positive peak (at 0.0 mm) corresponding to the air/coating interface and a negative peak for the coating/core interface. To reduce the effect of noise on the network while maintaining the main features of the tablet waveforms, all waveforms were preprocessed. In particular, the front of the main peak and the rear of the negative peak of offline waveforms were discarded. Furthermore, waveforms whose main peak's position outside of $\pm$ 0.2 mm, i.e., outside the length where thickness can be accurately measured, and with an amplitude less than 0.05 were discarded from the dataset. The in-line dataset for run B is used for external validation.

D. Training Dataset Augmentation

As there were significant variations in the number of selected tablet waveforms in cross-validation, thus indicating unstable training, the training dataset is increased. In particular, we

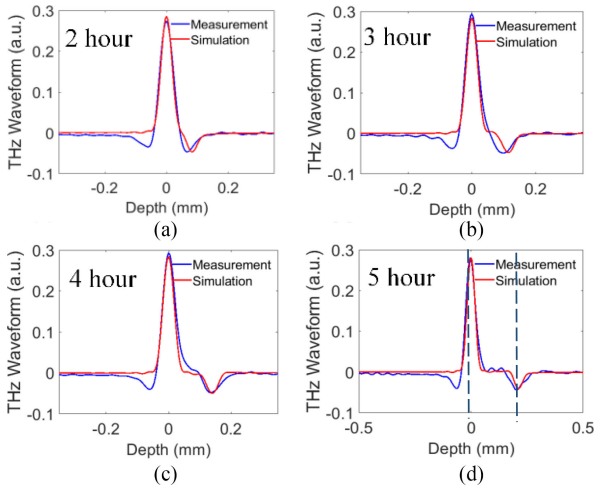

Fig. 3. Comparison of the measured offline and simulation tablet waveforms from different coating time. The dashed lines in (d) indicate the air/coating interface and coating/core interface, respectively.

simulated terahertz waveforms using Fresnel's equations, representative of reflections from a single-layer coated tablet acquired with the sample placed at the focus and normal to the terahertz sensor by a robotic arm [5]. Given the incident terahertz field E_i propagating through air, the equation for the reflection assuming plane wave approximation [46] is

$$E_r = \frac{r_{01} + r_{12}e^{-i2kd}}{1 + r_{01}r_{12}e^{-i2kd}} E_i \tag{6}$$

where d is the coating thickness, k is the wavenumber in coating material, r_{ab} is Fresnel reflection coefficient (0, 1 or 2 in this case corresponds to air, coating and core) that is defined as $r_{ab} = \frac{n_a - n_b}{n_a + n_b}$, where n_a and n_b is the refractive index of material a and b. In order to generate a large amount of tablet waveforms, three parameters: n_a (refractive index of coating); n_b (refractive index of core); and d (thickness of tablet coating) are used as variables. The averaged refractive indexes of the coating and core materials are obtained from offline measurements, which are 1.79 and 1.63, respectively. Then, we made n_a and n_b vary within $\pm$ 2% of the averaged values. According to the calculated coating thickness of the offline measurements, the values of d for each tablet waveform are randomly selected from 45–110 μm. Fig. 3 shows the examples of simulated and measured offline tablet waveforms for 2, 3, 4, and 5 h where generally good agreement can be observed. There are some differences between simulation and measurements due to the assumption that the refractive index is constant where in actual fact, it is frequency dependent [5] and settings used in deconvolution and filtering also affect the waveform profile [46]. Although the front part before the main peak (between -0.15 mm and -0.05 mm) is different, it will not affect training process as waveform pre-processing excludes this part. It should be noted that standard deconvolution and filtering settings are used here [46] and where these values deviate away the standard settings, high frequency oscillations and interface peak broadening will be introduced but with minor effect on the subsequent analysis. The distributions of data numbers used in the training (60%), validation (20%) and test (20%) according to classes are given in Table I.

TABLE I
NUMBER OF TRAINING AND TESTING SAMPLES USED IN THE OFFLINE TRAINING

Class	Classification Dataset			
	Number of training data (60%)	Number of validation data (20%)	Number of testing data (20%)	Total
Hit	90000	30000	30000	150000
Miss	90000	30000	30000	150000
Total	180000	60000	60000	300000

TABLE II
NUMBER OF TRAINING AND TESTING SAMPLES USED IN THE TRANSFER LEARNING

Class	Classification Dataset			
	Number of training data (60%)	Number of validation data (20%)	Number of testing data (20%)	Total
Hit	3600	1200	1200	6000
Miss	3600	1200	1200	6000
Total	7200	2400	2400	12000

E. Model Refinement

The training results using the augmented training dataset show that the averaged coating thicknesses increase over coating time, however, compared with offline measurements, the network tends to identify waveforms with thick profiles as tablets. This highlights that the offline measurements do not truly represent in-line waveforms. This is in part due to how the measurements are acquired. Such an idealized scenario would be impossible to maintain for measurements taken inside the rotating coating pan. The offline waveforms therefore serve only as a guide to what the in-line waveforms would look like. To accommodate this, we consider transfer learning, which can be employed to refine a trained model to be effective with a different task or data. This model is further trained as the starting point for another model for a second task [47]. To train the network to learn the differences between offline and in-line, transfer learning is used to refine the trained network. The training procedure is divided into two distinct steps: initial model training with simulated offline data based on run A and retraining the same BLSTM network weights using a small amount of in-line data of the run A selected by WSA at a learning rate of 0.0005 and a mini-batch size of 125 for 15 epochs. Table II gives the dataset distribution for the transfer learning. The training of the first step lasts about 30 min and the training time for the second step is in less than 5 min.

F. Postprocessing and Implementation

The offline training takes 30–40 min and the training of the transfer learning step is much faster at 5 min. After applying the trained network to the entire in-line dataset of run B, the

TABLE III
CONFUSION MATRIX OF BLSTM NETWORK. THE CLASSIFICATION RESULT OF THE WSA IS ASSUMED AS ACTUAL CLASS AND BLSTM IS NAMED AS PREDICTED CLASS

Predicted class / Actual Class	Hit	Miss	Evaluation parameters
Hit	6008	2330	Sensitivity: 72.06%
Miss	10897	48859	Specificity: 81.76%
			Accuracy: 80.57%

coating thickness of the selected tablet waveforms are calculated. Since the minimum resolution of TPI is 30–40 μm [9], the tablet waveforms with coating thickness less than 30 μm are manually removed from the hit class. All the analysis and algorithms developed in this study are implemented in MATLAB (MathWorks, USA).

III. RESULTS AND DISCUSSION

A. Network Performance

To evaluate the performance of the BLSTM network, we compare the tablet waveforms selected by the BLSTM network with the tablet waveforms selected by the WSA. The confusion matrix for the BLSTM network is given in Table III where results of the WSA are assumed to be correct. This assumption is valid because WSA used stringent thresholds to ensure only high quality waveforms are accepted [10]. We use accuracy, sensitivity and specificity [48] to evaluate the performance of the network, which are defined as follows

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{7}$$

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{9}$$

where true positives (TP) and true negatives (TN) are the number of the hits and miss correctly classified, while false positives (FP) and false negatives (FN) are the number of the incorrectly classified hits and miss, respectively. The classification results can then be evaluated, showing total accuracy is 80.57%, sensitivity is 72.06% and specificity is 81.76%, thus indicating that BLSTM can accurately select the correct tablet waveforms from the in-line dataset.

In order to examine these miss-classified waveforms further, Fig. 4 shows the coating thicknesses of the selected tablet waveforms at 30 s average for comparison against offline measurements where generally a good agreement against the offline TPI, weight gain measurements and WSA is observed. It should be noted that the waveforms selected by WSA and BLSTM are processed in the exact manner to determine the coating thickness [9]. As expected, a steady increase in coating

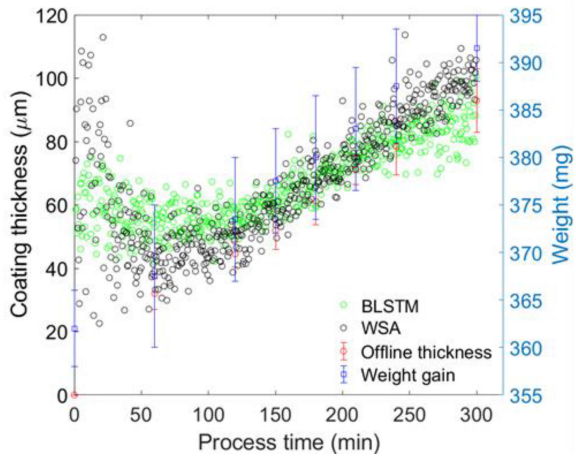

Fig. 4. 30 s average of the selected film coating thickness by BLSTM network (green circle) and WSA (black circle).

thickness is observed in-line with coating time [9]. Thicknesses from WSA and BLSTM before 80 min are not considered due to the fact that coating thicknesses are below the minimum TPI resolution of 30 μm [9]. At the ending stage of the coating process (250–300 min), selected thicknesses deviate in slope compared with the WSA, but are still consistent with offline measurements. Fig. 5 shows coating thickness distribution from 60 to 300 min, which loosely follows a Gaussian distribution with the mean increasing steadily over coating time. The intertablet coating uniformity, expressed as the coefficient of variation or the relative standard deviation of coating thickness is compared in Fig. 6 where there is a good agreement [10]. Although BLSTM is slightly lower in the early part of the process, Figs. 5 and 6 indicate that the coating uniformity evaluated by two methods is consistent. Compared to the WSA [10], BLSTM is able to select twice the number of tablet waveforms without compromising on the accuracy when compared against the offline measurement. This work therefore shows promise for in-line waveform selection.

To examine the selected waveforms further, Fig. 7 shows some examples of true and false positives. By comparing the amplitude of the second negative peak (coating/core interface) between two true positives and false positives, it can be seen that the waveforms with large second negative peak are selected as tablet by both WSA and BLSTM. However, WSA overlooks waveforms with second peak amplitude falling below certain threshold while BLSTM is less rigid and can discriminate based on additional features learnt from the dataset. Table IV gives the performance between WSA with stringent thresholds, WSA with relaxed thresholds [10] and BLSTM. R^2 here indicates the correlation of coating thickness between offline terahertz measurements made on coated tablets removed at regular intervals during the coating process and two selection methods. The data shows that BLSTM can select the same amounts of tablet waveforms, while still in good agreement with offline thickness measurements. To further assess the applicability of the algorithm for real-time processing, we examine the computational

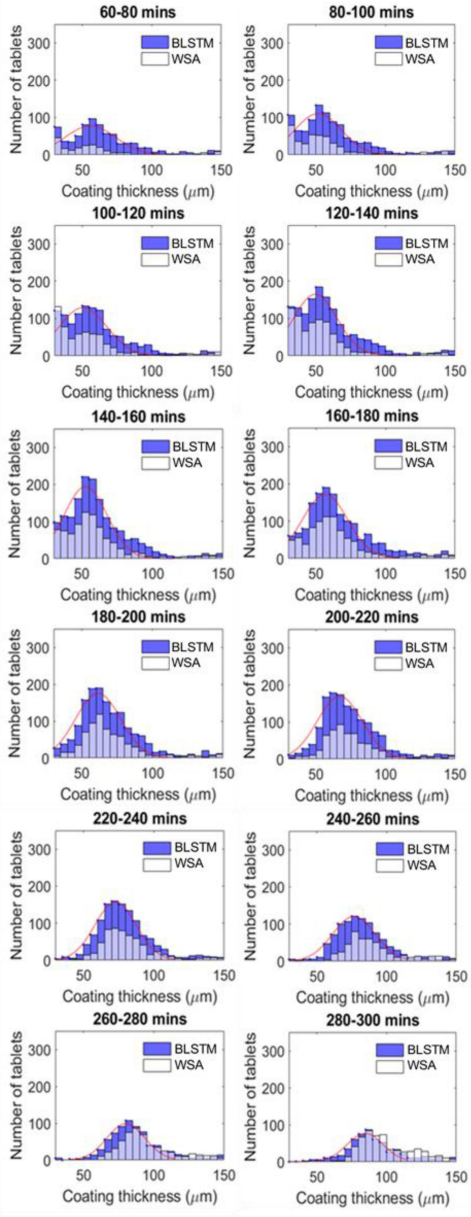

Fig. 5. Histogram of tablet coating thicknesses inside the coating pan from 60 to 300 min obtained by the WSA and BLSTM network.

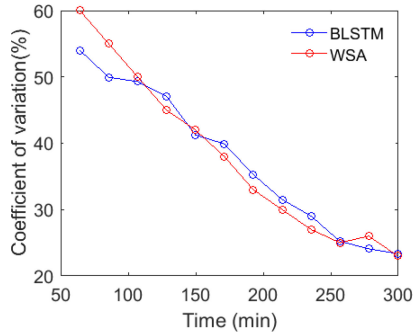

Fig. 6. Comparison of coefficient of variation between BLSTM network and WSA from [10].

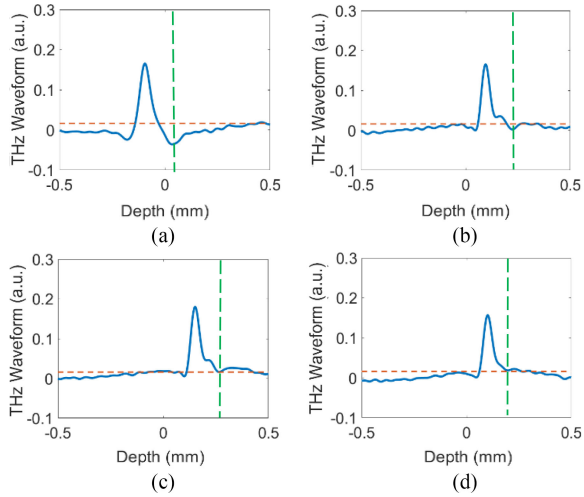

Fig. 7. Examples of two (a) and (b) true positives and (c) and (d) false positives from the in-line coating dataset. The green dashed lines indicate the positions of coating/core interface. The red dashed lines indicate the threshold value (0.016) for the amplitude criterion of the second peak. The waveforms whose amplitude is lower than 0.016 can pass this criterion in the WSA.

TABLE IV
PERFORMANCE COMPARISON OF TWO METHODS. FOR THE WSA, TWO DIFFERENT THRESHOLDS SETTING ARE SHOWN HERE ACCORDING TO [10]

	Total number hits	R^2
WSA using stringent thresholds	8200	0.91
WSA using relaxed thresholds	16660	0.80
BLSTM network	16965	0.91

time of the algorithms. In particular, a single waveform takes 1 ms to be processed in WSA while BLSTM takes 0.1 ms, an order of magnitude faster. This therefore suggests that BLSTM network could potentially be deployed for selecting terahertz waveform in real-time during the coating process.

B. Unforeseen Processes

The close agreement between coating thicknesses of BLSTM selected waveforms and offline measurements suggest BLSTM has learnt variations between coating processes with identical conditions. To further examine the ability to generalise in addition to process variations, we present the trained network with the dataset from run C [10], acquired from a process with conditions entirely artificial and vastly different to that of run A. This in turn resulted in very low classification accuracy (Sensitivity = 44.31%) especially for the already coated tablets shown in Fig. 8. Therefore, whilst the network can operate on similar processes i.e., run B, the network cannot generalize to processes with a complete change in process conditions. This can be improved if the network can be trained using data from coating runs with different process conditions supplemented by additional data of an empty production scale coater to better represent the miss class.

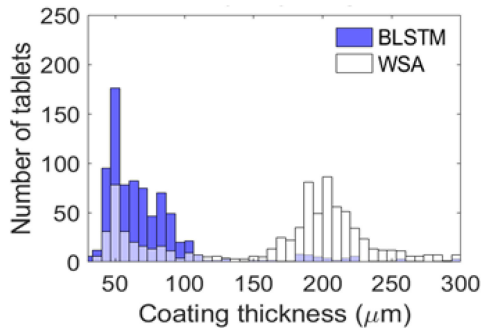

Fig. 8. Histogram of tablet coating thicknesses inside the coating pan of run C from 210 to 220 min obtained by the WSA and BLSTM network.

IV. Conclusion

In this article, we have demonstrated a BLSTM network for terahertz waveform selection as part of a pharmaceutical tablet coating process. Compared with conventional WSA, BLSTM is able to select the tablet waveforms whose features are less apparent and without compromising on accuracy when compared against the offline thickness data. The processing time is largely improved as well so that it can potentially be used for real-time monitoring of the pharmaceutical film coating process for both quality control and process investigation. However, the network has to be refined by transfer learning due to the lack of training data from an empty production-scale coater. This therefore restricts the generalization ability of the network to processes similar to what it has been trained with.

Acknowledgment

Xiaoran Li would like to thank Dr. Y. Zheng for many useful discussions.

References

[1] C. M. McGoverin, T. Rades, and K. C. Gordon, "Recent pharmaceutical applications of Raman and terahertz spectroscopies," *J. Pharmaceut. Sci.*, vol. 97, no. 11, pp. 4598–4621, 2008.

[2] C. Ricci, L. Nyadong, F. M. Fernandez, P. N. Newton, and S. G. Kazarian, "Combined Fourier-transform infrared imaging and desorption electrospray-ionization linear ion-trap mass spectrometry for analysis of counterfeit antimalarial tablets," *Anal. Bioanalytical Chem.*, vol. 387, no. 2, pp. 551–559, 2007.

[3] H. Lin, Z. Zhang, D. Markl, J. A. Zeitler, and Y. Shen, "A review of the applications of OCT for analysing pharmaceutical film coatings," *Appl. Sci.,* vol. 8, no. 12, 2018, Art. no. 2700.

[4] Y. Dong *et al.*, "Investigating intra-tablet coating uniformity with spectral-domain optical coherence tomography," *J. Pharmaceut. Sci.*, vol. 106, no. 2, pp. 546–553, 2017.

[5] Y.-C. Shen, "Terahertz pulsed spectroscopy and imaging for pharmaceutical applications: A review," *Int. J. Pharmaceutics*, vol. 417, no. 1/2, pp. 48–60, 2011.

[6] D. Alves-Lima *et al.*, "Review of terahertz pulsed imaging for pharmaceutical film coating analysis," *Sensors*, vol. 20, no. 5, 2020, Art. no. 1441.

[7] V. P. Wallace, E. MacPherson, J. A. Zeitler, and C. Reid, "Three-dimensional imaging of optically opaque materials using nonionizing terahertz radiation," *J. Opt. Soc. Amer. A*, vol. 25, no. 12, pp. 3120–3133, 2008.

[8] H. Lin, B. M. Fischer, S. P. Mickan, and D. Abbott, "Review of THz near-field methods," in *Proc. Smart Struct., Devices Syst. III*, 2007, pp. 144–151.

[9] R. K. May *et al.*, "Terahertz in-line sensor for direct coating thickness measurement of individual tablets during film coating in real-time," *J. Pharmaceut. Sci.*, vol. 100, no. 4, pp. 1535–1544, 2011.

[10] H. Lin *et al.*, "Impact of processing conditions on inter-tablet coating thickness variations measured by terahertz in-line sensing," *J. Pharmaceut. Sci.*, vol. 104, no. 8, pp. 2513–2522, 2015.

[11] H. Lin *et al.*, "Measurement of the intertablet coating uniformity of a pharmaceutical pan coating process with combined terahertz and optical coherence tomography in-line sensing," *J. Pharmaceut. Sci.*, vol. 106, no. 4, pp. 1075–1084, 2017.

[12] C. Pei *et al.*, "A quantitative comparison of in-line coating thickness distributions obtained from a pharmaceutical tablet mixing process using discrete element method and terahertz pulsed imaging," *Chem. Eng. Sci.,* vol. 192, pp. 34–45, 2018.

[13] H. Lin *et al.*, "Steps towards numerical verification of the terahertz in-line measurement of tablet mixing by means of discrete element modelling," *IET Microw., Antennas Propag.,* vol. 12, no. 11, pp. 1775–1779, 2018.

[14] M. Naftaly, *Terahertzmetrology*. London, U.K.: Artech House, 2015.

[15] J.-L. Coutaz, F. Garet, and V. P. Wallace, *Principles of Terahertz Time-Domain Spectroscopy*. Boca Raton, FL, USA: CRC Press, 2018.

[16] Y. Wang *et al.*, "Suppression of spectral interferences due to water-vapor rotational transitions in terahertz time-domain spectroscopy," *Opt. Lett.*, vol. 33, no. 12, pp. 1354–1356, 2008.

[17] D. Alves-Lima *et al.*, "Evaluation of water states in thin proton exchange membrane manufacturing using terahertz time-domain spectroscopy," *J. Membrane Sci.*, vol. 647, 2022, Art. no. 120329.

[18] Y. Shen, Y. Yin, B. Li, C. Zhao, and G. Li, "Detection of impurities in wheat using terahertz spectral imaging and convolutional neural networks," *Comput. Electron. Agriculture,* vol. 181, 2021, Art. no. 105931.

[19] X. Shen, C. R. Dietlein, E. Grossman, Z. Popovic, and F. G. Meyer, "Detection and segmentation of concealed objects in terahertz images," *IEEE Trans. Image Process.,* vol. 17, no. 12, pp. 2465–2475, Dec. 2008.

[20] M. Mikerov, J. Ornik, and M. Koch, "Removing water vapor lines from THz-TDS data using neural networks," *IEEE Trans. Terahertz Sci. Technol.,* vol. 10, no. 4, pp. 397–403, Jul. 2020.

[21] R. Liu, T. Kubiczek, P. Lehmann, D. Damyanov, and J. Balzer, "Material classification based on THz reflection mode measurements enabled by an artificial neural network," in *Proc. 45th Int. Conf. Infrared, Millimeter Terahertz Waves*, 2020, pp. 1–2.

[22] X. Yin, B. W.-H. Ng, B. M. Fischer, B. Ferguson, and D. Abbott, "Support vector machine applications in terahertz pulsed signals feature sets," *IEEE Sensors J.*, vol. 7, no. 12, pp. 1597–1608, Dec. 2007.

[23] J. Shi *et al.*, "Automatic evaluation of traumatic brain injury based on terahertz imaging with machine learning," *Opt. Exp.,* vol. 26, no. 5, pp. 6371–6381, 2018.

[24] A. Zahid *et al.*, "Machine learning driven non-invasive approach of water content estimation in living plant leaves using terahertz waves," *Plant Methods*, vol. 15, no. 1, pp. 1–13, 2019.

[25] K. Li, X. Chen, R. Zhang, and E. Pickwell-MacPherson, "Classification for glucose and lactose terahertz spectrums based on SVM and DNN methods," *IEEE Trans. Terahertz Sci. Technol.*, vol. 10, no. 6, pp. 617–623, Nov. 2020.

[26] D. Hou *et al.*, "Terahertz spectroscopic investigation of human gastric normal and tumor tissues," *Phys. Med. Biol.*, vol. 59, no. 18, 2014, Art. no. 5423.

[27] L. H. Eadie, C. B. Reid, A. J. Fitzgerald, and V. P. Wallace, "Optimizing multi-dimensional terahertz imaging analysis for colon cancer diagnosis," *Expert Syst. Appl.*, vol. 40, no. 6, pp. 2043–2050, 2013.

[28] H. J. Motlak and S. I. Hakeem, "Detection and classification of breast cancer based-on terahertz imaging technique using artificial neural network k-nearest neighbor algorithm," *Int. J. Appl. Eng. Res.*, vol. 12, no. 21, pp. 10661–10668, 2017.

[29] Y. Sun *et al.*, "Quantitative characterization of bovine serum albumin thin-films using terahertz spectroscopy and machine learning methods," *Biomed. Opt. Exp.*, vol. 9, no. 7, pp. 2917–2929, 2018.

[30] D. S. Bulgarevich, M. Talara, M. Tani, and M. Watanabe, "Machine learning for pattern and waveform recognitions in terahertz image data," *Sci. Rep.*, vol. 11, no. 1, pp. 1–8, 2021.

[31] I. Busboom *et al.*, "Towards neural network classification of terahertz measurements for determining the number of coating layers," in *Proc. 45th Int. Conf. Infrared, Millimeter Terahertz Waves*, 2020, pp. 1/2.

[32] H. Zhong, A. Redo-Sanchez, and X.-C. Zhang, "Identification and classification of chemicals using terahertz reflective spectroscopic focal-plane imaging system," *Opt. Exp.,* vol. 14, no. 20, pp. 9130–9141, 2006.

[33] B. S. Ferguson, S. Wang, H. Zhong, D. Abbott, and X.-C. Zhang, "Powder retection with T-ray imaging," in *Proc. Terahertz Mil. Secur. Appl.*, 2003, vol. 5070, pp. 7–16.
[34] F. Oliveira *et al.*, "Analysis of terahertz spectral images of explosives and bioagents using trained neural networks," in *Proc. Terahertz Mil. Secur. Appl. II*, 2004, vol. 5411, pp. 45–50.
[35] S. Zhou *et al.*, "Terahertz signal classification based on geometric algebra," *IEEE Trans. Terahertz Sci. Technol.,* vol. 6, no. 6, pp. 793–802, Nov. 2016.
[36] W. Ketterhagen *et al.*, "Modeling tablet film-coating processes," in *Predictive Modeling of Pharmaceutical Unit Operations*. New York, NY, USA: Elsevier, 2017, pp. 273–316.
[37] O. Hockwin, V. Dragomirescu, and H. Laser, "Measurements of lens transparency or its disturbances by densitometric image analysis of Scheimpflug photographs," *Graefe's Arch. Clin. Exp. Ophthalmol.*, vol. 219, no. 6, pp. 255–262, 1982.
[38] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2013, pp. 6645–6649.
[39] N. Kalchbrenner and P. Blunsom, "Recurrent continuous translation models," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2013, pp. 1700–1709.
[40] Ö. Yildirim, "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification," *Comput. Biol. Med.,* vol. 96, pp. 189–202, 2018.
[41] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.,* vol. 9, no. 8, pp. 1735–1780, 1997.
[42] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition," Feb. 2014, *arXiv:1402.1128*.
[43] S. Muzaffar and A. Afshari, "Short-term load forecasts using LSTM networks," *Energy Procedia,* vol. 158, pp. 2922–2927, 2019.
[44] S. Abdoli, P. Cardinal, and A. L. Koerich, "End-to-end environmental sound classification using a 1D convolutional neural network," *Expert Syst. Appl.,* vol. 136, pp. 252–263, 2019.
[45] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Netw.*, vol. 18, no. 5-6, pp. 602–610, 2005.
[46] D. Brock, J. A. Zeitler, A. Funke, K. Knop, and P. Kleinebudde, "Critical factors in the measurement of tablet film coatings using terahertz pulsed imaging," *J. Pharmaceut. Sci.*, vol. 102, no. 6, pp. 1813–1824, 2013.
[47] R. Asaoka *et al.*, "Using deep learning and transfer learning to accurately diagnose early-onset glaucoma from macular optical coherence tomography images," *Amer. J. Ophthalmol.*, vol. 198, pp. 136–145, 2019.
[48] A. T. Sahlol, M. Abd Elaziz, A. Tariq Jamal, R. Damaševičius, and O. Farouk Hassan, "A novel method for detection of tuberculosis in chest radiographs using artificial ecosystem-based optimisation of deep neural network features," *Symmetry*, vol. 12, no. 7, 2020, Art. no. 1146.

Xiaoran Li received the B.Eng. and Ph.D. degrees in electronic engineering from the University of Liverpool, Liverpool, U.K. in 2016 and 2020, respectively.

He is currently a Research Associate with the Department of Engineering, Lancaster University, Lancaster, U.K. His research interest involves ophthalmic optical imaging and machine learning of terahertz data in pharmaceutics.

Bryan M. Williams received the Ph.D. degree in mathematics from the University of Liverpool, Liverpool, U.K., in 2015. He is currently a Lecturer in biometrics and human identification with Lancaster University, Lancaster, U.K. Prior to that, he was a Research Associate with Universitat des Saarlandes, Saarbrücken, Germany. His research interests include developing computer vision research and application to the medical, pharmaceutical and forensic identification areas.

Robert K. May received the B.Sc. degree in experimental physics in 2002 and the Ph.D. degree in terahertz optics under J.A. Murphy in 2008 from the National University of Ireland Maynooth, Maynooth, Ireland.

In 2008, he was a Research Associate with the Terahertz Applications Group, University of Cambridge, developing terahertz imaging techniques for real-time inspection of pharmaceutical manufacturing processes. In 2010, he was with TeraView Ltd., as an Applications Scientist, where he has worked on a wide range of applications areas of terahertz imaging and now concentrates on the industrialization of coating thickness inspection of automotive paint processes.

Michael J. Evans received the Master's degree in semiconductor technology from the University of Manchester Institute for Science and Technology, Manchester, U.K., in 1990.

He is currently the Group Leader of Core Technology with TeraView Ltd., Cambridge, U.K. He holds 13 patents in semiconductor devices and physics. His current interests include commercialization of THz technology in the pharmaceutical field in particular.

Shuncong Zhong received the Ph.D. degree in mechanical engineering from The University of Manchester, Manchester, U.K., in 2007.

He had many-years industrial and academic career in Imperial College London, London, U.K., University of Liverpool, Liverpool, U.K., University of Strathclyde, Glasgow, U.K., Shanghai Jiao Tong University, Shanghai, China, and Mindray Company, Ltd., Shenzhen, China. He is currently a Chair Professor with Fuzhou University, China. He has authored 152 articles, 82 patents, two book/chapters, and one ISO standard. His research interests include on intelligent sensing and diagnosis, optics and terahertz systems, structural health monitoring, nondestructive testing and evaluation, signal/imaging processing, and pattern recognition for diagnosis and prognostics.

Dr. Zhong was elected as an IET Fellow in 2018. Due to his scientific achievements and social and economic impact to China, he was awarded Scientific Chinese "The People of the Year of 2017."

Lynn F. Gladden received the B.Sc. degree in chemical physics from the University of Bristol, Bristol, U.K., in 1982 and the Ph.D. degree in physical chemistry from the University of Cambridge, Cambridge, U.K., in 1987.

Since 2018, he has been the Executive Chair of the Engineering and Physical Sciences Research Council and is the Shell Professor of Chemical Engineering with the Department of Chemical Engineering and Biotechnology, University of Cambridge, and a Fellow of Trinity College. She currently heads up Cambridge's Magnetic Resonance Research Centre, and has a particular interest in applying magnetic resonance imaging techniques in the fields of heterogeneous catalysis and multiphase transport in porous media, as well as an interest in the terahertz region of the electromagnetic spectrum.

Dr. Gladden was appointed Fellow of Trinity College in 1999; Fellow of the Royal Academy of Engineering in 2003; a Fellow of the Royal Society in 2004; the Officer of the Order of the British Empire in 2001; and was the Commander of the Order of the British Empire in 2009.

Yaochun Shen received the B.Sc., M.Sc., and Ph.D. degrees in electronics from Nanjing University, Nanjing, China, in 1986, 1989, and 1992, respectively.

After that, he held various positions with Southeast University, China, Heidelberg University, Germany, Heriot-Watt University, U.K., Cambridge University, Cambridge, U.K., and TeraView, Ltd., Cambridge, U.K. Since 2006, he has been with the University of Liverpool, Liverpool, U.K., where he is currently a Chair Professor with the Department of Electrical Engineering and Electronics. He has authored or coauthored seven patents, five book chapters, and more than 200 conference and journal publications with an H-index of 45. His current research interests include the development of innovative terahertz and optical imaging technologies with a focus on the exploitation of their applications in industry and science.

J. Axel Zeitler received the Staatsexamen from Würzburg, Germany, in 2002, the PGCert and the Ph.D. degrees in pharmaceutical sciences under T. Rades and K. Gordon from the University of Otago, Dunedin, New Zealand, in 2003 and 2007, respectively.

He was the Chair of microstructure engineering with the Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge U.K., and is the Kenneth Denbigh Lecturer and Fellow at Gonville and Caius College. He is currently the Terahertz Application Group, Cambridge focusing on terahertz spectroscopy and imaging for material characterization, nondestructive imaging supported by complementary techniques. His work was recognized with the 2016 Royal Pharmaceutical Society Science Award and he is an Elected Fellow of the Royal Society of Chemistry.

Hungyen Lin (Senior Member, IEEE) received the Ph.D. degree in electronic engineering from the University of Adelaide, Adelaide SA, Australia, in 2012.

He is currently a Lecturer in Electronic Engineering with Lancaster University, Lancaster, U.K. Prior to that, he was a Research Associate with the Terahertz Applications Group, University of Cambridge, Cambridge, U.K. His research interests include developing terahertz and optical technologies and creating industrially relevant applications in pharmaceutical manufacturing, fuel cells and advanced materials.

Dr. Lin is an EPSRC Peer Review College Member and a Fellow of the Higher Education Academy.