

Optimising 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 (TPI) has been utilised 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 paper presents a recurrent neural network approach to optimise waveform selection. In comparison with the conventional method of WSA, our approach allows more than double the number of waveforms to be selected while maintain great agreement with off-line thickness measurement. 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 more advancement on the quality control for the pharmaceutical film coating.

Index Terms—Terahertz pulsed imaging; Machine learning; Neural network; Pharmaceutical film coating thickness; Coating uniformity.

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 fifteen years ago and has attracted interest in the pharmaceutical industry as a fast, non-destructive 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] process and consolidated against numerical modelling [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 off-line 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 waveform selected by relaxing the parameters [10] without significantly compromising on the accuracy to off-line 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 (CNN) [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 network [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 measurement geometric algebra was also 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 modelling using discrete element models are used yielding promising results [12, 13] As large volume of experimental dataset can be generated, the challenge is to investigate if 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 unit replace the traditional node in the hidden layer.

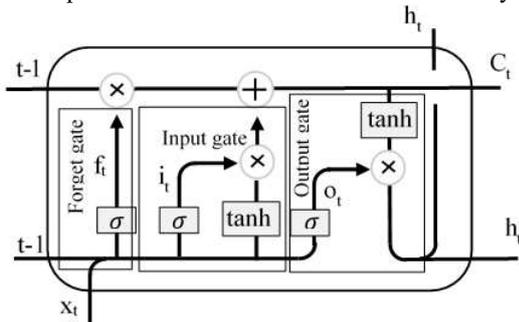


Fig. 1. Illustration of LSTM cell unit

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:

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

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 tanh activation function and the previous memory block. These signals can be expressed as

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

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

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

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

$$h_t = \tanh(C_t) \cdot o_t \quad (5)$$

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 analyse 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 off-line dataset acquisition

Details of coating experiments have been covered elsewhere [9]. In short, a production-scale perforated pan tablet coater (Premier 200; Oystar Manesty, Merseyside, UK) 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, UK) 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-minute coating process. The rotational speed of the coater was set as 6 rpm. The tablet geometry was bi-convex (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 minutes for both runs, and were measured using the off-line 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 were added into the coating process after 140 minutes of coating [10]. Dataset 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 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.

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.

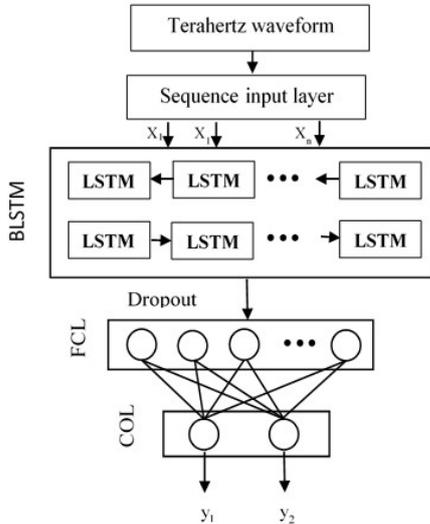


Fig. 2. LSTM network used for classification. FCL: fully connected layer; COL: classification output layer.

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 off-line measurement of run A at process endpoint. Off-line waveforms from run A are used to represent the hit class, the training data of miss class is generated from an empty coater.

Fig. 3 shows the examples of measured off-line terahertz time-domain waveform of the coated tablet at different coating timestamps, 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 pre-processed. In particular, the front of the main peak and the rear of the negative peak of off-line waveforms were discarded. Furthermore, waveforms whose main peak's position outside of ± 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. 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 an unstable training, the training dataset is increased. In particular, we 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

$$E_r = \frac{r_{01} + r_{12} e^{-izkd}}{1 + r_{01} r_{12} e^{-izkd}} E_i \quad (6)$$

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 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 coating and core materials are obtained from off-line 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 calculated coating thickness of off-line 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 off-line tablet waveforms for 2, 3, 4 and 5 hours where generally good agreement can be observed. There are some differences between simulation and measurements due to the assumption that 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 -0.15mm and -0.05mm) 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 peaks 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 provided in Table I.

TABLE I

Number of training and testing samples used in the off-line 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

E. Model refinement

The training results using augmented training dataset show that the averaged coating thicknesses increase over coating time, however, compared with off-line measurements, the network tends to identify waveforms with thick profiles as tablets. This highlights that the off-line measurements do not truly represent in-line waveforms. This is in part due to how the measurements are acquired. Such an idealised scenario would be impossible to maintain for measurements taken inside the rotating coating pan. The off-line 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 off-line and in-line, transfer learning is used to refine the trained network. The training procedure is divided into two distinct steps: 1) the initial model training with simulated off-line data based on run A; 2) retrain 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 shows the dataset distribution for the transfer learning. The training of the first step lasts about 30 minutes and the training time for the second step is in less than 5 min.

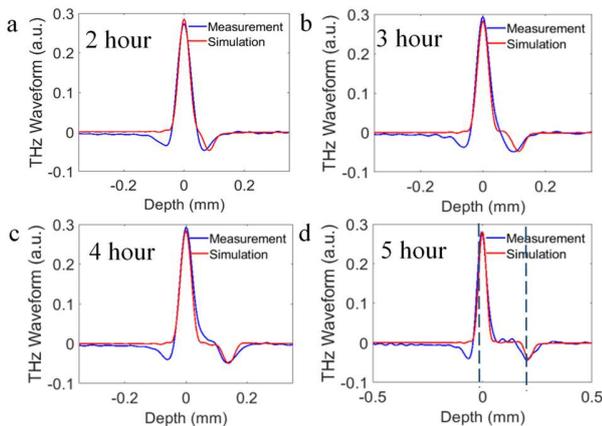


Fig. 3. Comparison of the measured off-line and simulation tablet waveforms from different coating time. The dashed lines in (d) indicate the air/coating interface and coating/core interface, respectively.

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

F. Post-processing and implementation

The off-line training takes 30-40 minutes and the training of transfer learning is much faster that is 5 minutes. After applying the trained network to the entire in-line dataset of run B, the 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 shown 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 network, which is defined as follows:

$$\text{accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

$$\text{sensitivity} = \frac{TP}{TP+F} \quad (8)$$

$$\text{specificity} = \frac{TN}{TN+FP} \quad (9)$$

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

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.

Actual Class \ Predicted class	Hit	Miss	Evaluation parameters
	Hit	6008	
Miss	10897	48859	Specificity: 81.76%
			Accuracy: 80.57%

In order to examine these miss classified waveforms further,

Fig. 4 shows the coating thicknesses of the selected tablet waveforms at 30s average for comparison against off-line measurements where generally a good agreement against the off-line 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 thickness is observed in-line with coating time [9]. Thicknesses from WSA and BLSTM network before 80 minute are not considered due to the fact coating thicknesses are below the minimum TPI resolution of 30 μm [9]. At the ending stage of the coating process (250-300 minutes), selected thicknesses deviate in slope compared with the WSA, but is still consistent with off-line measurement. Fig 5 shows coating thickness distribution from 60 to 300 minutes, which loosely follows Gaussian distribution with the mean increasing steadily over coating time. The inter-tablet 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 are 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 off-line measurement. This work therefore shows promise for in-line waveform selection.

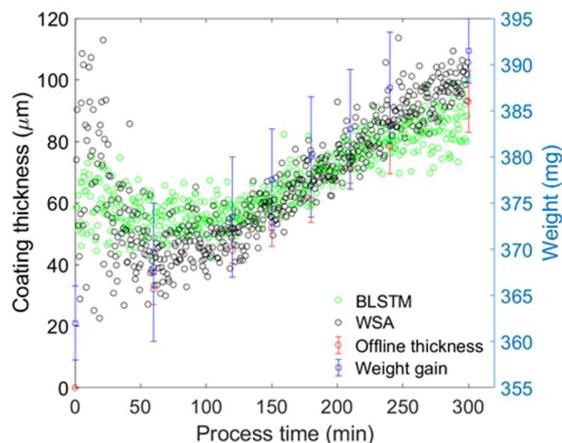


Fig. 4. 30 second average of the selected film coating thickness by BLSTM network (green circle) and WSA (black circle).

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 network. However, WSA overlooks waveforms with second peak amplitude falling below certain threshold while BLSTM network is less rigid and can discriminate based on additional features learnt from the dataset. Table IV summaries the performance between WSA with stringent thresholds, WSA with relaxed thresholds [10] and BLSTM. R^2 here indicates the correlation of coating thickness between off-line terahertz measurements made on coated tablets removed at regular intervals during the coating

process and two selection methods. The data shows that BLSTM network can select the same amounts of tablet waveforms, while still in good agreement with off-line thickness measurement. To further assess the applicability of the algorithm for real-time processing, we examine the computational 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.

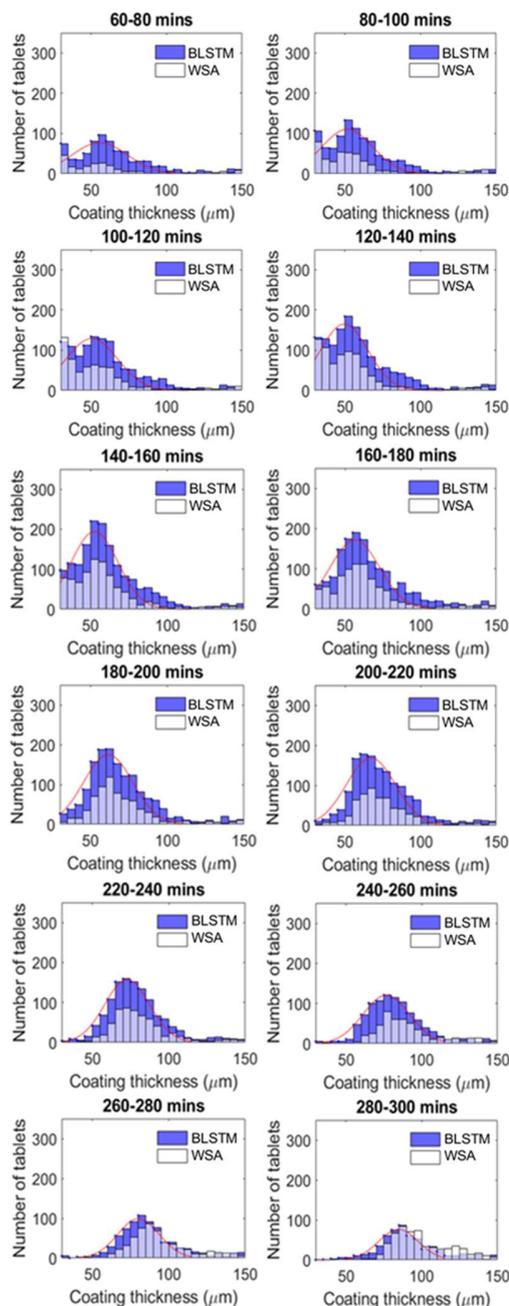


Fig. 5. Histogram of tablet coating thicknesses inside the coating pan from 60 to 300 minutes obtained by the WSA and BLSTM network.

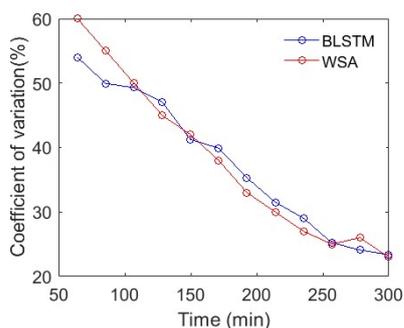


Fig. 6. Comparison of coefficient of variation between BLSTM network and WSA from [10].

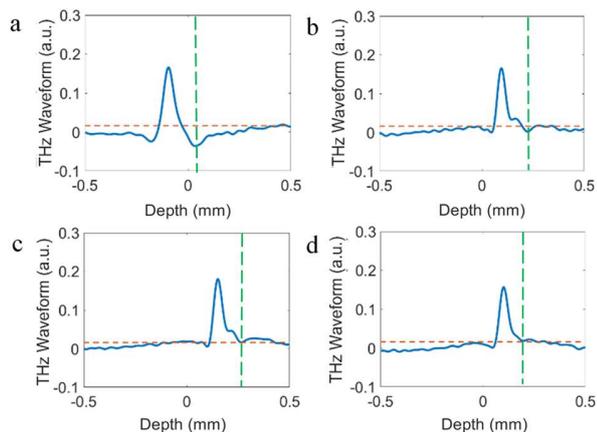


Fig. 7. Examples of two (a-b) true positives and (c-d) false positives from the in-line coating dataset. The green dashed lines indicate the positions of coating/core interface. The red dashed lines indicate the threshold value (0.016) for the amplitude criterion of the second peak. The waveforms whose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

B. Unforeseen processes

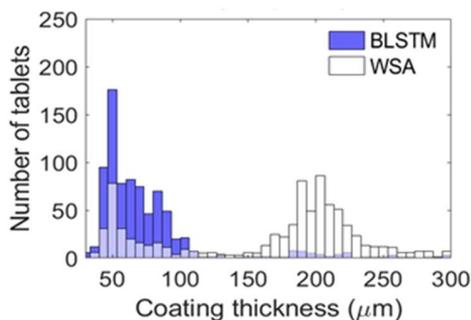


Fig. 8. Histogram of tablet coating thicknesses inside the coating pan of run C from 210 to 220 minutes obtained by the WSA and BLSTM network.

The close agreement between coating thicknesses of BLSTM selected waveforms and off-line measurements suggest BLSTM network has learnt variations between coating processes with identical conditions. To further examine the ability to generalise in addition to process variations, we presented the trained network with 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, network cannot generalise 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.

IV. CONCLUSION

In this study, we have demonstrated a BLSTM network for terahertz waveform selection as part of a pharmaceutical tablet coating process. Compared with conventional WSA, BLSTM network is able to select the tablet waveforms whose features are less apparent and without compromising on accuracy when compared against the off-line 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 empty production-scale coater. This therefore restricts the generalisation ability of the network to processes similar to what it has been trained with.

ACKNOWLEDGMENT

X.L. and H.L. acknowledge financial support from the EPSRC (Grant No. EP/R019460/1, H2FC Supergen Flexible Grant EP/P024807/1) and Australian Defence Science and Technology Group. X.L. also thank Dr Yalin Zheng for many useful discussions.

REFERENCES

- [1] C. M. McGovern, T. Rades, and K. C. Gordon, "Recent pharmaceutical applications of Raman and terahertz spectroscopies," *Journal of pharmaceutical sciences*, 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," *Analytical and bioanalytical chemistry*, 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," *Applied Sciences*, vol. 8, no. 12, p. 2700, 2018.
- [4] Y. Dong, H. Lin, V. Abolghasemi, L. Gan, J. A. Zeitler, and Y.-C. Shen, "Investigating intra-tablet coating uniformity with spectral-domain optical coherence tomography," *Journal of pharmaceutical sciences*, vol. 106, no. 2, pp. 546-553, 2017.
- [5] Y.-C. Shen, "Terahertz pulsed spectroscopy and imaging for pharmaceutical applications: a review," *International Journal of 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, p. 1441, 2020.

- [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. Am. 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 *Smart Structures, Devices, and Systems III*, 2007, vol. 6414, p. 64140L: International Society for Optics and Photonics.
- [9] R. K. May *et al.*, "Terahertz in-line sensor for direct coating thickness measurement of individual tablets during film coating in real-time," *Journal of Pharmaceutical Sciences*, 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," *Journal of Pharmaceutical Sciences*, 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," *Journal of Pharmaceutical Sciences*, vol. 106, no. 4, pp. 1075-1084, 2017.
- [12] C. Pei, H. Lin, D. Markl, Y.-C. Shen, J. A. Zeitler, and J. A. Elliott, "A quantitative comparison of in-line coating thickness distributions obtained from a pharmaceutical tablet mixing process using discrete element method and terahertz pulsed imaging," *Chemical Engineering Science*, vol. 192, pp. 34-45, 2018.
- [13] H. Lin, C. Pei, D. Markl, Y. Shen, J. A. Elliott, and J. A. Zeitler, "Steps towards numerical verification of the terahertz in-line measurement of tablet mixing by means of discrete element modelling," *IET Microwaves, Antennas & Propagation*, vol. 12, no. 11, pp. 1775-1779, 2018.
- [14] M. Naftaly, *Terahertz metrology*. Artech House, 2015.
- [15] J.-L. Coutaz, F. Garet, and V. P. Wallace, *Principles of Terahertz time-domain spectroscopy*. CRC Press, 2018.
- [16] Y. Wang, Z. Zhao, Z. Chen, Y. Zhang, L. Zhang, and K. J. O. I. Kang, "Suppression of spectral interferences due to water-vapor rotational transitions in terahertz time-domain spectroscopy," 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," p. 120329, 2022.
- [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," *Computers and Electronics in Agriculture*, vol. 181, p. 105931, 2021.
- [19] X. Shen, C. R. Dietlein, E. Grossman, Z. Popovic, and F. G. Meyer, "Detection and segmentation of concealed objects in terahertz images," *IEEE transactions on Image Processing*, vol. 17, no. 12, pp. 2465-2475, 2008.
- [20] M. Mikerov, J. Ormik, and M. Koch, "Removing Water Vapor Lines From THz-TDS Data Using Neural Networks," *IEEE Transactions on Terahertz Science and Technology*, vol. 10, no. 4, pp. 397-403, 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 *2020 45th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, 2020, pp. 1-2: IEEE.
- [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 Journal*, vol. 7, no. 12, pp. 1597-1608, 2007.
- [23] J. Shi *et al.*, "Automatic evaluation of traumatic brain injury based on terahertz imaging with machine learning," *Optics express*, 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 Transactions on Terahertz Science and Technology*, vol. 10, no. 6, pp. 617-623, 2020.
- [26] D. Hou *et al.*, "Terahertz spectroscopic investigation of human gastric normal and tumor tissues," *Physics in Medicine & Biology*, vol. 59, no. 18, p. 5423, 2014.
- [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 Systems with Applications*, 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," *International Journal of Applied Engineering Research*, 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," *Biomedical optics express*, 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," *Scientific Reports*, vol. 11, no. 1, pp. 1-8, 2021.
- [31] I. Busboom, N. Rohde, S. Christmann, V. Feige, H. Haehnel, and B. Tibken, "Towards Neural Network Classification of Terahertz Measurements for Determining the Number of Coating Layers," in *2020 45th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, 2020, pp. 01-02: IEEE.
- [32] H. Zhong, A. Redo-Sanchez, and X.-C. Zhang, "Identification and classification of chemicals using terahertz reflective spectroscopic focal-plane imaging system," *Optics Express*, vol. 14, no. 20, pp. 9130-9141, 2006.
- [33] B. S. Ferguson, S. Wang, H. Zhong, D. Abbott, and X.-C. Zhang, "Powder refection with T-ray imaging," in *Terahertz for Military and Security Applications*, 2003, vol. 5070, pp. 7-16: International Society for Optics and Photonics.
- [34] F. Oliveira *et al.*, "Analysis of terahertz spectral images of explosives and bioagents using trained neural networks," in *Terahertz for Military and Security Applications II*, 2004, vol. 5411, pp. 45-50: International Society for Optics and Photonics.
- [35] S. Zhou *et al.*, "Terahertz signal classification based on geometric algebra," *IEEE Transactions on Terahertz Science and Technology*, vol. 6, no. 6, pp. 793-802, 2016.
- [36] W. Ketterhagen *et al.*, "Modeling tablet film-coating processes," in *Predictive Modeling of Pharmaceutical Unit Operations*: 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," *Graefes Archive for Clinical and Experimental Ophthalmology*, 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 *2013 IEEE international conference on acoustics, speech and signal processing*, 2013, pp. 6645-6649: Ieee.
- [39] N. Kalchbrenner and P. Blunsom, "Recurrent continuous translation models," in *Proceedings of the 2013 conference on empirical methods in natural language processing*, 2013, pp. 1700-1709.
- [40] Ö. Yildirim, "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification," *Computers in biology and medicine*, vol. 96, pp. 189-202, 2018.
- [41] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
- [42] H. Sak, A. W. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," 2014.
- [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 Systems with Applications*, vol. 136, pp. 252-263, 2019.
- [45] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural networks*, 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," *Journal of Pharmaceutical Sciences*, 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," *American journal of ophthalmology*, 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, p. 1146, 2020.

Xiaoran Li received the B.Eng. degree and PhD degree in electronic engineering from University of Liverpool, UK in 2016 and 2020, respectively.

He is currently a research associate at the Engineering Department at Lancaster University. His research interest involves ophthalmic optical imaging and machine learning of terahertz data in pharmaceuticals.

Bryan M. Williams received the Ph.D. in Mathematics from University of Liverpool, UK, in 2015. He is currently a Lecturer in Biometrics and Human Identification at Lancaster University, UK. Prior to that he worked as a Research Associate at Universitat des Saarlandes, Germany. His research is focused on developing computer vision research and application to the medical, pharmaceutical and forensic identification areas.

Robert K. May received B.Sc. in Experimental Physics in 2002 and Ph.D. in terahertz optics under J.A. Murphy in 2008, both at the National University of Ireland Maynooth. In 2008 he joined the Terahertz Applications Group at the University of Cambridge as a Research Associate developing terahertz imaging techniques for real-time inspection of pharmaceutical manufacturing processes. In 2010 he joined 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 industrialisation of coating thickness inspection of automotive paint processes.

Michael J. Evans received a Masters Degree in Semiconductor Technology from the University of Manchester Institute for Science and Technology in 1990. He is currently the group leader of Core Technology at TeraView Ltd, Cambridge, UK. He holds thirteen patents in semiconductor devices and physics. His current interests are the commercialisation of THz technology in the pharmaceutical field in particular.

Shuncong Zhong received the Ph.D. degree 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 are 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 at the University of Cambridge, Cambridge, U.K., in 1987.

Dame Prof Gladden is the Executive Chair of the Engineering and Physical Sciences Research Council since

2018 and is the Shell Professor of Chemical Engineering at the University of Cambridge in the Department of Chemical Engineering and Biotechnology, 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.

Prof. Gladden was appointed Fellow of Trinity College in 1999; Fellow of the Royal Academy of Engineering in 2003; Fellow of the Royal Society in 2004; Officer of the Order of the British Empire (OBE) in 2001; and was appointed Commander of the Order of the British Empire (CBE) 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, U.K., and TeraView, Ltd., 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 PG Cert degree and the Ph.D. degree under T. Rades and K. Gordon from the University of Otago, Dunedin, New Zealand, in 2003 and 2007, respectively.

He holds the Chair of Microstructure Engineering at Department of Chemical Engineering and Biotechnology at University of Cambridge and is the Kenneth Denbigh Lecturer and Fellow at Gonville and Caius College.

Prof Zeitler currently leads the Terahertz Application Group at Cambridge focusing on terahertz spectroscopy and imaging for material characterization, non-destructive imaging supported by complementary techniques. His work was recognised with the 2016 Royal Pharmaceutical Society Science Award and he is an elected Fellow of the Royal Society of Chemistry.

Hungyen Lin received the Ph.D. degree in Electronic Engineering from University of Adelaide, Australia, in 2012.

He is currently a Lecturer in Electronic Engineering at Lancaster University, UK. Prior to that he worked as a Research Associate at the Terahertz Applications Group, University of Cambridge. His research interests is focused on developing terahertz and optical technologies and creating industrially relevant applications in pharmaceutical manufacturing, fuel cells and advanced materials.

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