Powder Technology 403 (2022) 117392

Contents lists available at ScienceDirect

Powder Technology

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

Deep neural network for generalizing and forecasting on-demand drying kinetics of droplet solutions

Oluwafemi Ayodele George [a,*], Aditya Putranto [b], Jie Xiao [c], Patrick Shola Olayiwola [a], Xiao Dong Chen [c], John Ogbemhe [a], Teminijesu Jesufemi Akinyemi [a], Abdolreza Kharaghani [d,*]

[a] Department of Systems Engineering, Faculty of Engineering, University of Lagos, Akoka, Nigeria
[b] Discipline of Chemical Engineering, School of Engineering, Monash University Malaysia, Bandar Sunway, Selangor 47500, Malaysia
[c] Suzhou Key Laboratory of Green Chemical Engineering, School of Chemical and Environmental Engineering, College of Chemistry, Chemical Engineering and Materials Science, Soochow University, Jiangsu 215123, PR China
[d] Thermal Process Engineering, Otto von Guericke University Magdeburg, P.O. 4120, 39016 Magdeburg, Germany

ARTICLE INFO

Article history:
Received 10 January 2022
Received in revised form 21 March 2022
Accepted 11 April 2022
Available online 15 April 2022

Keywords:
On-demand-to-produce powders
Long short-term memory
Drying kinetics
Single droplet drying
Artificial neural network
Generalized model

ABSTRACT

Recent trend in consumer preferences toward healthy and on-demand-to-make items has mandated food manufacturers to seek more efficient production processes in order to make their businesses sustainable. For powder producers in particular, the business models are drifting from 'ready-made' toward on-the-go items, thus requiring an urgent attention. Here, we present a long short-term memory (LSTM) model for forecasting drying kinetics histories of lactose, low-fat and high fat milk droplet solutions; covering a wide gas-and-material spectrum, utilizing only their initial conditions. Within the examined range, results show that the forecast lactose temperature and mass drying curves have accuracies within 0.9987 and 0.9841 respectively; that improve in the order of training-material-blend/training-data-rows as thus: lactose/3612, lactose-fat/4515, lactose-protein/5317 and lactose-fat-protein/6220. This indicates the impact of data size on the model accuracy and generalization of the trained LSTM network. For low/high fat milk (20–30% total solid), accuracy margins are within 0.0179 and 0.0655 in the range of 1–8 combined test samples. Beyond 8 samples up to 20 combined scenarios, accuracies are capped at 0.9749 for temperature and 0.9000 for mass profiles. Since only the initial conditions are required by the developed model to provide forecasting, this removes cost and time barriers inherent in traditional approaches during new product launch. Future application of deep learning models would integrate the presented LSTM network to consider actual characteristics of material mixing such as colour, texture and taste.

1. Introduction

By natural instinct, humans understand that foods will soon start to smell and attract microbes if they are wet or left open for a long time. Therefore, ancient man had to improvise by harnessing the use of nature (i.e., ice and sun), salt and fire for preserving meat, fish and other forms of food items [1–3]. Consequent upon the advent of electricity and advancement in drying technology, the narrative soon changed from what used to be the usual practice to automated drying. Although the old methods of drying are still much alive, their scale of applications is limited and are unable to cope with the expanding food stuffs that must be preserved for the consumption needs of the increasing population. Nowadays, drying techniques such as freeze drying, spray drying, super-critical drying, and drum drying are frequently employed to cater for several tons of food dehydration annually. Detailed information about each drying method can be found elsewhere [4–7]. Among the drying techniques, spray, freeze and drum drying appear to be the most prominent due to a high demand for powders and higher turnover in ton of dry items per annum compared to other drying techniques. Milk, fruits and vegetables are in the top chart of powders in high demand, with the global market share of about USD 31 billion worth of milk powders in 2020. This is expected to reach USD 42 billion by 2025, growing at a compound annual growth rate (CAGR) of 4.25% [8]. Similarly, the global market value of fruit powders is expected at USD 83 billion by the end of 2030 growing at a CAGR of ~7% [9]. The increased awareness for healthy products coupled with changing consumer preferences have all made commercialization of powders a lucrative business. Despite the growing demand, manufacturers are still obligated to deliver powders that can perform well both on the line and offline. This requires series of analyses on droplet drying for correct reconciliation of the material drying kinetics to the process

* Corresponding authors.
E-mail addresses: oageorge@unilag.edu.ng (O.A. George), abdolreza.kharaghani@ovgu.de (A. Kharaghani).

https://doi.org/10.1016/j.powtec.2022.117392
0032-5910/

variables. Thus, thorough understanding of droplet drying rate curve is a prerequisite for manufacturing high-grade powders. By definition, drying rate curve is a unique identity that characterizes material responses to moisture removal and its functionalities during post drying operation. It is particularly important in the food and pharmaceutical industries of dry items as it influences the mode of storage, packaging material and quality of the final product for market acceptability. Till date, four main set-ups have been used to investigate droplet drying histories for a wide array of products, namely single droplet drying by glass suspension, acoustic levitator, freefall experiment and droplet deposited on a heated substrate. To a large extent, these apparatuses have been successful at studying droplet drying kinetics (i.e., temperature, mass and size profiles) for milk, fruits and vegetables powders or other forms of powders. The experimental data have subsequently been used in formulating drying models categorized as physics-based and machine learning (ML) models.

The physics-based models are based on the conservation laws, and are commonly employed in the analyses of droplet drying. A well detailed review on the four main classes of physics-based drying models applicable to various types of solution droplets can be found in Langrish and Kockel [10], Chen and Lin [11], Mezhericher et al. [12], Dalmaz et al. [13], Kuts et al. [14], Nešić and Vodnik [15], among others. The physics-based techniques, though reliable at capturing the engineering approach to drying in great detail, are not without their own shortcomings. In recent time, there have been emergence of new trends in consumer preferences and the demand for fast paced food powders has skyrocketed. As general awareness on health increases, there is a rise in demand for healthy foods. However, consumers demand for healthy foods is unique and of wide variation in each product constituent. Powder formulation and development for such a scenario is not practicable with the traditional powder processing method since it requires distinct drying kinetics curves which cannot be extrapolated with any known physics-based approach. As an additional burden, when a blend of two or more dissimilar materials is desired, the traditional kinetics measurements approach is constrained by time and the associated material complexities. The pressure for timely product delivery placed on powder producers and the impact on revenues necessitate the need for the adoption of intelligent digital platform that can reliably predict in real-time the drying curves of a wide range of powder formulations.

There exists a large body of studies on intelligent drying systems which have been addressed to tackle the shortcomings of the physics-based techniques [16]. Notable mention among these studies includes: (1) Jinescu and Lavrić [17] for sebacic acid powder; (2) Kwapińska and Zbiciński [18] for Maltodextrin, cacao, and detergent solutions; (3) Chegini et al. [19] for orange juice powder; (4) Mihajlovic et al. [20] for naratriptan/ maltodextrin/lactose; (5) Keshani et al. [21] for lactose solution; (6) Vieira et al. [22] for milk powder; (7) Kamiński et al. [22] for sliced potatoes, green peas and silica gel saturated with ascorbic acid. An attempt for providing a generalized model for drying was made by Hussain et al. [23]. While we acknowledge the contributions of the past authors on the subject matter; from our observation, majority of the existing findings are aimed at introducing new ML techniques for the purpose of achieving a high-fidelity model. No doubt, R^2 score is a major metric used in determining the accuracy of a model. Notwithstanding, a marginal improvement in accuracy relative to the established threshold level may not constitute monumental development in the performance or otherwise of a model. Other factors that may be beneficial to a model include reproducibility, level of generalization, data dynamics/variation, model run-time, compatibility with modern software and ease of operation, etc. In addition, an upward and significantly varying trend has been projected for the consumption of food powders in the coming years [8,9]. And it is feared that the food manufacturers may go bankrupt if they do not respond promptly to the new market dynamics. In our opinion, a method that aims to optimize powder production at the least time expense irrespective of market fluctuations would have a positive influence on revenue. Yet, there is hardly any studies till date which has sought to quicken powder processing cycle by combining different material components for the purpose of predicting in real-time the drying kinetics of entirely new products. This repeatedly poses a huge drawback, and has remained a major reason as to why food digitalization is yet a matured industrial process. For this reason, a generic droplet drying model is herein proposed. Although, a generic model can be built using regression techniques, MLP/RNN, or other ML techniques, this study introduces the use of a special RNN, called Long-Short-Term-Memory (LSTM). It is well established that RNNs are particularly suited for input-output data with significant memory such as time series, i.e., an ordered sequence of the values of a variable measured in consistent, discrete time intervals [24]. For very long time series however, RNNs usually suffer from the vanishing gradient problem when fitting the training parameters. In contrast, LSTM is distinguished for using a gating approach to greatly minimize the vanishing gradient problem and achieve good performance irrespective of the length of the time series. Hence, it has been deemed appropriate to forecast the time evolution of liquid-solid scenario occurring in sequence [25].

The powder production life-cycle for two major drying kinetics models' deployment is shown in Fig. 1, while comparison between the current work and previously available studies is summarized in Table 1. One key attribute of this initiative is that the method does not always rely on measured material drying kinetics data to provide predictions. That is, reliable and/or 'finger-tip' forecasting of materials drying curves of any milk composition or material of interest is still possible even when the actual drying kinetics data are not known. Wide material spectra/constituents of lactose, fat and protein drying kinetics data have been combined as the training materials while lactose, high-fat and low-fat milk data represent the testing materials. Several combinations of window size, neurons and hidden layers were simulated to arrive at the optimum model performance. Apart from generalizing well, the model results also show that it is reliable at forecasting material drying rate history given only the initial conditions. This confers on LSTM network the potential for use as a model for on-demand real-time powder development.

2. Materials

2.1. Data gathering and preparation

Since this work is intended to be consumer-driven, it aims at providing quick response to varying demand for food powders. Under an ideal situation, the materials inventory should be done at the manufacturers-consumers interface (refer to Stage 1 in Fig. 1). In Stage 2, drying kinetics experiments of the ordered products are performed. In this stage, fresh experiments are usually a prerequisite in the formulation of physics-based approaches, and only under a rare scenario is new experiment required to develop a ML model. For this reason, Stage 2 has been skipped in the current study. Lactose, fat, protein, skim milk and whole milk materials drying kinetics data utilized in building LSTM network in this work have been sourced from the literature. Also, due to the difficulty in obtaining material drying history curves from large-scale drying operations (e.g., spray drying), SDD experiments are usually employed to collect data, which are then modelled for industrial powder processing. Various experimental options are available to generate SDD kinetics data. The materials drying kinetics data used in our analyses were obtained from the SDD experiments conducted with a glass filament suspension. Our readers are referred to [11,26–31], where materials drying kinetics data for this work have been extracted. Two categories of data were utilized for this study namely; (1) single material component (i.e., lactose, fat and protein) data for the purpose of training and testing and (2) multi-material components (i.e., whole or high-fat milk and skim or low-fat milk) data, which served as testing only. The

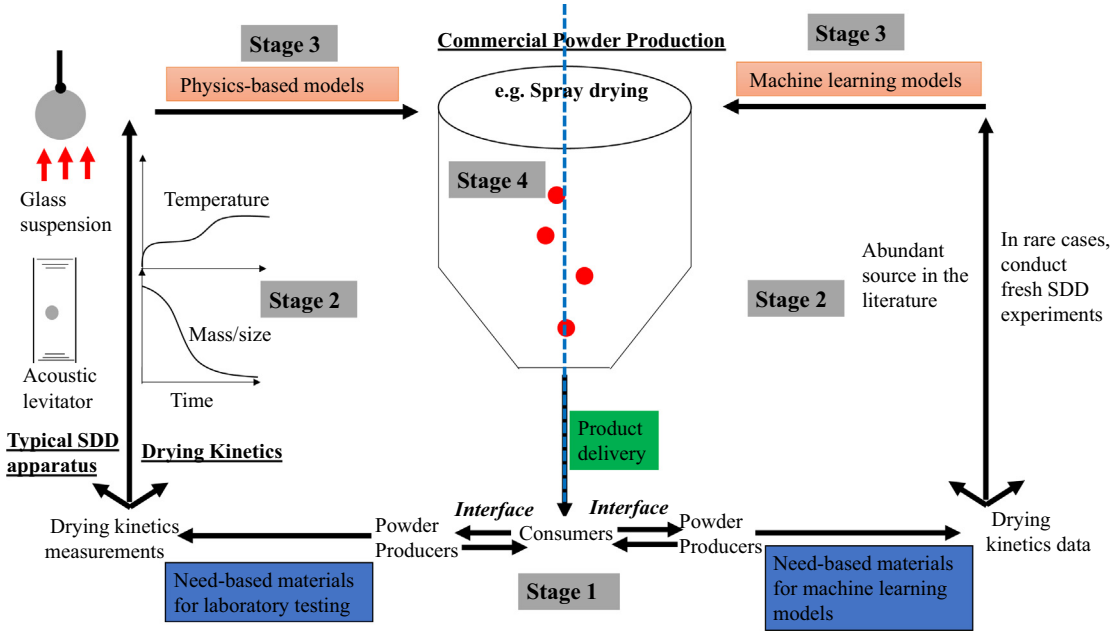


Fig. 1. Commercial powder production life-cycle with material drying kinetics models deployed via physics-based and machine learning techniques. In Stage 1, producers' interface with consumers to receive material orders. In Stage 2, drying kinetics data for ordered materials are obtained either through abundant source in the literature or in rare cases, by conducting fresh experimental analyses. Glass filament suspension, acoustic levitation, free fall method and droplet deposited on horizontal surfaces are typical experimental set-ups for single droplet drying. In Stage 3, the material drying kinetics data profiles are used to formulate relevant drying models. In Stage 4, the models are to guide in large scale production, ensuring product quality and yield are unaffected during the entire drying operation.

drying air temperature varied from 67.5 °C to 110 °C. Air velocity of 0.45–0.75 m/s, size in the range 1.24–2 mm and relative humidity of 1–2% have been considered. Others include the initial solid composition in the range 10–30% and initial droplet temperature of ~22 °C. Pressure was kept at 1 atm. Table 2 gives the summary of the three material components used for this study including their initial conditions. It should be noted however that two protein qualities; that is, whey protein concentrate (WPC) and whey protein isolate (WPI) were used for our analyses. In addition, both cream and the proteins contain traces of other materials, such as lactose, moisture and ash [30].

Table 1
Summary of the existing models in comparison with the current study. The existing methods comprise of physics-based models (e.g., CDRC, reaction engineering approach (REA), diffusion and receding interface-based models) and machine learning models (e.g., regression models and artificial neural networks such as multilayer perceptron and recurrent neural networks). The current approach is a special kind of recurrent neural network, called Long Short-Term Memory (LSTM). Material: Precursor solution e.g., Lactose, Protein or Fat solution.

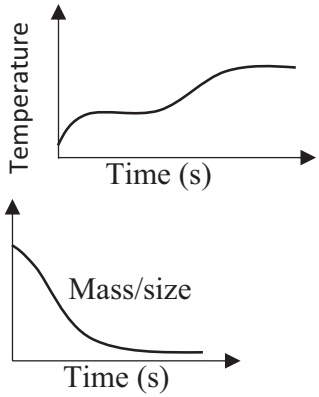
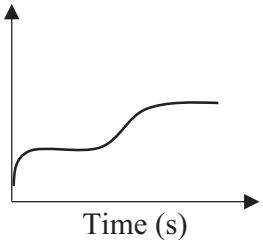
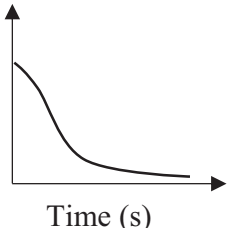
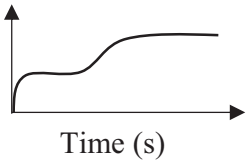
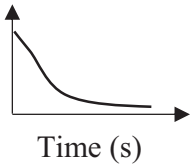
Existing models	Drying kinetics	Summary/challenges/solutions if any
		
(a) Physics-based Models CDRC, REA, diffusion and receding interface-based models	Individual for Lactose, Protein or Fat only. Drying kinetics models for material mixtures only possible via fresh experimental analyses.	Narrow validity window. Heavy reliant on experimental testing to obtain drying kinetics. Hence, physics-based models do not support on-the-go product for real-time powder processing and control.
(b) ML Models Regression models, ANN models, e.g., MLP, RNN	Individually for Lactose, Protein or Fat only	Cover large validity window. However, ML models are not yet extended to combine two or more dissimilar materials to predict drying kinetics for single or mixture products.
Current Approach/Solution Deep recurrent neural network models, e.g., LSTM	Individual product and mixture of product possible, e.g., Lactose, Protein, Fat or Lactose-Fat, Lactose-Protein and Lactose-Fat-Protein possible.	Cover large validity window. One or more individual materials can be combined to predict drying kinetics of single or mixture of products present in the training set, thus allowing for generalization. Deep recurrent neural network models can handle varying and fast-paced consumers need, and easily deplored for on-the-go/real-time powder processing and control.

Table 2
Conditions of the material components used to develop LSTM network. Input features correspond to the material compositions (i.e., lactose, protein and fat), air temperature, droplet initial size, gas velocity, gas relative humidity, time, pressure and droplet initial temperature; while droplet temperature and mass drying profiles represent the output features. Both pressure and droplet initial temperature are intentionally left out in the list of input features. Although droplet size profile is among the three-droplet drying kinetics, it has been excluded in the current study. Calculation of size evolution can be obtained using the mass-density relationship. The test cases are shown in the lower part of the table; with low/high fat milk comprising of lactose, protein and fat.

Input features							Output features	
Training cases								
Material	Air temperature (°C)	Composition (%)	Size (mm)	Velocity (m/s)	Relative humidity (%)	Time (s)	Droplet temperature (°C)	Droplet mass (mg)
Lactose	67.5–110	10, 20	1.24–1.79	0.45, 0.75	1	0–400		
Fat	67.5–110	30	1.45	0.45	1	0–400		
WPC	67.5–110	30	1.45	0.45	1	0–400		
WPI	65–85	9.3	2	0.5	2	0–400		
High-Fat milk	67.5–110	20, 30	1.24–1.79	0.45, 0.75	1	0–400		
Low-Fat milk	67.5–110	20, 30	1.24–1.79	0.45, 0.75	1	0–400		
Test cases								
Lactose	110	10	1.79	0.75	1	0–400		
High-Fat milk	67.5–110	20, 30	1.24–1.79	0.45, 0.75	1	0–400		
Low-Fat milk	67.5–110	20, 30	1.24–1.79	0.45, 0.75	1	0–400		

3. Methods

3.1. LSTM architecture

In this study, LSTM network was trained to predict one-step ahead temperature and mass drying profiles of lactose, skim milk and whole milk for different compositions of these materials and under different gas input conditions. Due to the nature of the employed drying kinetics data, i.e., constant input features for time-dependent and varied temperature and mass drying profiles of the output features, LSTM training was repeated several times until all time bounds are captured. As shown in Fig. 2, the first-time step, i.e., $LTM^{\langle 0\rangle}$ and $STM^{\langle 0\rangle}$ input vectors are essentially the same or equal zero since there are no long-term or short-term kinetics data memory information to be carried over at the initial time. However, for a specified number of window size at some advanced training times or data points, there always exist long-term and short-term dependencies. This special feature of utilizing network of layers to learn long-term dependencies while also efficient with short-term operation distinguishes LSTM from other recurrent neural networks. The experimented window size utilized for this study ranged from 1

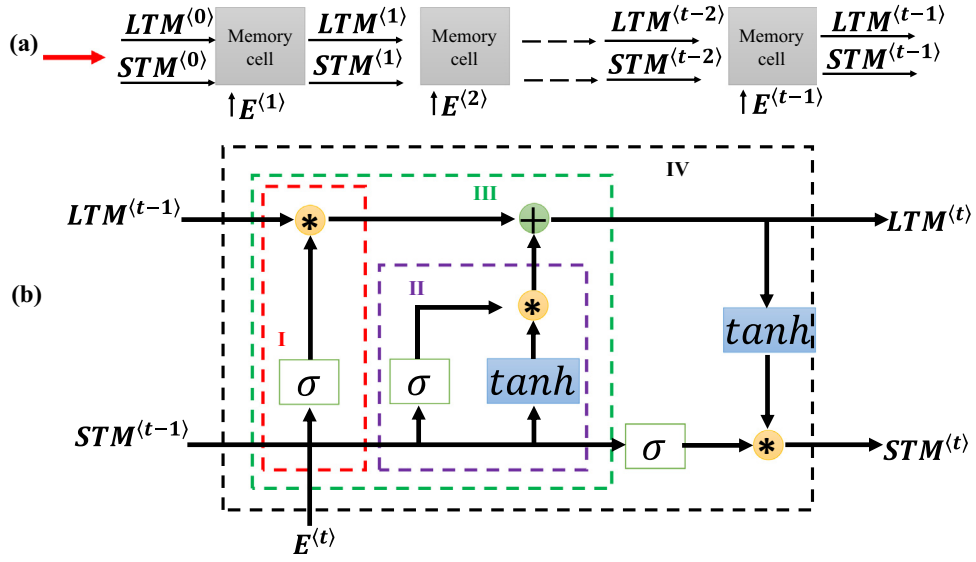


Fig. 2. LSTM architecture (a) Starting point for a network of repeated memory cell. (b) Enlarged view of the input vectors $LTM^{\langle t-1\rangle}$ and $STM^{\langle t-1\rangle}$ and current event matrix $E^{\langle t\rangle}$ representing stages I-IV. *tanh* is the hyperbolic tangent function, σ is the sigmoid activation function. The $LTM^{\langle 0\rangle}$ and $STM^{\langle 0\rangle}$ input vectors equal zero at the first-time step. ⊛ and ⊕ represent the element wise multiplication operator and matrix addition operator for long term memory output respectively.

to 51 in step of 10 (further information on window size will be provided in the results and discussion section). At each time step, the memory cell is fed with the previous cell state or long-term memory ($LTM^{\langle t-1\rangle}$) vector, previous hidden layer or short-term memory ($STM^{\langle t-1\rangle}$) vector and a matrix of current events ($E^{\langle t\rangle}$). The long-term memory is regarded as those parameter features, constant for the entire time, such as the input conditions of material compositions, air temperature, droplet initial size, gas velocity, gas relative humidity, pressure and droplet initial temperature. Short-term memory can be depicted by the output features such as changing temperature and mass drying profiles. Current events refer to the combination of input and output features of temperature and mass data points for the next LSTM input. The input vectors and the current event matrix interact through series of network gates in the memory cell to update the output vectors or input vectors for the next memory cell (see I-IV in Fig. 2). Representative materials and gas input conditions of defined window size were simultaneously passed through the forget gate (Stage I), the learn gate (Stage II), the remember gate (Stage III) and the use gate (Stage IV) to output the next data points of droplet temperature and mass conditions. The stage I of the training process is given by

$$\begin{cases} f^{\langle g\rangle} = \sigma\left(w_{f/STM} \times STM^{\langle t-1\rangle} + w_{f/E} \times E^{\langle t\rangle} + b_f\right) \\ f^{(g)}_{output} = LTM^{\langle t-1\rangle} \circledast f^{\langle g\rangle} \end{cases} \tag{1}$$

where $f^{\langle g\rangle}$ is the forget gate; b is the bias; w is the weight matrix mapping $STM^{\langle t-1\rangle}$and $E^{\langle t\rangle}$ to the sigmoid function σ. As can be seen, $\times$ and $+$ (i.e., matrix multiplication and matrix addition operators) first act on $STM^{\langle t-1\rangle}$ and $E^{\langle t\rangle}$ via a sigmoid function to produce $f^{\langle g\rangle}$. In the second expression of Eq.1, $LTM^{\langle t-1\rangle}$ is then combined with $f^{\langle g\rangle}$ using the element wise multiplication operator, $\circledast$ to produce the forget gate output. The sigmoid function assumes values between 0 and 1. A value of 0 implies that the memory cell caries all the current information to the next cell and none for $\sigma = 1$. Subscripts f, STM and E retain their usual meanings as forget, short-term memory and events, respectively.

In the learn stage (II), two functions via a separate route simultaneously act on $STM^{\langle t-1\rangle}$ and $E^{\langle t\rangle}$ to produce the learn gate results. (i) hyperbolic tangent function, tanh which introduces non-linearity to the cell with an output range value between 1 and -1 and (ii) the sigmoid activation function ($0 \le \sigma \le 1$), used for ignoring some insignificant information in the cell. The combination of *tanh* and σ functions produce the learn gate output represented by

$$\begin{cases} N^{\langle g\rangle} = \tanh\left(w_{N/STM} \times STM^{\langle t-1\rangle} + w_{N/E} \times E^{\langle t\rangle} + b_N\right) \\ i^{\langle g\rangle} = \sigma\left(w_{i/STM} \times STM^{\langle t-1\rangle} + w_{i/E} \times E^{\langle t\rangle} + b_i\right) \\ l^{\langle g\rangle}_{output} = N^{\langle g\rangle} \circledast i^{\langle g\rangle} \end{cases} \tag{2}$$

where b, w and subscripts STM and E are as defined previously. Through the element wise multiplication operator, $\circledast$ the vector matrix, $N^{\langle g\rangle}$ obtained as a result of the cell's non-linearity is combined with the ignored factor $i^{\langle g\rangle}$ generated due to sigmoid activation function to yield the learn gate output, $l^{\langle g\rangle}_{output}$. Subscripts N and i refer to the vector matrix and ignore, respectively.

The remember stage (III) is characterized by the move from the previous cell state ($LTM^{\langle t-1\rangle}$) to the current long-term memory ($LTM^{\langle t\rangle}$). In the first instance, $STM^{\langle t-1\rangle}$ and $E^{\langle t\rangle}$ candidates are passed via σ where the decision to forget irrelevant information is made. Moving further within the cell, $STM^{\langle t-1\rangle}$ and $E^{\langle t\rangle}$ candidates are fed through σ (i.e., to ignore trivial information) and tanh, which introduces non-linearity to the cell. In both scenarios, element wise multiplication operator $\circledast$ is used to obtain the output. Finally, the addition operator $\oplus$ is used to connect respective outputs, and update the long-term memory of the current time step ($LTM^{\langle t\rangle}$). The aforementioned procedure is simply achieved by the combination of Eqs. 1 and 2 given as;

$$LTM^{\langle t\rangle} = LTM^{\langle t-1\rangle} \circledast f^{\langle g\rangle} \oplus N^{\langle g\rangle} \circledast i^{\langle g\rangle} \tag{3}$$

where $\oplus$ denotes the matrix addition operator for long term memory output only.

To arrive at the use stage (IV) or the final stage, previous stage procedures are repeated, and with the addition of an extra tangent and sigmoid activation functions. In order to avoid repetition, Eq. 3 is taken as the output representation of stages I-III. To cater for the additional steps, (i) Eq. 3 is passed through the hyperbolic tangent function to produce an output labelled as $U^{\langle g\rangle}$ and (ii) $STM^{\langle t-1\rangle}$ and $E^{\langle t\rangle}$ are joined and passed through the sigmoid activation function to generate another output referred to as $V^{\langle g\rangle}$. Both $U^{\langle g\rangle}$ and $V^{\langle g\rangle}$ are then operated upon by the element wise multiplication operator, $\circledast$ to give predictions of the material temperature, size and moisture content profiles represented mathematically as;

$$\begin{cases} U^{\langle g\rangle} = \tanh\left(LTM^{\langle t-1\rangle} \circledast f^{\langle g\rangle} \oplus N^{\langle g\rangle} \circledast i^{\langle g\rangle}\right) \\ V^{\langle g\rangle} = \sigma\left(w_{V/STM} \times STM^{\langle t-1\rangle} + w_{V/E} \times E^{\langle t\rangle} + b_V\right) \\ STM^{\langle t\rangle} = U^{\langle g\rangle} \circledast V^{\langle g\rangle} \end{cases} \tag{4}$$

3.2. Steps for LSTM drying model development

Having worked through the LSTM equation derivation in Section 2.2.1., it is essential to provide detailed step-by-step procedures for the prediction of materials drying kinetics profiles passing through LSTM algorithm. The process is itemized as follows:

Step 1: Data preparation

A total of 10×10,220 (i.e., 10 columns by 10,220 rows) data points were collected for lactose, protein, fat, high-fat milk and low-fat milk. The collected drying kinetics data were cleaned and divided into different training and testing arrangements. Though, not set in stones, the commonest forms of data splitting are 70:10:20 (i.e., 70% training, 10% validation and 20% testing) and 80:20 (i.e., 80% training and 20% testing if validation is discarded) rule. In this study however, data splitting was done mainly on the basis of the initial conditions rather than dividing the total data points available for the given conditions. Unlike other sequential data where the input features vary alongside the target, only the initial features defined the time varying outputs for droplet drying examined in the current study. As such, there is a higher chance for errors minimization as opposed to random data splitting since specific input features must learn to map the entire drying profiles representing the target outputs. In constructing the training dataset, the total number of data points for each material was identified. Then the training-testing rule ratio was applied. A data point (e.g., the temperature value at a certain time during a drying case) is a case. In this study, there were respectively 10×4800; 10×1560 and 10×1600 total number of lactose, high-fat milk and low-fat milk data points. In addition, 10×7100 data points were used as the learning or training cases; with the remaining data points of 10×2260 corresponding to fat and protein data used solely for the purpose of training. About 817 data points representing 17.02% was used as the testing cases for lactose. The testing cases for high-fat milk and low-fat milk correspond to 22% and 23% respectively of the total learning cases.

Step 2: Programming language and computing software/hardware

The prepared data sets were saved in Excel format and processed using software that contain libraries of Python programming language i.e., Anaconda (run on the memory of a local computer) and cloud computing platform of Google Colab. The simulation exercise for each utilized software was performed independent of the other. Although other options of programming languages such as R, Java Script/Java, C/C++, Julia, are available for use, Python offers much flexibility in terms of compatibility with several software environments.

Step 3: Seeding of the data

Because weights/biases used to initialize LSTM models are generated by randomness, it is not uncommon for such models to produce different results when run multiple times. By design, neural networks employ randomness to guarantee that they learn the function being approximated for the problem efficiently. However, there are instances when it is desirous of a model to provide the same results regardless of when it is trained as long as the same training data is used. In such a situation, seeding helps to control variation in weights each time the data is trained by using the seed function to save the state of a random function. The same random weights are returned at each run of the data, thus, ensuring results predictability and reproducibility. Seed value can be any positive integers from 0 to 9 or even non integers. In this study, a value of 123 has been used.

Step 4: Scaling/normalization of the data

In this study we scaled the entire data to avoid convergence issue due to the significantly varying order of magnitude of the input/output features. Without distorting the ranges of values, each feature of the data was scaled separately using the same scaling algorithm to ensure equal contribution to the model fitting and learn function. Several scaling techniques have been discussed elsewhere [32]. Employed for the current work is the Min-Max Scaler. This normalization algorithm works by converting the numeric values of all features (both inputs and outputs) into the range [0,1]. This is summarized as follows:

Given X_i as the data point representing each feature; X_{min} and X_{max}, the minimum and maximum data points for each feature respectively. Then, the scaled data point is given as

$$X_{scaled} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \tag{5}$$

It is noteworthy that to achieve the objective of data scaling (i.e., ease of model convergence), Eq. (5) should be applied separately to scale each data feature making up the training/testing data.

Step 5: Training/testing the data and performance metric

LSTM models are trained by fitting the weights and biases to the network model (usually by iteration), until the discrepancies between the actual and the model (errors) values are minimized. Eqs. 1–4 are solved in successive iteration through a forward pass, followed by a loss function which compares the difference in values between the experimental and LSTM results. A backward pass process, thought of to be the opposite of the forward pass is applied to estimate the contribution of each network to the error. To initialize the forward pass process, all network weights identified with the LSTM drying model are randomly selected. Applicable loss functions for LSTM models include the mean absolute error (MAE), the mean percentage error (MAPE) and the mean square error (MSE). In this study, MSE was deemed appropriate to perform error analysis because it has no outlier predictions with huge errors. Mathematically, it is represented as

$$MSE = \frac{1}{N}\sum_{t=1}^{N}\left(y^{\langle t\rangle} - \hat{y}^{\langle t\rangle}\right)^2 \tag{6}$$

where N is the number of data points, $y^{\langle t\rangle}$ and $\hat{y}^{\langle t\rangle}$ are the actual and predicted data points respectively. To minimize the errors through the backward propagation, several optimizers can be employed. The preferred choice for this study is the efficient Adam optimizer. As mentioned in Step 2, the LSTM network was implemented in Python programming language using Keras with a Tensorflow as the backend. Other important hyperparameters to complete training the LSTM model include: (i) dropout, which defines the forget gate, $f^{\langle g\rangle}$ (ii) number of neurons and layers, to fit the model (iii) batch size, defining the number of training data passing through the LSTM network per iteration, and (iv) number of epochs, which determines the number of times LSTM network is trained to minimize the loss function. Since it is usual practice to set a high number for epoch, early stopping criterion ensures LSTM training does not continue indefinitely, and terminates depending on the defined 'patience' and 'mode' for the loss function. In this study, mode ='min' and patience = 10; implying that the training should stop if the loss function has not decreased after 10 epochs have been used. Details on the number of neurons and hidden layers used will be discussed in the next section. To validate the network, sets of test data were passed through the pretrained model. Coefficient of determination, R^2, was used to compare the training and testing accuracies with the measured data. Eq. 7 is used to estimate the coefficient of determination as

$$R^2 = 1 - \frac{\sum_{t=1}^{N}\left(y^{\langle t\rangle} - \hat{y}^{\langle t\rangle}\right)^2}{\sum_{t=1}^{N}\left(\bar{y}^{\langle t\rangle} - \hat{y}^{\langle t\rangle}\right)^2} \tag{7}$$

where $\bar{y}^{\langle t\rangle}$ denotes the mean of all the actual points.

4. Results and discussion

The aim of this work is to accelerate powder production process for on-the-go consumers by providing generalized data-driven models that can describe the drying kinetics of powders. One aspect of our analyses considers a scenario where the LSTM network is modelled for powders whose drying kinetics are known a priori. This scenario is termed as prediction. In the second case, only the material and air initial conditions are required to build the drying kinetics of powders. These initial conditions correspond to a window of one LSTM memory cell used to forecast subsequent data points. In both situations, the air and droplet temperatures, the air pressure, the air velocity, the relative humidity of air, the droplet size, the material compositions (i.e., lactose, protein and fat) and time represent the input features, whereas the droplet temperature and mass are the output features. There are two possibilities for the input-output features, namely multi-input single-output and multi-input multi-output. Here, a multi-input single-output has been used, i.e., more than one input features were used to model one output of droplet temperature and mass occurring separately. As previously noted, a total of 10×10,220 lactose, protein, fat, high-fat milk and low-fat milk drying kinetics data (for each output) obtained via single droplet drying (SDD) experiments were used for our analyses. Of the total data collected, lactose, high-fat milk and low-fat milk have been used as the testing case with a testing-training ratio of 17.02%, 22% and 23%, respectively (see Table 2 for the conditions of the test cases). Although the three material components and the range of conditions used for this study cannot be considered too narrow in itself, the possibility of a better performing model when the data samples are increased makes the extension of LSTM model to large variation in material compositions even more endless. In order to achieve optimum model fits, analyses have also been extended to investigate the correct combination of window size, number of neurons and hidden layers that produces the best results. The range of values considered for this procedure is shown in Table 3.

To round up this section, situations where the training and testing data are of different material input shapes have been considered using the following case studies: (i) input shape of three material compositions representing lactose, protein and fat training data versus input shape of one depicting only lactose as the testing data, (ii) input shape of three material compositions representing lactose, protein and fat

Table 3
Range of values for window size, neurons and hidden layers used to optimize the drying scenario. The window size is in step of 10; the hidden layer is in step of 1 and the neurons is in step of 2 from 2 to 30, step of 5 from 30 to 65. For each parameter, a loop operation is performed to cover the range of values of the other two parameters.

Windows		Layers		Neurons	
1–51		1–5		2–65	
Layers	Neurons	Windows	Neurons	Windows	Layers
1–5	2–65	1–51	2–65	1–51	1–5

training data versus input shape of one depicting mixture of lactose-protein-fat as the testing data. Unlike numerous other studies where the input shapes for the training and testing data are usually the same, this study is designed to generalize droplet drying irrespective of whether the drying kinetics of a material is known apriori or not. Results for both (i) and (ii) are summarized as follows, with consideration to 'predict' and 'forecast'.

4.1. Different material inputs versus single component output

To investigate the generalizability of the LSTM drying network, three material components were used at four different runs of training to output a single material component. The training materials correspond to lactose, lactose-protein, lactose-fat and lactose-protein-fat at separate run to output lactose temperature and mass drying kinetics in each case (refer to Table 2 for the conditions of the materials). Spray dried lactose, just like protein and fat, is a useful ingredient in many food (e.g., sweetened condensed milk) and drug preparations (e.g., penicillin) [33,34]. According to a published report by international market analysis research and consulting (IMARC) [35], the global lactose powder market reached an estimated volume of 865 kt in 2020. If 2026 predictions were anywhere to go by, it is expected at 1047 kt. Since lactose demand is scattered across various needs, its concentration in the final consumers product is also bound to differ. There have been previous studies, both experimental and theoretical, aimed at investigating the drying kinetics of lactose [33,36]. However, when air and material conditions so far apart and within are desired to generate relevant drying profiles, the previous approaches seem to pose a great difficulty. Provided in Fig. 3 are the temperature and mass profiles results when drying history curves of lactose, protein and fat materials are combined as the training data, with lactose serving as the testing case. As shown in Fig. 3, the 'predict' scenario gives the best trialed window size, neurons and layers of 31,

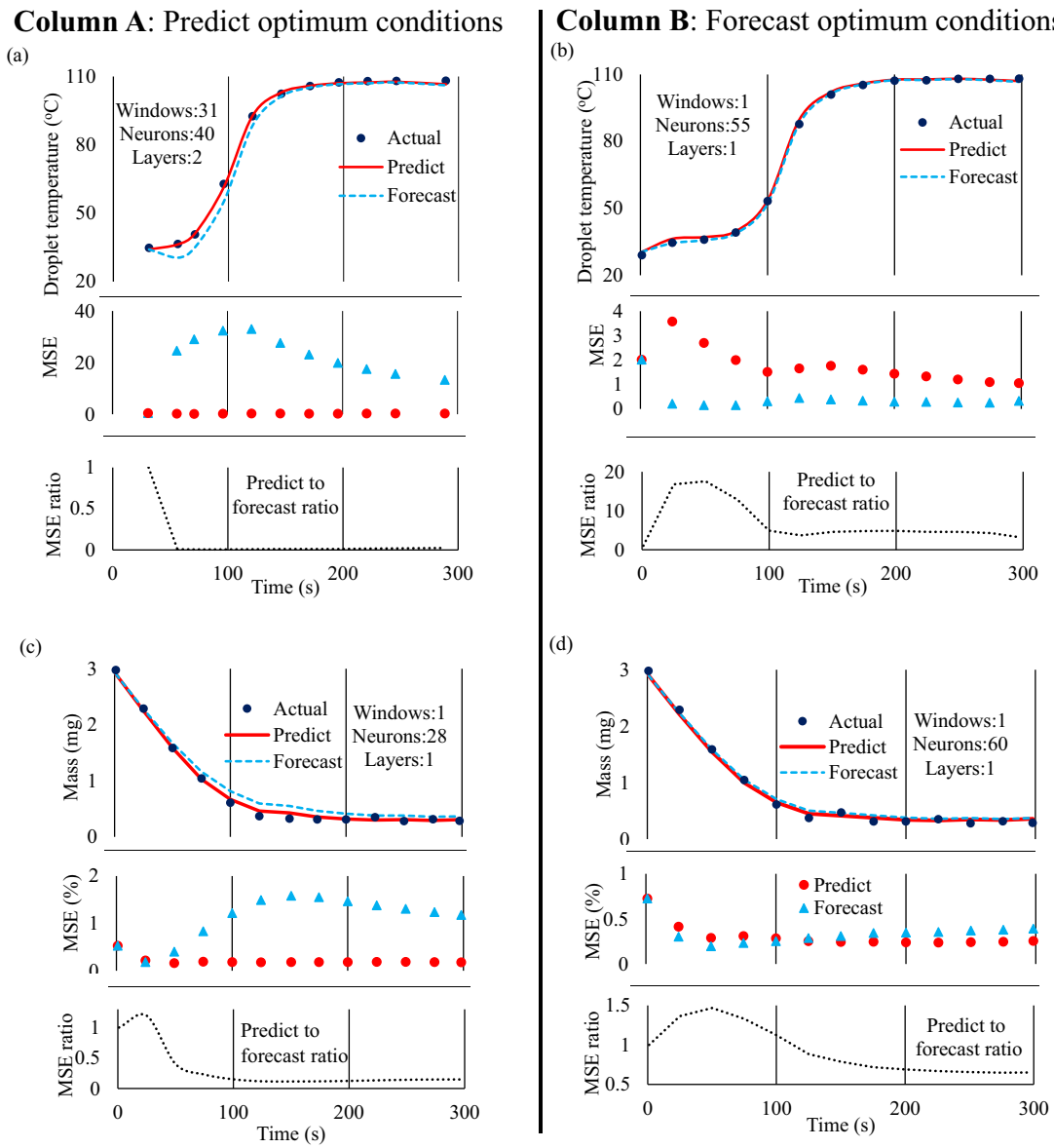


Fig. 3. Optimum trialed window size, neurons and hidden layers for the generation of lactose droplet temperature and mass profiles. (a & c) In Column A, optimum conditions for 'predict' serve as the benchmark, and the forecast plots (blue) are for comparison as shown (b & d) Column B gives the optimum conditions for the 'forecast' scenarios, representing the control curves, while the predict plots (red) are used for comparison under the same conditions. The conditions of the test case, i.e., lactose droplet correspond to 23 °C and 110 °C droplet and air temperatures respectively, 0.75 m/s air velocity, 1.79 mm size, 1% relative humidity and 10% initial solid concentration. Only a few data points have been plotted as true representatives of the entire data points. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

40, 2 and 1, 28, 1 for temperature and mass plots, respectively (Figs. 3 (a) and (c)). The optimum achievable temperature and mass drying profiles for 'forecast' corresponds to window size 1, neurons 55, layers 1 and window size 1, neurons 60, layers 1 shown in Figs. 3 (b) and (d), respectively. In each case, comparison 'predict' and 'forecast' plots for temperature and mass profiles as well as corresponding errors have been given. For the sake of emphasis, the term 'predict' is used to indicate the reliance of LSTM algorithm on the measured data at every step of the network to give the next prediction. On the other hand, 'forecast' refers to the dependence of LSTM network on the predicted value at every step to give the next data point during drying. The MSE observed for 'predict' has a mean of 0.2474 and variance of 0.0034 and a mean of 0.0020 and variance of 3.71×10^{-7} for temperature and mass profiles, respectively. Similarly, the mean 0.3115 and variance 0.0296 of MSE for forecasting temperature and mass drying curves are 0.0033 and 6.98×10^{-7}, respectively. As can be observed, the MSE mean errors and variances for temperature and mass differ significantly even though, the mean absolute change in temperature is within 0.5 °C (i.e., 0.8%) compared to mass at 0.06 (i.e., 7.3%). This is due to higher penalization of large values for errors generally attributed to MSE. To further uncover the pattern observed by 'forecast' scenario, we took the mean representing the optimum conditions for the temperature and mass curves. This is necessary seeing that for real application of drying, the drying kinetics occur simultaneously with the product, and is thus not practicable to generate the mass, size and temperature curves for the same product separately. Furthermore, among the simulated categories, 'forecast' with a window size of one is recommended for on-the-fly powder processing. The decision for this choice lies entirely on it being the closest similitude to the main idea of the current work. That is, discarding the heavy reliance on lengthy experimental trial to generate relevant drying kinetics curves when new product orders too numerous arrive at the manufacturers desk. Fig. 4 shows the 'forecast' results representing 1 window size, 58 neurons and 1 hidden layer. The simulated temperature and mass curves give close match with the experimental data with a mean MSE of 0.3179 and variance of 0.0247; and mean MSE of 0.0047 and variance 4.13×10^{-6}, respectively (Figs. 4(a) and (b)). Results for the generalizability of the test cases are shown further down in Figs. 4(c and e) and Figs. 4 (d and f) for different material sets investigated. In Figs. 4 (a and c), lactose constitutes the highest error and lowest R^2 score among the materials used for training, while in Figs. 4 (b and d), it only comes a close second after lactose-fat. A natural question that arises at this point is, why is it so given that purely lactose data were used for the training? Although it is expected that using a pure data should give the best prediction, however, a thorough examination of the training and testing conditions reveal the following: (i) air temperature is bounded by 70-90 °C for lactose training, whereas 110 °C air temperature was considered for testing. This implies that the LSTM training network for this category was stretched beyond the scope of validity in the air temperature range (ii) the addition of protein and then fat to lactose data extended the range of temperature used for the training set to 106.6 °C and (iii) the number of data rows for lactose, lactose-protein, lactose-fat and lactose-protein-fat correspond to 3612, 5317, 4515 and 6220, respectively, and is in consonance with the model accuracy. The lactose-protein-fat and lactose-protein maintain first and second positions

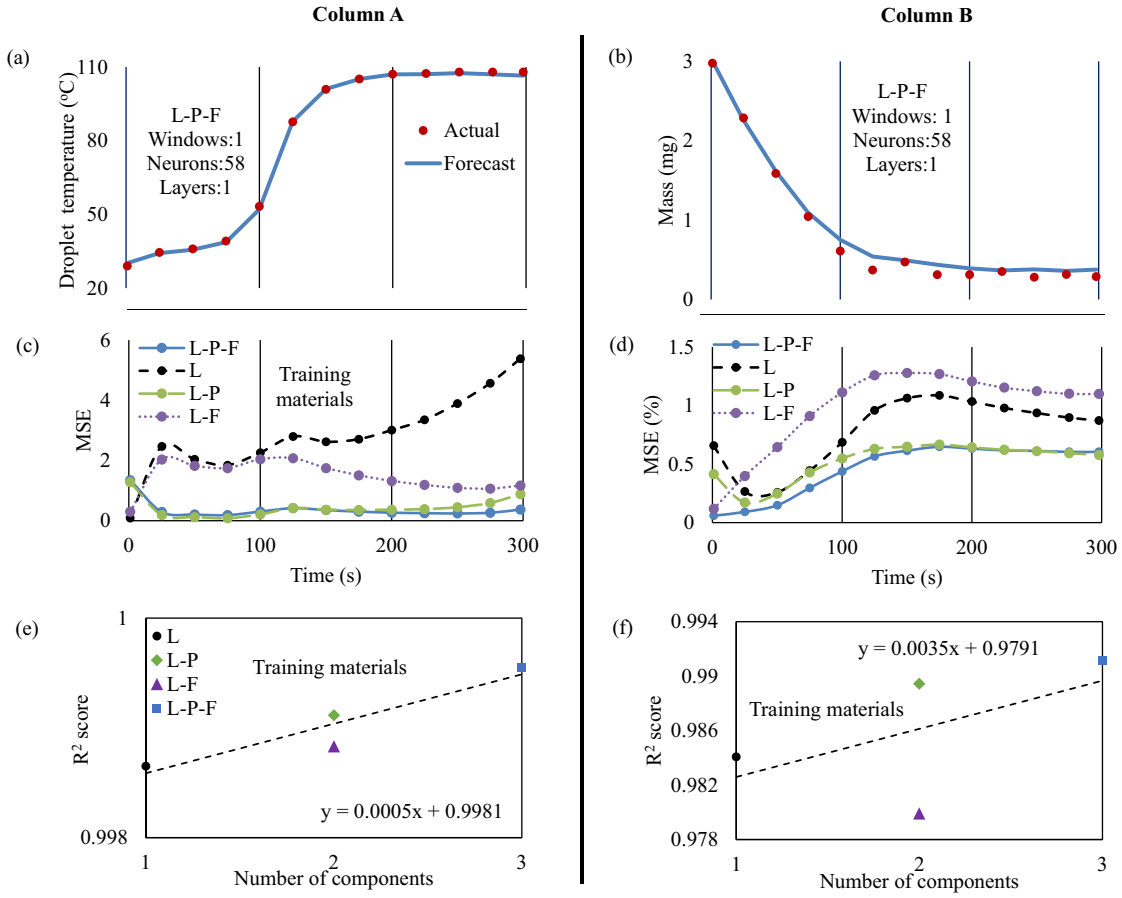


Fig. 4. Optimum trialed window size, neurons and hidden layers for the forecast scenario. (Column A) (a) Temperature versus time plot (c) Comparison error plots for L, L-P, L-F and L-P-F (e) R^2 score for different material components (Column B) (b) Mass versus time plot (d) Comparison error plots for L, L-P, L-F and L-P-F (f) R^2 score for different material components. L is lactose, L-P is lactose-protein, L-F is lactose-fat and L-P-F is lactose-protein-fat. The conditions of the test cases, i.e., lactose droplet correspond to 23 °C and 110 °C droplet and air temperatures respectively, 0.75 m/s air velocity, 1.79 mm size, 1% relative humidity and 10% initial solid concentration. Only a few data points have been plotted as true representatives of the entire data points.

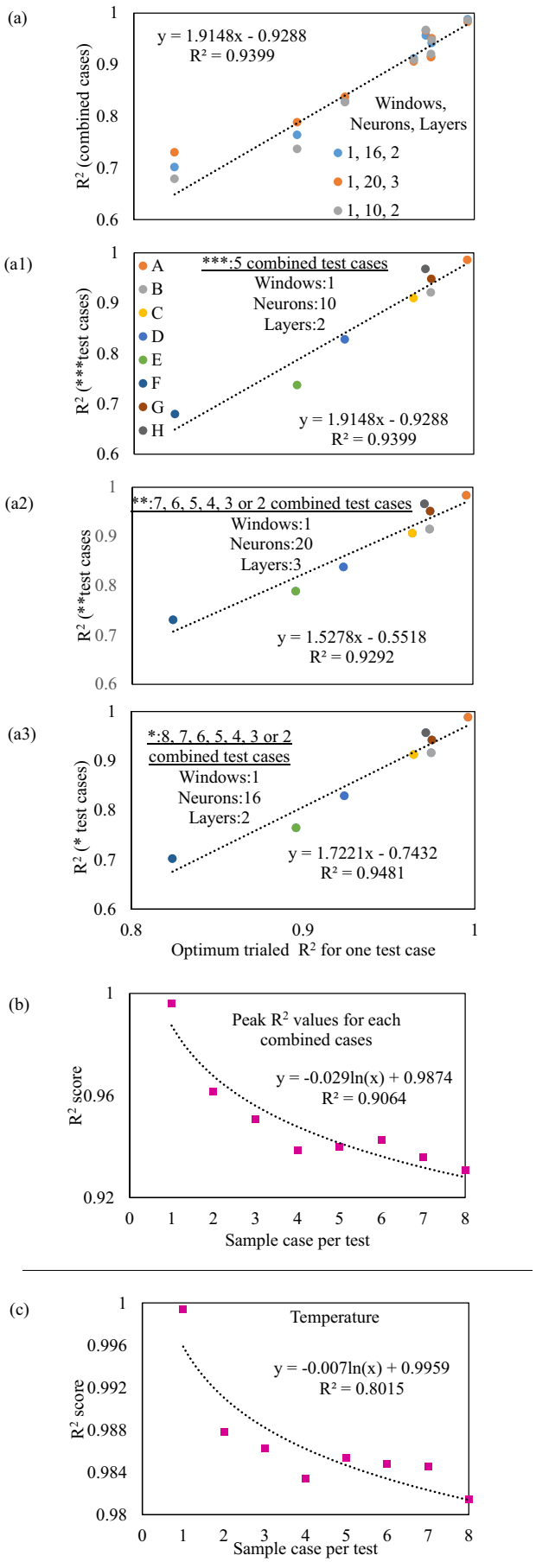
respectively with the least error and highest R^2 score both in the temperature and mass plots. Our analyses have not been able to establish yet the reason why there is a switch in positions occupied by lactose and lactose-fat in the temperature and mass plots for error and R^2 score. However, general results trend suggests that the deep learning model tends to perform better when the number of training material component and data size are increased. The simulated results can be used to guide updating of deep learning model parameters.

4.2. Different material inputs versus one output as multi-component mixture

Results given in Figs. 5 and 6 have shown the endless possibilities that can be derived from using several trained single material components to output one product; comprising of the mixture of all or some of the single material components. To explore this scenario, diverse range of whole milk and skim milk, which have been termed as the high-fat and low-fat milk for convenience, were used as the testing materials. Table 2 gives the conditions of lactose, protein and fat training materials used. Up to a total of eight, high-fat and low-fat milk materials with 20% solid concentration and temperature range of 67.5–106.6 °C have been considered for testing. To obtain individual single component from the mixture, the bulk data composition of the industrial milk powders taken from [11,27] was utilized. Final compositions of lactose, protein and fat were achieved by multiplying each composition value with the initial concentration of the high-fat and low-fat milk. Bearing in mind that whole milk and skim milk differ only in terms of component compositions, it would be counter-productive to take them as two different entities when testing to achieve the best accuracy. Thus, one LSTM network has been constructed for both material data. Furthermore, individual material droplet possesses a unique drying characteristic curve that often exhibits some discrepancies in accuracy if all the test datasets were simulated at once as opposed to testing individually. We also probed if addition of more cases to the test samples would result in huge shift in accuracy compared to testing for individual sample. The advantage of the scenarios presented above is that the same LSTM training algorithm can be deployed to generate new mixture product should there be need for additional material components. Given m as the total number of materials available for testing, and r as the number of material samples going through the LSTM network per test case. The number of material combination corresponds to $m_{\mathbb{C}r}$. For the case in context with $m = 8$ test samples and $r = 1, 2, \ldots, 7, 8$; we have $8_{\mathbb{C}1} = 8$ possibilities (one sample per test); $8_{\mathbb{C}2} = 28$ possibilities (two samples per test); ……, $8_{\mathbb{C}7} = 8$ possibilities (seven samples per test) and $8_{\mathbb{C}8} = 1$ possibility (eight samples per test). Ideally, all the representative number of material samples per test case should be simulated to determine the best accuracy. In this study however, the material combinations for $8_{\mathbb{C}1}$, $8_{\mathbb{C}7}$ and $8_{\mathbb{C}8}$ were exhausted; those from $8_{\mathbb{C}2}$ to $8_{\mathbb{C}6}$ have been randomly selected. The summary of this exercise is represented in Table 4. R^2 scores for combined test cases against optimum trialed R^2 scores for each of the eight materials are presented in Fig. 5(a). As shown, there are three possibilities for the mass profiles, i.e., 1, 16, 2; 1, 20, 3 and 1, 10, 2 corresponding to window size, neurons and hidden layers respectively. A line has been drawn to fit the data with an accuracy of 0.9399. When the three conditions were considered separately, the accuracy level differs by margin with window size of 1, neurons of 16 and hidden layers of 2 giving the best accuracy of 0.9481 (see Figs. 5 (a1, a2, and a3)). Average cumulative error, expressed as

$$error_{average\ cummulative} = \frac{\sum_{i=A}^{H}\left(\frac{\left|\hat{R}^2{}_i - R^2{}_i\right|}{R^2{}_i} \times 100\%\right)}{Total\ number\ of\ test\ samples} \tag{8}$$

was performed on each of the conditions to further buttress the choice for an optimum scenario. Estimated cumulative errors give 6.06%, 6.08% and 7.59% for 1, 16, 2; 1, 20, 3 and 1, 10, 2, respectively. In Eq. 8,



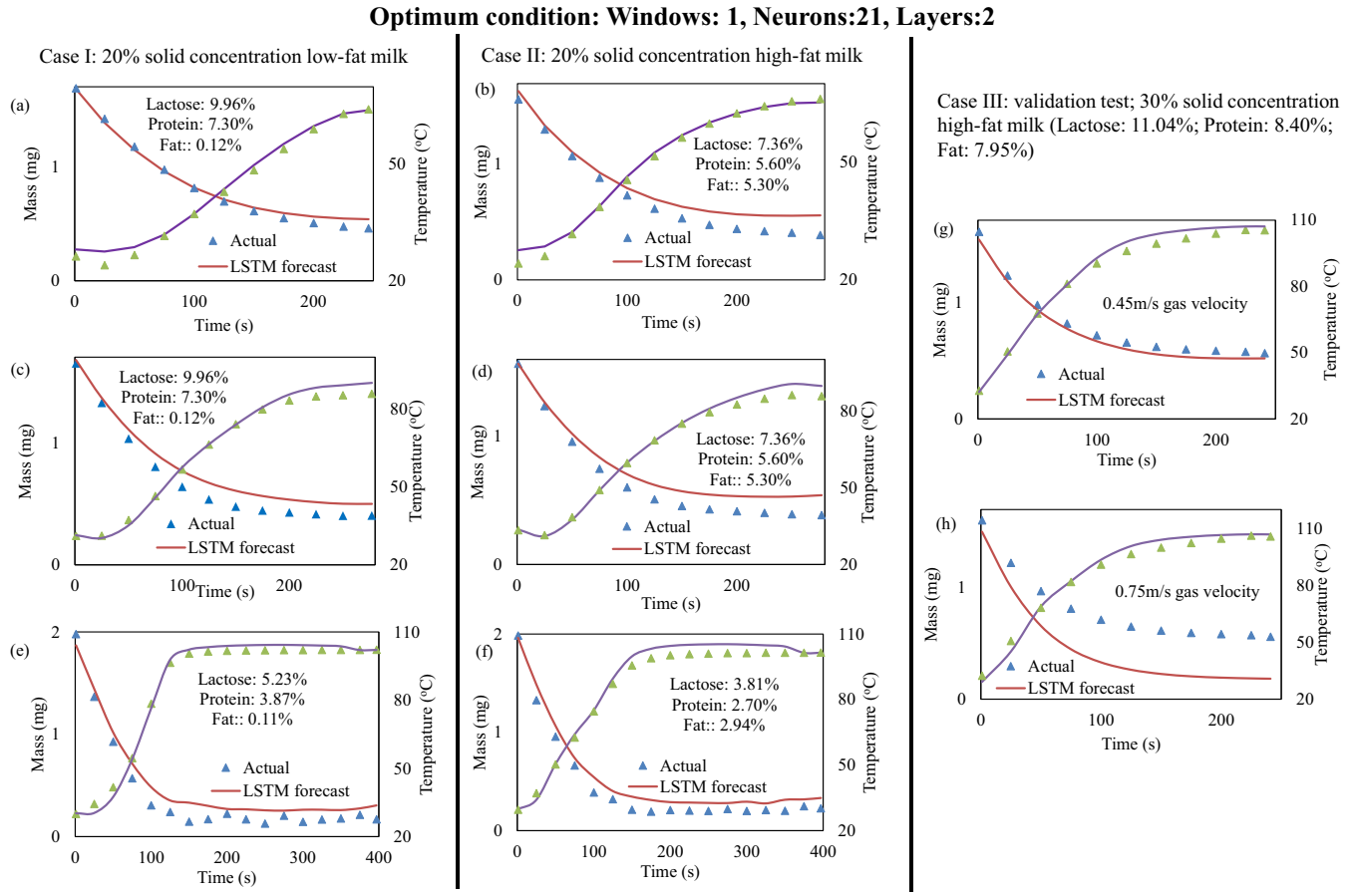


Fig. 6. Temperature and mass drying kinetics profiles for milk generated at 1 window size, 21 neurons and 2 hidden layers (**I**) 20% initial solid concentration corresponding to 9.96% lactose; 7.30% protein and 0.12% fat for plots (a) and (c); and 5.23% lactose; 3.87% protein and 0.11% fat for plot (e) (**II**) 20% initial solid concentration corresponding to 7.36% lactose; 5.60% protein and 5.30% fat for plots (b) and (d); and 3.81% lactose; 2.70% protein and 2.94% fat for plot (f) (**III**) 30% initial solid concentration corresponding to 11.04% lactose; 8.40% protein and 7.95% fat at 0.45 m/s and 0.75 m/s gas velocities for (g) and (h) respectively (test samples not part of the eight test cases used to arrive at the optimum scenario and serve as validation test). Only a few data points have been plotted as true representatives of the entire data points.

R^2 is the optimum coefficient of determination for one test case, $\widehat{R}^2$ is the realized coefficient of determination for one test case after material combination has been performed to obtain optimum window size, neurons and layers conditions. A-H denotes the identity of each sample test. Similar operation was conducted for temperature plot (with 8 possibilities) but skipped due to clustering of the data toward the same point. Instead, average cumulative error analysis served as the main criterion used, with errors ranging from 0.511% - 3.61%. For temperature profile, a window size of 1; 26 neurons and 2 layers gives the least average cumulative error at 0.511%. By gradually increasing the number of sample test per case, the R^2 score has been fitted with the number of sample test with a logarithmic function to investigate the possibility of a diminishing accuracy when more test data are considered at once. For simplicity, R^2 scores corresponding to the peak values in each sample test combination have been used. As represented in Figs. 5 (b and c), the logarithmic curve fitting functions have established that within the limit of one and eight sample tests, the margin in accuracy does not exceed 0.0179 and 0.0655 for temperature and mass profiles respectively. Beyond eight sample test cases up to 20, the margin for accuracy is within 0.0244 and 0.0956 for temperature and mass respectively. This suggests that if as many as 20 sample tests were to be passed through the trained LSTM network for testing simultaneously, the model accuracy would be not less than 0.9749 for temperature and 0.9000 for mass profiles. The average of 1, 16, 2 (mass) and 1, 26, 2 (temperature) gives the optimum realized condition (i.e.,1, 21, 2) used for the drying of milk solutions. Results of the test cases based on this condition is shown in Fig. 6 for high and low-fat milk at different initial conditions for the 'forecast' LSTM, with comparison made with the actual measurements. The mean MSE of LSTM model for temperature plots lies within 2.3866 and 5.1982; and 1.3617 and 10.6715 for low-fat and high-fat milk respectively. Similarly for mass profiles, the mean MSE is in the range $2.9382 \times 10^{-6} \leq \text{MSE} \leq 0.0130$; and $3.7552 \times 10^{-5} \leq \text{MSE} \leq 0.0126$ for low-fat and high-fat milk respectively (see Fig. 6 (I and II)). In case III (i.e., Fig. 6 (g and h)), two case scenarios are presented different from the eight observed test conditions to serve as validation cases. For the same material and initial conditions except the gas conditions, the effect of changing gas velocity from 0.45 m/s to 0.75 m/s can be observed on the temperature and mass profiles. Although no noticeable changes are observable for temperature plots, there is however a significant decline in accuracy moving from 0.45 m/s to 0.75 m/s for mass plots (Fig. 6 (g and h)). Examination of the training sets shows that the training features at 0.75 m/s correspond to low initial solid concentrations whose equilibrium mass profiles are essentially small as opposed to the training features for 0.45 m/s. This implies that 30% total

Fig. 5. R^2 scores for sample test A-H obtained from the optimum window size, neurons and layers conditions after material combination has been performed (a) plot of R^2 scores obtained from 1, 16, 2; 1, 20, 3 and 1, 10, 2; window size, neurons and layers respectively against optimum R^2 scores for materials A, B, C, D, E, F, G and H. (a1), (a2) and (a3) correspond to 1, 10, 2; 1, 20, 3 and 1, 16, 2 respectively. Plot showing highest realized R^2 score in each m_{C_r} combination versus sample case per test for (b) mass and (c) temperature.

Table 4
Summary of the R^2 scores for combined test cases against optimum trialed R^2 scores for each of the eight materials.

	Combined test samples	No of occurrences/no of trial	Optimum conditions		R^2 score optimum/R^2 score for one test sample	
			Windows neurons layers			
	2	0/4	1	20	1	A: 0.990/0.998
	3	0/3				B: 0.987/0.997
	4	1/2				C: 0.990/0.999
	5	1/4				D: 0.985/0.999
	6	1/2				E: 0.999/0.999
	7	0/8				F: 0.998/0.999
	8	1/1				G: 0.991/0.997
						H: 0.988/0.995
	2	1/4	1	26	2	A: 0.985/0.988
	3	1/3				B: 0.993/0.997
	4	1/2				C: 0.993/0.999
	5	2/4				D: 0.998/0.999
	6	0/2				E: 0.996/0.999
	7	8/8				F: 0.998/0.999
	8	0/1				G: 0.994/0.997
						H: 0.991/0.995
Temperature	2	0/4	1	24	1	A: 0.992/0.998
	3	0/3				B: 0.993/0.997
	4	0/2				C: 0.988/0.999
	5	0/4				D: 0.988/0.999
	6	1/2				E: 0.999/0.999
	7	0/8				F: 0.999/0.999
	8	0/1				G: 0.992/0.997
						H: 0.994/0.995
	2	1/4	1	16	1	A: 0.989/0.998
	3	0/3				B: 0.995/0.997
	4	0/2				C: 0.977/0.999
	5	1/4				D: 0.983/0.999
	6	0/2				E: 0.995/0.999
	7	0/8				F: 0.996/0.999
	8	0/1				G: 0.983/0.997
						H: 0.988/0.995
	2	0/4	1	20	2	A: 0.992/0.998
	3	1/3				B: 0.994/0.997
	4	0/2				C: 0.944/0.999
	5	0/4				D: 0.930/0.999
	6	0/2				E: 0.997/0.999
	7	0/8				F: 0.989/0.999
	8	0/1				G: 0.992/0.997
						H: 0.980/0.995
	2	0/4	1	22	1	A: 0.988/0.998
	3	1/3				B: 0.989/0.997
	4	0/2				C: 0.996/0.999
	5	0/4				D: 0.980/0.999
	6	0/2				E: 0.995/0.999
	7	0/8				F: 0.997/0.999
	8	0/1				G: 0.992/0.997
						H: 0.990/0.995
	2	0/4	1	10	5	A: 0.949/0.998
	3	0/3				B: 0.966/0.997
	4	1/2				C: 0.981/0.999
	5	0/4				D: 0.880/0.999
	6	0/2				E: 0.976/0.999
	7	0/8				F: 0.988/0.999
	8	0/1				G: 0.981/0.997
						H: 0.973/0.995
	2	1/4	1	26	1	A: 0.994/0.998
	3	0/3				B: 0.997/0.997
	4	0/2				C: 0.988/0.999
	5	0/4				D: 0.984/0.999
	6	0/2				E: 0.998/0.999
	7	0/8				F: 0.998/0.999
	8	0/1				G: 0.987/0.997
						H: 0.989/0.995
	2	0/4	1	10	2	A: 0.986/0.996
	3	0/3				B: 0.921/0.975
	4	0/2				C: 0.910/0.964
	5	2/4				D: 0.828/0.924
	6	0/2				E: 0.738/0.896
	7	0/8				F: 0.680/0.824
	8	0/1				G: 0.948/0.975
						H: 0.968/0.972

(continued on next page)

Table 4 (*continued*)

	Combined test samples	No of occurrences/no of trial	Optimum conditions		R^2 score optimum/R^2 score for one test sample	
			Windows neurons layers			
Mass	2	3/4	1	2	16	A: 0.989/0.996
	3	1/3				B: 0.917/0.975
	4	1/2				C: 0.913/0.964
	5	1/4				D: 0.830/0.924
	6	1/2				E: 0.765/0.896
	7	7/8				F: 0.702/0.824
	8	1/1				G: 0.943/0.975
						H: 0.957/0.972
	2	1/4	1	3	20	A: 0.984/0.996
	3	2/3				B: 0.915/0.975
	4	1/2				C: 0.907/0.964
	5	1/4				D: 0.838/0.924
	6	1/2				E: 0.789/0.896
	7	1/8				F: 0.731/0.824
	8	0/1				G: 0.952/0.975
						H: 0.966/0.972

solid concentration at 0.75 m/s is outside the validity window of the training set. Thus, passing it through the LSTM network resulted in over estimation of the mass profile. In general, a higher accuracy was observed for temperature plots compared to mass profiles, which can be attributed to the validity range of testing features such as gas velocity and component compositions for mass existing more outside the training validity range.

4.3. Potential industrial application

A reliable description of droplet drying characteristics plays a crucial role in large-scale powder production by spray drying. The knowledge as such ensures the integrity and yield of the final product stay unaltered. In this work, LSTM networks have been presented to forecast temperature and mass drying kinetics, which by extension, can be used to cover droplet size and moisture content profiles. Two case scenarios comprising single material component of lactose, protein and fat as inputs to output (i) only lactose and (ii) lactose-protein-fat mixture have been simulated. The work may be extended to train LSTM networks using the data obtained from well-controlled SDD experiments (free falling droplets, acoustically levitated droplets and horizontally deposited droplets) conducted for different materials and under a variety of process conditions.

Once trained, the LSTM networks may be used in a variety of ways. Two suggestions are given below. First, they may be used for real-time processing and development of new powders. As diverse new orders are growing on manufacturers' menu, they constantly seek solutions that bypass lengthy experimental drying kinetics measurements. For example, lycopene and riboflavin hold the path to a healthy heart. At the dictate of the market, the materials or any of their constituents may be added to lactose-protein-fat mixture, processed through the LSTM network and made available instantly as a new mixture product provided all the new material additions already form part of the trained datasets. Second, the LSTM network models may be re-configured to capture datasets of actual characteristics of material mixing (such as colour, texture and taste). It is yet unclear whether the combined material sets whose drying kinetics forecasting has been provided by LSTM would give rise to any coloration different from the one envisaged. Same holds true for texture and taste. Leveraging on the modern computer vision technology, collection of material images can serve as markers for colour and texture, and electronic tongue sensor to characterize taste. Datasets that include features of these parameters can be incorporated into the LSTM network models for real applications.

5. Conclusion

In this study, we employed a special kind of recurrent neural network, i.e., LSTM network that could be seen as a tool to accelerate powder production cycle, from conceptualization to realization. The distinguishing features of the built model include material generalization, independency of drying kinetics data (i.e., of desired output powder), and on-time tailor-made product delivery. From series of rigorous analyses conducted, the LSTM forecast model with a window size of one is herewith recommended for drying. In addition, the LSTM forecast network predicted values at every stage to give subsequent predictions, thereby obviating the need for kinetics data of intended powder. This is unlike other existing methods that suffer from lengthy processing time due to heavy reliant on drying kinetics obtained from experimental trials, and are thus unsuitable for keeping-up with the present-day trends and needs of consumers.

Going forward, new deep learning models would be sought to investigate the actual attributes of materials mixing such as colour, texture and taste. Attributes as such are important considerations in real application.

Notation

AI	Artificial intelligence
ANN	Artificial neural network
b	Bias
CDRC	Characteristic drying rate curve
$E^{<t>}$	Current event matrix
$f^{<g>}$	Forget gate
LSTM	Long short-term memory
LTM	Long-term memory
ML	Machine learning
MLP	Multilayer perceptron
MSE	Mean square error
N	Number of data points
R^2	Coefficient of determination
RNN	Recurrent neural network
SDD	Single droplet drying
STM	Short-term memory
t	Time (s)
$U^{<g>}$	LSTM output
$V^{<g>}$	LSTM output
w	Weight bias
WPC	Whey protein concentrate
WPI	Whey protein isolate
$y^{<t>}$	Actual data points
$\hat{y}^{<t>}$	Predicted data points

Credit authors statement

O. A. George conceived the project and supervised the project. O. A. George and T. J. Akinyemi performed the data analysis, machine learning and interpreted the data. O. A. George wrote the manuscript. All authors have read and approve the manuscript.

Declaration of Competing Interest

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

Acknowledgment

The authors appreciate the support of research and innovation department, University of Lagos. Also, partial financial support from project ″MASTER″ funded by the German Research Foundation (DFG) (Project number 465347282) is gratefully acknowledged.

References

[1] H. Hayashi, Drying technologies of foods - their history and future, Dry. Technol. 7 (2007) 315–369.
[2] T.P. Guinee, P.F. Fox, Salt in Cheese: Physical, Chemical and Biological Aspects, Chapman & Hall Republic of Ireland, 1993.
[3] M. Weiss-Adamson, Food in Medieval Times, Greenwood Press, Wesport, CT, 2004.
[4] K.J. Chua, S.K. Chou, Low-cost drying methods for developing countries, Trends Food Sci. Technol. 14 (2003) 519–528.
[5] A.S. Mujumdar, L.-X. Huang, X. Dong Chen, An overview of the recent advances in spray-drying, Dairy Sci. Technol. 90 (2010) 211–224.
[6] M. Mumenthaler, H. Leuenberge, Atmospheric spray-freeze drying: a suitable alternative in freeze-drying technology, Int. J. Pharm. 72 (1991) 97–110.
[7] O.A. George, X.D. Chen, J. Xiao, M. Woo, L. Che, An effective rate approach to modeling single-stage spray drying, AlCHE J. 61 (2015) 4140–4151.
[8] Global Milk Powder Market to 2027 - by Segment, Application, Regions and Company Analysis - ResearchAndMarkets.com, 2021.
[9] Global Market Study on Fruit Powder, Growing Clean Label Trend Favoring Sales Across Regions, 2021.
[10] T.A.G. Langrish, T.K. Kockel, The assessment of a characteristic drying curve for milk powder for use in computational fluid dynamics modeling, Chem. Eng. J. (2001) 69–74.
[11] X.D. Chen, S.X.Q. Lin, Air drying of milk droplet under constant and time-dependent conditions, AlCHE J. 51 (2005) 1790–1799.
[12] M. Mezhericher, A. Levy, I. Borde, Heat and mass transfer of single droplet/wet particle drying, Chem. Eng. Sci. 63 (2008) 12–23.
[13] N. Dalmaz, H.O. Ozbelge, A.N. Eraslan, Y. Uludag, Heat and mass transfer mechanisms in drying of a suspension droplet: a new computational model, Dry. Technol. 25 (2007) 391–400.
[14] P.S. Kuts, C. Strumillo, I. Zbicinski, Evaporation kinetics of single droplets containing dissolved bioaiass, Dry. Technol. 14 (1996) 2041–2060.
[15] S. Nešić, J. Vodnik, Kinetics of droplet evaporation, Chem. Eng. Sci. 46 (1991) 527–537.
[16] M. Aghbashlo, S. Hosseinpour, A.S. Mujumdar, Application of artificial neural networks (ANNs) in drying technology: a comprehensive review, Dry. Technol. 33 (2015) 1397–1462.
[17] G. Jinescu, V. Lavric, The artificial neural networks and the drying process modeling, Dry. Technol. 13 (2007) 1579–1586.
[18] M. Kwapińska, I. Zbiciński, Prediction of final product properties after cocurrent spray drying, Dry. Technol. 23 (2006) 1653–1665.
[19] G.R. Chegini, J. Khazaei, B. Ghobadian, A.M. Goudarzi, Prediction of process and product parameters in an orange juice spray dryer using artificial neural networks, J. Food Eng. 84 (2008) 534–543.
[20] T. Mihajlovic, S. Ibric, A. Mladenovic, Application of design of experiments and multilayer perceptron neural network in optimization of the spray-drying process, Dry. Technol. 29 (2011) 1638–1647.
[21] S. Keshani, W.R.W. Daud, M.W. Woo, M.Z.M. Talib, A.L. Chuah, A.R. Russly, Artificial neural network modeling of the deposition rate of lactose powder in spray dryers, Dry. Technol. 30 (2012) 386–397.
[22] G.N.A. Vieira, M. Olazar, J.T. Freire, F.B. Freire, Real-time monitoring of milk powder moisture content during drying in a spouted bed dryer using a hybrid neural soft sensor, Dry. Technol. 37 (2018) 1184–1190.
[23] M.A. Hussain, M.S. Rahman, C.W. Ng, Prediction of pores formation (porosity) in foods during drying: generic models by the use of hybrid neural network, Food Eng. 51 (2002) 239–248.
[24] T.I. Okedi, A.C. Fisher, Time series analysis and long short-term memory (LSTM) network prediction of BPV current density, Energy Environ. Sci. 14 (2021) 2408–2418.
[25] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780.
[26] S.X.Q. Lin, X.D. Chen, Changes in milk droplet diameter during drying under constant drying conditions investigated using the glass-filament method, Trans. IChemE Part C Food Bioprod. Process. 82 (2004) 213–218.
[27] S.X.Q. Lin, X.D. Chen, Improving the glass-filament method for accurate measurement of drying kinetics of liquid droplets, Trans. IChemE Part A 80 (2002).
[28] J.H. Chew, N. Fu, M.W. Woo, K. Patel, C. Selomulya, X.D. Chen, Capturing the effect of initial concentrations on the drying kinetics of high solids milk using reaction engineering approach, Dairy Sci. Technol. 93 (2013) 415–430.
[29] S.X.Q. Lin, X.D. Chen, A model for drying of an aqueous lactose droplet using the reaction engineering approach, Dry. Technol. 24 (2007) 1329–1334.
[30] S.X.Q. Lin, X.D. Chen, The reaction engineering approach to modelling the cream and whey protein concentrate droplet drying, Chem. Eng. Process. Process Intensif. 46 (2007) 437–443.
[31] M. Amdadul Haque, A. Putranto, P. Aldred, J. Chen, B. Adhikari, Drying and denaturation kinetics of whey protein isolate (WPI) during convective air drying process, Dry. Technol. 31 (2013) 1532–1544.
[32] A. Pandey, A. Jain, Comparative analysis of KNN algorithm using various normalization techniques, Int. J. Comput. Netw. Inform. Secur. 9 (2017) 36–42.
[33] K.C. Patel, X.D. Chen, Drying of aqueous lactose solutions in a single stream dryer, Food Bioprod. Process. 86 (2008) 185–197.
[34] J.G. Zadow, Lactose: properties and uses, J. Dairy Sci. 67 (1984) 2654–2679.
[35] Lactose powder market, Global Industry Trends, Share, Size, Growth, Opportunity and Forecast 2021–2026, International market analysis research and consulting, 2022.
[36] O.A. George, J. Xiao, R. Mercadé-Prieto, N. Fu, X.D. Chen, Numerical probing of suspended lactose droplet drying experiment, J. Food Eng. 254 (2019) 51–63.