A Predictive Integrated Framework Based on the Radial Basis Function for the Modelling of the Flow of Pharmaceutical Powders

Maen Alshafiee, Wafa' H. AlAlaween, Daniel Markl, Mithushan Soundaranathan, Ammar Almajaan, Karl Walton, Liam Blunt, Kofi Asare-Addo

PII: S0378-5173(19)30586-1
DOI: https://doi.org/10.1016/j.ijpharm.2019.118542
Article Number: 118542
Reference: IJP 118542

To appear in: *International Journal of Pharmaceutics*

Received Date: 25 February 2019
Revised Date: 28 June 2019
Accepted Date: 18 July 2019

Please cite this article as: M. Alshafiee, W.H. AlAlaween, D. Markl, M. Soundaranathan, A. Almajaan, K. Walton, L. Blunt, K. Asare-Addo, A Predictive Integrated Framework Based on the Radial Basis Function for the Modelling of the Flow of Pharmaceutical Powders, *International Journal of Pharmaceutics* (2019), doi: https://doi.org/10.1016/j.ijpharm.2019.118542

A Predictive Integrated Framework Based on the Radial Basis Function for the Modelling of the Flow of Pharmaceutical Powders

Maen Alshafiee[1], Wafa' H. AlAlaween[2*], Daniel Markl[3,4], Mithushan Soundaranathan[3,4], Ammar Almajaan[5] , Karl Walton[6], Liam Blunt[6], Kofi Asare-Addo[1*]

[1]*Department of Pharmacy, University of Huddersfield, Queensgate, Huddersfield, HD1 3DH*
[2]*Department of Industrial Engineering, The University of Jordan, Amman, Jordan*
[3]*Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, 161 Cathedral Street, G4 0RE Glasgow, UK*
[4]*EPSRC Centre for Innovative Manufacturing in Continuous Manufacturing and Crystallisation, University of Strathclyde, Glasgow, UK*
[5]*School of Pharmacy, Queen's University Belfast, Northern Ireland, UK*
[6]*EPSRC Future Metrology Hub, University of Huddersfield, Huddersfield, HD1 3DH, UK*

([*]Email addresses for corresponding authors: k.asare-addo@hud.ac.uk (Kofi Asare-Addo); w.alaween@ju.edu.jo (Wafa' H. AlAlaween))

Abstract:

This study presents a modelling framework to predict the flowability of various commonly used pharmaceutical powders and their blends. The flowability models were trained and validated on 86 samples including single components and binary mixtures. Two modelling paradigms based on artificial intelligence (AI) namely, a radial basis function (RBF) and an integrated network were employed to model the flowability represented by the flow function coefficient (FFC) and the bulk density (RHOB). Both approaches were utilized to map the input parameters (i.e. particle size, shape descriptors and material type) to the flow properties. The input parameters of the blends were determined from the particle size, shape and material type properties of the single components. The results clearly indicated that the integrated network outperformed the single RBF network in terms of the predictive performance and the generalization capabilities. For the integrated network, the coefficient of determination of the testing data set (not used for training the model) for FFC was $R^2 = 0.93$, reflecting an

acceptable predictive power of this model. Since the flowability of the blends can be predicted from single component size and shape descriptors, the integrated network can assist formulators in selecting excipients and their blend concentrations to improve flowability with minimal experimental effort and material resulting in the (i) minimization of the time required, (ii) exploration and examination of the design space, and (iii) minimization of material waste.

Keywords: Integrated network; Pharmaceutical powder; Powder flow; Radial basis function.

Abbreviations: RBF, radial basis function; IN, integrated network; FFC, flow function coefficient; PLS, partial least square; MCC, microcrystalline cellulose; DEM; discrete element method; PE, Polyethylene; PVC; Polyvinylchloride; PF; Phenylformaldheyde resin; SEM, scanning electron microscope; MISO; multi-input single output; RMSE, root mean square error; RHOB, bulk density

1. Introduction

Powder flow along with powder compression properties play a crucial role in the manufacturing of pharmaceutical tablets. Powder flow, in particular, is a critical issue of practical importance in those industries that primarily deal with granular materials such as the pharmaceutical industry, this being due to the fact that the flow behaviour can significantly affect the manufacturing efficiency and final product quality (e.g. dose uniformity) [1]. Poorly flowing powders can, for instance, lead to segregation during die filling before compaction [1]. Powders with good flowability characteristics (easy and free flowing powders) are therefore vital to prevent tableting issues and ensure a consistent quality of the final drug product [2]. In general, two main forces usually affect powder flow: (i) driving forces that consist of gravitation, powder mass and the angle of inclination of the powder in

relation to any bed; and (ii) dragging forces that usually include cohesion forces between similar surfaces, adhesion between unlike surfaces, water bridges and mechanical interlocking, and electrostatic forces [3-7]. Powders are, accordingly, classed to be free flowing when the driving forces are much more than the dragging ones, whereas poor powder flow occurs when the dragging forces are the primary forces in the powder bed [8].

A considerable body of research has been devoted to the understanding of particle and granular flow properties and the factors that affect these properties using various pharmaceutical powders [2-3, 8]. For instance, it has been found that particle flow properties are significantly affected by particle size and shape for both brittle and elastic pharmaceutical powders [9]. Garg *et al*. (2018) studied two commonly used brittle pharmaceutical powders, namely, Calcium Phosphate and Dicalcium Phosphate. It was shown that the Calcium Phosphate with a relatively larger particle size displayed good flow properties and less cohesiveness when compared to the Dicalcium Phosphate with a relatively large particle size [9]. Fu *et al*. (2012) investigated three grades of Lactose powders. The obtained results indicated that the powder flow properties of the three grades were significantly affected by both the particle size and shape [10]. The flow of elastic powders such as Microcrystalline Cellulose (MCC) was also sensitive to the changes in the particle size and shape [11]. Hou and Sun (2008) examined the flow of eleven grades of MCC. The results demonstrated a decrease in the powder flow rate with a decrease in the particle size even though the chemical nature and particle morphology were similar. In addition, it was found that a change in the particle morphology towards a more spherical morphology led to better flow and less cohesiveness. Furthermore, surface modifications (e.g. using silicified MCC) also led to better flow properties [11].

Modelling and predicting the powder flow properties of a material are essential in many pharmaceutical, chemical and agricultural applications. In general, modelling

paradigms can be classified as either mechanistic (or semi mechanistic) or data-driven models. For example, a data-driven model, partial least square (PLS) regression, was developed to linearly relate the particle size and shape distributions represented by multiple descriptors to the bulk powder flowability of various pharmaceutical materials [12]. Kachrimanis *et al*. (2003) implemented an artificial neural network, as a data-driven model, to map eight inputs to the powder flow rate in a circular orifice using three different pharmaceutical excipients [13]. In addition, the discrete element method (DEM), as a numerical method that is usually utilized to model/simulate the motion of a relatively large number of small particles, was utilized to simulate the flow behaviour of various powders [14-15]. Such a method allows one to model and consider the effect of equipment dynamics. Furthermore, a kinematic flow model, as a semi-mechanistic paradigm, was also established to characterize the particles flow in two-dimensional moving bed using three materials, namely, Polyethylene (PE), Polyvinylchloride (PVC) and Phenylformaldheyde (PF) resin [16]. The presented modelling paradigms (i.e. mechanistic- and data-driven models) have, in general, their limitations and strengths. On the one hand, mechanistic (or semi-mechanistic) based models can be implicitly built on some assumptions (e.g. monodisperse particle size distribution) that are not usually valid and may lead, as a result, to inaccurate results [17]. In addition, some of these models (e.g. DEM) are considered to be computationally taxing, particularly when more than billions of particles need to be considered, which is the actual case in powder flow [17]. Data-driven models, as the name indicates, rely significantly on the available data and its quality, which may include not only the number of the data points but also their distribution in the space under investigation [18]. As such, sparse and limited amount of data can decrease the performance of a data-driven models [18]. Modelling and predicting the powder flow behaviour is indeed a challenging task, this being due to (i) large number of parameters (e.g. several particle size and shape factors) that affect the powder

flow; (ii) a huge variety of excipients and APIs as well as mixtures of various excipients and API's that may possess different flow characteristics to their parent materials. Huge efforts are being towards the understanding of powders and as such their predictions. Authors such as Wang *et al.* (2016) have successfully established mathematical correlations between cohesion and the flow function coefficient. Their analysis of 41 powders using a ring shear tester enabled the proposed method that augmented shear cell data analysis and significantly reduced the complexity of the shear cell data also [19]. Leung *et al.* 2017 further studied 1130 powders to test this correlation. The authors identified a near-perfect inverse correlation between the flow function coefficient and cohesion. It was concluded that improving the flowability of pharmaceutical powder requires an alteration in the interparticlute properties rather than altering the friction properties of pharmaceutical powders [20]. A big data approached was also used by Megarry *et al*. 2019 where the authors examined 3909 historical experimental data from a shear cell. Their characterisation aided in establishing an operating space that can be used as a process flow map to guide formulators in future development [21].

In this research work, the ultimate aim is to develop a fast, cost effective and more accurate predictive model to represent the powder flow properties of various pharmaceutical powders and blends from single component data. This model can guide formulators to select excipients and their blend concentration that optimises the powder flowability. Firstly, a single radial basis function (RBF) network, as a relatively simple model, is implemented to map the particle size, shape and different blend ratios to the flow properties. The RBF network was, however, not able to describe the complex nature of powder flowability resulting in a poor prediction performance. This was addressed by developing an integrated network based on a combination of RBF models. Since the integrated network can mathematically be represented as a combination of superposition and composition functions

that are usually dense in a convex data space, it can circumvent the challenges posed by the single RBF network.

2. Materials and Methods

2.1. Materials

Three pharmaceutically-relevant powder materials having different grades were investigated in this research paper. These powder materials are MCC, Dicalcium Phosphate Dehydrate and Lactose. Eight grades of MCC having different particle size and shape were supplied by JRS Pharma (UK). These grades are VivaPur® MCC PH101, VivaPur® MCC PH102, VivaPur® MCC PH105, VivaPur® MCC PH100, VivaPur® MCC PH200, VivaPur® MCC PH302, MCC Prosolv 50 and MCC Prosolv 90. Five Dicalcium Phosphate Dihydrate grades, namely, DI-CAFOS A12, DI-CAFOS A60, DI-CAFOS A150, DI- DI-CAFOS D14 and DI-CAFOS D16, were supplied by Chemische Fabrik Budenheim KG (Germany). Seven Lactose Monohydrate grades were supplied by MEGGLE Group (Wasserburg, Germany). These are Flowlac 90, Flowlac 100, Granulac 70, Granulac 200, Inhalac 250, Inhalac 400 and Tablettose 80. In addition to being commonly used in the pharmaceutical industry, these powder materials were selected for this research work because of their different flow properties. The range of powder properties was further extended by mixing a combination of the different excipients at various ratios (3:1, 1:1, 1:3) as denoted as a note in Table 1 and also in the caption of Figure 7. Furthermore, such a range of pharmaceutical excipients and blends with different properties as utilized aids in the building of a robust flow model.

2.2. Particle Size and Morphology Analysis

Electron micrographs of all the excipient grades were obtained using a scanning electron microscope (SEM) (Quanta FEG 250), which was operated at 20kV. The samples

were mounted on a metal stub with double-sided adhesive tape and coated under vacuum with carbon in a nitrogen atmosphere. Several magnifications (i.e. ×100–500 and 1000) were used to observe the shape and surface topography of the particles.

A QICPIC instrument (Sympatec, UK) was utilized to characterise the particles in terms of size and shape. In order to ensure that the dispersing line was clean and free from contaminants, two spoonful of sand (40-100 mesh) were passed through it prior to analysis. The primary sample container containing each excipient grade was thoroughly mixed by rolling and inverted by hand as well as mixed using a spatula. Before starting the measurement, the sample, approximately 2 g, was gently inverted and agitated to evenly disperse it and, thus, reduce loss of material in the vials. The M7 lens was selected for this study, where each measurement was repeated three times. The WINDOX software was utilized to perform the statistical analysis of the obtained measurements. The following particle size and shape properties were determined and used as input parameters for the models:

- Particle size: D_{10}, D_{50}, D_{90}, $D_{4,3}$
- Aspect ratio: S_{10} and S_{50}

The particle size and shape properties of the binary blends were determined from the single component size and shape properties using a volume-based mixing rule. The physical property ($x_{\mathrm{mix},i}$) is calculated from the single component properties $x_{i,j}$ of material j and property i (more details about $x_{i,j}$ are provided in section 2.4.1):

$$x_{\mathrm{mix},i} = \sum_{j=1}^{N} f_{\varrho,j} x_{i,j} \quad (1)$$

with $N = 2$ as the number of components/materials. $f_{\varrho,i}$ is the volume based fraction considering particle true density, ϱ_i, and calculated by

$$f_{\varrho,i} = \frac{\varrho_i}{\sum_{j=1}^{N} \varrho_j} f_i. \quad (2)$$

with f_i as the weight based fraction of material i.

2.2.1. True Density Measurements

The true density of all the excipients, as detailed in section 2.1, was determined using a Micromeritics Accupyc II pycnometer 100 (Micromeritics, USA). The test was carried out using a multi-run system (10 runs) with a standard deviation of 0.005% for all the excipients.

2.3. Flow Properties Measurements

A Ring shear tester (RST-XS, Dietmar Schulze, Wolfenbuttel, Germany) was utilized to characterise the flow of the powders. The investigated powders also included a list of 66 powder blends in the ratio of 3:1, 1:1 and 1:3 of MCC, DCP and Lactose grades as detailed in section 2.1. In making the blends for FFC determinations, the appropriate powders were weighted out in their desired ratios as %w/w and blended in a Turbula mixer for 10 minutes to ensure homogeneity. These powders were then immediately analysed. The cell was over-filled with the sample powder of interest and then a spatula was used to gently smoothen the surface. The weight of the shear cell and the sample was determined and recorded using the software provided. A pre-shear stress of 4,000 Pa was applied to erase the powder history. Normal loads applied were 25%, 38%, 51%, 65% and 25% of 4,000 Pa. In order to assess the powder flowability, the flow function coefficient (FFC) can be used. Such a coefficient can be expressed as follows [12]:

$$FFC = \frac{\sigma_c}{\sigma_u} \quad (3)$$

where σ_c is the consolidation stress that compacts yield stress that makes the powder bed to flow. Yield locus is obtained by plotting the shear stress at failure as a function to the normal consolidation stress. By using Mohr cycles for each yield locus, unconfined yield stress and major principal stress are obtained [11]. The powder flowability can be classified into: (i) not flowable ($FFC < 1$); (ii) very cohesive powder ($1 \leq FCC < 2$); (iii) cohesive powder ($2 \leq FCC < 4$); (iv) easy flowing powder when the FFC value is in the range of ($4 \leq FCC < 10$); and (v) free flowing powder when the ($FCC > 10$) [12, 22].

The bulk density of the materials (RHOB in kg/m^3) was automatically determined by the ring shear tester (RST-XS, Dietmar Schulze, Wolfenbuttel, Germany). This parameter gives an indication of how these materials may pack and was thus used as one of the predicted output parameters for the model development.

2.4. Radial Basis Function Network

2.4.1. Model Development

This section briefly introduces an RBF network that is used in this research to model the powder flow properties. Readers are referred to various books and research papers for more in-depth reading, in particular references [23-24]. The RBF network usually maps an $\boldsymbol{N}$ -dimensional input space ($\mathbf{x}$) to a one-dimensional output space (y_T). In addition to the various powders (i.e. material type) used and explained in Table 1, the rest of the input variables are defined as follows:

$$\mathbf{x}_{\text{full}} = [x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6]$$
$$= [D_{10} \quad D_{50} \quad D_{90} \quad D_{4,3} \quad S_{10} \quad S_{50}] \tag{4}$$

The output parameter is either $y_T = FFC$ or $y_T = RHOB$.

Such a network typically consists of an input layer, basis functions acting as a hidden layer and an output layer [17, 23]. Basis functions ($\phi_i(\mathbf{x})$) are functions of the radial Euclidian distance from a defined centre. A Gaussian function is a common selection for the basis function, which can be written as follows [23]:

$$\phi_i(\mathbf{x}) = \exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu}_i)}{2\boldsymbol{\sigma}_i^2}\right)$$

(5)

where $\boldsymbol{\mu}_i$ and $\boldsymbol{\sigma}_i$ are the centre and the standard deviation of the i^{th} function, respectively. The output of the mapping can then be expressed as a linear combination of these basis functions [23]:

$$y(\mathbf{x}) = \sum_{i=1}^{I} w_i \phi_i(\mathbf{x}) + w_0$$

(6)

where w_0 and w_i are the bias and the coefficient connecting the i^{th} basis function to the output layer, respectively. The general structure of the RBF network is presented in Figure 1. The numbers of neurons in the input and output layers are determined by the process under investigation (i.e. the numbers of the inputs and outputs). In order to minimize the error of predicting each output, multi-input single output (MISO) model is commonly utilized. The optimal number of the basis functions is the one that achieves a trade-off between good training and good generalization capabilities. Thus, it corresponds to the minimum error usually measured via the root mean square error (RMSE).

The RBF parameters (e.g. connecting coefficients and bias) are usually optimized via the use of the back-propagation network. In general, back-propagation is a supervised learning algorithm that aims to minimize the mean squared error between the target output and the predicted output [17, 23, 25]. Such an algorithm typically involves two phases, namely, forward and backward phases. The forward phase calculates the network predicted output according to the inputs, whereas the backward phase adapts the network parameters (e.g. the connecting coefficients) based on the error performance via the use of an elicited optimization algorithm. Various optimization algorithms including, but not limited to, gradient descent, quasi-Newton optimisation, conjugate gradient, Levenberg-Marquardt and nature inspired optimization algorithms (e.g. Genetic algorithm), have been presented in the related literature [20, 25]. In this research paper, the scaled conjugate gradient (SCG) algorithm is utilized to optimize the RBF network parameters.

2.5. Integrated Network

2.5.1. Model Development

The integrated network, as a data based model, relies on predicting the output via two phases. The structure of such a network for MISO is depicted in Figure 2. In the first phase, the N-dimensional input space ($\mathbf{x}$) and the one-dimensional target space (y_T) are utilized to develop and train M models having different structures (e.g. number of basis functions). Then, the predicted outputs (i.e. the predicted flow properties from each model) from these models (y_{P1}, y_{P2}… y_{PM}) and the target output are used, in the second phase, to develop and train a single model leading to the final predicted output ($\hat{y}_P$) [26]. The idea of this integrated network is that the different model structures in the first phase can play a complementary role in representing the underlying patterns between the input parameters investigated and the

flowability parameters (i.e. FCC and RHOB). Furthermore, training the model in two phases helps in extracting the associated knowledge from the available limited data [26].

The predicted output of the integrated network can analytically be expressed as a combination of composition and superposition of the basis functions as follows [26]:

$$\hat{y}_P = \sum_{m=1}^{M} w_m^{(2)} \phi_m \left(\sum_{k=1}^{K} w_k \phi_k (\mathbf{x}) + w_0 \right) + w_0^{(2)} \quad (7)$$

where K is the number of the basis functions in each model in the first phase and M represents the number of the RBF models defined in the first phase. The rest of the parameters are as defined in Section 2.4, where the superscript number in Eq. 7 is utilized to distinguish the parameters used in Phase II from the ones used in Phase I. It has been proved that the superposition and composition functions are dense in a convex data space [27-28]. Thus, the function presented in Eq. 7 can minimize the difference between the predicted and the target outputs and can significantly improve the predictive performance [26]. The SCG algorithm is utilized with the backpropagation network to optimize the network parameters for the two phases.

3. Results and Discussions

3.1 Micrometric and flow properties

The micrometric properties of the three powder materials used in this research study are summarized in Table 1. The powders in this table are also the powders used in making the blends for FFC and RHOB determination. The electron micrographs obtained by SEM for the different grades of MCC, DCP and Lactose are depicted in Figures 3, 4 and 5, respectively.

As shown in Figure 3, the MCC microsphere 100 has, as expected, spherical particles with diameter values are in the range of approximately 150 μm to 300 μm. It is worth mentioning that a similar morphology can also be observed for the MCC microsphere 200, however, the particle diameter values are in the range of approximately 200 - 300 μm. These results were further confirmed by the QICPIC analysis. It was also observed that the MCC PH101 and MCC PH102 have elongated plate-like particles, with size in the range (40 – 350 μm). It is worth emphasising at this stage that the former has a smaller particle size when compared to the latter. A similar particle shape can also be observed for the MCC PH105 and MCC PH302. However, differences in the particle size are shown in Figure 3. For instance, particle size of the MCC PH105 is smaller than that of the MCC PH102 and larger than that of the MCC PH101, whereas the size of the MCC PH302 is larger than these grades. The silicified grades of MCC have also elongated plate-like particles. A look at Figure 1 and Table 1 suggests that the sphericity descriptors decrease in order: MCC PH200 > MCC PH100 >MCC PH302 > MCC prosolv 90 > MCC PH105 > MCC PH101 > MCC prosolv 50> MCC PH102.

Different particle morphologies can also be noticed for the DCP grades, as presented in Figure 4. For example, the D160 and A150 grades show an aggregated plate-like morphology with relatively large particle size (i.e. $D_{4,3}$ is approximately 160.3 μm). A similar morphology is noticeable for the D14 and A12 but the particle size is less than 100 μm. In contrast to these grades, the A60 grade has the most spherical particle shape with uniform size distribution ($D_{4,3}$ is approximately 76.12 μm). The sphericity descriptors for these grades are listed in Table 1. In Figure 5, it can be seen that Lactose shows versatile morphologies ranging from cubical to complete spherical particles. For instance, Flowlac 90 and Flowlac 100 have completely spherical particles with quite similar particle size distributions, as presented in Table 1. Granulac 70 and Granulac 200 show cubical morphology with different

particle size (i.e. Granulac 70 has a larger particle size ($D_{4,3}$=173.27 μm) when compared to Granulac 200 ($D_{4,3}$=58.90 μm).

The FFC values of the various grades (MCC, DCP and Lactose) of the powder materials investigated are shown in Figure 6a-c. It is apparent that D160 has the best flowability, with an FFC value of approximately 65. Among the MCC and Lactose grades, MCC PH200 and Flowlac 100 have the best flowability with FFC values equal to approximately 33 and 20, respectively. Statistical correlation analysis across the powders investigated was carried out between the size and shape descriptors of the powders investigated and the flowability represented by the FFC and RHOB. Reasonable linear correlation coefficient values among most of them are listed Table **2**. Different correlation values can also be observed in Table **2**. For instance the relationship between the D_{50} and the FFC is stronger than the relationship between D_{50} and the RHOB. In addition, the relationship between the D_{90} and the FFC is a strong direct relationship (i.e. the correlation coefficient is positive), whereas the relationship between the D_{90} and the RHOB is a weak inverse one (i.e. the correlation coefficient is negative). The analysis of variance showed that the various materials have significant effects on the flow properties, where the p-values were less than 0.05.

It was also interesting to note that of the true densities determined, the DCP samples had the highest values ranging from 2.38 – 2.92 kg/m^3 whereas the MCC and Lactose grades had values ranging from 1.40 – 1.97 kg/m^3 and 1.54 – 1.68 kg/m^3 respectively (Table 1).

The blends for the MCC, DCP and Lactose grades displayed a range of FFC values which were correlated to their particle size descriptors (i.e. $D_{4,3}$ values) (Figure 7a-e). It was noticeable that an increase or a decrease in one of the ratios of the blends (3:1, 1:1 or 1:3) had a significant influence on the FFC values. This was correlated directly to the calculated

particle size descriptors of the blend under investigation. This indicated that the development of the linear volume-based mixing rule (Eq. 1) for the blend in determining their particle size descriptors is possible and can be used to determine the particle size descriptor of any potential ratios which can be fed into a model and the FFC thus predicted.

3.2. Radial Basis Function Network

An RBF model was employed to model and predict the flowability of the various pharmaceutical powders investigated. The experimental data were randomly classified into two sets: training set (60), which allows the RBF model to learn the input/output relationships, and testing set (26), by which the generalization capabilities of the developed RBF model are tested. In addition to the various powders and the different powder blends used, particle size represented by its descriptors (i.e. D_{10}, D_{50}, D_{90} and $D_{4,3}$) and particle shape represented by its sphericity descriptors (i.e. S_{10} and S_{50}) were considered as input variables, whereas the powder flow represented by the FFC and RHOB was considered as an output. The number of basis functions selected was the one that corresponded to the minimum RMSE values for both training and testing sets. For the FFC, Figure 8 shows the RBF performance for both the training and the testing data sets using 6 basis functions, with a RMSE (training, testing) = [2.90, 5.16]. The testing RMSE value is approximately twice the training RMSE value, which, at first glance, could be attributed to an over-training problem. However, it was noted that one of the FFC values in the testing set was larger than 60, whereas, in the training set, most of the values are less than 30, thus, an error residual of 5 is actually less than 10% of the target value. The coefficient of determination (R^2) values for the training and testing sets are 0.80 and 0.79, respectively. The close R^2 values are an evidence that the over-training problem was not the case in this work. In a similar manner, an RBF model was developed for the RHOB. The performance measures presented by the R^2 (training, testing) and RMSE

(training, testing) values are [0.78, 0.77] and [112, 151], respectively, as summarized in Table 3. The results obtained indicate that the RBF network cannot represent and accurately predict the flow properties. This can be attributed to the limited number of data points (i.e. powder samples) and to the so-called "curse of dimensionality", which refers to the phenomenon that occurs when one deals with spaces of high dimensionality comprising of many input variables.

The predictive performance of the RBF model can, thus, be improved by reducing the dimensionality of the process (i.e. reducing the number of input variables). Therefore, an RBF model was developed using the materials and their mixtures, D_{50}, $D_{4,3}$ and S_{50}. The model performance values for the FFC and the RHOB are R^2 (training, testing) = [0.84, 0.85] and R^2 (training, testing) = [0.82, 0.83], respectively. The RMSE values for the FFC and the RHOB are [2.12, 4.72] and [109, 142], respectively, as listed in Table 3. These performance measures indicate that the RBF model developed using less inputs is superior to that of the RBF network developed using all the inputs, with an overall improvement of 7%. Although, such a model satisfactorily modelled the flow of the investigated powders, reducing the number of inputs may affect the generalization capabilities of the model. All the size descriptors should be included in the model, in order to take into account a multimodal or wide size distribution. Therefore, an integrated network is presented to capture the relationships between all the size and shape descriptors and the flow properties.

3.3. Integrated Network

In order to implement the integrated network, ten RBF networks, with different number of basis functions and different connecting coefficient values, and a single RBF one in the first and second phases, respectively, were developed (Table 3). For each model in the first phase, the data was randomly classified into two sets: training set (60) and testing set (26). For each flow property, the network parameters are listed in Table 3. It is worth

emphasising at this stage that the number of data points (i.e. powder samples) in training and testing data sets were the same for all RBF networks but their distributions in the space under investigation were different. For this reason, these models can play a complementary role in representing the possible patterns by considering the different areas in the space under investigation.

The integrated network performance measures for the FFC were R^2 (training, testing) = [0.92, 0.93] and RMSE (training, testing) = [1.41, 1.92], as shown in Figure 9, while the performance measures for the RHOB were R^2 (training, testing) = [0.91, 0.90] and RMSE (training, testing) = [75, 93], as shown in Figure 10. The output predictions in Figures 8 and 9 elucidate a satisfactory performance, where it was noticeable that most of the predicted values fitted properly within the 90% confidence interval. In addition, the prediction performance of the integrated network was superior to that of the single RBF network presented, with overall improvements of approximately 16% and 19% in R^2 for the FFC and RHOB, respectively. These results prove the ability of the integrated network in handling the difficulties of modelling the powders flowability and in dealing with the limited number of data points, this being due to the dense function represented by the superposition and composition functions presented in Eq. 7.

4. Conclusions

Modelling and predicting the flow properties of powder materials are essential in many industries, in particular the pharmaceutical industry, this being due to the fact that powder flow behaviour can affect the manufacturing efficiency and determine the final product quality. In this research work, data-driven models were developed to predict the flow properties of various commonly used pharmaceutical powders and their blends. Firstly, a radial basis function (RBF) network was utilized to map the size (i.e. D_{10}, D_{50}, D_{90} and $D_{4,3}$)

and shape (i.e. S_{10} and S_{50}) descriptors to the flow properties represented by the flow function coefficient (FFC) and the bulk density (RHOB). The simple RBF network, however, was not able to capture the highly nonlinear input/output relationships and the high dimensionality of the flowability. An integrated network was thus implemented to model the flow properties. In such a structure, the output was predicted by training and modelling the acquired data in two consecutive stages. The integrated network was successfully able to (i) capture the relationships between the particle size and shape and the flow properties, (ii) deal with the high dimensionality of the space under investigation, and (iii) predict the flow properties accurately. Furthermore, the integrated network thus outperformed the single RBF network in terms of the predictive performance and the generalization capabilities. Such a model has the ability to guide formulators in selecting excipients and their blend concentrations that can improve the flowability of a powder blend. Employing such a model can therefore reduce time and material waste. There is however a need to improve the interpretability of the input/output relationships, which can be achieved by incorporating fuzzy logic systems in the modelling structure.

5. Acknowledgment

The Authors would like to thank Andrew France at Sympatec for providing access to the QICPIC instrument that was used in this work. The first author would also like to thank the EPSRC DTP at the University of Huddersfield for funding this research work.

References

[1] Chattoraj, S., Sun, C.C., 2018. Crystal and Particle Engineering Strategies for Improving Powder Compression and Flow Properties to Enable Continuous Tablet Manufacturing by Direct Compression. J. Pharm. Sci. 107, 968–974.

[2] Navaneethan, C. V., Missaghi, S., Fassihi, R., 2005. Application of powder rheometer to determine powder flow properties and lubrication efficiency of pharmaceutical particulate systems. AAPS PharmSciTech 6, E398–E404.

[3] Visser, J., 1989. Van der Waals and other cohesive forces affecting powder fluidization. Powder Technol. 58, 1–10.

[4] Šupuk, E., Ghori, M.U., Asare-Addo, K., Laity, P.R., Panchmatia, P.M. and Conway, B.R., 2013. The influence of salt formation on electrostatic and compression properties of flurbiprofen salts. Int J Pharm, *458*(1), pp.118-127.

[5] Asare-Addo, K., Šupuk, E., Al-Hamidi, H., Owusu-Ware, S., Nokhodchi, A. and Conway, B.R., 2015. Triboelectrification and dissolution property enhancements of solid dispersions. Int J Pharm, *485*(1-2), pp.306-316.

[6] Adebisi, A.O., Kaialy, W., Hussain, T., Al-Hamidi, H., Nokhodchi, A., Conway, B.R. and Asare-Addo, K., 2016. Solid-state, triboelectrostatic and dissolution characteristics of spray-dried piroxicam-glucosamine solid dispersions. Colloids and Surfaces B: Biointerfaces, *146*, pp.841-851.

[7] Adebisi, A.O., Kaialy, W., Hussain, T., Al-Hamidi, H., Nokhodchi, A., Conway, B.R. and Asare-Addo, K., 2016. An assessment of triboelectrification effects on co-ground solid dispersions of carbamazepine. Powder Technol, *292*, pp.342-350.

[8] Gaisford, S., Saunders, M., 2013. Essentials of pharmaceutical preformulation. Wiley-Blackwell.

[9] Garg, V., Mallick, S.S., Garcia-Trinanes, P., Berry, R.J., 2018. An investigation into the flowability of fine powders used in pharmaceutical industries. Powder Technol. 336, 375–382.

[10] Fu, X., Huck, D., Makein, L., Armstrong, B., Willen, U., Freeman, T., 2012. Effect of particle shape and size on flow properties of lactose powders. Particuology 10, 203–208.

[11] Hou, H., Sun, C.C., 2008. Quantifying Effects of Particulate Properties on Powder Flow Properties Using a Ring Shear Tester. J. Pharm. Sci. 97, 4030–4039.

[12] Yu, W., Muteki, K., Zhang, L., Kim, G., 2011. Prediction of Bulk Powder Flow Performance Using Comprehensive Particle Size and Particle Shape Distributions. J. Pharm. Sci. 100, 284–293.

[13] Kachrimanis, K., Karamyan, V., Malamataris, S., 2003. Artificial neural networks (ANNs) and modeling of powder flow. Int. J. Pharm. 250, 13–23.

[14] Luding, S., 2005. Shear flow modeling of cohesive and frictional fine powder. Powder Technol. 158, 45–50.

[15] Fraige, F.Y., Langston, P.A., Chen, G.Z., 2008. Distinct element modelling of cubic particle packing and flow. Powder Technol. 186, 224–240.

[16] Wu, J., Chen, J., Yang, Y., 2008. A modified kinematic model for particle flow in moving beds. Powder Technol. 181, 74–82.

[17] AlAlaween, W.H., Mahfouf, M., Salman, A.D., 2017. Integrating physics with data analytics for the hybrid modelling of the granulation process. AIChE. 63, 4761-4773.

[18] AlAlaween, W.H., Khorsheed, B., Mahfouf, M., Gabbott, I., Reynolds, G.K., Salman, A.D, 2018. Transparent Predictive Modelling of the Twin Screw Granulation Process using a Compensated Interval Type-2 Fuzzy System. European Journal of Pharmaceutics and Biopharmaceutics. 124, 138-146.

[19] Wang, Y., Koynov, S., Glasser, B. J., & Muzzio, F. J. (2016). A method to analyze shear cell data of powders measured under different initial consolidation stresses. *Powder technology*, *294*, 105-112.

[20] Leung, L. Y., Mao, C., Srivastava, I., Du, P., & Yang, C. Y. (2017). Flow function of pharmaceutical powders is predominantly governed by cohesion, not by friction coefficients. *Journal of pharmaceutical sciences*, *106*(7), 1865-1873.

[21] Megarry, A. J., Swainson, S. M., Roberts, R. J., & Reynolds, G. K. (2019). A big data approach to pharmaceutical flow properties. *International journal of pharmaceutics*, *555*, 337-345.

[22] Rhodes, M.J., 2008. Introduction to particle technology. Wiley.

[23] Bishop, C., 1995. Neural networks for pattern recognition. Oxford: Clarendon Press, New York.

[24] Chen, S., Billings, S.A., Cowan, C.F.N., Grant, F.M., 1990. Non-linear systems identification using radial basis functions. International Journal of Systems Science. 21, 2513-2539.

[25] Duda, R.O., Hart, P.E., Stork, D.G., Pattern Classification. John Willey & Sons Inc., New York.

[26] AlAlaween, W.H., Mahfouf, M., Salman, A.D., 2016. Predictive modelling of the granulation process using a systems-engineering approach. Powder Technology. 302, 265-274.

[27] Cybenko, G., 1988. Continuous valued neural networks with two hidden layers are sufficient. Technical Report, Department of Computer Science, Tufts University.

[28] Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. Mathematics of Control, Signals, and Systems. 2, 303-314.

The authors declare no conflict of interest in this work

Figure captions

Figure 1. The Radial Basis Function (RBF) network structure.

Figure 2. The architecture of the integrated network [26].

Figure 3. SEM images for the different grades of MCC: (a) PH101, (b) PH102, (c) PH105, (d) PH302, (e) SMCC 50, (f) SMCC 90, (g) Microsphere 100 and (h) Microsphere 200.

Figure 4. SEM images for the different grades of Dicalcium Phosphate: (a) A12, (b) A60, (c) A150, (d) D14 and (e) D160.

Figure 5. SEM images for the different grades of Lactose: (a) Granulac 70, (b) Granulac 200, (c) Tablettose 80, (d) Inhalac 250, (e) Inhalac 400, (f) Flowlac 90 and (g) Flowlac 100.

Figure 6. FFC values of MCC grades (a), DCP grades (b) lactose grades (c) **and range of other pharmaceutically relevant excipients (d)**

Figure 7. FFC values for a range of pharmaceutical blends used in building the RBF and integrated network ; (a) Blends of MCC grades; C1 is MCC200, C2 is MCC101, C3 is Prsolov 90, C4 is MCC102 and C5 is MCC105; (b) Blends of DCP grades; D1 is A150, D2 is D14, D3 is D160, D4 is A12 and D5 is A60; (c) Blends of lactose grades; M1 is Granulac 70, M2 is Granulac 200, M3 is Inhalac 400, M4 is Flowlac 100, M5 is Tablettose 80, M6 is Inhalac 250 and M7 is Inhalac 90; (d) Blends of MCC and lactose grades; (e) Blends of MCC and DCP grades; Numbers after the coded names of the various excipients used in each blend refers to the ratios of materials used in manufacturing that specific blend.

Figure 8. The RBF network for the FFC (RBF-FFC1 in Table 3): (a) Training, (b) Testing (with 10% bands).

Figure 9. The integrated network based on 10 RBF networks for the FFC (IN-FFC in Table 3): (a) Training, (b) Testing (with 10% bands).

Figure 10. The integrated network based on 10 RBF networks for the RHOB (IN-RHOB in Table 3): (a) Training, (b) Testing (with 10% bands).

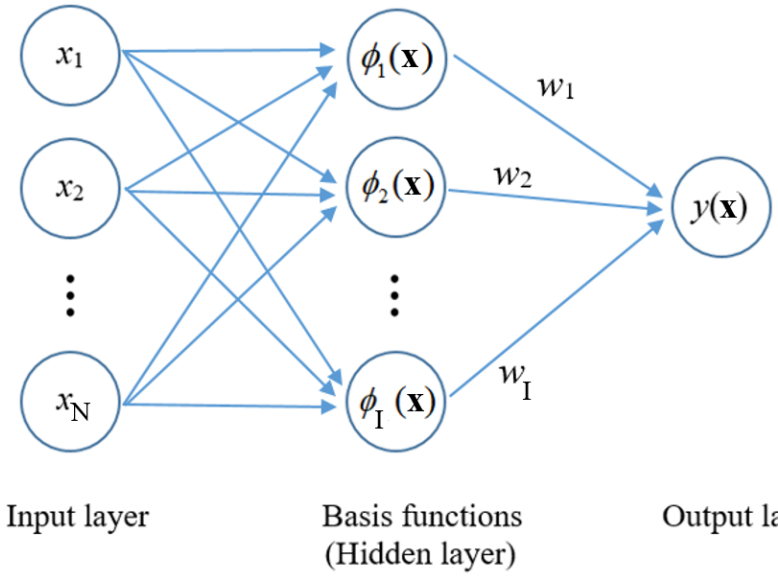


Figure 1

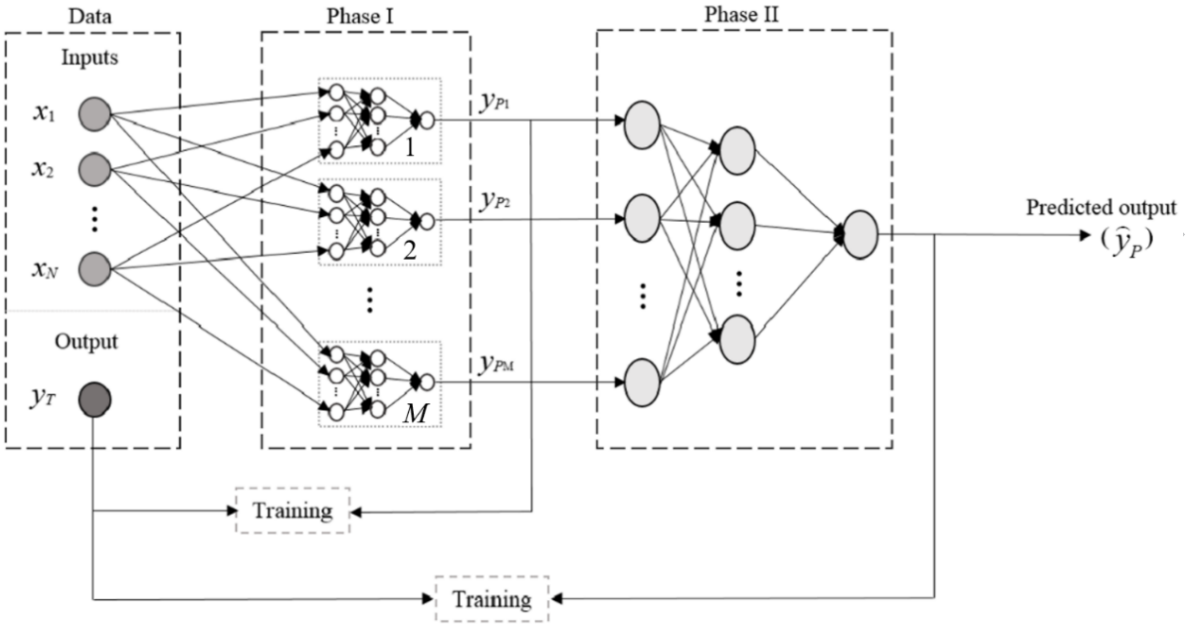


Figure 2

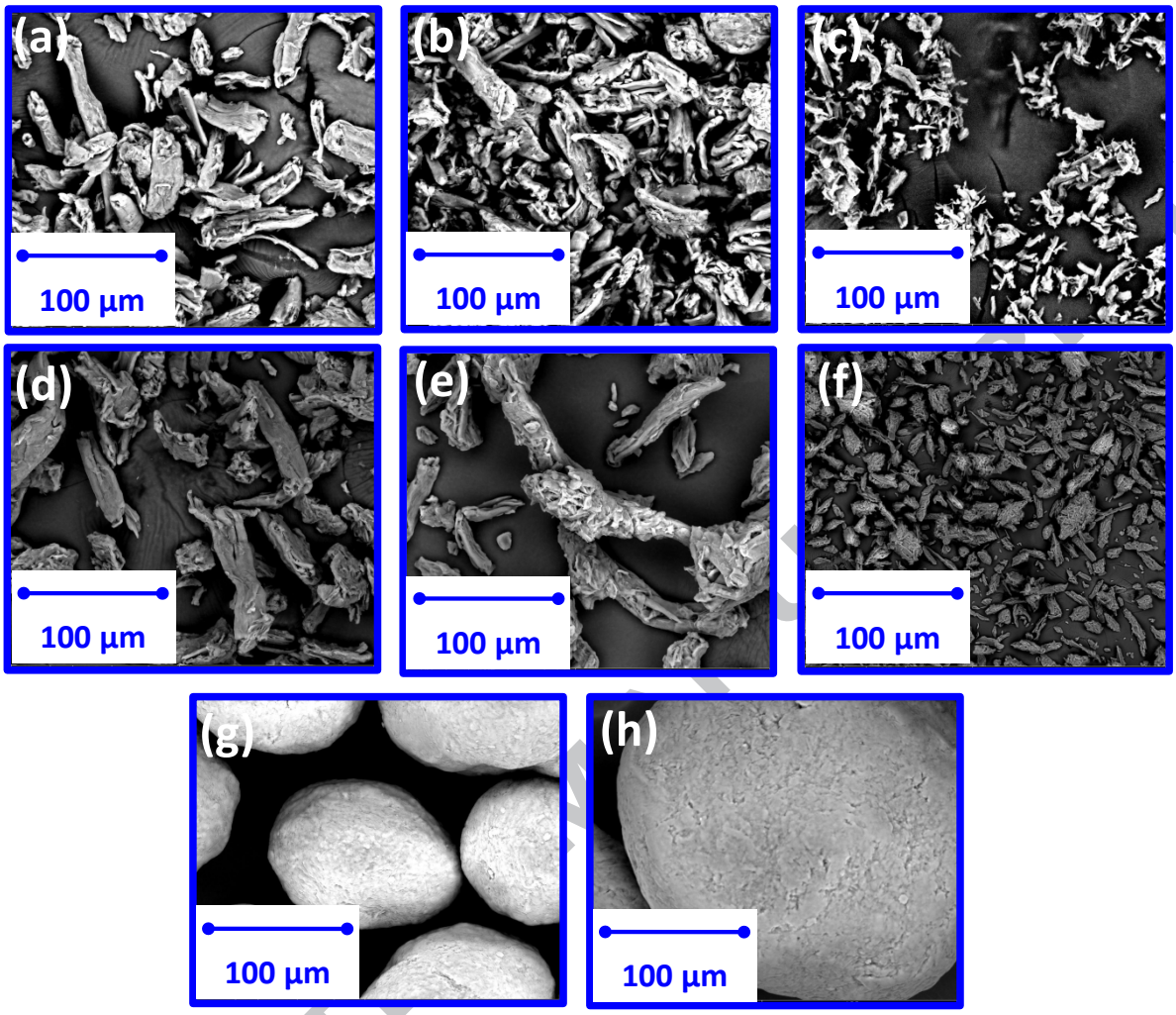


Figure 3

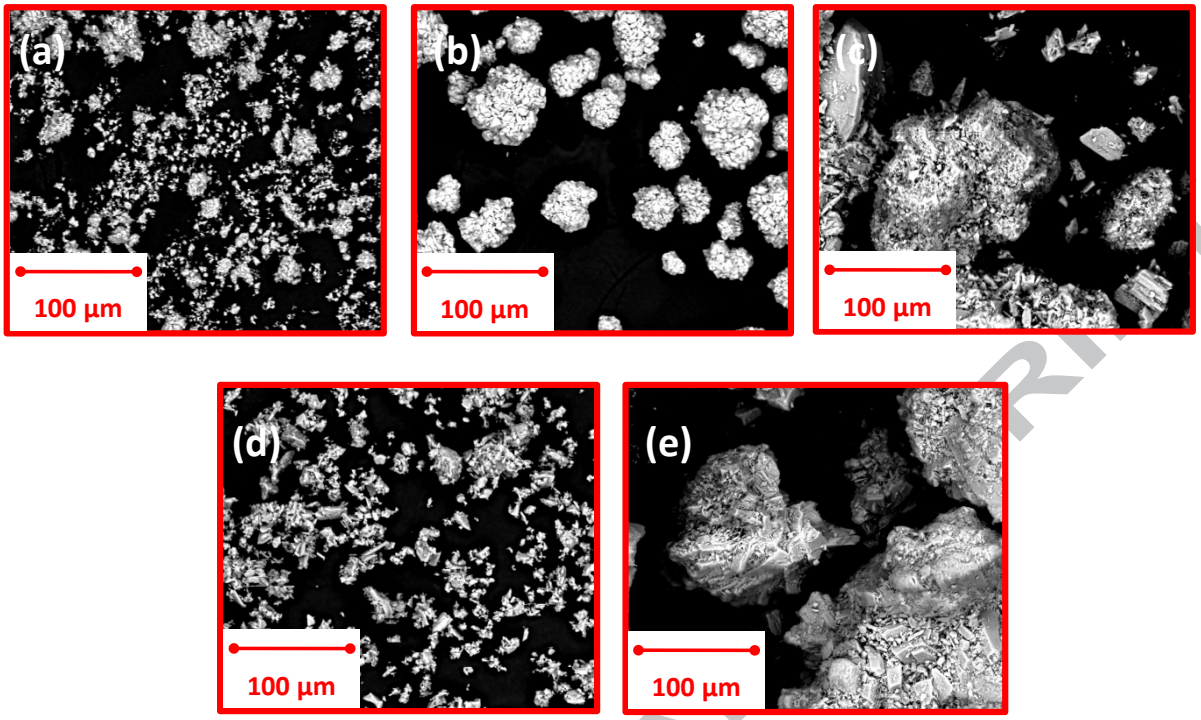


Figure 4

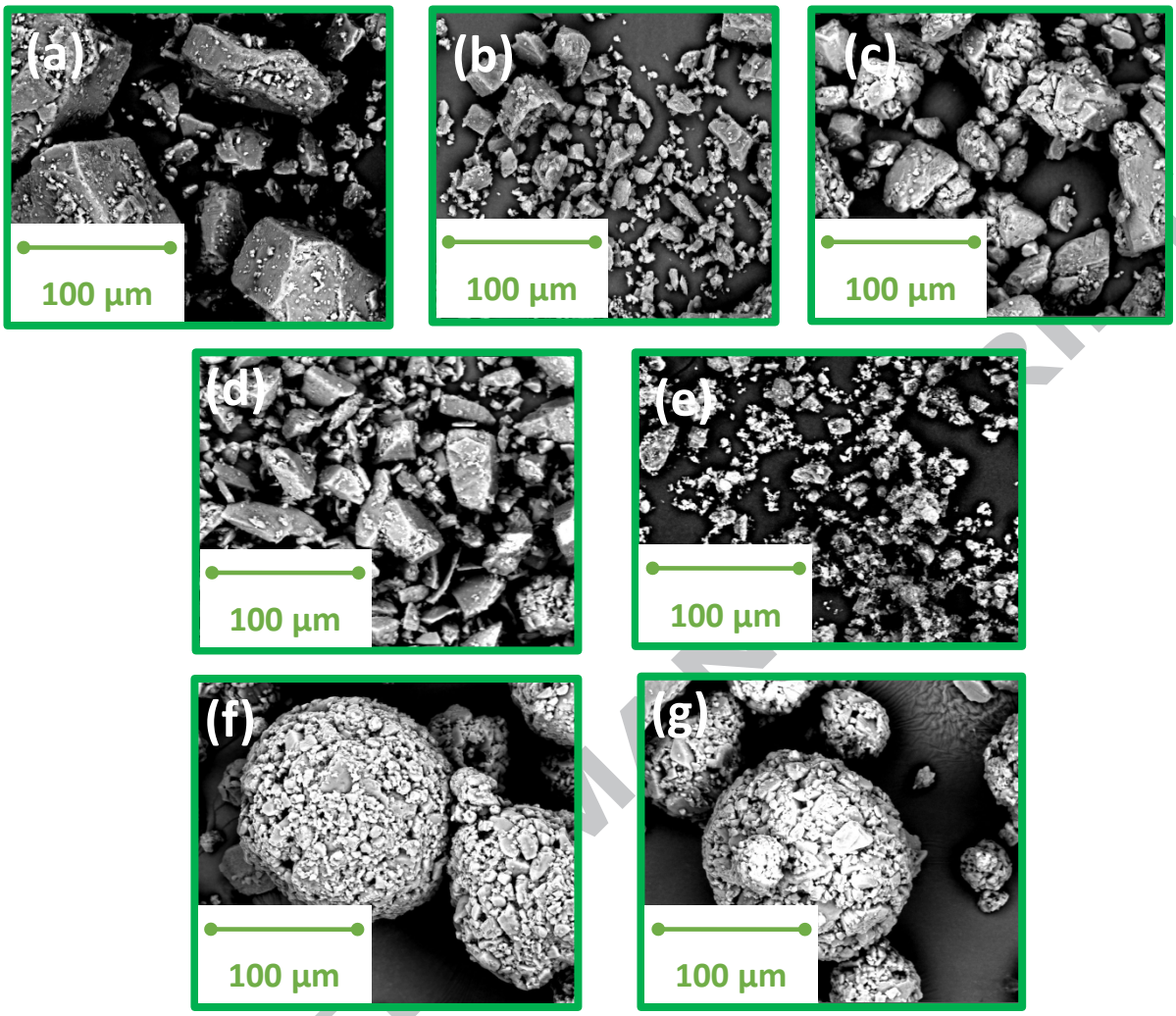


Figure 5

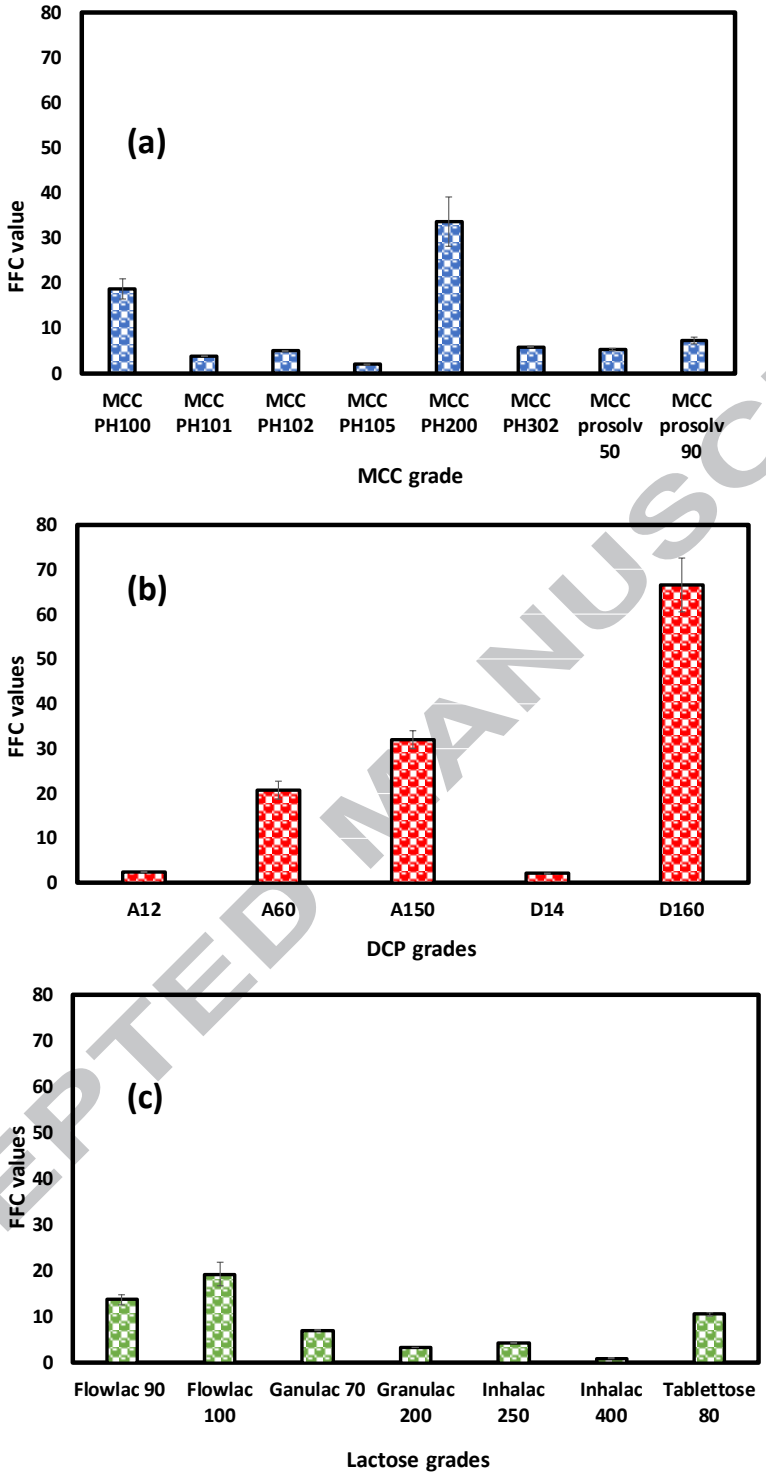


Figure 6

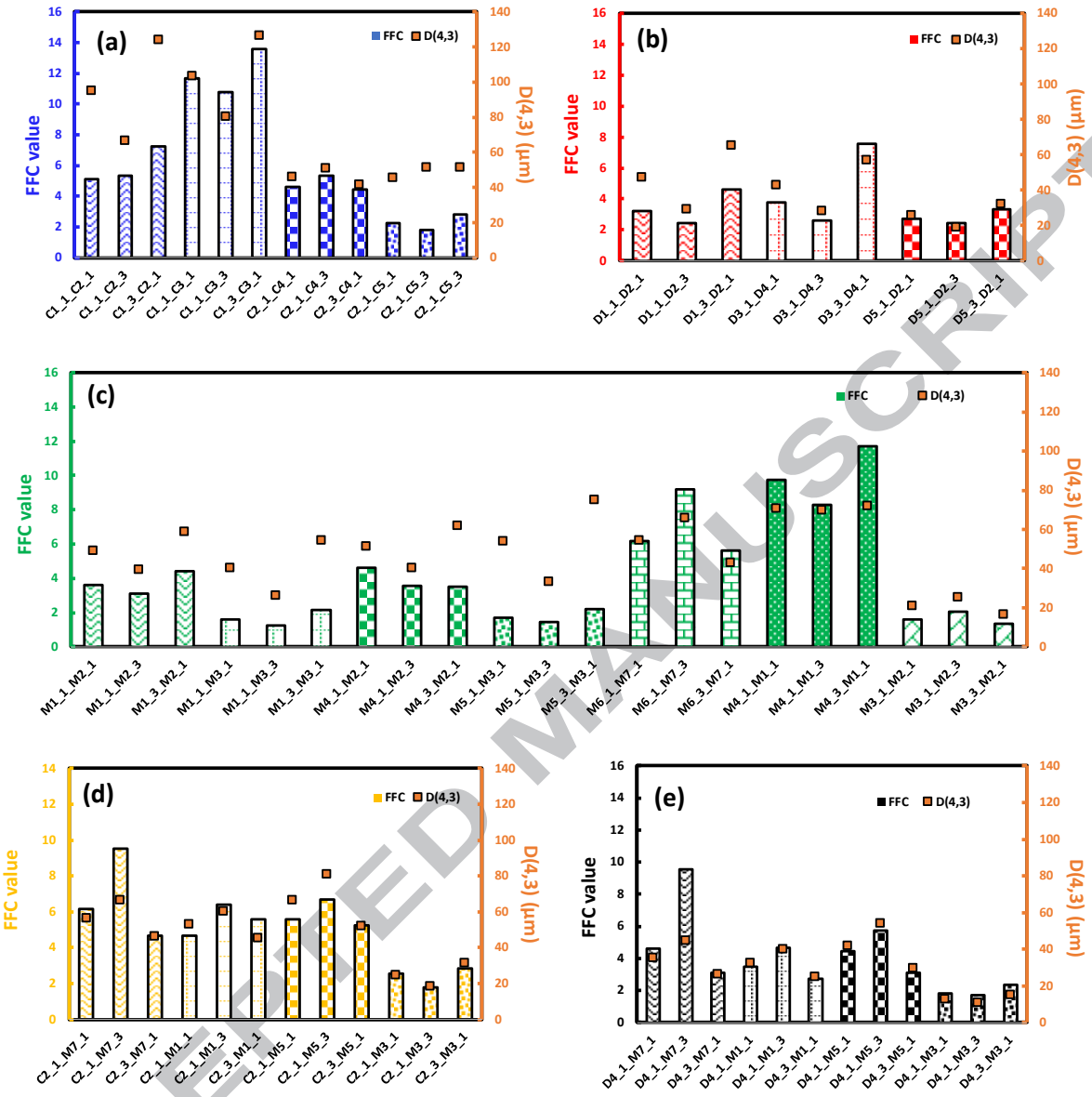
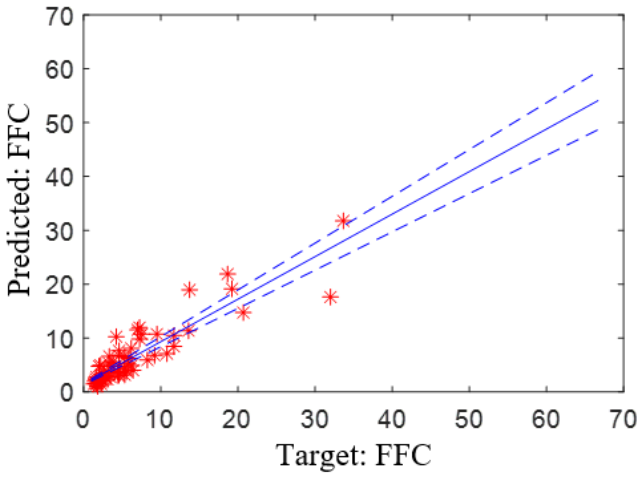
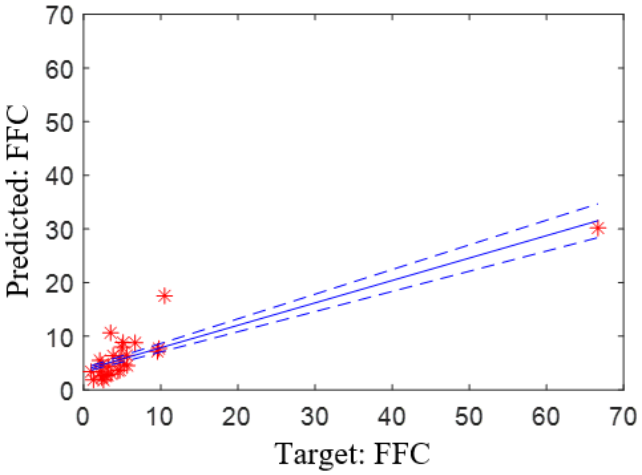


Figure 7

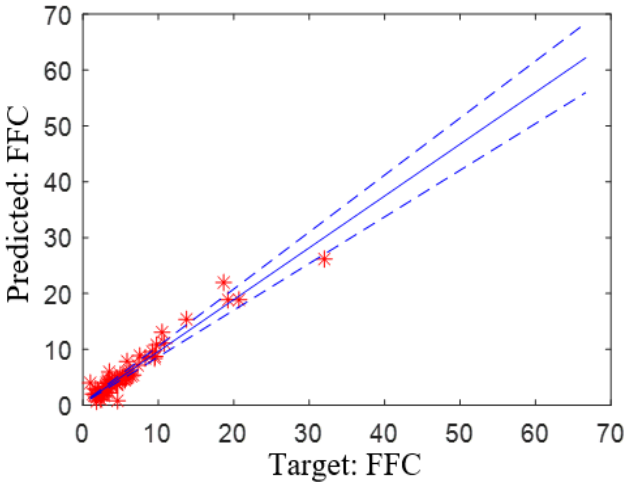


(a)

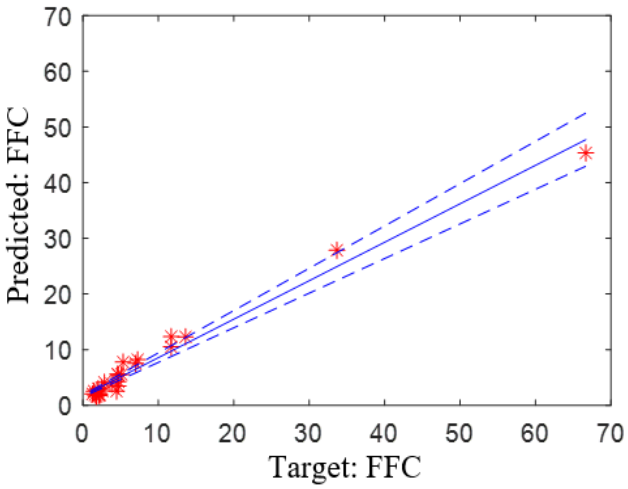


(b)

Figure 8



(a)



(b)

Figure 9

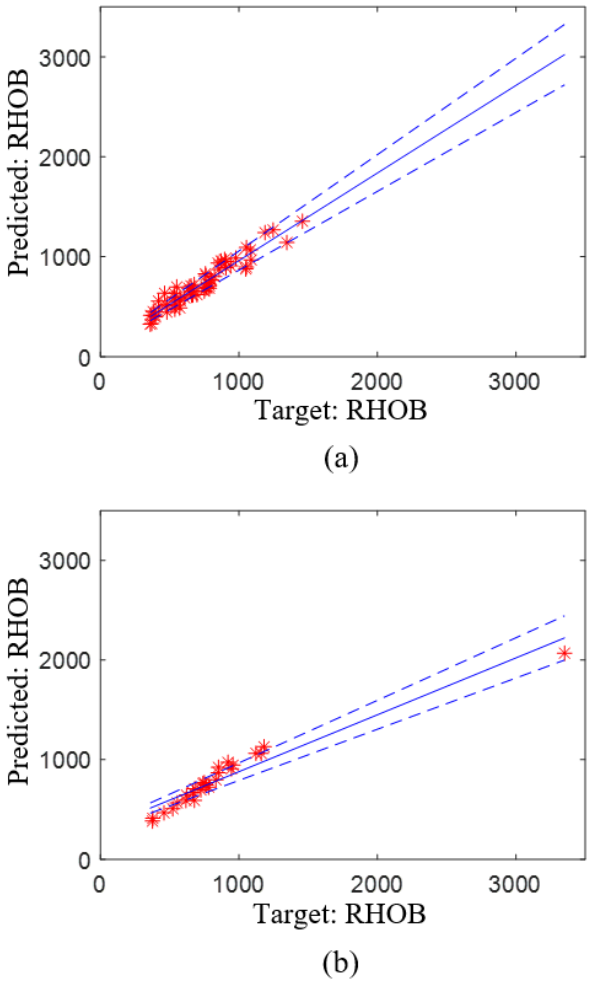


Figure 10

Table 1. Micrometric properties of the MCC, DCP and lactose grades.

Powders	Grades	Particle Size (µm)				Aspect Ratio		True density (g/cm^3)
		D_{10}	D_{50}	D_{90}	$D_{4,3}$	S_{10}	S_{50}	ϱ
Microcrystalline Cellulose	MCC PH100	177.16	217.91	272.19	222.16	0.61	0.78	1.56
	MCC PH101	32.51	68.09	126.21	75.32	0.39	0.6	1.57
	MCC PH102	39.98	89.96	213.06	111.07	0.36	0.58	1.58
	MCC PH105	13.52	34.15	235.41	104.74	0.44	0.61	1.97
	MCC PH200	241.02	313.56	384.33	311	0.71	0.87	1.52
	MCC PH302	37.32	104.49	241.62	122.38	0.44	0.64	1.56

	MCC prosolv 50	31.77	68.51	116.92	72.26	0.38	0.59	1.40
	MCC prosolv 90	40.04	97.64	201.77	111.16	0.41	0.62	1.64
Dicalcium Phosphate Dihydrate	A12	8.43	24.94	45.01	26.87	0.47	0.7	2.92
	A60	38.08	68.06	98.27	76.12	0.64	0.79	2.86
	A150	53.4	155.56	269.03	160.27	0.59	0.75	2.85
	D14	8.19	22.41	43.66	25.36	0.45	0.68	2.60
	D160	96.57	170.69	264.43	160.37	0.61	0.76	2.38
Lactose Monohydrate	Flowlac 90	67.02	145.56	231.67	148.51	0.68	0.88	1.68
	Flowlac 100	58.33	139.92	240	145.5	0.68	0.87	1.55
	Ganulac 70	42.02	131.02	238.36	137.27	0.56	0.72	1.54
	Granulac 200	17.6	50.32	113.87	58.9	0.52	0.69	1.54
	Inhalac 250	27.23	64.56	101.2	65.36	0.53	0.7	1.55
	Inhalac 400	8.44	22.38	42.46	24.6	0.45	0.67	1.56
	Tablettose 80	40.77	138.18	433.12	191.53	0.6	0.75	1.55

Note: These grades of excipients used were coded as follows in making the blends as explained in section 2.1 and 2.3.

Blends of MCC grades; C1 is MCC200, C2 is MCC101, C3 is Prsolov 90, C4 is MCC102 and C5 is MCC105

Blends of DCP grades; D1 is A150, D2 is D14, D3 is D160, D4 is A12 and D5 is A60

Blends of lactose grades; M1 is Granulac 70, M2 is Granulac 200, M3 is Inhalac 400, M4 is Flowlac 100, M5 is Tablettose 80, M6 is Inhalac 250 and M7 is Inhalac 90

Table 2. The correlation coefficients between size/shape descriptors and the flow properties.

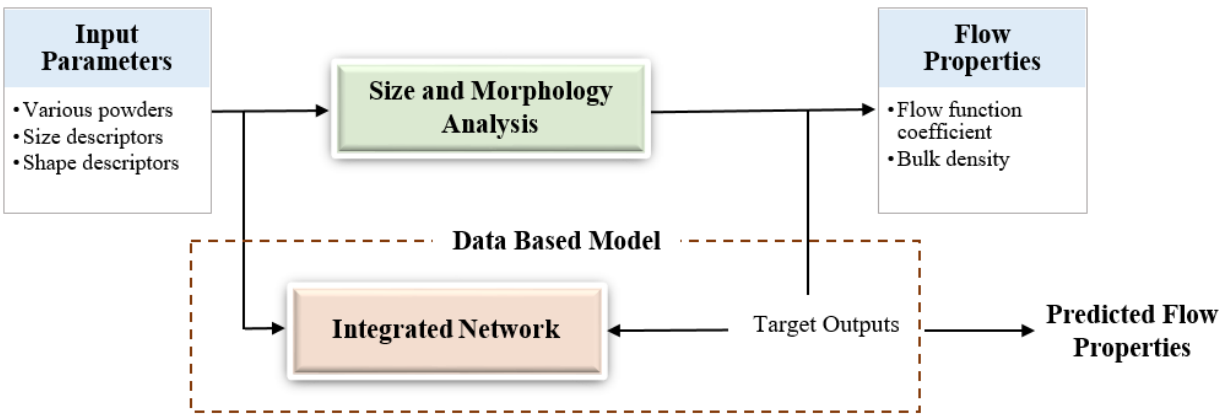
Particle characteristics		Flow properties	
		FFC	RHOB (kg/m^3)
Size descriptors	D_{10}	0.65	-0.01
	D_{50}	0.73	-0.05
	D_{90}	0.59	-0.17
	$D_{4,3}$	0.67	-0.11
Shape descriptors	S_{10}	0.61	0.06
	S_{50}	0.54	0.03

Table 3. Input and output parameters, coefficient of determination (R^2) and root mean squared error for the RBF and integrated networks (INs). $\mathbf{x}_{full}$ is defined in Eq. 4.

	Input	Output	Number of RBFs	R^2	RMSE
RBF-FFC1		*FFC*	*I = 6*	[0.80, 0.79]	[2.90, 5.16]
RBF-RHOB1		*RHOB*	*I = 6*	[0.78, 0.77]	[112, 151]
RBF-FFC2		*FFC*	*I = 7*	[0.84, 0.85]	[2.12, 4.72]
RBF-RHOB2		*RHOB*	*I = 9*	[0.82, 0.83]	[109, 142]
IN-FFC		*FFC*	*M = 10, K* = [3, 9]*	[0.92, 0.93]	[1.41, 1.92]
IN-RHOB		*RHOB*	*M = 10, K* = [3, 14]*	[0.91, 0.90]	[75, 93]

* The number of the basis functions of the RBF models defined in the first stage of the integrated network is in the range provided.

Graphical Abstract



Highlights

- A radial basis function (RBF) for pharmaceutical powders is presented
- An integrated modelling framework is presented to represent the flow properties of various pharmaceutical powders.
- The integrated modelling framework is validated.
- The flow properties are accurately predicted using the integrated network.