Received: 8 December 2019 | Revised: 16 August 2020 | Accepted: 27 September 2020

DOI: 10.1002/aic.17095

PROCESS SYSTEMS ENGINEERING

AIChE JOURNAL

A novel adaptive sampling based methodology for feasible region identification of compute intensive models using artificial neural network

Nirupaplava Metta[1] | Rohit Ramachandran[2] | Marianthi Ierapetritou[3]

[1]Applied Global Services, Applied Materials, Inc., Santa Clara, California

[2]Department of Chemical and Biochemical Engineering, Rutgers, The State University of New Jersey, Piscataway, New Jersey

[3]Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, Delaware

Correspondence
Marianthi Ierapetritou, Department of Chemical and Biomolecular Engineering, University of Delaware, 150 Academy St, Newark, DE 19716.
Email: mgi@udel.edu

Funding information
Consortium Agreement between Janssen Pharmaceutica, University of Ghent and Rutgers University; U.S. Food and Drug Administration, Grant/Award Numbers: 11695471, 1U01FD006487-01

Abstract

Identification of feasible region of operations in multivariate processes is a problem of interest in several fields. This is particularly challenging when the process model is black-box in nature and/or is computationally expensive, as analytical solutions are not available and the number of possible model evaluations is limited. An efficient methodology is required to identify samples where the model is evaluated for developing a computationally efficient surrogate model. In this work, an artificial neural network based surrogate model is proposed which is integrated with a statistical-based approach (Jack-knifing) to estimate the variance of the surrogate model prediction. This allows implementation of an adaptive sampling approach where new samples are identified close to the feasible region boundary or in regions of high prediction uncertainty. The proposed approach performs better than a previously published kriging based method for different dimensionality case studies.

KEYWORDS

adaptive sampling, black-box modeling, feasibility analysis, neural networks, surrogate model

1 | INTRODUCTION

Feasibility is the ability of a process to satisfy all production, operating, safety and quality constraints. Early concept of process feasibility was formulated[1] in terms of a flexibility test problem, which determines if a process can operate over a specified parameter range. The motivation was to incorporate operability considerations at design stage. The flexibility index problem determines a quantitative measure which indicates the extent of flexibility that can be attained in a given parameter range. The vertex solution method[2] and the active set method[3] require the constraints to be in closed form to apply traditional optimization methods. The simplicial approximation approach proposed by Goyal, Ierapetritou[4] measured the feasible region by evaluating inner and outer approximations of the feasible region. Points are determined at the boundary of the feasible region and a convex hull defined by these points is evaluated. However, this requires several function evaluations and is not applicable for nonconvex feasible regions. Building on this idea, Banerjee, Ierapetritou[5] proposed surface reconstruction approach by sampling points and constructing polygonal representation of the feasible region. While the method is successful in identifying disjoint and nonconvex feasible regions, large sampling costs are incurred which is a limitation for computationally expensive models. Laky et al.[6] applied the flexibility index concept to identify a probabilistic design space of pharmaceutical processes. Here, uncertainty in model parameters were considered to exist. The flexibility index problem was solved in terms of the model parameters and a probability map was generated over several process parameter values.

During the past decade, the need to develop predictive models for complex processes accompanied by advances in computational capabilities has led to the evolution of higher fidelity models, which incur high computational cost. Traditional feasibility analysis methods listed earlier cannot be applied to these models as they require high number of function calls which is prohibitive for such models. In addition, the constraints are not available in closed-form or function derivatives are difficult to evaluate. Surrogate based methods have gained

prominence to address problems of such nature. In this class of methods, a surrogate model is used to approximate the computationally expensive model for identifying the feasible region. Banerjee et al.[7] used a HDMR surrogate model to approximate the original model and used the computationally inexpensive surrogate to identify the feasible region. Kucherenko et al.[8] used HDMR as a metamodel to represent the design space in case of uncertainty in model parameters. Rogers, Ierapetritou[9] used kriging as the surrogate model to approximate the original function for identifying disjoint and nonconvex feasible regions with limited sampling. Wang, Ierapetritou[10] used RBF surrogate model, which performed comparable to kriging surrogate model. In some of these approaches, an adaptive sampling methodology is used, where samples are identified in regions that are of high interest such as regions where the surrogate model prediction errors are high or regions that are expected to be feasible. However, these approaches showed limitations or did not illustrate examples in identifying feasible regions in high dimensional problems.

In this work, we propose a surrogate based feasibility analysis method where an artificial neural network (ANN) is used as the surrogate model to address problems that are computationally expensive or do not have constraints in closed form. Owing to their function approximation, classification and pattern recognition properties, ANNs have attracted attention in several fields such as chemical engineering,[11,12] biochemical engineering[13] to optimize processes, and pharmaceutical research.[14] For ANN applications related to process optimization, ANN is used as the surrogate model to approximate the original process model or the objective function, and the developed surrogate model is used for subsequent optimization.[15,16] In this work, we propose to use ANN as the surrogate model to approximate the feasibility function, which is a characterization of constraint violation or infeasibility in a problem. An adaptive sampling methodology is used. A goal of adaptive sampling in this work is to improve accuracy of the surrogate model that represents the feasible region. Another goal is to limit the number of samples required to achieve this improvement in accuracy. Adaptive sampling in feasibility analysis problems entails identification of samples in regions of high surrogate model prediction variance or regions close to the feasible region boundary. In this work, the prediction error of the surrogate model is estimated using a statistical technique known as jack-knifing.[17] The identification of samples in the regions of interest is achieved by maximizing a modified expected improvement function proposed by Boukouvala, Ierapetritou.[18] The neural network based feasibility analysis method thus developed shows better performance compared to previously published Kriging based method[9] for low as well as high dimensional problems. Once the proposed feasibility analysis methodology is implemented, an algebraic form that represents the feasibility of the process at various values of the uncertain parameters is available. This is advantageous as it replaces computationally expensive model evaluations. Following this, approaches such as vertex solution method may be applied to evaluate the traditional flexibility index.[19] In addition, the developed surrogate model can be used to replace the exact feasibility constraints in the design optimization problem.

In the following sections, the mathematical foundation and formulation of the surrogate based feasibility analysis method is described in Section 2.1. Following this, the feasibility analysis methodology based on surrogate models used in this work, namely Kriging and ANNs are described in Section 2.2 and 2.3 respectively. Section 3 illustrates implementation of the methodology in several low and high dimensional test problems. A case study implementing the methodology on a pharmaceutical manufacturing process is then described in Section 3.4.

2 | METHODS

2.1 | Feasibility analysis methodology

The concept of process feasibility was introduced to have mathematical foundation for incorporating operability considerations at design stage.[20] For a process with design variables d, control variables z, and uncertain parameters θ, the constraints that represent feasible operation can be expressed as Equation (1).

$$g_j(d,z,\theta) \le 0, j \in J. \tag{1}$$

Feasibility of operation of a design d, operating at given values of uncertain parameters θ, is determined by establishing whether proper adjustment of control variables z allows each inequality given by Equation (1) to hold. To determine the feasibility of a process, a feasibility function ψ as given in Equation (2) is formulated. For fixed values of uncertain parameters θ, the control variable z is adjusted to minimize the largest value of g. If $\psi < 0$, the design is feasible for the given values of θ. If $\psi > 0$, that means any value of the control variable z cannot bring the process to a feasible operation. If $\psi = 0$, this indicates the process is at the boundary of feasible and infeasible operation.

$$\psi(d,\theta) = \min_z \max_{j \in J} g_j(d,z,\theta)\,. \tag{2}$$

If there are no control variables in the process that is, in an open loop system, the feasibility function in Equation (2) can be reformulated as Equation (3). In this article, the focus is on feasibility analysis for a given design d, which is the methodology to identify the region of the uncertain parameters θ, where the feasibility function given in Equation (3), $\psi(d, \theta) \le 0$.

$$\psi(d,\theta) = \max_{j \in J} g_j(d,\theta). \tag{3}$$

In cases where the constraints in the process model are black-box in nature or the process model is computationally expensive to run, surrogate based feasibility analysis methods are developed to fit an inexpensive surrogate model of the feasibility function.

2.1.1 | Adaptive sampling strategy

The sample points required to build the surrogate model need to be carefully chosen as the original model is expensive. Adaptive sampling strategies[21] are used to identify the samples that provide most useful information when building the surrogate model. At these sample points, the original process model is run, constraint violations are evaluated and the corresponding feasibility function value, which is the maximum of the constraint violation values is evaluated. The surrogate model is then updated based on the input and output of the feasibility function. In this work, a modified expected improvement function EI_{feas} as given in Equation (4) is used to implement the adaptive sampling strategy. Here, y and s are the surrogate model prediction and standard error of the prediction at x respectively, and ϕ represents the standard normal density function.

$$EI_{feas}(x) = s\,\phi\left(-\frac{y}{s}\right) = s\frac{1}{\sqrt{2\pi}}e^{-0.5\left(\frac{y^2}{s^2}\right)}. \tag{4}$$

This function is modified from the EI function proposed by Jones, Schonlau, Welch[22] for surrogate based optimization problems. The formulation of EI_{feas} function is explained in detail in.[18] For surrogate based optimization problems, maximization of the EI function directs search towards the regions of high uncertainty, or the regions where the objective function is minimum. Samples from these regions are used to improve the surrogate model accuracy, with the improvement focused towards identifying the optimum value of the objective function in the optimization problem. The modified EI function (EI_{feas}) applied in this work is used for feasibility analysis problems. It has similar properties such that maximization of EI_{feas} identifies sample points close to the feasible region boundary or in the region of high prediction uncertainty. In this case, addition of these samples leads to higher accuracy of the surrogate model, with improvement focused towards identification of feasible regions. To demonstrate how the EI_{feas} function can be used to identify the feasible region boundary, partial derivates may be taken as shown in Equation (5) and (6).

$$\frac{\partial EI_{feas}}{\partial s} = \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2s^2}}\left(1+\frac{y^2}{s^2}\right). \tag{5}$$

$$\frac{\partial EI_{feas}}{\partial y} = \frac{-1}{\sqrt{2\pi}}\frac{y}{s}e^{-\frac{y^2}{2s^2}}. \tag{6}$$

From Equation (5), it is evident that the right hand side term is always greater than 0, which implies that EI_{feas} monotonically increases with prediction error s. This indicates that maximization of EI_{feas} identifies sample points with high prediction error s that is, points with high uncertainty or points in unexplored regions. In Equation (6), the sign of the right hand side term is the opposite of the sign of y. When $y < 0$, the partial derivative is >0, which implies that as y approaches zero, EI_{Feas} increases. Similarly, when $y > 0$, the partial derivative is <0, which implies that EI_{Feas} increases as y approaches zero. Hence, from Equation (6), maximizing EI_{feas} identifies sample points where the surrogate model prediction is close to zero, namely sample points close to the feasible region boundary. Overall, maximization of EI_{feas} serves to find the sample points with high prediction uncertainty or points close to the feasible region boundary, both of which are the properties required to implement an adaptive sampling strategy in feasibility analysis.

In this work, an initial surrogate model is built based on initial sample points identified using grid sampling or Latin hypercube sampling. These are typically low in number as this requires evaluation of the computationally expensive process model. During model improvement step, new sample points are added iteratively by maximizing the EI_{feas} function. The adaptive sampling is terminated after itr_{max} number of iterations. Specific details on the initial number of samples used to build the surrogate model, the maximum number of iterations used for adaptive sampling are discussed in further sections. In the following sections, the surrogate modeling methods used in this work are explained, and their implementation in the feasibility analysis methodology is elucidated.

2.2 | Kriging based feasibility analysis

Kriging also referred to as gaussian process regression[23] or stochastic process model[22] is a popular interpolating surrogate model used in several applications. An attractive feature of Kriging model is that an estimate of the model prediction error is inherently provided by the model. This is a useful feature in the implementation of adaptive sampling strategies as seen in optimization literature.[24] In Kriging, the predictor at an untested sample point is modeled as a function of process responses at previously tested points or design points. The Kriging model prediction $\hat{f}(x^i)$ for a d dimensional input x^i is given as a realization of a regression model $\hat{\mu}(x^i)$ and an error term $\hat{\varepsilon}(x^i)$ as given in Equation (7). Here, $\hat{\mu}(x^i) = \sum_p \beta_p g_p(x^i)$ that is, a linear combination of p basis functions and represents the mean of process responses. The regression term takes the form of a polynomial, typically constant, linear or quadratic. The error ε is a deviation from the mean process response and is represented by a Gaussian process with zero mean and covariance $\sigma^2 R(\theta, x^{(i)}, x^{(j)})$ between $\varepsilon(x^{(i)})$, $\varepsilon(x^{(j)})$, where R is the correlation model.

$$\hat{f}(x^i) = \hat{\mu}(x^i) + \hat{\varepsilon}(x^i). \tag{7}$$

The correlation model R indicates that the errors in the predicted values are correlated as a function of the input variables. Several correlation models are available and used in literature.[25] In this work, linear, exponential or gaussian correlation models are considered as they have been successfully applied in previously published Kriging based feasibility analysis methods.[9] The Kriging model structure, specifically the order and type of regression and correlation model are chosen in the initial model building phase. In this phase, the domain is sampled using a space-filling design such as grid sampling or Latin hypercube sampling. Several Kriging models are fit with the combinations of regression and correlation models. The mean squared error (MSE)

between predicted data and target data for each Kriging model is computed. The model structure corresponding to the Kriging model that yields the least MSE is chosen. In this work, Kriging implemented in design and analysis of computer experiment (DACE),[26] is used. Overall, the Kriging based feasibility analysis algorithm is depicted in Figure 1. The maximum number adaptive samples added is represented by itr_{max}. Specific details on the implementation of the methodology in two-dimensional and high dimensional problems are discussed in further sections. Performance of the Kriging based feasibility analysis methodology is compared to the proposed ANN based methodology, which is explained in the next section.

2.3 | Artificial neural networks based feasibility analysis

ANNs constitute input and output variables as neurons that are connected using hidden neurons or nodes arranged in layers. In ANN, a neuron is weighed, and the weighted values are sent to neurons in the succeeding layer. In addition, a bias is applied with a constant weight of 1 and sent to neurons in the succeeding layer. All the inputs to a neuron in the succeeding layer are summed and a transfer function is applied, the result of which is weighed again and sent to the next layer. To build an ANN as a function approximator, the weights and biases are estimated such that a cost function, which constitutes differences between ANN predictions and target values, is minimized. In this work, a feed-forward neural network trained with back propagation algorithm is used.[27] One input layer, one hidden layer and one output layer are used, as a single hidden layer feed-forward ANN is known to be good function approximator provided sufficient number of neurons in the hidden layer, also known as hidden neurons are used.[28] The number of hidden neurons is a variable to be determined in the network structure. Higher number of neurons may lead to overfitting issues. Overfitting of the network parameters are generally avoided using methods such as early stopping and Bayesian regularization.[29] Early stopping requires the dataset to be divided into training, validation and testing sets. The training of network parameters is stopped based on model prediction performance on the training and validation dataset. In Bayesian regularization, the cost function includes a penalty term for high weights and does not require validation dataset to estimate optimum network parameters. In this work, Bayesian regularization is used for network parameter estimation, as a

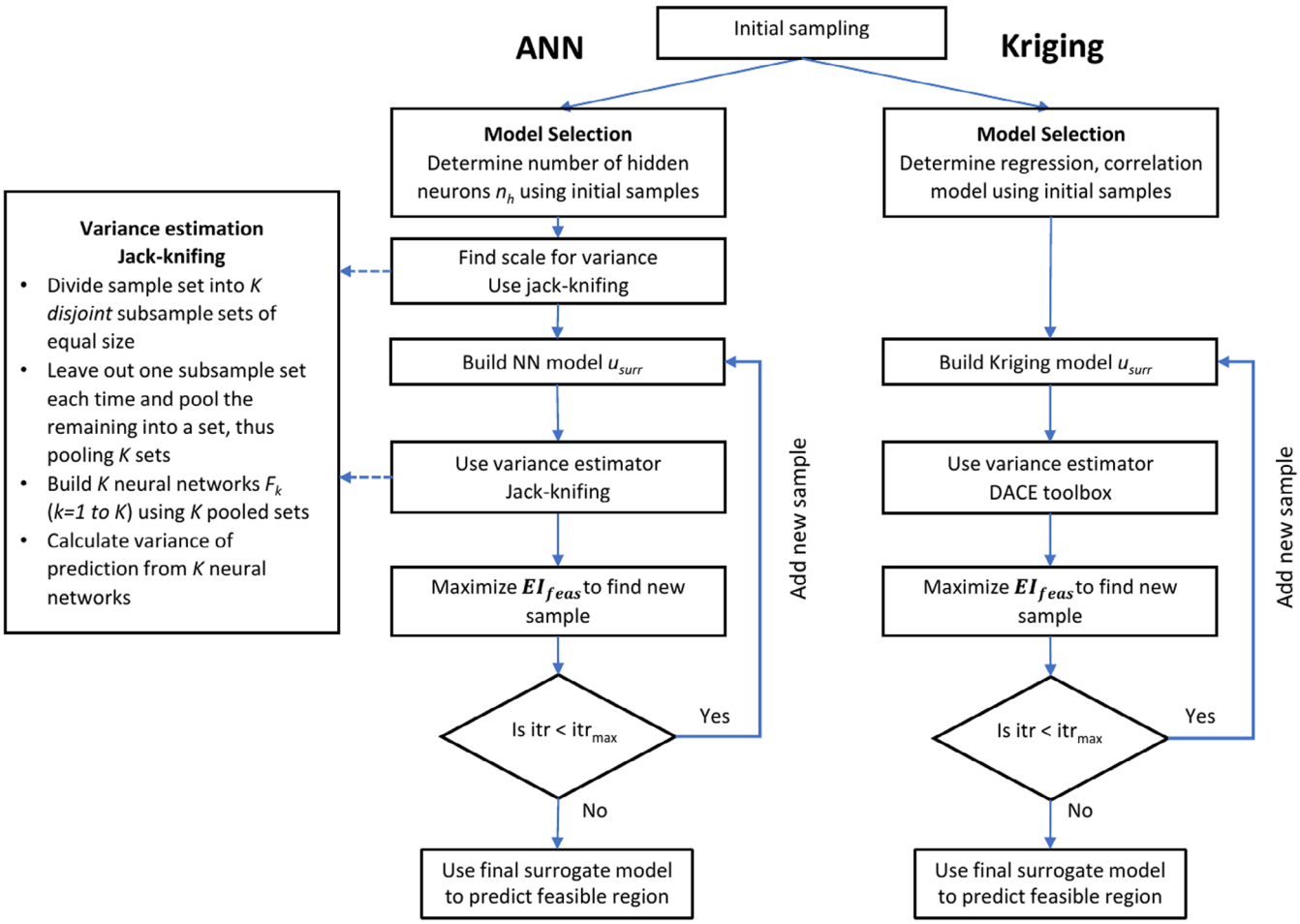

FIGURE 1 Algorithm for ANN and Kriging based feasibility analysis methodology. ANN, artificial neural network [Color figure can be viewed at wileyonlinelibrary.com]

validation dataset is not needed. This specifically serves as an advantage for computationally expensive models. In addition, it reduces potential oscillatory behavior between samples and reduces overfitting issues. For the hidden layer, a hyperbolic tangent sigmoid function also known as *tansig* is used as the activation function in this work. The number of hidden neurons is determined in the initial model building phase. In an approach similar to choosing the regression and correlation models in the Kriging based methodology, the domain is sampled using a space-filling design such as grid sampling or Latin hypercube sampling. Several ANNs are built with varying number of hidden neurons. The MSE between predicted data and target data for each ANN is computed. The number of hidden neurons corresponding to the neural network that yields the least MSE is chosen.

The implementation of feasibility analysis methodology includes maximization of the EI_{feas} function for adaptive sampling (Equation (4)) as explained in Section 2.1.1. The maximization of the EI_{feas} function requires the estimation of standard error of ANN model prediction s. In this work, a statistical technique known as jack-knifing[17,30] is used to estimate the variance σ of ANN model predictor y at a sample point x. The standard error s of the model predictor is then computed as $\sqrt{\sigma}$. This is evaluated at every sample point that requires estimation of the EI_{Feas} function. Computation of the variance σ is achieved by creating subsamples of the parent sample set. Following are the steps to implement the jack-knifing method for ANN model prediction variance:

1. Divide the sample set into K disjoint subsample sets of equal size
2. Leave out one subsample set each time and pool the remaining samples into a set. This leads to K sample sets
3. Build K neural networks F_k ($k = 1$ *to* K) using the K pooled sets
4. Calculate variance of prediction from K neural networks

In mathematical terms, at a sample point x, if $U_k(x)$ is the predictor from the kth ANN model, the variance of prediction is given by Equation (8).

$$\sigma(x) = \frac{1}{K}\sum_{k=1}^{K}\left(U_k(x) - \sum_{k=1}^{K}\frac{U_k(x)}{K}\right)^2. \tag{8}$$

The standard error s of the model predictor is further evaluated as $\sqrt{\sigma}$. The standard error s is used in the EI_{Feas} function and it appears in the exponential term $e^{-\frac{y^2}{2s^2}}$. In order to ensure the exponential term does not explode, in practice, we find that the scaling of the standard error s with respect to the model predictor y is needed. This is required to balance the magnitude of s compared to y. A similar scale factor has been used with a RBF based error prediction by Wang, Ierapetritou,[10] which has resulted in successful function maximization for feasibility analysis problems. The scale factor is determined based on model fit to the initial sample set in the initial model building phase. If $n_{initial}$ is the initial number of samples chosen, U_0 is the predictor from the ANN model that is fit using the initial sample set and σ_0 is the variance of the prediction using jack-knifing, the scale factor δ is given by Equation (9):

$$\delta = \frac{\max(U_0)}{\max(\sigma_0)^{0.5} * n_{initial}}, \tag{9}$$

where $\max(U_0)$ is the maximum value of the initial model predictor, and $\max(\sigma_0)$ is the maximum value of the initial model prediction variance. The EI_{feas} function for the ANN based methodology uses the scaled value, $\delta s(x) = \delta \sqrt{\sigma}(x)$ instead of $s(x)$. The algorithm used for the implementation of the proposed ANN based feasibility analysis is depicted in Figure 1.

2.4 | Surrogate model evaluation

The surrogate models are trained using the initial sample set and additional samples identified using the adaptive sampling strategy. In order to evaluate the effectiveness of the ANN based feasibility analysis methodology in comparison to the Kriging based methodology, the feasible region predicted by the surrogate models could be visually compared to the actual feasible region. However, such visual evaluation is possible for two- or three-dimensional problems. For higher dimensional problems, visual representation of feasible regions is a challenge. Hence, for this work, we use the performance of the surrogate model on a test sample set as a means to quantify effectiveness of the methodology. In this work, a set of samples generated using grid sampling are used as a test sample set. For two-dimensional problems, 100^2 samples are used. For three-, four-, five- and six- dimensional problems, $20,^3$ $7,^4$ $6,^5$ 5^6 samples respectively are used as test sample sets. Based on correct or incorrect identification of the test sample points as feasible or infeasible, the accuracy of the Kriging and ANN surrogate model can be evaluated and compared. Specifically, the test sample points can belong to one of the following categories:

1. Correctly identified feasible region (CF): the sample is classified as feasible by the original model and feasible by the surrogate model as well
2. Correctly identified infeasible region (CIF): the sample is classified as infeasible by the original model as well as the surrogate model
3. Incorrectly identified feasible region (IC-F): the sample is classified as infeasible by the original model and feasible by the surrogate model.
4. Incorrectly identified infeasible region (IC-IF): the sample is classified as feasible by the original model and infeasible by the surrogate model.

Based on the categories to which the points in the test sample set may belong, we define three metrics namely, percentage of correctly identified points in feasible region (CF%), percentage of correctly identified points in infeasible region (CIF%) and percentage of points in not conservative feasible region (NC%). The NC% metric is the percentage of points identified incorrectly as feasible, which is

undesirable as this is an over-estimation of feasible region by the model. In other words, the model may be considered 'not conservative' in identifying the feasible region. Hence, an accurate identification of feasible and infeasible regions corresponds to CF% and CIF% of 100 and NC% of 0. The respective calculations as given in Equation (10) are based on the number of sample points from the test sample set that belong to each of the four categories as explained earlier. The defined metrics may also be understood in terms of the percentage of points that are True Positive (TP), True Negative (TN) and False Positive (FP), which are more commonly used in the machine learning literature. The metrics using these terms are also added in Equation (10).

$$CF\% = \frac{CF}{CF + ICIF} * 100 = \frac{TP}{TP + FN} * 100.$$

$$CIF\% = \frac{CIF}{CIF + ICF} * 100 = \frac{TN}{TN + FP} * 100.$$

$$NC\% = \frac{ICF}{ICF + CF} * 100 = \frac{FP}{FP + TP} * 100. \quad (10)$$

In the following section, the proposed ANN based methodology is tested on several low and high dimensional problems and compared to the Kriging-based methodology.

3 | RESULTS AND DISCUSSION

A thorough evaluation of the proposed ANN based feasibility analysis method is discussed in this section. The performance of the proposed method is compared to the Kriging based method through several low and high dimensional problems. The test problems chosen are standard and have been used in the literature extensively. In addition to the standard test problems, two additional test problems are formulated in high dimensions (four and six dimensions). The formulated test problems serve to emulate a "flowsheet" problem, which is the case study discussed in section 3.4. In the following section, details on the test problems chosen for this work and the results of implementation of the Kriging and ANN based methodologies are discussed.

3.1 | Two dimensional test problems

Four two-dimensional test problems namely *branincon*, *camelback*, *ex3*[9] and *sasenacon*[31] are used in this work. The goal is to identify feasible regions for the test problems. The test problems are listed in Appendix A1. This table provides details on the variable bounds along with the constraints that define the feasible region for these problems. All of the problems include non-linear constraints, and the feasible regions are non-convex which are considered as difficult to identify. The *branincon*, *camelback*, and *sasenacon* test problems also have disjoint feasible regions which adds difficulty in identifying the feasible region boundary. The feasible region boundaries for these problems are shown in bold black color in Appendix A2. For these test problems, the performance of the ANN based and the Kriging based methodologies are compared through visual analysis of the feasible regions as well as through the validation metrics explained in Section 2.3.

For ANN as well as Kriging based methods, a grid sampling scheme is used to build the initial surrogate model in the initial model selection phase. The performance of the methods is compared for different number of initial samples that is, 49, 25, and 9, which correspond to 7, 5, and 3 levels respectively. After the initial surrogate model is built, the adaptive sampling strategy is implemented according to the algorithm in Figure 1 in order to identify additional sample points and to improve the surrogate model accuracy. To identify adaptive samples to be added at each iteration, a local solver *fmincon* in MATLAB is used to maximize the EI_{feas} function. To obtain a good initial guess, the EI_{feas} function is evaluated at densely sampled points in the variable space. Since evaluation of EI_{feas} function is computationally inexpensive, dense sampling is used. The point with the largest EI_{feas} value is selected as the initial guess. The dense sampling is obtained through Latin hypercube sampling of 1,000 points for two-dimensional problems. and 10,000 points for high dimensional problems. Several samples may be added at each iteration based on the number of samples chosen as initial guess. However, in this article, one sample is added per iteration (*itr*). The adaptive samples are added to the initial sample set and the Kriging and ANN models are trained and improved at every iteration. The accuracy of the surrogate model is evaluated every 10 iterations using CF, CIF and NC metrics as explained in Section 2.3. The test sample set used for model accuracy evaluation is generated through dense sampling of the variable space using 100^2 grid sampling. A maximum of 100 adaptive sample points is added to the initial sample set. If a desired surrogate model accuracy of CF > 99%, CIF > 99% and NC < 1% is achieved on the test set based on the model accuracy evaluation after every 10 iterations, the adaptive sampling is terminated earlier. Table 1 lists the initial surrogate model and the final surrogate model metrics for the ANN and Kriging based methods in the case of 49 initial samples. The number of adaptive samples required to reach an accuracy of 99% CF, 99% CIF, and 1% NC are also listed in Table 1. In addition, the number of hidden neurons used for the ANN model is also tabulated. The feasible region boundaries identified by the final surrogate model through dense sampling are also plotted for both methods to aid visual analysis.

From Table 1, in terms of the surrogate model accuracy achieved, Kriging as well as ANN models are able to identify the feasible regions with high accuracy (>97% feasible regions and > 99% infeasible regions are accurately identified) for *branincon* and *ex3* test problems, with ANN performing slightly better than Kriging in CF% and NC% and with fewer adaptive samples. Figures 2 and 3 show feasible region boundaries identified by the surrogate models compared to the actual feasible region boundaries for *branincon* and *ex3* problems respectively. The figures also show the location of adaptive samples. It is evident that the adaptive samples are placed close to the feasible region boundaries, which reflects the discussion in Section 2.1 regarding the maximization of the modified expected improvement function EI_{feas}.

TABLE 1 Surrogate model accuracy metrics for two-dimensional test problems

		Initial			Final				
Test problem	Surrogate model	CF%	CIF%	NC%	CF%	CIF%	NC%	Adaptive samples	Number of hidden neurons
Branincon	Kriging	26.10	100.00	0.00	97.86	99.89	1.20	100	–
	ANN	56.94	98.27	24.76	99.64	99.98	0.24	30	20
Camelback	Kriging	99.39	76.21	77.44	96.47	99.28	9.63	100	–
	ANN	87.88	98.55	19.18	99.85	100.00	0.00	20	24
Sasenacon	Kriging	58.03	98.06	17.85	88.14	99.62	2.73	100	–
	ANN	74.96	98.35	12.54	96.54	99.33	4.32	100	24
ex3	Kriging	91.53	98.25	3.22	98.38	99.50	0.88	100	–
	ANN	91.40	94.03	10.26	99.15	99.50	0.88	80	25

Abbreviation: ANN, artificial neural network; CF%, correctly identified points in feasible region; CIF%, percentage of correctly identified points in infeasible region; NC, percentage of points in not conservative feasible region.

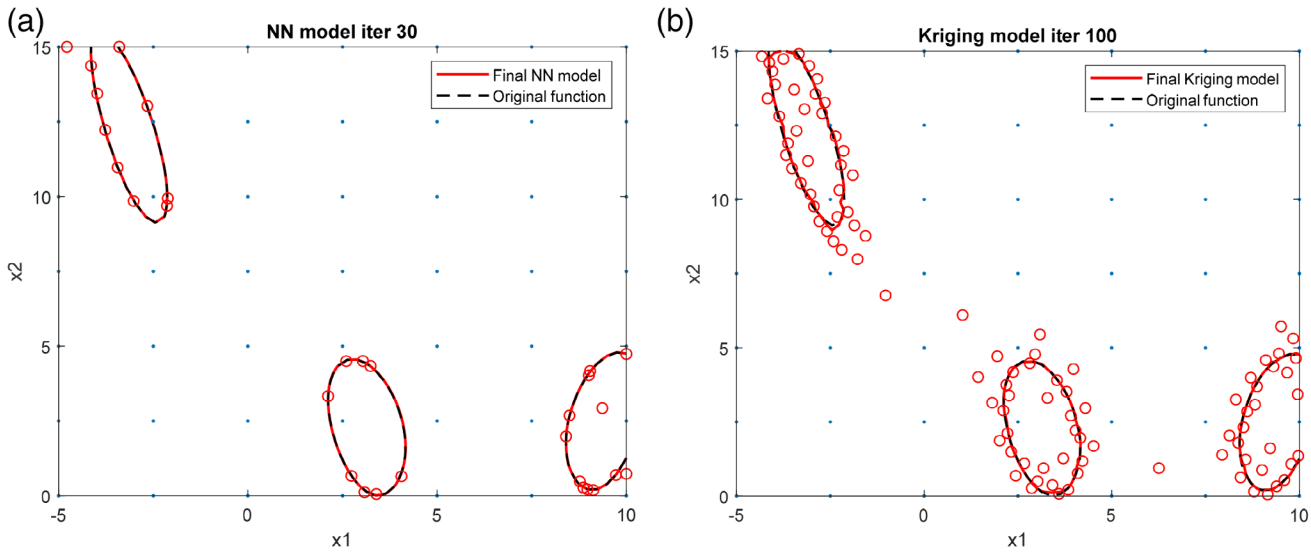

FIGURE 2 Visual comparison of the original model and the surrogate model feasible region boundaries (a) ANN for *branincon* and (b) Kriging for *branincon* (Initial samples are represented with dots and adaptive samples are represented with red circles. True feasible region boundary is represented as black dashed line and feasible region boundary predicted by the surrogate model is represented as red line). ANN, artificial neural network [Color figure can be viewed at wileyonlinelibrary.com]

For *camelback*, ANN performs significantly better in accurately identifying the feasible regions. The Kriging model over-estimates the feasible region indicated by NC of 9.63% compared to the ANN model that is very conservative in estimating the feasible region, as indicated by NC 0% This is also achieved by ANN with fewer adaptive samples. Figures 4a,b show feasible region boundaries identified by the ANN and the Kriging surrogate models respectively along with the placement of adaptive samples, which shows very good prediction of the feasible region boundaries by ANN.

For *sasenacon*, significantly larger feasible region is identified by ANN (ANN: 96.54 CF% and Kriging: 88.14 CF%), which is also accompanied by slight over-estimation when compared to Kriging (ANN: 4.32 NC% and Kriging: 2.73 NC%). Figure 5a,b show performance of ANN and Kriging methods respectively for this test problem. It can be observed that in the ANN method, the adaptive samples added are distributed between the larger and smaller feasible regions whereas in the Kriging method, the adaptive samples are concentrated more towards the larger feasible region. It is also worth noting that the ANN model requires a smaller number of adaptive samples for the *branincon* and the *camelback* test problems when compared to the *sasenacon* and the *ex3* test problems. The *branincon* and the *camelback* test problems constitute 1 non-linear constraint whereas the *sasenacon* and the *ex3* test problems are formulated using 2 non-linear and 1 linear constraint, as listed in Appendix A1. The complexity of the *sasenacon* and the *ex3* test problems from the additional constraints explains the need for more adaptive samples to accurately identify the feasible region.

The results discussed thus far include surrogate models built with 49 initial samples. In order to show the effectiveness of the proposed ANN based algorithm, results for the four test problems using

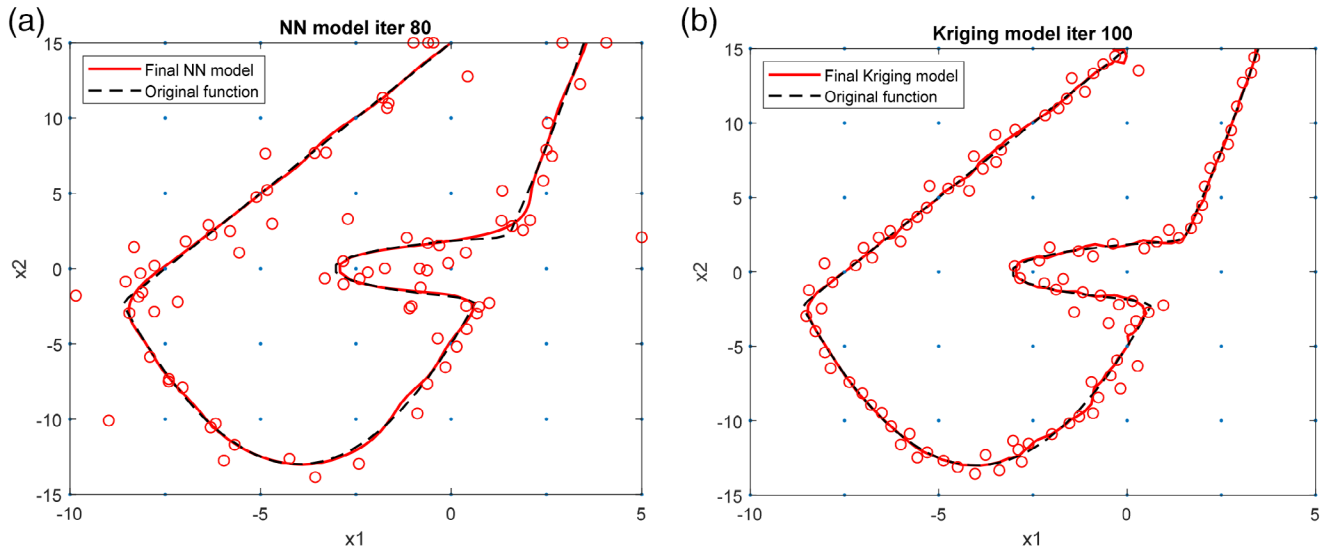

FIGURE 3 Visual comparison of the original model and the surrogate model feasible region boundaries (1) ANN for *ex3* (b) Kriging for *ex3* (Refer to the caption of Figure 2 for description of markers, lines and colors). ANN, artificial neural network [Color figure can be viewed at wileyonlinelibrary.com]

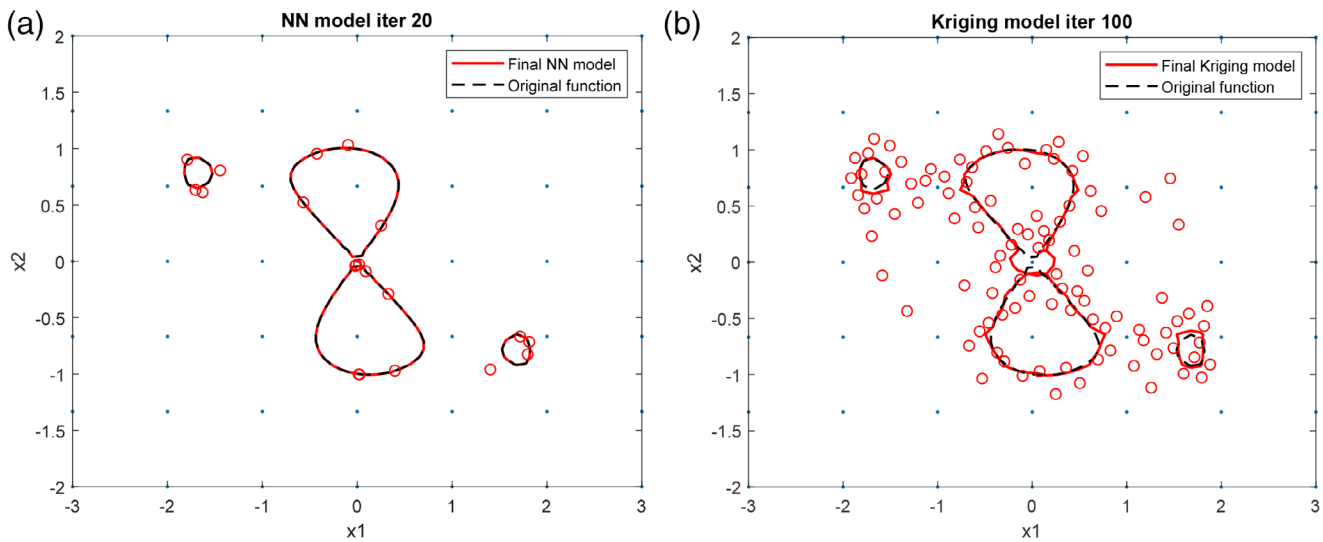

FIGURE 4 Visual comparison of the original model and the surrogate model feasible region boundaries (a) ANN for *camelback* and (b) Kriging for *camelback* (Refer to the caption of Figure 2 for description of markers, lines and colors). ANN, artificial neural network [Color figure can be viewed at wileyonlinelibrary.com]

different number of initial samples are plotted in Figure 6. To aid plotting of all the metrics that is, CF%, CIF% and NC% on the same scale, NC' = 100-NC is plotted instead of NC. A low value of NC indicates accurate identification of the feasible region. This implies a high value of NC' is desired. Figure 6 also shows the initial model accuracy and an improvement or degeneration in the accuracy attained due to the adaptive samples. From the plots, it can be observed that the accuracy of the ANN model is comparable or better than the Kriging model for the test problems at various levels considered. This is reflected in the CF%, CIF% and NC%. It is also worth noting that the performance of ANN is better than Kriging irrespective of the accuracy of the initial model. For branincon, at 25 initial samples (levels = 5), the initial ANN model accuracy reflected in CF% is significantly lower than the initial Kriging model accuracy. However, with adaptive sampling, the accuracy of the final ANN model is better than that obtained from Kriging. Similarly, for ex3 with 25 initial samples (levels = 5), the initial ANN model CF% is lower than initial Kriging model CF%. However, the final model metrics are comparable for both methods. Another interesting observation is the loss in model accuracy in the Kriging based approach as seen from the negative accuracy gain or the loss of accuracy. This is possibly due to interpolating nature of the Kriging model which may lead to oscillations with large number of closely located samples.

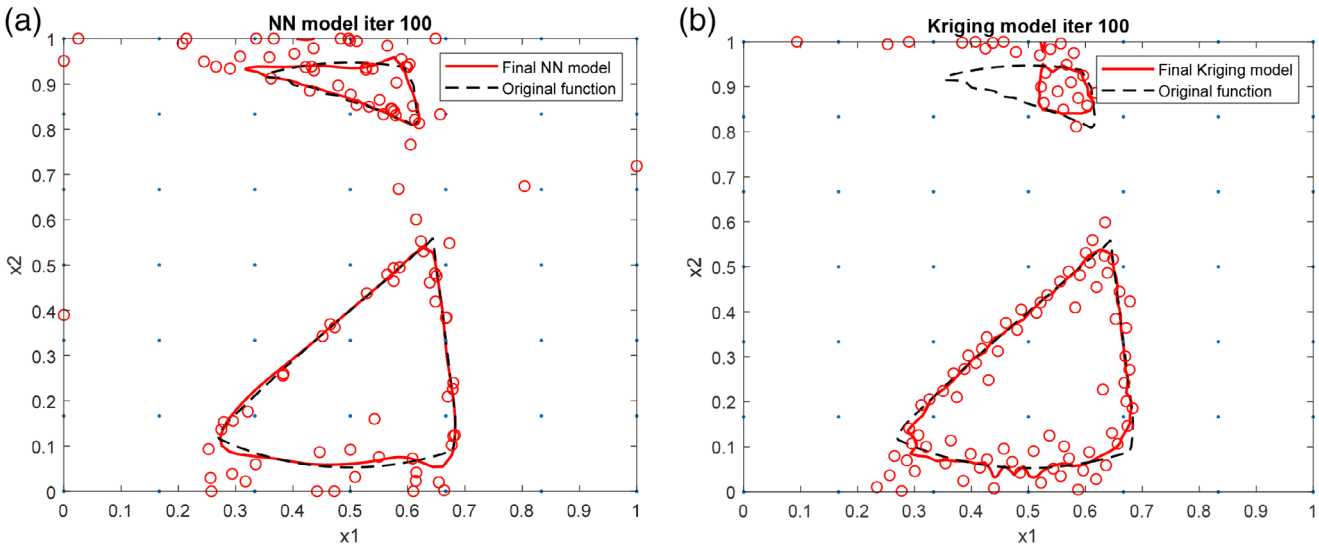

FIGURE 5 Visual comparison of the original model and the surrogate model feasible region boundaries (a) ANN for *sasenacon* and (b) Kriging for *sasenacon* (Refer to the caption of Figure 2 for description of markers, lines and colors). ANN, artificial neural network [Color figure can be viewed at wileyonlinelibrary.com]

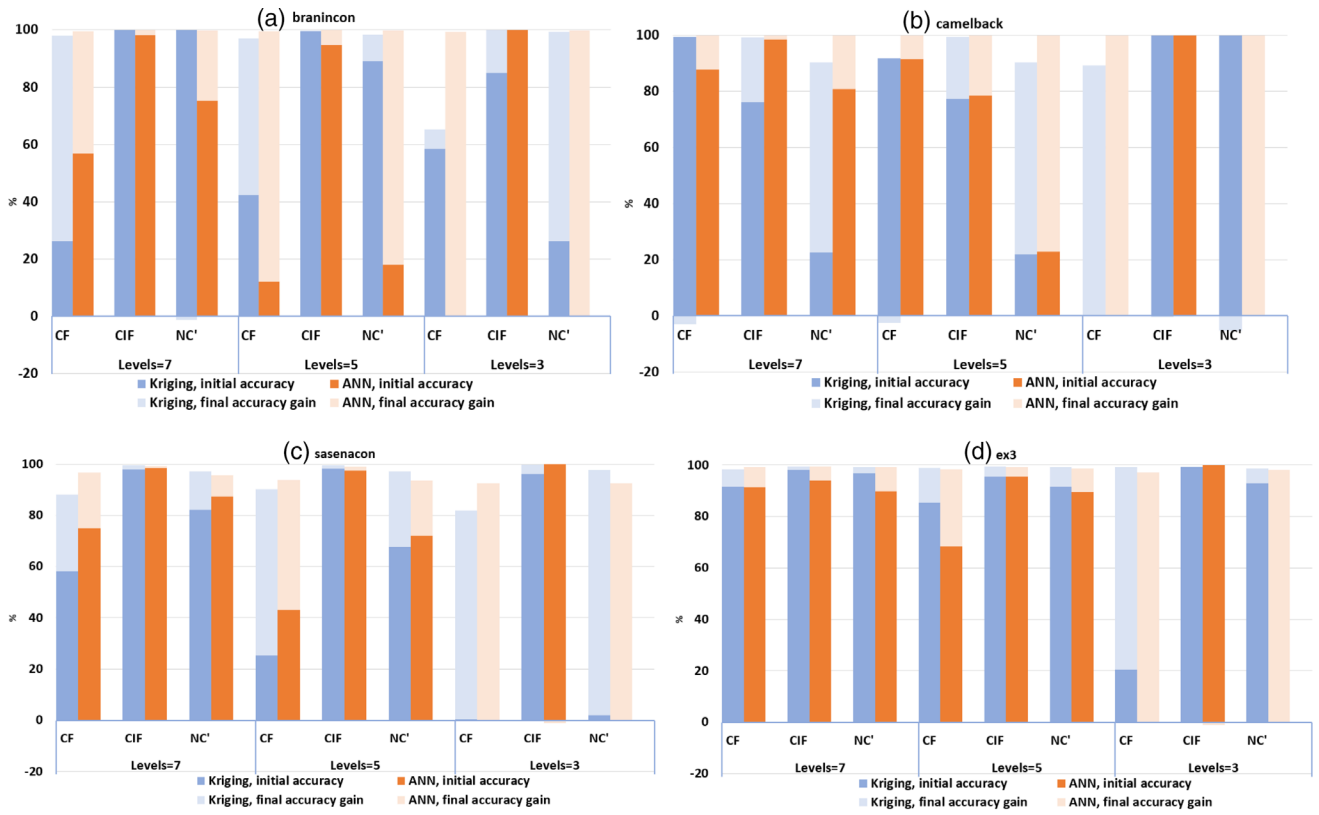

FIGURE 6 Comparison of ANN and Kriging model accuracy metrics for different levels of initial samples (a) *branincon*, (b) *camelback* (c) *sasenacon*, and (d) *ex3*. ANN, artificial neural network [Color figure can be viewed at wileyonlinelibrary.com]

The results listed in Table 1 show that generally ANN requires fewer adaptive samples compared to Kriging to attain a surrogate model accuracy of >99% CF, >99% CIF and < 1% NC. Specifically, Kriging utilizes the sampling budget of 100 adaptive samples whereas ANN requires considerably less number of adaptive samples for *branincon* (30 samples) and *camelback* (20 samples). For these test problems, this is also true when lower number of initial samples are used. The total number of adaptive samples used by the test problems

at different levels of initial samples are plotted in Figure 7. The plot shows that for *branincon*, at levels 5 and 3, ANN requires 30 and 40 adaptive samples respectively, whereas Kriging uses the entire sampling budget of 100 adaptive samples. Similarly, for camelback, Kriging uses 100 adaptive samples whereas ANN requires much lesser number of adaptive samples. In the *sasenacon* problem, Kriging as well as ANN require 100 adaptive samples at all the levels considered. Overall, the proposed ANN methodology uses the same or a smaller number of adaptive samples when compared to the Kriging methodology at various levels of initial samples considered. This is particularly useful for implementation of the methodology in identifying feasible regions for computationally expensive problems.

To take a closer look at the performance of the methods, the accuracy metrics of the ANN and Kriging surrogate models are also computed every 10 iterations. This also aids in evaluating the effect of adaptive sampling on the metrics. The evolution of the CF%, CIF% and NC% for the *camelback* and *ex3* test problems are plotted in Figures 8 and 9. Similar plots for the *branincon* and *sasenacon* test problems are added in Appendix A3. Generally, it may be observed that the metrics improve with the addition of adaptive samples that is, CF %, CIF% increase and NC% decreases. There may be oscillations in the trends which is the result of improvement in ability of the models to predict feasible regions as more samples are added. For example, in the Kriging based implementation for the *camelback* problem the initial model predicts a large portion of the variable space as feasible, as noted by the high CF%. However, this is an over-estimation of the feasible region which is reflected in the high initial NC% as well. As samples are added, CF% as well as NC% decrease that is, the model recognizes infeasible regions as well. After 20 iterations, the CF% begins to improve and the over-estimation of the feasible region is reduced as well.

In addition, one may observe slightly more oscillation in trends with the ANN based method when compared to the Kriging method, as seen in the *ex3* test problem. The oscillations in ANN based method can also be attributed to randomized nature of the ANN that is, the optimized ANN network parameters depend on the initial values of weights and biases which are randomly chosen in contrast to the deterministic nature of model parameters in the Kriging model. Overall, for two dimensional problems, the ANN based method performs better than the Kriging based method in identifying feasible regions, and with fewer adaptive samples.

3.2 | Higher dimensional test problems

In this section, three, five and six dimensional standard test problems namely *qcp4con*, *g4con* and *Hesse*[32] respectively are discussed. The test problems are tabulated in Appendix B1. Similar to the two dimensional problems, the variable bounds and constraints are listed in this table. In addition, two formulated test problems namely *4Dcomp* and *6Dcomp* are also used in this work. The *4Dcomp* is a two-dimensional test problem, whereas as the *6Dcomp* is a six dimensional test problem. The purpose of formulating the additional test problems is to simulate cases similar to a flowsheet model. Flowsheet models are mathematical representations of continuous processes and are used to simulate dynamic behavior of the entire manufactuirng line.[33] Several unit operation models connected such that relevant information is transferred from a unit to the the succeeding unit. A flowsheet model representing the wet granulation continuous manufacturing process is the subject of case study described in the Section 3.4. Flowsheet models are computationally expensive and hence serve as excellent candidates for implementing the proposed surrogate based feasibility analysis method for identification of feasible regions. To test effectiveness of the proposed methodology in flowsheet models, the test problems *4Dcomp* and *6Dcomp* are formulated. The test problem *4Dcomp* is formulated using the *sasenacon* and *qcp4con* functions as shown in Figure 10a. This leads to a four dimensional problem with six constraints. The test problem *6Dcomp* is formulated using the

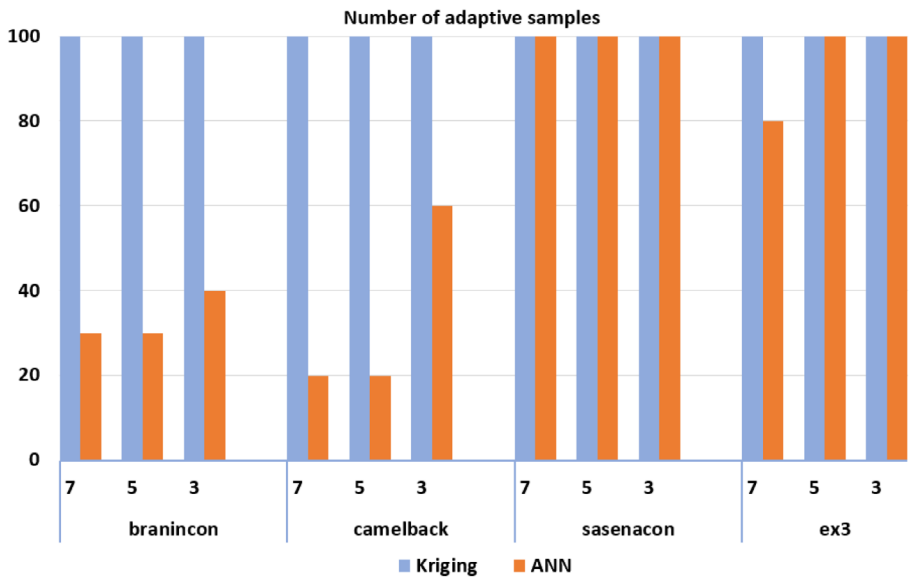

FIGURE 7 Comparison of number of adaptive samples required by ANN and Kriging based methodologies for different levels of initial samples. ANN, artificial neural network [Color figure can be viewed at wileyonlinelibrary.com]

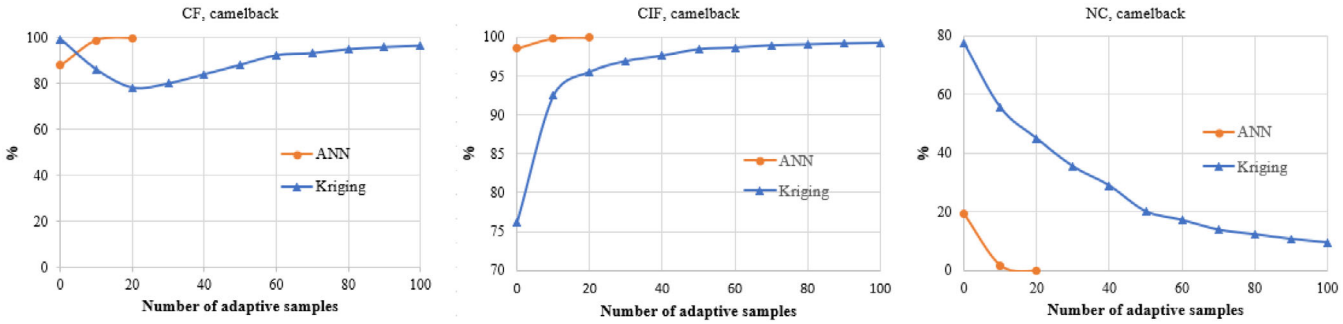

FIGURE 8 Comparison of ANN and Kriging model accuracy validation metrics by number of adaptive samples for camelback (a) CF%, (b) CIF %, and (c) NC%. ANN, artificial neural network; CF%, correctly identified points in feasible region; CIF%, percentage of correctly identified points in infeasible region; NC, percentage of points in not conservative feasible region [Color figure can be viewed at wileyonlinelibrary.com]

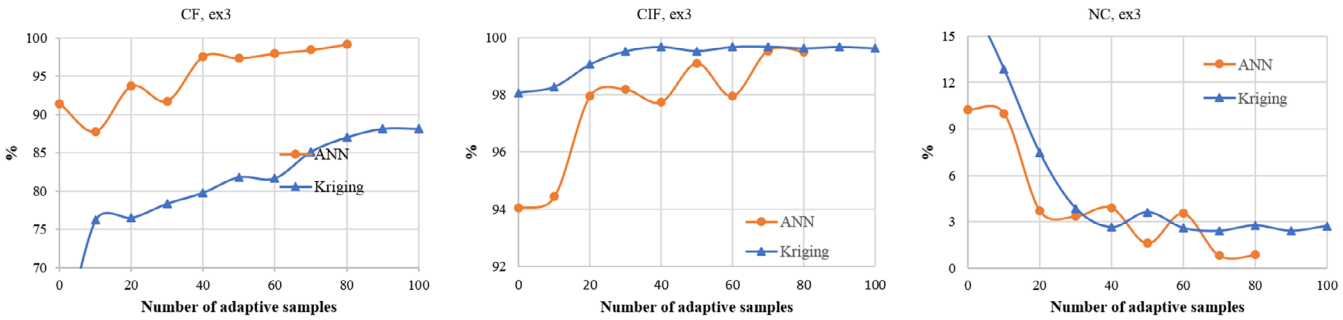

FIGURE 9 Comparison of ANN and Kriging model accuracy validation metrics by number of adaptive samples for ex3 (a) CF%, (b) CIF%, and (c) NC%. ANN, artificial neural network; CF%, correctly identified points in feasible region; CIF%, percentage of correctly identified points in infeasible region; NC, percentage of points in not conservative feasible region [Color figure can be viewed at wileyonlinelibrary.com]

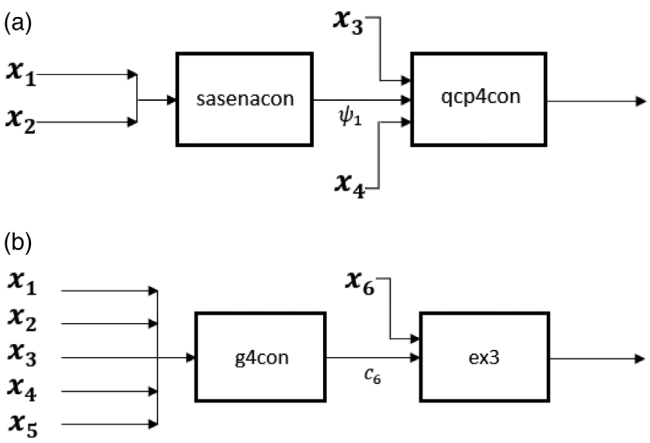

FIGURE 10 High dimensional test problems formulated (a) *4Dcomp* and (b) *6Dcomp*

g4con and *ex3* functions as shown in Figure 10b. This leads to a six dimensional problem with nine constraints. The performance of the ANN based and Kriging based methodologies for these test problems are compared through the validation metrics explained in Section 2.3.

To build the initial surrogate model in the model selection phase, Latin hypercube sampling scheme is used as it is known to provide better space filling property than grid sampling for high dimensional domain.[8,10] If d is the dimension of the variable domain, 2^d samples are chosen to fit the initial surrogate model. For the three-dimensional problem, 100 adaptive samples are used. For problems with four or higher dimensions, exhaustion of 500 adaptive samples is chosen as the stopping criteria. A higher number of adaptive samples are chosen when compared to the two-dimensional test problems as high dimensional problems are considered difficult to solve and are known to require large number of samples (*curse of dimensionality*). To identify adaptive samples to be added at each iteration, a local solver *fmincon* in MATLAB is used to maximize the EI_{feas} function. Similar to two dimensional problems, the EI_{feas} function is evaluated using 10,000 points in the variable space to obtain a good initial guess. The point with the largest EI_{feas} value is selected as the initial guess to find one adaptive sample per iteration. After addition of the adaptive samples, the accuracy of the surrogate models is evaluated using a test sample set in the variable space. For three, four, five and six dimensional problems, 20^3, 7^4, 6^5, and 5^6 points respectively are chosen through grid sampling. Table 2 lists the initial surrogate model and final surrogate model accuracy metrics for the ANN and the Kriging based methods. In addition, the number of hidden neurons used in the ANN model is also tabulated.

From the results, it is evident that the ANN based method performs better than Kriging in identifying feasible regions. ANN model yields good performance with the test sample set even when the

TABLE 2 Surrogate model accuracy metrics for high dimensional test problems

			Initial			Final				
d	Test problem	Surrogate model	CF%	CIF%	NC%	CF%	CIF%	NC%	Adaptive samples	Number of neurons hidden
3	qcp4con	Kriging	94.29	98.36	16.06	98.33	99.82	1.96	100	–
		ANN	94.08	99.35	7.19	98.89	99.50	5.39	100	19
4	4Dcomp	Kriging	56.35	76.36	74.10	86.23	99.00	7.39	500	–
		ANN	55.05	79.94	71.31	90.88	99.38	4.45	500	7
5	g4con	Kriging	83.49	95.85	12.40	97.92	99.22	2.22	500	–
		ANN	0.00	100.00	NA	99.26	99.67	0.94	500	35
6	Hesse	Kriging	74.84	87.12	48.24	96.11	99.87	0.79	500	–
		ANN	45.89	86.23	62.96	96.56	98.86	5.76	500	17
6	6Dcomp	Kriging	15.04	95.58	66.23	76.61	99.01	7.97	500	–
		ANN	3.44	98.94	67.29	90.52	98.04	11.73	500	21

Abbreviation: ANN, artificial neural network; CF%, correctly identified points in feasible region; CIF%, percentage of correctly identified points in infeasible region; NC, percentage of points in not conservative feasible region.

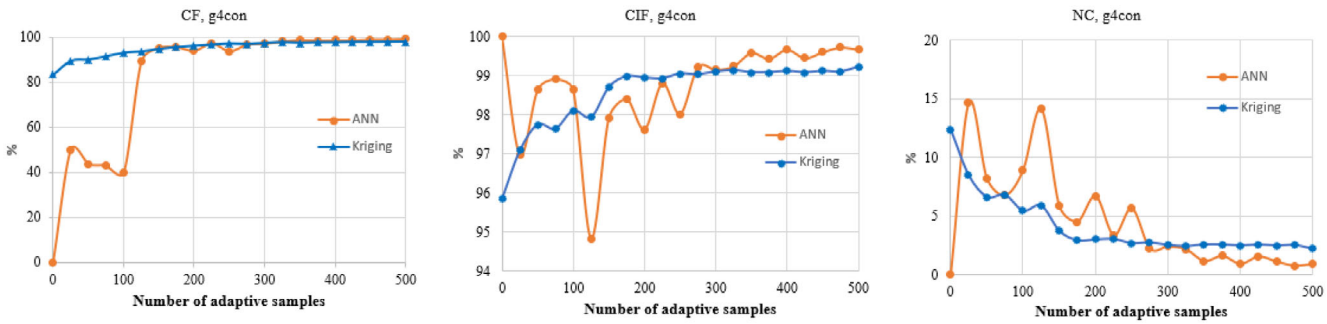

FIGURE 11 Comparison of ANN and Kriging model accuracy validation metrics by number of adaptive samples for g4con (a) CF%, (b) CIF%, and (c) NC%. ANN, artificial neural network; CF%, correctly identified points in feasible region; CIF%, percentage of correctly identified points in infeasible region; NC, percentage of points in not conservative feasible region [Color figure can be viewed at wileyonlinelibrary.com]

initial surrogate model accuracy is considerably poorer than the initial Kriging model. For example, in the five-dimensional *g4con* test problem, the initial Kriging model identifies the feasible points reasonably well. The initial ANN model for this test problem is unable to identify any of the feasible regions. However, with 500 adaptive sample points, the performance of the ANN model improves significantly, and the performance of the final Kriging model is worse than the final ANN model. In the *6Dcomp* problem, the initial ANN model performance is very poor compared to the initial Kriging model. However, with addition of 500 adaptive samples, the final ANN model is able to identify significantly more feasible points when compared to Kriging. However, this is also accompanied by an over-estimation of the feasible region by the ANN model.

In the *4Dcomp* problem, the initial ANN and Kriging surrogate model performance metrics are comparable. At the end of 500 adaptive samples, the final ANN model accuracy metrics are better than Kriging in the identification of feasible region as well as low over-estimation. For the six-dimensional *Hesse* problem, a lower initial ANN model performance is significantly improved through adaptive sampling. However, an over-estimation of the feasible region is observed for this case when compared to the Kriging model.

The ANN and the Kriging based methods are also compared through computation of the surrogate model accuracy metrics every 25 iterations to closely observe the effect of adaptive sampling. Figures 11 and 12 show trends of the surrogate model accuracy metrics every 25 iterations for *g4con and 6Dcomp* respectively. Similar plots for the *qcp4con, 4Dcomp and Hesse* test problems are provided in Appendix B2. Similar to the two-dimensional test problems, some oscillations are seen in the trends for Kriging and ANN models. The oscillations observed in the first few iterations may be attributed to the unreliability of the surrogate model. The initial surrogate model is unable to identify the feasible region as seen from the poor initial CF % metrics in the *g4con* and *Hesse* problems. In the first few iterations adaptive sampling explores the variable domain landscape to eventually identify more feasible regions. This is particularly noticeable in the oscillations with first few iterations in the *g4con* test problem. From here on, the adaptive sampling proceeds with exploitation phase where the focus is on identifying the feasible region boundary with

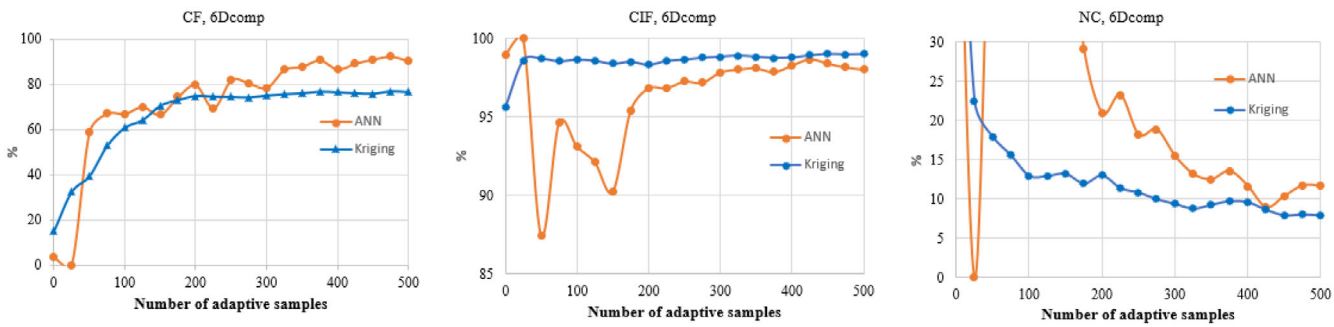

FIGURE 12 Comparison of ANN and Kriging model accuracy validation metrics by number of adaptive samples for 6Dcomp (a) CF%, (b) CIF %, and (c) NC%. ANN, artificial neural network; CF%, correctly identified points in feasible region; CIF%, percentage of correctly identified points in infeasible region; NC, percentage of points in not conservative feasible region [Color figure can be viewed at wileyonlinelibrary.com]

more accuracy. This also results in the oscillations being eventually dampened as more samples are added. Also, the oscillations generally are more for the ANN model which may be attributed to randomization effect of the ANN model building that is, initialization of the ANN network parameters being random in nature. The improvement in model performance with addition of adaptive samples is also seen in the *6Dcomp* problem (Figure 12) which serves as a test problem for a flowsheet feasibility analysis problem. Overall, the ANN based methodology performs comparable to better than the Kriging based method for the high dimensional test problems as well. The following section describes the implementation of the ANN based methodology on a four-dimensional problem of the continuous wet granulation process.

3.3 | Case study of a continuous wet granulation process

The purpose of the case study described in this section is to illustrate the effectiveness of the proposed ANN based methodolody in a manufacturing process problem. In this case study, a tablet manufacturing process model is used. The feasibility concept is applied to the process in steady state and no control variables, in order to understand the "design space" of the process. Design space is defined as "the multidimensional combination and interaction of input variables and process parameters that have been demonstrated to provide assurance of quality".[34] A tablet manufacturing process via continuous wet granulation is used in the pharmaceutical industry for the production of solid oral dosage products and consists of primarily a feeder, blender, twin screw granulator, fludized bed dryer, mill and tablet press. The formulation ingredients in powder form are blended and introduced as feed to a granulation unit. Liquid binder is added to the granulation unit, which aids in the production of uniform granules. The wet granules are dried in a fludized bed drying unit through contact with hot air for a certain duration. These granules are subsequently broken to smaller sizes in a comilling unit through the action of a rotating impeller. The smaller sized granules are compacted to produce tablets in a tablet press unit. A schematic of the continuous

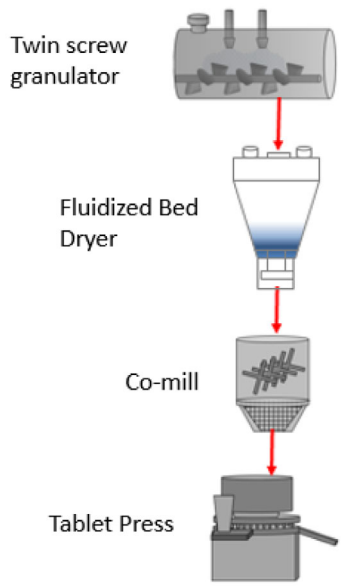

FIGURE 13 Schematic of a continuous wet granulation process [Color figure can be viewed at wileyonlinelibrary.com]

process used in this work is shown in Figure 13. The unit operations simulated by the process models are developed and validated with experimental data and are previously published by Metta et al.[33] They are briefly explained in Appendix C for the benefit of readers. The process models are used to develop a flowsheet model where relevant information is tranferred from a unit model to the downstream units. The flowsheet model is developed in gFormulatedProducts[35] and takes approximately 4.5 minutes to reach steady state.

3.3.1 | Feasibility problem for the continuous wet granulation process

The flowsheet model simulating the continuous wet granulation process can be used to capture and study the effects of several process variables. Techniques such as sensitivity analysis can be applied to the

TABLE 3 Variable bounds for the uncertain parameters θ used in the continuous wet granulation process feasibility case study

Factors	Units	Nominal	LB	UB
LS ratio	Kg/kg	0.12	0.114	0.126
Dryer air temperature	Deg C	42.1	40	44.2
Drying time	s	450	427.5	472.5
Mill impeller speed	Rpm	1,000	950	1,050

TABLE 4 Constraints for the continuous wet granulation process feasibility case study

Unit operation	Property	Units	Variation allowed based on % of nominal value	Multiplication factors used
Granulator	Mean d10, d50, d90	μm	+/− 10%	0.1, 0.01, 0.01
Dryer	Mean d10, d50, d90	μm	+/− 10%	0.1, 0.1, 0.01
	Moisture content	%	+/− 10%	10
Comill	Mean d10, d50, d90	μm	+/− 10%	0.1, 0.1, 0.1
	Mean bulk density	Kg/m^3	+/− 10%	0.1
	Mean tapped density	Kg/m^3	+/− 10%	0.1
Tablet press	Hardness	kP	+/− 5%	1

flowsheet model for identifying the critical process parameters. The feasible region is then identified considering the chosen critical process parameters as the variable space. For this work, the uncertain parameters used to define the feasibility problem are listed in Table 3. These factors are chosen based on the results of sensitivity analysis performed and published in Metta et al.[33] Specifically, the factors liquid to solid ratio (LS ratio) for the granulator, air temperature ($Temp_{dryer}$), drying time ($Time_{dryer}$) for the fludized bed dryer and mill impeller speed (RPM_{Mill}) are considered. The bounds on these variables, considered as +/− 5% of the operating values are also listed in the table. The process constraints that define the feasible region are listed in Table 4. Constraints related to the quality of the granules such as size (d_{10}, d_{50}, d_{90}), moisture content and density are considered. The final tablet hardness value is allowed to vary upto 5% that is, (84.9 N-93.8 N) from the targeted value of 89.4 N. In addition, it is general practice to ensure the response variables used in the problem are in the same order of magnitude. Since, various response variables such as particle sizes, densities and hardness are considered in this problem, appropriate multiplication factors are used to tranform them into variables in the same order of magnitude. The multiplication factors used are also provided in Table 4. The feasibility analysis problem thus formulated is solved using the proposed ANN based feasibility analysis methdology. The problem is four-dimensional and has 13 black-box constraints. 2^4 samples from a Latin hypercube sampling scheme are used to build the initial ANN model. As adaptive sampling points are identified, the flowsheet model is called to the MATLAB environment using gOMATLAB. A stopping criteria of 300 adaptive samples is chosen considering the computational expense of the flowhseet model. Due to the high computational expense of the flowsheet model, only 4^4 (=256) grid samples are run in order to be used as validation points. The feasibility analysis algorithm thus implemented takes approximately 36 hr to complete using 3 cores. After exhaustion of the sampling budget, the final ANN model the feasibility metrics are 98.75 CF%, 98.29 CIF% and 3.65 NC%. Since this is a high dimensional problem, the feasible region is represented using a matrix of contour plots of feasibility function values shown in Figure 14. For each contour plot, only two factors are varied and the remaining two factors are set at operating values. The feasible region boundary predicted by the final ANN model is represented using a red line. To aid visual analysis of the accuracy of the surrogate model, feasible region boundary from the original flowsheet model for $Time_{dryer}$ and $Temp_{dryer}$ contour plot is also plotted using green dashed line. Considering the computational expense of the process model, this comparision is made only for $Time_{dryer}$ vs $Temp_{dryer}$ plot. The true feasible region boundary is identified using 20^2 runs of the original process model with only $Time_{dryer}$ and $Temp_{dryer}$ as the variables. It is evident that the feasible region boundary predicted by the ANN model agrees very well with the feasible region boundary per the original flowsheet model. In addition, a better udnerstanding of the feasible region can be achieved by observing the matrix of contour plots. The process is feasible for the range of mill impeller speed considered. Dryer temperature is an important factor that needs to be controlled well for the process to remain in fesible region. Also, the drying time needs to set based on the dryer air temperature and LS ratio in the granulator. The ranges of feasible operation can thus be easily obtained and understood. In addition, the ANN surrogate model is computationally inexpensive and can be used to test feasibility of operation at desired combinations of the variables.

4 | CONCLUSIONS

In this work, a novel ANN based methodology using adaptive sampling is proposed to identify feasible regions of processes treated as a

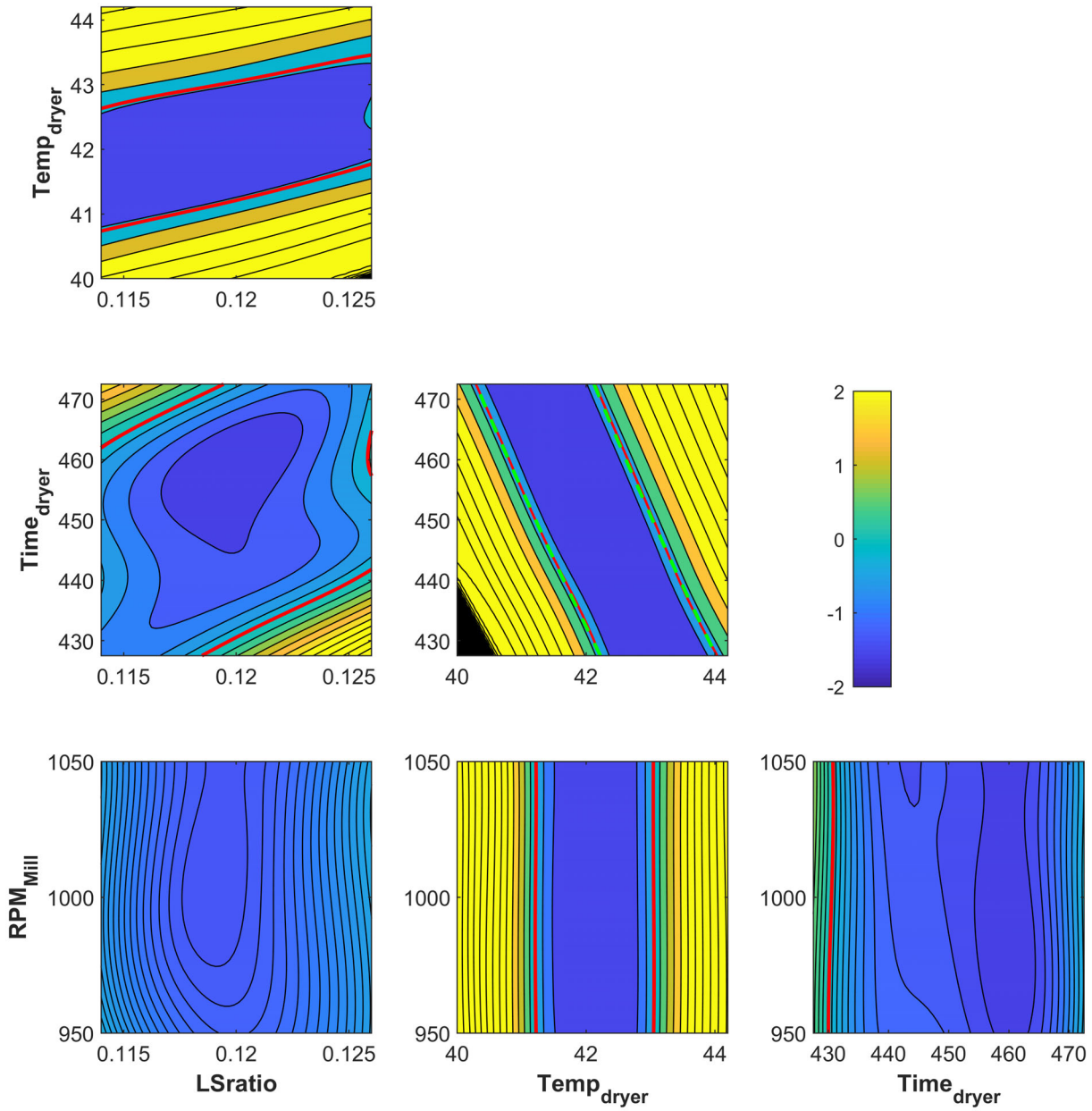

FIGURE 14 Matrix of feasibility function contour plots representing the feasible region of the continuous wet granulation process case study (Feasible region boundary from the original model and predicted by the ANN model are represented as green dashed line and red line respectively). ANN, artificial neural network [Color figure can be viewed at wileyonlinelibrary.com]

black-box. The adaptive samples are identified through the maximization of a modified Expected Improvement function. This requires estimation of the ANN model prediction variance, which is obtained using a statistical approach, Jack-knifing. The adaptive sampling allows exploration of the variable space to identify feasible regions and exploitation of the regions to identify the boundaries more accurately. The proposed ANN based methdology performs better than the Kriging based approach in identifying the feasible region boundaries for two dimensional as well as high dimensional problems. For low dimensional problems, the ANN based methdology requires lesser or comparable number of adaptive samples than the Kriging based methdology. The accuracy of the identified feasible region boundary is also better for the tested low dimensional problems. For high dimensional problems, the ANN based methodology performed better than the Kriging based approach for classifying feasible points. The superior performance of the ANN model may be attributed to its use of many parameters depending on the number of neurons used in the hidden layer when compared to the Kriging model. While this may lead to over-fitting issues, as is known to occur in machine learning models, this is also the model's strength and is the reason why ANN performs better than Kriging in several cases.

The effectiveness of the ANN based methodology is also tested to a realistic pharmaceutical process of a continuous wet granulation that is simulated using a computationally expensive model where

constraints are not available in closed form. The proposed ANN based methdology accurately identifies the feasible region boundary with limited number of samples. Currently, the proposed methdology uses a sampling budget as the stopping criteria for the algorithm. Future work needs to focus on determining alternative stoppping criterion that is based on model performance. Another area of research that needs focus is the effect of the number of hidden neurons on the methodology. Currently, the number of hidden neurons are determined using an initial sample set. Use of adaptive sampling strategies that also update the number of hidden neurons need to be explored.[21] In addition, current work is tested for determinstic scenarios that is, the original process model does not consider probabilistic uncertainties that may be associated with the model parameters. In stochastic scenarios, several model evaluations may be needed to map the feasible region boundaries as well as variance in the process responses. Further work is required to evaluate the applicability of the ANN model in such cases. The authors would also like to point that, in this article, feasibility is defined for processes with no control variables. The main goal was to understand feasibility while addressing the high computational of model evaluations. From the authors' knowledge there is very limited work on understanding feasibility of high cost models in the presence of control variables. This is a very challenging problem and solving the problem with control variables is a research area that merits further exploration.

ACKNOWLEDGMENTS

This work is supported by the US Food and Drug Administration (FDA), through grant 11695471 and 1U01FD006487-01, and a Consortium Agreement between Janssen Pharmaceutica, University of Ghent and Rutgers University. The authors gratefully acknowledge academic licenses from Process Systems Enterprise.

ORCID

Nirupaplava Metta https://orcid.org/0000-0001-8942-4982

REFERENCES

1. Grossmann IE, Morari M. *Operability, Resiliency, and Flexibility: Process Design Objectives for a Changing World*. Carnegie-Mellon University: Design Research Center; 1983.
2. Halemane KP, Grossmann IE. Optimal process design under uncertainty. *AIChE J*. 1983;29(3):425-433.
3. Grossmann IE, Floudas CA. Active constraint strategy for flexibility analysis in chemical processes. *Comput Chem Eng*. 1987;11(6):675-693.
4. Goyal V, Ierapetritou MG. Determination of operability limits using simplicial approximation. *AIChE J*. 2002;48(12):2902-2909.
5. Banerjee I, Ierapetritou MG. Feasibility evaluation of nonconvex systems using shape reconstruction techniques. *Indus Eng Chem Res*. 2005;44(10):3638-3647.
6. Laky D, Xu S, Rodriguez JS, Vaidyaraman S, García Muñoz S, Laird C. An optimization-based framework to define the probabilistic design space of pharmaceutical processes with model uncertainty. *Processes*. 2019;7(2):96.
7. Banerjee I, Pal S, Maiti S. Computationally efficient black-box modeling for feasibility analysis. *Comput Chem Eng*. 2010;34(9):1515-1521.
8. Kucherenko S, Giamalakis D, Shah N, García-Muñoz S. Computationally efficient identification of probabilistic design spaces through application of metamodeling and adaptive sampling. *Comput Chem Eng*. 2020;132:106608.
9. Rogers A, Ierapetritou M. Feasibility and flexibility analysis of black-box processes part 1: surrogate-based feasibility analysis. *Chem Eng Sci*. 2015;137:986-1004.
10. Wang Z, Ierapetritou M. A novel feasibility analysis method for black-box processes using a radial basis function adaptive sampling approach. *Aiche J*. 2017;63(2):532-550.
11. Leperi KT, Yancy-Caballero D, Snurr RQ, You F. 110th anniversary: surrogate models based on artificial neural networks to simulate and optimize pressure swing adsorption cycles for CO2 capture. *Indus Eng Chem Res*. 2019;58(39):18241-18252.
12. Fernandes FA. Optimization of Fischer-Tropsch synthesis using neural networks. *Chem Eng Technol: Indus Chem-Plant Equipment-Process Eng-Biotechnol*. 2006;29(4):449-453.
13. Smith JD, Neto AA, Cremaschi S, Crunkleton DW. CFD-based optimization of a flooded bed algae bioreactor. *Indus Eng Chem Res*. 2012;52(22):7181-7188.
14. Agatonovic-Kustrin S, Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *J Pharm Biomed Anal*. 2000;22(5):717-727.
15. Chambers M, Mount-Campbell CA. Process optimization via neural network metamodeling. *Int J Production Economics*. 2002;79(2):93-100.
16. Meireles MRG, Almeida PEM, Simoes MG. A comprehensive review for industrial applicability of artificial neural networks. *IEEE Transac Indus Electronics*. 2003;50(3):585-601.
17. Kleijnen JPC, WCMv B. Application-driven sequential designs for simulation experiments: kriging metamodelling. *J Operational Res Soc*. 2004;55(8):876-883.
18. Boukouvala F, Ierapetritou MG. Feasibility analysis of black-box processes using an adaptive sampling kriging-based method. *Comput Chem Eng*. 2012;36:358-368.
19. Bhosekar A, Ierapetritou M. Modular design optimization using machine learning-based flexibility analysis. *J Process Control*. 2020;90:18-34.
20. Grossmann IE, Calfa BA, Garcia-Herreros P. Evolution of concepts and models for quantifying resiliency and flexibility of chemical processes. *Comput Chem Eng*. 2014;70:22-34.
21. Jin Y, Li J, Du W, Qian F. Adaptive sampling for surrogate modelling with artificial neural network and its application in an industrial cracking furnace. *Can J Chem Eng*. 2016;94(2):262-272.
22. Jones DR, Schonlau M, Welch WJ. Efficient global optimization of expensive black-box functions. *J Global Opt*. 1998;13(4):455-492.
23. Quiñonero-Candela J. Rasmussen CE. A unifying view of sparse approximate Gaussian process regression. *J Machine Learning Res*. 2005;6:1939-1959.
24. Liu H, Ong Y-S, Cai J. A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design. *Struct Multidisciplinary Opt*. 2018;57(1):393-416.
25. Sacks J, Welch WJ, Mitchell TJ, Wynn HP. Design and analysis of computer experiments. *Statistical Sci*. 1989;4(4):409-423.
26. Lophaven SN, Nielsen HB, Sondergaard J, Dace A. A matlab kriging toolbox. Technical University of Denmark, Kongens Lyngby, Technical Report No IMM-TR-2002-12. 2002.
27. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Cognitive Modeling*. 1988;5(3):1.
28. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Netw*. 1989;2(5):359-366.
29. Burden F, Winkler D. Bayesian regularization of neural networks. *Artificial Neural Networks*. Silver Spring, MD, USA: Springer; 2008:23-42.
30. Eason J, Cremaschi S. Adaptive sequential sampling for surrogate model generation with artificial neural networks. *Comput Chem Eng*. 2014;68:220-232.
31. Sasena MJ, Papalambros P, Goovaerts P. Exploration of metamodeling sampling criteria for constrained global optimization. *Eng Opt*. 2002;34(3):263-278.

32. Amine Bouhlel M, Bartoli N, Regis RG, Otsmane A, Morlier J. Efficient global optimization for high-dimensional constrained problems by using the kriging models combined with the partial least squares method. *Eng Opt*. 2018;50(12):2038-2053.
33. Metta N, Ghijs M, Schäfer E, et al. Dynamic flowsheet model development and sensitivity analysis of a continuous pharmaceutical tablet manufacturing process using the wet granulation route. *Processes*. 2019;7(4):234.
34. FDA U. *Guidance for Industry. q8 (r2) Pharmaceutical Development*. Silver Spring, MD, USA: Department of Human and Human Services; 2009.
35. *gFormulatedProducts* [computer program]. 2019.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Metta N, Ramachandran R, Ierapetritou M. A novel adaptive sampling based methodology for feasible region identification of compute intensive models using artificial neural network. *AIChE J*. 2020;e17095. https://doi.org/10.1002/aic.17095