Minerals Engineering 170 (2021) 107026

Contents lists available at ScienceDirect

Minerals Engineering

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

Utilising a deep neural network as a surrogate model to approximate phenomenological models of a comminution circuit for faster simulations

Edwin J.Y. Koh [a,b,*], Eiman Amini [a], Geoffrey J. McLachlan [b], Nick Beaton [a]

[a] Cooperative Research Centre for Optimising Research Extraction (CRC ORE), Brisbane, Queensland 4069, Australia
[b] School of Mathematics and Physics, University of Queensland, Brisbane, Queensland 4072, Australia

ARTICLE INFO

Keywords:
Simulation and modelling
Surrogate model
Comminution
Neural networks
Evolutionary algorithm

ABSTRACT

Comminution circuits can be modelled using phenomenological models of individual unit operations to represent the operation performance. Process optimisation and block model variability analysis require millions of simulations to fully explore process efficiency through mine scheduling which is costly and time consuming. Surrogate modelling is a technique used in process engineering approach of approximating the behaviour of the underlying model by using a model which is computationally more feasible. Neural networks are a form of machine learning which can approximate complicated function mappings of inputs to outputs and are computationally parallelisable.

In this study, a neural network is used as a surrogate model to approximate a copper porphyry mine comminution circuit for faster simulations. The neural network was trained to predict throughput using data generated from the phenomenological models of the comminution circuit. The optimal neural network hyperparameters were determined using an evolutionary algorithm to minimise overfitting. The neural network predicted simulation results 3363 times quicker than phenomenological models with errors of 0.37%, 0.55% and 0.45% across three different test sets. The neural network only required a stratified training sample of 1 in 1000 data to interpolate the rest of the data.

1. Introduction

Increased environmental pressures and lack of high-grade ore deposits have pressured mine operations to explore various processing configurations to increase process efficiency. Process efficiency can be improved through mine scheduling, process optimisation, and block model variability analysis. However, this requires evaluation of all possible scenarios to run a mine value chain which results in millions of simulations of the resource model to determine the optimal mining and processing strategy. Simulations can be done using phenomenological models as they contain physical fundamentals which helps for interpretability and extrapolation. However, solving multiple phenomenological models together requires numerical solver-based techniques which are computationally intensive. Therefore, there is a common goal to reduce the simulation time to further explore different mine scheduling, process configurations, and analyse block model variability for optimisation purposes.

A surrogate model or metamodel is an engineering approach to approximate the behaviour of an underlying complex model using a more computationally feasible model. A neural network is ideal for surrogate modelling as it can approximate mathematical functions using matrix operations which allows speed-up through parallel computing. The aim of this paper is to develop a novel methodology of using a neural network surrogate model to approximate phenomenological models in a comminution circuit to reduce the time required to perform mass simulations on the resource block model. Although this study uses data from a previous study generated with the Integrated Extraction Simulator (IES), the same methodology of surrogate modelling can be used for other plant simulation software and combination of process unit models.

The phenomenological models are calibrated to daily process information from site to capture actual operating strategies and limitations. Only a sample, n of the population block model, N was evaluated with the phenomenological models as the training dataset for the neural network to predict throughput. The surrogate model is used to evaluate the rest of the N-n blocks (known as generalisation) which would be quicker than evaluating the original phenomenological models. The final developed surrogate model would be able to approximate the results from the N blocks as well as interpolating the continuous input

* Corresponding author.
E-mail address: edwin.koh@uq.net.au (E.J.Y. Koh).

https://doi.org/10.1016/j.mineng.2021.107026
Received 28 September 2020; Received in revised form 7 June 2021; Accepted 8 June 2021
Available online 18 June 2021

space within the discrete training data steps of the original *N* simulations. The capability of continuous interpolation unlocks optimisation capabilities at a higher resolution than the discrete data steps used in the phenomenological models.

To achieve this aim, the paper has several objectives:

i) To determine the minimum number of training samples required to generalise a resource block model using a neural network,
ii) For this number of samples, what is the optimum sampling method as the blocks are not uniformly distributed,
iii) To develop an algorithm which can automatically design the optimal neural network hyperparameters given the training dataset,
iv) To verify that the surrogate model has successfully generalised the block model across all operating scenarios,
v) To decompose the neural network into elementary operations such that it is easily deployable on any platform,
vi) To use the neural network to interpolate at high resolution and perform mass simulations for mine optimisation.

This section explains the need for mine operations to perform mass simulations to optimise performance and how simulations are typically done in industry. The chosen site of study, Anglo American Los Bronces Confluencia circuit is described along with the possible operating strategies on-site. The site simulation scenarios results reported by Amini et al. (2020) which were used in this study is highlighted. The history of the surrogate model development in minerals processing is reviewed as well. Finally, neural networks are introduced in further depth as a method for surrogate modelling.

1.1. Mine optimisation

Optimisation of the comminution circuit has been a large area of research as it accounts for most of the energy usage in a mineral processing plant (Ballantyne et al., 2012). Performance of a mineral processing plant is often measured by key production indicators (KPIs) such as total value chain throughput and total metal recovery. Improving the performance of mineral processing plants at the same energy usage indicates an improvement in efficiency. Mine operations can maximise KPIs through:

i) Optimising mine scheduling. Determining the optimum sequence of extraction requires simulation of the entire block model from different extraction points.
ii) Optimising process configuration variables. Determining the best extraction method requires simulation of the blocks at all possible process configurations.
iii) Risk analysis of the block model variation. Accounting for the variability in the block model information requires analysis of the differences in results from simulations of block models.

Optimising the entire operation is an iterative process between these three components as they are dependent on each other. As more conditions at increasing resolution of variables are explored, the number of required simulations easily increases to the order of millions (Amini et al., 2020).

In industry, it is common to perform simulations through specialised commercial software like METSIM (Bartlett et al., 2014), MODSIM (Ford and King, 1984), USIM PAC (Durance et al., 1994), SysCAD (Razavi-maneseh et al., 2006), CEET (Kosick et al., 2001) and FLEET (Dobby et al., 2002), JKSimMet (Morrison and Richardson, 2002), JKSimFloat (Harris et al., 2002), or through proprietary equations set up in Microsoft Excel sheets. These were done on a local computer and limited to a magnitude of thousands of simulations in a long period of time due to hardware limitations and lack of fully automated process. This is also due to the use of solver-based algorithms to solve simultaneous equations to find local minima.

To overcome this issue, the Cooperative Research Centre for Optimising Resource Extraction (CRC ORE) developed cloud-based simulation software known as the Integration Extraction Simulator (IES). Standardising the models into a single platform allows simulation of the entire flowsheet which contains phenomenological models of individual unit operations like in an actual process. Then, process constraints of the site are integrated in the simulations to mimic plant behaviour. By developing a platform on the cloud, IES is capable of using distributed cloud computing to divide the task to multiple computers to evaluate simulations. Because of this, Amini, et al. (2020) showed the optimisation of the comminution circuit through 14 million simulations of phenomenological models within a week. Although feasible, this "brute-force" solution is costly if the millions of simulations are to be repeated with block model variations or updates. Therefore, there is a common desire for an alternative approach to reduce the simulation time to fully explore the different mine schedules, process configuration, and block model variations.

1.2. Site operating strategy

The Confluencia comminution circuit located at Los Bronces operation was developed and fine-tuned in IES (Amini et al., 2020). The flowsheet and the block model populated with the hardness values were used to evaluate several technologies for Anglo American to increase the operation's productivity and profitability. Fig. 1 shows the flowsheet developed based on the Confluencia comminution circuit represented by unit models in IES. The simulation data was used to develop and test the surrogate neural network model in this study.

The input variables for the flowsheet can be divided into two categories of variables:

1. Ore Characteristic Variables. These are variables which represent the ore hardness. The mine resource is modelled into 25 × 25 × 15 m^3 blocks which were characterised for **density**, SAG (Semi-Autogenous Grinding) Power Index (**SPI**), Bond Work Index (**BWI**) and Crusher Index (**CI**) and grade. The **A**, **b**, and $\mathbf{t_a}$ values are determined empirically from the block model hardness values through polynomial regression using the results of SPI and Drop Weight Test (DWT) work done on about 270 samples from the Los Bronces deposit. These hardness indicators would be crucial inputs to the surrogate neural network model as the SAG Mill and ball mills are modelled on them.
2. Flowsheet Condition Variables. These are variables which describe the chosen operation condition for that simulation. In this study, all the phenomenological models are run in the same setting, except the Grade Engineering screen size which is only valid when Grade Engineering is used. The study was for investigating whether Grade Engineering configurations or Mine to Mill would maximise throughput. Therefore, there are three types of operating conditions for this comminution circuit:

i) Grade Engineering Natural Deportment
ii) Grade Engineering Differential Blasting
iii) Mine to Mill

1.2.1. Grade Engineering

Grade Engineering (GE) is a range of operational strategies that exploits the grade variability to separate low grade material prior to grinding to increase efficiency (Carrasco et al., 2016). Depending on the ore characteristics, different blasting strategies can be used to improve the separation efficiency.

1.2.1.1. Grade Engineering Natural Deportment. Grade Engineering Natural Deportment (GE ND) refers to the 'natural' preference for metal

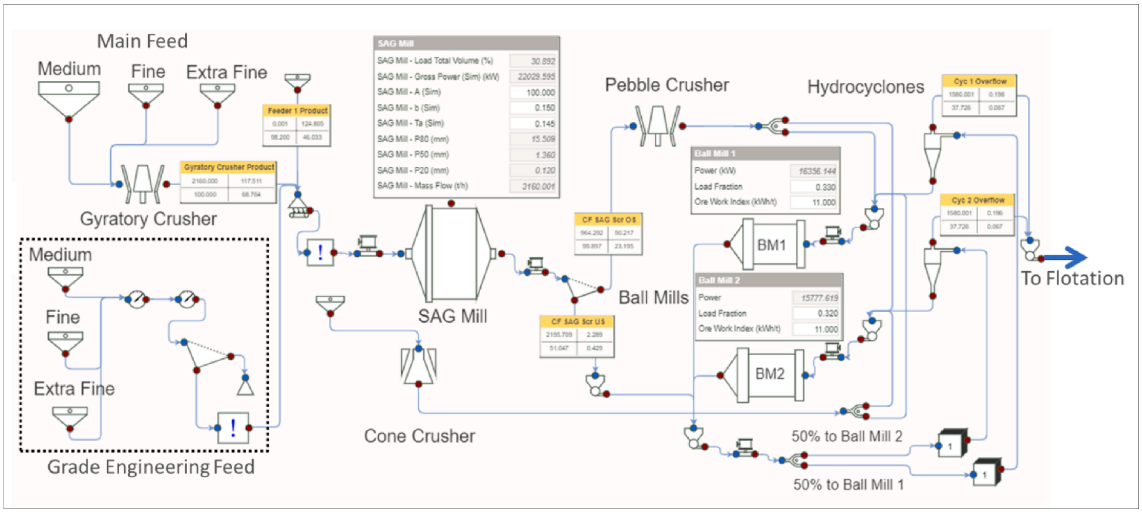

Fig. 1. Flowsheet summary in IES of the comminution circuit of the Anglo American Los Bronces Confluencia Las Tortolas.

proportion to preferentially deport into specific size fractions after breakage during blasting (Carrasco, 2013; Carrasco et al., 2014; Carrasco et al., 2016). These scenarios assessed the impact of blending upgraded GE screen undersize product with direct feed. Three GE feed blast fragmentations and three GE screen aperture sizes were applied in the simulations (nine scenarios). There was a total of 757,453 blocks that were identified suitable for the GE ND blasting strategy (Amini et al., 2020). The following are the simulated conditions in the study and hence the input variables for GE ND:

i) 3 × Direct feed fragmentations: Medium, Fine and Extra Fine size distributions
ii) 3 × GE feed fragmentation: Medium, Fine and Extra Fine
iii) Target Grind Size Output from the comminution circuit: 150–240 μm
iv) GE screen aperture size: 31.7–75 mm

1.2.1.2. Grade Engineering Differential Blasting. Grade Engineering Differential Blasting (GE DB) applies varying blasting intensities to exploit spatial grade heterogeneity (Carrasco et al., 2017). High levels of energy are applied to higher grade zones while low energy is applied to lower grade zones (Carrasco et al., 2016). This induces different particle size distributions based on the high and low grades which can be screened. These scenarios assessed the impact of blending differentially blasted ore that was upgraded as GE screen undersize product with the direct feed (three scenarios). There was a total of 167,582 blocks that were identified suitable for GE DB. The following are the simulated conditions in the study and hence the input variables for GE DB:

i) 3× Direct Feed fragmentations: Medium, Fine and Extra Fine size distributions
ii) GE feed fragmentation: Medium (low grade) mixed with Fine (high grade)
iii) Target Grind Size Output from the comminution circuit: 150–240 μm
iv) GE screen aperture size: 31.7–75 mm

1.2.2. Mine to Mill

Mine to Mill (M2M) is a process optimisation approach by reducing the ROM particle size distribution to increase process throughput throughout the system (Renner et al., 2006). Blast intensities can be controlled to yield varying plant feed size distributions. More detail on the M2M process can be found in Carrasco et al. (2019). These scenarios assessed varying plant feed size along with targeting different grind size distributions (12 scenarios). There was a total of 570,395 blocks that were identified suitable for the M2M blasting strategy (Amini et al., 2020). The following are the simulated conditions in the study and hence the input variables for M2M:

i) 3× Direct Feed fragmentations: Medium, Fine and Extra Fine size distributions
ii) Target Grind Size Output from the comminution circuit: 150–240 μm

1.3. Phenomenological models on the integrated extraction simulator

Each of the individual units in Fig. 1 is a phenomenological model which is calibrated to the plant data and constraints. Some of these models used are based on industry tested models found in JKTech, JKSimMet (Morrison and Richardson, 2002) and JKSimFloat (Harris et al., 2002). The specific phenomenological models used in the flowsheet are:

i) Gyratory Crusher: Andersen/Whiten (Andersen and Napier-Munn, 1990).
ii) Pebble Crusher: Andersen/Whiten (Andersen and Napier-Munn, 1990).
iii) SAG Mill: Variable Rates Model (Napier-Munn et al., 1996).
iv) CF SAG Screen: Single component efficiency curve (Napier-Munn et al., 1996).
v) Cone Crusher: Constant feed particle size distribution.
vi) Ball Mill: Perfect Mixing Model (Whiten, 1976).
vii) Hydrocyclone: Efficiency curve Nageswararao model (Nageswararao, 1978).

Amini et al. (2020) demonstrated the mass simulation capabilities of IES to simulate the 14,164,563 scenarios through AWS within a week (23.4 simulations/s). The phenomenological models do not have closed-form solutions and require iterative gradient-based solvers to obtain numerically approximated solutions which takes a long time.

1.4. Surrogate modelling in minerals processing

Surrogate modelling is a widely used approach in process engineering to approximate a chemical system or process for optimisation simulation (McBride and Sundmacher, 2019). On the contrary, surrogate models in minerals processing can only be found on simplifying CFD unit models in flotation (Rabhi et al., 2018) and thickening (Stephens et al., 2011; Stephens and Fawell, 2012). There is also a

pharmaceutical milling surrogate model paper by (Metta et al., 2020). Some of the commonly used surrogate models are polynomials, kriging, neural networks, radial basis functions, support and vector machines (Garud et al., 2017; Cisternas et al., 2020). There has been no existing literature in minerals processing which applies surrogate modelling to approximate a minerals processing comminution circuit or even multiple units in a process. This could be likely due to the large computational cost of mass simulating multiple phenomenological models to obtain a sizeable dataset suitable for surrogate modelling.

Although similar, there are more literature in minerals processing for response surface modelling (RSM) like in (Vieceli et al., 2016; Shahreza et al., 2015; Ebadnejad et al., 2013; Krajnik et al., 2005) among others. The aim of RSM is to define an empirical model which relates the input and output variables. In RSM, the underlying function between input and output variables are not known and must be obtained through experimentation. In contrast, surrogate modelling is where the underlying function between input and output variables is known and the focus is to simplify the computation.

1.5. Neural networks

Neural networks are an increasingly common form of machine learning used to approximate any continuous mapping function of inputs to output(s). The most basic form of neural networks is the multilayer perceptron (MLP) which is a dense feed-forward neural network (Almeida, 1997) shown in Fig. 2a.

The neural network consists of non-linear activation functions which provides its non-linear approximation capability (Hornik, 1989) as shown in Fig. 2b. Furthermore, neural networks can be represented as matrix operations which benefit from parallel computing. Neural networks are ultimately flexible empirical models which do not inherently capture underlying physical mechanisms. However, with proper care in sampling and experimental design for the training data, this study demonstrates that the physical behaviour of models can be approximated with good accuracy.

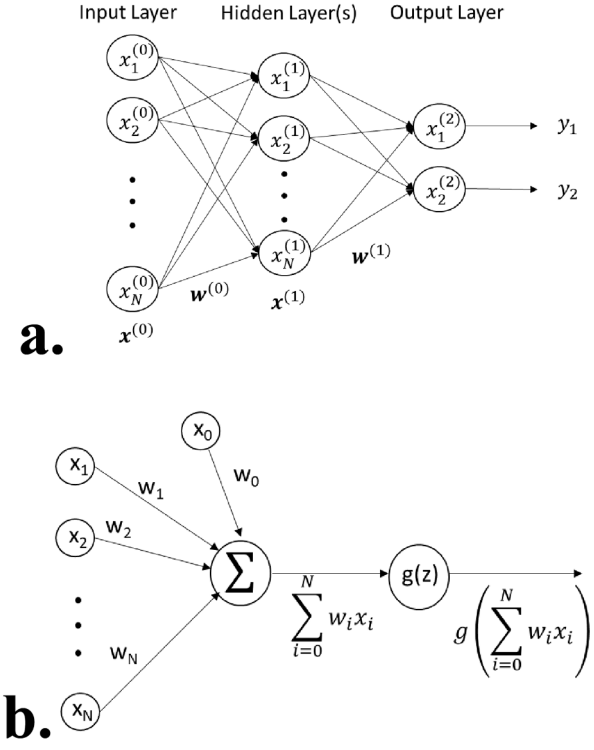

Fig. 2. a. Feed-forward neural network (multi-layer perceptron) with one hidden layer. Note that the number of neurons in each layer, N does not have to be equal. b. The dense neural network operation for an example node.

McCoy and Auret (2019) provides an extensive review of the existing attempts of using machine learning models in minerals processing. In their review, majority of neural networks developed in mineral processing are small in structure due to the lack of data. This is because most attempts were using neural networks to data obtained from laboratory-scale experiments. When the dataset is small, many benefits of deep learning are hindered as well as the increased likeliness for overfitting on a small dataset. This typically happens when the test set is a small fraction relative to the training set (for example, 10% test/90% train split). In this study, overfitting is minimized by allocating a test set which is 1000 times larger than the training set whilst still having enough training data for a deep neural network for representativeness.

2. Materials and methods

Fig. 3 summarises the entire process of developing the surrogate model used in this paper.

Given any block model population data, N the adequate number of samples, n required to generalise the ore hardness parameters needs to be determined. This is determined through trial and error by varying n (training set) to predict N-n (test set). Once the optimal n is determined, this is the required samples to predict the entire block model per operating condition. Since there were 174 operating scenarios to be investigated in this study, 174n samples were generated using phenomenological models. As the number of operating scenarios

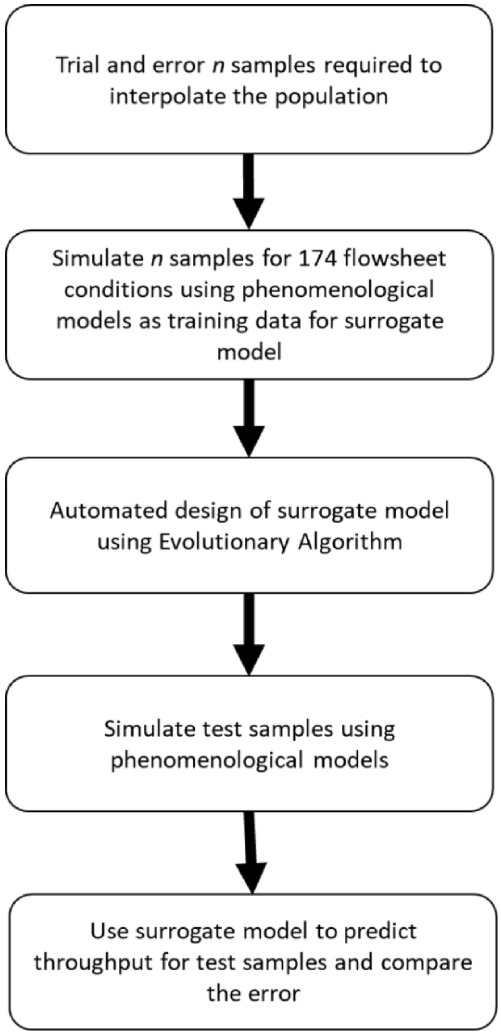

Fig. 3. Summary of the methodology used for this study.

change, the optimal surrogate model hyperparameters needs to be adjusted accordingly. To automate the optimisation of the hyperparameters, an evolutionary algorithm is used with measures to prevent overfitting. Then, randomly selected test data from the entire generalisation region (174 scenarios and any in between) is generated using phenomenological models and compared with the surrogate model predictions.

2.1. Hardware and software

All of the neural network training (including the evolutionary algorithm) and testing was performed on a local computer using Python 3.5.2 64-bit, Keras 2.2.4 (Chollet, 2015) and Tensorflow 1.13.1 Central Processing Unit (CPU) backend (Abadi et al., 2016). The computer had a 4-core Intel i7-4720HQ CPU chip and 16 GB of RAM. The test data generation was conducted on the IES cloud platform, using AWS (Amazon EC2 c5.large − 2 CPUs each) computing clusters. The number of clusters used depends on the scale of the simulation. The CPU backend (instead of graphical accelerated hardware) for Tensorflow was specifically chosen for the neural network prediction as it provides an equal comparison with the AWS CPU-based computations.

2.2. Training data sampling

The neural network must be able to interpolate in both the ore characteristics variable dimensions and the flowsheet condition variable dimensions. The following sub-sections detail the methodology used to sample training data in the ore characteristics and the flowsheet condition variable dimensions.

2.2.1. Ore characteristics variables sampling

There is a total of seven ore characteristics variables used for the neural network which are **density**, **SPI**, **BWI**, **CI**, **b**, **A**, and $\mathbf{t_a}$. These parameters were chosen because they are the most influential variables on the milling performance identified and embedded in the phenomenological grinding models in the flowsheet. Denoting the number of blocks in each operating strategy as $N_{\mathrm{M2M}} = 570{,}395$, $N_{\mathrm{GEND}} = 757{,}453$, $N_{GEDB} = 167{,}582$, and the population $N = N_{M2M} + N_{GEND} + N_{GEDB} = 1{,}495{,}430$. The number of samples of blocks taken for each operating strategy, n_{M2M}, n_{GEND}, n_{GEDB}, and $n = n_{M2M} + n_{GEND} + n_{GEDB}$ can be controlled to provide different degrees of generalisation accuracy. The computational time savings is directly proportional to N-n but the generalisation accuracy decreases as n decreases. This trade-off for n needs to be determined to yield a significant computational time savings without sacrificing too much generalisation accuracy.

To ensure that each ore characteristic "region" is represented in the training sample, stratified sampling is used. In this block model, it was determined that b was representative of ore characteristics for SAG milling and BWI was representative of the ore characteristics for ball milling. Therefore, the blocks were sorted by b and then by BWI. Then the sorted blocks were divided into strata/groups for sampling. This ensures that the sampled ore characteristics variables were representative of the population.

The number of samples taken in each stratum was determined through trial and error by sampling and training a neural network and evaluating the test error. Through trial and error, the final number of samples taken for each blasting strategy was $n_{M2M} = 552$, $n_{GEND} = 721$, and $n_{GEDB} = 337$ which is a sampling ratio of approximately 1 in 1000. Fig. 4 shows the sampled blocks, n in green among the population block model, N in blue. The samples are densely populated in the centre because that is reflective of the population block model which is densely populated in the centre as well.

The neural network was trained on simulation results from $n = 1610$ blocks and capable of predicting results from $N = 1{,}495{,}430$ blocks as well as intermediate values (between blue region) for a given flowsheet condition. Therefore, for the neural network to generalise to any given number of flowsheet conditions required, m, the number of training blocks required will be m*1610. Fig. 5 shows an illustration of this generalisation concept.

2.2.2. Flowsheet condition variables

The operation can switch between the three operating configurations: M2M, GE ND, GE DB. Therefore, this switch and other discrete variables must be represented numerically for the surrogate neural network to differentiate the conditions. To represent the circuit configuration and blasting methodology into numerical data inputs for the neural network, the following seven flowsheet condition variables were used:

i) $\mathbf{P_{80}}$, target grind size. This is a continuous variable between 150 and 240 μm.
ii) $\mathbf{D_{feed}}$, feed size to the SAG mill for M2M analysis. This is a continuous input with values of 1, 2, 3 for Extra Fine, Fine and Medium feed size respectively. This is done instead of one-hot encoding so that future interpolation between sizes is possible.
iii) $\mathbf{M_{GE}}$, GE feed throughput to the SAG mill feed. This is a continuous variable between 0 and 1200 tph. It is 0 for M2M scenarios and non-zero for GE ND and GE DB.
iv) $\mathbf{GE_{EF}}$, $\mathbf{GE_F}$, $\mathbf{GE_M}$, three levels of GE circuit product size. This is done using one-hot encoding for Extra Fine, Fine and Medium size because GE DB scenarios have a mix of both medium (low energy blast) and fine (high energy blast) which cannot be represented by one size variable. Therefore, for M2M scenarios it will be (0 0 0). For GE ND it can be (1 0 0), (0 1 0) or (0 0 1) for Extra Fine, Fine and Medium feed sizes, respectively. For GE DB, it is a

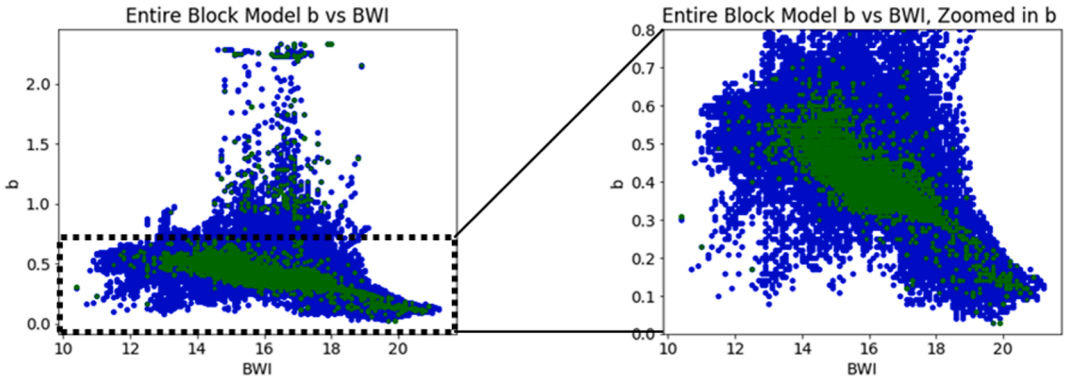

Fig. 4. Graph illustrating the region of the sampled blocks (green) compared to the non-sampled blocks (blue) for the entire block model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

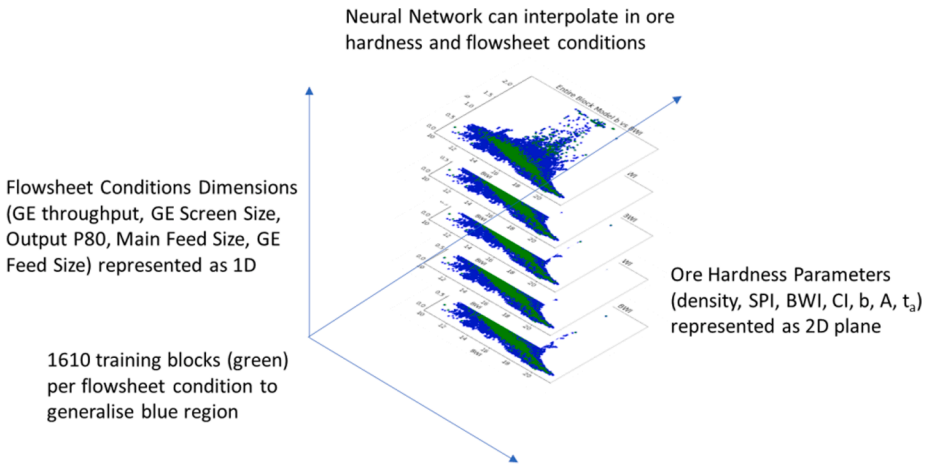

Fig. 5. Concept of using sampled 1610 training blocks (green) to generalise to the population (blue region) for each flowsheet condition. Note that the three-dimensional representation of the circuit and ore characteristics variables is an over-simplification to visualise the concept in 3D (there are 14 total input dimensions). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

percentage mix of Fine and Medium feed, (0 *j k*), where $j + k = 1$, e.g. (0 0.25 0.75).

v) **SS_{GE}**, GE screen size. This is a continuous variable representing the aperture size for the screen in the GE circuit. The value for this variable is from 31.75 to 75 mm for GE ND and GE DB scenarios and should be kept at zero for M2Ms.

The circuit variables that need to be interpolated are the main feed size, target output grind size, GE throughput, and GE screen size. For this study, there will be 174 flowsheet conditions, $m = 174$ which is detailed in Table 1. The results for these training scenarios were simulated on IES.

Table 1
Summary of training data used for the neural network.

Operation Strategy	Sampled Blocks per Flowsheet Condition	Flowsheet Conditions	Total Number of Blocks
Mine 2 Mill (M2M)	552	• 10 $\mathbf{P_{80}}$ (150, 160, 170, 180, 190, 200, 210, 220, 230, 240 μm) • 3 $\mathbf{D_{feed}}$ (Medium, Fines, Extra Fines) **Total 30 Scenarios**	16,560
Grade Engineering Natural Deportment (GE ND)	721	• 1 $\mathbf{P_{80}}$ (210 μm only) • 3 $\mathbf{D_{feed}}$ (Medium, Fines, Extra Fines) • 3 $\mathbf{SS_{GE}}$ (37.5, 57, 75 mm) • 4 $\mathbf{M_{GE}}$ (350, 650, 1000, 1200 tph) • 3 GE Sizes ($\mathbf{GE_{EF}} = 1$ or $\mathbf{G_F} = 1$ or $\mathbf{G_M} = 1$ for Medium, Fines, Extra Fines respectively) **Total 108 Scenarios**	77,868
Grade Engineering Differential Blasting (GE DB)	337	• 1 $\mathbf{P_{80}}$ (210 μm only) • 3 $\mathbf{D_{feed}}$ (Medium, Fines, Extra Fines) • 3 $\mathbf{SS_{GE}}$ (37.5, 57, 75 mm) • 4 $\mathbf{M_{GE}}$ (350, 650, 1000, 1200 tph) **Total 36 Scenarios**	12,132
Total 174 Scenarios, Total Training Data			106,560

2.3. Test data generation

The test data was divided into two datasets:

1. The first dataset was used to determine the appropriate number of samples required to interpolate the hardness parameters in the block model, i.e. *For the 174 Scenarios, can the surrogate neural network model predict the entire block model hardness variation by only training on a sub-sample of the block model hardness variables?*
2. The second dataset was used to test the interpolation capabilities of the neural network, i.e. *If the model can generalise the block model for the 174 Scenarios, can it interpolate for scenarios between the discrete input values of the 174 Scenarios?*

2.3.1. Ore characteristics generalisation test data

To test the predictive capability of the neural network, a mass simulation for the entire block model was performed for 24 flowsheet conditions. The total simulations performed was 14,164,563, summarised in Table 2 and documented in a previous publication by Amini et al. (2020). This mass simulation was computed on 3500 Amazon EC2 c5. large clusters (two virtual cores each) which cost AU$18,000 and took a week to complete through AWS. This is an average of AU$0.00127/simulation (or 787 simulations/AU$) and 23.4 simulations/s.

This dataset was used in the trial-and-error methodology to determine the appropriate *n* used to generalise for *N*. Variations of sampling ratios were used to determine appropriate trade-off between *n* and generalisation accuracy.

2.3.2. Flowsheet condition interpolation test data

To test the flowsheet condition interpolation capabilities, a smaller number of simulations were generated with phenomenological models. The following methodology was used to generate the 86,846 test data:

i) M2M Circuit Interpolation: 21,128 scenarios using randomly sampled blocks from the strata (non-training blocks) at varying target Grind Size ($\mathbf{P_{80}} = 150, 155, 160, 165 \ldots. 240$ μm) with $\mathbf{D_{feed}} = 1, 2, 3$ (Medium/Fines/Extra Fines SAG mill feed size). These test P80 sizes were chosen as they were the furthest interpolation points between the training data. Therefore, this test shows the highest possible error would be obtained and it is a heavy stress test to check the fidelity of the neural network.

ii) GE Circuit Interpolation: 65,718 scenarios using randomly sampled blocks from the strata (non-training blocks) with uniformly random

Table 2
Summary of the test data generated for Anglo American (after Amini et al., 2020).

Operation Strategy	Flowsheet Condition	Number of Blocks	Number of Scenarios = Total Flowsheet Condition Combinations * Number of Blocks
Mine 2 Mill (M2M)	• 4 $\mathbf{P_{80}}$ (150, 180, 210, 240 μm) • 3 $\mathbf{D_{feed}}$ (Medium, Fines, Extra Fines) • $\mathbf{SS_{GE}} = 0$ • $\mathbf{M_{GE}} = 0$ Total Flowsheet Condition Combinations = 4 × 3 = **12**	N_{M2M} = 570,395	$12{*}N_{M2M}$ = 6,844,740
Grade EngineeringNatural Deportment(GE ND)	• 1 $\mathbf{P_{80}}$ – 210 μm only • 1 $\mathbf{D_{feed}}$ – Medium only • 3 GE Sizes ($\mathbf{GE_{EF}} = 1$ or $\mathbf{G_F} = 1$ or $\mathbf{G_M} = 1$ for Medium, Fines, Extra Fines respectively) • 3 $\mathbf{SS_{GE}}$ (31.75, 57, 75 mm) • 1 $\mathbf{M_{GE}}$ – 1000 tph only Total Flowsheet Condition Combinations = 3 × 3 = **9**	N_{GEND} = 757,453	$9{*}N_{GEND}$ = 6,817,077
Grade Engineering Differential Blasting (GE DB)	• 1 $\mathbf{P_{80}}$ – 210 μm only • 1 $\mathbf{D_{feed}}$ – Medium only • 3 $\mathbf{SS_{GE}}$ (31.75, 57, 75 mm) • Variable GE Sizes from block model ($\mathbf{G_F} + \mathbf{G_M} = 1$, $\mathbf{GE_{EF}} = 0$) • 1 $\mathbf{M_{GE}}$ – 1000 tph only Total Flowsheet Condition Combinations = **3**	N_{GEDB} = 167,582	$3{*}N_{GEDB}$ = 502,746
Total Flowsheet Condition Combinations = 24, Total Simulations			**14,164,563**

continuous Screen Sizes (SS_{GE} between 37.5 and 75 mm), uniformly random continuous GE Throughput ($\mathbf{M_{GE}}$ between 0 and 1200 tph) with $\mathbf{D_{feed}} = 1, 2, 3$ (Medium/Fines/Extra Fines SAG mill feed size). For GE ND, the GE feed size was randomly chosen between Medium, Fine and Extra Fine ($\mathbf{GE_{EF}} = 1$ or $\mathbf{G_F} = 1$ or $\mathbf{G_M} = 1$). For GE DB, the GE Feed high-grade percentage blend was randomly generated ($\mathbf{G_F}$ between 0 and 1), and the low-grade was the remainder ($\mathbf{G_M} = 1\text{-}\mathbf{G_F}$), as they both must sum to 100%. The GE circuit variables were randomly generated within the range to heavily test whether the neural network can truly interpolate within all possible combinations of GE circuit variables.

2.4. Data pre-processing

To demonstrate the feasibility of this methodology, there is only one output which is the throughput of the comminution circuit. The neural network takes 14 input values (seven ore characteristics variables and seven flowsheet condition variables) to predict one output, the throughput of the comminution circuit. A neural network can have multiple outputs which can include other comminution circuit outputs like particle size distribution if necessary, in the future work. The input and output data were min-max normalised to be between 0 and 1 prior to training to ensure the loss function is well-behaved and allows for quicker convergence (Hastie et al., 2009).

For neural network training, it is common to allocate a small portion of the training data as a validation dataset (Hastie et al., 2009). The validation dataset acts as a "test" set during training to optimise hyperparameter and for model selection. But since the validation set is used to infer hyperparameters, it is not a true test set. Hence, there is a separate test set which is used after the final model is chosen to evaluate the true performance. The validation fraction of the training dataset used in this study is 10%.

2.5. Evolutionary algorithm for optimising neural network

The search space for the optimal neural network configuration is very large given that there are 14 inputs, 3 outputs and 106,560 training data. In traditional statistics, this would yield a degree of freedom of 106,560 – 3, i.e., a model with a maximum of 106,557 possible parameters. Evolutionary algorithms are great for searching optimal neural network solutions due to its global search capability without the use of gradient information. The evolutionary algorithm is an algorithm inspired by the principles of evolution and natural selection (Holland, 1992). The evolutionary algorithm methodology used in this paper was explained in detail by Real et al (2017). As the methodology used by Real et al. (2017) was for deep convolutional neural networks, some of the mutations will not be used as they are specific to convolutional layers and deep residual networks. In this study, the mutations available during reproduction are:

i) Alter learning rate (randomly alter learning rate)
ii) Insert dense layer (randomly insert a new dense hidden layer at any position between the input and output layers)
iii) Delete dense layer (randomly delete a hidden layer)
iv) Alter number of nodes in layer (randomly alter the number of nodes of a hidden layer)
v) Change activation function (randomly change the activation function in a hidden layer). The set of available functions are sigmoidal, ReLU, tanh, softplus, and linear.

These mutations were chosen based on their similarity to actions a human designer would take to improve the architecture.

For a reasonable computational time, each individual was only trained for 50 epochs to evaluate their fitness. This was done because improvements due to hyperparameter changes can be seen within the first 50 epochs. Early stopping (Prechelt, 2012) was used, and the stopping criteria was based on the validation loss plateauing. The training was done using the Adamax algorithm (Kingma and Ba, 2014) with a mean-squared loss error and a batch size of 200.

The population size was limited to 110 individuals. The initial population were randomly generated individuals with a single hidden layer with a random number of nodes that scaled with the dataset size, learning rate = 0.1, and a random choice of the available activation functions. The evolutionary algorithm terminated when the best individual in the population plateaued between generations. Once the best individual is determined through the evolutionary algorithm, the individual with the highest fitness was trained for 1000 epochs with early-stopping.

3. Results

This section shows the results for the evolutionary algorithm used for

automatically designing the surrogate neural network model and the final neural network model testing.

3.1. Evolutionary algorithm and training

The evolutionary algorithm took 10 h to run when using all four cores of the local computer. The results of the evolutionary algorithm are shown in Fig. 6.

The evolutionary algorithm terminated at generation 24 as the best individual fitness did not improve over 10 generations (although the average population continued to increase). The best individual found by the evolutionary algorithm was a 3-hidden layer neural network with the following configuration: 71 ReLU, 72 sigmoid, 39 sigmoid, a learning rate of 0.0111. This architecture had 9136 variables. The fitness of this neural network after 50 epochs was 8920, i.e. the average validation and training squared loss was 1.121×10^{-4}. These hyperparameters will be used as the final surrogate neural network model for full training.

The results of the full training for the surrogate neural network model is shown in Fig. 7. The training took one hour and 17 min. The algorithm terminated at 249 epochs as the validation loss plateaued.

3.2. Surrogate model testing

This section shows the results of the final surrogate model evaluated on the two generalisation test sets.

3.2.1. Ore characteristics generalisation test results

The neural network predicted all the 14,164,563 results from the phenomenological mass simulations in three minutes. Fig. 8 shows the parity chart of the neural network prediction against the phenomenological model results. As there are 14 million data points, it is difficult to show the density of the points, so the chart has been colour coded according to error thresholds. Fig. 9 shows the histogram of relative errors according to the three different blasting scenarios.

The parity chart shows that majority of the predictions (98.5%) of the neural network is within 2% error. If the average error of these 98.5% data is taken, it is 0.34%. Similarly, 99.8% of the data is within 4% error, yielding an average of 0.36% and all the data is within 10% error, yielding an overall average error of 0.37%. The histogram shows a more detailed breakdown of the errors within the 2% bins according to the operating strategies. Majority of the data (73%, 80% and 76% for M2M, GE ND, GE DB respectively) lies within the 0 and 0.5% error bin.

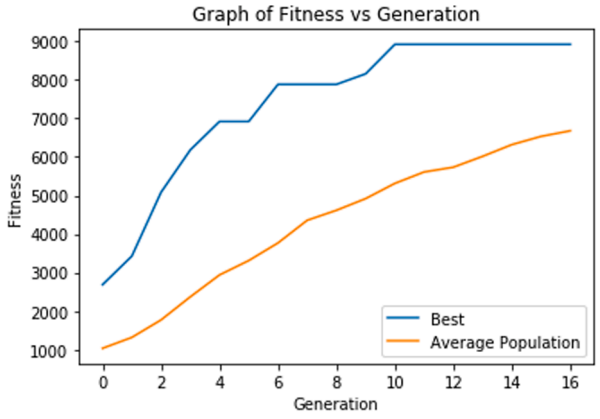

Fig. 6. The progress of the evolutionary algorithm through the generations showing the average fitness (orange) and the best individual fitness (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

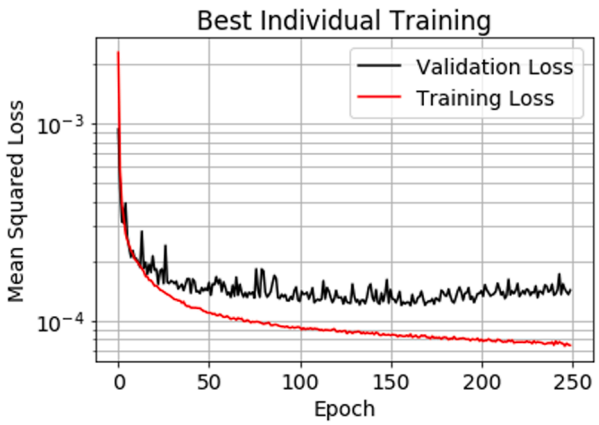

Fig. 7. Full training progress of the surrogate model (best neural network architecture found by the evolutionary algorithm) showing the training loss (red) and the validation loss (black). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

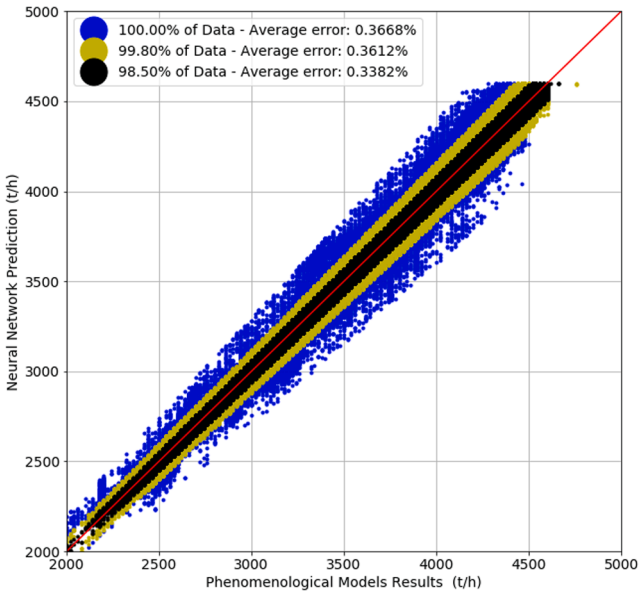

Fig. 8. Parity chart of the neural network prediction against the phenomenological model results for the ore characteristic test data according to different thresholds, black (errors $\leq$ 2%), yellow (errors $\leq$ 4%), blue (entire dataset). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2.2. Flowsheet condition interpolation testing results

The neural network predicted the 86,846 test data within two seconds on the local computer. Fig. 10 shows the parity charts of the neural network prediction and phenomenological model results for the two test scenarios.

4. Discussion

The test results show that sampling a small fraction (approximately 1/1000) of the block model is possible to interpolate the entire block model at a cost of slightly reduced accuracy. The solver-based solutions of throughput has a convergence threshold of 20 t/h which is around 0.5% for a throughput of 4000 t/h. Therefore, surrogate model solutions with errors below 0.5% is indistinguishable from the phenomenological model solutions due to round up errors. Given that the average error of the entire surrogate model results is 0.3668%, this is well within the acceptable limits for simulation purposes. This is especially true for

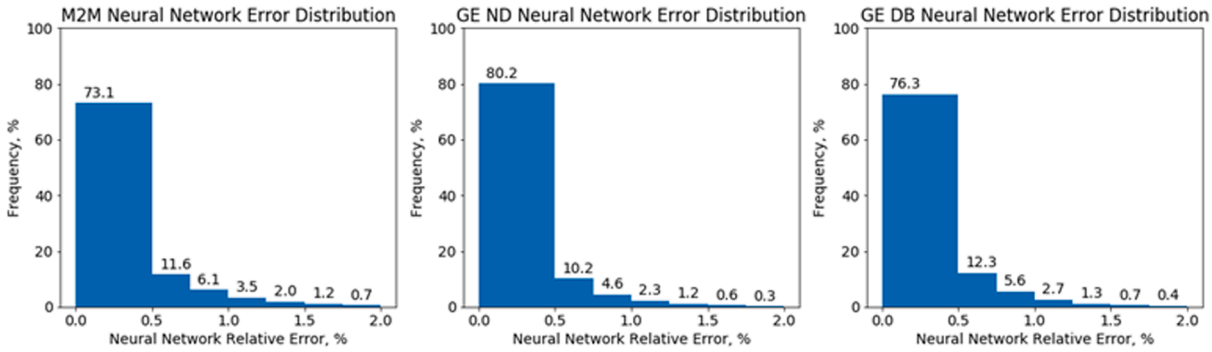

Fig. 9. Histogram of error distribution according to the operating strategies.

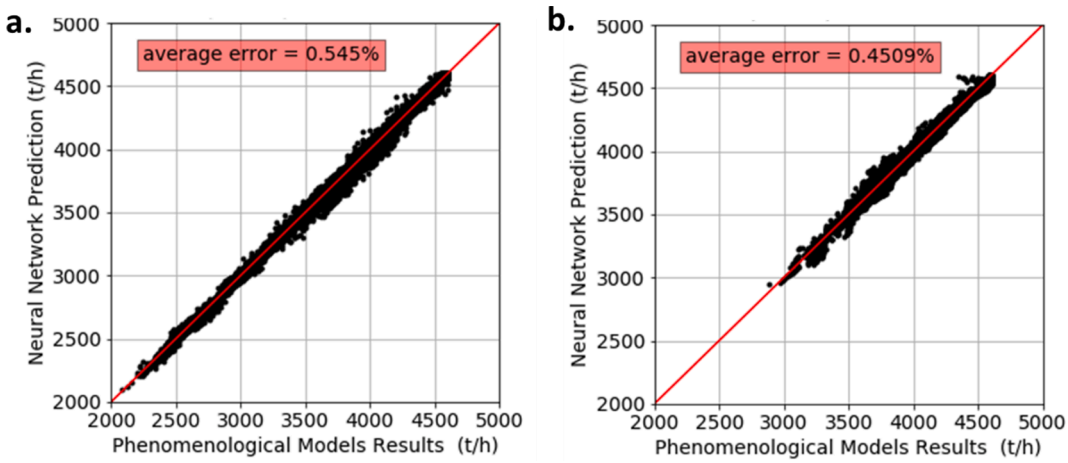

Fig. 10. Parity chart for the neural network predictions against phenomenological model results for interpolating the flowsheet condition variable test results, a. M2M product size interpolation, b. GE circuit variables interpolation.

minerals processing modelling and simulations where the accuracy of input data/sensors are greater but only the mathematical error between the original simulations and the surrogate model results are reported here which is expected to be less than round up error.

Significant care was taken in the methodology to minimise overfitting of the final surrogate model. The main reason overfitting occurs in neural network development is the training dataset is usually much larger than the test dataset. In this study the contrast is true, and the computational speed-up is more apparent as the training dataset becomes smaller. As a result, test set 1 is 1000 times larger than the training dataset which makes overfitting impossible. On top of this, the input variables for test set 2 have been randomized within the input bounds which is completely different from the training dataset. Furthermore, a validation set with early stopping was used in full training and the evolutionary algorithm which is standard practice in neural network literature to minimise overfitting (Hastie et al., 2009). The final neural network architecture chosen also only has 9136 parameters which is much less than the 106,557 degrees of freedom in the proposed problem. Part of the reason why the neural network generalised well can also be attributed to the stratified sampling methodology used in this paper. By carefully sampling training data from each of the ore hardness regions (Fig. 4) with a full factorial design, the neural network is exposed to all the input conditions.

The benefit of using a neural network surrogate model becomes very clear as it greatly reduces computational cost and time required at negligible loss in accuracy. Table 3 is a summary of the computational time and cost required for the phenomenological models compared to the surrogate model. The phenomenological model result is based on the results from the 14,164,563 simulations evaluated on AWS cloud servers.

Table 3
Summary of time required and cost for the phenomenological model results and the neural network methodology proposed in this study.

Methodology	Cost (A$)	Time	Simulations/ second	Simulation Capabilities
Phenomenological models	18,000	7 days (604,800 s) predict 14,164,563 simulations	23.4	24 Scenarios with no interpolation
Surrogate Neural Network model	Negligible	3 min (180 s) predict 14,164,563 simulations 10.5 h (37,800 s) train	78,700	Interpolation within 174 Scenarios

Therefore, simulations using the surrogate neural network model are 3363 times quicker than simulating phenomenological models. The block model attributes or process applications get updated frequently in operations, and mass simulation of the entire block model using surrogate models is a quick and feasible alternative to re-run the phenomenological model flowsheets. This will also allow the exploration of mine conditions at a larger scale which was constraint to time and budget before. However, phenomenological models are still required to generate the training data for the neural network albeit at ≈ 1000 times less.

5. Conclusion

This study investigated the possibility of using a neural network as a surrogate model to approximate phenomenological models to improve the number of simulations possible per second. The block model of a large copper porphyry mine was analysed to determine the number of representative blocks required to interpolate the entire block model. Through heuristics and expert analysis, it was determined that the blocks can be sampled according to groups of b and BWI to account for the SAG mill and ball mills. There was a total of $N_{M2M} = 570,395$, $N_{GEND} = 757,453$, and $N_{GEDB} = 167,582$ blocks for the Mine to Mill (M2M), Grade Engineering Natural Deportment (GE ND) and Grade Engineering Differential Blasting (GE DB) scenarios. Through stratified sampling it was found that only $n_{M2M} = 552$, $n_{GEND} = 721$, and $n_{GEDB} = 337$ blocks were required from the M2M, GE ND and GE DB scenarios to interpolate the rest of the blocks, which is a sampling ratio of approximately 1:1000. This meant only a small fraction of blocks was required to generalise the entire block model region for each given flowsheet condition.

A total of 174 scenarios was used, yielding a training dataset of 106,560 scenarios. The optimum neural network hyperparameters were obtained using an evolutionary algorithm. The evolutionary algorithm took 10 h to search and found the best architecture was a 3-hidden layer (71 ReLU, 72 sigmoid, 39 sigmoid) with a learning rate of 0.0111. This neural network was then fully trained on the dataset which took 30 min and tested on three test sets. The first test set of 14,164,563 scenarios was the full block model simulated on 24 scenarios using phenomenological models. This took a week to generate on IES with an average speed of 23.4 scenarios/second. In comparison, the surrogate model predicted all 14,164,563 scenarios in 180 s (78,700 simulations/second) on a local computer with an average error of 0.37%. The other two test sets were designed around interpolating the M2M (21,128 scenarios) and GE (65,718 scenarios) flowsheet conditions. The surrogate model predicted these with an average error of 0.55% and 0.45% respectively.

The entire methodology has been designed to be automated based on the block model input except for the block model sampling procedure. This would be part of future research to determine the best stratified sampling variable based on the block model data. Besides that, the methodology can be repeated automatically for any changes in the block model or simulation of other mine operation value chains. The neural network can approximate phenomenological models at a small cost of accuracy but greatly reduces the time required for simulations.

CRediT authorship contribution statement

Edwin J.Y. Koh: Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Visualization. **Eiman Amini:** Conceptualization, Methodology, Validation, Writing - review & editing, Visualization, Project administration. **Geoffrey J. McLachlan:** Writing - review & editing, Supervision. **Nick Beaton:** Conceptualization, Resources, Supervision, Funding acquisition.

Declaration of Competing Interest

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

Acknowledgement

This research would not have been possible without the financial support from CRC ORE. CRC ORE is part of the Australian Government's CRC Program, which is made possible through the investment and ongoing support of the Australian Government. The CRC Program supports industry-led collaborations between industry, researchers, and the community.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al., 2016. Tensorflow: A System for Large-Scale Machine Learning. In: Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, Savannah, USA.

Almeida, L.B., 1997. Multilayer perceptrons. In: Fiesler, E., Beale, R. (Eds.), Handbook of Neural Computation. IOP Publishing Ltd, New York.

Amini, E., Becerra, M., Bachmann, T., Beaton, N., Shapland, G., 2020. Development and Fine-tuning of a Mine Operation Value Chain Flowsheet in IES to Enable Grade Engineering and Process Mass Simulations for Scale-Up and Strategic Planning Analysis. Society for Mining, Metallurgy and Exploration Annual Conference.

Andersen, J.S., Napier-Munn, T.J., 1990. The Influence of liner condition on cone crusher performance. Minerals Eng. 3 (1–2), 105–116.

Ballantyne, G.R., Powell, M.S., Tiang, M., 2012. Proportion of Energy Attributable to Comminution. In: Proceedings of the 11th Australasian Institute of Mining and Metallurgy Mill Operator's Conference, Hobart, Tasmania, pp. 25-30.

Bartlett, J., Holtzapple, A., Rempel, C., 2014. A Brief Overview of the Process Modelling/Simulation and Design Capabilities of METSIM. Canadian Institute of Mining, Metallurgy and Petroleum: COM 2014 - Conference of Metallurgists Proceedings.

Carrasco, C., 2013. Development of geometallurgical tests to identify, rank and predict preferential coarse size by size Au deportment to support feed preconcentration at Telfer Au-Cu mine. Brisbane Australia: Newcrest Western Australia MPhil Thesis. University of Queensland (JKMRC).

Carrasco, C., Gahona, M., La Rosa, D., Shapland, G., Beaton, N., Amini, E., 2019. Augmenting Traditional Mine to Mill with Cloud-Based Simulation and Reconciliation Capabilities. Vancouver, Canada: SAG Conference Vancouver.

Carrasco, C., Keeney, L., Walters, S.G., 2014. Development of geometallurgical laboratory tests to characterise metal preconcentration by size. In: Proceedings XXVII International Mineral Processing Congress, Santiago, Chile, pp. 1-24.

Carrasco, C., Keeney, L., Walters, S.G., 2016a. Development of a novel methodology to characterise preferential grade by size deportment and its operational significance. Minerals Eng. 91, 100–107.

Carrasco, C., Keeney, L., Napier-Munn, T.J., Bode, P., 2017. Unlocking additional value by optimising comminution strategies to process Grade Engineering streams. Minerals Eng. 103–104, 2–10.

Carrasco, C., Keeney, L., Scott, M., Napier-Munn, T.J., 2016b. Value driven methodology to assess risk and operating robustness for grade engineering strategies by means of stochastic optimisation. Minerals Eng. 99, 76–88.

Chollet, F., 2015. Keras. Retrieved from GitHub repository: https://github.com/fchollet/keras.

Cisternas, L.A., Lucay, F.A., Botero, Y.L., 2020. Trends in modelling, design, and optimization of multiphase systems in minerals processing. Minerals 1 (22), 10.

Dobby, G., Kosick, G., Amelunxen, R., 2002. A focus on variability within the orebody for improved design of flotation plants. In: 34th Annual Canadian Mineral Processing Conference, Ottawa, Ontario.

Durance, M.-V., Guillaneau, J.-C., Villeneuve, J., Brochot, S., Fourniguet, G., 1994. USIM PAC 2 for Windows: Advanced simulation of mineral processes. In: Proceedings of the 5th International Mineral Processing Symposium, Cappadocia, Turkey.

Ebadnejad, A., Karimi, G.R., Dehghani, H., 2013. Application of response surface methodology for modelling of ball mills in copper sulphide ore grinding. Powder Technol. 245, 292–296.

Ford, M.A., King, R.P., 1984. The simulation of ore-dressing plants. Int. J. Mineral Process. 12, 285–304.

Garud, S.S., Karimi, I.A., Kraft, M., 2017. Design of computer experiments: a review. Computers Chem. Eng. 106, 71–95.

Harris, M.C., Runge, K.C., Whiten, W.J., Morrison, R.D., 2002. JKSimFloat as a practical tool for flotation process design and optimisation. Mineral Processing Plant Design Practice and Control Conference. Society for Mining Metallurgy & Exploration, New York, USA.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second ed. Springer, New York.

Holland, J.H., 1992. Genetic Algorithms. Scientific American 267 (1), 66–73.

Hornik, K., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359–366.

Kingma, D.P., Ba, J., 2014. Adam: A method for Stochastic Optimization. arXiv: 1412.6980.

Kosick, G., Dobby, G., Bennett, C., 2001. CEET (Comminution Economic Evaluation Tool) (For Comminution Circuit Design And Production Planning). In: Proceedings of 2001 SME Annual Meeting, Denver, Colorado, USA.

Krajnik, P., Kopac, J., Sluga, A., 2005. Design of grinding factors based on response surface methodology. J. Mater. Processing Technol. 162, 629–636.

McBride, K., Sundmacher, K., 2019. Overview of surrogate modelling in chemical process engineering. Chemie Ingenieur Technik 91 (3), 228–239.

McCoy, J.T., Auret, L., 2019. Machine learning applications in minerals processing: a review. Minerals Engineering 132, 95–109.

Metta, N., Ramachandran, R., lerapetritou, M. A., 2020. A Computationally Efficient Surrogate-Based Reduction of a Multiscale Comill Process Model. J. Pharmaceutical Innovation, 15, 424-444.

Morrison, R.D., Richardson, J.M., 2002. JKSimMet: A simulator for analysis, optimisation and design of comminution circuits. Mineral Processing Plant Design Practice and Control. SME: Society for Mining, Metallurgy and Exploration, New York, USA, pp. 442–460.

Nageswararao, K., 1978. Further developments in the modelling and scale-up of industrial hydrocyclones. JKMRC, University of Queensland, Brisbane, Australia.
Napier-Munn, T., Morell, S., Morrison, R., Kojovic, T., 1996. Mineral Comminution Circuits: Their Operation and Optimisation. Julius Kruttschnitt Mineral Research Centre University of Queensland, Brisbane.
Prechelt, L., 2012. Early Stopping — But When? In: Montavon, G., Orr, G.B., Müller, K.R. (Eds.), Neural Networks: Tricks of the Trade. Springer, Berlin, pp. 53–67.
Rabhi, A., Chkifa, A., Benjelloun, S., Latifi, A., 2018. Surrogate-based modelling in flotation processes. Computer Aided Chem. Eng. 43, 229–234.
Razavimaneseh, A., Rumball, J., Tadel, M., & Pareek, V., 2006. Steady-state simulation of hybrid nickel leaching circuit using SysCAD. Auckland, New Zealand: Chemeca 2006: Knowledge and Innovation.
Renner, D., La Rosa, D., DeKlerk, W., Valery, W., Sampson, P., Noi, S., Jankovic, A., 2006. Anglogold Ashanti Iduapriem mining and milling process integration and optimisation. In: International Conference on Autogenous and Semi-autogenous Grinding Technology, Vancouver, pp. 249-264.
Shahreza, S., Shafaei, S., Noaparast, M., Sarvi, M., 2015. Optimization of Galena Flotation Process of Irankouh Complex Ore Using A Statistical Design of Experiments. Curr. World Environ. 10, 626–636.
Stephens, D.W., Gorissen, D., Crombecq, K., Dhaene, T., 2011. Surrogate based sensitivity analysis of process equipment. Appl. Math. Model. 35, 1676–1687.
Stephens, D., Fawell, P., 2012. Optimization of Process Equipment Using Global Surrogate Models. In: Proceedings of the Conference on CFD in the Minerals and Process Industries CSIRO, Melbourne, Australia.
Vieceli, N., Durão, F.O., Guimarães, C., Nogueira, C.A., Pereira, M.F., Margarido, F., 2016. Grade-recovery modelling and optimization of the froth flotation process of a lepidolite ore. Int. J. Minerals Processing, 157, 184-194.
Whiten, W.J., 1976. Ball mill simulation using small calculators. In: Proceedings of the AusIMM 258, pp. 47-53.