Optimization of Inputs for the Application of ANN to Rail Track Granular Materials

Buddhima Indraratna

Distinguished Professor of Civil Engineering and Director of Transport Research Centre, University of Technology Sydney, Ultimo, Australia

Email: buddhima.indraratna@uts.edu.au

Haydn Hunt

PhD Student, School of Civil and Environmental Engineering, University of Technology Sydney, Ultimo, Australia

Email: haydn.r.hunt@student.uts.edu.au

Rakesh Sai Malisetty

Postdoctoral Research Associate, School of Civil and Environmental Engineering, University of Technology Sydney, Ultimo, Australia

Email: rakeshsai.malisetty@uts.edu.au

Srinivas Alagesan

PhD Student, School of Civil and Environmental Engineering, University of Technology Sydney, Ultimo, Australia

Email: srinivas.alagesan@student.uts.edu.au

Yujie Qi

Senior Lecturer, Program Co-leader, Transport Research Centre (TRC), University of Technology Sydney (UTS), 15 Broadway, Ultimo NSW 2007, Australia

E-mail: yujie.qi@uts.edu.au

Cholachat Rujikiatkamjorn

Professor, Transport Research Centre, School of Civil and Environmental Engineering, University of Technology Sydney, Ultimo, NSW 2007, Australia

Email: cholachat.rujikiatkamjorn@uts.edu.au

Corresponding Author:
Buddhima Indraratna, Distinguished Professor, School of Civil and Environmental Engineering
Faculty of Engineering and Information Technology, University of Technology Sydney, Australia.
Ph: +61 (0) 400 213 046, Email: Buddhima.indraratna@uts.edu.au

Manuscript details:
Number of Words: 11950
Number of Figures: 13
Number of Tables: 11

Abstract

Machine Learning (ML) models such as Artificial Neural Networks (ANN) have gained increasing popularity in geotechnical engineering applications as an alternative to conventional empirical and computational models. At present, very few ML models exist for predicting the mechanical responses of track granular materials such as ballast and subballast which may even comprise of composite mixtures of blended granular materials. Moreover, the performance of any ML model depends not only on the quality and quantity of available data but also on the selection process for input parameters, which often lacks adequate justification in the past literature. In this context, the current study introduces ANN models for track granular materials based on published laboratory data with special emphasis on the selection of an optimal set of input parameters. Two applications of ANN are considered to: (i) predict the peak friction angle (ϕ'_{peak}) of a variety of granular mixtures under static loading and (ii) predict ballast breakage under cyclic loading. The selection process involves prudent analysis of key influential parameters in a geotechnical perspective, while also ensuring that they are conveniently measurable. Performance evaluation of these models with various input combinations is carried out, while proposing optimal input parameters for both applications.

Keywords: Artificial Neural Networks, granular materials, friction angle, ballast breakage

1. Introduction

Granular layers such as ballast and subballast (capping) play an important role in providing the necessary bearing capacity to conventional ballasted railway tracks. Under repeated passes of freight trains operating in heavy-haul corridors, these materials are prone to degradation and deformation thus decreasing the long-term load-bearing capacity of the track substructure. A large portion of the track operational costs are associated with granular layer maintenance such as ballast replenishment and realignment of track geometry. For instance, more than AU$2.1

58 billion was spent on track maintenance by Australian railway asset owners during the period
59 2021-22 (BIS Oxford Economics, 2022). Accurate estimation of physical and mechanical
60 properties such as shear strength, particle breakage of ballast and their variation with time is
61 necessary for achieving a trade-off between the cost of maintenance and engineering safety of
62 tracks. Moreover, with the recent increase in demand for circular economy measures, the use
63 of marginal materials such as granular wastes from mining industry and recycled rubber
64 granules/crumbs derived from scrap tyres has gained much attention. Laboratory investigations
65 have shown that mechanical characteristics such as compaction and deformation response
66 under cyclic loading of these marginal materials can significantly differ from conventional or
67 natural geologic materials (Arulrajah et al., 2014, Indraratna et al., 2018, Indraratna et al.,
68 2019). In this regard, it is important to develop predictive models embedding the complex
69 behavior of such granular composites.

70 Empirical models are widely used to predict the shear strength of granular mixtures
71 (Rujikiatkamjorn et al., 2013, Abozraig et al., 2022) and ballast breakage (Lackenby et al.,
72 2007, Hussaini and Sweta, 2020) encompassing data trends from numerous laboratory testing
73 projects. The alternative practice has been to adopt mathematical models based on constitutive
74 stress-strain responses with model parameters calibrated with experimental data (Einav, 2007,
75 Malisetty et al., 2020). However, these two types of approaches have limitations as most
76 models have practical challenges when determining the required model parameters. Machine
77 Learning models via Artificial Neural Networks (ANN) have evolved swiftly in recent times
78 as a viable alternative to ultra-end human intelligence through a network of interconnected
79 nodes and layers representing complex underlying relationships within the training datasets.
80 In this respect, ANN models are often found to be superior to traditional statistical regression
81 models in capturing the non-linear behavior of geomaterials that exhibit high variability in their

engineering characteristics, apart from the fact that they do not require prior assumptions (Phoon and Zhang, 2023).

In recent years, there have been numerous examples of ANN models developed for different applications in geotechnical engineering. These include bearing capacity and settlement prediction models for pile foundations (Chan et al., 1995, Samui, 2008), shallow foundations and compaction (Shahin et al., 2002, Gomes Correia et al., 2013), models predicting the constitutive behavior of geomaterials (Penumadu and Zhao, 1999, Shahin and Indraratna, 2006), amongst other applications. Of these studies, two studies have focused on ballast which include the ANN model developed by Shahin and Indraratna (2006) to predict the stress-strain response under monotonic loading, while Indraratna et al. (2023a) predicted resilient modulus under cyclic loading conditions using ANN and ANFIS methods. However, it is to be noted that most of these models are developed for a specific type of soil, so there is a desirable need to extend the ANN applications to composite materials such as steel furnace slag (SFS), coal wash (CW) and rubber crumbs (RC), and ballast-rubber mixtures. Moreover, particle breakage evaluated using the Ballast Breakage Index, BBI (Indraratna et al., 2005) under cyclic loading needs quantification through ANN to effectively capture the relationship between particle breakage and the associated influential factors.

In the development of ML models, quality of data and selection of optimum input parameters/datasets is crucial for a superior performance. Phoon and Zhang (2023) highlighted that any data model developed should be practical to offer valuable insights to assist critical real-world decision-making. Moreover, the selection of input parameters should encompass the geotechnical context of the material and be based on an evaluation process that analyses the input-output relationships in line with the fundamental concepts. Furthermore, the number of input parameters should also be optimized to avoid the computational complexities and

106 inaccuracies attributed to redundant input parameters. Following these principles, ANN models
107 are developed in this study for: (i) predicting the peak friction angle of granular composites
108 including waste materials such as SFS, CW, and recycled rubber crumbs and granules, as well
109 as rubber mixed with sands and gravels, (ii) prediction of BBI with number of loading cycles.
110 A careful selection of input parameters is conducted such that they can be easily synthesized
111 from the field or laboratory and that they can be readily interpreted on a physical basis. By
112 incorporating specific material- and loading-based parameters, this study focuses on both the
113 accuracy and generalization of these developed models, rather than improving accuracy alone
114 via increased modelling complexities.

115 ## 2. Background of ANN

116 The structure of ANN consists of several nodes that are arranged into different interconnected
117 layers: input, hidden (intermediate) and output layers (Fausett 1994). At each node in the
118 hidden layer, the inputs from the previous layer are multiplied with weighting coefficients,
119 summed (with an additional bias), and then propagated to the next layer using a transfer
120 function. In this study, a sigmoid activation function is used as the transfer function between
121 the input and hidden layer, and a hyperbolic tangent function between hidden and output layer.
122 The training subset is used to calibrate the weights and biases between the layers where the
123 values are randomly initialized and iterated to reduce the mean squared error (MSE). For the
124 present study, Levenberg-Marquardt (LM) and Bayesian Regularisation (BR) algorithms are
125 used to train the ANN network and minimise MSE. LM is generally the fastest to train and
126 efficient in error minimisation, whereas BR, which is a modification of LM, is often better
127 adept at improving generalisation where smaller data sizes are used (Foresee and Hagan, 1997).
128 Afterwards, the testing subset is used to independently evaluate the model's performance on
129 data that was not used for training.

2.1. Model Evaluation Metrics

Five evaluation metrics are employed to assess the performance of the ANN models. These are the mean absolute relative error (MARE), root mean squared error (RMSE), coefficient of determination (R^2), variance account for (VAF), and A-10 indices, calculated by:

$$\text{MARE}=\frac{1}{n}\sum_{i=1}^{n}\frac{|O_p-O_t|}{O_m}\text{ x }100\% \quad (1)$$

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(O_p-O_t\right)^2} \quad (2)$$

$$R^2=1-\frac{\sum_{i=1}^{n}\left(O_p-O_t\right)^2}{\sum_{i=1}^{n}(\overline{O_t}-O_t)^2} \quad (3)$$

$$VAF=\left(1-\frac{var(O_p-O_t)}{var(O_t)}\right)\times 100\% \quad (4)$$

$$A\text{-}10=\frac{m10}{n} \quad (5)$$

where O_p and O_t are the predicted and target output values, respectively; $\overline{O_t}$ is the average of the target values; n is the total number of data samples; var refers to variance; $m10$ are the number of samples where O_p is within $\pm$10% of O_t.

3. Modelling of Friction Angle of Granular Materials (Static Loading)

Shear strength of track granular materials is usually represented using the peak friction angle (ϕ'_{peak}) which can be obtained through laboratory testing, although this process is often time-consuming. In the first application of this study, ANN is applied to develop a data model that is capable of accurately predicting ϕ'_{peak} for a variety of track granular materials. Different material types are investigated, as opposed to most of the literature where the focus is typically on one or two material types. Here, both non-waste (ballast [latite basalt] and sand) and waste (steel furnace slag [SFS], coal wash [CW], and granulated waste rubber) materials and mixtures

are investigated. Since large rubber chips and shreds have been shown to produce a reinforcing effect and induce completely different deformations and shearing behavior (Edil and Bosscher, 1994, Foose et al., 1996, Mashiri et al., 2015, Qi et al., 2018), the rubber used in this study is confined to granular rubber with comparable sizes to the rigid host material. A total of 154 consolidated drained (CD) static triaxial tests using these materials were synthesized from the literature, where approximately 85% of these data samples can be considered as waste mixtures (i.e., rigid waste material and/or rubber inclusion).

3.1. Input Parameter Selection

For a soil mechanical model to be comprehensive and practical, the input variables should represent the three key categories of (i) soil stress state, (ii) soil physical state, and (iii) soil type. This facet is where the application of geotechnical knowledge into a machine learning model is critical to avoid the model 'learning' relationships which fundamentally do not exist. Table 1 summarizes the nine input variables initially investigated to predict the peak friction angle of the materials considered in this study. Based on the Mohr-Coulomb model, the shear strength of a granular material is a function of the normal stress at failure, which in turn is a function of the principal stresses. For a linear shear failure envelope, ϕ'_{peak} is independent of the confining stress, as it is a constant and intrinsic material property. However, it is well-established that this envelope diverges from the linear model when granular materials, particularly with rubber inclusion, are considered (Tawk and Indraratna, 2021, Hunt et al., 2023, Indraratna et al., 1998, Sheikh et al., 2013). At lower confining stresses, this nonlinearity is more pronounced with a higher ϕ'_{peak} value. Moreover, as the database is synthesised from CD triaxial tests, the principal stresses σ'_2 and σ'_3 are equal. Therefore, the effective confining pressure, σ'_3, was selected as a key input parameter to represent the stress state of the granular materials.

Dry unit weight, γ_d, was also selected since the density of a granular matrix is directly correlated to its peak shear strength. Although, this relationship is not as straightforward when materials of differing specific gravities and degrees of compaction are considered. To account for weight discrepancy, Indraratna et al. (2019) adopted the initial void ratio, e_0, to better represent the state of compaction of lighter CW-rubber mixtures. However, prior ML studies are often constrained to a singular material or mixture type and tend to exclude at least one of e_0 and γ_d as little variance exists in their respective databases (e.g., Penumadu and Zhao, 1999, Habibagahi and Bamdad, 2003, Edincliler et al., 2012, Zhou et al., 2019). Therefore, as this study is concerned with a variety of material types (i.e., ballast, sand, SFS, CW, rubber), both e_0 and γ_d were selected to represent the physical, or compactive, state. Six additional variables were considered to further account for the variation in these materials and represent the overall soil type. Gravimetric rubber content, $RC\%$, was included as it has been shown to reduce a mixture's shear strength when blended at comparable sizes (Hunt et al., 2023). The coefficients of uniformity, C_u, and curvature, C_c, were also selected to incorporate particle size distribution (PSD) and gradation characteristic information into the model. Lastly, the mixture's effective particle size, D_{10}, its median particle size, D_{50}, and the maximum particle size of the rigid (i.e., non-rubber) constituent(s), $D_{max,rigid}$, were considered to include additional particle size information. For the latter, the rigid constituent was isolated since the majority of shear resistance is attributed to the stronger rigid particles. C_u, C_c, and D_{50} were used by Penumadu and Zhao (1999) and Shahin and Indraratna (2006) to predict the loading response of sand and gravel, and ballast, respectively, while Zhou et al. (2019) included D_{10} to predict the shear strength of rockfill material. The individual input parameters considered in this current study are summarised and categorised in Table 1, while the initial modelling phase can be expressed as the function $\phi'_{peak} = f(RC\%, \sigma'_3, e_0, \gamma_d, C_u, C_c, D_{10}, D_{50}, D_{max,rigid})$.

3.2. Database Statistics

After the removal of outliers, a total of 154 CD triaxial test data samples were incorporated into the modelling process, as summarized in Table 2. This test type was used since the drainage conditions replicate those experienced in the field for granular track materials. Additionally, 58% of the database consists of materials with rubber inclusion to capture the effect of soil-rubber mixtures with varying rubber contents.

Statistical information for the minimum, maximum, mean, and standard deviation of each of the nine input variables and single output variable is documented in Table 3, with the frequency histograms illustrated in Figure 1. A wide range of values for each variable is utilized in this study because of incorporating various material types and loading (confining pressure) conditions. This introduced variance assists the modelling process since a lack of variation can hinder a model's ability to learn the underlying relationships between predictor and output variables. This is a key reason why a large amount of geotechnical-based ML studies in the past literature had omitted important physical and material properties, since they are often focused on a singular material type. For example, Indraratna et al. (2023a) did not consider density, gradation, or other physical properties to predict the resilient modulus of ballast as there is little variation since they are effectively kept constant for track design based on Australian Standards (2015).

3.3. Modelling Process

To achieve a consistent division of data between training and validation datasets, a semi-random trial-and-error approach was implemented in this study, where the overall database was randomly partitioned 20 times and the most consistent in terms of mean, standard deviation, minimum, and maximum, across all ten variables was selected. The statistical properties for

221 the training and testing data subsets across each variable are illustrated in Figure 2. Here, the
222 individual datasets are partitioned with a satisfactory degree of statistical consistency,
223 particularly for $RC\%$, σ_3', γ_d, C_u, C_c, and, most importantly, ϕ_{peak}'. Although, some deviations
224 between the means for e_0 and the variables related to particle size are observable, which can
225 be attributed to the different compaction methods and targets, materials, and the array of
226 particle size distributions used, increasing the difficulty in maintaining consistency across all
227 nine input variables. However, these differences are adjudged to be suitable for modelling and,
228 moreover, an effort was made so that each material type was incorporated into both the training
229 and testing datasets. Additionally, since the ANN method struggles to extrapolate beyond the
230 limits of the data that the model is trained on (Flood and Kartam, 1994), it was also ensured
231 that the minimum and maximum values for each variable were contained within the training
232 set, which is clearly observable in the aforementioned plots.

233 A single hidden layer is often sufficient to approximate any continuous function for
234 geotechnical applications, if enough nodes in the hidden layer are used (Shahin et al., 2002).
235 No explicit rule exists for determining the optimal number of hidden nodes, therefore, it is
236 common practice to employ an iterative stepwise trial-and-error approach from 1 to n number
237 of hidden nodes, where n =2(number of input variables)+1 (Hecht-Nielsen, 1987). However,
238 due to the small number of training samples in the current study (i.e., 123), a relatively large
239 number of connection weights can increase the risk of overfitting the data and subsequently
240 impede the network's ability to obtain a generalized relationship. Therefore, a threshold of 16
241 hidden nodes was implemented for the modelling process in this study.

3.4. Model Performance

3.4.1. Larger ANN Model with 9 Inputs

Firstly, a larger ANN model is developed to predict the peak friction angle considering all the input variables as discussed in the previous section. Concerning the five evaluation metrics previously discussed, the results for each number of hidden nodes, from 1 to 16, are presented in Table 4. It is clear from these results that the performance improves significantly when the number of hidden nodes increases from one to three, after which R^2 values greater than 0.95 are observed. Where three or greater hidden nodes are present, the difference between each network becomes negligible, making it difficult to ascertain which network is optimal. In fact, each model containing at least four hidden nodes was able to predict all 123 ϕ'_{peak} values in the training set to within $\pm$10%. To overcome this, a ranking method, modified after the method introduced by Zorlu et al. (2008), is implemented. Each network receives a score from 1 (worst) to 9 (best) based on its performance on both the training and testing sets, with a 1.5 weight factor applied to the testing score to emphasize the model's performance on the 'unseen' testing data, reflecting its generalization capability. Based on these results, the network containing 8 hidden nodes is the best-performing among the configurations trialed, achieving R^2 values of 0.987 and 0.945, and A-10 indices of 1.000 and 0.935 for the training and testing sets, respectively. However, it is important to note from Table 4 that the difference between the 8-node configuration and several others remains marginal when comparing the actual values of these metrics, rather than the score. Therefore, further analysis is required to determine the desired ANN configuration.

Although the testing set contains unseen data from the ANN model's perspective, it is still related to the data used in the training phase since the same studies and material types were

divided across both datasets. Therefore, a better indicator to isolate the best network is to test its performance on a completely external set of data that is independent of the original database, in conjunction with the original database. This external data was synthesized from Disfani et al. (2017) and Indraratna et al. (2018) using mixtures of fine recycled glass and rubber crumbs (FRG-RC) and traditional subballast (crushed rock), respectively, under CD static triaxial compression. Again, to emphasize the importance of the final model's generalization, a multiplicative weight of 2 was attached to the scores associated with this external dataset. These results are summarized in Table 5, where the model containing 10 hidden nodes (i.e., 9-10-1 input-hidden-output layer node architecture) is deemed the best-performing network. Interestingly, the two best-performing networks on the original database alone (8 and 7 nodes) dropped significantly to 6th and 8th, respectively, when incorporating the external dataset. In fact, the latter returned an A-10 index of only 0.364, meaning almost two thirds of the external dataset had a relative prediction error greater than 10%. This highlights the caution and scrutiny required when assessing these ANN approaches as a model that is slightly less accurate on the original dataset but more generalized with external data is often more attractive for practical purposes. The predictive plots with respect to the target ϕ'_{peak} values for the 9-10-1 ANN model are illustrated in Figure 3, highlighting the strong predictive performance for a variety of granular material types.

3.4.2. Refined ANN Models

Although the larger 9-10-1 (inputs-hidden nodes-outputs) model architecture discussed previously exhibited a strong predictive performance on all three datasets (i.e., training, testing, external), nine input variables may be considered a substantial number. For practical purposes, a reasonably accurate predictive model with fewer model inputs is far more beneficial and feasible. Moreover, for smaller training data sizes, a reduction in the number of input variables

289 may increase a model's generalization since there is less risk of oversaturating the model with
290 degrees of freedom (connection weights). In fact, Rogers and Dowla (1994) suggested that the
291 number of connection weights should not exceed the number of training samples to avoid
292 overfitting of the data. Therefore, the ANN modelling process was repeated using 3–6 input
293 variables. For brevity, only the best-performing models with 5 and 6 input variables (Figure 4)
294 are discussed as the smaller configurations returned inferior predictive performances, likely
295 due to the higher variation and nonlinearity of the materials used in this study.

296 These variables were retained based on the initial requirement of satisfying the three
297 categories of (i) soil stress state, (ii) soil physical state, and (iii) soil type, as discussed
298 previously. Due to the large focus on rubber mixtures and the inferior shear strength of rubber,
299 $RC\%$ was retained. Additionally, out of the three initial particle size indices used, D_{50} provides
300 the greatest information on the type of soil the mixture can be described as. e_0 and γ_d were
301 retained as they provide the model with key information regarding the physical state of the
302 mixtures. Lastly, σ_3' was also retained as it is the only variable representing the stress state and
303 is also well-known to be related to the shear strength of a material as discussed previously. For
304 the 6-input model, C_c was included as an additional input to better describe the mixture type
305 and gradation characteristics. Although C_u has been reported to influence shear strength
306 (Indraratna et al., 2016), its high coefficient of variance (i.e., standard deviation divided by
307 mean) within the dataset used raised potential issues and therefore it was not considered for the
308 refined models.

309 *3.4.3. Performance Comparison*

310 Following the same scoring and ranking process, the performance between each number of
311 input variables is compared in Table 6 and Figure 5, with the predictive plots illustrated in
312 Figure 6. Interestingly, the 5-7-1 model is comparable to the model with 9-10-1 model, even

313 though it is trained using almost half of the number of input variables and less than half of the
314 weight connections (42 vs. 100). This result illustrates the ability of the ANN technique in
315 finding patterns in data and developing relationships between predictor and output variables.
316 However, it also highlights the inherent risk of overcomplicating an ML model with the sole
317 aim of improving training accuracy at the cost of generalization, particularly in applications
318 such as this study where data sizes are limited. Including C_c in the 6-input model clearly has a
319 significant increase on the predictive capabilities compared to the the other two models. Its
320 performance across all three data subsets is either better or equivalent to that of the more
321 complex 9-10-1 model, with the most significant improvements observed for the MARE and
322 A-10 metrics. In fact, only two predictions lie outside the 10% relative error range for the
323 external dataset, with a corresponding mean absolute relative error of 4.0%. Furthermore, while
324 the performance of the 5-7-1 and 9-10-1 models is noticeably diminished from testing to
325 external testing data, these results remain consistent across the two data subsets with the 6-10-
326 1 model. This result is a strong indicator of its generalization ability since the two material
327 types used in the external set are completely independent of the training set. Overall, based on
328 these modelling results, the 6-10-1 ANN model is the optimal configuration to predict the peak
329 friction angle, as a function of $RC\%$, σ_3', e_0, γ_d, C_c, and D_{50}, for a variety of granular materials
330 and mixtures including ballast, sand, subballast, SFS, CW, FRG, and rubber.

331 *3.4.4. Model Sensitivity and Adherence to Geotechnical Principles*

332 Although a model may perform well on training and testing data, an ANN that is validated
333 based on its predictive error alone may lack robustness for varying conditions if the underlying
334 relationships themselves are not accurately captured by the model. Therefore, additional checks
335 are carried out using a parametric study based on a synthetic database where all but one input
336 variable is fixed, with the free variable varied to examine the model's response to this variation.

337 The robustness of the model is then assessed based on its compliance with known geotechnical
338 relationships and behavior. A similar approach is used in this study, although $RC\%$, e_0, and γ_d
339 are interrelated, meaning it is impractical for two of these variables to remain constant while
340 the other does not. Therefore, to assess the robustness of the 6-10-1 model with respect to these
341 three variables, the rubber content is varied, and weight-volume relationships are used to better
342 represent the corresponding void ratio and dry unit weight. For brevity, the sensitivity response
343 for ϕ'_{peak} with respect to $RC\%$, e_0, γ_d, and σ'_3 are included in Figure 7. Expectedly, ϕ'_{peak} is
344 observed to decrease as the rubber content increases, at a diminishing rate, similar to results
345 reported by Indraratna et al. (2018) for SFS-CW-RC mixtures. Likewise, an increase in void
346 ratio also corresponds to a reduction in ϕ'_{peak}, which is again expected since a higher void ratio
347 is representative of a looser matrix, which provides less resistance against particle movement
348 and associated shearing due to reduced particle contacts. Following the same reasoning, ϕ'_{peak}
349 is also expected to increase as the density of the matrix increases, as is reflected in these results.
350 Lastly, the sensitivity response with respect to σ'_3 also adheres to known geotechnical principles
351 since an increase in the confining pressure, and hence normal stress, results in a decrease in
352 ϕ'_{peak} due to the nonlinear failure envelope exhibited by granular materials, particularly at low
353 normal stresses.

354 Attributed to its improved predictive performance based on predictive error, generalization
355 to external data, and adherence to underlying geotechnical principles, the 6-10-1 ANN model
356 is adjudged to be the optimal configuration in predicting the peak friction angle of a granular
357 material. This is enabled by utilizing key geotechnical parameters, coupled with the application
358 of geotechnical engineering knowledge. The rubber content, median particle size, and
359 coefficient of curvature, combined with the initial void ratio and dry unit weight, are able to

satisfactorily describe and differentiate between various material types such as ballast, sand, SFS, CW, and rubber.

4. Modelling of Ballast Particle Breakage Under Cyclic Loading

Accurate estimation of particle breakage is essential for scheduling the maintenance for ballast in railway tracks. A widely used method for measuring and quantifying the particle breakage of ballast is the ballast breakage index (BBI), which is computed based on the shift in the PSD curve (Indraratna et al., 2005, Gu et al., 2022). In empirical models, since multiple parameters such as gradation characteristics, loading conditions, and tamping effort (Selig and Waters, 1994, Indraratna et al., 2016), can influence particle breakage, it is challenging to accurately and comprehensively incorporate all of their effects. In view of this, the second application of ANN in this study discusses the development of a novel predictive BBI model. The database for the current study contains 78 data samples synthesized from previously published research on railway ballast under cyclic loading from large-scale cylindrical triaxial and track process simulation apparatuses (TPSA) (Lackenby et al., 2007, Tennakoon and Indraratna, 2014, Hussaini et al., 2016, Indraratna et al., 2016, Sun et al., 2016, Navaratnarajah and Indraratna, 2017, Jayasuriya et al., 2019, Arachchige et al., 2022, Ngo et al., 2022, Indraratna et al., 2023b). It is noteworthy that BBI can only be measured following the completion of a test, which prevents the collection of intermediate measurements throughout and restricts the size of data available to train these ANN models. To overcome this limitation, a rigorous analysis on the process of selecting the input parameters is presented.

4.1. Parameter Selection for ANN Modelling

From previous experimental investigations using cylindrical and cubical triaxial apparatuses, several parameters, including loading conditions and granular assembly, have been highlighted

383 as influential on BBI. Indraratna et al. (2005) showed that effective track confinement (σ_3') has
384 significant influence on BBI, where they reported a convex (downward) trend with an optimum
385 range of σ_3'. At very low and very high σ_3', BBI increased due to angular corners and particle
386 splitting/crushing, respectively. On the other hand, increasing the magnitude of cyclic
387 deviatoric stress (q_{max}) and frequency (f) increases the particle breakage due to the high
388 interparticle contact stresses at larger deviatoric stress magnitudes (Aursudkij et al., 2009, Sun
389 et al., 2016). Furthermore, the number of loading cycles (N) was considered as an input in the
390 current study as it has a considerable effect on the evolution of breakage, particularly in the
391 initial stages of loading.

392 Besides the loading conditions, the initial state and granular assembly characteristics also
393 affect particle breakage (Selig and Waters, 1994, Nålsund, 2010, Indraratna et al., 2016, Sun
394 and Zheng, 2017). To represent the initial state of ballast, dry density (γ_d) and void ratio (e_o)
395 are considered, while four specific parameters are considered to represent the particle size and
396 gradation characteristics, namely the coefficient of uniformity (C_u), coefficient of curvature
397 (C_c), median particle size (D_{50}), and maximum particle size (D_M). In summary, 10 geotechnical
398 input parameters are identified to develop a predictive ANN model for BBI using 78 large-
399 scale experimental datasets, with their statistical information provided in Table 7. Prior to
400 modelling, outliers in the data were identified and removed to suppress elevated errors.
401 Referring to the maximum values in Table 7, the loading cycles are between three to six orders
402 of magnitude greater than all other parameters. Therefore, to lower the risk of overfitting and
403 incorporate the logarithmic relationship that has been previously reported (Indraratna et al.,
404 2005), the natural logarithm of N (i.e., ln(N)) is used as the input instead of N. For better data
405 visualization, the distribution of input parameters is shown as box plots in Figure 8a, where the
406 input parameters are normalized within the range [0, 1] based on their individual data ranges.

Referring to Figure 8b, the specific experimental apparatus has a bearing on the magnitude of BBI, even if the loading conditions and ballast gradations used are identical, which could be explained by their differing boundary conditions. A categorical input is also included to account for this discrepancy, which uses a binary value of either 0 or 1 to categorize the data from the TPSA and cylindrical triaxial tests, respectively. Though previous studies (Asadi et al. 2022; Koohmishi and Guo, 2023) reported that parent rock type improved the predictions of breakage and abrasive resistance under impact loading, it is not considered in this study as an input due to limited availability of data under cyclic loading. The datasets considered for this study predominantly consists of ballast with latite basalt as parent rock type with a uniaxial compressive strength of 130 MPa.

Figure 9 presents the pairwise relationships and correlation strengths between the modelling parameters within the data used in this study, as well as the distribution of each parameter. As these distributions are non-Gaussian, the Spearman correlation coefficient (SCC) is adopted as it is effective in identifying non-linear relationships between parameters for values within the range [-1, 1]. A higher magnitude of SCC corresponds to a higher correlation strength, while the accompanying sign indicates its direction (i.e., positive or negative). Of the ten numerical input parameters, q_{max} and f possess the strongest relationships with BBI, having SCC values of 0.57 and 0.54, respectively. These observations are corroborated by the empirical models previously developed by Lackenby et al. (2007) and Hussaini and Sweta (2020). Conversely, the correlation analysis indicates there is no direct pairwise relationship between D_M and BBI (SCC = 0.03). This is understandable as BBI is determined on the basis that the largest ballast particles remain unchanged before and after loading (e.g., Indraratna et al. 1998). Therefore, the variable D_M is discarded from the database before modelling. Interestingly, SCC between σ_3' and BBI is also very low at 0.08, although this can be attributed to the low distribution and

variance of σ_3' used in this database. Nevertheless, as discussed earlier, previous investigations showed this parameter has a considerable influence on BBI, thus σ_3' is still considered for the analysis.

4.2. Modelling Framework

The overall modelling framework to predict and forecast ballast breakage is depicted in Figure 10, where there are two main phases. Firstly, as with the friction angle model, the database is partitioned into training and testing sets at proportions of 80% and 20%, respectively. As the LM training algorithm is used, 20% of this training dataset is used for validation to avoid overfitting during the training phase. This use of k-fold (5-fold in this case) validation reduces the risk of biased results for smaller databases (Hastie et al., 2009). The second phase consists of model simulations to determine the optimal input layer configuration. Applying the exhaustive search approach to determine the optimal set of input parameters where ten variables are considered would be computationally large with 1023 possible combinations. Therefore, the forward selection of variables is adopted in this study, where 16 total combinations are analyzed (see Table 8). Since q_{max} and f possess higher SCC values with BBI, these two loading parameters are included in all 20 combinations. These combinations are separated into two groups based on the presence (Group 1) or absence (Group 2) of the binary input parameter.

4.3. Results and Discussion

4.3.1. Original Database

The performance of each ANN model is assessed using the same five evaluation metrics previously stipulated (Eqs. 1-5). Using the results presented in Table 9, these performances are used in conjunction with the weighted ranking method to determine the best input layer

454 configuration. Here, a weight of 1.0 was applied to the calibration sets (training and validation),
455 and 1.5 to the testing set, for reasons similar to the previous friction angle model. The final
456 scores for each model are calculated and summarized in Table 10. Figure 11(a) presents the
457 evolution of each model's MSE across the overall database for the various input configurations,
458 with and without the binary data labelling parameter. Including this binary input significantly
459 improved the prediction accuracy of the model, where the largest relative reduction in MSE is
460 observed at around 40%, and around 15-20% for the configuration with the lowest MSE (i.e.
461 2H). This illustrates the necessity of data distinction in ML models based on the type of
462 laboratory testing and boundary conditions associated with it, especially when multiple sources
463 of data are used for training.

464 Expectedly, the number of input parameters also directly affected the accuracy of the models,
465 where an increase in parameters (up to Group H inclusively) reduced the MSE (Figure 11a)
466 and improved the overall score (Figure 11b). Including only q_{max} and f as the primary input
467 parameters (Combination A) returned the largest errors, while the addition of σ_3' and N
468 (Combinations B-C) significantly reduced this error by 50%, confirming the need for additional
469 loading information. Incorporating gradation parameters (D_{50}, C_u, C_c) with Combinations D-G
470 further reduced this error by around 30% as this provided the model with new information
471 regarding the material's PSD characteristics. Including γ_d (Combination H) minimized the
472 MSE for Group 2, after which a slight increase in error was observed. As the one material (i.e.,
473 latite basalt) is considered in this study, γ_d, in conjunction with the PSD parameters, is therefore
474 sufficient in describing the material's physical state to predict the extent of particle breakage.
475 This is emphasized in Figure 11b, where combination H achieved the highest score in each
476 group, with model 2H achieving the highest overall score.

To further quantify the performance of the ANN models under different input combinations, their predictive capability is visualized using Taylor diagrams (Figure 12). Three indices are used for the assessment and comparison of predictive models: the standard deviation (SD), correlation coefficient (R), and centered root-mean square difference (CRMSD) (Taylor, 2001). Taylor diagrams are especially useful in evaluating multiple aspects of complex models or in gauging the relative skill of many different models. The CRMSD between the predictions and observed BBI is proportional to the point on the x-axis labelled as "RG". The grey contours correspond to the RMSE and SD of the model, which are proportional to the radial distance from the origin. CRMSD quantifies the degree of deviation from the target BBI to the predicted BBI, and R and SD describe the comparability between the predictions and observations. An ideal model should have CRMSD and SD close to "RG", and R close to 1. As shown in Figure 12(a-c) for Group 1 models, R increased from 0.86 to 0.96, while CRMSD decreased from 1.67 to 0.69 when more input parameters were considered for training the model. Likewise, providing information on the testing conditions (Figure 12d-f) generates less error (CRMSD) and higher correlation (R). Therefore, based on both the scoring analysis and Taylor diagram results, it can be deduced that model 2H outperforms all other models and has the optimal configuration of input parameters.

4.3.2. External Validation

Although models 1H and 2H achieved the highest scores using the established database, secondary validation is encouraged to examine the generalized applicability of the models. Therefore, to eliminate the data insufficiency problem, these two optimised ANN models were validated against an external database completely independent of the original calibration (Aursudkij et al., 2009, Nålsund, 2010, Nimbalkar and Indraratna, 2016, Sun and Zheng, 2017, Hussaini and Sweta, 2020). The target (measured) BBI values from this external dataset are

plotted against those predicted by the ANN models 1H and 2H in Figure 13. The predictive ability of both models is in good agreement with the external data samples, although model 2H comfortably surpasses model 1H with an R^2 of 0.89 compared to 0.71. Table 11 shows the overall comparison of both models using all the five evaluation metrics, and it is clear that incorporating the testing condition as an input achieves a model that is more accurate and generalized compared to the model without this binary input, where the number of samples beyond 10% deviations are significantly less.

Meanwhile, for their application to field data from Australian tracks (Nimbalkar and Indraratna, 2016), model 2H outperforms 1H by a very large margin and closely predicts the measured values (see Figures 13a and b). As the cubical triaxial setup more closely resembles the actual field conditions compared to the cylindrical triaxial setup, including the binary input significantly improved the model's performance. Also, as the ballast type used by Nimbalkar and Indraratna (2016) is the same as that of the training data (i.e., basalt), these model predictions closely corroborate with the measured data. Although the model 2H also performs better for ballast having quartzite (Nalsund 2010) and limestone (Aursudkij et a., 2009) as the parent rock with low-to-medium BBI values, a relatively high error can be observed for granite aggregates with high BBI (Hussaini and Sweta, 2022). As the parent rock type and rock minerology are not considered for training, these deviations are expected and caution should be further exercised when using this model to extrapolate beyond the material types used for training. Based on detailed analysis of the prediction performances of the ANN models under different input combinations, the ANN model with N, σ_3', q_{max}, f, D_{50}, C_u, C_c, γ_d and binary input DT performs the best. With prior information on loading conditions, ballast gradation and tamping density, this model can be useful to practitioners to directly estimate the breakage in railway tracks.

5. Limitations of the Study

Despite the accuracy of the two ANN models developed in this study in predicting peak friction angle and degradation of track granular materials, there are some notable limitations as follows:

1. For both models, a relatively small number of data samples were used for development due to the limited availability of data and hence, the models are constrained to the data ranges considered. For example., the size of the database in the friction angle ANN model was restricted to not oversaturate it with traditional aggregates (i.e., pure sand and ballast). Therefore, it is suggested that the proposed model be recalibrated in the future with the availability of more data.
2. Material properties such as particle hardness and shape, and the relative size of rubber granules were not considered in this study which may also influence a mixture's shear strength.
3. For the breakage ANN model, the data considered for this study is limited to latite basalt and the extrapolation to any other parent rock type such as granite, limestone and quartzite yields deviations in predictions. Also, as the model is developed based on cylindrical triaxial testing and cubical triaxial testing (plane strain conditions), the influence of principal stress rotation that can exacerbate BBI under high train speeds cannot be predicted by this model (Malisetty et al. 2020, Powrie et al., 2007).

6. Conclusions

In this paper, two applications of ANN for track granular materials were developed with critical evaluation of the selection process of input parameters. The first application presented a generalised data model that predicts the peak friction angle of several track granular materials made of different mixtures (ballast, sand, sub-ballast, steel furnace slag, coal wash, fine recycled glass, and granulated rubber). The second application presented an ANN model for

549 predicting the particle breakage of ballast (via BBI) under repeated load cycles. Adhering to
550 the geotechnical context of the materials, optimised input parameter selection process was
551 conducted through two steps: (i) identifying geotechnically important parameters that are
552 practically easy to obtain, and (ii) deciding the optimum combination of these parameters based
553 on model metrics. Salient findings from this study are as follows:

554 1. For predicting ϕ'_{peak} of track granular materials, this study proposes using the six
555 optimum input parameters of $RC\%$, σ'_3, e_0, γ_d, D_{50}, C_c, with at least one representing
556 each of the three key modelling categories of soil stress state, physical state and type.
557 The 6-input model resulted in superior performance in all three (training, testing and
558 external) data sets achieving R^2, RMSE, MARE, VAF and A-10 values within the range
559 of 0.90-0.99, 0.85-2.44, 1.3-4.0, 88.8-98.8, and 0.909-1, respectively. The parametric
560 study also revealed that the 6-input model response complied with known geotechnical
561 principles with respect to changes in $RC\%$, σ'_3, e_0, γ_d.
562 2. Considering both an insufficient and excess number of parameters representing particle
563 gradation characteristics resulted in reduced model performance. For example, the 5-
564 input model with D_{50} as the only gradation parameter and the 9-input model with 5
565 gradation parameters ($D_{10}, D_{50}, D_{max}, C_u, C_c$) both performed at about 65% accuracy
566 (i.e., overall performance score) to that of the optimum 6-input model. This leads to the
567 conclusion that a simplified model using an optimum number of variables is more
568 efficient in predicting the peak friction angle and reducing the error associated with
569 parameter redundancy.
570 3. For the prediction of ballast breakage under cyclic loading, nine key input parameters
571 are identified such as q_{max}, $f, N, \sigma'_3, C_c, C_u, D_{50}, \gamma_d$ and e_0, mentioned in the order of
572 importance. Considering only loading parameters (q_{max}, f, N) produced high error

(MSE~2.0) with low predictive capability, while the inclusion of gradation and initial state properties of ballast reduced the error (MSE~0.5-1). The combination of 8-inputs ($N, \sigma_3', q_{max}, f, D_{50}, C_u, C_c$ and γ_d) yielded superior model performance, with a notable exclusion of e_0 which had marginal or no improvement in performance. In essence, this result highlights the significance of including the initial physical state, loading and gradation parameters together for accurately predicting ballast breakage.

4. Furthermore, the inclusion of laboratory testing type (cylindrical triaxial or cubical triaxial) as the 9th input improved the accuracy by 20% during testing when compared to models without this distinction. By considering cubical triaxial testing data with plain strain conditions, the 9-input model predicted the BBI measured from Australian railway tracks and other external data sets with an R^2 of 0.87, significantly higher than that of the 8-input model (0.71). This emphasises the importance of incorporating the boundary conditions used for data collection and their resemblance to actual conditions.

7. Acknowledgment

The authors would like to thank the financial support from the Australian Research Council (LP200200915) and acknowledge the financial and technical support from industry partners including Sydney Trains, Australian Rail Transport Corporation (ARTC), SMEC Australia Pty. Limited, Bentley Systems Pty. Limited, and Bestech Australia Pty. Limited.

References

Abozraig, M., Ok, B. and Yildiz, A. 2022. Determination of shear strength of coarse-grained soils based on their index properties: a comparison between different statistical approaches. Arabian Journal of Geosciences, 15, 593.

595 Arachchige, C. M. K., Indraratna, B., Qi, Y., Vinod, J. S. and Rujikiatkamjorn, C. 2022.
596 Deformation and degradation behavior of rubber intermixed ballast system under cyclic
597 loading. Engineering Geology, 307, 106786.
598 Arulrajah, A., Disfani, M. M., Horpibulsuk, S., Suksiripattanapong, C., and Prongmanee, N.
599 (2014). Physical properties and shear strength responses of recycled construction and
600 demolition materials in unbound pavement base/subbase applications. Construction and
601 Building Materials, 58, 245–257.
602 Australian Standards, 2015. As 2758.7 aggregates and rock for engineering purposes, Part 7:
603 Railway Ballast. Sydney, Nsw, Australia.
604 Aursudkij, B., Mcdowell, G. and Collop, A. 2009. Cyclic loading of railway ballast under
605 triaxial conditions and in a railway test facility. *Granular Matter,* 11**,** 391.
606 BIS Oxford Economics, 2022. Australian rail market outlook. Australian Railway Association.
607 Chan, W., Chow, Y. and Liu, L. 1995. Neural network: an alternative to pile driving formulas.
608 Computers And Geotechnics, 17, 135-156.
609 Disfani, M. M., Tsang, H.-H., Arulrajah, A. and Yaghoubi, E. 2017. Shear and compression
610 characteristics of recycled glass-tire mixtures. Journal of Materials in Civil Engineering,
611 29, 06017003.
612 Edil, T. and Bosscher, P. 1994. Engineering properties of tire chips and soil mixtures.
613 Geotechnical Testing Journal, 17, 453-464.
614 Edincliler, A., Cabalar, A. F., Cagatay, A. and Cevik, A. 2012. Triaxial compression behavior
615 of sand and tire wastes using neural networks. Neural Computing and Applications, 21,
616 441-452.
617 Einav, I. 2007. Breakage mechanics—Part II: Modelling granular materials. Journal of the
618 Mechanics and Physics Of Solids, 55, 1298-1320.

619 Fausett, L. V. 1994. Fundamentals neural networks: Architecture, algorithms and applications,
620 Prentice-Hall, Englewood Cliffs, New Jersey.
621 Flood, I. and Kartam, N. 1994. Neural networks in civil engineering. I: Principles and
622 understanding. Journal of Computing in Civil Engineering, 8, 131-148.
623 Foose, G. J., Benson, C. H. and Bosscher, P. J. 1996. Sand reinforced with shredded waste tires.
624 Journal of Geotechnical Engineering, 122, 760-767.
625 Foresee, F. D. and Hagan, M. T. Gauss-newton approximation to bayesian learning.
626 Proceedings of International Conference on Neural Networks (ICNN'97), 12-12 June
627 1997. 1930-1935 Vol.3.
628 Gomes Correia, A., Cortez, P., Tinoco, J., and Marques, R. (2013). Artificial intelligence
629 applications in transportation geotechnics. Geotechnical and Geological Engineering,
630 31(3), 861–879.
631 Gu, Q., Zhao, C., Bian, X., Morrissey, J. P. and Ooi, J. Y. 2022. Trackbed settlement and
632 associated ballast degradation due to repeated train moving loads. Soil Dynamics and
633 Earthquake Engineering, 153, 107109.
634 Habibagahi, G. and Bamdad, A. 2003. A neural network framework for mechanical behavior
635 of unsaturated soils. Canadian Geotechnical Journal, 40, 684-693.
636 Hastie, T., Tibshirani, R., Friedman, J. H. and Friedman, J. H. 2009. The elements of statistical
637 learning: data mining, inference, and prediction. Vol. 2, pp. 1-758. New York: Springer.
638 Hecht-Nielsen, R. Kolmogorov's mapping neural network existence theorem. Proceedings of
639 the International Conference on Neural Networks, 1987. IEEE Press New York, NY, USA,
640 11-14.
641 Hunt, H., Indraratna, B. and Qi, Y. 2023. Ductility and energy absorbing behavior of coal wash
642 – rubber crumb mixtures. International Journal of Rail Transportation, 11, 508-528.

643 Hussaini, S. K. K., Indraratna, B. and Vinod, J. S. 2016. A laboratory investigation to assess
644 the functioning of railway ballast with and without geogrids. Transportation Geotechnics,
645 6, 45-54.
646 Hussaini, S. K. K. and Sweta, K. 2020. Investigation of deformation and degradation response
647 of geogrid-reinforced ballast based on model track tests. Proceedings of the Institution of
648 Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 235, 505-517.
649 Indraratna, B., Armaghani, D. J., Gomes Correia, A., Hunt, H. and Ngo, T. 2023a. Prediction
650 of resilient modulus of ballast under cyclic loading using machine learning techniques.
651 Transportation Geotechnics, 38, 100895.
652 Indraratna, B., Ionescu, D. and Christie, H. 1998. Shear behavior of railway ballast based on
653 large-scale triaxial tests. Journal of Geotechnical and Geoenvironmental Engineering, 124,
654 439-449.
655 Indraratna, B., Lackenby, J. and Christie, D. 2005. Effect of confining pressure on the
656 degradation of ballast under cyclic loading. Géotechnique, 55, 325-328.
657 Indraratna, B., Ngo, T. and Rujikiatkamjorn, C. 2020. Performance of ballast influenced by
658 deformation and degradation: laboratory testing and numerical modeling. International
659 Journal of Geomechanics, 20, 04019138.
660 Indraratna, B., Qi, Y. and Heitor, A. 2018. Evaluating the properties of mixtures of steel furnace
661 slag, coal wash, and rubber crumbs used as subballast. Journal of Materials in Civil
662 Engineering, 30, 04017251.
663 Indraratna, B., Rujikiatkamjorn, C. and Salim, W. 2023b. Advanced Rail Geotechnology –
664 Ballasted Track (2nd Ed.), CRC Press.
665 Indraratna, B., Rujikiatkamjorn, C., Tawk, M. and Heitor, A. 2019. Compaction, degradation
666 and deformation characteristics of an energy absorbing matrix. Transportation
667 Geotechnics, 19, 74-83.

668 Indraratna, B., Sun, Y. and Nimbalkar, S. 2016. Laboratory assessment of the role of particle
669 size distribution on the deformation and degradation of ballast under cyclic loading.
670 Journal of Geotechnical and Geoenvironmental Engineering, 142, 04016016.
671 Jayasuriya, C., Indraratna, B. and Ngoc Ngo, T. 2019. Experimental study to examine the role
672 of under sleeper pads for improved performance of ballast under cyclic loading.
673 Transportation Geotechnics, 19, 61-73.
674 Kaastra, I. and Boyd, M. 1996. Designing a neural network for forecasting financial and
675 economic time series. Neurocomputing, 10, 215-236.
676 Kaliboullah, C. I. 2016. Behavior of compacted coalwash under saturated condition
677 incorporating particle breakage. Doctor of Philosophy, University of Wollongong,
678 Australia.
679 Lackenby, J., Indraratna, B., Mcdowell, G. and Christie, D. 2007. Effect of confining pressure
680 on ballast degradation and deformation under cyclic triaxial loading. Géotechnique, 57,
681 527-536.
682 Malisetty, R. S., Indraratna, B. and Vinod, J. 2020. Behavior of ballast under principal stress
683 rotation: multi-laminate approach for moving loads. Computers and Geotechnics, 125,
684 103655.
685 Mashiri, M. S., Vinod, J. S., Sheikh, M. N. and Tsang, H.-H. 2015. Shear strength and dilatancy
686 behavior of sand–tyre chip mixtures. Soils And Foundations, 55, 517-528.
687 Nålsund, R. 2010. Effect of grading on degradation of crushed-rock railway ballast and on
688 permanent axial deformation. Transportation Research Record, 2154, 149-155.
689 Navaratnarajah, S. K. and Indraratna, B. 2017. Use of rubber mats to improve the deformation
690 and degradation behavior of rail ballast under cyclic loading. Journal of Geotechnical and
691 Geoenvironmental Engineering, 143, 04017015.

692 Ngo, T., Indraratna, B. and Ferreira, F. 2022. Influence of synthetic inclusions on the
693 degradation and deformation of ballast under heavy-haul cyclic loading. International
694 Journal of Rail Transportation, 10, 413-435.
695 Nimbalkar, S. and Indraratna, B. 2016. Improved performance of ballasted rail track using
696 geosynthetics and rubber shockmat. Journal of Geotechnical and Geoenvironmental
697 Engineering, 142, 04016031.
698 Noorzad, R. and Raveshi, M. 2017. Mechanical behavior of waste tire crumbs–sand mixtures
699 determined by triaxial tests. Geotechnical and Geological Engineering, 35, 1793-1802.
700 Penumadu, D. and Zhao, R. 1999. Triaxial compression behavior of sand and gravel using
701 artificial neural networks (ANN). Computers and Geotechnics, 24, 207-230.
702 Phoon, K.-K. and Zhang, W. 2023. Future of machine learning in geotechnics. Georisk:
703 Assessment and Management of Risk for Engineered Systems and Geohazards, 17, 7-22.
704 Powrie, W., Yang, L. A., and Clayton, C. R. I. (2007). Stress changes in the ground below
705 ballasted railway track during train passage. Proceedings of the Institution of Mechanical
706 Engineers, Part F: Journal of Rail and Rapid Transit, 221(2), 247–262.
707 Qi, Y., Indraratna, B., Heitor, A. and Vinod, J. S. 2018. Effect of rubber crumbs on the cyclic
708 behavior of steel furnace slag and coal wash mixtures. Journal of Geotechnical and
709 Geoenvironmental Engineering, 144, 04017107.
710 Rogers, L. L. and Dowla, F. U. 1994. Optimization of groundwater remediation using artificial
711 neural networks with parallel solute transport modeling. Water Resources Research, 30,
712 457-481.
713 Rujikiatkamjorn, C., Indraratna, B. and Chiaro, G. 2013. compaction of coal wash to optimise
714 its utilisation as water-front reclamation fill. Geomechanics and Geoengineering, 8, 36-45.
715 Samui, P. 2008. Prediction of friction capacity of driven piles in clay using the support vector
716 machine. Canadian Geotechnical Journal, 45, 288-295.

717 Selig, E. T. and Waters, J. M. 1994. Track geotechnology and substructure management,
718 Thomas Telford.
719 Shahin, M. A. and Indraratna, B. 2006. Modeling the mechanical behavior of railway ballast
720 using artificial neural networks. Canadian Geotechnical Journal, 43, 1144-1152.
721 Shahin, M. A., Maier, H. R. and Jaksa, M. B. 2002. Predicting settlement of shallow
722 foundations using neural networks. Journal of Geotechnical and Geoenvironmental
723 Engineering, 128, 785-793.
724 Sheikh, M. N., Mashiri, M. S., Vinod, J. S. and Tsang, H.-H. 2013. Shear and compressibility
725 behavior of sand–tire crumb mixtures. Journal of Materials in Civil Engineering, 25, 1366-
726 1374.
727 Sun, Q. D., Indraratna, B. and Nimbalkar, S. 2016. Deformation and degradation mechanisms
728 of railway ballast under high frequency cyclic loading. Journal of Geotechnical and
729 Geoenvironmental Engineering, 142, 04015056.
730 Sun, Y. and Zheng, C. 2017. Breakage and shape analysis of ballast aggregates with different
731 size distributions. Particuology, 35, 84-92.
732 Tasalloti, A. 2015. Behavior of blended waste materials for land reclamation for port extension.
733 Doctor of Philosophy, University of Wollongong, Australia.
734 Tawk, M. and Indraratna, B. 2021. Role of rubber crumbs on the stress-strain response of a
735 coal wash matrix. Journal of Materials in Civil Engineering, 33, 04020480.
736 Tennakoon, N. and Indraratna, B. 2014. Behavior of clay-fouled ballast under cyclic loading.
737 Géotechnique, 64, 502-506.
738 Zhou, J., Li, E., Wei, H., Li, C., Qiao, Q. and Armaghani, D. J. 2019. Random forests and cubist
739 algorithms for predicting shear strengths of rockfill materials. Applied Sciences, 9, 1621.

740 Zorlu, K., Gokceoglu, C., Ocakoglu, F., Nefeslioglu, H. and Acikalin, S. 2008. Prediction of
741 uniaxial compressive strength of sandstones using petrography-based models. Engineering
742 Geology, 96, 141-158.

743 **List of Tables:**

744 Table 1 Input variables investigated for peak friction angle predictive model

Soil Model Category	Variable	Symbol
(i) Stress state	Effective confining pressure	σ_3'
(ii) Physical state	Initial void ratio	e_0
	Dry unit weight	γ_d
(iii) Soil type	Gravimetric rubber content	$RC\%$
	Coefficient of uniformity	C_u
	Coefficient of curvature	C_c
	Effective particle size	D_{10}
	Median particle size	D_{50}
	Maximum particle size of rigid material	$D_{max,rigid}$

745

746 Table 2 Summary of material types and number of tests synthesised for peak friction angle
747 database

Rigid Material	Tests Without Rubber	Tests With Rubber	References
Ballast	17	9	Arachchige et al. (2022) Indraratna et al. (1998)
Sand	7	59	Edincliler et al. (2012) Sheikh et al. (2013) Noorzad and Raveshi (2017)
SFS	4	0	Tasalloti (2015)
CW	21	11	Tasalloti (2015) Kaliboullah (2016) Indraratna et al. (2019) Tawk and Indraratna (2021)
SFS-CW	15	11	Tasalloti (2015) Qi et al. (2018)

748

749

750 Table 3 Database statistics for peak friction angle model

Variable	Variable Type	Unit	Min.	Max.	Mean	Std. Dev.
$RC\%$	Input	%	0	40	9.5	11.3
σ_3'	Input	kPa	8	483	102.2	97.4
e_0	Input	-	0.29	0.88	0.57	0.17
γ_d	Input	kN/m^3	9.1	20.4	14.96	2.1
C_u	Input	-	1.5	83.1	16.9	23.8
C_c	Input	-	0.9	4.1	1.6	0.90
D_{10}	Input	mm	0.013	27.1	3.6	7.7
D_{50}	Input	mm	0.19	38.9	6.6	12.2
$D_{max,rigid}$	Input	mm	0.30	53	15.0	18.6
ϕ'_{peak}	Output	°	26.4	66.5	44.9	8.1

751

752

Table 4 Ranking of number of hidden nodes based on performance (friction angle model) on individual training and testing data subsets

Hidden	R^2			RMSE			MARE (%)			VAF (%)			A-10			Final	
Nodes	**Train**	**Test**	**Score**	**Train**	**Test**	**Score**	**Train**	**Test**	**Score**	**Train**	**Test**	**Score**	**Train**	**Test**	**Score**	**Score**	**Rank**
1	0.850	0.807	2.5	3.01	4.12	2.5	5.12	7.32	2.5	85.0	80.6	2.5	0.846	0.677	2.5	12.5	16
2	0.890	0.858	5	2.58	3.66	5	4.43	6.92	5	89.0	85.6	5	0.919	0.774	5	25	15
3	0.957	0.908	9	1.62	3.00	7.5	2.82	5.31	7.5	95.7	90.8	9	0.992	0.871	10.5	43.5	14
4	0.969	0.906	8.5	1.37	2.98	10	2.41	5.09	10	96.9	90.5	8.5	1.000	0.839	8.5	45.5	13
5	0.981	0.944	26	1.08	2.49	21.5	1.82	4.42	17	98.1	94.2	26	1.000	0.903	16	106.5	8
6	0.983	0.945	28.5	1.01	2.35	28.5	1.68	3.96	27	98.3	94.5	30	1.000	0.903	16	130	5
7	0.987	0.939	33.5	0.89	2.45	35	1.42	3.89	38	98.7	93.9	33.5	1.000	0.903	16	156	2
8	**0.987**	**0.945**	**40**	**0.88**	**2.34**	**40**	**1.42**	**3.96**	**37.5**	**98.7**	**94.4**	**38.5**	**1.000**	**0.935**	**26.5**	**182.5**	**1**
9	0.985	0.923	17.5	0.94	2.68	17.5	1.50	4.55	17	98.5	92.3	17.5	1.000	0.839	8.5	78	12
10	0.987	0.935	31.5	0.89	2.52	30	1.49	4.38	26	98.7	93.3	30	1.000	0.903	16	133.5	4
11	0.986	0.934	28	0.90	2.47	31	1.48	4.35	30	98.6	93.4	29.5	1.000	0.903	16	134.5	3
12	0.985	0.933	24.5	0.94	2.54	24.5	1.46	4.38	29.5	98.5	93.3	24.5	1.000	0.903	16	119	7
13	0.984	0.937	25	1.00	2.45	26.5	1.63	4.41	20.5	98.4	93.7	25	1.000	0.935	26.5	123.5	6
14	0.985	0.928	17	0.96	2.62	17	1.50	4.46	19.5	98.5	92.8	17	1.000	0.871	11.5	82	11
15	0.985	0.928	19.5	0.95	2.60	21	1.49	4.23	29.5	98.5	92.8	19.5	1.000	0.903	16	105.5	9
16	0.986	0.931	24	0.91	2.62	22.5	1.40	4.59	23.5	98.6	93.0	24	1.000	0.871	11.5	105.5	9

Table 5 Ranking of number of hidden nodes based on performance (friction angle model) including external testing data

		R^2		RMSE		MARE (%)		VAF (%)		A-10			
Hidden Nodes	**Original Score**	**External Test**	**Total Score***	**External Test**	**Total Score***	**External Test**	**Total Score***	**External Test**	**Total Score***	**External Test**	**Total Score***	**Final Score**	**Rank**
1	12.5	0.823	18.5	4.30	12.5	10.11	12.5	82.2	20.5	0.455	8.5	72.5	16
2	25	0.806	15	4.00	19	8.46	19	80.5	17	0.591	15	85	15
3	43.5	0.630	11	4.11	19.5	8.03	23.5	63.0	13	0.773	30.5	97.5	14
4	45.5	0.752	14.5	3.98	26	7.79	28	71.3	14.5	0.773	28.5	111.5	13
5	106.5	0.730	30	6.92	23.5	14.77	19	6.5	28	0.500	24	124.5	12
6	130	0.807	40.5	5.11	34.5	12.38	33	80.6	44	0.444	20	172	11
7	156	0.911	57.5	5.61	39	13.97	42	82.2	49.5	0.364	18	206	8
8	182.5	0.754	48	4.80	48	10.46	45.5	71.6	46.5	0.636	38.5	226.5	6
9	78	0.941	49.5	1.90	49.5	4.06	49	92.6	49.5	0.955	40.5	238	5
10	**133.5**	**0.894**	**53.5**	**2.69**	**58**	**6.16**	**54**	**88.9**	**58**	**0.727**	**32**	**255.5**	**1**
11	134.5	0.862	46	3.14	55	6.63	54	86.1	53.5	0.773	36	244.5	4
12	119	0.913	50.5	3.03	50.5	7.09	51.5	86.8	50.5	0.818	42	245	3
13	123.5	0.938	55	3.90	44.5	9.18	32.5	84.6	45	0.682	40.5	217.5	7
14	82	0.869	37	3.25	39	7.35	39.5	85.6	39	0.727	27.5	182	10
15	105.5	0.815	33.5	3.72	41	6.61	55.5	73.5	29.5	0.864	44	203.5	9
16	105.5	0.928	52	2.51	52.5	5.86	53.5	92.5	54	0.864	39.5	251.5	2

* Total score is the sum of training, testing, and external testing scores

760 Table 6 Performance comparison between initial 9-input model and refined 5-input and 6-input friction angle models

Model	R^2			RMSE			MARE (%)			VAF (%)			A-10			Final	
	Train	Test	Ext.	Train	Test	Ext.	Train	Test	Ext.	Train	Test	Ext.	Train	Test	Ext.	Score	Rank
5-7-1	0.97	0.94	0.83	1.27	2.46	2.80	2.0	4.0	5.2	97.3	94.0	79.1	1.000	0.968	0.864	35.5	2
6-10-1	**0.99**	**0.94**	**0.90**	**0.85**	**2.41**	**2.44**	**1.3**	**4.1**	**4.0**	**98.8**	**94.0**	**88.8**	**1.000**	**0.903**	**0.909**	**53.5**	**1**
9-10-1	0.99	0.94	0.89	0.89	2.52	2.69	1.5	4.4	6.2	98.7	93.3	88.9	1.000	0.903	0.727	34.5	3

761 Note: 'Ext.' refers to the external dataset

762

763 Table 7 Statistical information of the input parameters and BBI used in the present study

Parameter	Cylindrical triaxial tests				TPSA tests				Remarks
	Min	Max	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	
N	1000	500000	464172	126610	5000	500000	337000	2169256	
σ_3' (kPa)	10	240	43.8	32.5	7	20	13.4	3.6	Category 1: Loading characteristics
q_{max} (kPa)	230	370	249.3	48.7	230	460	301	99.4	
f (Hz)	5	60	24.7	12.4	15	25	17	3.4	
D_M (mm)	53	53	53	0	53	63	62.5	2.2	
D_{50} (mm)	27.2	49.4	38.8	4.3	33.5	43	35.8	3.2	Category 2: Particle size distribution
C_u	1.2	4.5	1.9	0.70	1.6	2.5	1.9	0.42	
C_c	1	1.4	1.1	0.12	1	1.3	1.1	0.13	
γ_d (kN/m^3)	15	15.6	15.3	0.1	15.2	15.3	15.3	0	Category 3: Ballast initial state
e_o	0.53	0.82	0.73	0.05	0.72	0.77	0.74	0.02	
BBI (%)	1.2	17.7	6.6	3.3	6	13.4	10.3	2.6	Output

764

765 Table 8 Different input combinations considered for the breakage model

Groups	**Inputs**	**Remarks**
Reference group	$N, \sigma_3', q_{max}, f, D_{50}, C_u, C_c, \gamma_d, e_o, DT$	referred as "RG"
Group 1A	q_{max}, f	Combination 1: Loading conditions alone
Group 1B	q_{max}, f, σ_3'	
Group 1C	q_{max}, f, σ_3', N	
Group 1D	$q_{max}, f, \sigma_3', N, D_{50}$	Combination 2: Loading + PSD parameters
Group 1E	$q_{max}, f, \sigma_3', N, D_{50}, C_c$	
Group 1F	$q_{max}, f, \sigma_3', N, D_{50}, C_u$	
Group 1G	$q_{max}, f, \sigma_3', N, D_{50}, C_u, C_c$	
Group 1H	$N, \sigma_3', q_{max}, f, D_{50}, C_u, C_c, \gamma_d$	Combination 3: Loading + PSD + Initial state properties
Group 1I	$N, \sigma_3', q_{max}, f, D_{50}, C_u, C_c, e_o$	
Group 1J	$N, \sigma_3', q_{max}, f, D_{50}, C_u, C_c, \gamma_d, e_o$	
Group 2 (A to J)	In Group 2, DT (binary input) is added to the ANN model	

766 Note: DT refers to binary input

767 Table 9 Performance indices of breakage models under Group 1 and 2

	R^2			RMSE			MARE (%)			VAF (%)			A-10		
Group	**Train**	**Test**	**Val.**	**Train**	**Test**	**Val.**	**Train**	**Test**	**Val.**	**Train**	**Test**	**Val.**	**Train**	**Test**	**Val.**
1A	0.79	0.73	0.82	1.72	1.33	1.31	32.7	27.1	24.3	78.5	72.5	81.3	0.500	0.440	0.630
1B	0.71	0.83	0.79	1.95	0.95	1.39	30.7	17.0	21.8	70.7	82.4	79.0	0.610	0.500	0.630
1C	0.87	0.75	0.92	1.30	1.14	0.90	20.9	21.2	14.8	86.8	74.2	91.9	0.670	0.630	0.810
1D	0.88	0.75	0.94	1.29	1.18	0.83	17.6	24.6	14.5	87.3	75.3	93.1	0.670	0.560	0.750
1E	0.90	0.66	0.89	1.16	1.30	1.02	14.9	26.5	16.3	89.5	66.0	88.5	0.760	0.560	0.690
1F	0.87	0.73	0.90	1.40	1.15	0.95	14.7	20.5	13.5	85.8	73.4	90.1	0.780	0.500	0.810
1G	0.93	0.81	0.90	0.99	1.03	0.97	16.9	20.0	16.7	92.6	79.5	89.8	0.830	0.690	0.630
1H	0.97	0.83	0.98	0.67	0.94	0.69	11.4	19.1	12.7	96.8	82.5	97.4	0.910	0.630	0.810
1I	0.95	0.81	0.96	0.83	0.99	0.62	10.5	18.0	8.7	94.6	80.7	96.1	0.870	0.630	0.880
1J	0.97	0.80	0.97	0.67	1.02	0.58	10.1	19.5	7.9	96.6	80.6	96.5	0.910	0.560	0.810
2A	0.79	0.88	0.85	1.66	0.84	1.18	27.7	17.0	21.3	78.6	86.1	84.5	0.700	0.690	0.750
2B	0.80	0.86	0.84	1.60	0.86	1.19	28.6	16.0	20.1	80.1	85.5	84.1	0.670	0.690	0.690
2C	0.90	0.76	0.94	1.17	1.17	0.87	16.9	26.3	14.5	89.4	76.4	92.8	0.670	0.560	0.690
2D	0.94	0.76	0.91	0.85	1.22	1.00	11.7	22.3	13.0	94.3	70.7	90.6	0.850	0.630	0.750
2E	0.93	0.79	0.92	0.95	1.08	0.86	11.9	20.5	10.7	93.1	76.8	91.7	0.850	0.560	0.810
2F	0.91	0.78	0.94	1.08	1.04	0.74	14.0	22.2	12.3	90.9	78.3	94.3	0.800	0.630	0.690
2G	0.95	0.78	0.97	0.78	1.10	0.52	9.2	21.1	8.4	95.3	75.8	97.0	0.910	0.500	0.940
2H	0.97	0.90	0.98	0.64	0.70	0.52	9.1	13.9	9.0	96.9	90.2	97.5	0.930	0.690	0.880
2I	0.94	0.82	0.98	0.85	0.96	0.57	10.7	19.8	9.9	94.4	82.1	97.6	0.870	0.630	0.750
2J	0.96	0.86	0.96	0.77	0.99	0.74	10.7	17.9	10.4	95.7	80.2	96.4	0.890	0.630	0.810

768

769

770 Table 10 Weighted scores and ranking of breakage models under Group 1 and 2 based on performance metrics

	Score																
	R^2			RMSE			MARE (%)			VAF (%)			A-10			Total	
Group	Train[a]	Test[a]	Val.[b]	Train[a]	Test[a]	Val.[b]	Train[a]	Test[a]	Val.[b]	Train[a]	Test[a]	Val.[b]	Train[a]	Test[a]	Val.[b]	Score	Rank
1A	2	3	2	2	1.5	2	1	1.5	1	2	4.5	2	1	1.5	1	28	19
1B	1	15	1	1	22.5	1	2	25.5	2	1	24	1	2	3	1	103	14
1C	4	4.5	8	6	10.5	9	5	10.5	7	6	7.5	10	3	6	4	101	16
1D	5	4.5	9	7	6	12	6	6	8	7	9	12	3	4.5	3	102	15
1E	6	1.5	5	9	3	5	9	3	6	9	1.5	5	5	4.5	2	74.5	18
1F	4	3	6	5	9	8	10	13.5	10	5	6	7	6	3	4	99.5	17
1G	8	12	6	11	16.5	7	7	16.5	5	11	16.5	6	8	7.5	1	139	10
1H	12	15	12	17	24	14	14	21	12	19	25.5	18	12	6	4	225.5	2
1I	10	12	10	14	19.5	15	17	22.5	18	15	21	14	10	6	5	209	5
1J	12	10.5	11	17	18	16	18	19.5	20	18	19.5	16	12	4.5	4	216	3
2A	2	18	4	3	27	4	4	27	3	3	28.5	4	4	7.5	3	142	9
2B	3	16.5	3	4	25.5	3	3	28.5	4	4	27	3	3	7.5	2	137	11
2C	6	6	9	8	7.5	10	8	4.5	9	8	12	11	3	4.5	2	108.5	13
2D	9	6	7	13	4.5	6	13	7.5	11	13	3	8	9	6	3	119	12
2E	8	9	8	12	13.5	11	12	15	14	12	13.5	9	9	4.5	4	154.5	7
2F	7	7.5	9	10	15	13	11	9	13	10	15	13	7	6	2	147.5	8
2G	10	7.5	11	15	12	18	19	12	19	16	10.5	17	12	3	6	188	6
2H	**12**	**19.5**	**12**	**18**	**28.5**	**18**	**20**	**30**	**17**	**20**	**30**	**19**	**13**	**7.5**	**5**	**269.5**	**1**
2I	9	13.5	12	13	21	17	16	18	16	14	22.5	20	10	6	3	211	4
2J	11	16.5	10	16	19.5	13	15	24	15	17	18	15	11	6	4	211	4

771 [a] weight = 1.0; [b] weight = 1.5

772 Table 11 Performance comparison of breakage models to external validation dataset

ANN Model	RMSE	R^2	MARE (%)	VAF (%)	A-10
1H	1.91	0.71	17.9	65.1	0.400
2H	1.31	0.89	10.0	84.3	0.900

773

774

List of Figures:

List of figures:

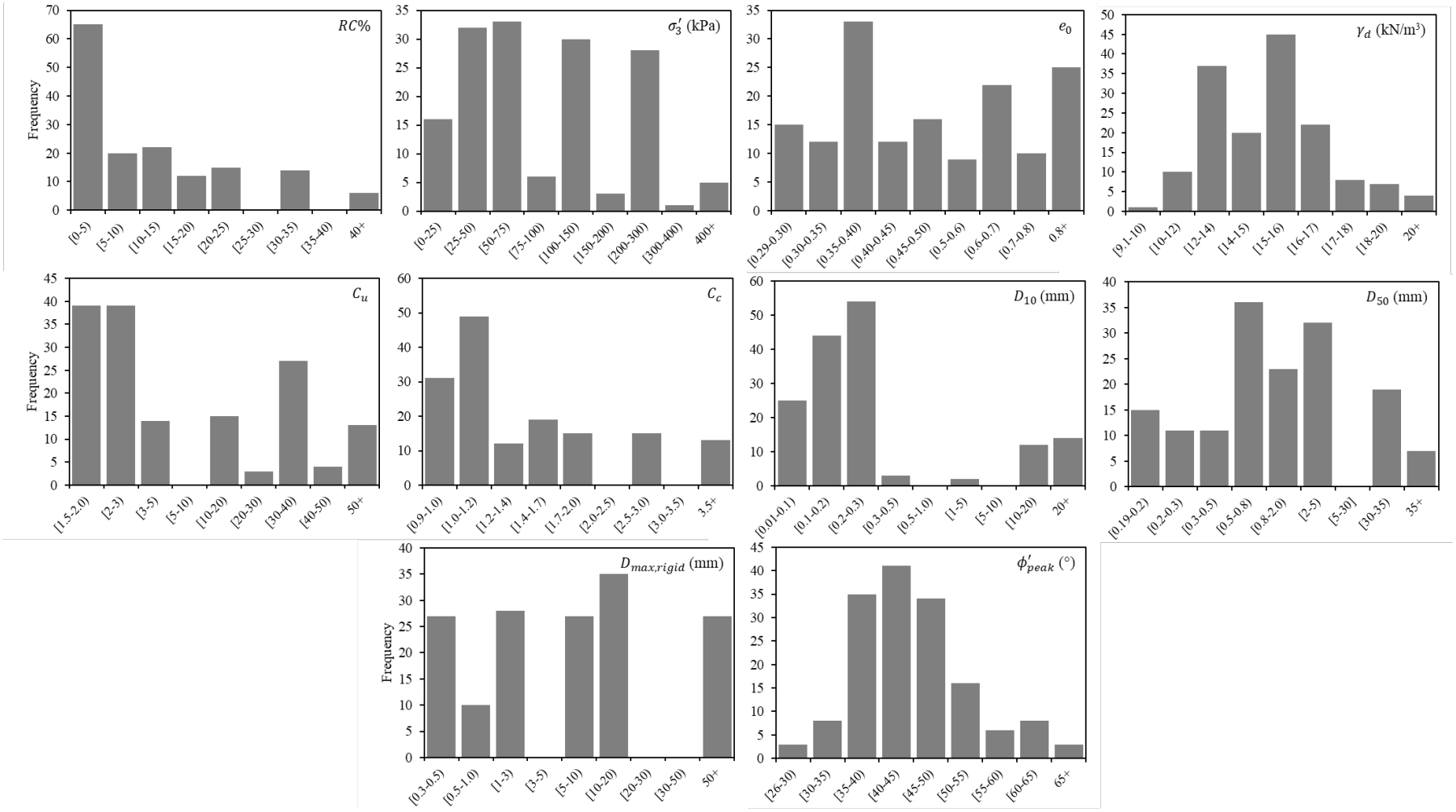
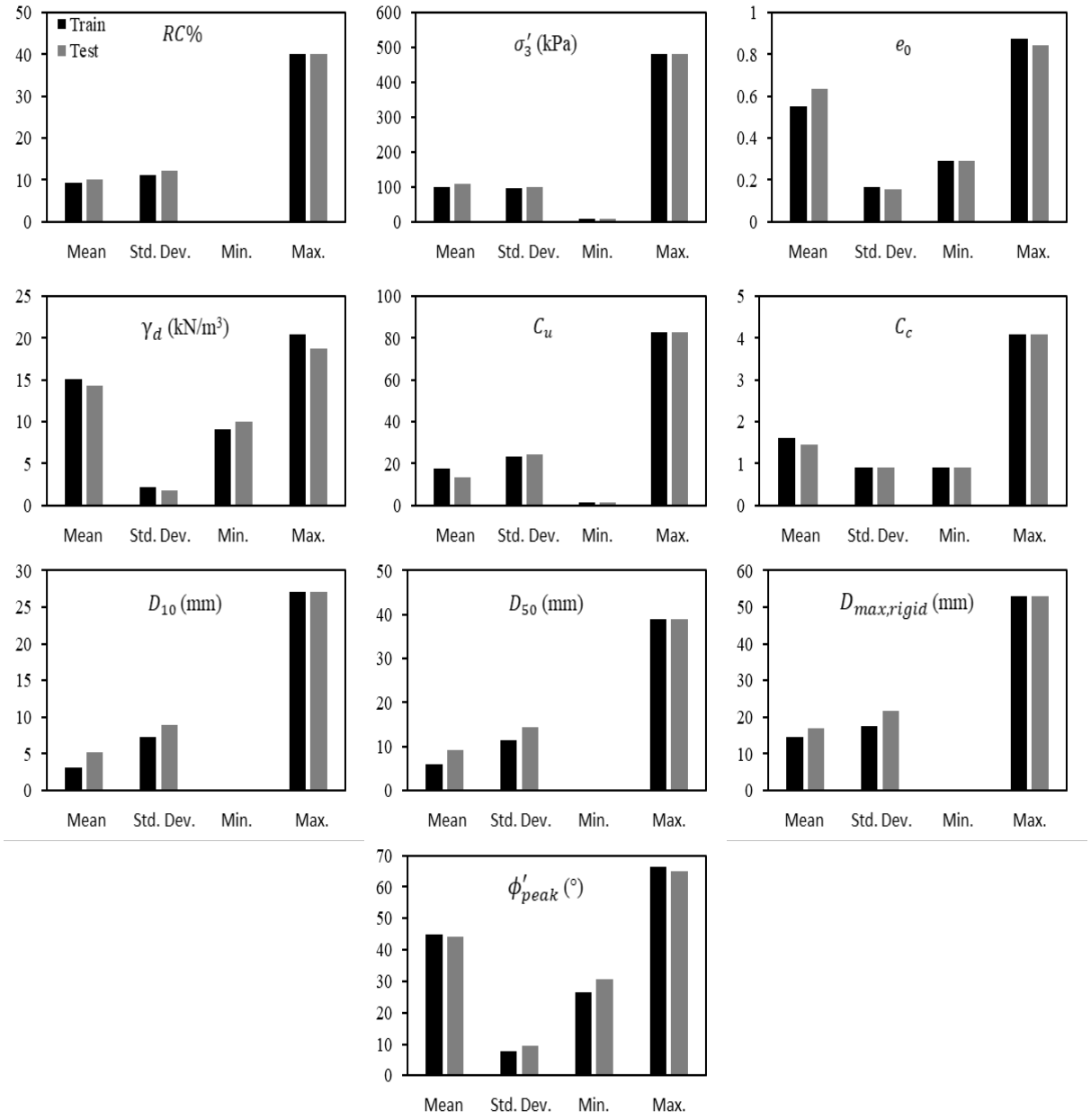


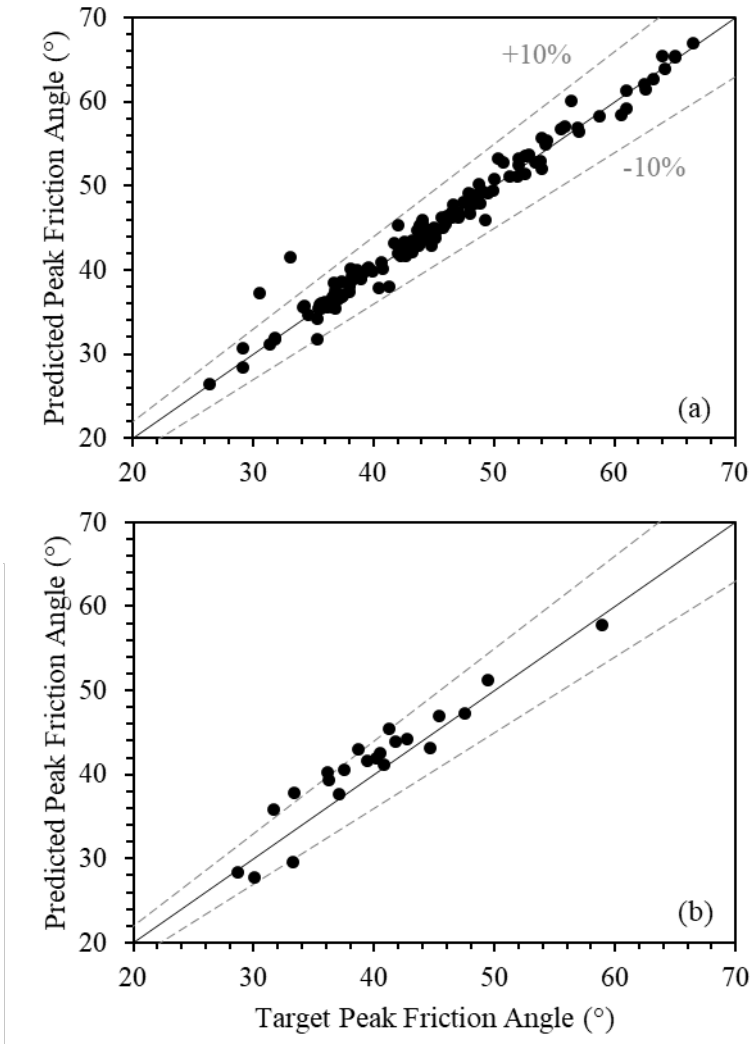
Figure 1 Frequency histograms for input and output parameters of peak friction angle model



802

803 Figure 2 Statistical properties for training and testing friction angle data subsets

804

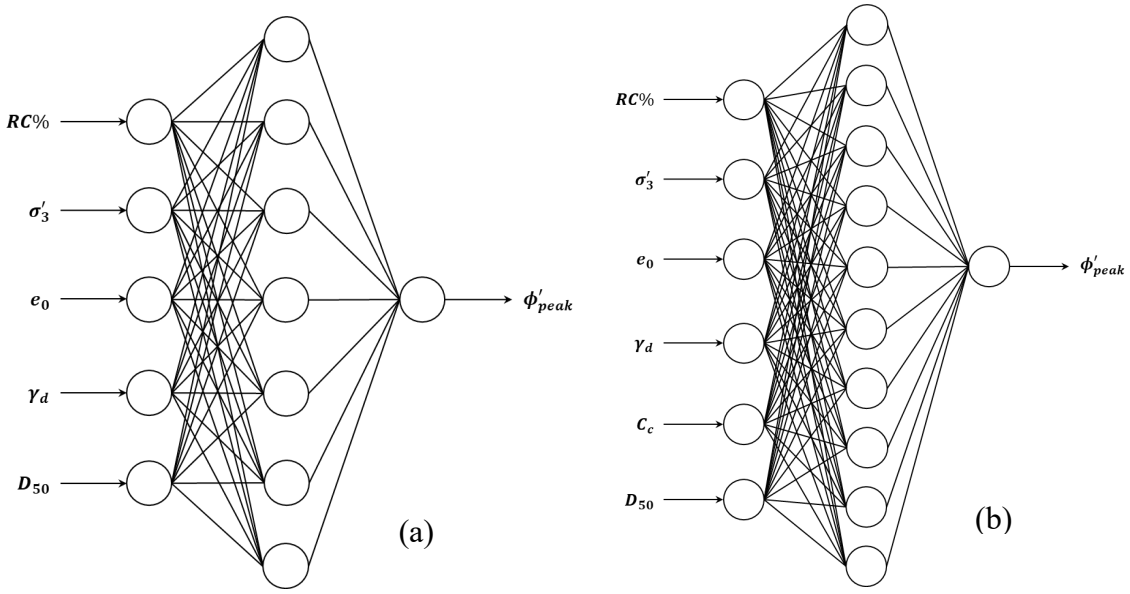


805

806 Figure 3 Predicted vs. target peak friction angle values for 9-10-1 model with (a) original

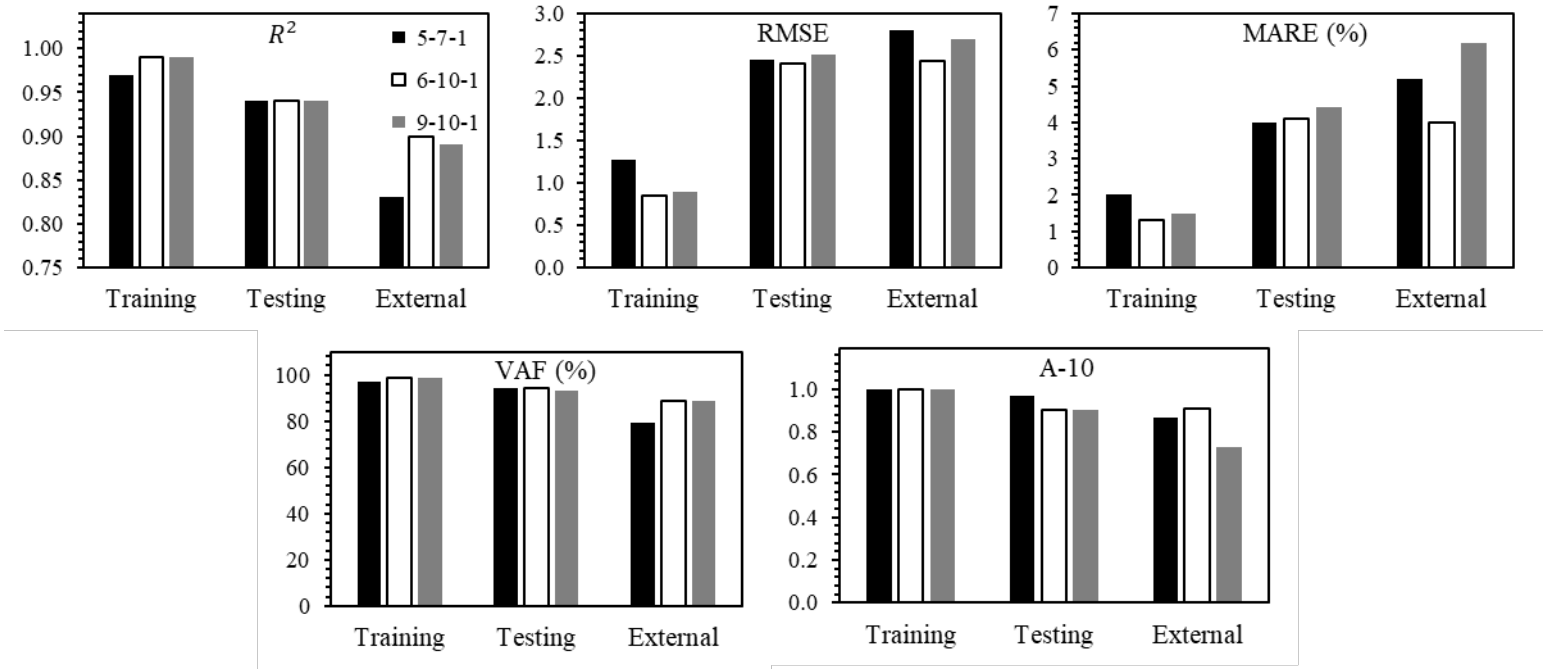
807 database and (b) external data

808



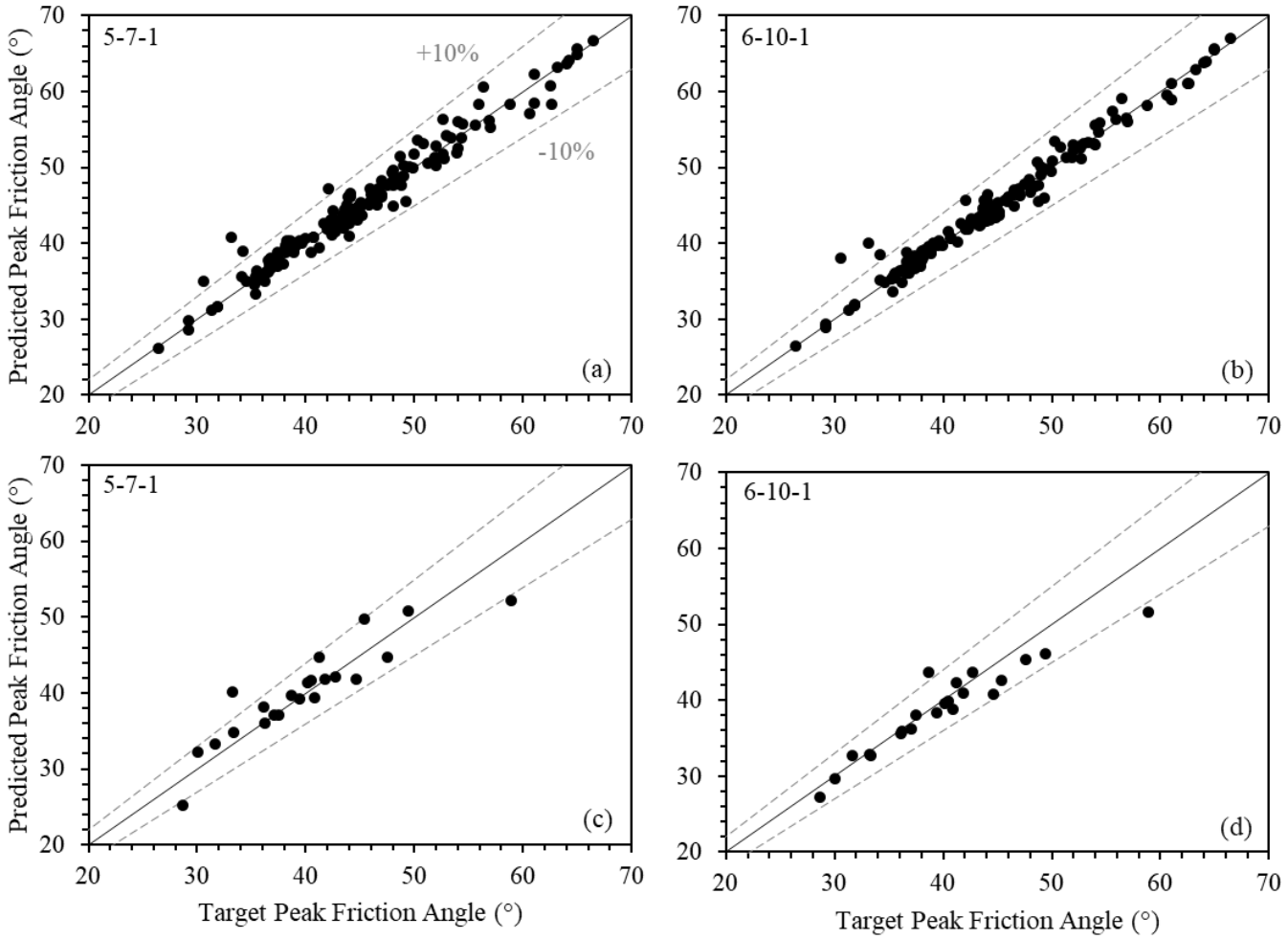
809

810 Figure 4 ANN structure diagrams for best (a) 5-input (5-7-1) and (b) 6-input (6-10-1)



811

812 Figure 5 Comparison of evaluation metrics between 5-7-1, 6-10-1, and 9-10-1 friction angle models for each data subset



813

814 Figure 6 Predicted vs. target peak friction angle values for 5-7-1 (a,c) and 6-10-1 (b,d) models with original and external data, respectively

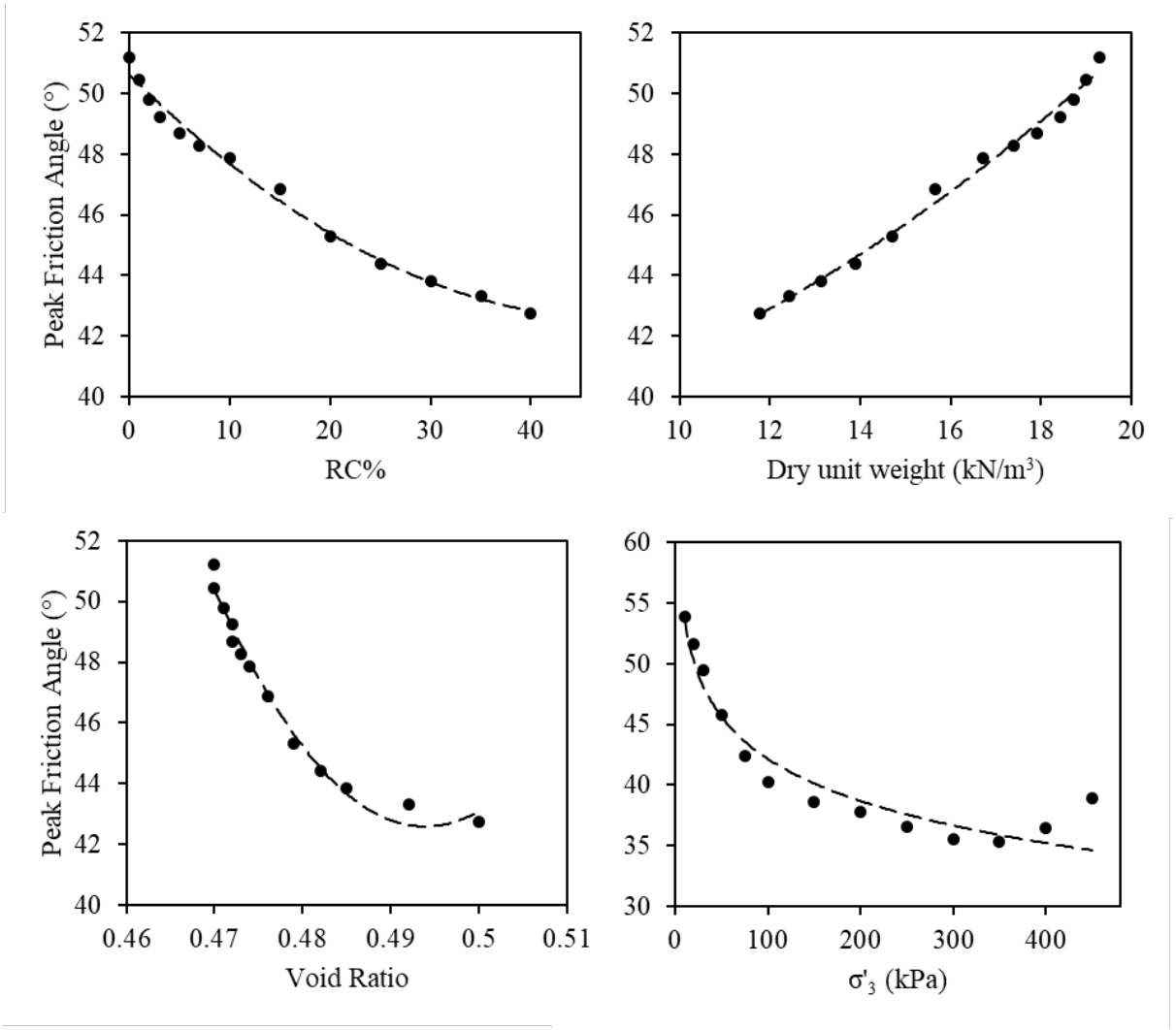
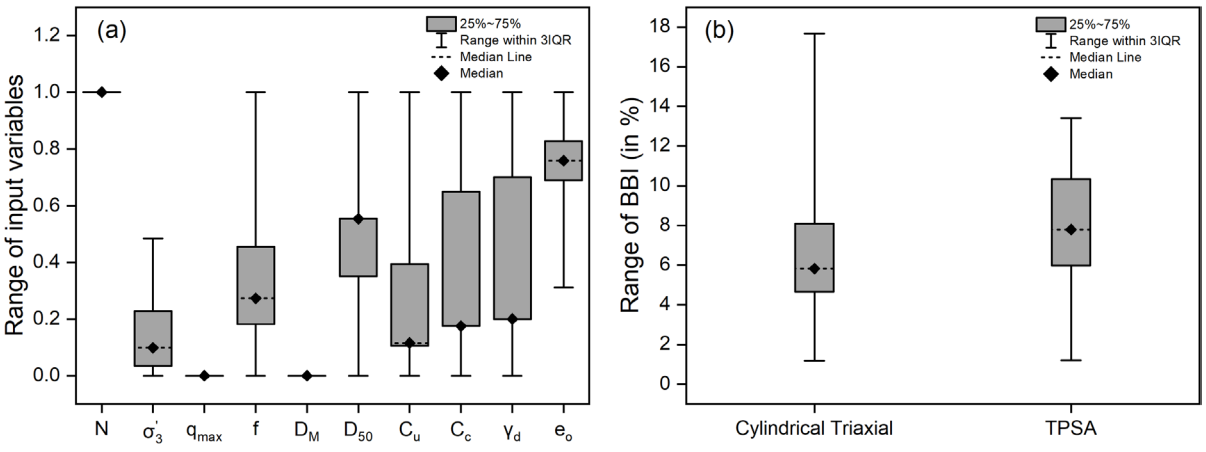


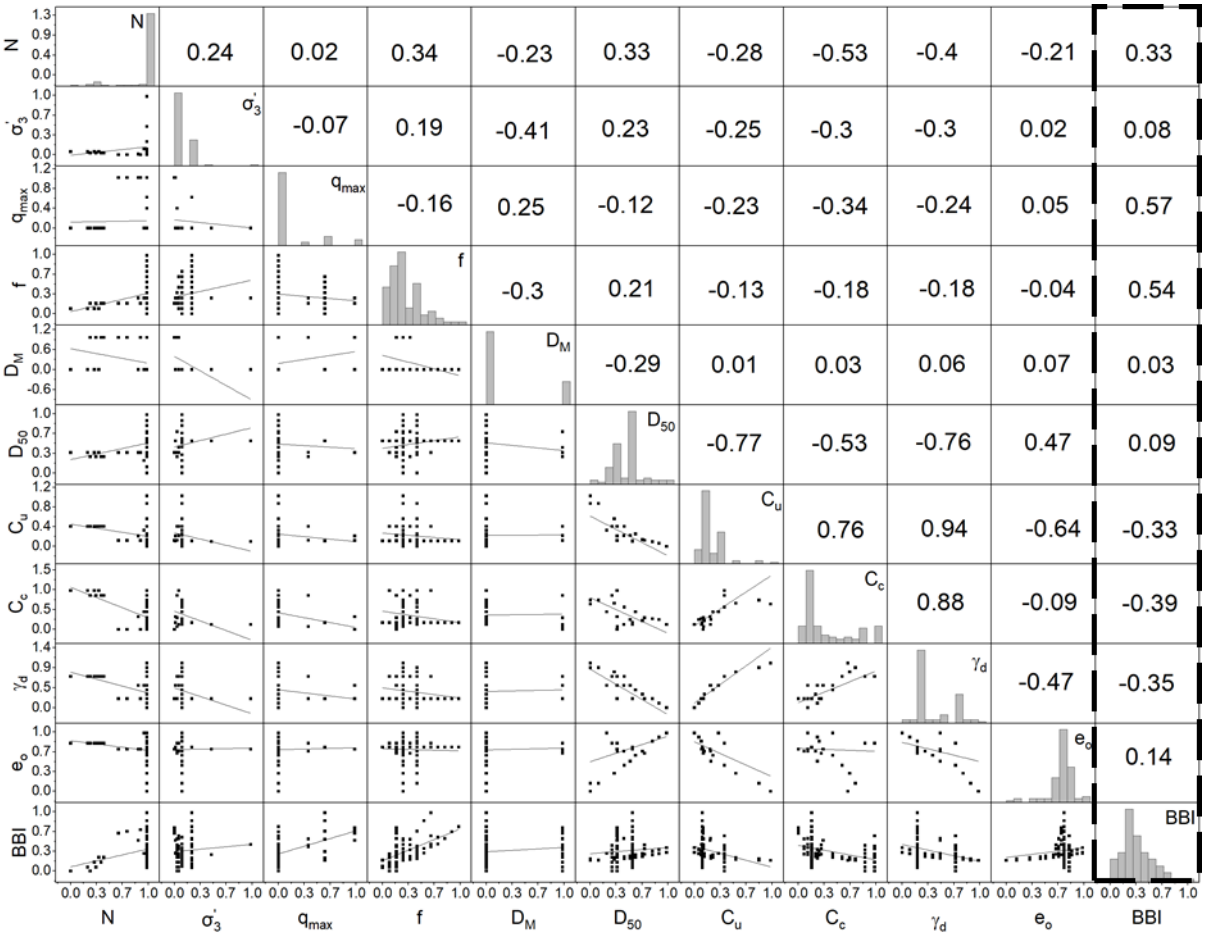
Figure 7 Sensitivity of peak friction angle with respect to varying rubber content, dry unit weight, void ratio, and effective confining pressure



820

821 Figure 8 (a) Data distribution of inputs parameters in the database (b) variation in BBI

822 magnitude from different experimental set up



823

824 Figure 9 Pairwise correlation matrix of breakage database (diagonal shows distributions of

825 variables as histograms; scatterplots are shown below the diagonal; Spearman correlation

826 coefficients are shown above the diagonal)

827

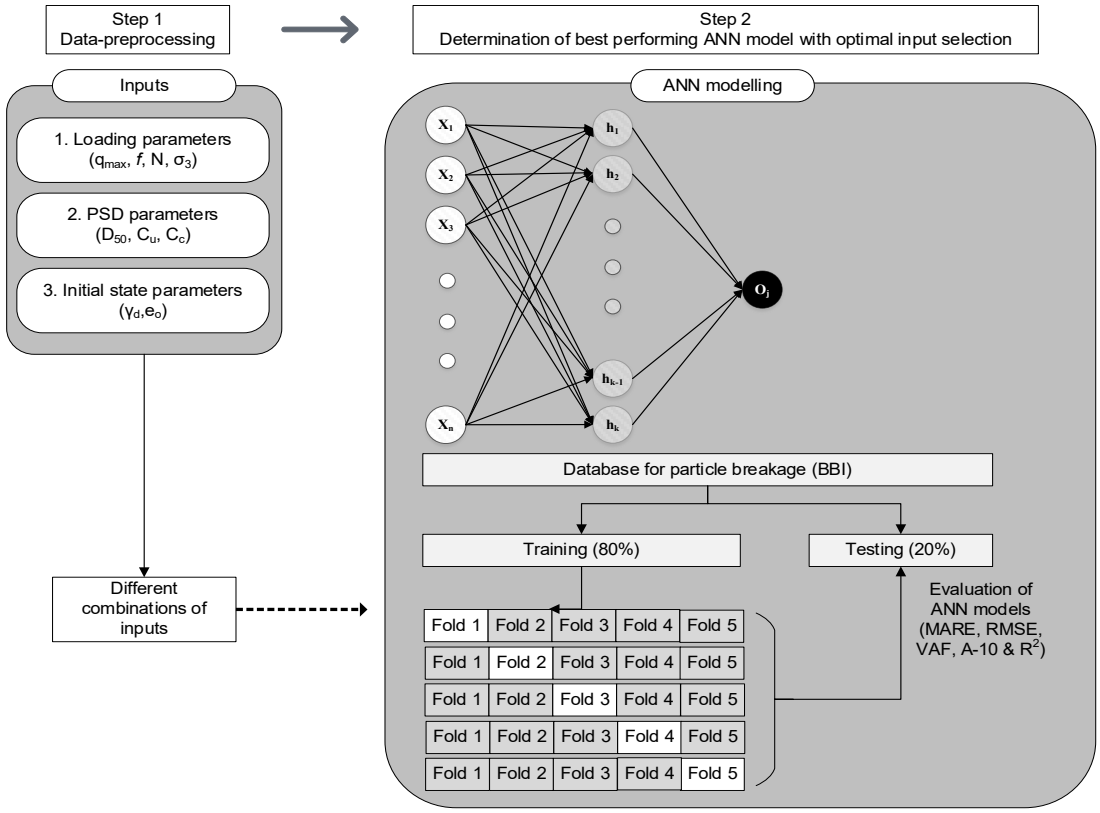


Figure 10 ANN framework for breakage model

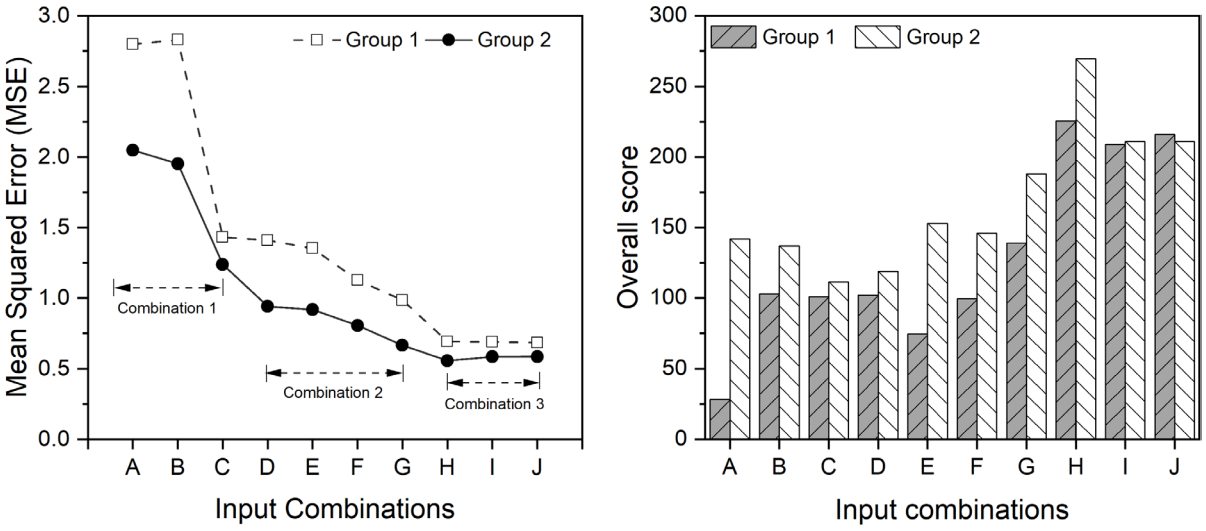
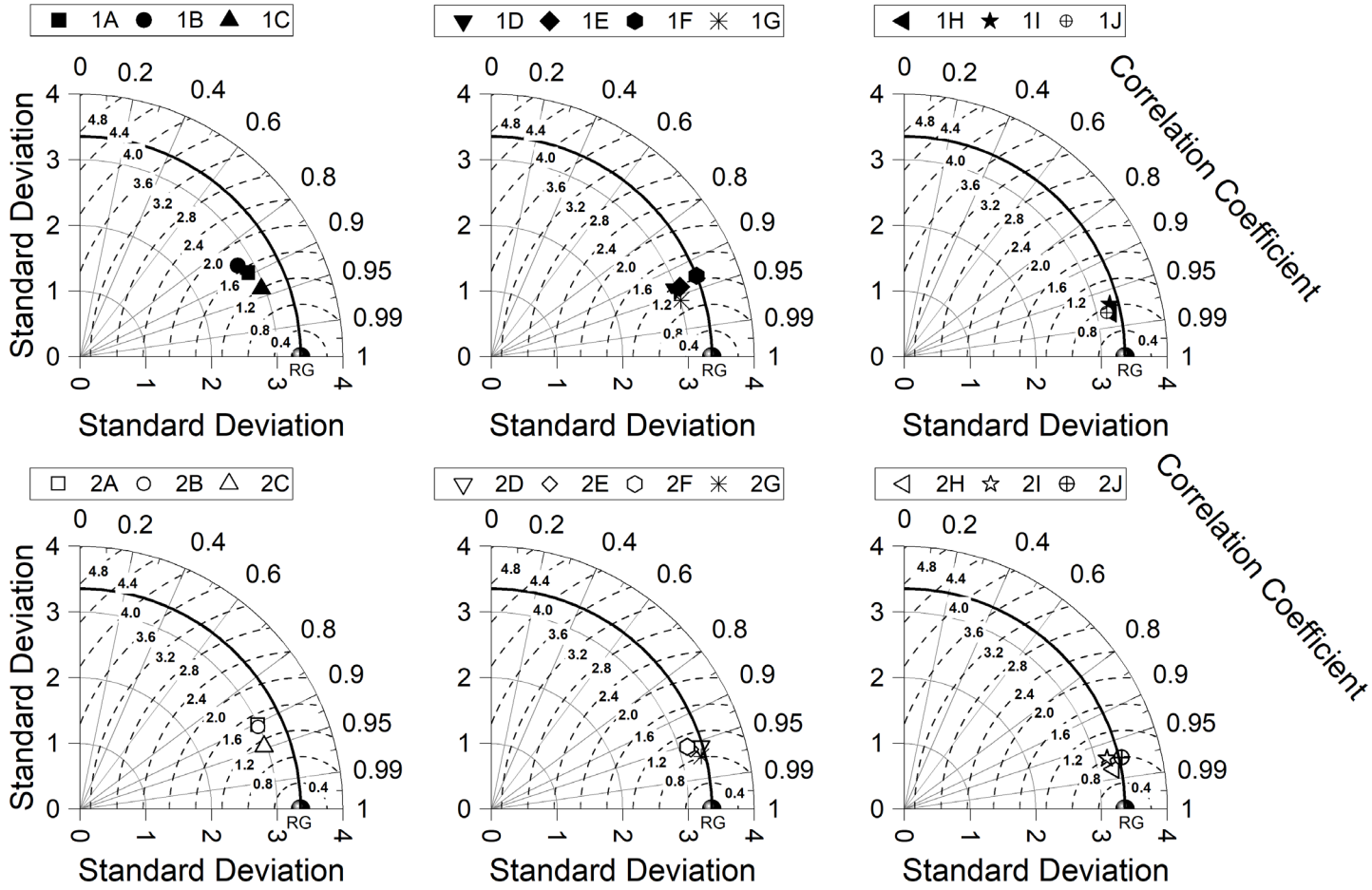


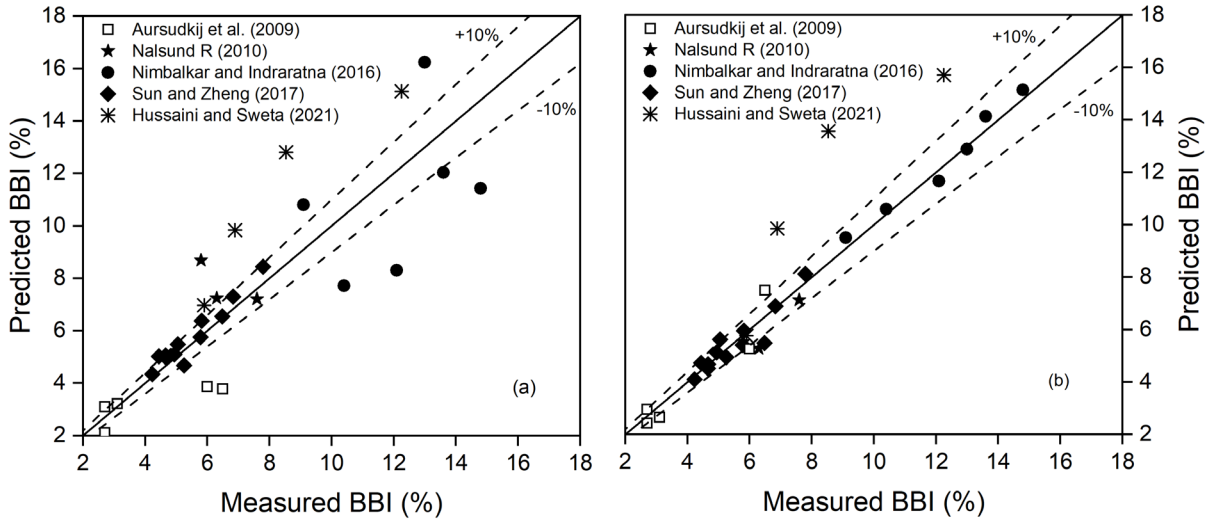
Figure 11 Breakage model comparison with different input combinations, with and without data labelling (a) mean squared error (b) overall score



835

836 Figure 12 Taylor diagram for breakage models with different input combinations of Group 1 and 2

837 (a) 1A-C (b) 1D-F (c) 1G-H (d) 2A-C (e) 2D-F (f) 2G-H



838

839 Figure 13 Predicted vs Measured BBI (a) without data label (Group 1H) (b) with data label

840 (Group 2H)

841

842