Evaluating the effect of curing conditions on the glass transition of the structural adhesive using conditional tabular generative adversarial networks

Evaluating the effect of curing conditions on the glass transition of the structural adhesive using conditional tabular generative adversarial networks

Songbo Wang [a,b,*], Haixin Yang [a,b], Tim Stratford [c], Jiayi He [a], Biao Li [a,b], Jun Su [a,b]

[a] School of Civil Engineering, Architecture and Environment, Hubei University of Technology, Wuhan 430068, China
[b] Key Laboratory of Intelligent Health Perception and Ecological Restoration of Rivers and Lakes, Ministry of Education, Hubei University of Technology, Wuhan 430068, China
[c] School of Engineering, Institute for Infrastructure and Environment, The University of Edinburgh, Edinburgh EH9 3FG, UK

Email: wangsongbo@hbut.edu.cn (Songbo Wang)

Abstract

Owing to its structural advantages, adhesively bonding fibre-reinforced polymers have been a promising solution for strengthening constructions. However, the effectiveness of this technique is significantly influenced by the material properties of the adhesive layer, which are largely determined by its curing condition. A comprehensive analysis of the effect of curing conditions on structural adhesives is hampered by the lack of sufficient experimental data. To mitigate such a limitation, this present study utilises a deep machine learning (ML) tool, the conditional tabular generative adversarial networks (CTGAN), to generate plausible synthetic dataset for developing a robust data-driven model. An artificial neural network (ANN) was trained on synthetic data and tested on real data, following the "Train on Synthetic – Test on Real" philosophy. The ultimately developed CTGAN-ANN model was validated by newly conducted experiments and several published studies ($R^2 \geq 0.95$), which demonstrated the ability to provide accurate estimates of the glass transition temperature values of the polymer adhesive. A comprehensive evaluation of the effect of each curing condition variable on the adhesive was performed, which revealed the underlying relationships, indicating that curing temperature and curing time have a positive effect, but that curing humidity has a negative

effect. The ML model developed could inform the practical use of the structural adhesive in civil engineering.

Keywords: Curing condition; Structural adhesive; Glass transition temperature; Conditional tabular generative adversarial networks; Artificial neural network

1 Introduction

Engineering adhesives have played an essential role in various industries (Heshmati et al., 2015; Higgins, 2000; Karbhari and Shulley, 1995). For civil engineering, adhesively bonding fibre-reinforced polymers (FRP) has been a mainstream method of structural strengthening due to the advantages of low additional weight, low labour and time costs, and uniform stress distribution (Cadei et al., 2004; He et al., 2021; Zhao and Zhang, 2007). The structural adhesive layer usually functions as a load transfer agent between the structure and the FRP, consequently the performance of the polymer adhesive, which is primarily determined by the curing condition, can considerably affect the effectiveness of the strengthening system (Agha and Abu-Farha, 2021; Teng et al., 2012; S. Wang et al., 2022; Wang et al., 2021).

Most previous research on the influence of curing conditions on structural adhesives has been constrained to consider only a few cases due to insufficient experimental data (I. M. Foletti et al., 2020; Jahani et al., 2022; Savvilotidou et al., 2017; Sousa et al., 2019), which inevitably results in conclusions that suffer from a lack of representativeness. Thus, a more profound investigation using the machine learning (ML) approach is conducted in the current paper to overcome this challenge.

1.1 Structural adhesives in civil engineering

For economic and practical reasons, the application of the FRP-bonded strengthening technique in constructions often involves the use of an ambient-cured (cold-curing) epoxy adhesive with a glass transition temperature (T_g) of around 50 °C. When solar heating causes temperatures to approach the T_g, it will induce a transition in the polymer adhesive from a rigid, glassy state to a soft, rubbery state, which can lead to a significant loss of strength and stiffness of the applied adhesive, ultimately impacting the performance of the FRP-bonded strengthening system (Ascione et al., 2022, 2021; Sahin and Dawood, 2016; Stratford and Bisby, 2012; S Wang et al., 2022; Wang et al., 2021).

Existing studies have demonstrated that the T_g as well as the mechanical properties of the structural adhesive is decisively affected by the applied curing condition. For instance, Carbas et al. (Carbas et al., 2014) studied six curing temperature conditions ranging from 23 °C to 120 °C , with the results showing that the T_g, strength, and stiffness of the structural adhesive increased with increasing curing temperature before reaching the fully cured status. Jahani et al. (Jahani et al., 2022) conducted experiments to demonstrate that both curing and post-curing temperatures can have a significant effect on the T_g values and mechanical properties of the structural adhesive. They found that a higher curing temperature (50 °C) would be beneficial, but an excessively high temperature (70 °C) may have a detrimental effect. Foletti et al. (I. M. Foletti et al., 2020) studied the impact of 4 curing conditions on the structural adhesive and demonstrated that curing at the ambient temperature accompanied by degassing can improve the performance of the material, while curing at a high temperature (90 °C) can reduce its long-term performance.

However, studies based on conventional experimental work are unlikely to be able to examine all possible scenarios comprehensively or to systematically summarise the patterns of influence of various curing condition variables. Whereas innovative machine learning methods

offer the potential to address this issue, as this approach can extract complex relationships between the inputs (the applied curing conditions) and the outputs (the T_g results of the adhesive) and summarise the implied non-linear laws for building a unified model that can be utilised to produce new predictions without involving labour-intensive work.

1.2 Machine learning approach

ML was introduced in 1959 as a branch of artificial intelligence (AI) that enables computers to learn intelligently without being explicitly programmed (Salehi and Burgueño, 2018; Samuel, 1959; Yin et al., 2020). Since the early 21st century, engineers have been using ML techniques to explore the complex behaviour of structural materials (Naser, 2019; Yin et al., 2020). With the rapid development of this technology in recent years, mature ML techniques encompass models such as artificial neural network (ANN), support vector machine and decision tree have become widely accepted and effectively used in academic research (Erdal, 2013; Huang et al., 2021; Su et al., 2021; Wang et al., 2023b).

However, in many cases, the limited available experimental data is one of the major obstacles to applying data-driven ML models to structural materials issues, since engineering experimental work is often time-consuming, costly, and labour-intensive. Therefore, data augmentation approaches are becoming increasingly attractive as the solution for developing ML models based on small datasets (Peres et al., 2021; Zhang and Ling, 2018). Among them, generative adversarial networks (GAN) represent the cutting-edge technology and have been widely used to generate synthetic images from real images (Goodfellow et al., 2014). Conditional tabular generative adversarial networks (CTGAN), on the other hand, have been developed as an extension of GAN, possessing the ability to generate tabular datasets suitable for engineering research (Xu et al., 2019). Almustafa and Nehdi in 2022 (Almustafa and Nehdi, 2022), for example, have used CTGAN to generate 300 synthetic data relating to steel fibre

reinforced concrete (SFRC) beams to systematically investigate the structural response of SFRC beams when subjected to far-field blast loads.

There has been limited research to date using ML methods to comprehensively explore the effect of curing conditions on structural adhesives. The authors (Wang et al., 2023b) previously developed an ANN model based on a dataset containing 157 experimental results (80% for training) to investigate the relationship between the modulus response of the adhesive and the curing conditions; however, the generalisability of the established model was somewhat limited to the small dataset used, which was considered to be a kind of over-fitting.

1.3 Research significance

The aims of this research are to investigate the impact of curing conditions on the glass transition of the structural epoxy adhesive using a state-of-the-art machine learning technique, and to develop a predictive model to map the relationship between the glass transition temperatures of the adhesive and the applied curing conditions. This is achieved by employing a deep learning data generation method called conditional tabular generative adversarial networks, which overcomes the obstacle of insufficient experimental data and allows the further construction of a robust data-driven ML model.

The constructed model can provide reference values for practical civil engineering applications, while the demonstrated methodology can be adopted to solve similar problems in structural materials where experimental data is too limited to enable effective analysis and modelling.

2 Background knowledge

The CTGAN was employed to produce ample synthetic data for the development of a reliable data-driven ML model that integrated the ANN technique. The essential background

knowledge regarding the applied ML methods, model performance evaluation, and data scaling is outlined below, while a more comprehensive explanation can be sourced in the cited scientific papers.

2.1 Conditional tabular generative adversarial networks

GAN are based on the zero-sum game strategy and enable the learning of data distributions via an adversarial learning process. A GAN consists of 2 models which are generator (*G*) and discriminator (*D*). The purpose of a *G* model is to learn the distribution pattern of the data and produce accurate fake samples to "fool" the discriminator, while, in contrast, a *D* model aims to distinguish between real and fake samples without being deceived (He and Zhou, 2022; Moon et al., 2020; Zhao et al., 2023). These conflicting objectives leads to a competitive training process (Goodfellow et al., 2014; Moon et al., 2020):

$$\mathrm{V(D,G)} = E_{x\sim p_r}[\log D(x)] + E_{z\sim p_z(z)}\left[\log\left(1 - D\left(G(z)\right)\right)\right] \quad (1)$$

where x is the real sample following the real dataset distribution p_r, z is the noise variable following the noise distribution p_z, $G(z)$ represents the generated sample, and D stands for the discriminator function with an output of 1 if the given data is (100%) real and 0 if the given data is (100%) fake (Goodfellow et al., 2014; Moon et al., 2020). Whilst the generator and discriminator struggle to fully accomplish their objectives during the adversarial training process, at the end of the training, the generator can be upgraded to have the ability to generate realistic-looking data, which is usually the purpose of applying a GAN (He and Zhou, 2022; Moon et al., 2020; Zhao et al., 2023).

CTGAN was proposed by Xu et al. (Xu et al., 2019) in 2019 as an innovative extension of GAN, which compensates for the difficulty of applying the original GAN method to tabular

data and enables the generation of both continuous and discrete data through two techniques (Habibi et al., 2023; Moon et al., 2020).

The first is the *mode-specific normalisation*, designed to deal with columns (of datasets) with complex distributions. For each continuous column, CTGAN utilised a variational Gaussian mixture model to estimate the number of modes and fit a Gaussian mixture. The value in the column is then normalised according to the corresponding mode. When generating synthetic data after the training process, CTGAN will convert the generated data back to the initial scale. The second is the *conditional generator and training-by-sampling*, designed to deal with the imbalanced discrete data columns (Almustafa and Nehdi, 2022; Habibi et al., 2023; Moon et al., 2020; Xu et al., 2019). Since the datasets used in this study contain only continuous columns (variables), this technique is not explained for brevity (can be found in Xu et al. (Xu et al., 2019)).

The fundamental architecture of CTGAN is illustrated in Figure 1, and the corresponding loss function can be expressed as (Moon et al., 2020; Xu et al., 2019):

$$\mathrm{L} = E_{G(z)\sim p_g}\left[D\left(G(z)\right)\right] - E_{x\sim p_r}[D(x)] + \lambda E_{\hat{x}\sim p_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\| - 1)^2] \quad (2)$$

where p_g is the generated dataset distribution, λ refers to the gradient penalty coefficient, $\hat{x} \sim p_{\hat{x}}$ refers to the random samples. The noise variable ($\mathrm{z} \sim \mathrm{P_z(z)}$) introduces randomness independent of the raw data. It serves as a seed for the generator, which learns how to generate data resembling the real distribution from these random inputs, refining its output through training (Moon et al., 2020; Xu et al., 2019).

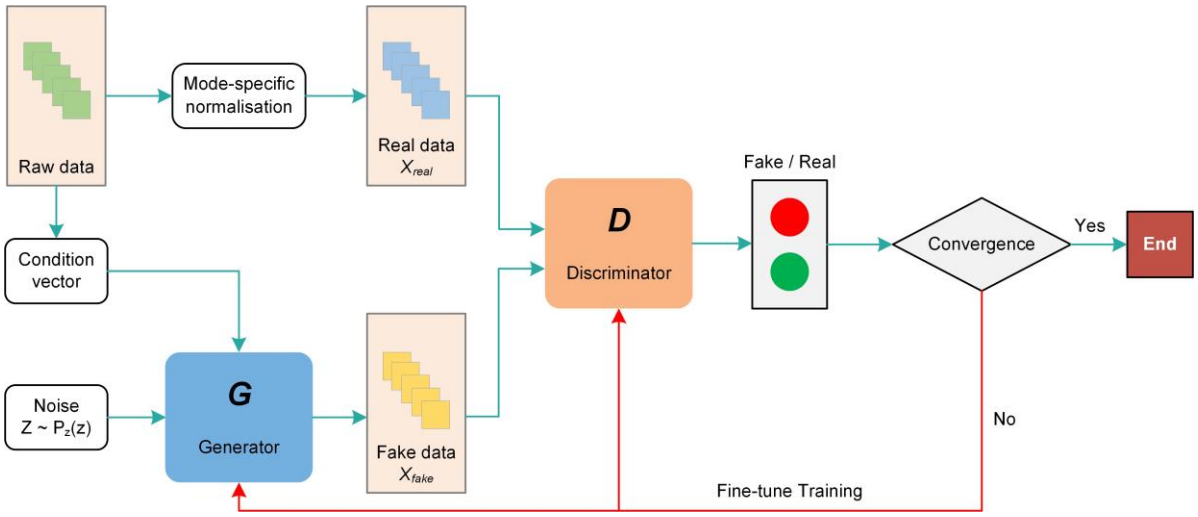


Figure 1: Architecture of conditional tabular generative adversarial networks (CTGAN)

2.2 Artificial neural network

ANN is a traditional machine learning technique inspired by the structure and function of biological neural networks in the brain. As illustrated in Figure 2, the ANN system is composed of numerous interconnected neurons that process and transmit signals between various layers. This process can be described as:

$$y_i = \sigma\left(\sum_{j=1}^{n} W_{ij}^{k} I_j + b_i\right) \tag{3}$$

where y_i represents the output of neuron i, I_i represents its input, σ is the activation function, W_i^k is the weight, and b_i stands for the bias (Hisham et al., 2021; Wang et al., 2023b; Yan et al., 2020).

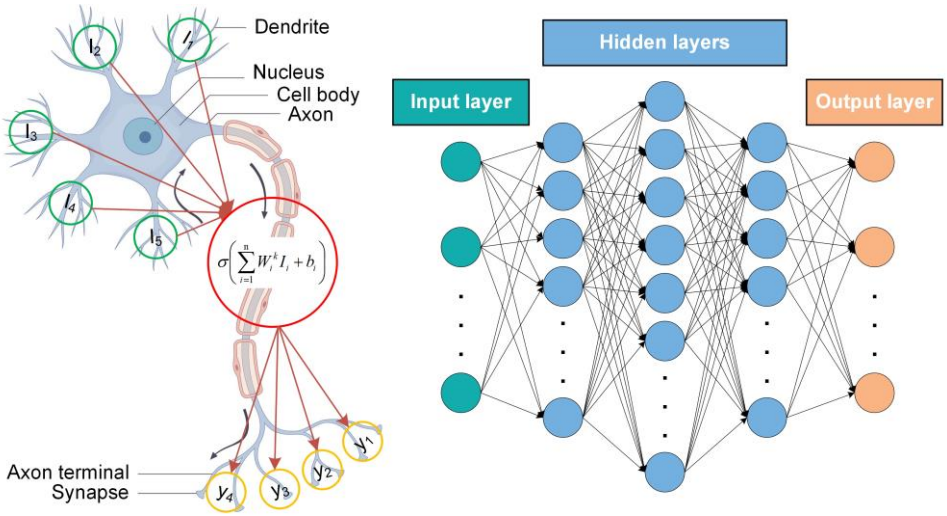


Figure 2: Schematic illustration of artificial neural networks (ANN)

The ANN can consist of a single input layer, a single output layer and single or multiple hidden layers, where as an ANN with multiple hidden layers is also referred to as a deep neural network (DNN) (Marani et al., 2023; Marani and Nehdi, 2022).

In this study, the ML procedures were executed using Python 3.9.The Keras library (Chollet and others, 2015) was used to build the multi-layer ANN, and the CTGAN library (Xu et al., 2019) was employed to develop the CTGAN-based model.

2.3 Performance measurement

A proper evaluation of the model performance is critical in ensuring the accuracy and robustness of the developed ML model. The well-known statistical performance metrics, including the coefficient of determination (R^2), mean absolute error (*MAE*), and mean squared error (*MSE*) were used in this study, as presented below:

$$R^2 = 1 - \frac{\sum_{i=1}^{m} \left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{m} (Y_i - \bar{Y})^2} \quad (4)$$

$$MAE = \frac{1}{m} \sum_{i=1}^{m} \left|Y_i - \hat{Y}_i\right| \quad (5)$$

$$MSE = \frac{1}{m}\sum_{i=1}^{m}\left(Y_i - \hat{Y}_i\right)^2 \quad (6)$$

where Y_i is the real target value, $\hat{Y}_i$ is the predicted value, and $\bar{Y}$ is the mean of $\sum Y_i$. The R^2 results measure the relationship between the Y_i and $\hat{Y}_i$ values, with closer to 1 indicating better model performance. Whereas the lower *MAE* and *MSE* values signify the lower differences between the Y_i and $\hat{Y}_i$ values, with *MSE* being more sensitive to outliers (Almustafa and Nehdi, 2022; Ben Chaabene and Nehdi, 2021; Zhao et al., 2023). The *MSE* served as the objective function during the training process of the following ML models.

2.4 Data scaling

Data scaling is a crucial step in the machine learning process, as it can enhance prediction accuracy, mitigate overfitting, and expedite training speed. As mentioned in Section 2.1, the raw data will be automatically scaled (normalised) during the CTGAN training process; however, for the ANN training, it is necessary to perform data scaling manually (or add the appropriate code) in the pre-processing phase. The Min-Max scaling method is therefore adopted in the current study:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (7)$$

where X is the raw data, and the minimum and maximum values of the variable represented as X_{min} and X_{max}, respectively (Gu et al., 2021; Hisham et al., 2021).

3 Establishment of the CTGAN-ANN model

This section presents the various stages involved in developing the ML model to enable the assessment of the impact of curing conditions on the structural adhesive. The CTGAN

approach was first utilised for data generation to generate sufficient synthetic data to train the ML model and, subsequently, an ANN-based predictive model was established by using TSTR.

The use of CTGAN to generate data and ANN for predictive modelling was strategically selected when evaluating the effect of curing conditions on structural adhesives. This is because CTGAN is capable of generating statistically significant synthetic data, which is essential for representing a wide range of curing conditions. At the same time, ANN is adept at modelling the non-linear relationships inherent in the curing process and is able to handle high-dimensional data and automatically learn the interactions between features without the need for manual design. The use of TSTR ensures that the generality of the model is assessed and is less prone to overfitting. These approaches together provide a robust framework for developing accurate, generalisable ML models for the complexity of adhesive curing conditions. The corresponding data generation, hyperparameter tuning, and model development processes are described. A general flowchart depicting the establishment is presented in Figure 3.

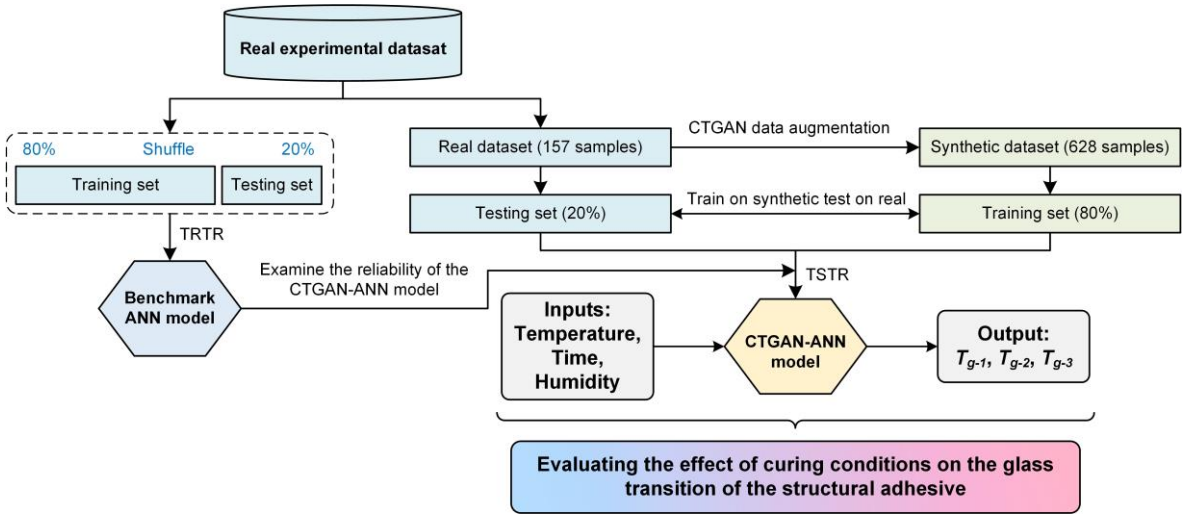


Figure 3: Workflow of the establishment of the CTGAN-ANN model

3.1 Real experimental dataset

The complete CTGAN-ANN model development framework is shown in Figure 4, where having an initial real experimental dataset is the first step in the machine learning process.

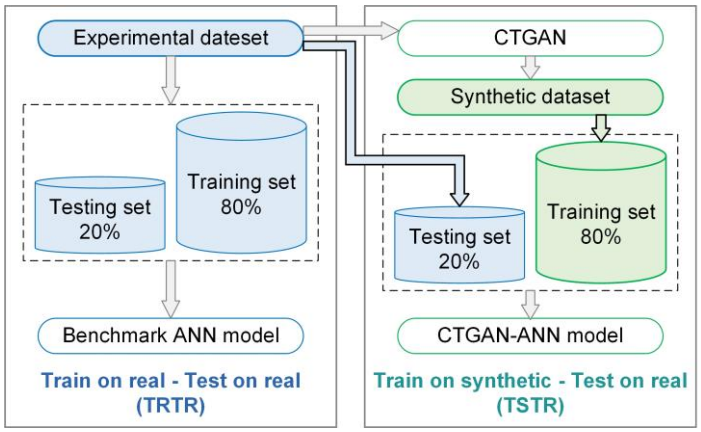


Figure 4: CTGAN-ANN model development framework

The authors (Wang et al., 2023b) recently compiled a dataset on the structural adhesive, Sikadur-330 (SIKA, 2017), consisting of 157 experimental results gathered through a comprehensive literature review. It was utilised to develop a preliminary single-hidden-layer ANN model to study the adhesive's modulus response at elevated temperatures (Wang et al., 2023b). The current study employed this dataset to (a) develop a benchmark ANN (DNN) model with multiple hidden layers; (b) serve as a reference for generating the synthetic dataset using CTGAN; and (c) serve as a testing set for establishing the CTGAN-ANN model to further investigate the effects of curing conditions on the glass transition of the adhesive.

Table 1 shows the features of the dataset considered for the development of the corresponding models. The curing condition variables include curing temperature, curing time, and curing humidity. Whilst the glass transition of a polymer adhesive occurs over a range of temperatures and is accompanied by a continuous reduction in modulus, which affects the performance of the bonded joint, it is usually quoted as a single T_g value. However, depending on the purpose of the engineering application, T_g has different definitions (British Standards Institutions, 2019; Menard and Menard, 2020):

- Onset T_g ($T_{g\text{-}1}$): the temperature at which the material begins its transition from the glassy to the rubbery state. It is the lowest T_g result and is often used as a reference temperature in engineering because it provides a relatively conservative value;

- Inflection point T_g ($T_{g\text{-}2}$): the temperature at which the material reaches the midpoint of the glass transition. It is normally higher than onset T_g and represents a more advanced stage of the glass transition;
- Peak Tan δ T_g ($T_{g\text{-}3}$): the temperature at which the material shows a distinct rubbery response and is regularly used as a reference temperature in material science as it is easily identifiable and provides a well-defined T_g value.

Table 1: Statistical summary of the real dataset

Variables	Unit	Min	25%	50%	75%	Max	Mean	Standard deviation	Coefficient of variation
Inputs									
Curing temperature	°C	13	24	35	65	80	43.6	22.8	52.1%
Curing time	days	3	6	7	28	34	13.2	9.7	73.5%
Curing humidity	%	0	0	40	100	100	49.5	48.1	97.0%
Outputs									
Onset T_g ($T_{g\text{-}1}$)	°C	28.5	44.3	50.2	60.7	74.9	52.3	11.9	22.6%
Inflection point T_g ($T_{g\text{-}2}$)	°C	34.9	49.7	58.0	66.4	80.3	58.8	12.5	21.1%
Peak Tan δ T_g ($T_{g\text{-}3}$)	°C	46.5	57.3	66.3	73.0	89.1	67.3	11.8	17.5%

Note that the benchmark ANN model used for the comparison purpose was trained on 80% of the real dataset and tested with the remaining 20%, an approach known as "Train on Real – Test on Real (TRTR)" (Esteban et al., 2017; Marani and Nehdi, 2022). In contrast, the CTGAN-ANN model was trained on the synthetic dataset (628 results) and tested with the entire real dataset (157 results), an approach known as "Train on Synthetic – Test on Real (TSTR)" (Esteban et al., 2017; Marani and Nehdi, 2022).

The aim of developing the benchmark ANN model is to examine whether the data generated by CTGAN is reliable and whether the CTGAN-ANN model can achieve satisfactory accuracy. Table 2 lists the hyperparameters of the benchmark ANN model after tuning using the grid search strategy, although variations in the hyperparameters had limited impact on the performance of the ANN model. Three hidden layers were ultimately chosen, with 32 neurons in the first and third layers and 64 neurons in the second layer. The best training performance

was achieved at the 495th epoch with the minimum *MSE* [Eq. (6)] equals to 9.92×10^{-4}. The source code of the benchmark ANN model is available in the supplementary material provided.

Table 2: Hyperparameters of the benchmark ANN model

Hyperparameters	Grid search ranges	Ultimate selection
Number of Hidden layers and neurons in each layer	[(64, 64), (128, 128), (16, 32, 16), (32, 64, 32), (64, 128, 64)]	(32, 64, 32)
Activation function in the output layer	[Linear, Sigmoid]	Linear
Activation function in other layers	[Sigmoid, ReLU, Tanh]	ReLU
Loss function	[*MSE*, *MAE*]	*MSE*
Optimizer	[Adam, SGD]	Adam
Epochs	[100, 300, 500]	500
Batch size	[16, 32, 64]	32

3.2 Generating synthetic dataset using CTGAN

The quality of the synthetic dataset generated by the generator (*G*) in a CTGAN (see Figure 1) is highly influenced by the hyperparameters used during the adversarial training process (Marani et al., 2023; Marani and Nehdi, 2022). Thus, a tuning process was conducted to determine the best combination of the CTGAN hyperparameters. Figure 5 depicts the tuning process, which involves testing various combinations of the hyperparameters selected from the grid search range and evaluating the corresponding generated synthetic dataset using a baseline ANN model with the same parameters as the benchmark ANN model developed in Section 3.1.

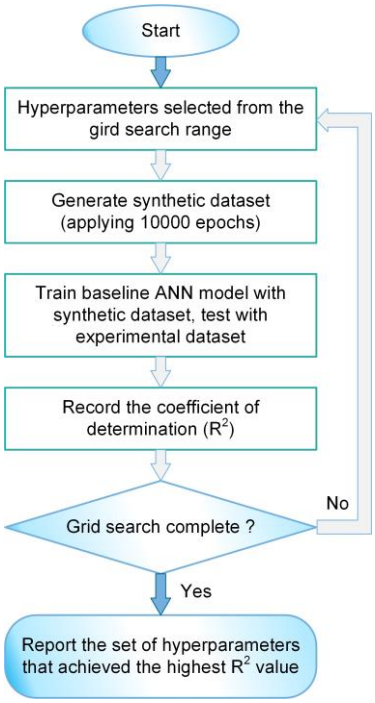


Figure 5: Hyperparameter tuning process for CTGAN

The dataset generated by the CTGAN was used to train a baseline ANN model. The quality of the synthetic dataset was then evaluated by comparing the coefficient of determination (R^2) [Eq. (4)] between the predicted outcomes provide by the baseline ANN model and the real experimental results. The hyperparameters yielding the highest R^2 values were identified as optimal, as they would guarantee the generation of a dataset capable of training a precise ANN-based predictive model. The final tuned hyperparameters are shown in Table 3, and these were next used to train the final CTGAN-ANN model for subsequent analyses.

Table 3: Hyperparameters of the CTGAN

Hyperparameters	Grid search ranges	Ultimate selection
Size of the random sample passed to the *G*	[64, 128]	64
Size of the output samples for each one of the *G* layers	[(128, 128), (256, 256), (512, 512)]	(256, 256)
Size of the output samples for each one of the *D* layers	[(128, 128), (256, 256), (512, 512)]	(512, 512)
Learning rate for the *G*	[2×10^{-5}, 2×10^{-4}, 2×10^{-3}]	2×10^{-5}
G weight decay for the Adam optimizer	[1×10^{-6}, 1×10^{-5}]	1×10^{-5}
Learning rate for the *D*	[2×10^{-5}, 2×10^{-4}, 2×10^{-3}]	2×10^{-4}
D weight decay for the Adam optimizer	[1×10^{-6}, 1×10^{-5}]	1×10^{-5}
Batch size	[100, 500]	500

3.3 CTGAN-ANN model

The CTGAN-ANN model consists of two key components: the data generation component and the predictive model component (refer to Figure 4).

In the data generation component, the CTGAN approach was utilised with the hyperparameters listed in Table 3. Following 10^6 epochs of adversarial training, the resulting generator was able to accurately capture the distribution patterns of the original real dataset and produce synthetic data that were virtually indiscernible. The generated dataset is listed in the supplementary material.

In the predictive model component, the TSTR method was utilized to train an ANN-based predictive model. The model was trained on 628 synthetic data (80%) and tested on 157 real experimental data (20%). The source code has been made public in the supplementary material.

The proposed CTGAN-ANN model can therefore:

- Generate additional credible synthetic dataset using the trained generator (G) in the data generation component, without quantitative restrictions;
- Provide estimates of the T_g values of the structural adhesive according to the planned curing condition, using the predictive model component;
- Contribute to revealing the underlying intricacies and evaluating the extent of the impact of each curing variable, owing to the convenience of the ML model in obtaining a large number of informative estimates (discussed further in Section 5).

4 Results and discussions

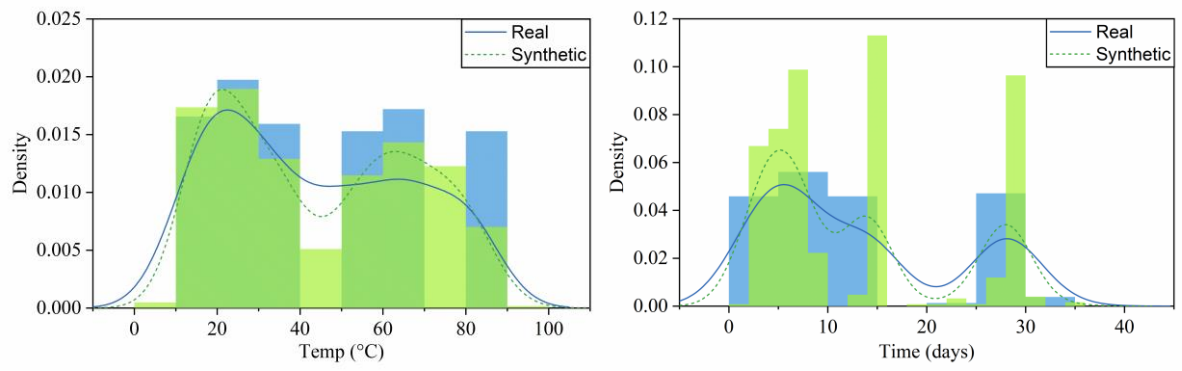
In this section, the reliability of the generated synthetic dataset and the performance of the CTGAN-ANN model are explored using statistical measures and visual representations. Moreover, the robustness of the model is verified by comparing the prediction results with newly conducted experiments and some published studies.

4.1 Generated synthetic dataset

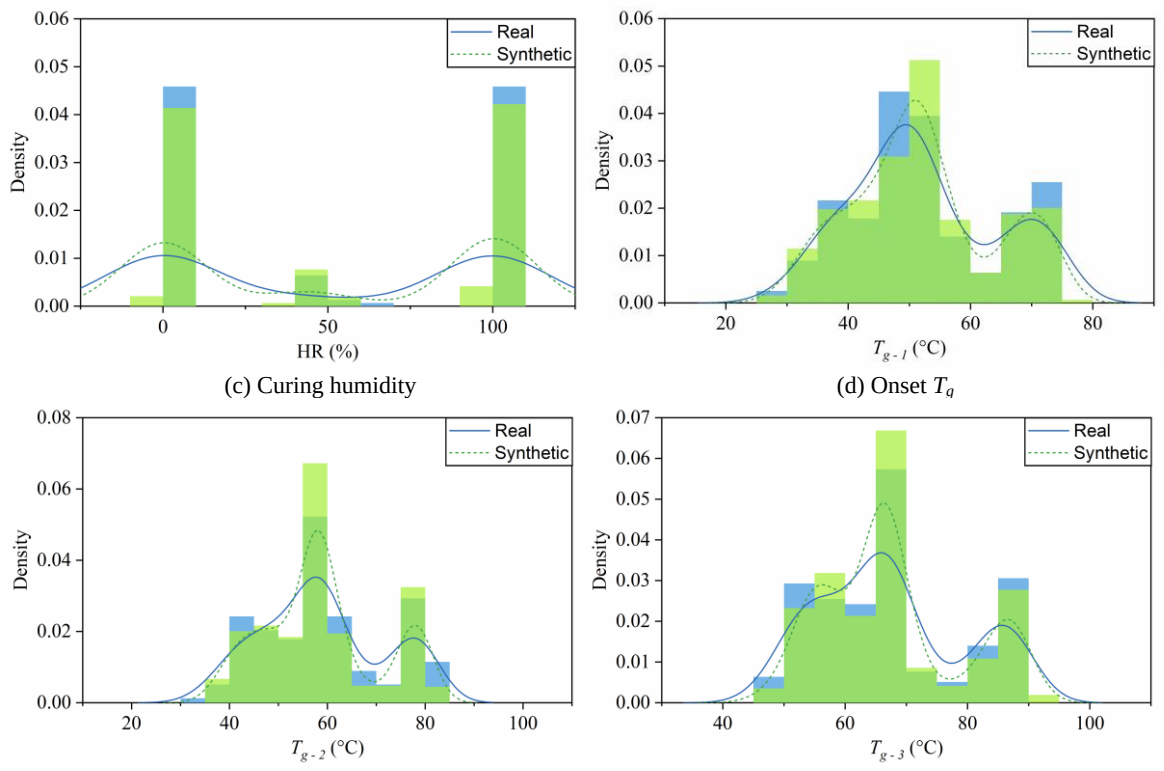
The features of the synthetic dataset generated using the trained CTGAN generator are shown in Table 4, which can be compared with those of the real dataset in Table 1. Additionally, Figure 6 provides a visual comparison of the histograms and kernel density estimates between the real and synthetic datasets. The successful learning of distribution patterns from the experimental dataset by the CTGAN is evident in its ability to generate realistic synthetic data with similar densities (see Figure 6). Whilst the generated dataset exhibits a slightly expanded range for each feature (comparing Table 4 to Table 1), the distribution-related metrics, such as standard deviation and coefficient of variation, demonstrate a discrepancy of less than 6%.

Table 4: Statistical summary of the synthetic dataset

Variables	Unit	Min	25%	50%	75%	Max	Mean	Standard deviation	Coefficient of variation
Inputs									
Curing temperature	°C	7	23	40	65	93	44.6	22.9	51.3%
Curing time	days	1	5	8	14	35	12.6	9.4	74.0%
Curing humidity	%	-1	0	45	100	102	51.1	47.5	92.9%
Outputs									
Onset T_g ($T_{g\text{-}1}$)	°C	26.8	44.2	50.9	57.4	76.1	52.0	11.5	22.1%
Inflection point T_g ($T_{g\text{-}2}$)	°C	36.2	50.3	58.0	63.1	80.7	58.8	11.7	20.0%
Peak Tan δ T_g ($T_{g\text{-}3}$)	°C	45.4	58.5	65.9	71.3	92.2	67.4	11.3	16.7%



(a) Curing temperature (b) Curing time



(c) Curing humidity (d) Onset T_g

(e) Inflection point T_g (f) Peak Tan δ T_g

Figure 6: Histograms and kernel density estimates of real and synthetic datasets

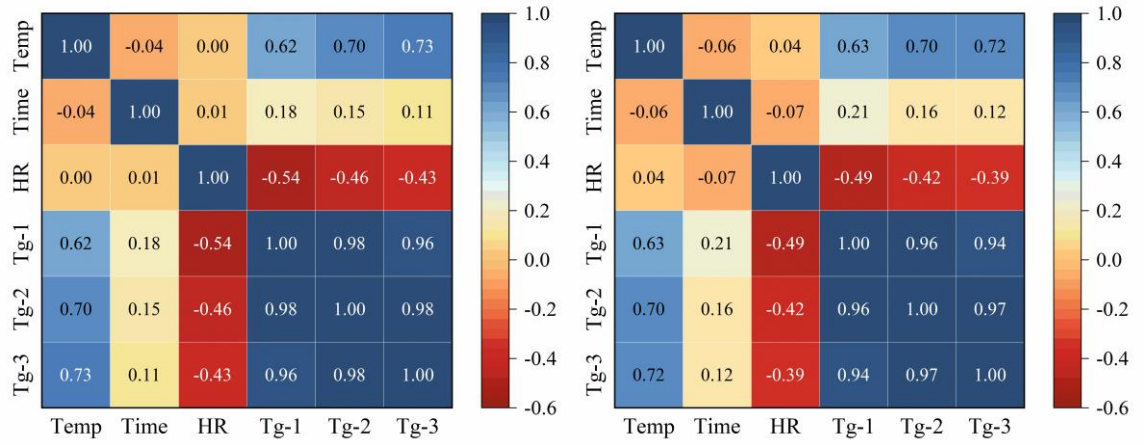
Note that the generated dataset includes results where the curing humidity level falls slightly below 0% or exceeds 100%. These anomalies were intentionally not corrected or filtered out to allow for a thorough evaluation of the reliability of the generator (G) and the feasibility of developing a CTGAN-ANN model using the generated dataset. Nevertheless, when applying the trained G from this study for engineering reference, such unrealistic data points should be excluded. The occurrence of out-of-bound results is a common issue in AI-generated data or images. Fortunately, this issue can be relatively easily addressed, either manually or by adding constraints during the data generation process.

Figure 7 illustrates the Pearson correlation coefficient (ρ) matrix for both the real and synthetic datasets. The formula for ρ is:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{8}$$

where cov () stands for the covariance, while σ_X and σ_y represent the standard deviation of X and Y, respectively.

As shown in the figure, the relationships between the variables in the synthetic dataset closely align with those of the real dataset, which again demonstrates the reliability of the trained CTGAN *G*. It is noticeable that the $T_{g\text{-}1}$, $T_{g\text{-}2}$, and $T_{g\text{-}3}$ show significant correlations with coefficients close to 1.00, as anticipated. All the T_g values of the structural adhesive exhibit a positive correlation with curing temperature and curing time, but a negative correlation with curing humidity. By comparing the absolute values, the sensitivity between the T_g values and the curing condition variables can be determined, where the most influential factor is the curing temperature, followed by the curing humidity and finally the curing time.

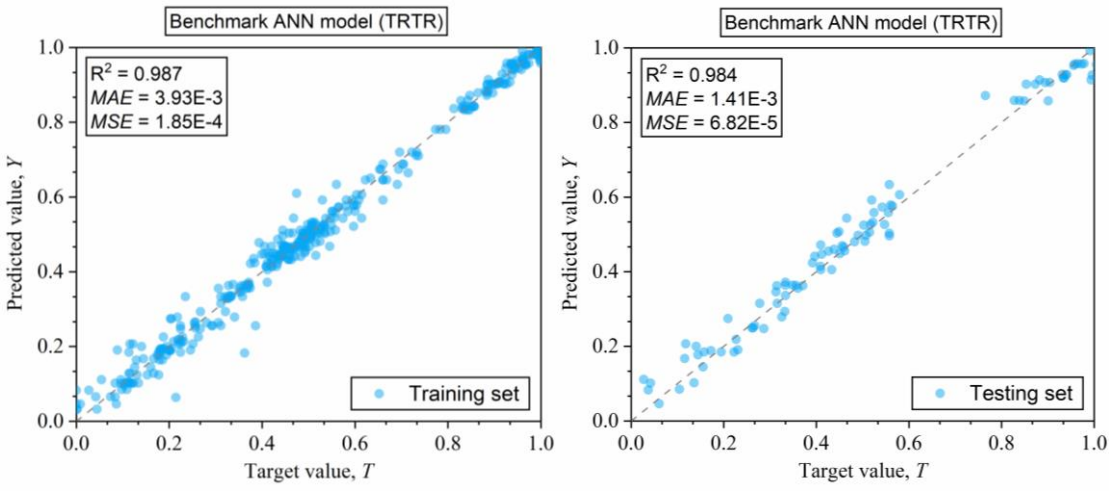


(a) Real dataset (b) Synthetic dataset

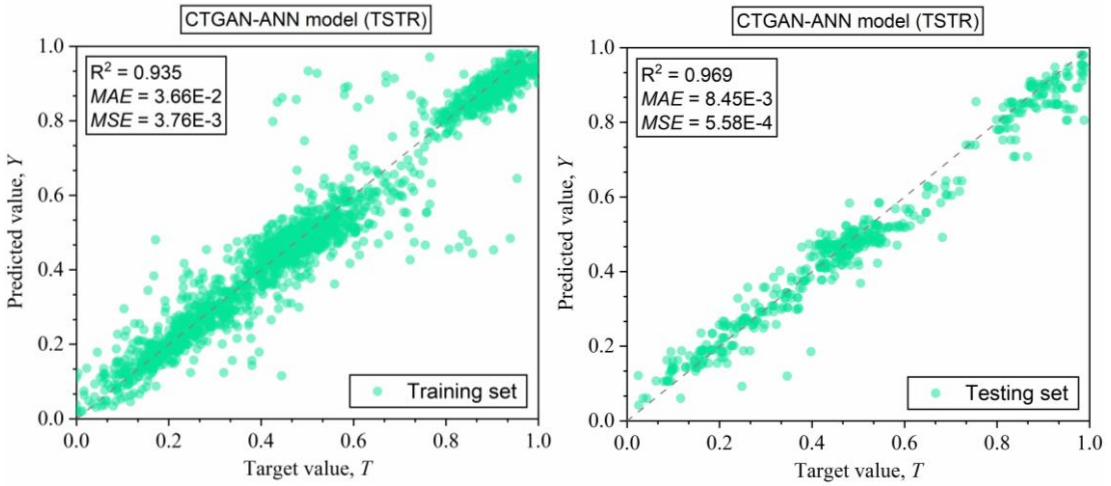
Figure 7: Correlation coefficient matrix of real and synthetic datasets

4.2 Performance of the CTGAN-ANN model

The CTGAN-ANN model achieved its best training performance at the 498th epoch with an *MSE* [Eq. (6)] equals to 3.98×10^{-3}. Figure 8 graphically shows the best performance of the CTGAN-ANN model as well as the benchmark ANN model, where *Y* and *T* represent the normalised predicted and target values, respectively.



(a) Benchmark ANN model training set (b) Benchmark ANN model testing set



(a) CTGAN-ANN model training set (b) CTGAN-ANN model testing set

Figure 8: Regression plots of the benchmark ANN model and the developed CTGAN-ANN model

As mentioned in Section 3, the benchmark ANN model was trained with 80% and tested with 20% of the experimental dataset (TRTR), while the CTGAN-ANN model was trained with a much higher number of synthetic data points and tested with 100% of the experimental results (TSTR). Both models showed high accuracy in predicting the T_g values of the structural adhesive based on the curing conditions, as evidenced by the coefficient of determination (R^2) [Eq. (4)] being close to 1. The slightly out-of-range results mentioned previously have no significant impact on the performance of the CTGAN-ANN model.

The statistical metrics [Eq. (4) – (6)] of the CTGAN-ANN model were better for the testing set as compared to the training set, indicating that the model has excellent generalisation capability and is able to predict future unseen data with greater accuracy. This impressive

performance is attributed to the model's unique architecture, which excels at uncovering and exploiting hidden patterns in the data. The use of sophisticated data augmentation techniques during the training phase enhances the model's ability to generalise, ensuring that it not only remembers the training data, but also truly understands the underlying features and can efficiently apply this knowledge to new and unseen datasets, resulting in superior prediction accuracy.

By comparison, whilst the benchmark ANN model showed better statistical performance metrics on both the testing and training sets, there is concern that the model may have been overfitted to the small experimental dataset used for training and testing. In real-world scenarios, where generalisation ability is critical, the outstanding performance of the CTGAN-ANN model on the test set was no fluke, but rather a testament to its well-designed structure and training methodology. These elements combine to produce a model that not only excels at preventing overfitting, but also demonstrates exceptional accuracy in predicting future unseen data.

The next section will present the validation work on the benchmark ANN model and the CTGAN-ANN model to further compare their capabilities fairly, and to determine whether the CTGAN approach is valid for developing data-driven ML models.

4.3 Model validation

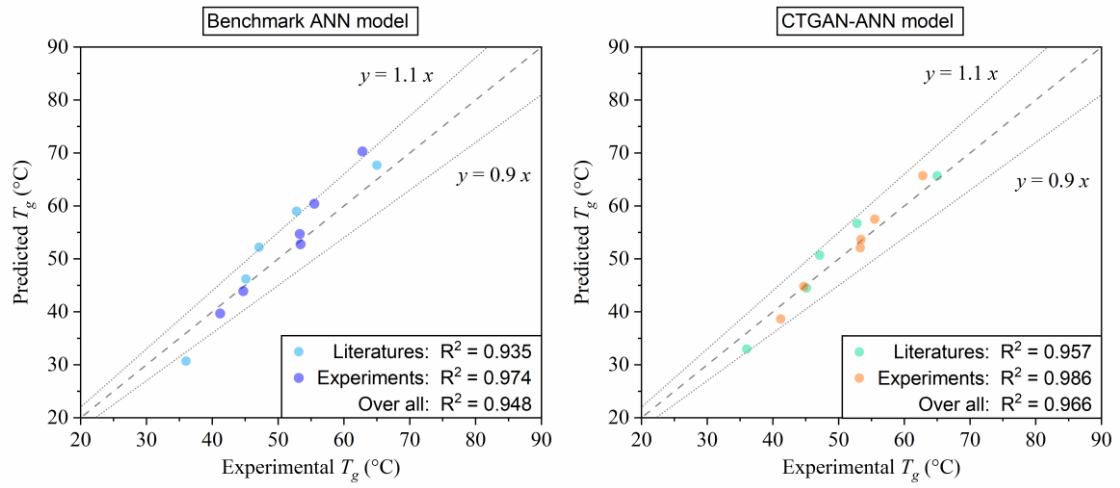
Table 5 lists the data utilised for the model validation, comprising 5 data points from 3 academic papers which were not used in the model training process and 6 data points from 2 independent experiments to ensure impartiality of the validation. The newly conducted dynamic mechanical analysis (DMA) experiments conform to the requirements of ISO 6721 (British Standards Institutions, 2019). The temperature scans were carried out using a single cantilever configuration with a dynamic strain of 0.05 mm at 1.0 Hz and a temperature ramp

of 2 °C/min (British Standards Institutions, 2019; Wang et al., 2023a). The resulting storage modulus and Tan δ responses were used to finalise the glass transition temperatures.

Table 5: Data used for model validation

Sources	Curing conditions			Glass transition temperatures		
	Temperature (°C)	Time (days)	Humidity (%)	T_{g-1} (°C)	T_{g-2} (°C)	T_{g-3} (°C)
Stratford and Bisby (Stratford and Bisby, 2012)	21	15	40	47.1	52.8	65.0
Ke et al. (Ke et al., 2020)	25	10	65	45.1	N/A	N/A
Savvilotidou et al. (Savvilotidou et al., 2017)	13	5	40	36.0	N/A	N/A
Experiment 1	20	28	60	53.3	55.5	62.8
Experiment 2	20	7	60	41.2	44.7	53.4

The listed curing conditions are the common ambient curing conditions seen on construction sites. The T_g values achieved through the different curing periods range from 36.0 °C to 65.0 °C. Figure 9 illustrates the prediction performance of the benchmark ANN model and the CTGAN-ANN model, respectively. The CTGAN-ANN model showed better performance, with the obtained coefficients of determination [Eq. (4)] close to 1.0 for both literature and experimental data, indicating a better generalisation capability. In contrast, although the benchmark ANN model showed excellent training results in Section 4.3, it had a limited predictive ability when dealing with unseen data, as evidenced by errors of over 10% (Figure 9 (a)). Therefore, the CTGAN data augmentation approach used in this study is necessary for the development of a robust data-driven machine learning model.

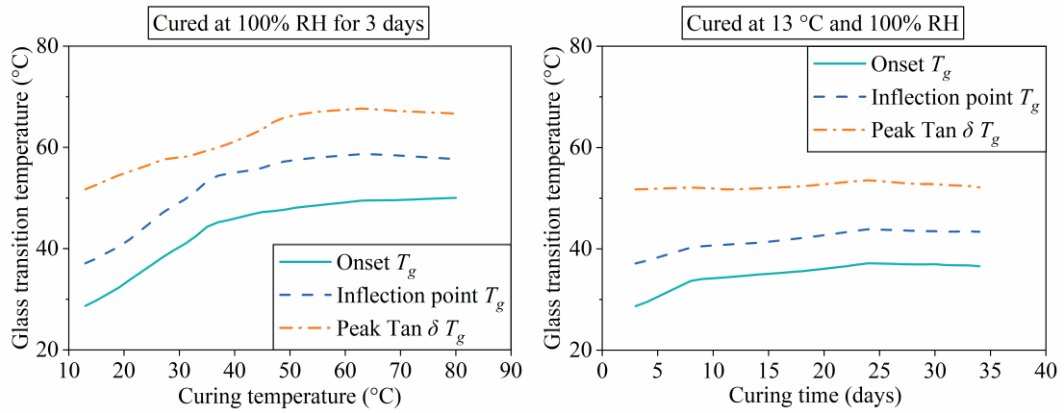


(a) Benchmark ANN model (b) CTGAN-ANN model

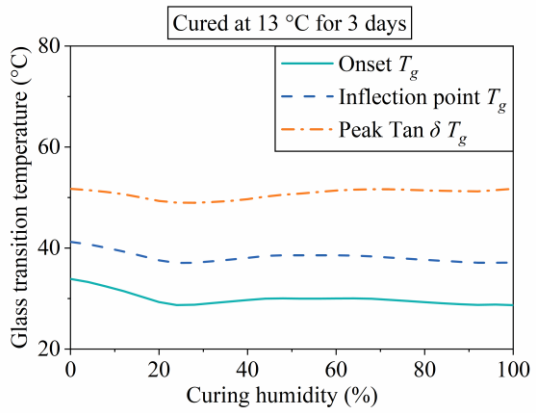
Figure 9: Prediction performance of the developed ML models

5 The effect of curing conditions on the structural adhesive

The CTGAN-ANN model developed in this study enables a large number of predictions to be obtained, facilitating a comprehensive investigation of the underlying relationship between the curing conditions and the glass transition temperatures (T_g) of the structural adhesive. The glass transition property of the polymer adhesive is of critical importance as it is closely related to the strength and stiffness that the adhesive exhibits in various engineering environments. Figure 10 shows the individual partial dependence plots of the curing variables. The value of each variable varies relative to the theoretical worst curing condition (curing at 13 °C and 100% RH for 3 days) over the range of the experimental dataset (see Table 1).



(a) Curing temperature (b) Curing time

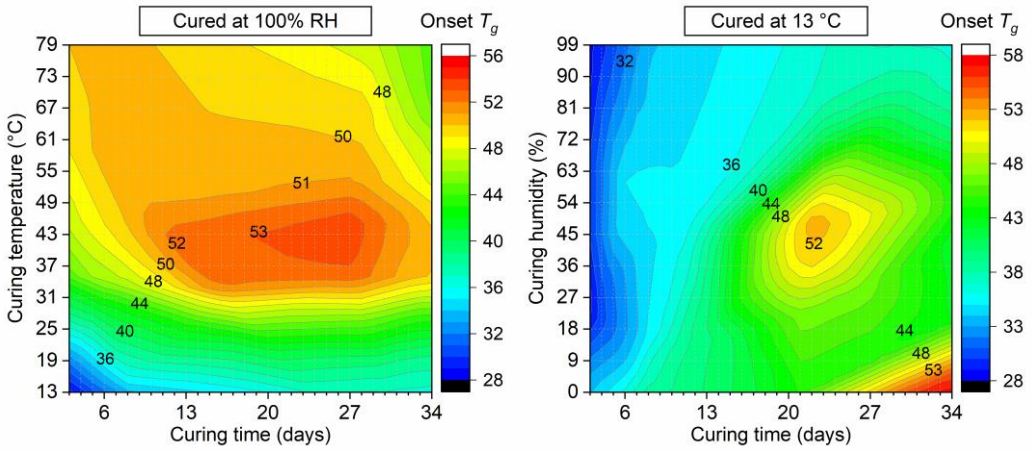


(c) Curing humidity

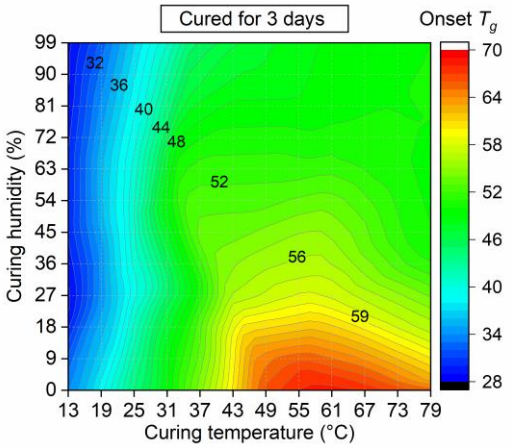
Figure 10: Individual partial dependence plots of the curing variables

The plots reveal that the curing temperature is the most significant factor influencing the T_g of the epoxy resin adhesive, which is consistent with the results of the Pearson correlation coefficient matrix in Figure 7. With an increase in curing temperature from 13°C, the cross-linking of the epoxy adhesive also increases, causing the free volume to decrease, ultimately leading to an increase in the T_g values until they reach a plateau. An increase in curing time leads to a similar trend, however, with an increase in curing humidity, an almost opposite trend is observed.

Figure 9 further illustrates the 2-feature partial dependence plots of the curing variables, which indicates that increasing the curing temperature or time, or decreasing the curing humidity, does not always result in higher T_g values. In fact, extremely high curing temperatures or extremely long curing periods may have a negative impact on the T_g of the structural adhesive due to the thermal degradation or oxidative cross-linking effects. Practical engineering applications can benefit from using the presented plots to select potentially optimal curing conditions that can achieve the desired T_g values. This can save valuable time and resources compared to conducting numerous trial experiments.



(a) Effect of curing temperature and curing time (b) Effect of curing time and curing humidity



(a) Effect of curing temperature and curing humidity

Figure 11: 2-feature partial dependence plots of the curing variables

A three-dimensional scatter plot was also created using the powerful ML model to visualise the relationship between the curing conditions and the T_g values of the structural adhesive more comprehensively, as shown in Figure 12. The plot includes 36992 data points representing nearly all possible combinations of the curing variables within the range of the experimental dataset, which would be impossible to obtain by conventional experimental methods. The scatter plot demonstrates the correlation with the partial dependency plots (Figure 10 and Figure 11) and provides the engineer with valuable insight into the effects of different curing conditions.

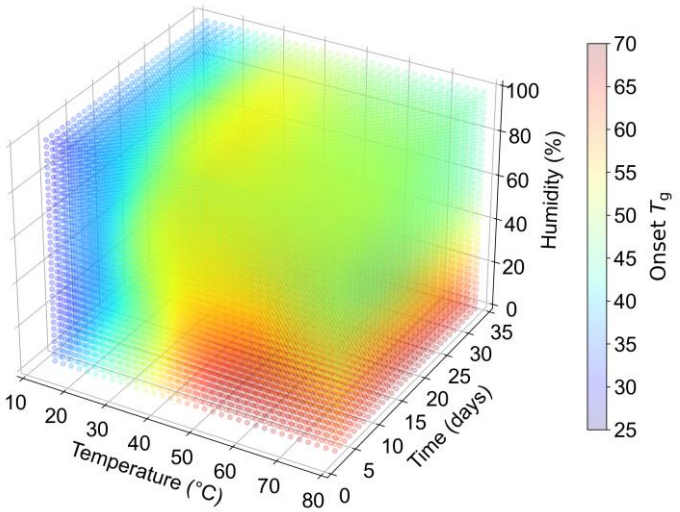


Figure 12: 3D scatter plot showing the combine effects of three curing variables

6 Conclusions and recommendations

This study employs a deep ML approach to evaluate the impact of curing conditions on the glass transition of the structural epoxy adhesive. To overcome the issue of limited experimental data, the CTGAN technique was applied to generate sufficient synthetic data, which was then integrated with the ANN technique to build a robust predictive model. The CTGAN-ANN model was developed using the TSTR approach and validated against the results of both newly conducted experiments and several published studies. The model has demonstrated impressive generalisation and predictive capabilities, indicating its potential to be a robust tool for predicting curing outcomes. A comprehensive evaluation of the effect of each curing variable was therefore conducted based on a large number of predicted results.

- The curing temperature is the most influential factor on the T_g values of the adhesive, followed by the curing humidity and the curing time. A positive correlation was observed between the T_g values and the curing temperature and curing time, while a negative correlation was observed between the T_g values and the curing humidity.

- The presented partial dependence plots can be used to select potentially optimal curing conditions to achieve the desired T_g values, which can be expected to save time and resources in practical applications.

The developed CTGAN-ANN ML model is currently constrained to the analysis of the Sikadur-330 structural adhesive. However, as more experimental data becomes available, the model can be upgraded to include different structural adhesives, since the CTGAN can also deal with discrete data such as the types of adhesives. Nevertheless, the ML model developed could inform the practical use of the structural adhesive in civil engineering, and the demonstrated approach using CTGAN and ANN can be applied to solve similar problems in structural materials where the available experimental data are insufficient for effective analysis and modelling.

CRediT authorship contribution statement

Songbo Wang: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Funding acquisition. **Haixin Yang:** Investigation, Visualization, Writing - original draft. **Tim Stratford:** Conceptualization. **Jiayi He:** Investigation. **Biao Li:** Writing - review & editing, Validation. **Jun Su:** Resources, Project administration.

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.

Acknowledgments

The authors are grateful for the support from Key Laboratory of Intelligent Health Perception and Ecological Restoration of Rivers and Lakes. This work was funded by the

Natural Science Foundation of Hubei Province of China (Project No. 2023AFB388) and the Doctoral Research Starting Foundation of Hubei University of Technology under (Project No. XJ2022001301).

Appendix A: Supplementary material

Supplementary material to this article can be found in the attachments.

References

Agha, A., Abu-Farha, F., 2021. Viscoelastic model to capture residual stresses in heat cured dissimilar adhesive bonded joints. Int. J. Adhes. Adhes. 107, 102844. https://doi.org/10.1016/j.ijadhadh.2021.102844

Almustafa, M.K., Nehdi, M.L., 2022. Machine learning prediction of structural response of steel fiber-reinforced concrete beams subjected to far-field blast loading. Cem. Concr. Compos. 126, 104378. https://doi.org/10.1016/j.cemconcomp.2021.104378

Ascione, F., Granata, L., Guadagno, L., Naddeo, C., 2021. Hygrothermal durability of epoxy adhesives used in civil structural applications. Compos. Struct. 265, 113591. https://doi.org/10.1016/j.compstruct.2021.113591

Ascione, F., Granata, L., Lombardi, A., 2022. The influence of the hygrothermal aging on the strength and stiffness of adhesives used for civil engineering applications with pultruded profiles: an experimental and numerical investigation. J. Adhes. 98, 1733–1771. https://doi.org/10.1080/00218464.2021.1936507

Ben Chaabene, W., Nehdi, M.L., 2021. Genetic programming based symbolic regression for shear capacity prediction of SFRC beams. Constr. Build. Mater. 280, 122523. https://doi.org/10.1016/j.conbuildmat.2021.122523

British Standards Institutions, 2019. BSI Standards Publication Plastics — Determination of dynamic mechanical properties, BS ISO 6721:2019. https://doi.org/https://doi.org/10.3403/BSISO6721

Cadei, J.M.C., Stratford, T.J., Hollaway, L.C., Dcukett, W.G., 2004. Strengthening metallic structures using externally bonded fibre-reinforced polymers. Ciria.

Carbas, R.J.C., Marques, E.A.S., Da Silva, L.F.M., Lopes, A.M., 2014. Effect of cure temperature on the glass transition temperature and mechanical properties of epoxy adhesives. J. Adhes. 90, 104–119. https://doi.org/10.1080/00218464.2013.779559

Chollet, F., others, 2015. Keras documentation. keras. io 33.

Erdal, H.I., 2013. Two-level and hybrid ensembles of decision trees for high performance concrete compressive strength prediction. Eng. Appl. Artif. Intell. 26, 1689–1697. https://doi.org/10.1016/j.engappai.2013.03.014

Esteban, C., Hyland, S.L., Rätsch, G., 2017. Real-valued (medical) time series generation with recurrent conditional gans. arXiv Prepr. arXiv1706.02633.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. arXiv. arXiv Prepr. arXiv1406.2661.

Gu, Z., Liu, Y., Hughes, D.J., Ye, J., Hou, X., 2021. A parametric study of adhesive bonded joints with composite material using black-box and grey-box machine learning methods: Deep neuron networks and genetic programming. Compos. Part B Eng. 217, 108894. https://doi.org/10.1016/j.compositesb.2021.108894

Habibi, O., Chemmakha, M., Lazaar, M., 2023. Imbalanced tabular data modelization using CTGAN and machine learning to improve IoT Botnet attacks detection. Eng. Appl. Artif. Intell. 118, 105669. https://doi.org/10.1016/j.engappai.2022.105669

He, J., Xian, G., Zhang, Y.X., 2021. Numerical modelling of bond behaviour between steel and CFRP laminates with a ductile adhesive. Int. J. Adhes. Adhes. 104, 102753. https://doi.org/10.1016/j.ijadhadh.2020.102753

He, Z., Zhou, W., 2022. Generation of synthetic full-scale burst test data for corroded pipelines using the tabular generative adversarial network. Eng. Appl. Artif. Intell. 115, 1–14. https://doi.org/10.1016/j.engappai.2022.105308

Heshmati, M., Haghani, R., Al-Emrani, M., 2015. Environmental durability of adhesively bonded FRP/steel joints in civil engineering applications: State of the art. Compos. Part B Eng. 81, 259–275. https://doi.org/10.1016/j.compositesb.2015.07.014

Higgins, A., 2000. Adhesive bonding of aircraft structures. Int. J. Adhes. Adhes. 20, 367–376.

Hisham, M., Hamdy, G.A., El-Mahdy, O.O., 2021. Prediction of temperature variation in FRP-wrapped RC columns exposed to fire using artificial neural networks. Eng. Struct. 238, 112219. https://doi.org/10.1016/j.engstruct.2021.112219

Huang, J.S., Liew, J.X., Liew, K.M., 2021. Data-driven machine learning approach for exploring and assessing mechanical properties of carbon nanotube-reinforced cement composites. Compos. Struct. 267, 113917. https://doi.org/10.1016/j.compstruct.2021.113917

I. M. Foletti, A., Sena Cruz, J., Vassilopoulos, A.P., 2020. Fabrication and curing conditions effects on the fatigue behavior of a structural adhesive. Int. J. Fatigue 139, 105743. https://doi.org/10.1016/j.ijfatigue.2020.105743

Jahani, Y., Baena, M., Barris, C., Perera, R., Torres, L., 2022. Influence of curing, post-curing and testing temperatures on mechanical properties of a structural adhesive. Constr. Build. Mater. 324, 126698. https://doi.org/10.1016/j.conbuildmat.2022.126698

Karbhari, V.M., Shulley, S.B., 1995. Use of Composites for Rehabilitation of Steel Structures—Determination of Bond Durability. J. Mater. Civ. Eng. 7, 239–245. https://doi.org/10.1061/(asce)0899-1561(1995)7:4(239)

Ke, L., Li, C., He, J., Dong, S., Chen, C., Jiao, Y., 2020. Effects of elevated temperatures on mechanical behavior of epoxy adhesives and CFRP-steel hybrid joints. Compos. Struct. 235. https://doi.org/10.1016/j.compstruct.2019.111789

Marani, A., Nehdi, M.L., 2022. Predicting shear strength of FRP-reinforced concrete beams using novel synthetic data driven deep learning. Eng. Struct. 257, 114083.

https://doi.org/10.1016/j.engstruct.2022.114083

Marani, A., Zhang, L., Nehdi, M.L., 2023. Design of concrete incorporating microencapsulated phase change materials for clean energy: A ternary machine learning approach based on generative adversarial networks. Eng. Appl. Artif. Intell. 118, 105652. https://doi.org/10.1016/j.engappai.2022.105652

Menard, K., Menard, N., 2020. Dynamic Mechanical Anaylsis: a practical introduction. CRC Press.

Moon, J., Jung, S., Park, S., Hwang, E., 2020. Conditional tabular GaN-based two-stage data generation scheme for short-term load forecasting. IEEE Access 8, 205327–205339. https://doi.org/10.1109/ACCESS.2020.3037063

Naser, M.Z., 2019. AI-based cognitive framework for evaluating response of concrete structures in extreme conditions. Eng. Appl. Artif. Intell. 81, 437–449. https://doi.org/10.1016/j.engappai.2019.03.004

Peres, R.S., Guedes, M., Miranda, F., Barata, J., 2021. Simulation-Based Data Augmentation for the Quality Inspection of Structural Adhesive with Deep Learning. IEEE Access 9, 76532–76541. https://doi.org/10.1109/ACCESS.2021.3082690

Sahin, M.U., Dawood, M., 2016. Experimental Investigation of Bond between High-Modulus CFRP and Steel at Moderately Elevated Temperatures. J. Compos. Constr. 20, 1–11. https://doi.org/10.1061/(ASCE)CC.1943-5614.0000702

Salehi, H., Burgueño, R., 2018. Emerging artificial intelligence methods in structural engineering. Eng. Struct. 171, 170–189. https://doi.org/10.1016/j.engstruct.2018.05.084

Samuel, A.L., 1959. Machine learning. Technol. Rev. 62, 42–45.

Savvilotidou, M., Vassilopoulos, A.P., Frigione, M., Keller, T., 2017. Development of physical and mechanical properties of a cold-curing structural adhesive in a wet bridge environment. Constr. Build. Mater. 144, 115–124. https://doi.org/10.1016/j.conbuildmat.2017.03.145

SIKA, 2017. Sikadur®-330 Data Sheet.

Sousa, J.M., Correia, J.R., Gonilha, J., Cabral-Fonseca, S., Firmo, J.P., Keller, T., 2019. Durability of adhesively bonded joints between pultruded GFRP adherends under hygrothermal and natural ageing. Compos. Part B Eng. 158, 475–488. https://doi.org/10.1016/j.compositesb.2018.09.060

Stratford, T.J., Bisby, L.A., 2012. Effect of warm temperatures on externally bonded FRP strengthening. J. Compos. Constr. 16, 235–244. https://doi.org/10.1061/(ASCE)CC.1943-5614.0000260

Su, M., Zhong, Q., Peng, H., Li, S., 2021. Selected machine learning approaches for predicting the interfacial bond strength between FRPs and concrete. Constr. Build. Mater. 270, 121456. https://doi.org/10.1016/j.conbuildmat.2020.121456

Teng, J.G., Yu, T., Fernando, D., 2012. Strengthening of steel structures with fiber-reinforced polymer composites. J. Constr. Steel Res. 78, 131–143. https://doi.org/10.1016/j.jcsr.2012.06.011

Wang, S., Stratford, T., Reynolds, T., 2023a. International Journal of Adhesion and Adhesives

636 Viscoelastic model for analysing the behaviour of adhesive-bonded FRP-to-steel joints in
637 civil engineering applications. Int. J. Adhes. Adhes. 123, 103359.
638 https://doi.org/10.1016/j.ijadhadh.2023.103359

639 Wang, S., Stratford, T., Reynolds, T.P.S., 2022. Creep of an FRP-strengthened metallic beam
640 under cyclic temperature and cyclic load. J. Constr. Steel Res. 196, 107417.
641 https://doi.org/10.1016/j.jcsr.2022.107417

642 Wang, S, Stratford, T., Reynolds, T.P.S., 2022. A comparison of the influence of nonlinear and
643 linear creep on the behaviour of FRP-bonded metallic beams at warm temperatures.
644 Compos. Struct. 281, 115117. https://doi.org/10.1016/j.compstruct.2021.115117

645 Wang, S., Stratford, T., Reynolds, T.P.S., 2021. Linear creep of bonded FRP-strengthened
646 metallic structures at warm service temperatures. Constr. Build. Mater. 283, 122699.
647 https://doi.org/10.1016/j.conbuildmat.2021.122699

648 Wang, S., Xu, Z., Stratford, T., Li, B., Zeng, Q., Su, J., 2023b. Machine learning approach for
649 analysing and predicting the modulus response of the structural epoxy adhesive at elevated
650 temperatures. J. Adhes. 00, 1–19. https://doi.org/10.1080/00218464.2023.2183851

651 Xu, L., Skoularidou, M., Cuesta-Infante, A., Veeramachaneni, K., 2019. Modeling tabular data
652 using conditional GAN. Adv. Neural Inf. Process. Syst. 32.

653 Yan, S., Zou, X., Ilkhani, M., Jones, A., 2020. An efficient multiscale surrogate modelling
654 framework for composite materials considering progressive damage based on artificial
655 neural networks. Compos. Part B Eng. 194, 108014.
656 https://doi.org/10.1016/j.compositesb.2020.108014

657 Yin, L., Gao, Q., Zhao, L., Zhang, B., Wang, T., Li, S., Liu, H., 2020. A review of machine
658 learning for new generation smart dispatch in power systems. Eng. Appl. Artif. Intell. 88,
659 103372. https://doi.org/10.1016/j.engappai.2019.103372

660 Zhang, Y., Ling, C., 2018. A strategy to apply machine learning to small datasets in materials
661 science. npj Comput. Mater. 4, 28–33. https://doi.org/10.1038/s41524-018-0081-z

662 Zhao, X.L., Zhang, L., 2007. State-of-the-art review on FRP strengthened steel structures. Eng.
663 Struct. 29, 1808–1823. https://doi.org/10.1016/j.engstruct.2006.10.006

664 Zhao, X.Y., Chen, J.X., Chen, G.M., Xu, J.J., Zhang, L.W., 2023. Prediction of ultimate
665 condition of FRP-confined recycled aggregate concrete using a hybrid boosting model
666 enriched with tabular generative adversarial networks. Thin-Walled Struct. 182, 110318.
667 https://doi.org/10.1016/j.tws.2022.110318

668