DOI: 10.1111/mice.12965

INDUSTRIAL APPLICATION

COMPUTER-AIDED CIVIL AND INFRASTRUCTURE ENGINEERING WILEY

A knowledge transfer enhanced ensemble approach to predict the shear capacity of reinforced concrete deep beams without stirrups

Hongrak Pak | **Samuel Leach** | **Seung Hyun Yoon** | **Stephanie German Paal**

Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, Texas, USA

Correspondence
Stephanie German Paal, Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX 77843, USA.
Email: spaal@civil.tamu.edu

Abstract

This paper proposes a novel learning algorithm, the transfer ensemble neural network (TENN) model, to increase the performance of shear capacity predictions on small datasets, illuminating the usefulness of advanced machine learning techniques in general. By incorporating ensemble learning and transfer learning, the TENN model is designed to control the high variability inherent in machine learning models trained on small amounts of data. The novel TENN model is validated to predict the shear capacity of deep reinforced concrete (RC) beams without stirrups across varying data availability levels. Knowledge acquired through pretraining a model on slender RC beams is utilized for training a model to better predict the shear capacity of deep RC beams without stirrups. To evaluate the performance of the TENN model, three baseline models are developed and examined across multiple data availability levels. The novel TENN model outperforms the baseline models, particularly when trained on a very limited dataset. Furthermore, the proposed algorithm achieves a higher accuracy than the currently accepted design standards in accurately predicting deep RC beams' shear capacity and demonstrates the capabilities of the TENN model to extrapolate in other domains where large-scale or physical testing is cost-prohibitive.

1 | BACKGROUND

Generally, reinforced concrete (RC) beams fail in flexure or shear. Shear failure occurs in a relatively brittle manner compared to the more ductile flexural failure. As sudden brittle failure of RC beams may cause severe collapse of structures, loss of properties, and casualties, accurate prediction of shear behavior is critical. Although RC beams are designed to fail in flexure to avoid such hazards, deep beams are usually governed by shear. In addition, considerable research has shown that the American Concrete Institute (ACI) provisions for estimating the shear capacity of large, narrow, lightly reinforced beams without stirrups are not accurate (Bentz, 2005; Lubell et al., 2004). Therefore, further understanding of the shear behavior of deep beams and methods for accurate prediction of shear capacity is necessary.

Flexural behavior can be predicted with reasonable accuracy; however, shear behavior still lacks an equation that is universally accepted. To predict the shear capacity of slender beams, which have a/d values approximately higher than 2.5, a consensus has been reached that shear capacity is taken as the sum of the shear strengths of the concrete and transverse reinforcement (Hu & Wu, 2018). However, although many experiments and numerical analyses have been carried out to understand the influence of various parameters such as shear span-to-depth ratio, reinforcement ratio, failure mode, and concrete strength on the shear capacity of deep beams, there is still disagreement regarding the shear transfer mechanisms and critical influencing parameters, causing discrepancies in the predicted shear capacity between different codes and researchers (J.-H. Zhang et al., 2020). Thus, there is still a need to develop techniques that can more accurately predict the shear capacity of RC beams.

Recent advancements in the field of machine learning (ML) have enabled many applications in estimating structural behavior in civil engineering (CE) (Adeli, 2001; Gao & Mosalam, 2018; German et al., 2012; Rafiei et al., 2017a; Rafiei et al., 2017b). By using databases based on experimental results, real-world data, and/or simulations, various ML models have been proposed to predict the behavior of different structural components and systems with better accuracy. However, one crucial and inevitable assumption exists in previous studies. Traditionally, ML algorithms assume that both training and testing data have identical explanatory features and have the same, or almost identical, distributions to obtain a good model from the established data. Accurately predicting the response variable(s) when this assumption is not satisfied is difficult for conventional ML models. One obvious solution is to collect more data of interest; however, this often requires a lot of time and money, and it is sometimes impractical or impossible. To avoid unnecessary time or cost, transfer learning (TL) or ensemble learning (EL) can be adopted. TL is a specific subfield of ML focusing on breaking the previous assumption by learning from nonidentical but relevant domains. It is often called by different names (e.g., covariate shift, domain adaptation, instance-based transfer) based on the primary approach used to transfer knowledge. EL is a subfield of ML which focuses on aggregating multiple learners trained on different subsets of datasets. In this study, by incorporating TL and EL, a new algorithm is proposed to overcome the problem of data scarcity and provide a more efficient ML model with low prediction variance even when trained on a small number of training samples.

2 | RELATED WORK

2.1 | Traditional approaches

The shear behavior of RC deep beams has been a prime focus of investigation over the last few decades, as the calculation of the shear capacity of RC deep beams needs further improvement due to the complexity of the shear transfer mechanism (Bazant & Sun, 1987; Cladera & Marí, 2004a, 2004b; Mansour et al., 2004). As a result of extensive studies, a large number of experimental data are available in the literature. Based on the results, understanding of shear behavior has been notably improved, and various theories and design codes have been realized (Reineck et al., 2013, 2014; Reineck & Todisco, 2014). The strut-and-tie model (STM) (Schlaich et al., 1987) was developed based on the failure pattern of deep beams, and compression field theory, modified compression field theory (MCFT) (Vecchio & Collins, 1986), and so on were developed following the development of STM. STM has been generally recognized as the most rational method for designing deep beams since its development. Accordingly, design standards such as ACI 318 and Canadian Standards Association (CSA) A23.3 recommend STM as a design approach for deep beams.

STM is more appropriate for deep beams because a significant portion of shear is transferred directly from the loading point to supports via diagonal compression struts (Hu & Wu, 2018). In STM, the diagonal strut and truss concept is used to describe the shear failure of an RC deep beam. While the diagonal strut represents compression stress fields that develop in the concrete web among diagonal cracks, the truss mechanism accounts for the horizontal and vertical web reinforcement (Chetchotisak et al., 2014). In other words, to apply STM and provide a safe design, both horizontal and vertical web reinforcement is desired; however, this is not applicable when considering shear in RC deep beams without stirrups because stirrups act as vertical reinforcement. Some theories like MCFT have proved their applicability to members without shear reinforcement; however, the use of such theories in practice remains complicated (Muttoni & Fernández Ruiz, 2008).

Because most of the current equations for calculating the shear capacity of beams without stirrups are based on empirical observation, they do not represent the actual shear failure mechanism and often are highly conservative despite identifying the important variables that affect shear capacity (T. Zhang et al., 2016). Therefore, in this study, an ML approach will be adopted to estimate the shear capacity of RC deep beams without shear reinforcement. Accurate predictions of the shear capacity of RC deep beams will provide a better understanding of the shear failure mechanism. Then, the estimated shear capacity will be compared with the results based on the equations for the shear capacity of nonprestressed concrete members listed in ACI 318-19 (ACI Committee 318, 2020) and CSA A23.3 19 (Canadian Standards Association , 2019).

In ACI 318-19 (ACI Committee 318, 2020), the shear strength of nonprestressed concrete members is given as

$$V_n = \left[0.66\lambda_s\lambda(\rho_w)^{\frac{1}{3}}\sqrt{f'_c} + \frac{N_u}{6A_g}\right]b_w d + \frac{A_v f_y d}{s}, \quad (1)$$

where λ is a factor that accounts for lightweight concrete (for normal weight concrete, λ is taken as 1), f_c' is the concrete compressive strength, ρ_w is the flexural reinforcement ratio, N_u is the axial load, A_g is the cross-sectional area, b_w is the width of the web, d is the effective depth of the section, A_v is the area of transverse reinforcement, f_y is the yield strength of transverse reinforcement, and λ_s is the size effect modification factor.

In CSA A23.3-19 (Canadian Standards Association, 2019), the shear strength of RC beams without shear reinforcement is considered as a function of the concrete compressive strength and sectional dimensions. The shear strength in an RC beam without shear reinforcement is given as

$$V_c = \varphi_c \lambda \beta \sqrt{f_c'} b_w d_v, \tag{2}$$

where φ_c is the resistance factor for concrete, which is taken as 0.65, λ is the strength reduction factor, taken as 1 for normal density concrete, d_v is the effective depth, and β is a factor to account for aggregate interlock in concrete members.

Furthermore, β can be calculated as follows:

$$\beta = \begin{cases} 0.18, & \text{if minimum transverse reinforcement is provided} \\ \dfrac{230}{1000 + d_v}, & a_g \geq 20 \text{ mm} \\ \dfrac{230}{1000 + S_{ze}}, & a_g < 20 \text{ mm} \end{cases} \tag{3}$$

Here, S_{ze} can be conservatively taken as d_v and a_g is the maximum nominal size of coarse aggregate.

2.2 | Machine learning approaches

In recent years, ML techniques have been widely adopted in CE, and most of them have shown excellent performance to find the optimal values in a given condition (Adeli & Panakkat, 2009; Butcher et al., 2014; Rafiei et al., 2016; Chen & Feng, 2022). In the field of structural engineering, many researchers have focused on developing ML techniques to accurately predict the shear behavior of RC beams. Sanad and Saka (2001) trained an artificial neural network (ANN) to predict the ultimate shear capacity of deep beams with 111 experimental results and concluded that reliable predictions can be made even if equations used to calculate the ultimate shear capacity of deep beams are not provided. Other researchers not only proved the reliability of ANNs in predicting the shear capacity but also compared the prediction with empirical equations in the literature (Mansour et al., 2004; Yang et al., 2007). Mansour et al. (2004) trained an ANN with 176 experiments to predict the shear capacity of RC beams, and Yang et al. (2007) used 631 experiments to train an ANN with the early stopping technique to predict the shear capacity of deep beams. Abdalla et al. (2007) used a multilayer ANN trained with 164 experimental observations to predict the shear capacity of RC beams. Koo et al. (2021) also used an ANN to predict shear capacity and tried to increase prediction accuracy by preprocessing the data with a principal component analysis. They compared the prediction to the existing building codes and conducted a parametric study to determine the effects of the variables. In addition to making a prediction, (Cladera & Marí, 2004a, 2004b) studied the shear design procedure of RC beams with and without stirrups using ANNs with 123 experiments. Oreta (2004) examined the influence of the size effect on the shear capacity of RC beams without stirrups using a multilayer ANN. In total, 118 experimental observations were used, and the trained model could successfully simulate the effects of input variables on the shear capacity of RC beams without stirrups. Moving away from ANNs, Mohammed and Ismail (2021) predicted the shear capacity of RC beams using a random forest (RF). The data they used comprised 349 experimental samples, and they validated the prediction by comparing it to the results of a support vector machine (SVM) and other empirical equations. They concluded that the proposed model was a robust approach to predict the shear capacity of RC beams. Although several of the aforementioned researchers have successfully demonstrated the applicability of ML approaches to predict the shear capacity of various types of RC beams, the number of available data samples is still lacking compared to that of other fields, such as image processing, transportation, or natural language processing, where ML approaches are easily and actively adopted. One dilemma is that collecting a large amount of data in structural engineering is often difficult, considering that setting up a full-scale structural experiment requires enormous cost, time, and space. However, EL and TL provide one way to circumvent these challenges and may offer a remedy for the data scarcity problem in structural engineering.

2.3 | Ensemble learning and transfer learning

EL is the process of combining predictions from multiple learners to achieve more powerful performance. Through the aggregation of the predictions of multiple models, it is possible to achieve a more robust model with higher accuracy compared to that of a single learner (Sagi & Rokach, 2018). An ensemble method can be applied to any ML algorithm, such as a decision tree, SVM, neural network (NN), etc. The main premise of EL is that by aggregating multiple weak learners, errors of a single learner will likely be compensated for by other weak learners. Consequently, the overall prediction of the model would become better than that of a weak learner. Thus, EL is considered as a state-of-the-art approach to solving a plethora of ML challenges. While a single ML learner is more frequently used, EL has been used in several areas within structural engineering. Huang et al. (2018) utilized EL to increase the performance of models in structural damage detection. They compared these EL prediction results to those of SVM, XGBoost, and RF models. Based on their results, EL increased the classification accuracy by over 20%. Feng et al. (2020) used several ensemble learners to predict failure modes and the bearing capacity of RC columns. They compared single learners, including a classification and regression tree, SVM, and ANN, to EL algorithms including RF and a gradient-boosting decision tree. They proved that the EL methods provide better performance than the single learners. There have also been more studies related to EL for predicting various structural performances, for example, shear capacity prediction of RC deep beams (Feng et al., 2021), Fiber reinforced polymer (FRP)-concrete bond strength (Chen, Zhang et al., 2021), and Carbon fiber reinforced polymer (CFRP)-steel bond strength (Chen, Feng et al., 2021). These studies have demonstrated the capabilities of EL in both classification and regression problems.

TL is an ML technique that seeks to apply knowledge learned from one or more domains to another domain (Pan & Yang, 2009). Traditional ML algorithms implicitly assume that the training and testing data should share the same features, distributions, and tasks. Thus, it is necessary to gather appropriate data samples and train a new ML model if any of the three are different. This requires extensive time and money, and it may be impractical if the number of new data samples is still insufficient. TL has the ability to address this problem while maintaining good predictive performance. The primary goal of TL is to reduce the amount of data necessary for training a high-performance model in a specific domain (J.-H. Zhang et al., 2020). A common issue, especially in the field of structural engineering, is the difficulty of gathering large, labeled datasets suitable for training an ML model. TL aims to train a model on a dataset similar to the domain in which data is difficult to obtain and then transfer that knowledge to the desired domain for further training. While various ML theories and techniques have been adopted in CE, few TL studies have been conducted (Azimi & Pekcan, 2020; Luo & Paal, 2021; Pak & Paal, 2022). Furthermore, the majority are focused on image-based models (Cha et al., 2017; Cheng et al., 2021; Leach et al., 2021; Li et al., 2019).

In contrast to the number of studies regarding ML, there has been a very limited number of studies combining EL and TL. Liu et al. (2017) designed a framework for ensemble TL to increase classification accuracy when the number of training data is insufficient. This study dealt with weighted-resampling-based, bagging-based, and multiboosting-based TL algorithms. They proved that the performance of their algorithm was excellent compared to that of TrAdaBoost, naïve Bayes, and SVM. Zhu et al. (2020) also extended the classic TL model through EL and demonstrated its superior generalization ability. After dividing their target dataset into 10 folds, deep neural networks (DNN) were trained on the source dataset and then refined on the target dataset to make an ensembled prediction from the 10 models. Although some researchers have demonstrated the efficacy of combining EL and TL, there remain many areas that require more thorough investigation. First, the previous studies have primarily addressed the task of classification. Combined ensemble and TL algorithms need to be developed for the task of regression. Second, no research has been conducted on the simultaneous application of bagging and random subspace sampling. Third, the application of these combination methods should be investigated in civil engineering.

2.4 | Significance of this study

To increase the accuracy of shear capacity predictions of ML models and achieve better generality than those of the existing empirical equations, collecting more data and understanding the complex relationship between parameters is important. However, continuously conducting additional structural experiments is difficult due to the aforementioned limitations. Therefore, because large amounts of data already exist for slender RC beams, the application of ML techniques specifically designed to utilize these existing experiments to predict the shear capacity of deep RC beams is beneficial. The hypothesis of this work is that it is possible to transfer knowledge from slender to deep beams, hence validating this type of approach for other components or materials where even less data is available. In this study, a new algorithm is proposed to predict the shear capacity of deep beams without stirrups and then evaluated by comparing the shear capacity calculated based on ACI 318-19 (ACI Committee 318, 2020) and CSA (Canadian Standards Association , 2019). Although many studies related to ML have been made to accurately predict the shear capacity of RC beams, the scarcity of the experimental data limits the efficiency in the training process and decreases the prediction accuracy. A new learning algorithm integrating EL and TL is proposed to mitigate the data scarcity problem and reduce necessary expenses for gathering additional samples or training a new ML model, all while increasing the robustness of the resulting prediction. The proposed algorithm does not require data preprocessing and is able to train a good ML model with a small training dataset. The application to deep RC beams without stirrups is illustrative of the wider impact of this research. As different mechanisms are responsible for the shear capacity of slender and deep beams, historically, these capacities have also been calculated via different, disparate processes. With the recent advancement of ML, and, in specific, the emphasis of TL in this work, we can use one approach to predict the capacity of both beams. This could be very useful in the analysis of existing structures where the beams may not be apparent but instead, occluded by nonstructural components, and thus, it is not clear if the beams are slender or deep.

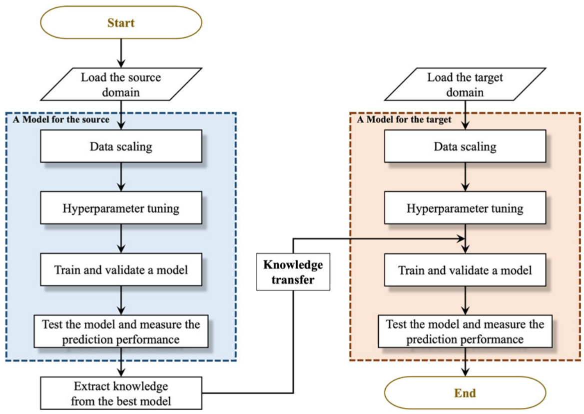


FIGURE 1 Schematic procedure of transferring knowledge in neural networks

3 | FUNDAMENTALS OF THE PROPOSED METHODOLOGY

3.1 | Knowledge transfer and fine-tuning

A domain, $\mathscr{D}$, consists of a feature space and its probability distribution, and a task, $\mathscr{T}$, is what the ML model learns from $\mathscr{D}$. TL aims to extract knowledge from the source domain, and then properly utilize the pretrained knowledge to more efficiently learn the target task with additional training processes on the target domain. Given the source domain, $\mathscr{D}^S$, and source learning task, $\mathscr{T}^S$, TL helps in understanding the target learning task, $\mathscr{T}^T$, trained on the target domain, $\mathscr{D}^T$, where $\mathscr{D}^S \neq \mathscr{D}^T$ and/or $\mathscr{T}^S \neq \mathscr{T}^T$. $\mathscr{D}^S$ can be represented as $\{(\mathbf{x}_i^S, y_i^S)_{i=1}^n\}$, where n is the number of samples in the source domain. Similarly, $\mathscr{D}^T$ can be represented as $\{(\mathbf{x}_i^T, y_i^T)_{i=1}^m\}$, where m is the number of samples in the target domain. Here, $\mathbf{x}_i \in \mathbb{R}^d$ is the ith explanatory variable vector, where d indicates the vector dimension, and y_i denotes the response variable, which could be a discrete variable for a classification problem or a continuous variable for a regression problem. In most cases, m is much smaller than n, and probability distributions of the source and target domains are different but somewhat related.

A powerful advantage of TL is the ability to learn a target task from not only the target domain but also the source domain. This advantage is most compelling when overfitting problems are encountered due to a limited number of training samples. To efficiently transfer the pretrained knowledge from a NN model trained on the source domain, the parameter-transfer approach was adopted in this study. The fine-tuning strategy is the most popular application of TL for a NN model. Figure 1 shows the schematic procedure for transferring knowledge from the source model to the target model. In Figure 1, the dotted line box on the left represents the process for training an ANN model on the source domain and the solid line box on the right represents the process for training an ANN model on the target domain. The same nomenclature is used later in Figure 3. The first step of the procedure is to develop an ANN model that can accurately predict the response variable(s) in the source domain. The second step is to extract the pretrained knowledge from the developed model and transfer it into another ANN model created for the target domain. In this process, the number of transferred layers is automatically determined by hyperparameter tuning, which will be introduced in Section 4.3, and thus varies based on the complexity of the problem. The ANN model for the target task has the same model structure for the layers containing the knowledge from the source model. In other words, the first n hidden layers of the source model are identical to those of the target model. Additional layers may or may not be added to the target model according to the results from hyperparameter tuning. The final step is to train the target model on the target domain. During this process, additional layers and an output layer are randomly initialized and trained from scratch, while the layers with the pretrained knowledge are fine-tuned or frozen.

3.2 | Ensemble learning

EL is the process of training multiple models for the same task, then merging the outputs to produce the desired response. The intuition behind EL lies in the concept of the "wisdom of the crowd." By gathering the opinions of a collection of individuals and taking a majority vote or an average across them, one can construct a more unbiased response than that formed by a single expert. In general, EL can improve the performance of a model by reducing overfitting as well as assisting to address the issue of class imbalance (Sagi & Rokach, 2018). This method is applicable for both regression and classification tasks.

Bootstrap aggregating, also called bagging, is an ML ensemble method to reduce model bias and variance as well as to avoid overfitting. Introduced by Breiman (1996),

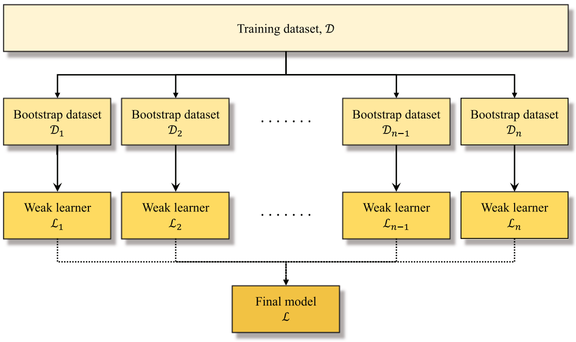


FIGURE 2 Overview of the Bagging algorithm used in this study

it is a method of generating subsets from a dataset for use in training a set of ensemble learners. After randomly excluding the test set from a dataset, suppose a given training dataset $\mathscr{D}$, as can be seen in Figure 2, with i number of samples. The bagging algorithm generates n different bootstrap datasets $\mathscr{D}_k$ for $k = 1, \ldots, n$, with i' number of samples. An integer value i' is determined by the sample bagging ratio. This ratio is a scalar value ranging from 0 to 1, which represents the percent of samples included in $\mathscr{D}_k$. As data samples in $\mathscr{D}_k$ are selected with replacement, some data samples in $\mathscr{D}$ could be selected more than once. The samples not selected in $\mathscr{D}_k$ can be used to construct an out-of-bag validation set. Each bootstrap dataset, $\mathscr{D}_k$, is then used to train a weak learner. Finally, the prediction results from each learner are aggregated to form the end prediction value.

Random subspace sampling, also referred to as feature bagging, describes the process of only utilizing a subset of attributes associated with the samples for each weak learner in an EL problem. Intuitively, this is a similar process to bagging applied to the explanatory variables instead of the samples themselves. The random subspace sampling method developed in this study follows the same process as the bagging method. Given the training dataset $\mathscr{D}$ with j number of explanatory variables (the dimension of explanatory variables vector is j), each of the bootstrap datasets, $\mathscr{D}_k$ for $k = 1, \ldots, n$ has j' number of explanatory variables by random selection with replacement. Thus, one explanatory variable could be selected more than once, and each bootstrap dataset $\mathscr{D}_k$ has a different combination of explanatory variables.

4 | TRANSFER ENSEMBLE NEURAL NETWORK

This study proposes a new neural network, called the transfer ensemble neural network (TENN), to properly take advantage of the synergetic effects of EL and TL. To effectively demonstrate the performance of the proposed methodology, three additional models are developed as baselines: a single ANN, a single ANN with knowledge transfer, and an ensemble ANN model. Section 4.1 describes the three baseline models, and Section 4.2 introduces the method newly proposed in this study. Each model is outlined in Figure 3.

4.1 | Baseline models

A general feedforward NN, denoted as ANN in Figure 3, is used to directly compare the model performances from the other models. This model is the simplest among those implemented in this study and is trained on the target domain without using EL or TL. The architecture of the model, including the number of hidden layers, the number of units in a layer, and the amount of regularization, is determined by the hyperparameter optimization process. The output of this model directly predicts the response variable without aggregating multiple ANN models. The second model, denoted as TLNN (transfer learning neural network) in Figure 3, is a single ANN designed to transfer knowledge from the source domain to the target domain. The ANN model is first trained on the source domain, and the pretrained knowledge is then transferred to a new model designed to learn the target task. The target model is then trained on the training set of the target domain in a similar fashion to the single ANN. Hyperparameter optimization is performed first on the source model to determine the core model architecture and then performed again on the target model to augment the core architecture. Thus, the target model may or may not have additional layers based on the results from the hyperparameter optimization process. As with the single ANN, the output provided by this model directly predicts the response variable. For the third model, denoted as ELNN (ensemble learning neural network) in Figure 3, both bagging and random subspace sampling are performed on the target domain to construct n different bootstrap datasets. To maximize the number of samples available for training, out-of-bag sampling was utilized to construct the validation set for each model. In this method, the samples which were not selected in $\mathscr{D}_k$ for $k = 1, \ldots, n$, are used as the validation set. Each of these datasets is then used to train n different ANN models. Hyperparameter tuning is performed a single time on the original training set and applied uniformly to each of the n weak learners. The ensemble algorithm developed in this study aggregates the models with good performance from the validation set. This means models with performance below a prespecified threshold are discarded before aggregation. The broken lines between $\mathscr{L}_k$ for $k = 1, \ldots, n$, and $\mathscr{L}$ in Figure 2 represent that each weak learner may or may not be included in the final model to predict the response variable. Thus,

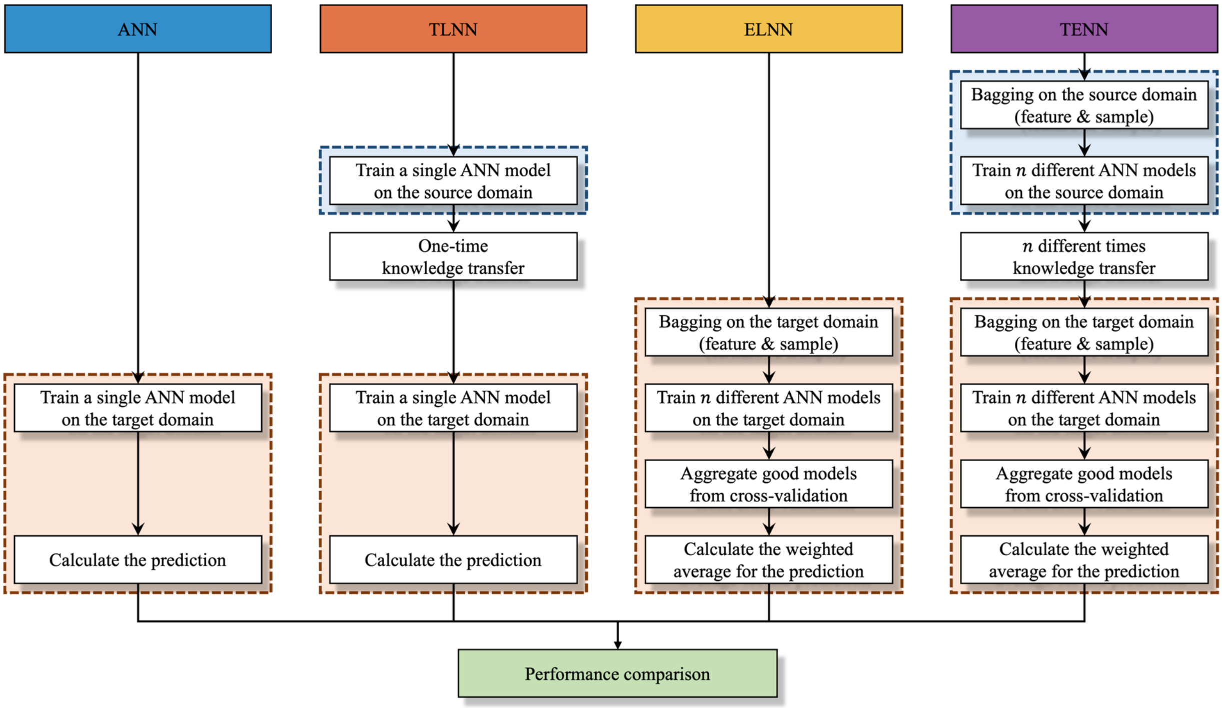


FIGURE 3 Overview of the models implemented and developed in this study

the number of aggregated models is equal to or less than n. Finally, the reciprocal of the validation loss from each weak learner is used as a weight value for that specific weak learner. The final prediction is calculated as the weighted average of each weak learner's prediction.

4.2 | Transfer ensemble neural network

The TENN model proposed in this study merges the training approaches described in Section 4.1. First, bagging and random subspace sampling are performed on the source domain to construct n bootstrap datasets. These datasets are used to train n different models. As with the models introduced in Section 4.1, hyperparameter tuning is performed before bagging and random subspace sampling. Then, the acquired knowledge is transferred to the n different target models. Hyperparameter tuning is again performed. Finally, bagging is applied to the target domain to construct n target training sets. To adequately utilize the pretrained knowledge from the n different source models, the randomly selected explanatory variables for the source domain are maintained to train the target models. The models are then trained on the target datasets as in the ELNN model. As before, models that exhibit performance below a prespecified threshold are discarded, and the remaining models are aggregated using a weighted average with the reciprocal of the validation loss supplying each learner's weight.

The TENN model is designed to mitigate the data scarcity problem and to address the high prediction variances associated with small datasets while maintaining robust model performance. Since bagging and random subspace sampling are both used to produce multiple bootstrap datasets, there is no need to include a separate procedure for choosing or extracting the appropriate features for the learning task. The final aggregated learner in the TENN model will automatically exclude a weak learner trained with irrelevant features because its performance will be worse than that of the other weak learners. Furthermore, thanks to the knowledge acquired from the source domain, individual weak learners of the TENN model can be initialized better than the baseline models introduced in Section 4.1. Therefore, such advantages from EL and TL integrated into the TENN model are expected to perform robustly and reduce the variances caused by small-size datasets. Especially when developing an ML model to predict the shear capacity of RC beams, deep RC beams are underrepresented in a database, since the majority of RC beams are slender. Thus, if the shear capacity of deep RC beams can be accurately predicted with limited training data, the TENN model's effectiveness can be highlighted.

4.3 | Hyperparameter tuning

The TENN model and three baseline models introduced in earlier sections are all based on an ANN algorithm.

TABLE 1 Possible values or ranges of hyperparameters

Training domain	Hyperparameter	ANN	TLNN	ELNN	TENN
Source	Number of layers	-	1, 2, 3	-	1, 2, 3
	Number of units	-	10, 30, 50, 100	-	10, 30, 50, 100
	Amount of regularize	-	[1e-4, 1e-1]	-	[1e-4, 1e-1]
Target	Number of transferred layers	-	1, 2, 3	-	1, 2, 3
	Transferred layer trainable	-	Yes, No	-	Yes, No
	Number of layers	1, 2, 3	0, 1, 2, 3	1, 2, 3	0, 1, 2, 3
	Number of units	10, 30, 50, 100	10, 30, 50, 100	10, 30, 50, 100	10, 30, 50, 100
	Amount of regularize	[1e-4, 1e-1]	[1e-4, 1e-1]	[1e-4, 1e-1]	[1e-4, 1e-1]
	Number of learners	-	-	10, 30, 50	10, 30, 50

Generally, an ANN model has one or more hidden layers consisting of a number of units, edges, and activation functions. An ANN with multiple sequential hidden layers is typically referred to as a DNN. In many engineering problems, the input and output variables are generally governed by the design requirements. On the other hand, as there is no rule for selecting the number of hidden layers and units in a layer, they must be appropriately selected by the user. In this study, before training and testing NN-based models, the Bayesian optimization process developed by Bergstra et al. (2013) has been implemented to find a reasonable combination of hyperparameters. This algorithm generally requires fewer iterations when compared to the random search or grid search methods, assuming the same search space. The algorithm developed in this study automatically chooses the optimal number of hidden layers and units in each layer within the given range of hyperparameters in each trial. Thus, every trial can have a unique model architecture with a different number of hidden layers and units. The possible values or ranges defined in the hyperparameter tuning process are presented in Table 1. The most prominent value is chosen within the possible values or ranges for the corresponding hyperparameter. For the TLNN and TENN models, the number of layers in the target domain denotes the number of added layers built upon layers transferred from the source model. The possible range of the amount of regularization is defined as the minimum and maximum values. More details of the dataset, model architectures, and hyperparameter tuning results are accessible through the NSF NHERI DesignSafe-CI portal.

4.4 | Model training procedures

Once a model architecture is determined based on the results of hyperparameter tuning, the TENN and three baseline models are trained by the procedures presented in Figure 3. In this study, 20 trials were performed for each model with the training and testing data being randomly selected each time. Seventy percent of the source domain is used to train the source model in the TLNN and TENN models, and the remaining 30% is used to test the trained model. To rigorously evaluate the TENN and three baseline models, the splitting ratio of the target domain varies depending on the experiment, and the details are explained in Section 5. During the training procedure, the Adam optimizer (Kingma & Ba, 2014) is implemented to minimize a loss function, specifically, the root mean square error (RMSE). To reduce the risk of overfitting problems and properly terminate the training procedure, 10% of the training set is defined as a validation set, and the early stopping method is adopted.

In this study, four performance metrics are utilized to measure the model performance and estimate the error. Each metric provides insights not provided by the other. The RMSE is monitored during the model training and validation process (Equation 4). Because of a square root, this metric is sensitive to larger differences between the predicted and actual values. Thus, when the RMSE is used during the training, the model becomes more robust to outliers. Nevertheless, the drawback is that this metric is dependent on the range of the response variable associated with the dataset. Therefore, they cannot be directly compared across different datasets. For this reason, this study does not report the RMSE as a metric for the model evaluation. It is only used in the training of the models as an optimization metric. The symmetric mean absolute percentage error (sMAPE) can measure a relative error in the predicted values (Equation 5). Contrary to MAPE, sMAPE is symmetric and has a lower and upper bound. The coefficient of determination (R^2) is a metric describing the model's capability to account for the variation in the response variable (Equation 6). R^2 is well suited for comparing multiple regression models across a single dataset and a single regression model across multiple datasets. The coefficient of variation (CV) is a measure of the dispersion of a set of values around their mean (Equation 7). Formally, it is the ratio of the standard deviation to the mean of the data samples. CV is utilized to measure the variation in model performance across multiple trials. This provides a means to assess a model's robustness to changes in the dataset. The following equations show the definitions of the RMSE, sMAPE, R^2, and CV:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{4}$$

$$\mathrm{sMAPE} = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{|y_i| + |\hat{y}_i|}, \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{6}$$

$$\mathrm{CV} = \frac{1}{\bar{x}}\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n-1}}, \tag{7}$$

where n is the number of samples, y_i is the true value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of y_i, s_x is the standard deviation, x_i is a random variable, and $\bar{x}$ is the mean of x_i.

The R^2 for each trial was recorded, and the average across 20 trials is presented. The CV of R^2 results across 20 trials is also calculated for each model and presented as a metric of model robustness. A higher CV value indicates that the performance of a model has a relatively greater level of dispersion in 20 trials. Therefore, by directly comparing the CV values from different training approaches, the effectiveness and robustness of the proposed approach can be demonstrated.

5 | EXPERIMENT AND RESULTS

5.1 | Dataset description

To investigate the feasibility of the proposed approach and compare it with other training approaches, an RC beam database (Collins et al., 2008) has been used in this study. The dataset consists of extensive shear experiments on RC beams of rectangular or T-section without stirrups. It has 1849 shear experiment specimens collected over 60 years, primarily composed of normal weight and non-prestressed RC beams. The dataset comprises two failure modes, 1696 of shear failure and 153 of flexural failure, and a wide range of dimensions and reinforcement ratios. By using this dataset, the usefulness of the TENN model in addressing a problem in a challenging domain, particularly the accurate prediction of the shear capacity of deep RC beams with a small training dataset, is illustrated. Sixteen explanatory variables are provided as inputs to the ANN and TLNN models, whereas a subset of randomly selected explanatory variables is provided to the ELNN and TENN models. Two variables among them are categorical: loading configuration and rebar cutoffs. The detailed information and statistics of the continuous explanatory variables are presented in Table 2. To precisely evaluate the performance of each model, training and test sets are mutually exclusive from each other. For each model, 20 trials have been conducted to adequately capture the generalization capabilities of the proposed approach.

5.2 | Validating the TENN model

Prior to applying the proposed algorithm to the prediction of the shear capacity with a small dataset, a validation study was performed. The purpose of this section is to validate the model's performance with regard to increased accuracy and decreased model variance as compared to the baseline models. A derived parameter, $M/(\rho_w V d)$, was used as the response variable. This parameter represents the coexisting moment at the critical shear section, which is a measure of the stress in the longitudinal reinforcement and significantly affects the shear capacity of RC beams (Collins et al., 2008).

$$\frac{M}{\rho_w V d} = \begin{cases} \dfrac{1}{\rho_w}\dfrac{a-d}{d} & \text{(for Type P)} \\ \dfrac{1}{\rho_w}\left(\dfrac{0.25L}{d} - 0.5\right) & \text{(for Type U1)} \\ \dfrac{1}{\rho_w}\dfrac{L/d-1}{L/d-2} & \text{(for Type U2)} \end{cases}, \tag{8}$$

where a is the shear span length, d is the effective depth, L is the span length, and ρ_w is the horizontal reinforcement ratio.

Although it is not an observed result from the experiments, the derived parameter is meaningful in verifying the proposed approach because it is calculated in varying manners depending on the loading configuration (Collins et al., 2008). The dataset used in this study contains three different loading configurations, as can be seen in Figure 4. Equation (8) shows three different mathematical expressions to calculate $M/(\rho_w V d)$. There are 1701 samples of Type P and 148 samples of Types U1 and U2, corresponding to less than 10% of the number of Type P samples. This presents a natural case for the application of a TL approach such as the TENN model. For the validation process, the

TABLE 2 Statistics of the continuous explanatory variables

Parameter	Description	Unit	Average	Standard deviation	Minimum	Maximum
b_w	Web width	mm	213.01	212.94	21	3000
b	Flange width	mm	256.76	230.21	21	3000
h	Depth	mm	364.35	254.05	51	3140
d	Effective depth	mm	320.25	237.77	41	3000
$a : M/V$	Distance from point of inflection to maximum moment location	mm	953.22	823.7	80	9000
a: STM	Distance from center of applied load to center of support	mm	1000.97	845.97	80	9000
a/d	Span-to-depth ratio	-	3.2	1.82	0.25	15.06
B	Length of bearing plate	mm	107.13	74.15	0	600
ρ_w	Horizontal reinforcement ratio	%	2.24	1.52	0.1	9.5
A_s	Area of steel reinforcement	mm^2	1280.43	1350.5	16.42	18450
f'_c	Concrete compressive strength	MPa	34.86	18.34	6.1	127.5
a_g	Maximum aggregate size	mm	18.48	6.96	1	50
f_y	Steel yield strength	MPa	462.37	172.14	267	1779
$M/\rho_w Vd$	Derived parameter	-	138.62	156.01	0	2768.3

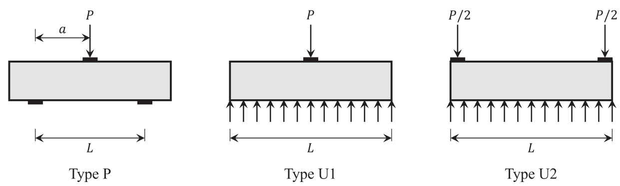
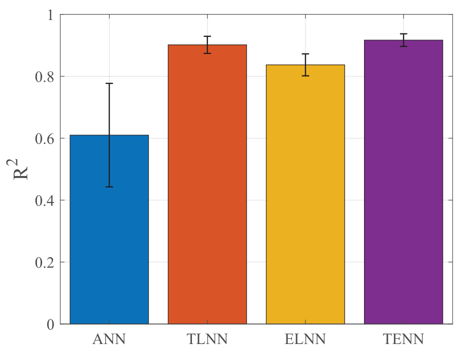


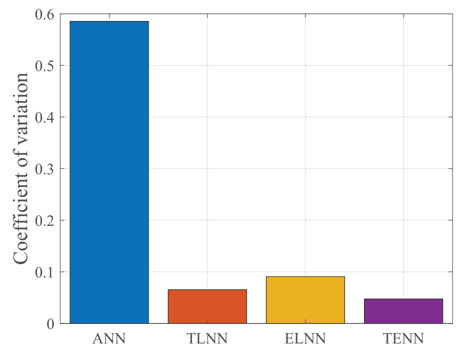
FIGURE 4 Types of loading configurations present in the dataset

data samples with Type P loading configuration are used as the source domain, and the pretrained knowledge is transferred to provide more accurate predictions of Types U1 and U2 loading configurations. Seventy percent of data samples in each domain is used as the training set, and the remaining 30% is used as the testing set. Thus, given that the number of samples in the target domain is not enough to train a good ML model, the proposed approach can be verified if it shows higher accuracy with lower variance.

Following the procedures of the four different approaches presented in Figure 3, 20 trials were implemented to verify the efficacy of combining EL and TL. The performance of the four different models is presented in Figure 5 in terms of their prediction ability and stability. Figure 5a shows the average R^2 values within the 20 trials with 95% confidence interval bounds. The variation of the model performance is shown in Figure 5b, represented by the CV values across the 20 trials. The ANN model has the lowest performance in both accuracy and variance. These results are expected, as the number of samples in the target domain, Types U1 and U2 loading configuration, is insufficient to train a robust ML model. The results



(a) The coefficient of determination (R^2) with 95% confidence interval



(b) The coefficient of variation (CV)

FIGURE 5 Performance comparison of four different models to predict $M/\rho_w Vd$

from the ELNN model demonstrate the effect of EL in a small-size dataset. Although the results from the ELNN model are not as good as those from the TLNN or TENN models, an increased ability to predict the response variable over the traditional ANN model can be observed.

TABLE 3 Available data in $\mathscr{D}^S$ and $\mathscr{D}^T$ in different situations

Situation	$\mathscr{D}^S$	$\mathscr{D}^T$ (train)	$\mathscr{D}^T$ (test)
Mild	1244	302	303
Moderate	1244	151	454
Severe	1244	90	515

The highest predictive performance and its robustness can be observed in the novel TENN model. It achieved an R^2 value of 0.9168 and presented the lowest CV across the four models (0.0424). This means that the model has the best prediction ability with the lowest variance. The TLNN model also shows good performance with low variance, though less effective than the TENN model. The proposed TENN model exhibited a CV of more than 14 times less than that of the traditional ANN model. These results validate the proposed TENN model and demonstrate its usefulness in addressing the data scarcity problem.

5.3 | Predicting the shear capacity of deep RC beams

To comprehensively investigate the feasibility of the proposed methodology, 1244 samples of slender beams, which have a/d values greater than or equal to 2.5, are used as the source domain, $\mathscr{D}^S$. The remaining 605 samples of deep beams are used as the target domain, $\mathscr{D}^T$. $\mathscr{D}^S$ and $\mathscr{D}^T$ are physically correlated since they consist of similar specimens with similar cross sections and the same material. Furthermore, $\mathscr{T}^S$ and $\mathscr{T}^T$ are also similar since eventually the tasks in both domains are to accurately predict the shear capacity of RC beams. Thus, transferring the acquired knowledge on $\mathscr{D}^S$ is expected to show better performance on $\mathscr{D}^T$. As introduced in the earlier section, the ANN and ELNN models are trained only on $\mathscr{D}^T$. The TLNN and TENN models adopt the pretrained knowledge from $\mathscr{D}^S$ to better predict the response variable in $\mathscr{D}^T$. The main purpose of the proposed novel methodology is to maintain high model performance and low variance even when the size of the dataset is small. Thus, three different situations are designed to compare the model performance under varying levels of data scarcity. Table 3 summarizes the three data availability scenarios in this study. In the *mild* scenario, the training set for the target domain consists of 50% of the deep beam samples. In this case, all four approaches are expected to have an acceptable ability to predict the shear capacity of the deep beams. In the *moderate* scenario, 25% of the target domain is used for training. Finally, in the *severe* scenario, only 15% of the target samples are used in training. Obtaining a good NN model with this little amount of training data is almost impossible. The samples not selected for use in the training set are used as the corresponding test set; thus, the training and test set are mutually exclusive such that model performances can be accurately estimated and compared with one another. Regardless of the situation, all 1244 samples in the source domain are utilized to extract meaningful knowledge that will be transferred to the target model. The four approaches introduced in the earlier sections are implemented with 20 trials for each scenario.

The results of the four different models are summarized in Table 4. As expected, the models in the *mild* condition show good test results regardless of the approaches used for training. It is observed that every model yields average R^2 values of approximately 0.9 with a small 95% confidence interval range. Although the ANN model has the highest CV value compared to those of the other three models in the *mild* condition, all four models can be treated as robust and stable ML models. However, these favorable outcomes can hardly be found in the *moderate* and *severe* situations, since the available training data samples correspondingly decrease. Under the *moderate* situation, the ANN and ELNN models are trained on 151 samples of deep beams, whereas the TLNN and TENN models are trained on 151 samples of deep beams after they were assisted by the pretrained knowledge from 1244 samples of slender beams. The ANN model's average performance falls below 0.8, and it has a CV value an order of magnitude larger than seen in the *mild* case. Among the four models trained in the *moderate* condition, the ANN model has the lowest R^2 value and largest confidence interval. Notably, its confidence interval is much larger than that of the other three models. Although the average R^2 values of the other three models decreased to approximately 0.8, their predictions of the shear capacity of deep RC beams remain relatively accurate. Moreover, the ranges of their confidence intervals are well controlled compared to that of the ANN model. Similar results can also be found in the *moderate* situation. The CV value of the ANN model is roughly 10 times higher than that of the other three models. The TENN model proposed in this work achieves lower CV values than both the TLNN and ELNN models. Based on these results, the effectiveness of the novel approach is evident in cases where only moderate amounts of data are available. The worst data availability scenario in this study is the *severe* situation. Only 90 data samples are available to train a model and predict the shear capacity of deep beams. Generally, it is difficult to obtain an adequate NN model without additional techniques under these circumstances. Accordingly, the ANN model experiences the worst prediction performance. The ANN model achieved an average R^2 value of approximately 0.5. By adopting EL, the predictive performance and model variance can be enhanced to some extent. In the *severe* situation, TL is more effective at

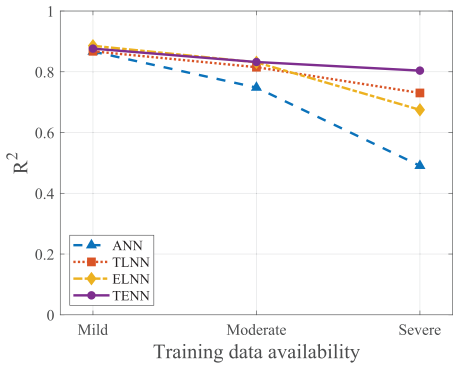
TABLE 4 Result summary in three different data availabilities

Model	*Mild* R^2	CI	CV	*Moderate* R^2	CI	CV	*Severe* R^2	CI	CV
ANN	0.87	[0.854, 0.880]	0.032	0.75	[0.630, 0.866]	0.337	0.49	[0.235, 0.746]	1.114
TLNN	0.87	[0.861, 0.874]	0.015	0.80	[0.804, 0.826]	0.029	0.73	[0.704, 0.757]	0.077
ELNN	0.89	[0.877, 0.894]	0.020	0.83	[0.813, 0.849]	0.046	0.67	[0.511, 0.839]	0.520
TENN	0.88	[0.869, 0.883]	0.017	0.83	[0.824, 0.840]	0.021	0.81	[0.793, 0.815]	0.030

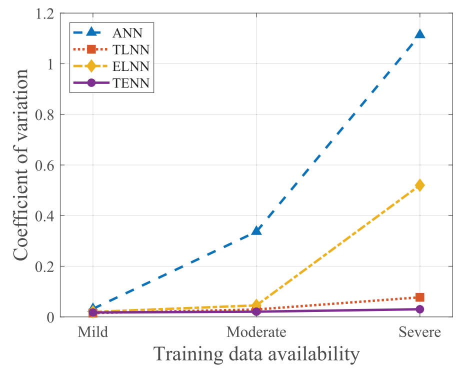
C.I. denotes the confidence interval.

remedying the data scarcity problem than EL. This is evidenced by the superior model performance of the TLNN model compared to that of the ELNN model. In comparison, the approach proposed in this work saw little change in prediction ability between the *moderate* and *severe* scenarios. By combining EL and TL in the novel proposed approach, a 70% increase in predictive performance can be observed with a much smaller range of confidence interval as compared to a traditional ANN model. The TENN model shows a remarkable ability to predict the shear capacity and robustness when trained on only 90 samples. According to the results presented in Table 4, the TENN model outperforms the other three models in terms of R^2 and model variance.

As training data availability decreases, it is generally accepted that model performance will deteriorate due to the lack of available information from the data. The results reported in this study are in accordance with this generally accepted fact. Figure 6 shows the variation of the performance metrics in *mild*, *moderate*, and *severe* data scarcity conditions. As the situations progress from *mild* to *severe*, R^2 values from all four models decrease without exception. However, the proposed model proves extremely effective at mitigating this deterioration. As seen in Figure 6, every model considered in this study is capable of predicting the shear capacity of deep beams with low model variance in the *mild* situation. The most drastic decline of model performance is found in the ANN model: its R^2 value falls to about 0.49 from around 0.87. This is because the baseline ANN model is sensitive to the number of available training samples. If the ANN model is trained in the *moderate* or *severe* training data scarcity conditions, it is likely to obtain an inadequate model. This drastic drop of R^2 value and surge of CV value can be mitigated by adopting EL, TL, or both. When the training data availability is changed from *mild* to *moderate*, the performances of the TLNN, ELNN, and TENN models marginally decrease. The performance deterioration from the three models is significantly less than that of the ANN model. This evidence supports the idea that EL and TL make the ML model more accurate and stable in the face of data scarcity problems. In the *severe* situation, it is impossible to get a good ML model through the use of EL or TL alone. It is



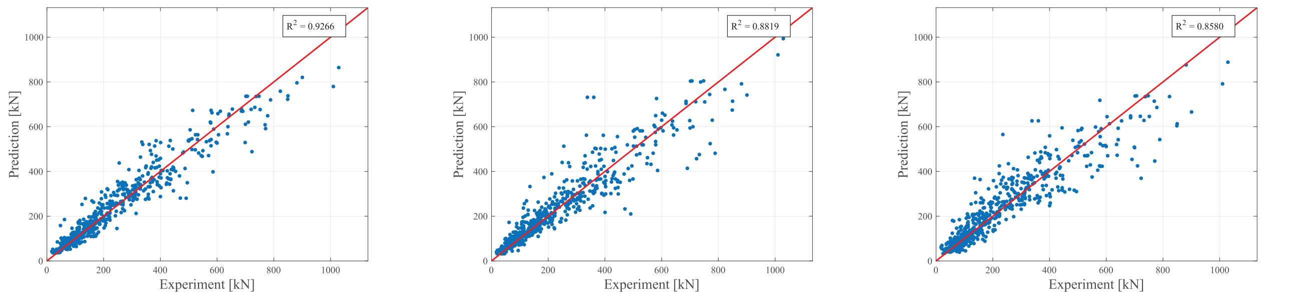
(a) The coefficient of determination (R^2) with 95% confidence interval



(b) The coefficient of variation (CV)

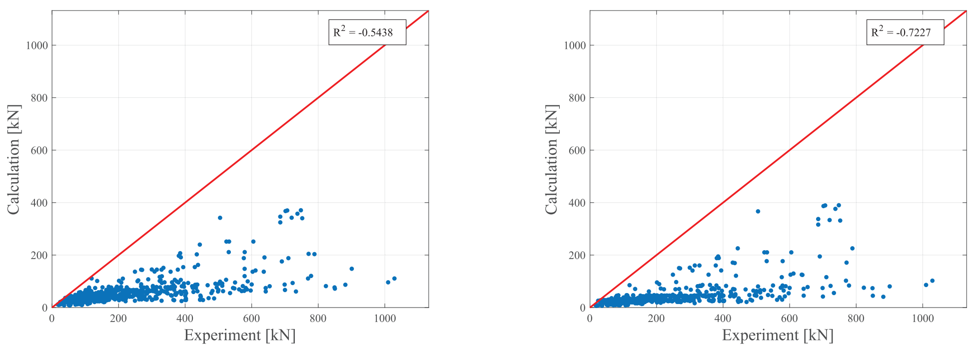
FIGURE 6 Variation of the performance metrics in different training data availability

observed that the ANN model has the lowest R^2 and highest CV values, and the ELNN model has a high CV which cannot be treated as a stable model. If TL is applied to the baseline ANN model, a slightly better model can be obtained although the performance deterioration remains unacceptable. The most impressive results are found in the novel TENN model proposed in this study. Even when subjected to a very limited number of training samples, the TENN model not only blocks the reduction of R^2 due to the small data but also maintains a low model variance comparable to that of the *mild* or *moderate* scenarios. As can be seen in Figure 6, the TENN model preserves its high R^2 and low CV values across all data scarcity levels. The knowledge acquired from the slender RC beams enables the TENN model to accurately predict the shear



(a) *Mild* condition (b) *Moderate* condition (c) *Severe* condition

FIGURE 7 Comparison of the experimental and predicted results from the TENN models in different situations



(a) ACI 318-19 (b) CSA

FIGURE 8 Comparison of the experimental and calculated results based on the design standards

capacity of deep RC beams, and simultaneously, multiple weak learners can significantly reduce the model variance. Consequently, by incorporating the advantages from EL and TL, the TENN model can be initialized in a superior manner compared to the baseline models, and, accordingly, provide a better estimation from the aggregated learners.

5.4 | Comparison with the existing standards

The shear capacity of all deep beams in the dataset is also calculated using the provisions of ACI 318-19 and CSA and the calculated values are compared to the results predicted by the best TENN model out of 20 trials. The purpose of comparisons in this section is to demonstrate the accuracy of the TENN model in terms of properly estimating the shear capacity of deep RC beams. Figure 7 shows the comparison of the experimental and predicted results from the TENN models trained in the different data scarcity conditions. The *mild* condition, Figure 7a, shows the best agreement between the experimental and predicted results with an R^2 value of 0.93. As the data availability condition degrades, the performance of the TENN model slightly decreases due to the nature of ML approaches. However, it is noteworthy that even in the *severe* case, the model still maintains its performance with an R^2 value of 0.86. Moreover, the TENN models provide more stable predictions, since they have lower CV values than ACI 318-19 and CSA. The comparisons of the experimental and calculated results based on ACI 318-19 and CSA are depicted in Figure 8. More details of the model evaluation metrics are summarized in Table 5. Because the R^2 value from ACI 318-19 is higher than that from CSA, it is believed ACI 318-19 yields marginally more accurate estimations for the shear capacity of deep beams. Notably, there are considerable differences between the experimental results and the calculated values from those two existing standards. Most of the calculated values based on the existing standards are much lower than the experimental results. For the purpose of precisely estimating the shear capacity of deep beams, the TENN model outperforms the existing standards even in the *severe* data availability condition by 82.56% and 102.51% sMAPE for ACI and CSA, respectively. Load factors and resistance factors were not considered when the shear capacity values were calculated based on the standards. If those factors were included, the calculated values

TABLE 5 The performance of the TENN models and the design standards

Model	Mean of V_{cal}/V_{exp}	CV of V_{cal}/V_{exp}	R^2	sMAPE (%)
TENN (*Mild*)	1.08	0.25	0.93	16.02
TENN (*Moderate*)	1.10	0.27	0.88	18.63
TENN (*Severe*)	1.11	0.34	0.86	21.72
ACI 318	0.34	0.52	−0.54	104.28
CSA	0.25	0.61	−0.72	124.23

would further shift down compared to those depicted in Figure 8. This is understandable as sufficient amounts of safety against service or structural limit states should be ensured. Nevertheless, the ACI and CSA equations may be less effective methods for accurately estimating the shear capacity of deep beams. The proposed TENN model provides significantly better performance than the existing standards, even with very limited amounts of training data.

6 | CONCLUSIONS

This study has presented a new ML algorithm, the TENN, which can accurately predict the shear capacity of deep beams without stirrups, when a limited amount of data is available. The proposed algorithm integrates TL and EL after randomly sampling subtraining datasets based on bagging and random space sampling. The algorithm is tested on an extensive RC beam database. The shear capacity of deep beams without stirrups is estimated in three data availability conditions and compared with the values calculated by ACI 318-19 and CSA. In accordance with the results, the following conclusions are drawn:

1. The TENN model shows outstanding performance in terms of accuracy of variance, especially in the *severe* data availability condition. The TENN model has the highest R^2 value with the smallest confidence interval range and the lowest model variance.
2. When it comes to accurate estimation of the shear capacity, the TENN model not only outperforms ACI 318-19 and CSA but it also provides an efficient method for estimating the shear capacity of RC beams without stirrups.
3. By incorporating a knowledge transfer technique and aggregating multiple neural network models, the TENN model can mitigate the data scarcity problem and provide better predictions with low variance with a very limited number of training data.
4. This work demonstrates the overall ability of the TENN model to accurately predict structural response when there is little data available, as long as a large dataset can be identified to extract the common knowledge.

Expanding the proposed algorithm on RC beams with stirrups and comparing the TENN model with other models such as the dynamic EL algorithm should be a focus of future research. In addition, further research specific to structural engineering should be conducted to develop more interpretable models, to augment data with physics-based models or simulations, and to evaluate the extrapolative knowledge transfer capabilities associated with the dataset.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under grant CMMI # 1944301.

REFERENCES

Abdalla, J. A., Elsanosi, A., & Abdelwahab, A. (2007). Modeling and simulation of shear resistance of R/C beams using artificialneural network. *Journal of the Franklin Institute*, *344*(5), 741–756.

ACI Committee 318 (2020). Building code requirements for structural concrete (ACI 318-19): An aci standard; commentary on building code requirements for structural concrete (ACI-318R-19). American Concrete Institute.

Adeli, H. (2001). Neural networks in civil engineering: 1989–2000. *Computer-Aided Civil and Infrastructure Engineering*, *16*(2), 126–142.

Adeli, H., & Panakkat, A. (2009). A probabilistic neural network for earthquake magnitude prediction. *Neural Networks*, *22*(7), 1018–1024.

Azimi, M., & Pekcan, G. (2020). Structural health monitoring using extremely compressed data through deep learning. *Computer-Aided Civil and Infrastructure Engineering*, *35*(6), 597–614.

Bazant, Z. P., & Sun, H.-H. (1987). Size effect in diagonal shear failure: influence of aggregate size and stirrups. *ACI Materials Journal*, *84*(4), 259–272.

Bentz, E. C. (2005). Empirical modeling of reinforced concrete shear strength size effect for members without stirrups. *ACI Structural Journal*, *102*(2), 232.

Bergstra, J., Yamins, D., & Cox, D. (2013). Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In *International conference on machine learning* (pp. 115–123). PMLR. MIT Press.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123–140.

Butcher, J. B., Day, C., Austin, J., Haycock, P., Verstraeten, D., & Schrauwen, B. (2014). Defect detection in reinforced concrete using random neural architectures. *Computer-Aided Civil and Infrastructure Engineering*, *29*(3), 191–207.

Canadian Standards Association (CSA) (2019). Design of concrete structures. CSA standard A23.3-19. CSA Group.

Cha, Y.-J., Choi, W., & Büyüköztürk, O. (2017). Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering*, *32*(5), 361–378.

Chen, S.-Z., & Feng, D.-C. (2022). Multifidelity approach for data-driven prediction models of structural behaviors with limited data. *Computer-Aided Civil and Infrastructure Engineering*, *37*, 1566–1581.

Chen, S.-Z., Feng, D.-C., Han, W.-S., & Wu, G. (2021). Development of data-driven prediction model for CFRP-steel bond strength by implementing ensemble learning algorithms. *Construction and Building Materials*, *303*, 124470.

Chen, S.-Z., Zhang, S.-Y., Han, W.-S., & Wu, G. (2021). Ensemble learning based approach for FRP-concrete bond strength prediction. *Construction and Building Materials*, *302*, 124230.

Cheng, C.-S., Behzadan, A. H., & Noshadravan, A. (2021). Deep learning for post-hurricane aerial damage assessment of buildings. *Computer-Aided Civil and Infrastructure Engineering*, *36*(6), 695–710.

Chetchotisak, P., Teerawong, J., Yindeesuk, S., & Song, J. (2014). New strut-and-tie-models for shear strength prediction and design of RC deep beams. *Computers and Concrete*, *14*(1), 19–40.

Cladera, A., & Marí, A. (2004a). Shear design procedure for reinforced normal and high-strength concrete beams using artificial neural networks. Part I: Beams without stirrups. *Engineering Structures*, *26*(7), 917–926.

Cladera, A., & Marí, A. (2004b). Shear design procedure for reinforced normal and high-strength concrete beams using artificial neural networks. Part II: Beams with stirrups. *Engineering Structures*, *26*(7), 927–936.

Collins, M. P., Bentz, E. C., & Sherwood, E. G. (2008). Where is shear reinforcement required? Review of research results and design procedures. *Structural Journal*, *105*(5), 590–600.

Feng, D.-C., Liu, Z.-T., Wang, X.-D., Jiang, Z.-M., & Liang, S.-X. (2020). Failure mode classification and bearing capacity prediction for reinforced concrete columns based on ensemble machine learning algorithm. *Advanced Engineering Informatics*, *45*, 101126.

Feng, D.-C., Wang, W.-J., Mangalathu, S., Hu, G., & Wu, T. (2021). Implementing ensemble learning methods to predict the shear strength of RC deep beams with/without web reinforcements. *Engineering Structures*, *235*, 111979.

Gao, Y., & Mosalam, K. M. (2018). Deep transfer learning for image-based structural damage recognition. *Computer-Aided Civil and Infrastructure Engineering*, *33*(9), 748–768.

German, S., Brilakis, I., & DesRoches, R. (2012). Rapid entropy-based detection and properties measurement of concrete spalling with machine vision for post-earthquake safety assessments. *Advanced Engineering Informatics*, *26*(4), 846–858.

Hu, B., & Wu, Y.-F. (2018). Effect of shear span-to-depth ratio on shear strength components of RC beams. *Engineering Structures*, *168*, 770–783.

Huang, D., Hu, D., He, J., & Xiong, Y. (2018). Structure damage detection based on ensemble learning. In *2018 9th International Conference on Mechanical and Aerospace Engineering (ICMAE)* (pp. 219–224). IEEE.

Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization*. arXiv. https://doi.org/10.48550/arXiv:1412.6980

Koo, S., Shin, D., & Kim, C. (2021). Application of principal component analysis approach to predict shear strength of reinforced concrete beams with stirrups. *Materials*, *14*(13), 3471.

Leach, S., Xue, Y., Sridhar, R., Paal, S., Wang, Z., & Murphy, R. (2021). Data augmentation for improving deep learning models in building inspections or postdisaster evaluation. *Journal of Performance of Constructed Facilities*, *35*(4), 04021029.

Li, S., Zhao, X., & Zhou, G. (2019). Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network. *Computer-Aided Civil and Infrastructure Engineering*, *34*(7), 616–634.

Liu, X., Liu, Z., Wang, G., Cai, Z., & Zhang, H. (2017). Ensemble transfer learning algorithm. *IEEE Access*, *6*, 2389–2396.

Lubell, A., Sherwood, T., Bentz, E., & Collins, M. (2004). Safe shear design of large wide beams. *Concrete International*, *26*(1), 66–78.

Luo, H., & Paal, S. G. (2021). Reducing the effect of sample bias for small data sets with double-weighted support vector transfer regression. *Computer-Aided Civil and Infrastructure Engineering*, *36*(3), 248–263.

Mansour, M. Y., Dicleli, M., Lee, J.-Y., & Zhang, J. (2004). Predicting the shear strength of reinforced concrete beams using artificial neural networks. *Engineering Structures*, *26*(6), 781–799.

Mohammed, H. R. M., & Ismail, S. (2021). Random forest versus support vector machine models' applicability for predicting beam shear strength. *Complexity*, *2021*, 9978409.

Muttoni, A., & Fernández Ruiz, M. (2008). Shear strength of members without transverse reinforcement as function of critical shear crack width. *ACI Structural Journal*, *105*, 163–172.

Oreta, A. W. C. (2004). Simulating size effect on shear strength of rc beams without stirrups using neural networks. *Engineering Structures*, *26*(5), 681–691.

Pak, H., & Paal, S. G. (2022). Evaluation of transfer learning models for predicting the lateral strength of reinforced concrete columns. *Engineering Structures*, *266*, 114579.

Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359.

Rafiei, M. H., Khushefati, W. H., Demirboga, R., & Adeli, H. (2016). Neural network, machine learning, and evolutionary approaches for concrete material characterization. *ACI Materials Journal*, *113*(6), 781–789.

Rafiei, M. H., Khushefati, W. H., Demirboga, R., & Adeli, H. (2017a). Novel approach for concrete mixture design using neural dynamics model and virtual lab concept. *ACI Materials Journal*, *114*(1), 117–127.

Rafiei, M. H., Khushefati, W. H., Demirboga, R., & Adeli, H. (2017b). Supervised deep restricted Boltzmann machine for estimation of concrete. *ACI Materials Journal*, *114*(2), 237.

Reineck, K.-H., Bentz, E., Fitik, B., Kuchma, D. A., & Bayrak, O. (2014). ACI-DAfStb databases for shear tests on slender reinforced concrete beams with stirrups. *ACI Structural Journal*, *111*(5), 1147–1156.

Reineck, K.-H., Bentz, E. C., Fitik, B., Kuchma, D. A., & Bayrak, O. (2013). ACI-DAfStb database of shear tests on slender reinforced concrete beams without stirrups. *ACI Structural Journal*, *110*(5), 867–876.

Reineck, K.-H., & Todisco, L. (2014). Database of shear tests for non-slender reinforced concrete beams without stirrups. *ACI Structural Journal*, *111*(6), 1363.

Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1249.

Sanad, A., & Saka, M. (2001). Prediction of ultimate shear strength of reinforced-concrete deep beams using neural networks. *Journal of Structural Engineering*, *127*(7), 818–828.

Schlaich, J., Schäfer, K., & Jennewein, M. (1987). Toward a consistent design of structural concrete. *PCI Journal*, *32*(3), 74–150.

Vecchio, F. J., & Collins, M. P. (1986). The modified compression-field theory for reinforced concrete elements subjected to shear. *ACI Journal*, *83*(2), 219–231.

Yang, K.-H., Ashour, A. F., & Song, J.-K. (2007). Shear capacity of reinforced concrete beams using neural network. *International Journal of Concrete Structures and Materials*, *1*(1), 63–73.

Zhang, J.-H., Li, S.-S., Xie, W., & Guo, Y.-D. (2020). Experimental study on shear capacity of high strength reinforcement concrete deep beams with small shear span–depth ratio. *Materials*, *13*(5), 1218.

Zhang, T., Visintin, P., & Oehlers, D. J. (2016). Shear strength of RC beams without web reinforcement. *Australian Journal of Structural Engineering*, *17*(1), 87–96.

Zhu, Y., Brettin, T., Evrard, Y. A., Partin, A., Xia, F., Shukla, M., Yoo, H., Doroshow, J. H., & Stevens, R. L. (2020). Ensemble transfer learning for the prediction of anti-cancer drug response. *Scientific Reports*, *10*(1), 1–11.

How to cite this article: Pak, H., Leach, S., Yoon, S. H., & Paal, S. G. (2023). A knowledge transfer enhanced ensemble approach to predict the shear capacity of reinforced concrete deep beams without stirrups. *Computer-Aided Civil and Infrastructure Engineering*, 1–16. https://doi.org/10.1111/mice.12965