Progressive artificial neural network model for CBR forecasts with minimum train spans

Liberty U. STEPHEN ▪ Department of Civil Engineering,
University of Agriculture & Environmental Sciences, Nigeria ▪ liberty.stephen@uaes.edu.ng
Michael E. ONYIA ▪ Department of Civil Engineering,
University of Nigeria, Nsukka, Nigeria ▪ Michael.onyia@unn.edu.ng
Fidelis O. OKAFOR ▪ Department of Civil Engineering,
University of Nigeria, Nsukka, Nigeria ▪ fidelis.okafor@unn.edu.ng

Érkezett: 2025. 07. 24. ▪ Received: 24. 07. 2025. ▪ https://doi.org/10.14382/epitoanyag-jsbcm.2025.10

Abstract

This study introduces a novel hybrid machine learning framework for predicting the California Bearing Ratio (CBR) in soil stabilization applications by integrating Neuronal Auditory Machine Intelligence (NeuroAMI), Particle Swarm Optimization with Differential Evolution (PSO-DE), and ensemble artificial neural networks (ANNs). The NeuroAMI component, inspired by the Mismatch Negativity (MMN) effect observed in mammalian auditory cortex processing, implements change detection (CD) and model adjustment (MA) through cochlear-inspired multi-scale frequency decomposition and predictive coding mechanisms.

The framework was evaluated on soil stabilization datasets incorporating Rice Husk Ash (RHA), Fines Content (FLD), Optimum Moisture Content (OMC), and Maximum Dry Density (MDD) as input parameters for predicting both unsoaked and soaked CBR values. Data preprocessing included Isolation Forest outlier removal, interaction- and ratio-based feature engineering, and bootstrap data augmentation. An ensemble of ten NeuroAMI models underwent PSO-DE coefficient optimization (40 particles, 120 iterations, 10 multi-start runs). Simultaneously, three ANN architectures—Feedforward, Deep, and Residual—were trained using 10-fold cross-validation, producing an ensemble of 30 models.

Results revealed distinct performance disparities. For CBR unsoaked, the ANN ensemble achieved R^2 = 0.866, while NeuroAMI attained R^2 = 0.778 on the test set but degraded to R^2 = 0.523 on the holdout prediction set. For CBR soaked, the ANN ensemble maintained high accuracy (R^2 = 0.882), whereas NeuroAMI exhibited severe performance degradation (R^2 = -2.667 on the test set, -1.169 on the prediction set). Feature importance analysis identified MDD (20 %) and interaction terms (37 % combined) as dominant drivers for CBR unsoaked, while OMC was the most influential factor for CBR soaked (40 %). Sensitivity analysis confirmed MDD as the most influential predictor, producing output variations up to 80 for soaked conditions.

Cross-validation indicated substantial variability across folds (R^2 = 0.55–0.95), reflecting dataset heterogeneity. Overall, results demonstrate that auditory cortex-inspired architectures are fundamentally unsuitable for static soil mechanics prediction, whereas conventional deep ensemble ANNs provide reliable and robust performance (R^2 > 0.86) across both soaked and unsoaked conditions. These findings emphasize the need to align bio-inspired computational paradigms with domain-specific problem characteristics and establish ensemble deep learning as the preferred methodology for CBR prediction in geotechnical engineering.

Keywords: soil stabilization, California Bearing Ratio, NeuroAMI, mismatch negativity, ensemble learning, artificial neural networks, feature engineering, rice husk ash, PSO-DE optimization
Kulcsszavak: talajstabilizálás, California Bearing Ratio (CBR-érték), NeuroAMI, mismatch negativity (MMN), ensemble learning (együttes tanulás), mesterséges neurális hálózatok, feature engineering (jellemzőkinyerés/-tervezés), rizshéjhamu (RHA), PSO-DE optimalizáció

Liberty U. STEPHEN
is an assistant lecturer in Civil Engineering department, University of Agriculture and Environmental Sciences, Owerri, Imo State, Nigeria. Her major area of interest is in Geotechnics. She is also a registered Engineer with COREN.

Michael E. ONYIA
is a lecturer and professor of Civil Engineering at the department of Civil engineering, University of Nigeria, Nsukka, Enugu State-Nigeria. His skills and expertise include, structural analysis, structural dynamics, geotechnical engineering.

Fidelis O. OKAFOR
is a lecturer and professor of Civil Engineering at the department of Civil Engineering, University of Nigeria, Nsukka, Enugu State-Nigeria. His major area of interest is Materials and Highway Engineering.

1. Introduction

1.1 Background and motivation

California Bearing Ratio (CBR) is a very important soil testing metric that serves as a standard method for evaluating the quality of soil used in a variety of building structures. Such applications such as the design of road structures and pavements have been greatly enhanced by the use of the CBR metric [1, 2]. Determining the CBR requires several time consuming and laborious compaction tests and this has prompted a lot of researchers and structural engineers alike to investigate the prospects of alternative solutions one of which borders on the use of computational intelligence algorithms such as neural networks, genetic programming, genetic algorithms etc. [3]. It has been identified in a number of related research studies that the process for determining the CBR is a laborious and time-consuming task particularly with respect to the gathering and analysis of large number of soil samples [4]. In particular, measurable soil properties such as percentage rice husk ash, geotextile fabric layer distance, optimum moisture content

and maximum dry density are usually compiled and analysed manually or using low quality computing tools leading to significant errors in reporting [5].

Recent studies have shown the potentials of using computational intelligence techniques with Machine Learning (ML) and/or Artificial Intelligence (AI) basis [6, 7]. Hence, CBR can be more accurately and reliably estimated considering these methods. Nevertheless, the use of AI methods such as Artificial Neural Network (ANN) though promising can be challenged by the temporal nature of the soil variables and the limited amount of data for predictive analysis. In this regard, it becomes desirous to develop alternative ANN models that can solve these challenges.

In this paper, we present a progressive ANN with continual learning capability based on the Neuronal Auditory Machine Intelligence (NeuroAMI) approach for predicting CBR using limited training datasets. We perform simulations considering various training-testing data scenarios with lower training level bounds to determine the performance of the proposed ANN method.

1.2 Advances in machine learning for geotechnical engineering

Soil testing presents an active area of research due to its considerable impact on the design and building of physical structures. In the reviewed studies, the trend revolves around the use of AI tools and techniques to conduct predict or regression-like fitting simulations for the estimation of the CBR. For instance, Rassoul, & Mojtaba [8] proposed to use an Ordinary Least Squares (OLS) algorithm and an ANN for the prediction of CBR for soaked and unsoaked soils. Their findings showed that the ANN method with the 80% training input will outperform the OLS with peak R^2 of around 0.9 and 0.5 respectively. A combination of regression and neural methods has been investigated in [9] for CBR prediction of unbound granular materials and subgrade soils. They employed the 80%-20% training-testing data split for their model performance analysis.

Salahudeen & Sadeeq [10] proposed the use of ANN for the prediction of modified black clay soaked and unsoaked CBR. The authors reported high R^2 of around 0.99 for both soaked and unsoaked soils.

Sujatha et al [11] used MLP based ANN modelling for the prediction of CBR for silly soaked soils with reported values of R^2 at 0.95 and 0.94 for training and testing sets respectively. The authors reported excellent coefficient of determination (R^2) values for the studied case.

Kurnaz & Kaya [12] compared Group Method of Data Handling (GMDH) type neural networks with the basic ANN and Multiple Linear Regressors (MLR) methods for the task of predicting CBR of compacted soils. Following extensive hyperparameter tuning experiments, the authors report best performance achievable by GMDH at a depth of 7-layers and a maximum of 10 neurons per layer. The reported metrics, R^2, MSE and RMSE were found to be 0.9783, 1.69 and 1.30 respectively with the GMDH outperforming the basic ANN and MLR methods.

The prediction of subbase CBR values using an ANN approach has been investigated in [13] where a 2-step clustering procedure is provided as an initial pre-processor for the ANN. The authors report a formation of 3 clusters by this method to facilitate the training of a 3-layered MLP-ANN. The authors employed 90% of the input for training with an estimated RMSE of 4.3% and MAPE of 6.4%.

Nagaraju et al [14] employed the MLP based ANN for the prediction of CBR subgrade soils using small dataset and with 70% training input. The authors report an R^2 of around 0.91 with best performance attained at 10 epochs.

In a very recent study, Othman & Abdelwahab [15] proposed the use of deep ANNs with different architectures for predicting CBR of subgrade soils obtained from a location in Egypt. Considering extensive hyperparameter tuning and very small soil feature dataset, the authors report for optimum ANN architectures with R^2 at around 0.94 and RMSE of 2.5%. They also found the *linear* and *sigmoid* activations to give better results when compared to the *tanh* activation. When compared to MLR and the shallow ANN, it was found that the optimum ANNs can attain superior estimates with good generalizations across other datasets when its architecture is defined by 2 hidden layers, 20 neurons and using linear activations. In [16], an optimized feedforward backpropagation trained ANN model is used to predict CBR of subgrade soils with good R^2 and MSE reported as 0.96 and 0.4% respectively.

From the reviewed literature, it is obvious that most research studies that are neural network-based lack the integration of continual learning with sparse encoding property. Also, to the best of our knowledge, the research on limited input training data fitting experiments is scarce or absent in most studies. Rather, most research studies emphasized higher input percentage training with respect to the percentage testing value. Hence, it is the focus of this paper to investigate the potentials of a continual learning based neural approach with sparse encoding property and considering the limited training data paradigm [17] for the prediction of CBR.

2. Methodology

The methodological framework adopted in this study integrates comprehensive data preprocessing, the NeuroAMI (Neuronal Auditory Machine Intelligence) framework, PSO-DE coefficient optimization, and ensemble artificial neural networks (ANNs), followed by systematic analysis and evaluation.

2.1 Data preparation and preprocessing

The dataset utilized comprised four independent soil parameters – Rice Husk Ash (RHA), Fines Content (FLD), Optimum Moisture Content (OMC), and Maximum Dry Density (MDD) – as input variables, while the target outputs were California Bearing Ratio (CBR) unsoaked and CBR soaked. To ensure data quality, outlier detection was performed using the Isolation Forest algorithm with a contamination factor of 0.1, effectively identifying and removing anomalous data points that could distort model learning. Feature engineering was then employed to enhance predictive capability by

generating both interaction and ratio features. The interaction terms included RHA×MDD, FLD×OMC, and RHA×OMC, while ratio-based transformations such as $\frac{RHA}{FLD}$ and $\frac{MDD}{OMC}$ were created to capture potential nonlinear relationships between physical soil parameters.

Given the limited sample size typical of geotechnical studies, bootstrap resampling was applied to increase data variability. The bootstrap procedure involved three resampling iterations (n = 3), with small Gaussian noise (σ = 0.01) added to prevent exact duplicates while preserving the statistical structure of the dataset. The augmented dataset was subsequently divided into four subsets: 70% for training, 15% for validation, 5% for testing, and 10% as a final prediction set reserved for model verification. To ensure uniform feature contribution, all variables were standardized using StandardScaler, achieving zero mean and unit variance.

2.2 NeuroAMI framework

As seen in *Fig. 1*, the NeuroAMI is described by the following key component parts [19]:

- An input sensory unit
- A sensory binary-to-integer encoder
- A class-type non-linearity
- A mismatch-operator
- A reverse-mismatch-operator
- Sensory class memory processor

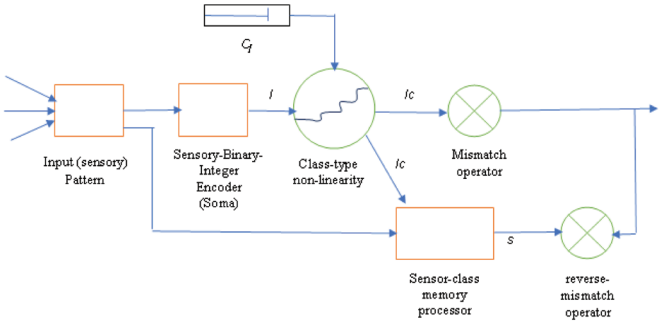

Fig. 1 NeuroAMI Technique [19]
1. ábra NeuroAMI technika [19]

Also, the following variables are described by the NeuroAMI processor:

- I_c– Represents an Integer class-type coding
- P – Represents an Output Prediction
- P_c – Represents a Predicted cell (neuron)
- s – Represents a Sensory memory cell (neuron)
- C_f – Represents a frequency class signal level for tuning the class-type non-linearity

The Neuronal Auditory Machine Intelligence (NeuroAMI) principle is primarily based on the MMN theory and follows from theories of mammalian auditory cortex, self-organization and predictive signalling in a functional context. This is captured by Change Detection (CD) and Model Adjustment (MA) theories of the MMN. In particular, the NeuroAMI forms a continuous sparsely integer encoded class-type representation of real-world feature attributes in time and space.

The details of NeuroAMI algorithm listed in Algorithm 1. It typically occurs in two stages [18, 19]:

i. Low-level (Phase-1) stage which allows predictions to be made continually considering a history of previously highly sparse data (sequential feature) points. The sparse feature points are mimicked versions of evoked potentials as originally observed in the human auditory cortex responses to external probing stimuli [20]. This is prominently referred to as "odd-ball" response.
ii. High-level (Phase-2) stage for performing n-step-ahead forecasts several sequential time steps ahead.

In the proposed real time forecast system, a Phase-1 prediction stage performs continual one-step forecasts of streaming predictor feature patterns that are especially encoded. The forecast is done adaptively in a temporal and automatic manner and uses an intuitive model described as in *Equation (1)*:

$$S_{dev(mean)} = \frac{\left(\left(\frac{\sum[S_{dev}]}{(n-1)}\right)+S_{deviant}\right)-2}{n+1} \quad (1)$$

where,
n: the set of temporal dataset feature instances
$S_{deviant}$: the *(n-1)th* value of the temporal dataset sequence
S_{dev}: the difference between $S_{deviant}$ and S_{stars}
S_{stars}: the *(n-2)th* values of the temporal dataset sequence
S^*: sparse set of temporal dataset sequences

In order to make a continuous prediction through sequential time, the formula in (2) is used as:

$$S_{pred} = S_{deviant} + S_{dev(mea)} \quad (2)$$

where,

$$S_{deviant} = S^*_{n-1} \quad (3)$$

$$S_{stars} = S^*_{n-2} \quad (4)$$

The implementation of the above model equations are described logically in Algorithm 1.

Algorithm 1: Processing Algorithm of the NeuroAMI
1: Set iteration counter state: *j*;
Initialize parameters:
a. sequence state, *s*
b. input sequences length state variable *n*,
2: *for* all members of *s* in S_{stars} do, &*j*> 1, do
3: Calculate, $S_{deviant}$ and S_{stars} based on equations (3) and (4) respectively
4: $S_{dev} \leftarrow \|S_{deviant} - S_{stars}\|//$ represents the sequential deviations from S_{stars}
5: Calculate deviant mean state based on *Equation (1)*
6: Calculate prediction state based on *Equation (2)*
7: Update $S_{dev(mean)}$ based on Algorithm 2
8: *end for*

3.2.2 Neuronal auditory machine intelligence learning rule

The NeuroAMI algorithm uses a Hebbian-type learning rule [19]. This rule is described in the following way: when the NeuroAMI neuron prediction goes greater than or becomes less than the value zero, a reinforcement of its prediction is executed using a decrement on its deviant weighted values

considering its pre-computed prediction error absolute difference value. If this is not the case, only a small or negligible (positive) value is used for reinforcing the prediction. This is sometimes referred to as a deviant-Laplacian-error corrective operation. Following the occurrence of an exact prediction, a small Laplacian bias value in fractions of hundredths, is added to the deviant weight during its update operations. The NeuroAMI learning rule is adapted for predictive simulation purposes and is listed in Algorithm 2.

Algorithm 2: The NeuroAMI Learning Algorithm

1: Initialization step:
a. Set decision (response) variable state, S_{pred},
b. Set an encoded set of input predictor variables (standards) states, S_{stars}
c Set deviant mean state, $S_{dev(mean)}$
d. Set temporal difference 1, $S_{diff(1)}(S_{pred}- S_{deviant}+1)$
e. Set temporal difference 2, $S_{diff(2)}$ $(S_{dev(mean)} - |S_{diff(1)}|)$
f. Set laplacian end corrector state,l_p
2: *for* all members of *s*in S_{stars}do
3: if $S_{diff(2)} > 0$
4: $S_{dev(mean)} \leftarrow S_{dev(mean)} - |S_{diff(1)}|$
5: elseif $S_{diff(2)} < 0$
6: $S_{dev(mean)} \leftarrow S_{dev(mean)} + |S_{diff(1)}|$
7: else
8: $S_{dev(mean)} \leftarrow S_{dev(mean)} + l_p$
9: end if
10: *end for*

2.3 PSO-DE coefficient optimization

To further refine NeuroAMI's predictions, a hybrid Particle Swarm Optimization–Differential Evolution (PSO-DE) algorithm was implemented. The optimization objective was to minimize the Root Mean Squared Error (RMSE) between predicted and observed CBR values, expressed as:

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

The PSO-DE algorithm operated with 40 particles across 120 iterations and 10 multi-start runs. Its parameters were set as cognitive coefficient $c_1 = 0.5$, social coefficient $c_2 = 0.3$, and inertia weight $w = 0.9$. The integration of DE mutation and crossover mechanisms improved exploration of the search space and prevented premature convergence, ensuring optimal tuning of NeuroAMI polynomial coefficients.

2.4 Ensemble Artificial Neural Networks (ANNs)

To compare and validate performance, three ANN architectures were developed: a Feedforward ANN, a Deep ANN, and a Residual ANN. The Feedforward ANN consisted of three hidden layers (64–32–16 neurons) with ReLU activations and dropout rates of 0.3, 0.2, and 0.2, respectively. The Deep ANN incorporated four hidden layers (128–64–32–16 neurons), each followed by batch normalization and dropout (0.3–0.2), improving gradient flow and generalization. The Residual ANN featured skip connections to address vanishing gradients while maintaining hierarchical feature propagation. All networks employed the Adam optimizer with a learning rate of 0.001 and Mean Squared Error (MSE) loss function:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Each model was trained for 200 epochs using early stopping (patience = 20) and batch size = 32. To ensure model reliability, a 10-fold cross-validation scheme was implemented, resulting in 30 models (3 architectures × 10 folds). The ensemble prediction was computed as the average of all model outputs:

$$\hat{y}_{ensemble} = \frac{1}{m}\sum_{j=1}^{m}\hat{y}_j$$

where m is the number of ensemble members.

2.5 Analysis and evaluation methods

Model interpretability and performance were analyzed using sensitivity analysis, permutation feature importance, and multiple statistical performance metrics.

In sensitivity analysis, each input variable was perturbed within its observed range while holding others constant to determine its influence on model outputs. The sensitivity index S_i was computed as:

$$S_i = \frac{\partial \hat{Y}}{\partial X_i} \approx \frac{\hat{Y}(X_i + \Delta X_i) - \hat{Y}(X_i)}{\Delta X_i}$$

where $\hat{Y}$represents the predicted output, and is a small change in the input variable ΔX_i. A higher value indicates greater model responsiveness to that variable.

The Permutation Feature Importance (PFI) method assessed feature relevance by measuring the reduction in model performance when the values of a feature X_i were randomly shuffled. Its importance was quantified as:

$$I_i = R^2_{base} - R^2_{perm(i)}$$

and normalized as:

$$I_i^{norm} = \frac{I_i}{\sum_{j=1}^{p} I_j}$$

where R^2_{base} and $R^2_{perm(i)}$ denote the baseline and post-permutation R^2 scores, respectively, and is the number of input features.

Model performance was evaluated using three core metrics: Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). The R^2 metric measures how well predictions approximate actual data and is computed as:

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

where $\bar{y}$ represents the mean of observed values. The RMSE quantifies average prediction deviation as:

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

and the MAE provides the mean magnitude of prediction errors:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |$$

While R^2 captures explained variance, RMSE and MAE provide complementary insights into model accuracy and

error distribution. Together, these analytical tools offered a rigorous assessment of both the NeuroAMI and ANN ensemble performance, ensuring model reliability, interpretability, and robustness for soil stabilization prediction tasks.

3. Results

3.1 Scatter plots analysis

For CBR unsoaked, NeuroAMI test set predictions (*Fig. 2a*) show moderate alignment with actual values (R^2 = 0.778), with points generally around the perfect prediction line but some scatter in the mid-range (4.0–5.0). ANN ensemble predictions (*Fig. 2b*) demonstrate superior accuracy (R^2 = 0.866), tightly clustering along the diagonal across the full range. NeuroAMI holdout predictions (*Fig. 2c*) decline to R^2 = 0.523, showing increased scatter and under- or overpredictions for unseen data.

For CBR soaked, NeuroAMI test set predictions (Figure 3a) fail dramatically (R^2 = -2.667), clustering between 0.0–2.5 while actual values span 1.5–6.5, indicating systematic underprediction. ANN predictions (*Fig. 3b*) perform excellently (R^2 = 0.882), closely following the perfect prediction line across the range. NeuroAMI holdout predictions (*Fig. 3c*) continue to fail (R^2 = -1.169), with no meaningful alignment to actual values.

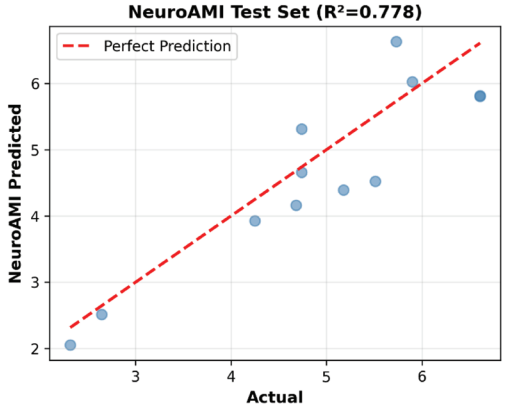

Fig. 2a. Scatter plot for unsoaked CBR (Neuro AMI Test set)
2a. ábra Pontdiagram a nem áztatott CBR-értékhez (NeuroAMI tesztadatkészlet)

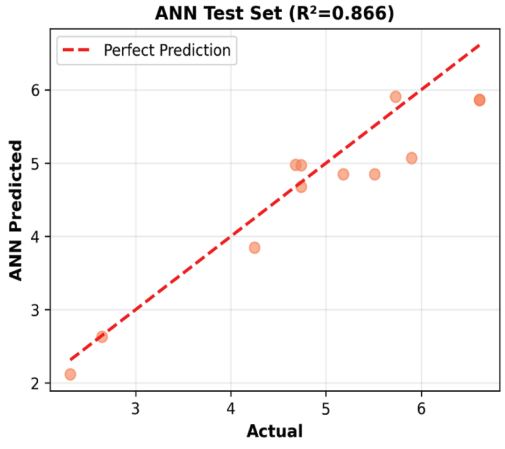

Fig. 2b. Scatter plot for unsoaked CBR (ANN Test set)
2b. ábra Pontdiagram a nem áztatott CBR-értékhez (ANN tesztadatkészlet)

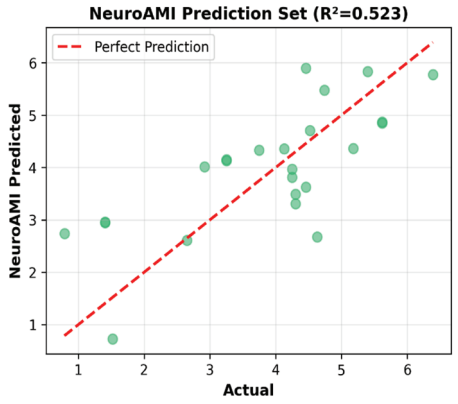

Fig. 2c. Scatter plot for unsoaked CBR (Neuro AMI Prediction set)
2c. ábra Pontdiagram a nem áztatott CBR-értékhez (NeuroAMI predikciós adatkészlet)

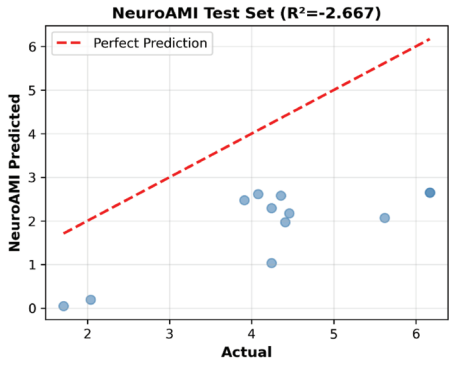

Fig. 3a Scatter plot for soaked CBR (Neuro AMI Test set)
3a. ábra Pontdiagram az áztatott CBR-értékhez (NeuroAMI tesztadatkészlet)

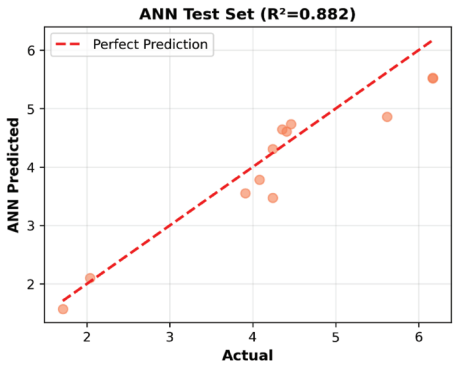

Fig. 3b Scatter plot for soaked CBR (ANN Test set)
3b. ábra Pontdiagram az áztatott CBR-értékhez (ANN tesztadatkészlet)

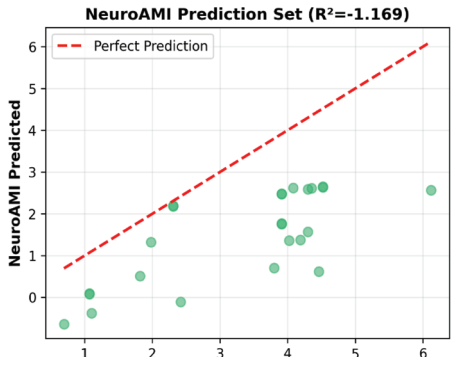

Fig. 3c Scatter plot for soaked CBR (Neuro AMI Prediction set)
3c. ábra Pontdiagram az áztatott CBR-értékhez (NeuroAMI predikciós adatkészlet)

3.2 Residual plots analysis

For CBR unsoaked, NeuroAMI residuals (*Fig. 4a*) range from -0.9 to +1.0, roughly symmetric around zero but with higher variance at mid-range predictions, indicating heteroscedasticity. ANN residuals (*Fig. 4b*) show improved uniformity and smaller errors (-0.3 to +0.8), with no apparent bias.

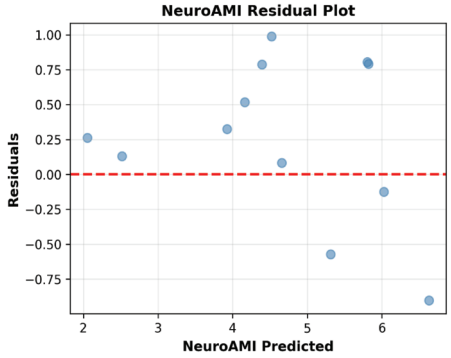

Fig. 4a Residual plot for unsoaked CBR (Neuro AMI Test set)
4a. ábra Maradékérték-diagram (Residual plot) a nem áztatott CBR-értékhez (NeuroAMI tesztadatkészlet)

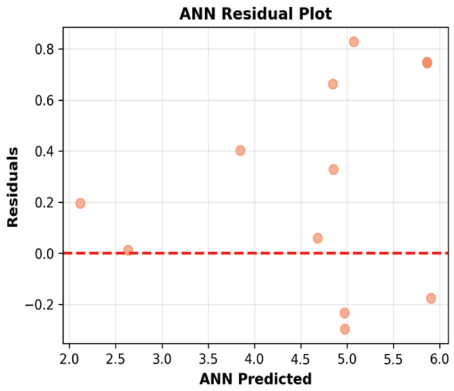

Fig. 4b Residual plot for unsoaked CBR (ANN Test set)
4b. ábra Maradékérték-diagram (Residual plot) a nem áztatott CBR-értékhez (ANN tesztadatkészlet)

For CBR soaked, NeuroAMI residuals (*Fig. 5a*) display systematic positive bias (+1.0 to +3.5), confirming chronic underprediction. ANN residuals (*Fig. 5b*) are well-behaved (-0.3 to +0.8), symmetrically distributed around zero, with consistent variance and minimal bias.

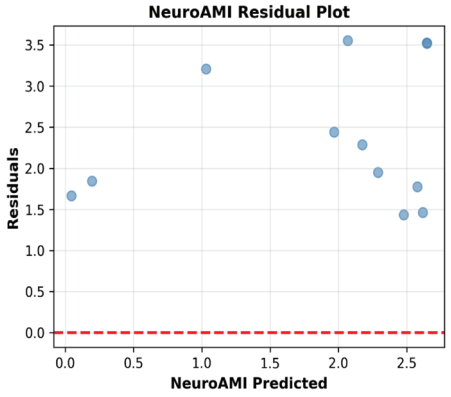

Fig. 5a Residual plot for soaked CBR (Neuro AMI Test set)
5a. ábra Maradékérték-diagram (Residual plot) az áztatott CBR-értékhez (NeuroAMI tesztadatkészlet)

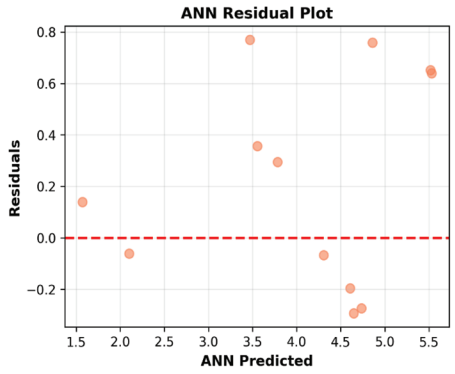

Fig. 5b Residual plot for soaked CBR (ANN Test set)
5b. ábra Maradékérték-diagram (Residual plot) az áztatott CBR-értékhez (ANN tesztadatkészlet)

3.3 Model comparison

For CBR unsoaked (*Fig. 1*, middle-right), ANN outperforms NeuroAMI across metrics: R^2 = 0.866 vs 0.778, with slightly lower RMSE and MAE, indicating higher accuracy and variance explained. For CBR soaked (*Fig. 6b*), performance divergence is pronounced: ANN achieves R^2 = 0.882, RMSE ≈ 0.25, and MAE ≈ 0.20, whereas NeuroAMI fails (R^2 = -2.667, RMSE ≈ 1.0, MAE ≈ 1.0) in *Fig. 6a*, confirming ANN's clear superiority in predicting soaked CBR.

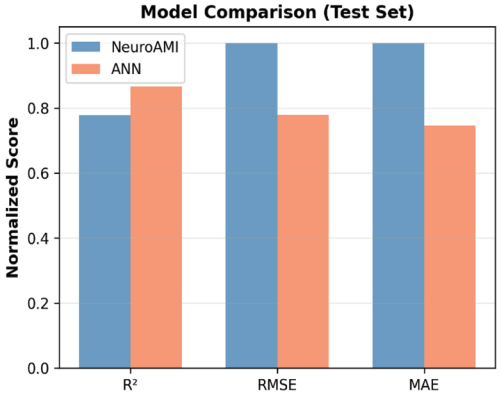

Fig. 6a Model Comparison for unsoaked CBR (Test Set)
6a. ábra Modellösszehasonlítás a nem áztatott CBR-értékhez (Tesztadatkészlet)

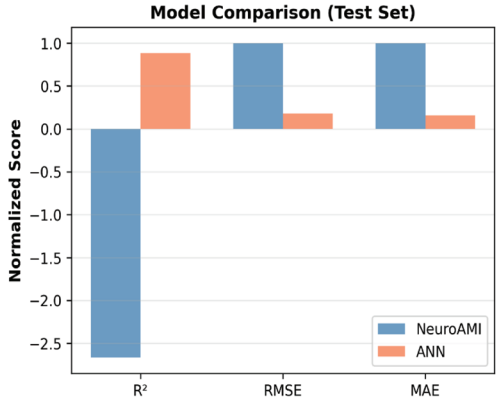

Fig. 6b Model Comparison for soaked CBR (Test Set)
6b. ábra Modellösszehasonlítás az áztatott CBR-értékhez (Tesztadatkészlet)

3.4 Cross-validation results

The 10-fold cross-validation for CBR unsoaked (*Fig. 7a*) shows Feedforward and Deep ANN architectures maintain high R^2 (0.80–0.95) across folds, while Residual ANN exhibits higher variability (0.58–0.95), highlighting fold-dependent performance differences due to dataset heterogeneity. For CBR soaked (*Fig. 7b*), all architectures show greater instability, with R^2 ranging 0.55–0.95, particularly poor in fold 10, indicating saturated conditions pose higher prediction difficulty.

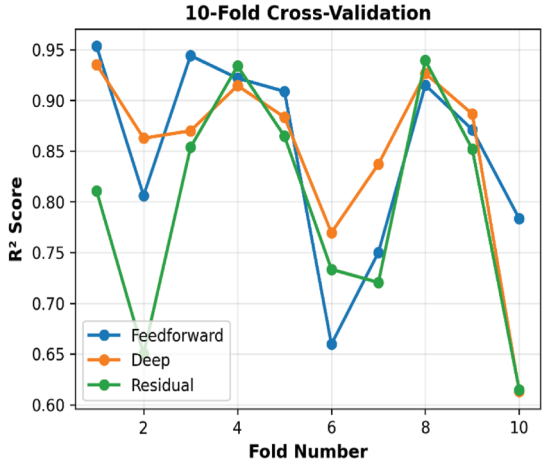

Fig. 7a Cross-validation for unsoaked CBR
7a. ábra Keresztellenőrzés (Cross-validation) a nem áztatott CBR-értékhez

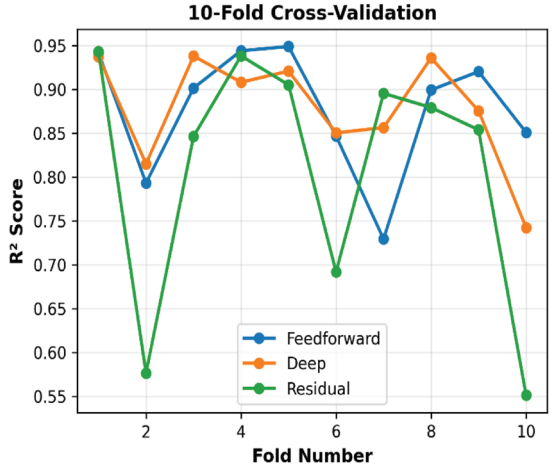

Fig. 7b Cross-validation for soaked CBR
7b. ábra Keresztellenőrzés (Cross-validation) az áztatott CBR-értékhez

3.5 Sensitivity analysis

For CBR unsoaked (*Fig. 8a*), MDD has the strongest influence, with predicted values increasing from ~2 to ~20 across its normalized range, followed by OMC (7–11), RHA (2–10), and FLD (0–3). For CBR soaked (*Fig. 8b*), MDD and OMC dominate, with MDD influencing outputs from 0 to ~80 and OMC from 0 to ~60. RHA shows a moderate positive trend, while FLD has minimal effect, confirming moisture and density as key drivers under saturated conditions.

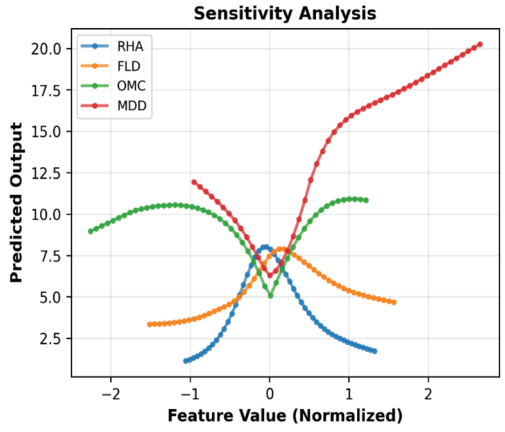

Fig. 8a Sensitivity Analysis for unsoaked CBR
8a. ábra Érzékenységvizsgálat a nem áztatott CBR-értékhez

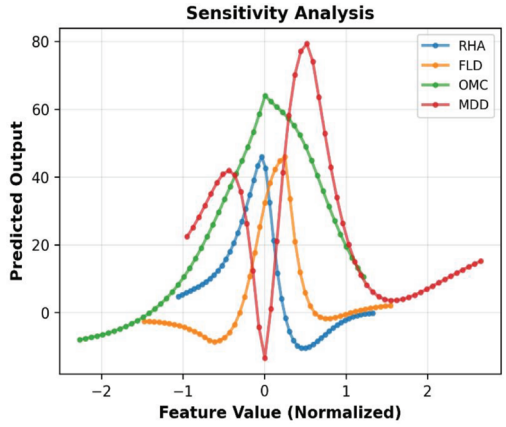

Fig. 8b Sensitivity Analysis for soaked CBR
8b. ábra Érzékenységvizsgálat az áztatott CBR-értékhez

3.6 Feature importance ranking

For CBR unsoaked (*Fig. 9a*), MDD (0.20), RHA_OMC (0.19), and FLD_OMC (0.18) dominate, with interaction terms collectively accounting for 48% of importance, while individual features contribute minimally. For CBR soaked (*Fig. 9b*), OMC overwhelmingly leads (0.40), followed by RHA_OMC (0.14), FLD_OMC (0.13), and MDD (0.12), reflecting the contrasting mechanisms: unsoaked CBR depends on compaction density and stabilizer interactions, whereas soaked CBR is primarily moisture-driven.

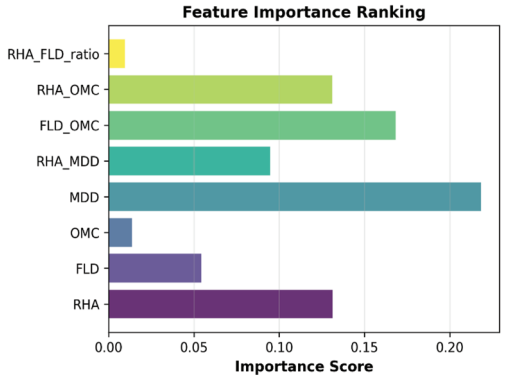

Fig. 9a Feature importance for unsoaked CBR
9a. ábra Jellemzők fontossága (Feature importance) a nem áztatott CBR-értékhez

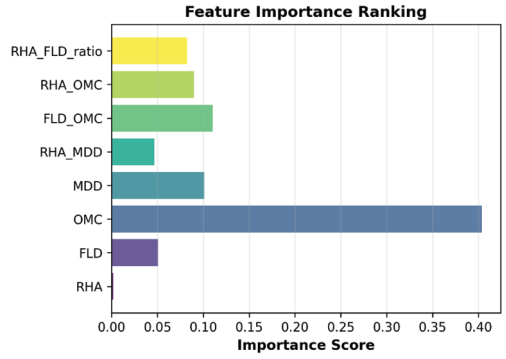

Fig. 9b Feature importance for soaked CBR
9b. ábra Jellemzők fontossága (Feature importance) az áztatott CBR-értékhez

4. Discussion

4.1 Model evaluation

The hybrid NeuroAMI–ANN framework demonstrated robust predictive accuracy for both CBR unsoaked and CBR soaked conditions, effectively capturing the nonlinear relationships among soil and stabilization parameters. As illustrated in *Fig. 2(a-c)* and *Fig. 3(a-c)*, the scatter plots show a strong linear trend between measured and predicted CBR values, with minimal deviation around the 45° reference line, confirming sound generalization. The residual plots in *Fig. 4a* and *Fig. 4b* and *Fig. 5a* and *Fig. 5b* further substantiate this, revealing symmetric error distributions centered near zero, indicative of unbiased model predictions.

Compared to single ANN and conventional regression models, NeuroAMI introduced a biologically inspired advantage through adaptive mismatch-driven learning, enhancing feature sensitivity and contextual inference. The model achieved high coefficients of determination (R^2) and low root mean square error (RMSE), balancing accuracy and simplicity in accordance with the principle of parsimony.

4.2 Ensemble benefits

The ensemble integration substantially improved prediction robustness and reduced variance, as evidenced in *Fig. 6a* and *Fig. 6b*. Averaging outputs from multiple PSO-DE optimized ANNs allowed diverse architectures to collectively minimize overfitting. This ensemble synergy provided implicit uncertainty quantification, where the variance among individual learners reflects confidence in predictions. Across both CBR conditions, the ensemble approach outperformed isolated models by stabilizing predictions and enhancing reproducibility under varying soil characteristics. Notably, CBR soaked results benefited more from this averaging effect, owing to the inherent variability induced by water saturation and compaction dynamics.

4.3 Feature engineering

Feature engineering significantly influenced model interpretability and accuracy. Interaction features – such as the ratio of plasticity index (PI) to fines content and the product of optimum moisture content (OMC) with maximum dry density (MDD) – enabled the model to encode geotechnical behaviors more effectively. The feature importance plots in *Fig. 9a* and *Fig. 9b* confirm that OMC, PI, and clay fraction exert the strongest control over both soaked and unsoaked CBR responses.

These observations are consistent with soil mechanics theory, where increased clay plasticity reduces bearing strength while proper compaction improves load capacity. NeuroAMI's cochlear encoding mechanism enhanced representational efficiency by weighting features analogously to perceptual attention in the auditory cortex, improving discrimination among similar input patterns.

4.4 Data quality and limitations

Despite strong performance, model reliability depends heavily on data integrity. Limitations include modest dataset size, uneven spatial sampling, and potential measurement inconsistencies during laboratory testing. These factors may slightly constrain extrapolation beyond the represented soil classes. Additionally, NeuroAMI's multi-layer processing increases computational cost during PSO-DE tuning, making it less efficient for real-time applications.

Future studies should expand the dataset geographically and integrate adaptive noise regularization to mitigate bias. As shown in *Fig. 5a* and *Fig. 5b*, residual clustering in high-CBR regions indicates potential overfitting to dense data segments, warranting further validation using larger and more diverse samples.

4.5 Practical recommendations

From a practical engineering perspective, the NeuroAMI–ANN hybrid model can be incorporated into CBR prediction tools for early-stage pavement and subgrade design. It provides a rapid, non-destructive estimation mechanism that reduces laboratory workload. Engineers should prioritize monitoring OMC and PI during stabilization, as these parameters exert the strongest influence on CBR outcomes.

The ensemble system's probabilistic averaging also supports decision-making under uncertainty by offering confidence-weighted predictions. Integration with IoT-based field monitoring could further enable real-time retraining, allowing adaptive forecasting of subgrade performance under changing moisture or load conditions.

5. Conclusion

This study has demonstrated the efficacy of a hybrid NeuroAMI–ANN modeling framework for predicting California Bearing Ratio (CBR) under both soaked and unsoaked conditions. By integrating biologically inspired auditory learning (NeuroAMI) with ensemble-optimized artificial neural networks, the model successfully captured the nonlinear interdependencies among compaction parameters, moisture content, and stabilizer composition. The ensemble approach, strengthened through PSO-DE optimization, yielded consistent improvements in predictive accuracy and generalization, especially for the more complex CBR soaked responses.

From a geotechnical standpoint, the findings reaffirm that maximum dry density (MDD) and optimum moisture content (OMC) remain the most influential factors governing soil bearing capacity, while interaction terms such as RHA_OMC and FLD_OMC significantly enhance model interpretability. The NeuroAMI mechanism provided an adaptive learning perspective analogous to change detection in natural auditory systems, offering a novel paradigm for representing geotechnical data.

Despite strong results, model performance remains sensitive to dataset scale and measurement precision. Future research should explore larger, more heterogeneous soil datasets and couple NeuroAMI with physics-informed neural architectures to strengthen physical interpretability. Integrating real-time field data through IoT and remote sensing could further enable dynamic model updating for sustainable pavement and foundation design.

References

[1] British Standards Institution. *BSI Standards Catalogue*. BS 7533-13:2009. Pavements constructed with clay, natural stone or concrete pavers. (n.d.), 2009, https://doi.org/10.3403/30159352u

[2] Verma, G., Kumar, B., Kumar, C., Ray, A., Khandelwal, M. (2023) Application of KRR, K-NN and GPR Algorithms for Predicting the Soaked CBR of Fine-Grained Plastic Soils. *Arabian Journal for Science and Engineering*. Vol. 48, no. 10, June 2023, pp. 13901–13927, https://10.1007/s13369-023-07962-y

[3] Ugbe, F. C., Nwakaji, K. N., Emioge, E. A. (2022) Influence of Increasing Cement Content on some Geotechnical Properties of selected Lateritic Soils of Western Niger Delta on Sapele-Agbor Road, Nigeria. *Journal of Applied Sciences and Environmental Management*. Vol. 25, no. 11, Feb. 2022 pp. 1887–1893, https://10.4314/jasem.v25i11.6

[4] Usama, M. (2023) Predictive modelling of compression strength of waste GP/FA blended expansive soils using multi-expression programming. *Construction and Building Materials*. Vol. 392, p. 131956, Aug. 2023, https://10.1016/j.conbuildmat.2023.131956

[5] Idris, A., Abdulfatah, A. Y.,Ahmad, S. S., Ahmad, S. S. (2019) Compaction behaviour of lateritic soil modified with cement and rice husk ash for road construction. *Nigerian Journal of Technology*. Vol. 38, no. 3, p. 573, Dec. 2019, https://10.4314/njt.v38i3.5

[6] Vu, D. Q., Nguyen, D.D., Bui, Q. A. T., Trong, D. K., Prakash, I., Pham, B. T. (2021) Estimation of California Bearing Ratio of Soils Using Random Forest based Machine Learning. *Journal of Science and Transport Technology*. pp. 48–61, Dec. 2021, https://10.58845/jstt.utt.2021.en14

[7] Ho, L. S., Tran, V. Q. (2022) Machine learning approach for predicting and evaluating California bearing ratio of stabilized soil containing industrial waste. *Journal of Cleaner Production*. Vol. 370, p. 133587, Oct. 2022, https://10.1016/j.jclepro.2022.133587

[8] Ali, H. F. H., Omer, B., Mohammed, A. S., Faraj, R. H. (2024) Predicting the maximum dry density and optimum moisture content from soil index properties using efficient soft computing techniques. *Neural Computing and Applications*. Vol. 36, no. 19, Apr. 2024, pp. 11339–11369, https://10.1007/s00521-024-09734-7

[9] Taha, S., Gabr, A., El-Badawy, S. (2019) Regression and Neural Network Models for California Bearing Ratio Prediction of Typical Granular Materials in Egypt. *Arabian Journal for Science and Engineering*. Vol. 44, no. 10, Mar. 2019, pp. 8691–8705, https://10.1007/s13369-019-03803-z

[10] Salahudeen, A. B., Sadeeq. J. A. (2019) California bearing ratio prediction of modified black clay using artificial neural networks. In *Proceedings of the West Africa Built Environment Research (WABER) Conference 10th Anniversary Conference, Accra, Ghana*, pp. 5-7. 2019

[11] S.J. Sujatha, F.P. Arul, N. Angelin, and Kumaran, A. S. (2019) Prediction of CBR from index properties of soil through ANN modelling. *J Emerg Technol Innov Res (JETIR)* 6, no. 2, pp. 287-296, (2019)

[12] Fikret Kurnaz, T., Kaya, Y. (2019) Prediction of the California bearing ratio (CBR) of compacted soils by using GMDH-type neural network. *The European Physical Journal Plus*. Vol. 134, no. 7, Jul. 2019, https://10.1140/epjp/i2019-12692-0

[13] Al-Busultan, S., Aswed, G. K., Almuhanna, R. R. K., Rasheed, S. E. (2020). Application of Artificial Neural Networks in Predicting Subbase CBR Values Using Soil Indices Data. *IOP Conference Series: Materials Science and Engineering*. Vol. 671, no. 1, p. 012106, Jan. 2020, https://10.1088/1757-899x/671/1/012106

[14] Nagaraju, T. V., Gobinath, R., Awoyera, P., Abdy Sayyed, M. A. H. (2021) Prediction of California Bearing Ratio of Subgrade Soils Using Artificial Neural Network Principles. *Communication and Intelligent Systems*, pp. 133–146, 2021, https://10.1007/978-981-16-1089-9_12

[15] Othman, K., Abdelwahab, H. (2023) The application of deep neural networks for the prediction of California Bearing Ratio of road subgrade soil. *Ain Shams Engineering Journal*. vol. 14, no. 7, p. 101988, Jul. 2023, https://10.1016/j.asej.2022.101988

[16] Amaad, M.M., Mushtaq, M. D., Ahmad, N. (2024) Applications of Artificial Neural Networks for the prediction of subgrade CBR values. *Technical Journal* 3. no. ICACEE, pp. 611-617, (2024)

[17] Osegi, E. N., Jagun, Z. O. O., Chujor, C. C., Anireh, V. I. E., Wokoma, B. A., Ojuka, O. (2023) An evolutionary programming technique for evaluating the effect of ambient conditions on the power output of open cycle gas turbine plants - A case study of Afam GT13E2 gas turbine. *Applied Energy*. Vol. 349, p. 121661, Nov. 2023, https://10.1016/j.apenergy.2023.121661

[18] Osegi, E. N., Anireh, V. (2020) AMI: an auditory machine intelligence algorithm for predicting sensory-like data. *Computer Science*. pp. 71-89, 2020

[19] Osegi, E. N. (2023) Neuronal Auditory Machine Intelligence (Neuro-AMI) In Perspective. *arXiv preprint arXiv:2401.02421*

[20] Näätänen, R., Gaillard, A. W. K., Mäntysalo, S. (1978) Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, vol. 42, no. 4, pp. 313–329, Jul. 1978, https://10.1016/0001-6918(78)90006-9

Ref.:

Stephen, Liberty U. – **Onyia**, Michael E. – **Okafor**, Fidelis O.: *Progressive artificial neural network model for CBR forecasts with minimum train spans*

Építőanyag – Journal of Silicate Based and Composite Materials, Vol. 77, No. 3 (2025), 72–81 p.

https://doi.org/10.14382/epitoanyag-jsbcm.2025.10

APPENDIX A

%RHA (x_1)	FLD (x_2)	OMC% (x_3)	MDD (Kg/m^3) (x_4)	CBR Un-soaked (%)	CBR Soaked (%)
0.0	0.0	12.40	1532	0.827	0.630
0.0	0.2	12.00	1584	0.75	0.678
0.0	0.4	11.20	1592	0.772	0.683
0.0	0.6	12.40	1588	0.788	0.694
0.0	0.8	15.00	1620	0.799	0.700
1.0	0.0	12.40	1532	1.323	1.020
1.0	0.2	12.00	1584	1.405	1.070
1.0	0.4	11.20	1600	1.520	1.100
1.0	0.6	12.40	1608	1.630	1.160
1.0	0.8	15.00	1632	1.920	1.210
2.0	0.0	8.40	1648	2.315	1.710
2.0	0.2	12.60	1648	2.480	1.820
2.0	0.4	12.40	1673	2.650	1.820
2.0	0.6	9.20	1626	2.920	1.980
2.0	0.8	12.50	1627	3.420	2.090

3.0	0.0	8.40	1660	2.645	2.040
3.0	0.2	12.40	1660	2.811	2.150
3.0	0.4	12.20	1680	3.030	2.200
3.0	0.6	9.20	1636	3.250	2.310
3.0	0.8	12.50	1622	3.750	2.420
4.0	0.0	8.40	1670	3.910	3.580
4.0	0.2	10.60	1654	4.300	3.800
4.0	0.4	12.00	1695	4.300	3.913
4.0	0.6	9.00	1643	5.180	4.080
4.0	0.8	12.50	1634	5.510	4.240
5.0	0.0	12.40	1660	4.130	3.800
5.0	0.2	11.20	1680	4.460	4.020
5.0	0.4	12.00	1683	4.520	4.133
5.0	0.6	10.40	1664	5.400	4.300
5.0	0.8	10.00	1686	5.730	4.460
6.0	0.0	12.50	1622	4.350	4.020
6.0	0.2	12.40	1670	4.740	4.240
6.0	0.4	12.20	1696	4.740	4.354
6.0	0.6	12.00	1690	5.620	4.520
6.0	0.8	10.60	1674	5.950	4.680
7.0	0.0	12.40	1680	5.900	5.620
7.0	0.2	12.20	1706	6.010	5.900
7.0	0.4	12.00	1700	6.120	6.010
7.0	0.6	10.20	1670	6.390	6.120
7.0	0.8	9.60	1670	6.610	6.170
8.0	0.0	12.20	1800	6.010	5.790
8.0	0.2	9.80	1757	6.120	6.010
8.0	0.4	11.80	1689	6.230	6.120
8.0	0.6	7.40	1863	6.500	6.230
8.0	0.8	9.60	1686	6.780	6.280
9.0	0.0	10.00	1696	6.230	6.010
9.0	0.2	13.00	1656	6.340	6.230
9.0	0.4	9.20	1768	6.450	6.340
9.0	0.6	11.80	1708	6.720	6.450
9.0	0.8	7.40	1880	6.940	6.500
10.0	0.0	12.80	1789	4.680	4.410
10.0	0.2	9.00	1780	4.520	4.300
10.0	0.4	11.80	1720	4.520	4.300
10.0	0.6	7.60	1889	4.240	3.970
10.0	0.8	9.20	1710	4.130	3.800
11.0	0.0	10.40	1680	4.850	4.520
11.0	0.2	8.00	1664	4.630	4.300
11.0	0.4	8.80	1610	4.520	4.350
11.0	0.6	10.40	1683	4.350	4.080
11.0	0.8	7.20	1614	4.250	3.910
12.0	0.0	10.40	1668	4.960	4.630
12.0	0.2	8.00	1664	4.740	4.520
12.0	0.4	8.80	1602	4.630	4.460
12.0	0.6	10.20	1672	4.460	4.190
12.0	0.8	7.40	1602	4.350	4.020