

USE OF NEURAL NETWORKS WITHIN CONSTITUTION MODELS OF SOILS

Filip Cigán^{*1}

*filip.cigan@uniza.sk

¹University of Žilina, Univerzitná 8215/1 010 26 Žilina

Abstract

This paper focuses on the innovative use of machine learning and neural networks in constitutive modelling of soils, a material with complex and nonlinear behaviour. Traditional constitutive models, based on Hooke's law or the Mohr-Coulomb model, often show significant discrepancies from the real-world behaviour of soils, leading to high costs and uncertainties in construction projects. The aim of this work is to lay the groundwork for a neural network capable of learning and reproducing results that are closer to the real behaviour of soils than current constitutive models. This approach could bring about a revolutionary change in the fields of geotechnics and construction by providing more accurate and efficient models for analysis and design of structures. The results could lead to the optimization of materials, cost reduction, and increased safety and sustainability of construction projects. This interdisciplinary approach opens up new possibilities for research and applications, with the potential to significantly contribute to innovations in geotechnics and construction.

Keywords

Machine learning, neural networks, constitutive model, geotechnics

1 INTRODUCTION

This paper focuses on the application of machine learning principles and neural networks in the constitutive modelling of soils. Constitutive models are mathematical representations that describe how materials behave under various physical conditions. In construction and geotechnical engineering, these models are critical for predicting the performance of materials like concrete, steel, and soil under different loading conditions. Soil presents a unique challenge as it is predominantly an inhomogeneous, anisotropic material with highly nonlinear behaviour.

A well-known example of a constitutive model is Hooke's law for linearly elastic materials [1]. However, even more complex models such as Cam-Clay are insufficient to capture all aspects of soil behaviour. The existing models often lead to computational instability and necessitate additional inputs, significantly increasing costs.

The aim of this research is to harness neural networks to predict soil deformation without the need for explicitly defining individual aspects of constitutive models. The study will utilize synthetic data that represents the Mohr-Coulomb elastic-plastic constitutive model of soil. This model, despite not being the ideal representation for soil behaviour, has been selected due to its relative simplicity. This simplicity facilitates the testing and validation of various methodologies and results, essential in the initial phase of our research. The methodology will involve data collection, training of neural networks, and comparison of the results with existing models.

This approach could revolutionize geotechnical engineering and construction by shifting from traditional linear models to more sophisticated nonlinear models. It has the potential to lead to optimized materials and designs, reducing construction costs while enhancing safety and sustainability. This interdisciplinary exploration might pave the way for new research directions and innovative methodologies in constitutive modelling [2].

2 MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

In 2023, Artificial Intelligence (AI) is at the forefront of technological development, greatly propelled by OpenAI's release of ChatGPT in November 2022. This large language model has elevated AI as one of the fastest-growing fields, building on foundational work like Alan Turing's 1950 paper introducing the "Turing Test." [3].

Media often conflates terms like Machine Learning (ML), Deep Learning (DL), and AI. However, experts differentiate them as follows:

- AI: The broadest term, encompassing any effort to simulate human intelligence using machines. It can include problem-solving, language understanding, voice recognition, and social intelligence.
- ML: A subset of AI focusing on methods that allow machines to learn from data and improve performance without explicit programming. Commonly used in predictive analytics and fraud detection.
- DL: A specialized subset of ML inspired by human neural networks, used in applications like real-time language translation and image recognition.

Machine learning

Machine Learning (ML) is a field under Artificial Intelligence (AI) that leverages datasets to create models. In this context, “data” refers to a variety of information types like text, audio, and images. The “model” is the result of analysing these data. Unlike traditional techniques, ML does not require human intervention. The algorithm learns to identify patterns and relationships in the data on its own, making it analogous to training—hence the term “training data”.

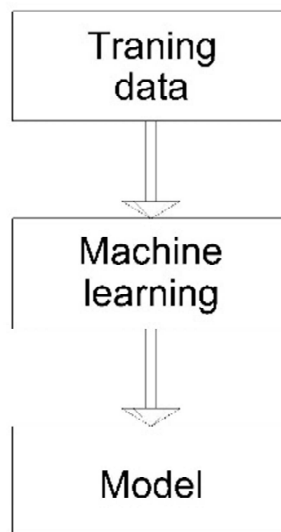


Fig. 1 Machine learning process.

Fig. 1 demonstrates the ML process. The “model” can be thought of as the end goal. For instance, in developing an automatic spam email filter, the spam filter itself is the model. Some prefer to term it a “hypothesis”, particularly those with a statistical background.

ML is not the only modelling technique. Classical mechanics employs modelling based on Newton’s laws, while expert systems in AI are based on specialized knowledge. However, ML excels in areas where laws and logical reasoning are less applicable, such as image recognition, speech recognition, and natural language processing [4].

Neural Networks and Nodes

Analogous to the human brain’s vast neural network, a neural network in AI consists of nodes. Depending on how these nodes are connected, various types of neural networks can be formed. A commonly used type is the layered structure, as illustrated in Fig. 2.

The set of nodes forming the grey structure is called the input layer, primarily serving as channels for transmitting input signals without performing weighted calculations or activation functions. In contrast, the blue nodes constitute the output layer, whose output represents the result of the network. Layers between the input and output are known as hidden layers, named for their inaccessibility from the neural network’s external environment.

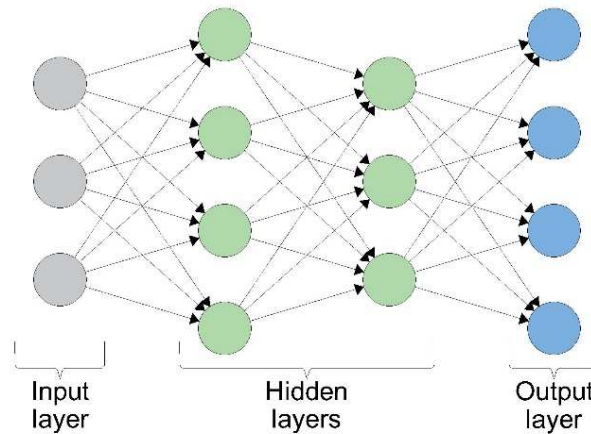


Fig. 2 Structure of layers.

Artificial Neurons: Principles and Functions

Artificial neurons or nodes are mathematical functions designed to mimic biological neurons. The basic model involves multiplying input values by weights (analogous to synaptic weights in biological neurons), summing these weighted inputs, and then applying an activation function (akin to the threshold function in biological neurons). Despite significant differences between biological and artificial neurons, their foundational principles are similar, enabling the development of more sophisticated and efficient artificial neural networks. A neuron is illustrated in Fig. 3.

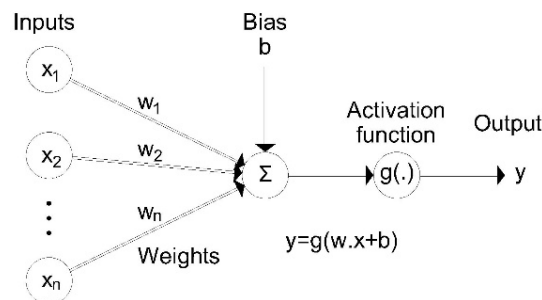


Fig. 3 An artificial neuron.

In this setup, each input signal is multiplied by a corresponding weight before reaching the node. Once inside the node, these weighted signals are summed to form a weighted sum. The weighted sum is calculated in relation (1) as follows:

$$\Sigma = (w_1 \cdot x_1) + (w_2 \cdot x_2) + (w_n \cdot x_n) + b \tag{1}$$

here x_1, x_2, \dots, x_n are the input signals (-) and w_1, w_2, \dots, w_n are the weights (-), with b representing the bias (-). The weighted sum then undergoes an activation function $g()$, resulting in the node's output. Activation functions vary and can include types such as sigmoid, hyperbolic tangent, or rectified linear unit (ReLU). We will refer to the activation function (2) as:

$$y = g(\Sigma) \tag{2}$$

In summary, the information in an artificial neural network is stored in the form of weights and biases, and the activation function determines the node's behaviour. The model allows for dynamic learning similar to how the brain alters neuronal associations [5].

The resulting formula for the output from the neural network then looks like:

$$y = g([w_1 \cdot x_1] + [w_2 \cdot x_2] + [w_n \cdot x_n] + b) \tag{3}$$

We can generalize (3) equation into the following modified form (4) as:

$$y = g(w \cdot x + b) \tag{4}$$

3 CONSTITUTIVE MODEL IN GEOTECHNICS

The constitutive model serves as a fundamental building block in the analysis and design of geotechnical structures. Its primary aim is to quantify and describe the complex behaviour of geotechnical materials, particularly deformations and stresses. These models enable predictions of responses to various mechanical loads. They incorporate key characteristics such as material properties (cohesion, angle of internal friction, elastic modulus, etc.), modes of deformation (e.g., elastic, plastic, or elastoplastic), soil-water interactions (considering factors like saturation and total/effective stress), and deformation history (accounting for prior load cycles affecting current material behaviour) [6].

Mohr-Coulomb Plasticity Condition

The most used plasticity criterion in soil mechanics is the Mohr-Coulomb condition. It represents the envelope of Mohr's circles at yield stress and is usually expressed as follows:

$$\tau + \sigma \cdot \tan(\varphi) - c = 0 \tag{5}$$

where τ (kPa) is the shear stress and σ is the normal stress (kPa) on the shear plane. The Mohr envelope describes the increase in yield shear stress with rising normal stress, represented by the internal friction angle φ (°). Cohesion c (kPa) indicates the shear strength at zero normal load [7].

Although ϕ and c are often considered standard soil or rock parameters, it is important to note that these are only associated with the Mohr-Coulomb failure criterion Fig. 4, one among many plasticity conditions [8].

The form of Equation (5) is not ideal for use in a constitutive relation as it requires finding the plane with maximum τ/σ ratio. A more useful expression would be:

$$f = \frac{1}{2}(\sigma_1 - \sigma_3) + \frac{1}{2}(\sigma_1 + \sigma_3) \sin(\varphi) - c \cdot \cos(\varphi) = 0 \tag{6}$$

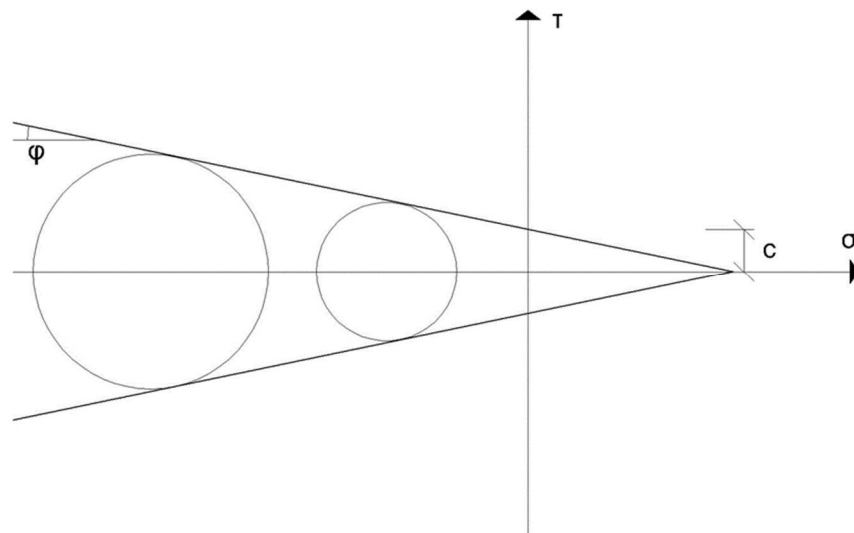


Fig. 4 Mohr-Coulomb criterion of plasticity expressed in the plane in the plane $\tau - \sigma$.

In the case of isotropic material, the order of the principal axes can be arbitrarily changed. Therefore, Equation (6) represents a regular hexagonal pyramid in the principal stress space, as depicted in Fig. 5.

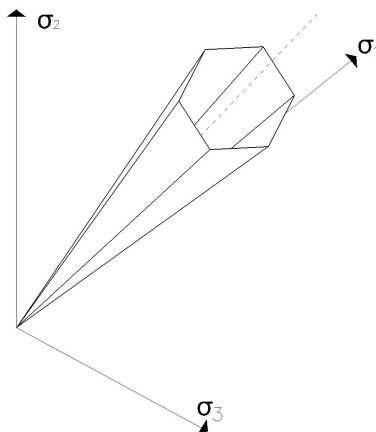


Fig. 5 Mohr-Coulomb criterion of plasticity in the space of principal stresses.

4 METHODOLOGY

Neural Network Architecture

- Example 1: Initial neural network setup with a single input node.
- Example 2: Three input nodes representing deformation (ϵ), internal friction angle (ϕ), and cohesion (c).
- Example 3: Modified structure with two input nodes for deformation difference ($\Delta\epsilon$), and stress state (σ).

Training and Evaluation

- Training data generated for each example, specifying deformation and stress values within a defined range.
- RMSE (Root Mean Square Error) used as the evaluation metric.
- Identification of problematic areas with high error rates for further investigation.

Output Analysis

- Visualization of predictions.
- Analysis of anomalies and discrepancies in prediction.

Future Work

- Focus on minimizing or eliminating identified error zones.
- Solve two-dimensional problems using constitutive models.

This section aims to provide an overview of the methodology employed in our experiments and highlight areas for future work.

5 PRACTICAL BASIS OF THE SOLUTION

The practical part of this work sets the foundational steps for future application and training of neural networks within the context of advanced soil models. For simplicity, this section focuses on a one-dimensional example. Although a one-dimensional model is simpler than multi-dimensional ones, it offers a good overview of the problem and highlights challenges that will need to be addressed in the future.

Example 1

The initial example introduces one of the potential methodological approaches for stress and strain prediction. Given the one-dimensional nature of the model, the stress trajectory is considered as the only possible one, enabling neural network training on the direct relationship between deformations and stress. Mohr-Coulomb constitutive model is used to generate the dataset in the first experiment, visualized in Fig. 6. Deformations and stresses range from 0 to 1, with a transition from elastic to plastic deformation occurring at a stress value of 0.5.

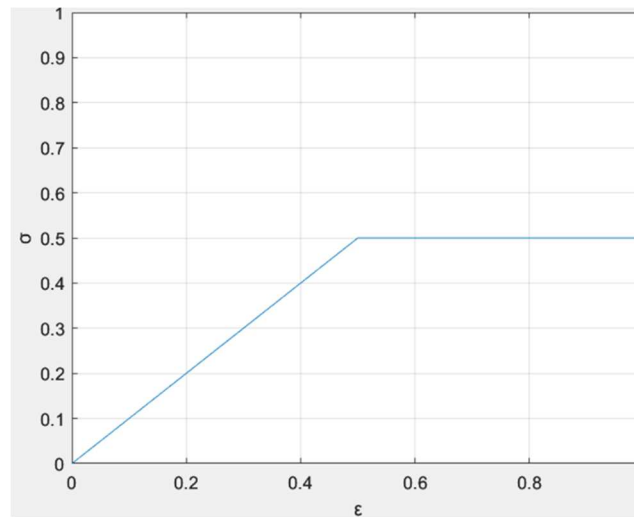


Fig. 6 Generated data.

A total of 10000 data points mapping the relationship between deformation and stress are generated and serve as the training set for the neural network. The resultant data visualized in Fig. 7 will provide the constitutive model for subsequent neural network training phases. The network architecture is a feed-forward neural network, consisting of one input layer, two hidden layers with eight neurons each, and one output layer, based on general recommendations and heuristics [9].

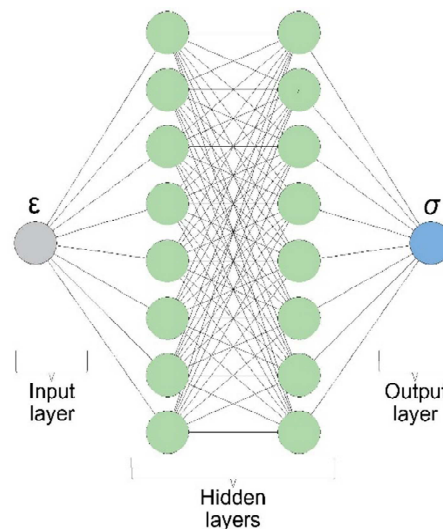


Fig. 7 The structure of a neural network.

The neural network achieving an RMSE accuracy of 0.005. Fig. 8 illustrates the reference data in black, representing the deformation and stress graph, while predictions made by the neural network are shown in red. It is important to note that the data used for prediction is different from the training data, as explicitly recorded in Tab. 1.

Tab. 1 Data comparison example 1.

Data for training	Data for prediction
$\varepsilon [0; 0.0001; 0.0002; 0.0003 \dots 1]$	$\varepsilon [0.00005; 0.00015; 0.00025; \dots 0.99995]$

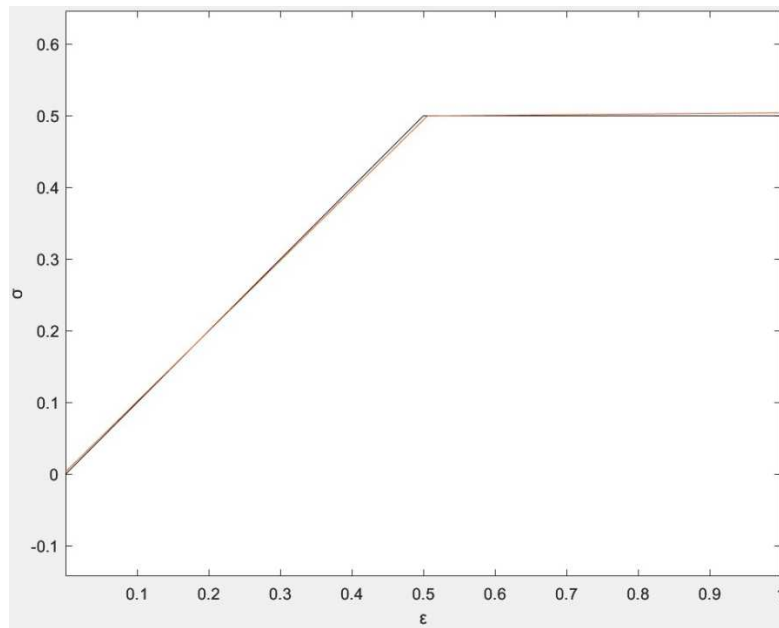


Fig. 8 The result of the neural network – red predictions, black correct answers.

Example 1 yielded fairly successful results, confirming that a feed-forward neural network architecture can effectively predict stress in the context of deformations in a one-dimensional model.

Example 2

Like Example 1, Example 2 will also focus on a one-dimensional problem. The key difference is an attempt to effectively predict stress using multiple types of soils, represented by variable parameters φ (angle of internal friction) and c (cohesion). Specifically, 100 values each for φ , ranging from 30 to 35, and c , ranging from 10 to 15, will be generated. Additionally, 100 values for deformations ε in the range of 0 to 25 will be generated.

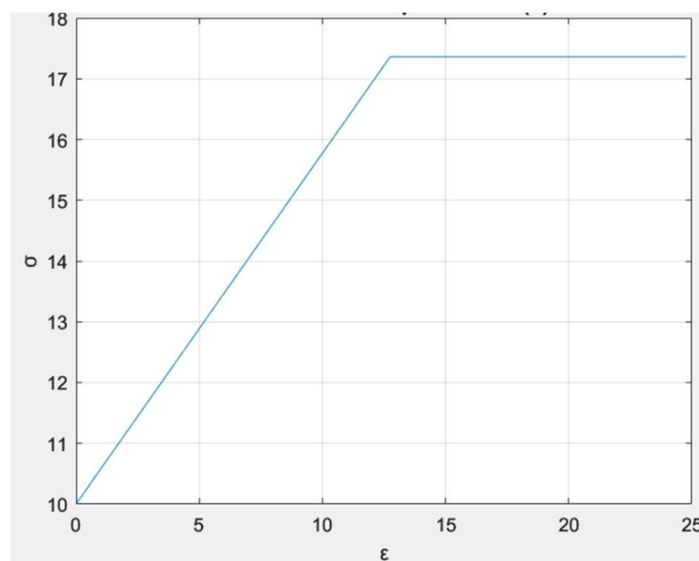


Fig. 9 Example of generated soil i(1).

The total number of hypothetical soils generated will be 10,000, achieved by combining the values of ϕ , c , and the corresponding stress and deformation levels. This will create a robust dataset for subsequent training and validation of the neural network. This training data are shown in Fig. 9.

In Example 2, the neural network architecture largely remains the same as in Example 1, except the number of input nodes is increased to three. These will represent deformation (ϵ), internal friction angle (ϕ), and cohesion (c). This modified structure is shown in Fig. 10.

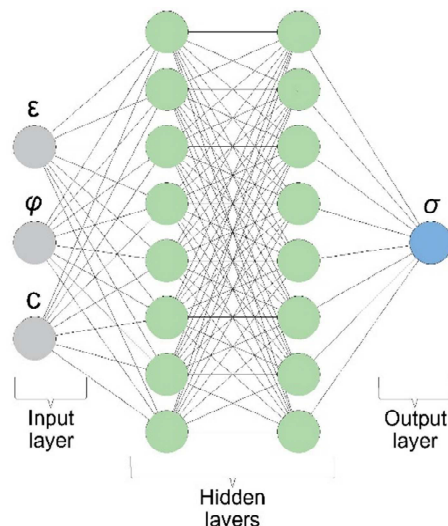


Fig. 10 The structure of a neural network 2.

The RMSE error fluctuated around 0.12 by the end of training. In Fig. 11, correct data points are shown in black, while predictions made by the network are highlighted in red. Errors were observed, with the most critical segment identified during the initial loading phase, further detailed in Fig. 12. Future iterations of this work plan to focus on mitigating these high-error areas.

The increased error is partly attributed to the fact that the network had to generate predictions for stress-deformation values not included in the training set, similar to Example 1. Additional complexity arose when geotechnical characteristics ϕ and c were introduced into the model within the ranges of ϕ (30 to 35) and c (10 to 15) used during the training phase. However, the prediction values were within but not part of the training set. For a better understanding, these data are displayed in the Tab. 2.

Tab. 2 Data comparison example 2.

ϕ learning	c learning	ϕ predicting	c predicting
ϕ [30; 30.05; 30.10; 30.15 ... 35]	c [10; 10.05; 10.10; 10.15 ... 15]	ϕ [30.025; 30.075; 30.125 ... 34.975]	c [10.025; 10.075; 10.125 ... 14.975]

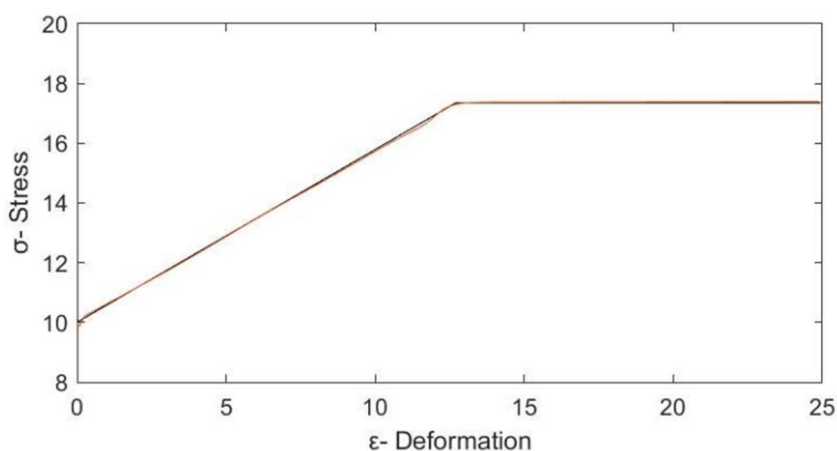


Fig. 11 Neural network result 2– red predictions, black correct answers.

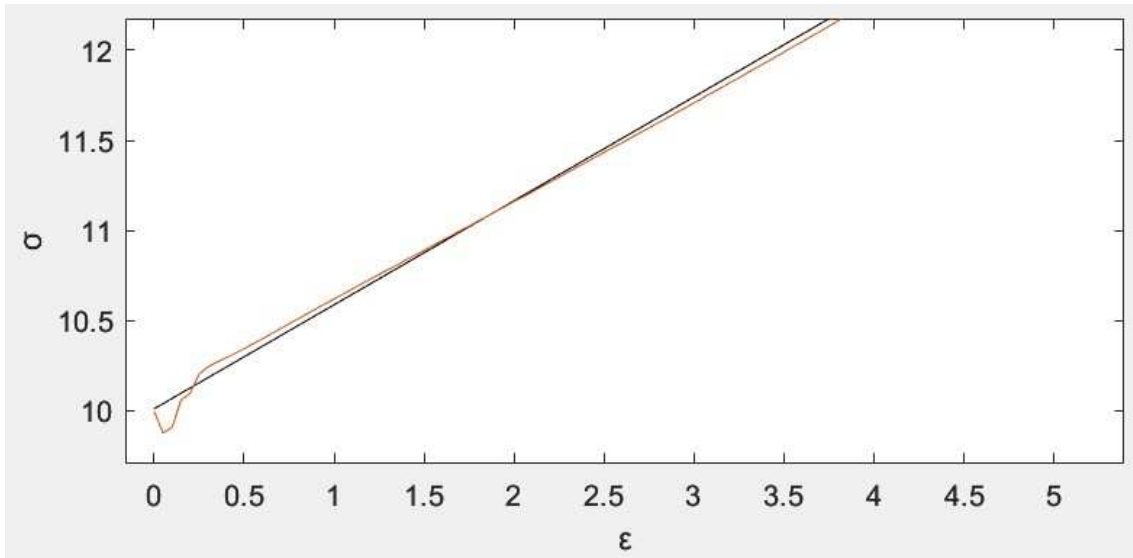


Fig. 12 Detail of the beginning of the graph example 2– red predictions, black correct answers.

Example 2 demonstrated that the trained neural networks can interpolate effectively within the soil region on which they were trained. However, the example also revealed local prediction issues that will need to be addressed or minimized in the next work.

Example 3

In Examples 1 and 2, the neural network directly predicted the relationship between deformation and stress. However, Example 3 differs by focusing on multi-dimensional stress trajectories rather than just the material's behaviour under stress or deformation changes. While the outcomes should be similar to Examples 1 and 2, the methodological approach will be different and adaptable to multi-dimensional issues.

Data generation for training in Example 3 will be like Example 1, with deformations (ϵ) and stresses (σ) ranging from 0 to 1. At a stress value of 0.5, the material will transition from elastic to plastic deformation. A total of 10,000 data points will be generated, which includes both stress values $\sigma(i)$ and corresponding changes in deformation $\Delta\epsilon$ and stress $\Delta\sigma$.

$$\Delta\sigma = NN(\sigma; \Delta\epsilon) \quad (7)$$

In Example 3, the neural network architecture which can be written as (3) remains similar to that used in Example 1, with the key difference being an increase to two input nodes. These nodes will represent the change in deformation ($\Delta\epsilon$) and the stress state (σ). This modified network structure is illustrated in Fig. 13.

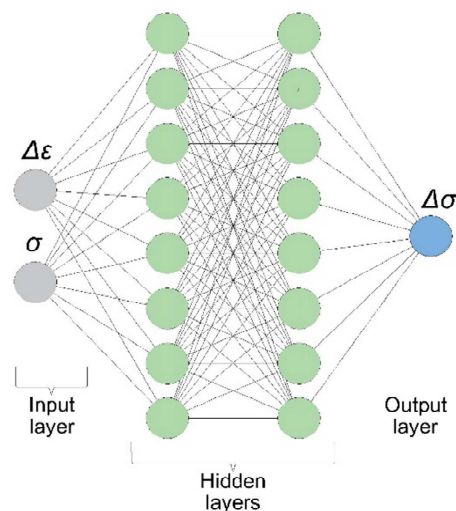


Fig. 13 The structure of a neural network 3.

The RMSE variability stabilized around 0.005 by the end of training. In Fig. 14 and Fig. 15, correct data points are marked in black, while predictions made by the neural network are highlighted in red. Errors are apparent, most notably in the initial loading phase. Future work aims to minimize these error zones.

This increase in error is partly due to the neural network having to predict stress-strain values not included in the training set. Further complications arose when stress state (σ) and change in deformation ($\Delta\varepsilon$) values were included in the training range of σ (from 0 to 0.5) and $\Delta\varepsilon$ (from 0 to 0.5) but were not part of the training set for prediction. For a better understanding, these data are displayed in the Tab. 3.

Tab. 3 Data comparison example 3.

σ learning	$\Delta\varepsilon$ learning	σ predicting	$\Delta\varepsilon$ predicting
$\sigma [0; 0.005; 0.01; 0.015 \dots 0.5]$	$\Delta\varepsilon [0; 0.005; 0.01; 0.015 \dots 0.5]$	$\sigma [0.0025; 0.0075; 0.0125; 0.0175 \dots 0.4975]$	$\Delta\varepsilon [0.0025; 0.0075; 0.0125; 0.0175 \dots 0.4975]$

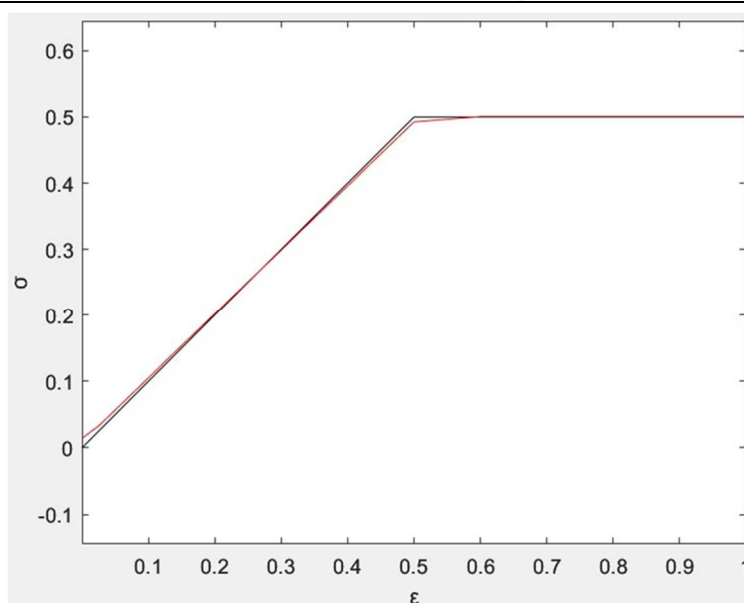


Fig. 14 Neural network result 3– red predictions, black correct answers.

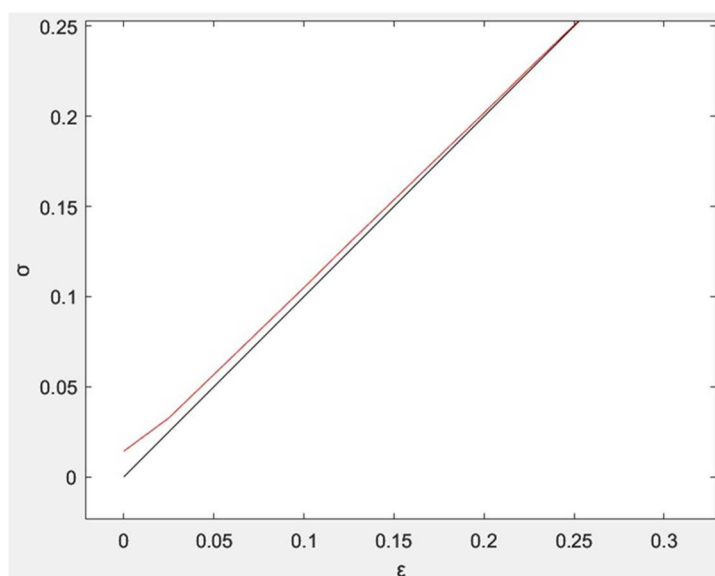


Fig. 15 Detail of the beginning of the graph example 3 – red predictions, black correct answers.

Example 3 empirically showed that the implemented neural network can approximate constitutive properties in one-dimensional cases. However, output analysis revealed local anomalies in the predictive models. These discrepancies represent critical areas that the dissertation will focus on eliminating or minimizing.

6 DISCUSSION

This research represents a preliminary investigation into the domain of applying neural networks to soil constitutive modelling, with a primary focus on assessing the neural network's ability to replicate the constitutive model of soils. Our findings serve as a foundational step in this innovative approach, underscoring the potential and limitations of neural networks in geotechnical engineering.

Neural Networks Versus Traditional Models

A key distinction between neural networks and traditional models lies in their foundational approach to soil behaviour modelling. Unlike traditional models, which are often hampered by the inherent errors of constitutive models, neural networks offer a more direct and adaptable approach. They can be trained using real soil test data, providing a practical learning method that potentially leads to models more reflective of actual soil behaviour under various conditions.

This capability of neural networks to learn from real-world data marks a significant advancement. Traditional models may rely heavily on theoretical assumptions or parameters derived from laboratory settings, which might not fully capture the complexities of soil behaviour in real-world scenarios.

Limitations and Future Research

While this study presents encouraging outcomes, it is essential to recognize its initial stage. One of the key challenges encountered was the neural network's limitations in processing data beyond its initial training set. This issue was particularly evident in replicating complex, multi-dimensional soil behaviours, underscoring the urgent need for an expanded and diversified training dataset.

An integral aspect of future research will be to incorporate real-world data from soil testing into the neural network's training process. Specifically, triaxial tests will be utilized, considering their comprehensive data provision capability. However, this approach comes with significant limitations due to the technical and time-intensive nature of triaxial testing, which can extend to an entire day or even more per test. For optimal neural network training, a high volume of tests – in the tens of thousands to millions – is preferable. This requirement presents a challenge: how to effectively incorporate such extensive real-world data into the learning process of neural networks.

The discrepancies observed in the initial loading phase when compared to traditional models also highlight an area ripe for improvement. Future endeavours will focus on developing more sophisticated neural network models. These models aim to more accurately simulate the initial behaviour of soil under load, moving beyond the current limitations and offering a more nuanced understanding of soil mechanics.

By addressing these limitations and focusing on these specific future research directions, we aim to enhance the accuracy and applicability of neural network models in soil behaviour analysis, paving the way for more robust and reliable predictions in geotechnical engineering.

7 CONCLUSION

This research paper embarked on an ambitious journey to integrate neural networks into constitutive modelling of soils, striving to overcome the limitations of traditional models like Hooke's law and the Mohr-Coulomb model. The main goal was to establish a neural network that could more accurately mimic the complex and nonlinear behaviour of soils, a pivotal factor in geotechnics and construction.

By ability to transform from linear to sophisticated nonlinear models, it paves the way for more efficient and accurate predictions in construction projects, potentially leading to optimized material use, cost reductions, and enhanced safety and sustainability.

However, it is important to recognize that this study is just the start. The successful replication of the constitutive model using a neural network sets the stage for more comprehensive research. Future work should focus on expanding the model's capabilities to handle multi-dimensional problems and a broader range of soil types and conditions, leveraging real-world data for training.

In conclusion, the use of neural networks in this context is a promising development, offering a new paradigm in soil constitutive modelling. As this technology continues to evolve, it has the potential to significantly enhance the accuracy and efficiency of predictions in geotechnics and construction, paving the way for more optimized and sustainable engineering solutions.

Acknowledgement

This work was supported by Grant National Agency VEGA of the Slovak Republic. Project number G1/0009/2023.

References

- [1] SMYRNIOU, Eleni. Soil Constitutive Modelling Using Neural Networks [online]. Master of Science Thesis. Department of Geo-Engineering, Civil Engineering and Geosciences, Delft University of Technology. Supervisors: Dr. P. J. Vardon, Dr. ir. F. C. Vossepoel, Dr. ir. T. Schweckendiek, TU Delft; Dr. J. Nuttall, Dr. B. Zuada Coelho, Deltas. 2018. Available at: <http://resolver.tudelft.nl/uuid:278a44fc-086c-4f7d-8d67-dfd07f12edbb>
- [2] KIM, Phil. MATLAB Deep Learning: With Machine Learning. Neural Networks and Artificial Intelligence [online]. Seoul: Soul-t'ukpyolsi, Korea. 2017. ISBN-13 (pbk): 978-1-4842-2844-9. ISBN-13 (electronic): 978-1-4842-2845-6. DOI 10.1007/978-1-4842-2845-6
- [3] S. N. Sivanandam, S. Sumathi, S. N. Deepa. Introduction to Neural Networks Using Matlab 6.0. New Delhi: Tata McGraw-Hill Publishing Company Limited 2006. ISBN 0-07-059112-1
- [4] CHOI, Rene Y., COYNER, Aaron S., KALPATHY-CRAMER, Jayashree, CHIANG, Michael F., CAMPBELL, J. Peter. Introduction to Machine Learning, Neural Networks, and Deep Learning. *Translational Vision Science & Technology* [online]. February 2020, Vol.9, No. 2. DOI 10.1167/tvst.9.2.14. Available at: <https://tvst.arvojournals.org/article.aspx?articleid=2762344>
- [5] DORNHEIM, Johannes, MORAND, Lukas, NALLANI, Hemanth Janarthanam, HELM, Dirk. Neural Networks for Constitutive Modeling – From *Universal Function Approximators to Advanced Models and the Integration of Physics* [online]. 2023 [accessed 2024-01-13]. Available at: <https://arxiv.org/abs/2302.14397>
- [6] Ivo Herle. Základy matematického modelování v geomechanice. Nakladatelství Karolinum. Praha. 2003. ISBN 80-246-0745-X
- [7] BORJA, Ronaldo I., AYDIN, Atilla. Computational modeling of deformation bands in granular media. I. Geological and mathematical framework. *Computer Methods in Applied Mechanics and Engineering* [online]. 2004 [Accessed 2024-01-13]. Vol. 193, Issues 27–29, pp. 2667–2698. ISSN 0045-7825. DOI 10.1016/j.cma.2003.09.019
- [8] DRUCKER, Daniel C., GIBSON, Robert E., HENKEL, David J. Soil Mechanics and Work-Hardening Theories of Plasticity. *Transactions of the American Society of Civil Engineers* [online]. 1955 [accessed 2024-01-13]. Vol. 122, pp. 338–346. Available at: <https://api.semanticscholar.org/CorpusID:138053635>
- [9] THE MATHWORKS, INC. Deep Learning Toolbox™ Getting Started Guide. [online] Natick, MA: The MathWorks, Inc. 2022 [accessed 2024-01-13]. Available at: <https://www.mathworks.com>