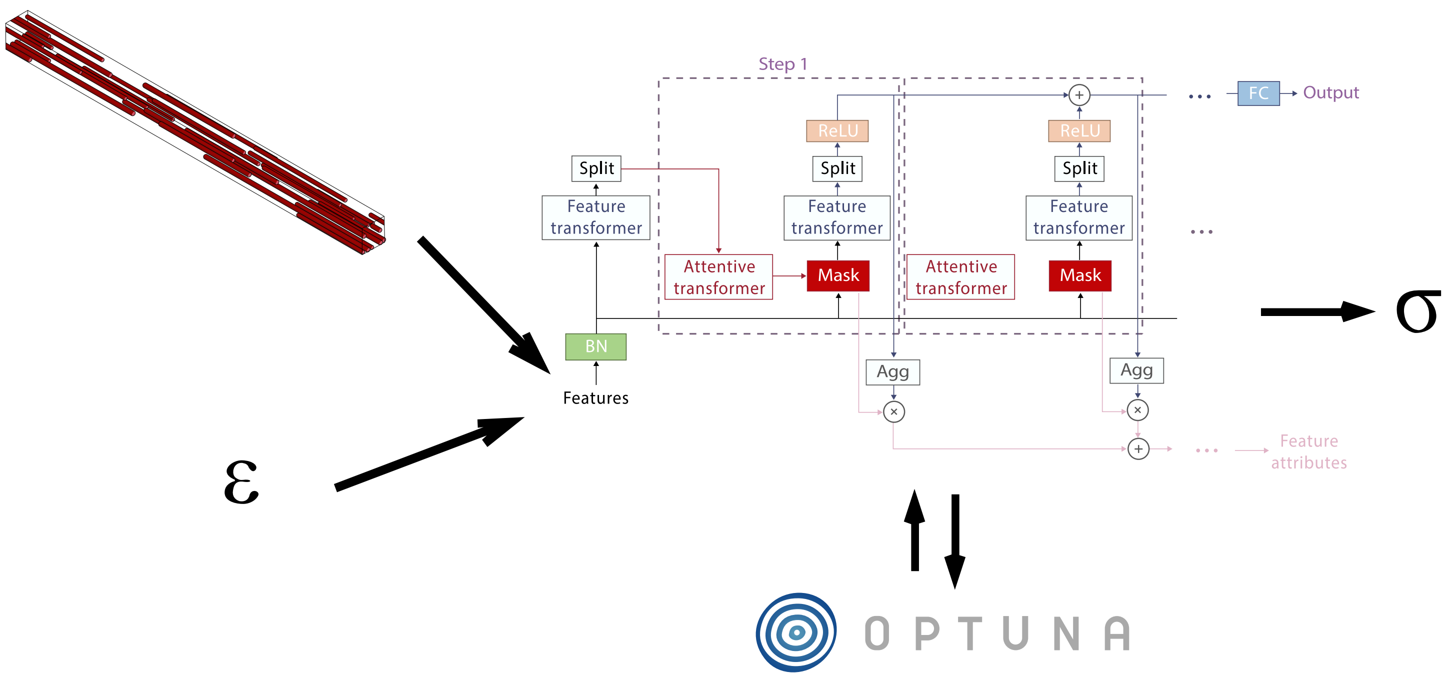
CHALMERS
UNIVERSITY OF TECHNOLOGY



Deep Learning-based Stress Prediction for Short Fiber Reinforced Composites Using TabNet

VARIATION PREDICTION BY TABNET

Master's thesis in Master Material Engineering

CHAO WU

DEPARTMENT OF IMS
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

MASTER'S THESIS 2026

Deep Learning-based Stress Prediction for Short Fiber Reinforced Composites Using TabNet

VARIATION PREDICTION BY TABNET

CHAO WU

Mechanics and Physics
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026

Deep Learning-based Stress Prediction for Short Fiber Reinforced Composites Using TabNet
VARIATION PREDICTION BY TABNET
CHAO WU

Supervisor and Examiner: Mohsen Mirkhalaf, Mechanics and Physics, Gothenburg University

Master's Thesis 2026
Mechanics and Physics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2026

Abstract

This thesis investigates how deep learning models can predict stress uncertainties in short fiber-reinforced composites (SFRCs). The central question is whether machine learning models, particularly TabNet, can accurately predict stress variations within SFRCs under different fiber distributions. To address this, full-field data were generated using Digimat and expanded through data augmentation to train the TabNet model. Model performance was evaluated using root mean square error (RMSE), and the results show that the TabNet model effectively predicts stress variations across different realizations of representative volume elements (RVEs). The model captures the stress uncertainties arising from the microstructural variability of the material while maintaining high accuracy. This study demonstrates that combining TabNet with data augmentation significantly reduces the computational resources required for traditional full-field simulations while providing accurate predictions of stress uncertainties in SFRCs, highlighting its potential applications in composite material design and manufacturing.

Keywords: Short Fiber Reinforced Composites, Deep Learning, TabNet, Uncertainty, Stress Prediction, Data Augmentation, Full-field Simulation.

Acknowledgements

First of all, I would like to express my deepest gratitude to Professor Mohsen Mirkhalaf. His unwavering support and valuable feedback have greatly improved the quality of this work. His meticulous attention to detail is an important source of motivation for me to move forward. Secondly, I would like to thank my friend Wentao Wang for his continued support and patience. This journey would not have been possible without his unwavering encouragement

Chao Wu, Gothenburg, 05,08,2024

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

ADAM	Adaptive Moment Estimation
AutoML	Automated machine learning
AVG	Average
FE	Finite Element
FFT	Fast-Fourier-Transforms
MSE	Mean Squared Error
RNN	Recurrent neural network
RVE	Representative Volume Element
SFRCs	Short Fiber Reinforced Composites
STD	Standard Deviation
FFT	Fast-Fourier-Transform

Contents

List of Figures

List of Tables

1

Introduction

Composites are engineered from two or more constituent materials with significantly different physical or chemical properties [1]. These materials work synergistically to provide unique macroscopic properties, leading to widespread applications in industries such as aerospace, automotive, construction, and marine engineering [2, 3, 4, 5]. The primary advantages of composites include high specific strength, high stiffness, low weight, and excellent corrosion resistance [6].

Short Fiber Reinforced Composites (SFRCs) have become increasingly important in modern engineering applications because they combine strong mechanical performance with flexible manufacturing routes [7]. SFRCs consist of a polymer matrix reinforced by discontinuous fibers [8]. These fibers can be oriented in specific or random patterns to significantly enhance the mechanical performance of the resulting composite. SFRCs offer distinct advantages over continuous fiber-reinforced polymers: they are more cost-effective to produce as they do not require complex manufacturing machinery [9, 10]; they can exhibit quasi-isotropic properties; and they are highly suitable for injection molding into complex geometries [11].

The mechanical performance of SFRCs is primarily governed by fiber orientation, fiber length, and volume fraction. Generally, a higher fiber concentration leads to higher load-bearing capacity, as the fibers carry the majority of the external force rather than the matrix [12]. When fibers are aligned in a specific direction, SFRCs exhibit anisotropic behavior, with superior mechanical performance along the alignment axis [13]. However, even with identical fiber alignment and volume fractions, local variations in fiber distribution can cause significant stress fluctuations [14]. This underscores the critical importance of investigating fiber distribution to achieve optimal and predictable mechanical properties.

By analyzing Representative Volume Elements (RVEs) at the microscale, computational homogenization can provide accurate estimates of the effective material response. In this study, the mechanical response is characterized by the longitudinal normal stress, denoted as S_{11}. As illustrated in Figure 1.1, different stochastic realizations of an RVE, even those sharing the same macroscopic parameters, can result in variations in the S_{11} response. This discrepancy highlights the inherent uncertainties in RVE simulations that this study aims to investigate. RVE simulation has become an essential tool in both science and industry, enabling the study of material properties across scales (from micro to macro) and reducing the reliance on costly, time-consuming experimental cycles during early-stage material development.

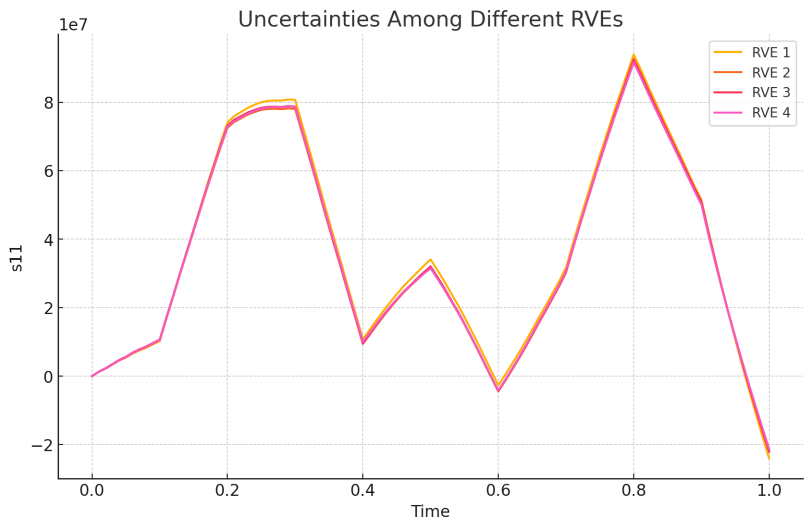


Figure 1.1: Variations in the longitudinal normal stress component (S_{11}) over time across four distinct RVE realizations. Each line represents a different stochastic fiber distribution, highlighting the uncertainty despite identical input parameters.

Table 1.1: Input parameters and the resulting peak stress for a sample RVE realization. The fiber geometry is modeled as a cylinder to avoid numerical stress concentrations at the tips. Fiber alignment is represented through the individual entries of the second-order orientation tensor a_{ij}.

Orientation Tensor (a_{ij})	Fiber VF (ϕ)	Fiber Diameter (d)	Fiber Shape	Strain (ε_{11})	Peak Stress (S_{11})
diag(1, 0, 0)	0.14	1×10^{-5} m	cylinder	4.2×10^{-3}	6.47×10^{6} Pa

To connect microscale descriptions with the effective response observed at the macroscopic level, homogenization techniques are commonly applied. Among these approaches, full-field (FF) simulations [16] based on Finite Element (FE) analysis [18] or Fast Fourier Transforms (FFT) [20] resolve the local microstructural fields and can therefore deliver detailed, high-fidelity predictions. In contrast, mean-field (MF) methods [15] offer computational efficiency but often suffer from lower fidelity due to simplifying assumptions. While FF methods are accurate, they are computationally expensive, often requiring hours or days to solve depending on the complexity [21]. This computational bottleneck is a major obstacle to deploying multi-scale FE analysis in real-world industrial production.

To address this problem, deep learning-based surrogate models have been introduced. These models, once properly trained, can accurately and instantaneously predict the mechanical output based on microstructural inputs. A successful example is found in the modeling of progressive damage behavior in large composite structures, where a pre-computed surrogate model based on an artificial neural network was used to represent complex nonlinear stress-strain relationships [22]. Similarly, Wu et al. [24] developed an improved variant of the Recurrent Neural Network (RNN) [25], known as the Gated Recurrent Unit (GRU), to serve as a surrogate model for micromechanical analysis. They applied this model to 2D RVEs of continuous fiber-reinforced elastoplastic matrix composites subjected to random strain path loadings.

Building upon these research foundations, this study utilizes a TabNet-based model [31] to predict the statistical uncertainties of unidirectional composites across different RVE realizations. The workflow involves using Digimat-FE to generate high-fidelity RVE datasets via the full-field method. Subsequently, a data augmentation script is employed to expand the training set and enhance model robustness [23]. The results indicate that this approach can significantly reduce computational costs while achieving high predictive accuracy for the homogenized response.

The subsequent chapters are structured to follow the main workflow of the study. Chapter 2 presents the relevant micromechanical background and the principles of the TabNet architecture. Chapter 3 describes how the simulation data are generated and expanded through augmentation. Chapter 4 focuses on the network design, training strategy, and evaluation procedure. The obtained results and their interpretation are discussed in Chapter 5, while Chapter 6 summarizes the main findings and outlines possible directions for future work.

2

Theory

This chapter introduces the micromechanical concepts and the TabNet model components needed for the present stress-uncertainty prediction task. The emphasis is placed on how each concept is used in this thesis, rather than on a complete review of the underlying numerical or machine-learning literature.

2.1 Modeling Methods

The numerical data in this thesis are generated from representative volume elements (RVEs). An RVE is a microscale sample that contains enough information about the fiber distribution to represent the effective response of the composite at a larger scale. In this work, each RVE is used as a controlled realization of a short-fiber microstructure, and the differences between realizations are used to quantify uncertainty in the stress response.

2.1.1 RVE And Full-field simulation

Full-field simulations are used because they resolve the local response inside the RVE instead of only estimating averaged behavior. After the RVE geometry, material properties, and loading path are defined, the numerical solver computes the local fields and the homogenized stress response. This response is then used as training data for the surrogate model. Figure 2.1 shows an example of a generated RVE by Digimat-FE.

In FE analysis, the RVE domain is divided into finite elements, and the element-level equations are assembled into a global system that represents the prescribed boundary value problem. In the present study, this procedure provides the reference micromechanical response for the selected unidirectional RVE configurations. Further details on RVE boundary conditions and numerical implementation can be found in [27].

2.1.2 Orientation Tensors

Fiber orientation is represented by a unit vector $\mathbf{p}$, described by the angles θ and ϕ, as shown in Figure 2.1.

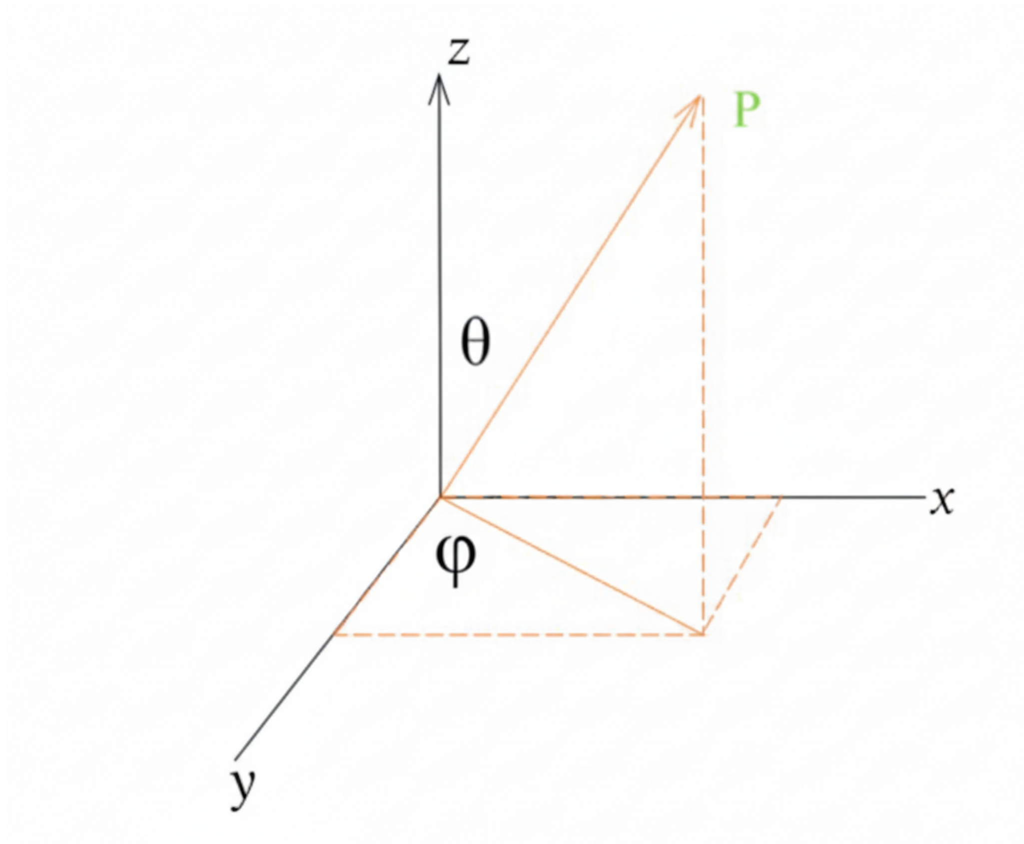


Figure 2.1: Unit vector **p** defined by θ and ϕ. *Adapted from Mentges et al. [28].*

$$\boldsymbol{p} = \left[\sin(\theta)\cos(\phi) \;\; \sin(\theta)\sin(\phi) \; \cos(\theta)\right]^T. \tag{2.1}$$

Because each inclusion in an RVE may have its own orientation, the orientation distribution function (ODF) $\psi(\boldsymbol{p})$ describes the probability of finding fibers in a given direction. Since the two directions **p** and $-$**p** represent the same fiber axis, the distribution satisfies

$$\psi(\boldsymbol{p}) = \psi(\text{-}\boldsymbol{p}), \tag{2.2}$$

and the total probability over all orientations is normalized as

$$\oint \psi(\boldsymbol{p})\, d\boldsymbol{p} = 1. \tag{2.3}$$

For the data used here, the RVE generation is controlled by a second-order orientation tensor rather than by the full ODF:

$$a_{ij} = \oint p_i p_j \psi(\boldsymbol{p})\, d\boldsymbol{p}. \tag{2.4}$$

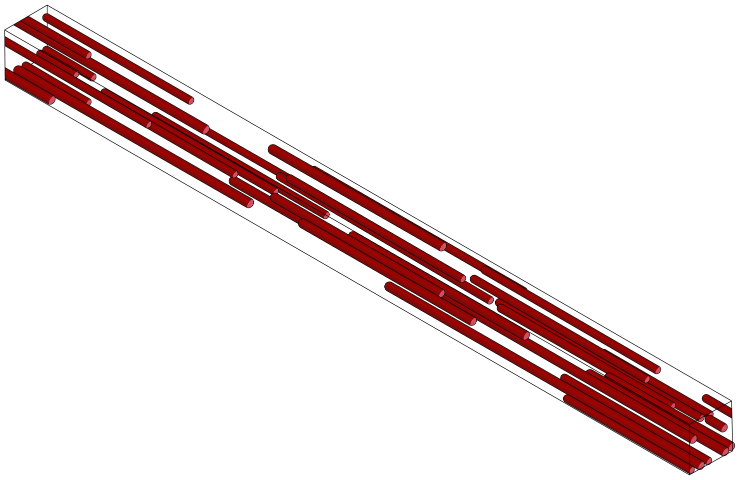


Figure 2.2: Example of a generated RVE.

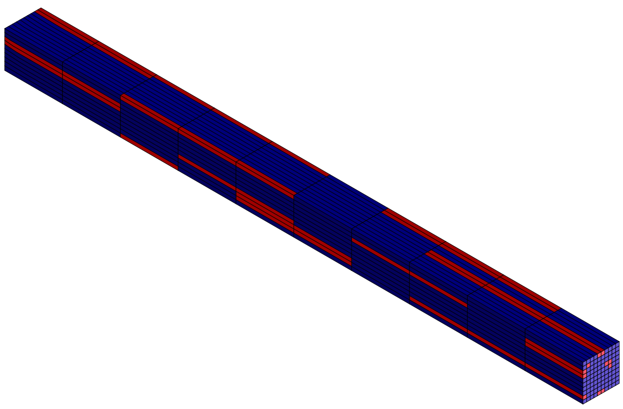


Figure 2.3: Voxel-based mesh of the RVE. The domain is discretized into a regular grid of hexahedral elements to support full-field computation.

FFT-based solvers provide another route for micromechanical full-field analysis [29]. The distinction relevant to this thesis is practical: FE simulations provide detailed local fields through an explicit mesh, while FFT-based approaches can be efficient for voxelized microstructures. This thesis uses the generated full-field response as high-fidelity training information and does not benchmark FE against FFT; further discussion can be found in [30].

2.2 TabNet

TabNet is used in this thesis as a tabular surrogate model for mapping RVE descriptors and strain histories to statistical stress responses [31]. The input vector contains the orientation tensor components, fiber volume fraction, and strain components at each time step. The targets are the mean and standard deviation of the six stress components. Therefore, the model is not used as a general-purpose classifier; it is used to learn which mechanical descriptors are most informative for a particular loading state and then produce continuous stress predictions.

2.2.1 Feature selection

Feature selection is useful in this problem because the dominant mechanical input can change along the loading path. For example, one strain component may dominate during a nearly uniaxial segment, while shear components become more important during coupled loading. TabNet represents this changing importance through a learned mask:

$$M[i] \in \mathbb{R}^{B\times D}. \tag{2.5}$$

Here, B is the batch size and D is the number of input features. The mask is applied directly to the feature vector:

$$M[i] \cdot f. \tag{2.6}$$

At decision step i, the mask is computed from information carried from the previous step:

$$M[i] = \mathrm{sparsemax}(P[i-1] \cdot h(a[i-1])). \tag{2.7}$$

In this expression, $h(a[i-1])$ is the transformed representation from the previous step and $P[i-1]$ controls how strongly previously used features remain available. The sparsemax function [33] is important because it can set some feature weights exactly to zero, producing a compact feature selection instead of a dense weighting of all inputs.

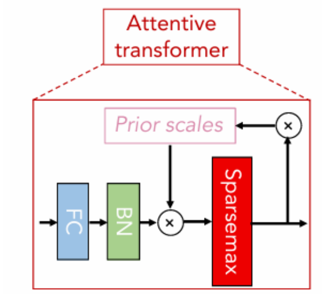


Figure 2.4: Attentive transformer block used in TabNet. The prior scale influences the next feature mask, while sparsemax encourages a small set of active features. *Adapted from Arik and Pfister [31].*

Feature reuse is controlled by the prior scale term:

$$P[i] = \prod_{j=1}^{i} (\gamma - M[j]). \tag{2.8}$$

The relaxation parameter γ determines how easily a feature can be reused. In the stress-prediction problem, this is relevant because some quantities, such as fiber volume fraction, may remain informative throughout the full path, while individual strain components may only be important during specific loading segments. Initially, $P[0]$ is set to one so that no feature is penalized before the first decision step.

2.2.2 Feature processing

A feature transformer processes the masked input and separates the representation into two parts: one part contributes to the current prediction, and the other is passed to the next decision step:

$$[d[i], a[i]] = f_i(M[i] \cdot f). \tag{2.9}$$

Here, $d[i]$ supports the current stress prediction and $a[i]$ carries information forward:

$$d[i] \in \mathbb{R}^{B \times N_d} \text{ and } a[i] \in \mathbb{R}^{B \times N_a}. \tag{2.10}$$

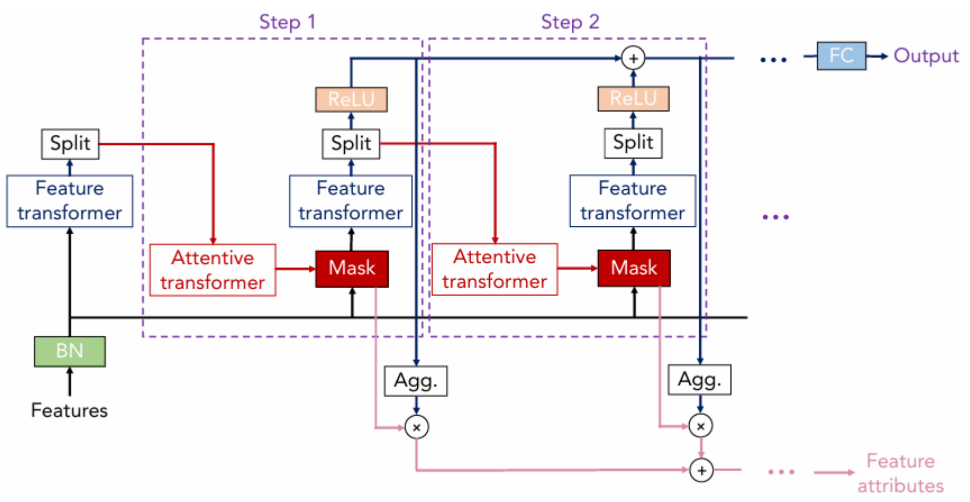


Figure 2.5: TabNet decision-step structure. Masked features are processed and split into prediction information and information passed to the following attentive transformer. *Adapted from Arik and Pfister [31].*

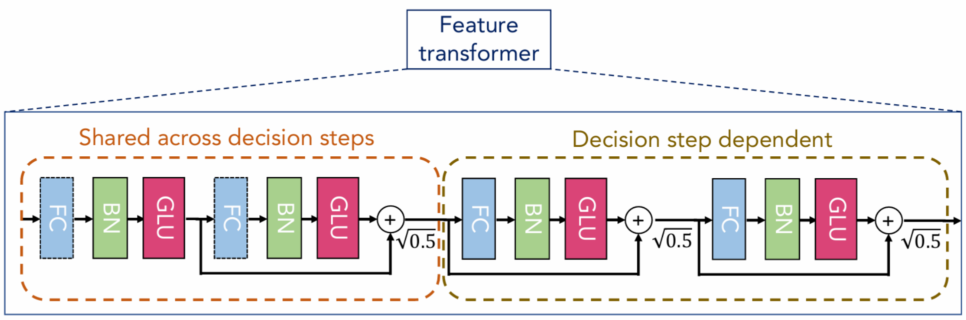


Figure 2.6: Example feature transformer block with shared and decision-step-specific layers. *Adapted from Arik and Pfister [31].*

For this thesis, the practical role of the feature transformer is to build a compact mechanical representation from the selected input features. Shared layers allow common transformations to be reused across decision steps, while step-specific layers let the model refine the representation as the loading state changes. The transformed outputs are aggregated as

$$d_{\text{out}} = \sum_{i=1}^{N_{\text{steps}}} \text{ReLU}(d'[i]). \tag{2.11}$$

The aggregated vector d_{out} is then passed to the final regression layer. Because the present task is stress regression rather than classification, the final layer is interpreted as a continuous mapping from the learned representation to the stress statistics.

2.2.3 Optimizing

The Adam optimizer [38] is used to train the TabNet models. Adam updates each parameter using estimates of the first and second moments of the gradient, which helps stabilize training when different input features and stress components have different numerical scales:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t. \tag{2.12}$$

Here, η is the learning rate, $\hat{m}_t$ is the bias-corrected first moment estimate, $\hat{v}_t$ is the bias-corrected second moment estimate, and ϵ prevents division by zero. The learning rate and weight decay are later tuned separately for the AVG and STD models because the mean stress and stress fluctuation have different target distributions.

3

Data Generation And Augmentation

To study uncertainties, various fiber distributions and volume fractions are considered. The constitutive properties of the fiber and matrix are provided in [39, 40].

Table 3.1: Fiber and matrix properties used in this study.

Property	**Fiber**	**Matrix**
Young's modulus	76 GPa	3.1 GPa
Poisson's ratio	0.22	0.35
Length	240 μm	-
Diameter	10 μm	-
Fiber volume fraction	0.1-0.15	-
Yield strength	-	25 MPa
Linear hardening modulus	-	150 MPa
Hardening modulus	-	20 MPa
Hardening exponent	-	325

J2 plasticity is used for the matrix. Detailed information about the matrix and fiber is listed in Table 3.1. The dataset is built from unidirectional RVE simulations because this thesis focuses on the uncertainty caused by different fiber distributions under a fixed alignment assumption. For each selected fiber volume fraction, a random 6D strain path is applied and Digimat-FE is used to compute the micromechanical stress response.

3.1 Uni-directional RVE's Orientation Tensor

To generate a unidirectional (UD) composite RVE, the microstructure is defined such that the fibers are aligned along the longitudinal axis. In terms of the second-order orientation tensor $\mathbf{a}$, this configuration is represented by setting the first diagonal component a_{11} to 1 and all other components to 0. This choice isolates the effect of stochastic fiber placement from the effect of changing the global fiber-orientation state. Detailed methodology regarding tensor sampling can be found in [39].

3.2 Random Strain Path Generation

The strain paths are generated using the procedure of Friemann et al. [39, 41]. In this thesis, the purpose of these paths is to expose the model to coupled loading states rather than only monotonic uniaxial loading. The number of drift directions is fixed to 10, which gives a varied path while keeping all simulations comparable. Each path contains 100 time steps. The drift directions are sampled from a standard normal distribution, perturbed with a noise term scaled by γ, accumulated over time, and rescaled so that the maximum strain reaches the selected value of $\varepsilon_{\max}$. Table 3.2 summarizes the parameters used in this thesis. Figure 3.1 shows one generated 6D strain path, where each curve corresponds to one strain component.

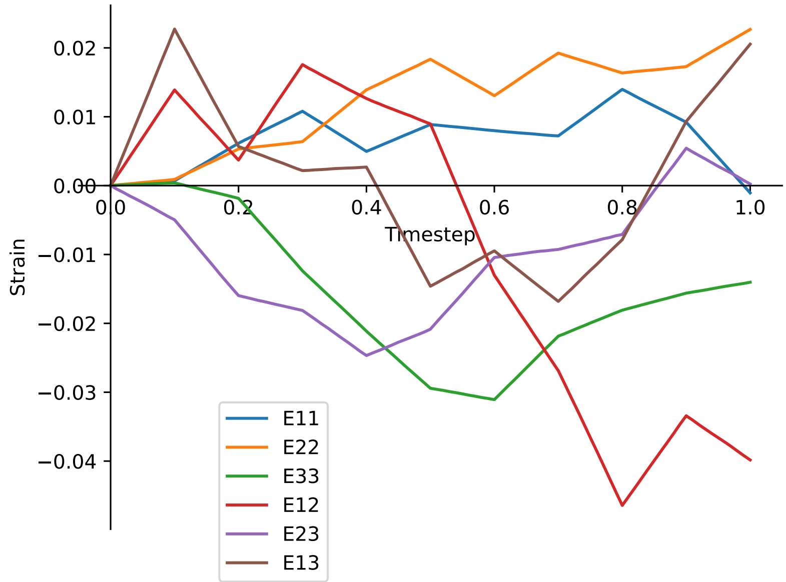


Figure 3.1: Example of a generated 6D random strain path. Different lines represent strain components evolving over 100 time steps.

Table 3.2: Parameters used for random strain path generation.

Parameter	Distribution	Range/Value
n – TimeSteps	-	100
n – Drift	uniformly sampled	(1, 2, 5, 10)
γ	uniformly sampled	(0–1)
$\varepsilon_{\max}$	uniformly sampled	(0.01–0.05)

3.3 Post-Processing

After the RVE simulations, the orientation tensor, fiber volume fraction, strain history, and micromechanical stress response are extracted into a table used by the learning pipeline. The original simulation set contains 10 fiber volume fractions, listed in Table 3.3.

For each volume fraction, 20 fiber-distribution realizations are simulated. The mean and standard deviation of the six stress components are then computed across these realizations. These two statistics define the AVG and STD targets used by the neural-network models.

Table 3.3: Average and standard deviation of stress components (s_{11}, s_{22}, s_{33}, s_{12}, s_{23}, s_{13}) over timesteps for different RVE configurations.

Fibre Volume Fraction
0.105
0.110
0.115
0.120
0.125
0.130
0.135
0.140
0.145
0.149

The 200 original simulations are sufficient for estimating the stress variation within each volume-fraction group, but they are too limited for training a robust deep-learning surrogate. Figure 3.1 illustrates the variation of stress components for one fiber volume fraction, while Figure 3.2 summarizes the corresponding mean and standard deviation. Ten representative realizations are shown for visual clarity, although all 20 realizations are used in the calculations. The data augmentation step below expands this original simulation set before model training.

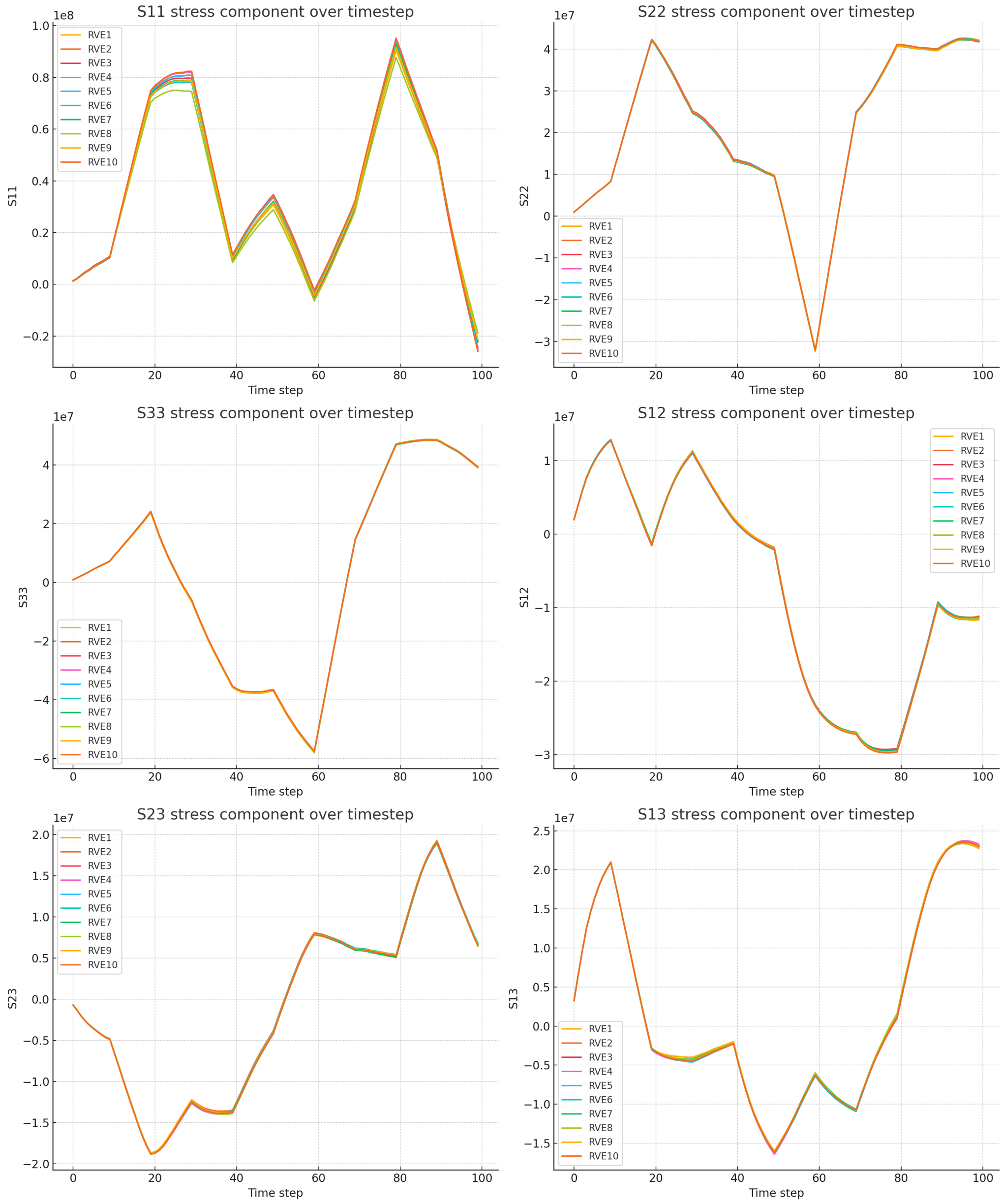


Figure 3.2: Variation of stress components (s_{11}, s_{22}, s_{33}, s_{12}, s_{23}, s_{13}) over timesteps for different RVE configurations.

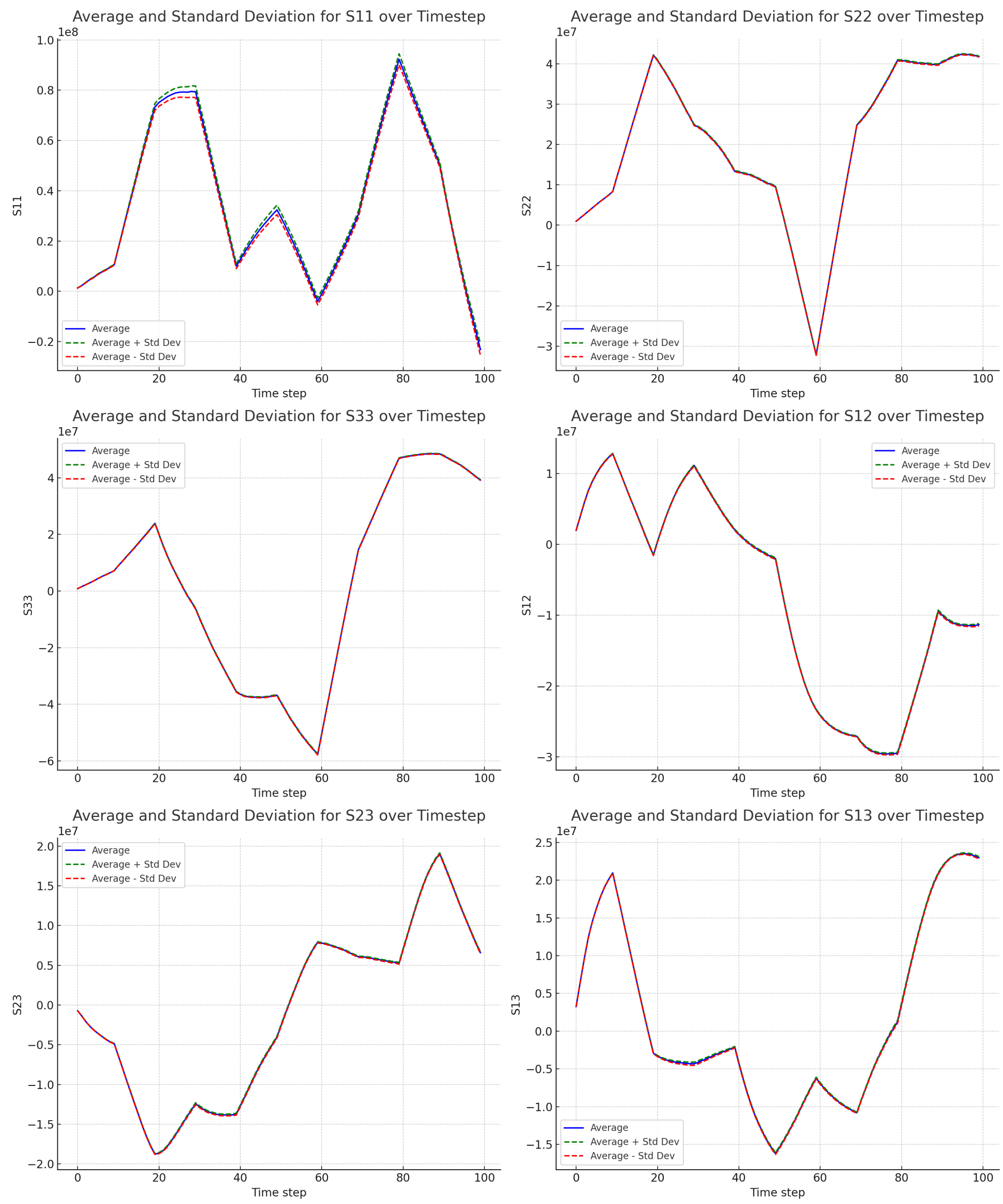


Figure 3.3: Average and standard deviation of stress components (s_{11}, s_{22}, s_{33}, s_{12}, s_{23}, s_{13}) over timesteps for different RVE configurations.

Figure 3.4: Ten representative realizations for one fiber volume fraction.

3.4 Data Augmentation Method

The dataset is expanded using the tensor-rotation augmentation method proposed by Cheung et al. [23]. In this thesis, the method is used to increase the number of training samples without running additional Digimat-FE simulations. Each original orientation, strain, and stress tensor is first written in matrix form:

$$\mathbf{a} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} \\ \varepsilon_{21} & \varepsilon_{22} & \varepsilon_{23} \\ \varepsilon_{31} & \varepsilon_{32} & \varepsilon_{33} \end{bmatrix}, \quad \boldsymbol{\sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}. \tag{3.1}$$

Random rotation matrices are generated using Arvo's algorithm [42]. The first rotation uses an angle θ sampled from $[0, 2\pi]$:

$$\mathbf{R} = \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.2}$$

A unit vector $\mathbf{v}$ is then generated using $\phi \in [0, 2\pi]$ and $z \in [0, 1]$:

$$\mathbf{v} = \begin{bmatrix} \cos(\phi)\sqrt{z} \\ \sin(\phi)\sqrt{z} \\ \sqrt{1-z} \end{bmatrix}. \tag{3.3}$$

The corresponding Householder term is

$$-\mathbf{H} = 2\mathbf{v}\mathbf{v}^T - \mathbf{I}, \tag{3.4}$$

where **I** is the second-order identity tensor. The full rotation matrix is

$$\mathbf{M} = -\mathbf{H} \cdot \mathbf{R}. \tag{3.5}$$

The same rotation is applied to the orientation, stress, and strain tensors so that input and output remain mechanically consistent:

$$\begin{Bmatrix} \mathbf{a}_r \\ \sigma_r \\ \varepsilon_r \end{Bmatrix} = \mathbf{M} \cdot \begin{Bmatrix} \mathbf{a} \\ \sigma \\ \varepsilon \end{Bmatrix} \cdot \mathbf{M}^T. \tag{3.6}$$

This creates additional equivalent coordinate descriptions of the same mechanical response, as illustrated in Figure 3.4.

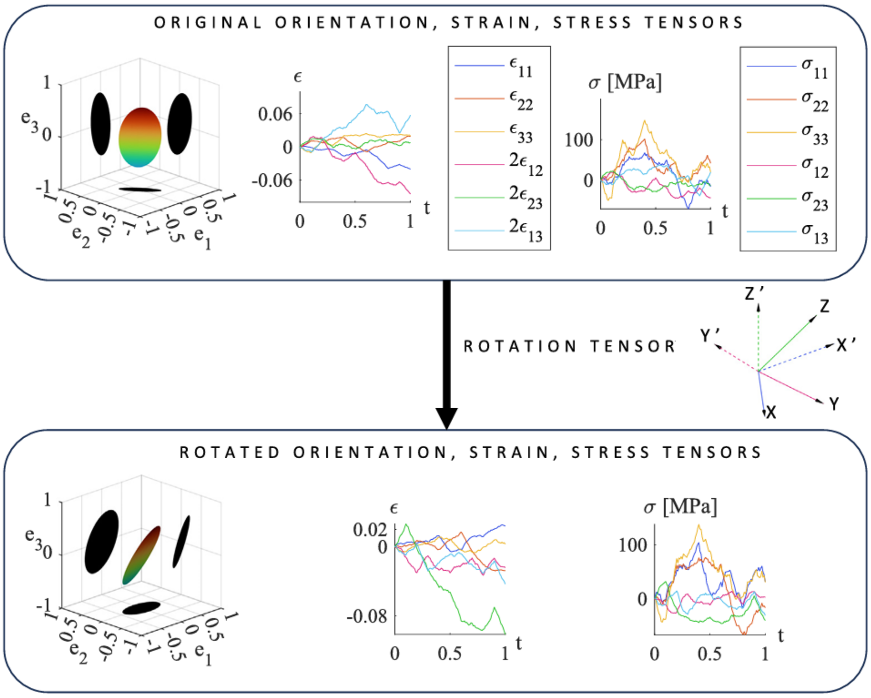


Figure 3.5: Tensor-rotation augmentation applied to the second-order orientation, strain, and stress tensors. *Reproduced from Cheung et al. [23].*

The 200 original sequences are augmented by a factor of 1,000, resulting in 200,000 sequences. Since each sequence contains 101 time steps, the final training table contains approximately 20 million rows. This expansion is the main reason the TabNet models can be trained on high-fidelity full-field information without requiring 200,000 separate RVE simulations.

For model training, the augmented dataset is split into training, validation, and test sets with a ratio of 80%, 10%, and 10%. The split is stratified by fiber volume fraction so that each subset preserves the same volume-fraction distribution as the full dataset.

4

Network Development

This chapter describes how the augmented RVE data are converted into two TabNet regression models. The choice of TabNet is motivated by the structure of the dataset: each row combines tabular descriptors of the microstructure with the strain state at a specific time step, and the target is a continuous stress statistic. The model therefore needs to handle coupled mechanical inputs while still allowing inspection of which features are emphasized during prediction.
RNN-based surrogate models have been used successfully for path-dependent composite data [39, 24]. In this thesis, TabNet is explored as an alternative because the training data are organized as large tabular sequences after augmentation, and because feature masks provide a direct way to analyze the influence of orientation, volume fraction, and strain components. The comparison in this work is therefore focused on the present AVG and STD stress-prediction task rather than on a general claim that TabNet is always superior to RNNs.

4.1 Architecture

To account for material uncertainties, the network must predict the mean stress response (μ) as well as the upper and lower bounds defined by the standard deviation (σ). Consequently, two distinct models were developed to predict the average (AVG) and standard deviation (STD) values separately, ensuring higher precision for each statistical measure.
The overall workflow is illustrated in Figure 4.1. Input data is first normalized using a Min-Max scaling method before being passed to a coordinator. This coordinator utilizes a thread pool executor to distribute the data to the AVG and STD models in parallel, allowing for simultaneous prediction. The models accept 13 input features (orientation tensors, fiber volume fraction, and strain components) and output 12 variables representing the predicted mean and standard deviation for each of the six stress components, as shown in Figure 4.2.
The Mean Squared Error (MSE) loss function was selected for training:

$$\text{Loss} = \frac{1}{2S} \sum_{i=1}^{S} \sum_{j=1}^{R} (t_{ij} - O_{ij})^2. \tag{4.1}$$

In this formulation, S denotes the sequence length and R represents the six stress components $s_{11}, s_{22}, s_{33}, s_{12}, s_{23}, s_{13}$. The variable t is the target value obtained from post-processing the simulations, and O is the model prediction. MSE is used because it penalizes large stress errors strongly and gives a single objective that can be used consistently for both AVG and STD models.

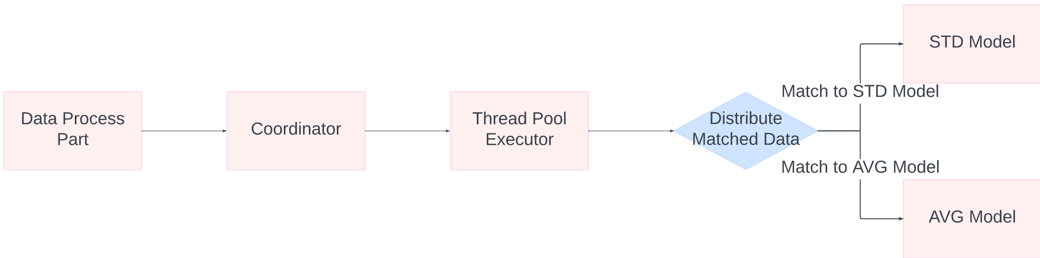


Figure 4.1: A data processing pipeline where data is processed, coordinated, and distributed to either the STD or AVG model using parallel execution.

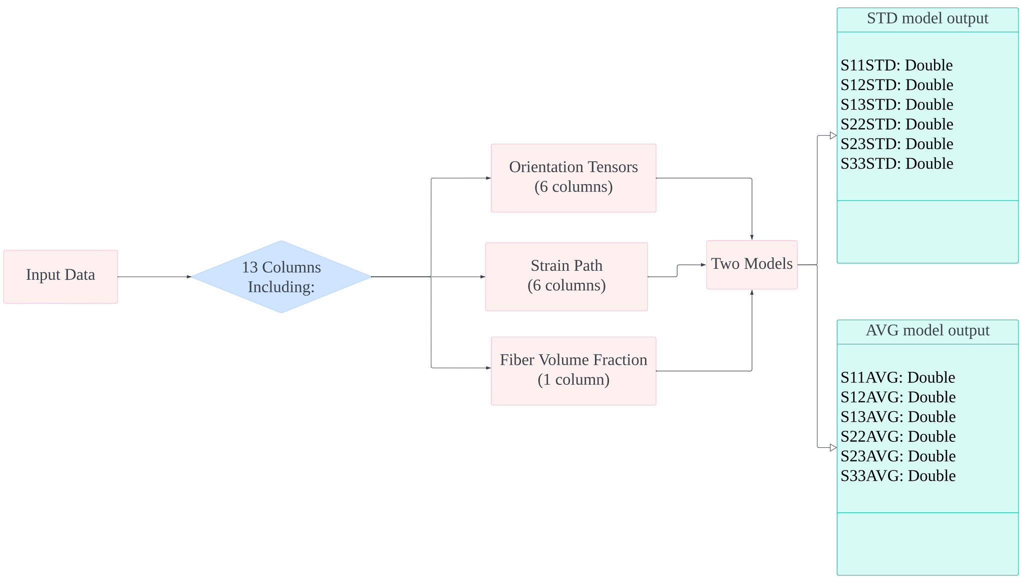


Figure 4.2: Diagram illustrating the prediction of input data into stress components S_{ij}, where $S11$, $S12$, etc., represent the stress in different directions. The STD model calculates the standard deviation (STD) for each component, while the AVG model calculates their average (AVG) values. The double means double-precision.

4.2 Training Parameters

The same MSE definition is used during hyperparameter tuning:

$$\text{loss} = \frac{1}{2S}\sum_{i=1}^{S}\sum_{j=1}^{R}(t_{ij} - O_{ij})^2. \tag{4.2}$$

Here, the symbols have the same meaning as in Section 4.1. Adam [38] is used to minimize the loss. The tuned optimizer-related parameters are the learning rate and weight decay.

CosineAnnealingLR [45] is used to vary the learning rate during training. The parameters $T_{\max}$ and $\eta_{\min}$ define the length of the cosine cycle and the lowest learning rate reached in that cycle. The TabNet representation sizes n_a and n_d, the number of decision steps, the batch size, and the optimizer parameters are tuned separately for the AVG and STD models. The batch-size range is chosen according to the available GPU memory. Detailed search ranges are listed in Table 4.1.

Table 4.1: AVG and STD Model Parameters.

Parameter	**AVG model**	**STD model**
batch size	5050 - 10100	5050 - 7070
eta_min	0.00021 - 0.001982	0.0002801 - 0.0019859
learning_rate	0.00521 - 0.019947	0.005067 - 0.0099971
n_d	8 - 32	16 - 256
n_a	8 - 32	16 - 256
weight_decay	0.00002316 - 0.00099839	1e-8 - 0.0008
n_steps	3 - 32	3 - 50

Since the whole dataset is huge, only a small portion (20%) of the dataset is used during hyperparameter search to reduce tuning time. Once the best hyperparameters are found, the full dataset is used to train the final model.

4.3 AutoML with Optuna

Optuna [46] is used to tune the model and training parameters. Each trial trains a candidate configuration and returns the validation loss, which is then used to guide the next suggestions.

The Tree-structured Parzen Estimator (TPE) [47] is selected as the sampling strategy. In practical terms, this allows the search to spend more trials near parameter regions that have already produced low validation error. The optimization workflow, including the interaction between the Optuna API and independent trial workers, is summarized in Figure 4.3.

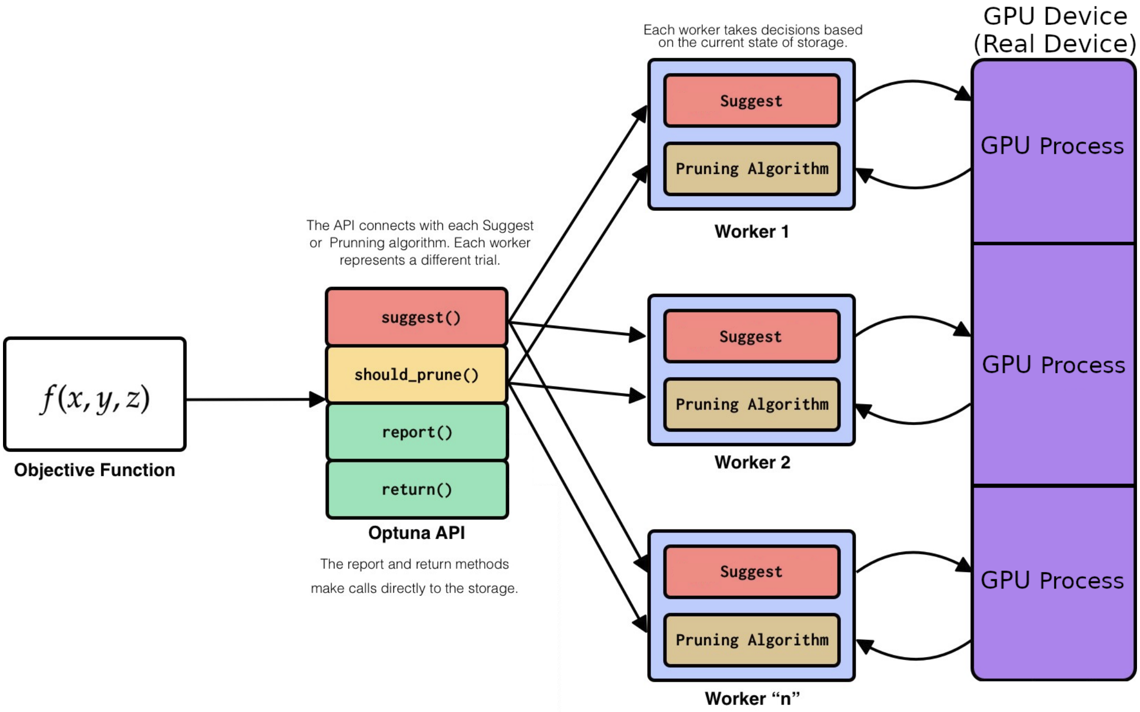


Figure 4.3: Workflow of the Optuna optimization framework used for hyperparameter search. *Adapted from Kaggle discussion material [48].*

Finally, a parallel coordinate figure of detailed trial information and a hyperparameter importance figure can be obtained. Thus, the best hyperparameters can be found. More possibilities can be explored in the future.

4.4 Network Performance Evaluation

4.4.1 Evaluation Metrics

To assess the fitting accuracy between the predicted stress values ($\hat{y}$) and the actual simulation results (y), the Root Mean Squared Error (RMSE) is employed:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}. \tag{4.3}$$

Here, n represents the total number of observations. RMSE provides a clear measure of the average error magnitude, with lower values indicating a higher degree of predictive accuracy.

4.4.2 General Loading Test

In the general loading test, the network is evaluated using random 6D strain paths. While the initial sampling for general paths in this specific dataset focuses on a strain range of 0 to 0.05 to establish a baseline for monotonic response, the complexity of the paths ensures that all six stress components are activated simultaneously.

4.4.3 Specific Loading Test

To further test the network's robustness, particularly its ability to handle cyclic loading and reverse plasticity, specific loading scenarios were designed. Unlike the general test, these cases explicitly include negative strain components. The loading cycle begins at 0, ramps up to 0.035, reverses through zero to a compressive strain of -0.035, and finally returns to 0. This cyclic approach allows for the evaluation of the model's performance under both tensile and compressive states across five scenarios: (i) uniaxial normal stress (σ_{11}), (ii) uniaxial shear stress (σ_{12}), (iii) biaxial normal stress, (iv) combined normal and shear stress, and (v) plane strain.

5

Results and Discussion

In this chapter, the predictive performance of the trained TabNet models is evaluated and discussed. The analysis begins with hyperparameter optimization and training convergence, followed by a comprehensive evaluation of the model's accuracy on the test dataset. Crucially, we compare the predicted stress evolution against the ground truth (Micro-mechanics simulations) for all six stress components. Finally, the advantages of the proposed TabNet architecture over traditional Recurrent Neural Networks (RNNs) are discussed.

5.1 Neural Network Training and Hyperparameter Analysis

The hyperparameter optimization process was conducted using Optuna, an automated framework that identifies the most influential parameters for model performance. The importance of each hyperparameter, as shown in Figure 5.1 for the AVG model and Figure 5.2 for the STD model, is determined using a functional Analysis of Variance (f-ANOVA) method. This technique quantifies the fraction of the objective function's variance that can be attributed to each hyperparameter.

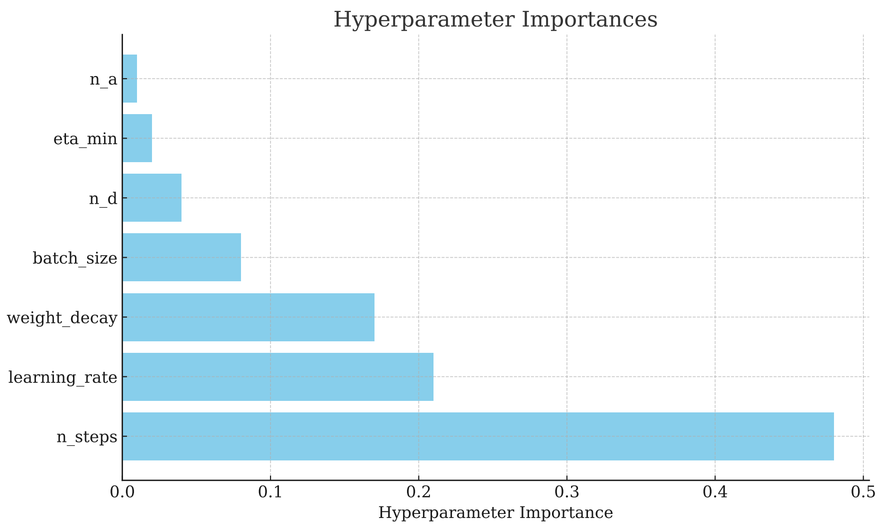


Figure 5.1: Hyperparameter importance scores for the AVG model, identifying *n_steps* and learning rate as the primary drivers of performance.

According to Figure 5.1, `n_steps` is the most critical hyperparameter for the AVG model, with an importance score close to 0.5. This suggests that the depth of the sequential attention mechanism in TabNet is vital for capturing the complexity of the stress-strain relationship. The `learning_rate` and `weight_decay` also exhibit significant influence, emphasizing the importance of optimization stability.

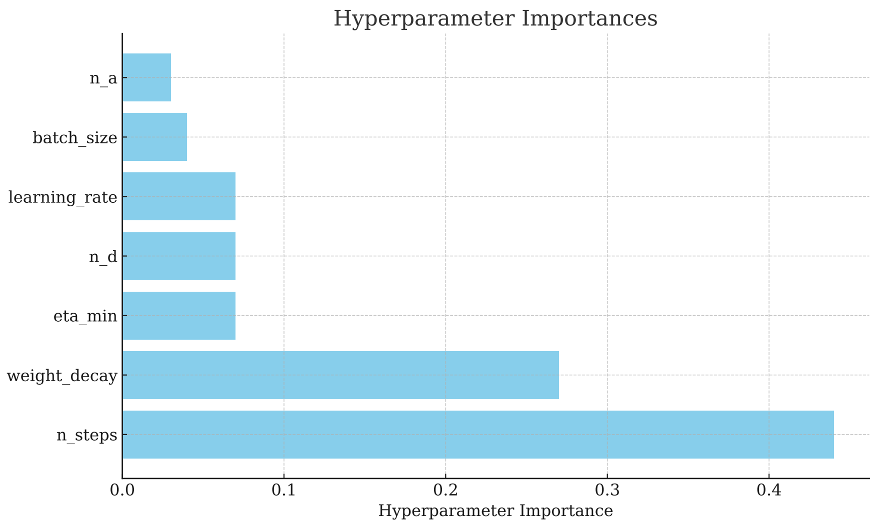


Figure 5.2: Hyperparameter importance scores for the STD model.

A similar trend is observed for the STD model in Figure 5.2, where `n_steps` remains the primary focus. The optimal hyperparameter configurations derived from these analyses are summarized in Table 5.1.

Table 5.1: Optimal hyperparameter configurations for the AVG and STD models determined by Optuna.

Parameter	**Best AVG model**	**Best STD model**
Batch size	6060	5050
eta_min	0.011283	0.014553
Learning_rate	0.018087	0.009960
n_d	16	64
n_a	32	128
Weight_decay	0.000996	0.000774
n_steps	7	9

Using these optimized settings, early stopping regularization was employed to prevent overfitting. The AVG and STD models reached their optimal states after 12 and 7 epochs, respectively, demonstrating rapid convergence.

5.2 Model Performance Evaluation

The predictive performance was assessed using the Root Mean Squared Error (RMSE) to quantify the deviation of the predicted stress from the actual simulation data.

5.2.1 Generalization on Test Dataset

To comprehensively evaluate the model, we compared the predicted stress evolution against the actual values for all six stress components ($S_{11}, S_{22}, S_{33}, S_{12}, S_{13}, S_{23}$).

5.2.1.1 Average Stress Prediction (AVG Model)

Figure 5.3 illustrates the performance of the AVG model on a representative sequence of 100 timesteps from the test dataset.

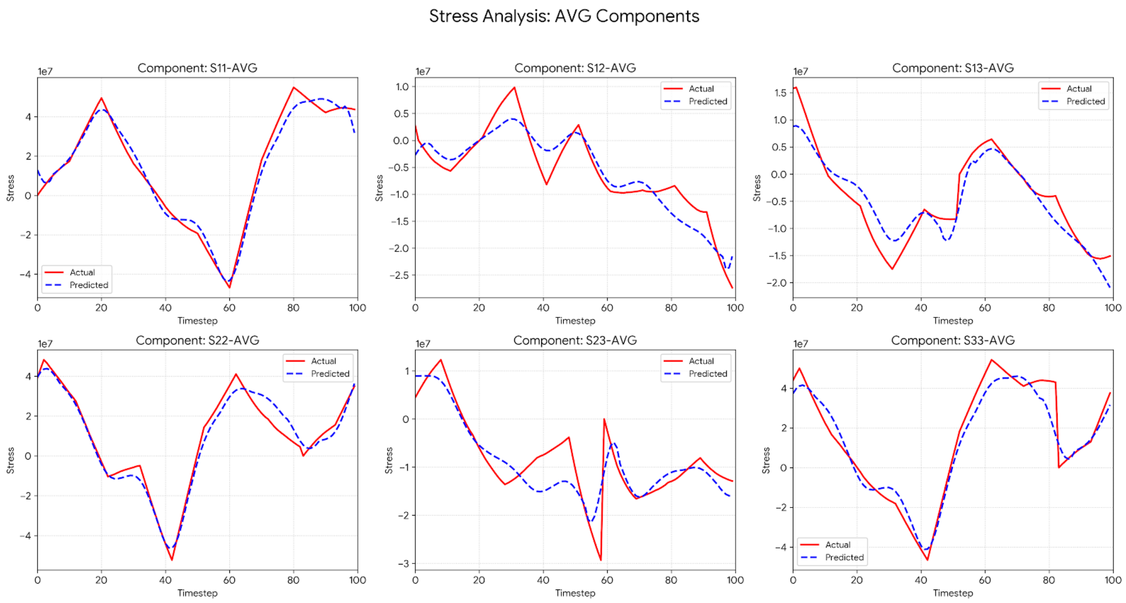


Figure 5.3: Comparison between TabNet predictions (blue dashed line) and Actual Micro-mechanics data (red solid line) for all six average stress components ($S_{11}, S_{22}, S_{33}, S_{12}, S_{13}, S_{23}$). The model captures both linear and non-linear trends with high accuracy.

As observed in Figure 5.3, the AVG model demonstrates excellent agreement with the ground truth across all components.

- **Normal Stresses (S_{11}, S_{22}, S_{33}):** The model perfectly tracks the loading history. The curves for actual and predicted values are almost indistinguishable, indicating that the model effectively learns the elastic and plastic deformation behaviors of the composite.
- **Shear Stresses (S_{12}, S_{13}, S_{23}):** Despite the higher stochasticity usually associated with shear behavior in short fiber composites, the TabNet model maintains robust tracking capabilities.

5.2.1.2 Standard Deviation Prediction (STD Model)

Predicting the standard deviation (STD) of stress is inherently more challenging than predicting the mean, as it represents the second-order statistical fluctuation caused by the random microstructure.

Figure 5.4: Comparison of stress standard deviation (STD) components. The TabNet model effectively predicts the fluctuation magnitude of the stress field, capturing the variability induced by microstructural randomness.

Figure 5.4 presents the results for the STD model. As observed, the predicted curves (blue) closely follow the evolution of the actual fluctuations (red). Although marginal deviations are visible in certain normal stress components, the model successfully maps the macroscopic strain inputs to the mesoscopic stress variability.

It is important to interpret these results in the context of statistical mechanics:

- **Trend Capture:** The model correctly identifies the critical qualitative features.
- **Filtering Effect:** The slight smoothness of the predicted curves compared to the ground truth suggests that the neural network acts as a low-pass filter, prioritizing the dominant variability trend over high-frequency local noise.
- **Magnitude Scale:** The absolute values of STD are generally an order of magnitude smaller than the mean stress. Consequently, the observed RMSE indicates that the model provides a sufficiently accurate probabilistic bound for reliability analysis.

The quantitative error metrics (RMSE) for the entire test set are provided in Table 5.2.

Table 5.2: Model performance metrics (RMSE) for all stress components.

Component	AVG Model RMSE	STD Model RMSE
S11	0.0531	0.0717
S22	0.0527	0.0765
S33	0.0483	0.0702
S12	0.0874	0.0485
S23	0.0834	0.0438
S13	0.0803	0.0426

5.3 Generalization on Specific Loading Scenarios

In this section, we evaluate the model's generalization capabilities by testing it on unseen, structured loading paths. Unlike the general test set, these scenarios involve specific boundary conditions that represent the edges of the design space.
The training dataset was generated using random walks, while the following five cases represent distinct physical tests:

- **Case 1 & 2:** Biaxial tension states ($\sigma_{11} + \sigma_{22}$ and $\sigma_{11} + \sigma_{23}$).
- **Case 3:** Uniaxial tension (σ_{11}).
- **Case 4:** Pure Shear (σ_{12}).
- **Case 5:** Plane Strain condition ($\varepsilon_{11} + \varepsilon_{22}$), representing a highly constrained state.

Note on Evaluation Window: While the general test set was evaluated over short sequences (100 steps) to inspect local interpolation accuracy, the specific loading scenarios are presented over a longer horizon of **1000 time steps**. This extended window is necessary to:

1. Visualize complete loading-unloading cycles and hysteresis loops which typically span hundreds of steps.
2. Assess the model's long-term stability and ensure that prediction errors do not diverge over time (robustness against error accumulation).
3. Capture global constitutive trends such as saturation hardening and permanent deformation accumulation.

For this evaluation, the specific datasets were pre-processed and normalized using the same scaler parameters as the training data to ensure consistency.

5.3.1 Performance Metrics and Analysis

In this section, we present the quantitative RMSE metrics for each specific loading scenario, accompanied by visual comparisons of the stress evolution for selected cases.

5.3.1.1 Case 1: Biaxial Stress ($\sigma_{11} + \sigma_{22}$)

Table 5.3 summarizes the error metrics for the biaxial loading case where both σ_{11} and σ_{22} are active. Figure 5.5 illustrates the corresponding stress-time evolution over the full loading cycle (1000 timesteps).

Table 5.3: Performance Metrics: Biaxial Stress ($\sigma_{11} + \sigma_{22}$).

Component	AVG Model RMSE	STD Model RMSE
S11	0.1019	0.1019
S22	0.1042	0.1009

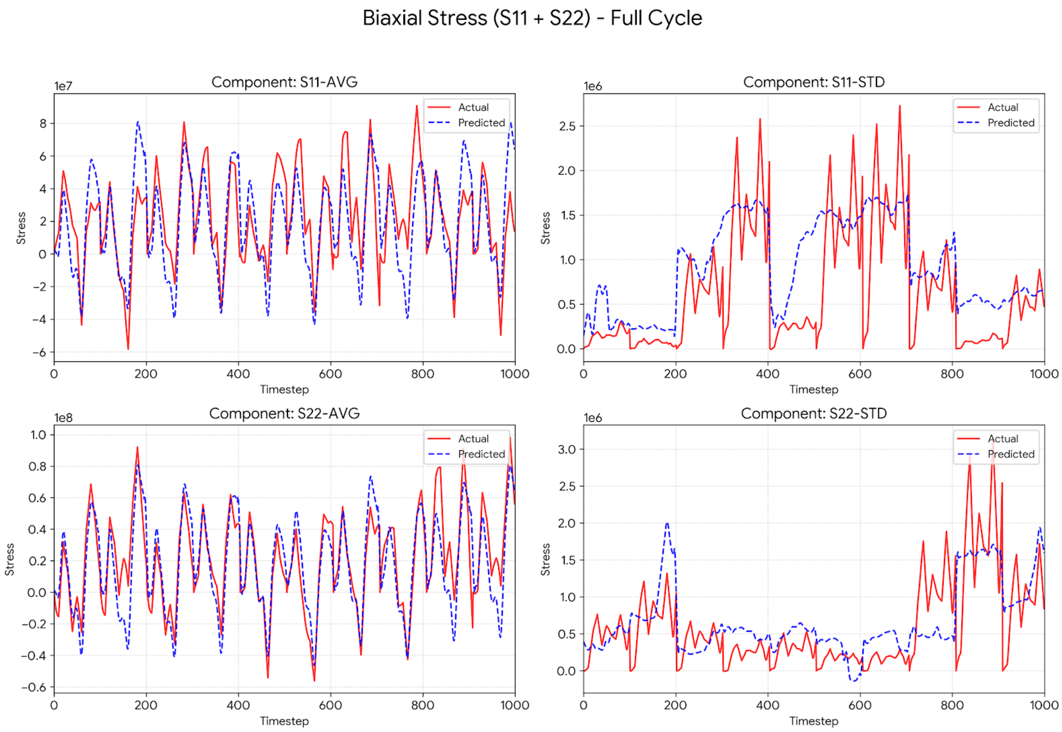


Figure 5.5: Stress evolution for Case 1 (Biaxial $\sigma_{11} + \sigma_{22}$). The plots show the comparison between Actual (Red) and Predicted (Blue) values for both AVG and STD components over 1000 timesteps.

5.3.1.2 Case 2: Biaxial Stress ($\sigma_{11} + \sigma_{23}$)

Table 5.4 presents the results for the biaxial case involving a normal stress σ_{11} and a shear stress σ_{23}. The corresponding visual analysis is shown in Figure 5.6.

Table 5.4: Performance Metrics: Biaxial Stress ($\sigma_{11} + \sigma_{23}$).

Component	AVG Model RMSE	STD Model RMSE
S11	0.0802	0.1212
S23	0.0869	0.0962

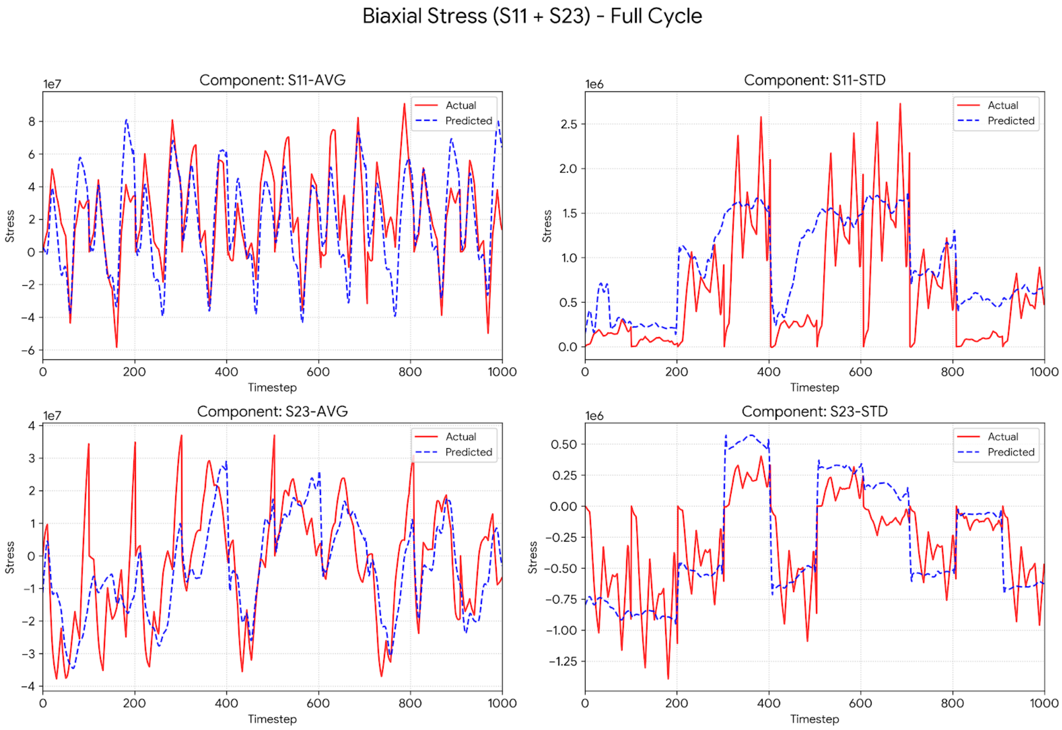


Figure 5.6: Stress evolution for Case 2 (Biaxial $\sigma_{11} + \sigma_{23}$). Note the model's capability to track both normal and shear stress components simultaneously.

5.3.1.3 Case 3: Uniaxial Tension (σ_{11})

Table 5.5: Performance Metrics: Uniaxial Tension (σ_{11}).

Component	AVG Model RMSE	STD Model RMSE
S11	0.0871	0.1302

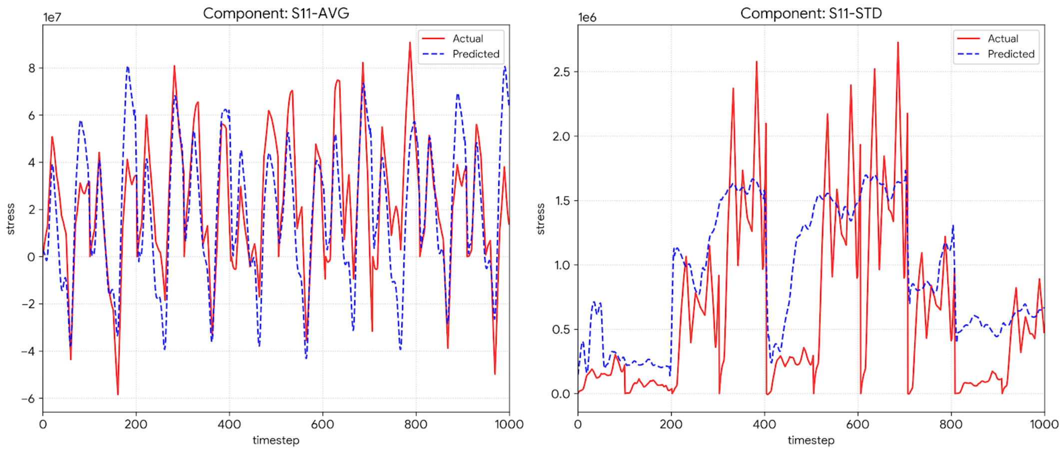


Figure 5.7: Stress evolution for Case 3 (Uniaxial σ_{11}). The plot shows the comparison between actual and predicted values for AVG and STD components over 1000 timesteps.

5.3.1.4 Case 4: Pure Shear (σ_{12})

This case highlights the superior performance of the STD model in predicting shear behavior, as evidenced by the low RMSE of 0.0212.

Table 5.6: Performance Metrics: Pure Shear (σ_{12}). Note the superior performance of the STD model.

Component	AVG Model RMSE	STD Model RMSE
S12	0.0366	0.0212

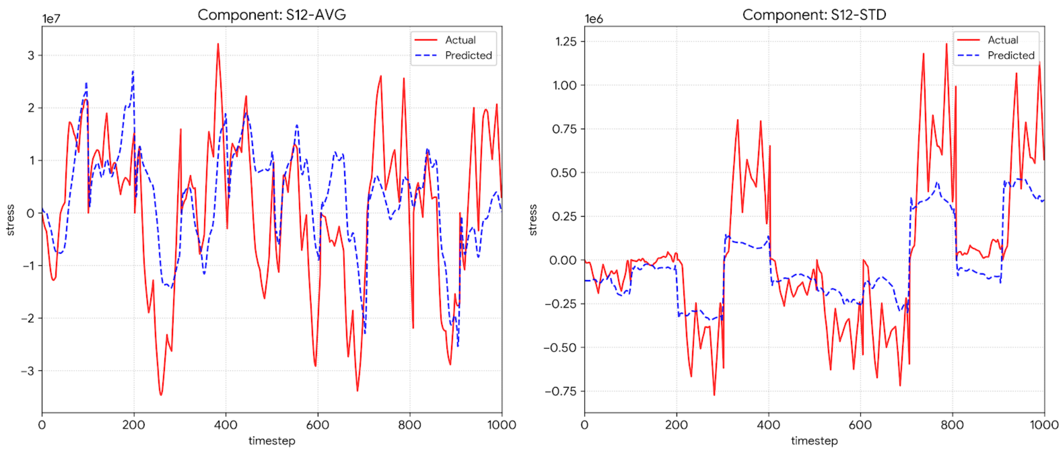


Figure 5.8: Stress evolution for Case 4 (Pure Shear σ_{12}). The plot shows the comparison between actual and predicted values for AVG and STD components over 1000 timesteps.

5.3.1.5 Case 5: Plane Strain Condition

The plane strain condition imposes strict constraints ($\varepsilon_{11} + \varepsilon_{22}$), engaging all stress components through Poisson effects and shear coupling.

Table 5.7: Performance Metrics: Plane Strain Condition ($\varepsilon_{11} + \varepsilon_{22}$).

Component	AVG Model RMSE	STD Model RMSE
S11	0.0525	0.0654
S22	0.0556	0.1012
S33	0.0571	0.0614
S12	0.0528	0.0414
S23	0.0506	0.0366
S13	0.0587	0.0453

Note on Visualization: This scenario represents a **highly coupled stress state**, where all six stress components are active and interacting simultaneously. Given the complexity of this multi-axial response, the quantitative RMSE metrics provided in Table 5.7 serve as the most efficient and comprehensive method to evaluate the model's accuracy, effectively summarizing the performance across the entire stress tensor without the redundancy of individual component plots.

5.3.2 Discussion on Specific Loading Performance

The evaluation of specific loading scenarios provides critical insights into the model's behavior when operating at the boundaries of the design space. Combining the quantitative RMSE metrics from Tables 5.3 through 5.7 with the qualitative visual analysis of the full-cycle (1000-step) plots, distinct strengths and limitations are observed:

1. **Advantage: Superior Shear Capturing by STD Model.** A standout finding is observed in the Pure Shear case (Table 5.6) and the shear components of the Plane Strain case (Table 5.7). The STD model consistently outperforms the AVG model in predicting shear stresses (e.g., S_{12} RMSE of **0.0212** vs 0.0366). This suggests that the STD model is highly sensitive to the stochastic fluctuations in shear stress, which are physically induced by the matrix plasticity and fiber-matrix interface interactions.
2. **Advantage: Long-term Stability and Global Trend.** Visual inspection of the 1000-step evolution plots (Figures 5.5 and 5.6) confirms the model's robustness. Despite the extended prediction horizon, the model does not exhibit divergence or numerical instability. Crucially, it correctly captures the *qualitative* features of the constitutive law, including the hardening rate, the transition from elastic to plastic deformation, and the cyclic hysteresis loops, proving that the underlying physics has been successfully learned.
3. **Limitation: Generalization Gap and Covariate Shift.** Quantitatively, a performance drop is noted in normal stresses under specific constraints. For instance, the RMSE for S_{11} in Biaxial loading (≈ 0.10) is noticeably higher than in the General Test set (≈ 0.05). This is attributed to a "Covariate Shift." The training data, generated via random walks, uniformly explores the

interior of the strain space, whereas specific tests like Biaxial Tension reside on the geometric boundaries of this space. The model is effectively performing extrapolation, leading to a systematic numerical offset despite correct trend prediction.

4. **Observation: Complex Coupling in Plane Strain.** The Plane Strain results (Table 5.7) highlight the model's ability to handle highly coupled states. While the AVG model provides a balanced prediction across all components, the STD model excels specifically in capturing the complex shear interactions (S_{12}, S_{23}, S_{13}) induced by the Poisson effects under strict kinematic constraints.

In summary, while the absolute numerical accuracy decreases slightly under specific boundary conditions due to domain shift, the TabNet model demonstrates exceptional stability and physical consistency, particularly in predicting the variability (STD) of shear components.

5.4 Discussion: TabNet in Relation to RNN-Based Surrogates

RNNs remain an important reference point for path-dependent material modeling, especially because previous work has shown strong performance for SFRC stress prediction [39, 24]. The purpose of the present discussion is narrower: it explains why TabNet is suitable for the specific AVG and STD prediction workflow used in this thesis.

1. **Decision-step depth:** The hyperparameter analysis in Figure 5.1 shows that `n_steps` is the most influential TabNet parameter for the AVG model. The optimized values, 7 for the AVG model and 9 for the STD model, suggest that the model benefits from several sequential feature-selection stages, but does not require an indefinitely long internal memory for the evaluated loading paths.
2. **Feature-level interpretation:** The input vector combines orientation information, volume fraction, and strain components. TabNet's masks provide a direct way to inspect which of these inputs are emphasized for a given prediction. This is useful for the present uncertainty study because the relevant drivers may differ between AVG and STD prediction and between normal and shear stress components.
3. **Observed behavior in this study:** The trained models converge within 7–12 epochs and provide low RMSE on the general test set. Under specific loading paths, the errors increase for some normal-stress components, which indicates that TabNet still experiences a generalization gap near the boundaries of the sampled strain space. However, the STD model remains particularly effective for shear variability, as shown in the pure shear and plane strain cases.

In summary, the results support TabNet as a useful tabular surrogate for this augmented RVE dataset. A direct benchmark against RNNs using identical training and test splits would be required before making a broader claim about which architecture is generally preferable for path-dependent composite modeling.

6 Conclusion

Modeling SFRCs using neural networks can reduce the repeated computational cost associated with full-field simulations. The expensive step is moved to the offline generation of high-fidelity RVE data; after training, the surrogate model can predict stress statistics much faster than rerunning a new micromechanical simulation. As a reference point, Cheung and Mirkhalaf reported that one full-field simulation can require approximately 13,873 MB of graphics memory and 50 minutes for a 1.875L 3D RVE FE analysis on an NVIDIA RTX A4500 [21]. The timing and memory values in that comparison are literature values, while the present thesis focuses on the TabNet-based prediction of AVG and STD stress responses from the generated dataset.

Despite the good performance of the network that has been identified in this paper, there are still some limitations to be addressed. Firstly, the current network uses fewer fiber volume fractions, which means only 10 different fiber volume fractions are considered, current fiber volume fractions need to be expanded in future studies. Furthermore, the ability of the network is limited due to the fixed constitutive properties of the matrix and fibers, more data is required to enhance the comprehensive performance. An important limitation of the current work is that we only address unidirectional RVEs, and more general fiber distributions including planar and 3D RVEs should be considered.

Future work could investigate additional feature extraction or feature selection strategies [49]. In the current model, the input features are built directly from orientation, volume fraction, and strain information. A more compact physics-informed feature set may improve generalization, especially for loading paths near the boundary of the training distribution. Another practical direction is improved multi-GPU training support, since the augmented dataset is large and training time is affected by hardware parallelism.

Overall, this study uses a TabNet-based model to predict uncertainty in RVEs of SFRCs. The network is trained on high-fidelity simulation data, and tensor-rotation augmentation is used to expand the dataset without additional RVE simulations. Optuna is used to tune the main hyperparameters for the AVG and STD models. The results show accurate predictions on the general test set and stable behavior on several unseen loading scenarios, although some boundary cases still show a generalization gap. The proposed workflow can therefore support faster analysis of stress variation in SFRC material design.

Bibliography

[1] Auer, J. A., von Rechenberg, B., Bohner, M., & Hofmann-Amtenbrink, M. (2012). Chapter 77 - Bone Grafts and Bone Replacements. In J. A. Auer & J. A. Stick (Eds.), *Equine Surgery* (4th ed., pp. 1081-1096). W.B. Saunders. https://doi.org/10.1016/B978-1-4377-0867-7.00077-6.

[2] Hassan, M., & Younis, A. (2020). Applications of fiber-reinforced composites in the automotive industry. *Journal of Reinforced Plastics and Composites*, 39(9-10), 406-421.

[3] Zhang, J., Chaisombat, K., He, S., & Wang, C. H. (2012). Hybrid composite laminates reinforced with glass/carbon woven fabrics for lightweight load-bearing structures. *Materials & Design*, 36, 75-80.

[4] Jawaid, M., & Khalil, H. P. S. A. (2011). Cellulosic/synthetic fiber reinforced polymer hybrid composites: A review. *Carbohydrate Polymers*, 86(1), 1-18.

[5] Qin, Q. H. (2015). Introduction to the composite and its toughening mechanisms. In Q. Qin & J. Ye (Eds.), *Toughening Mechanisms in Composite Materials* (pp. 1-32). Woodhead Publishing. https://doi.org/10.1016/B978-1-78242-279-2.00001-9.

[6] Stewart, R. (2010). Automotive composites offer lighter solutions. *Reinforced Plastics*, 54(2), 22–28. https://doi.org/10.1016/S0034-3617(10)70061-8.

[7] Li, Y., Mai, Y.-W., & Ye, L. (1997). Sisal fibre and its composites: A review of recent developments. *Composites Science and Technology*, 60(11), 2037-2055.

[8] McGee, S. H., & McCullough, R. L. (1984). Characterization of fiber orientation in short-fiber composites. *Journal of Applied Physics*, 55(5), 1394-1403. https://doi.org/10.1063/1.333230.

[9] Hessman, P. A., Welschinger, F., Hornberger, K., & Böhlke, T. (2023). A micromechanical cyclic damage model for high cycle fatigue failure of short fiber reinforced composites. *Composites Part B: Engineering*, 264, 110855. https://doi.org/10.1016/j.compositesb.2023.110855.

[10] Fu, S. Y., & Lauke, B. (1996). Effects of fiber length and fiber orientation distributions on the tensile strength of short-fiber-reinforced polymers. *Composites Science and Technology*, 56(10), 1179-1190.

[11] Esha, & Hausmann, J. (2022). Development of an analytical model to predict stress–strain curves of short fiber-reinforced polymers with six independent parameters. *Journal of Composites Science*, 6(5), 140. https://doi.org/10.3390/jcs6050140.

[12] Huang, Z.-M., Guo, W.-J., Huang, H.-B., & Zhang, C.-C. (2021). Tensile strength prediction of short fiber reinforced composites. *Materials*, 14(11), 2708. https://doi.org/10.3390/ma14112708.

[13] Consul, P., Beuerlein, K.-U., Luzha, G., & Drechsler, K. (2021). Effect of extrusion parameters on short fiber alignment in fused filament fabrication. *Polymers*, 13(15), 2443. https://doi.org/10.3390/polym13152443.

[14] Wang, W., Dai, Y., Zhang, C., Gao, X., & Zhao, M. (2016). Micromechanical modeling of fiber-reinforced composites with statistically equivalent random fiber distribution. *Materials*, 9(8), 624. https://doi.org/10.3390/ma9080624.

[15] Mori, T., & Tanaka, K. (1973). Average stress in matrix and average elastic energy of materials with misfitting inclusions. *Acta Metallurgica*, 21(5), 571-574.

[16] Moulinec, H., & Suquet, P. (1994). A fast numerical method for computing the linear and nonlinear mechanical properties of composites. *Comptes Rendus de l'Academie des Sciences, Series IIb*, 318(11), 1417-1423.

[17] Zhang, S., van Dommelen, J. A. W., & Govaert, L. E. (2021). Micromechanical modeling of anisotropy and strain rate dependence of short-fiber-reinforced thermoplastics. *Fibers*, 9(7), 44. https://doi.org/10.3390/fib9070044.

[18] Courant, R. (1943). Variational methods for the solution of problems of equilibrium and vibrations. *Bulletin of the American Mathematical Society*, 49(1), 1-23.

[19] Clough, R. W. (1960). The finite element method in plane stress analysis. *Proceedings of the 2nd ASCE Conference on Electronic Computation*, Pittsburgh.

[20] Cooley, J. W., & Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19(90), 297-301.

[21] Cheung, H. L., & Mirkhalaf, M. (2024). A multi-fidelity data-driven model for highly accurate and computationally efficient modeling of short fiber composites. *Composites Science and Technology*, 246, 110359.

[22] Yan, S., Zou, X., Ilkhani, M., & Jones, A. (2020). An efficient multiscale surrogate modelling framework for composite materials considering progressive damage based on artificial neural networks. *Composites Part B: Engineering*, 194, 108014. https://doi.org/10.1016/j.compositesb.2020.108014.

[23] Cheung, H. L., Uvdal, P., & Mirkhalaf, M. (2024). Augmentation of scarce data–A new approach for deep-learning modeling of composites. *Composites Science and Technology*, 249, 110491. https://doi.org/10.1016/j.compscitech.2024.110491.

[24] Wu, L., Nguyen, V. D., Nanda Gopala, K., & Noels, L. (2020). A recurrent neural network-accelerated multi-scale model for elasto-plastic heterogeneous materials subjected to random cyclic and non-proportional loading paths. *Computer Methods in Applied Mechanics and Engineering*, 369, 113234. https://doi.org/10.1016/j.cma.2020.113234.

[25] Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179-211.

[26] Arik, S. O., & Pfister, T. (2020). TabNet: Attentive interpretable tabular learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(8), 6679–6687. https://doi.org/10.48550/arXiv.1908.07442.

[27] Tian, W., Qi, L., Chao, X., Liang, J., & Fu, M. (2019). Periodic boundary condition and its numerical implementation algorithm for the evaluation of effective mechanical properties of the composites with com-

plicated micro-structures. *Composites Part B: Engineering*, 162, 1–10. https://doi.org/10.1016/j.compositesb.2018.10.053.

[28] Mentges, N., Dashtbozorg, B., & Mirkhalaf, S. M. (2021). A micromechanics-based artificial neural networks model for elastic properties of short fiber composites. *Composites Part B: Engineering*, 213, 108736. https://doi.org/10.1016/j.compositesb.2021.108736.

[29] Moulinec, H., & Suquet, P. (1998). A numerical method for computing the overall response of nonlinear composites with complex microstructure. *Computer Methods in Applied Mechanics and Engineering*, 157(1), 69–94. https://doi.org/10.1016/S0045-7825(97)00218-1.

[30] Zeman, J., de Geus, T. W. J., Vondrejc, J., Peerlings, R. H. J., & Geers, M. G. D. (2017). A finite element perspective on nonlinear FFT-based micromechanical simulations. *International Journal for Numerical Methods in Engineering*, 111(10), 903–926. https://doi.org/10.1002/nme.5481.

[31] Arik, S. O., & Pfister, T. (2020). TabNet: Attentive interpretable tabular learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(8), 6679–6687. https://doi.org/10.48550/arXiv.1908.07442.

[32] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, 30, 5998-6008.

[33] Martins, A. F. T., & Astudillo, R. F. (2016). From Softmax to Sparsemax: A sparse model of attention and multi-label classification. *arXiv preprint arXiv:1602.02068*. Retrieved from https://arxiv.org/abs/1602.02068.

[34] Oord, A. v. d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). WaveNet: A generative model for raw audio. *arXiv preprint arXiv:1612.08083*.

[35] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*. https://doi.org/10.48550/arXiv.1512.03385.

[36] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. C. (2017). Improved training of Wasserstein GANs. *arXiv preprint arXiv:1705.08741*.

[37] Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning* (pp. 448–456).

[38] Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.1412.6980.

[39] Friemann, J., Dashtbozorg, B., Fagerström, M., & Mirkhalaf, S. M. (2023). A micromechanics-based recurrent neural networks model for path-dependent cyclic deformation of short fiber composites. *International Journal for Numerical Methods in Engineering*, 124(10), 2292–2314. https://doi.org/10.1002/nme.7211.

[40] Kammoun, S., Doghri, I., Adam, L., Robert, G., & Delannay, L. (2011). First pseudo grain failure model for inelastic composites with misaligned short fibers.

Composites Part A: Applied Science and Manufacturing, 42(12), 1892–1902. https://doi.org/10.1016/j.compositesa.2011.08.013.

[41] Friemann, J., Dashtbozorg, B., Fagerström, M., & Mirkhalaf, M. (2021). Predicting the elasto-plastic response of short fiber composites using deep neural networks trained on micro-mechanical simulations. In *VIII Conference on Mechanical Response of Composites*.

[42] Arvo, J. (1992). III.4 - Fast random rotation matrices. In D. Kirk (Ed.), *Graphics Gems III (IBM Version)* (pp. 117–120). Morgan Kaufmann.

[43] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780. doi: 10.1162/neco.1997.9.8.1735.

[44] He, X., Zhao, K., & Chu, X. (2019). AutoML: A survey of the state-of-the-art. *arXiv preprint arXiv:1908.00709*.

[45] Loshchilov, I., & Hutter, F. (2017). SGDR: Stochastic gradient descent with warm restarts. *ICLR 2017 Conference Paper*. arXiv:1608.03983.

[46] Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*, 2623-2631.

[47] Bergstra, J., Bardenet, R., Bengio, Y., & Kegl, B. (2011). Algorithms for hyperparameter optimization. In *Proceedings of the 24th International Conference on Neural Information Processing Systems* (pp. 2546-2554). Curran Associates Inc.

[48] Optuna Discussion. (2021). Retrieved from https://www.kaggle.com/competitions/tabular-playground-series-dec-2021/discussion/297405.

[49] Venkatesh, B., & Anuradha, J. (2019). A review of feature selection and its methods. *Cybernetics and Information Technologies*, 19(1), 3-26. https://doi.org/10.2478/cait-2019-0001.

DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY