Comparative Analysis of Deep Learning Approaches for Predicting Thermomechanical Behavior of Shape Memory Polymers

1st Khan Raqib Mahmud
Department of Computer Science
Louisiana Tech University
Ruston, USA
krm070@latech.edu

2nd Lingxiao Wang
Department of Electrical Engineering
Louisiana Tech University
Ruston, USA
lwang@latech.edu

3rd Jinyuan Chen
Department of Electrical Engineering
Louisiana Tech University
Ruston, USA
jinyuan@latech.edu

4rd Xiyuan Liu
Department of Mathematics and Statistics
Louisiana Tech University
Ruston, USA
liuxyuan@latech.edu

5th Sunzid Hassan
Department of Computer Science
Louisiana Tech University
Ruston, USA
sha040@latech.edu

SoutheastCon 2025 | 979-8-3315-0484-7/25/$31.00 ©2025 IEEE | DOI: 10.1109/SOUTHEASTCON56624.2025.10971553

***Abstract*—Thermomechanical constitutive modeling is crucial for understanding and designing shape memory polymers (SMPs) for advanced engineering applications. Traditional approaches are often time-consuming and computationally expensive which requires the development of more efficient and accurate methods. In this paper, we propose a Transformer-based deep learning model to predict the thermomechanical behavior of semicrystalline two-way shape memory polymers (2W-SMPs) under thermomechanical cycles. By leveraging its ability to capture complex, time-dependent patterns, the framework accurately models the intricate relationships between polymer strain, time, temperature, and stress. A comparative analysis with other deep learning models demonstrates that the Transformer excels in accuracy and robustness by capturing intricate dependencies and non-linearities in the data. The test results demonstrate that our proposed Transformer model achieved a root mean squared error (RMSE) of 0.3365, outperforms the traditional deep learning models such as the feedforward neural network (FNN), convolutional neural network (CNN), and long short-term memory (LSTM) models which achieved RMSE values of 6.78, 17.35, and 13.85, respectively. This study highlights the potential of the Transformer model as a powerful tool for predicting material behavior and reducing the time and resources required for constitutive modeling.**

***Index Terms*—Shape Memory Polymer, Thermomechanical Behavior Prediction, Transformer Model, Deep Learning Models.**

I. Introduction

Shape memory polymers (SMPs) are smart materials that can return to their original shape from a deformed shape in response to external stimuli such as temperature, pH, or magnetic fields [1]–[3]. Their versatility and low cost make them popular in various applications such as medicine, aerospace, and self-healing materials [2], [3]. In particular, the two-way SMPs (2W-SMPs) are known because of their ability to switch between two different shapes under cyclic thermal conditions, which makes them different compared to one-way SMPs and provides them multiple functionalities. Although SMPs show significant potential, the process of developing and validating new SMPs remains labor intensive and time-consuming, as it is highly dependent on experimentation through trial and error [4].

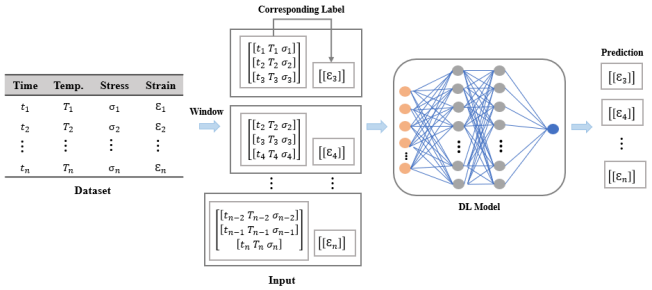


Fig. 1: An overview of the proposed methodology for predicting the thermomechanical behavior of SMPs. The method generates sequences using a sliding window approach to extract overlapping data windows. Each sequence includes n consecutive time steps of stress and temperature as inputs and the strain value at the last time step as the output. These sequences are used to train deep learning models for strain prediction.

Traditional approaches to modeling the thermomechanical behavior of SMPs often involve mathematical and constitutive models that depend on predefined parameters and curve-fitting to experimental data [5]. Although these techniques have provided important insights, they are often time-consuming

This work is supported by NSF EPSCoR LAMDA Seed Funding Track 1B with the contract number NSF(2024)-LAMDATr1B-22.

and not very flexible when it comes to adapting to new materials and changing conditions. Furthermore, thermomechanical analysis (TMA), which is a crucial experimental technique for measuring variations in stress and strain under different temperatures and loads, becomes resource-heavy when conducted under various conditions. The complexity of semicrystalline 2W-SMPs further underscores the need for effective predictive approaches that can reduce experimental demands and speed up material validation. Additionally, inconsistencies between model forecasts and experimental results highlight the requirement for more precise and efficient methodologies.

By offering data-driven strategies to address these issues, recent developments in machine learning (ML) and deep learning (DL) have demonstrated great promise in the field of materials science [6]–[8]. Researchers have been able to speed up the development and optimization of new materials by using ML to predict material attributes including mechanical strength, thermal conductivity, and glass transition temperature [4], [9]. In particular, DL models have shown great promise in simulating complicated, time-dependent behaviors, especially feedforward neural networks (FNNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs) [10], [11]. However, existing research on SMPs primarily focuses on discovering new materials rather than predicting their thermomechanical behavior under dynamic conditions. This gap highlights the opportunity to apply DL models to simulate complex material behavior, reduce reliance on experimental trials, and accelerate material validation.

In this study, we introduce a Transformer-based DL framework to predict the thermomechanical behavior of semicrystalline 2W-SMPs under varying temperature and stress conditions. Transformers, known for their ability to model long-range dependencies in time series data, offer a novel approach to capturing complex temporal dependencies in material behavior. Unlike traditional models the Transformer framework uses its self-attention mechanism to identify complex temporal patterns which makes it particularly well suited for time series predictions. As illustrated in Fig. 1, our approach involves generating data sequences using a sliding window mechanism, where overlapping windows are systematically extracted from the dataset. Each sequence contains a set of consecutive stress and temperature values, serving as inputs, and the strain value at the final time step as the output. These structured sequences are utilized to train deep learning models and enable accurate prediction of strain over time under varying thermomechanical conditions. To assess the performance of our approach, we compare the Transformer model with other DL models, such as FNNs, CNNs, and Long Short-Term Memory (LSTM) networks. Using these models, our goal is to provide a comprehensive analysis of their predictive capabilities and determine the most accurate and robust model for predicting the thermomechanical behavior of SMPs.

The key contributions of this study are as follows:

- We introduce a Transformer-based DL model to predict the thermomechanical behavior of SMPs, demonstrating its ability to capture complex temporal dependencies.
- We compare our Transformer model with other deep learning approaches and observed improved accuracy and enhanced generalization capability in predicting strain behavior under varying thermomechanical conditions.
- We propose a generalized framework for applying DL techniques to model the dynamic behavior of SMPs, offering a scalable and efficient alternative to traditional constitutive modeling.

The remaining of this work is organized as follows: Section II provides an overview of related work on modeling the thermomechanical behavior of SMPs with a focus on traditional constitutive modeling methods and recent advancements in ML and DL techniques. Section III details the architecture and preliminaries of the DL models evaluated in this study and includes the proposed Transformer-based model for predicting the thermomechanical behavior of SMPs. Section IV describes the methodology for data collection and preprocessing, followed by a comparative evaluation of the proposed approach against other DL models. Finally, Section V summarizes the key findings and outlines future research directions.

II. Related Works

Traditional modeling techniques have significantly contributed to understanding the thermomechanical behavior of SMPs. Micromechanics-based models, such as the generalized Maxwell model combined with Mori-Tanaka theory, have been effective in predicting stress-strain distributions in particulate reinforced SMP composites [12], [13]. Finite element analysis (FEA) has been employed to simulate coupled electro-thermo-mechanical responses by incorporating material properties like viscoelasticity and electrical conductivity [14], [15]. While continuum damage mechanics frameworks address deformation and damage during thermomechanical cycles, nonlinear thermomechanical constitutive models have also been widely used to describe SMP behaviors such as glass transition and shape recovery processes [16] [17]. However, as these models rely on predefined parameters, they are not adaptable to complex and dynamic conditions.

DL offers a promising alternative for modeling SMP behavior by reducing experimental dependencies and enhancing predictive accuracy [18]. FNNs and CNNs have demonstrated high accuracy in simulating semicrystalline SMPs under thermomechanical cycles [18], [19]. LSTM networks excel at modeling time-dependent behaviors, outperforming traditional models under large strain conditions [20]. Advanced approaches like Deep Material Networks (DMNs) and physics-informed DL models integrate domain knowledge to predict complex thermo-elasto-viscoplastic behavior and ensure thermodynamic consistency [21], [22].

ML frameworks have accelerated SMP discovery by exploring chemical spaces and predicting properties like recovery stress and glass transition temperature using notations like BigSMILES and dual-convolutional networks [4]. Techniques like transfer learning and variational autoencoders have facilitated efficient material screening and design [4]. Furthermore, rapid characterization of SMPs through scalable ML combined

with video data analysis enables sensitive and specific predictions of shape recovery behavior [23].

Neural networks have also been used in real-world SMP applications, such as controlling light-activated shape memory polymer actuators for dynamic stiffness adjustment [24]. Despite progress, there are still issues, such as the requirement for multilength scale structural integration and larger datasets [4]. By filling in these gaps with more datasets and integrating DL with experimental techniques, SMP applications across sectors can progress and predictive capabilities can be improved.

III. Deep Learning based Thermomechanical Behavior Prediction

A. Deep Learning Models

1) FNN: A FNN model consists of an input layer, one or more hidden layers, and an output layer, all of which are fully connected. In this architecture, each neuron in a layer is connected to every neuron in the next layer. This design enables the network to capture complex relationships and patterns in the data by transforming input features through weights, biases, and activation functions [25]. In this study, we implemented a multi-headed FNN architecture to process three distinct input features: time, temperature, and stress. Each input is processed independently through a dedicated head. Within each head, the input is first passed through an embedding layer that maps the input sequence to a 250-dimensional vector. This embedding is followed by a fully connected layer with 100 neurons and a ReLU activation function, along with a dropout layer with a rate of 0.1 to prevent overfitting. The outputs from the three heads are concatenated to form a combined feature representation of 300 dimensions. This combined representation is passed through an additional fully connected layer with 100 neurons and a ReLU activation, and then mapped to the target output through a final fully connected layer [18]. The model was trained for 150 epochs with a batch size of 16 using the Adam optimizer with a weight decay of 0.01. The structure of the FNN is illustrated in Fig. 2.

2) CNN: We developed a CNN to predict the thermomechanical behavior of SMPs by processing sequences of input features, such as time, temperature, and stress values. The model begins with an input layer that accepts the structured sequence data as a matrix. This input is passed through an initial convolutional layer, which extracts key features by applying learnable filters that detect patterns in the data, followed by a pooling layer to reduce dimensionality and computational complexity [26]–[28]. To enhance feature extraction and capture complex patterns, the architecture integrates multiple convolutional and pooling layers. The CNN architecture in this study comprises four convolutional layers with 128, 64, 64, and 32 filters respectively, each followed by a max pooling layer and a kernel size of 2 to preserve input dimensions, as illustrated in Fig. 3. The output of the final pooling layer is flattened and processed through a dense layer with 100 neurons to refine feature extraction, followed by a dropout layer with a rate of 0.3 to prevent overfitting and mapped to the target output through a fully connected layer. The model was trained for 100 epochs with a batch size of 8 using the Adam optimizer with a learning rate of 0.0001 and weight decay of 0.0001.

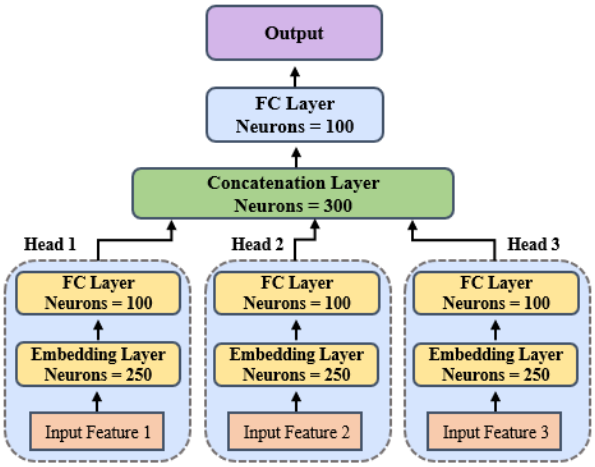


Fig. 2: Architecture of multi-headed FNN, where each input feature is processed through a dedicated head, comprising an embedding layer and a fully connected (FC) layer.Outputs from the heads are concatenated and passed through fully connected layers to generate the final prediction.

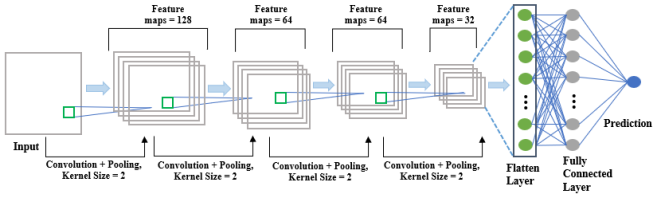


Fig. 3: Architecture of the CNN model for predicting thermomechanical behavior. The model processes sequences of time, temperature, and stress values through multiple convolutional and pooling layers for feature extraction and dimensionality reduction. The output is flattened and passed through fully connected layers to produce the final prediction.

3) LSTM: The LSTM model is designed to capture temporal dependencies in time series data [29]. The model begins with an input layer that processes sequences of time, temperature, and stress values. These sequences are passed through two stacked LSTM layers with 128 hidden units each, which use memory cells and gating mechanisms to retain long-term dependencies and extract meaningful patterns from the data [30]. To reduce overfitting, a dropout layer with a rate of 0.1 is applied before the LSTM layers. The output from the last time step of the LSTM layers is passed through three fully connected layers, each with 128 neurons and ReLU activation, followed by a final dense layer that predicts the strain values. The architecture of the LSTM model is illustrated in Fig. 4. The model was trained for 100 epochs with a batch size of 32 using the Adam optimizer (learning rate 0.01, weight decay 0.01).

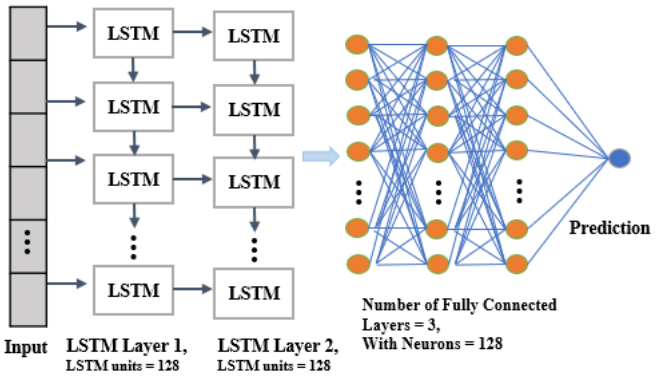


Fig. 4: Architecture of the LSTM model for predicting the thermomechanical behavior of SMPs. The model processes sequences of time, temperature, and stress values through stacked LSTM layers to capture temporal dependencies, followed by fully connected layers to refine features and generate the strain prediction.

B. Transformer-Based Framework

In this study, we implemented a Transformer model designed for time-series forecasting. The proposed model, illustrated in Fig. 5, consists of an encoder-decoder structure that relies on multi-head attention and positional encodings to capture temporal dependencies. The input sequences organized as overlapping windows are projected into a higher-dimensional space through a fully connected layer with positional information added to preserve the order of observations [31]. These sequences are processed through the encoder and decoder layers to learn complex patterns and generate predictions for the strain at the final time step of each input window [32].

1) Encoder Model: The encoder processes overlapping sequences of input features (e.g., time, temperature, stress) by first mapping them into a d-dimensional embedding space via a projection layer. Positional encodings are added to preserve the temporal order of observations since the attention mechanism does not inherently track sequence order. The embeddings are passed through multiple identical encoder layers, each consisting of two sub-layers: a multi-head self-attention mechanism that captures dependencies within the sequence, and a position-wise feed-forward network that refines the learned representations. Residual connections and layer normalization are applied in each sub-layer to stabilize training and preserve information flow. The encoder outputs a context-rich, d-dimensional representation of the input sequences.

2) Decoder Model: The decoder takes the encoder output, referred to as memory, along with the input sequence to predict the strain at the final timestep of each input window. Similar to the encoder, the decoder first maps its input through a projection layer and positional encodings. Each decoder layer includes three sub-layers: a masked multi-head self-attention mechanism for focusing on relevant parts of the decoder input, a multi-head cross-attention mechanism for attending to the encoded memory, and a position-wise feed-forward network for refining predictions. Residual connections and layer normalization ensure stability and efficient learning. A fully connected output layer maps the decoder's representation to the desired strain values for each input window.

3) Model Training: Before training, the time series data was normalized and segmented into overlapping windows to capture temporal dependencies. Each window consists of a fixed number of past observations (e.g., time, temperature, stress) as input, with the strain at the final time step as the target output. These structured windows are used to create training batches, where the input tensor has dimensions (batch size, input width, number of features). Here, $input\ \ width$ refers to the number of time steps in each sequence, and $number\ \ of\ \ features$ represents the number of variables in the input sequence. The labels are aligned accordingly to ensure that the model is accurately learning to predict the target strain. The Transformer architecture employed six stacked layers in both the encoder and decoder, with multi-head attention mechanisms utilizing eight heads and a model dimension of 512. Positional encodings preserved the temporal order of observations, while feed-forward layers with an intermediate dimension of 2048 and a dropout rate of 0.1 enabled non-linear transformations and regularization. The model was trained for 100 epochs with a batch size of 256 using the Adam optimizer with a learning rate of 0.0001 and a weight decay of 0.0001.

IV. Experiments and Results

A. Data Collection

To conduct this investigation, we utilized data from TMA experiments on a 2W-SMP known as crosslinked cis-polybutadiene (cPBD) [33]. The shape memory effect of 2W-SMPs, such as cPBD, is more sophisticated than that of conventional one-way SMPs, which depend on glass transition or crystallization-melting transition. The deformation of cPBD is specifically controlled by melt-crystallization transitions at low temperatures as well as rubber elasticity at high temperatures. Because of these special qualities, cPBD is a perfect choice for assessing how well ML models can represent complicated material behavior, especially in light of classic constitutive models' inability to match experimental data [34], [35].

Three time-dependent variables: stress, temperature, and strain were captured in each of the two datasets produced by the TMA trials. Strain, which represents the deformation of the polymer in response to mechanical and thermal stimuli, was assessed as the dependent variable during the tests while temperature and stress were controlled. With maximum temperatures above the melting point and lowest temperatures below the crystallization point, the records show how cPBD behaves under cyclic loading and temperature fluctuations. The range of stress that varies is the primary difference between the datasets. While Dataset 2 included a lower stress range of 0.1 MPa to 0.2 MPa, Dataset 1 included stress values between 0.3 MPa and 0.4 MPa. Each dataset has a different strain curve as a result of these variations. Dataset 1 has 19,820 data points,

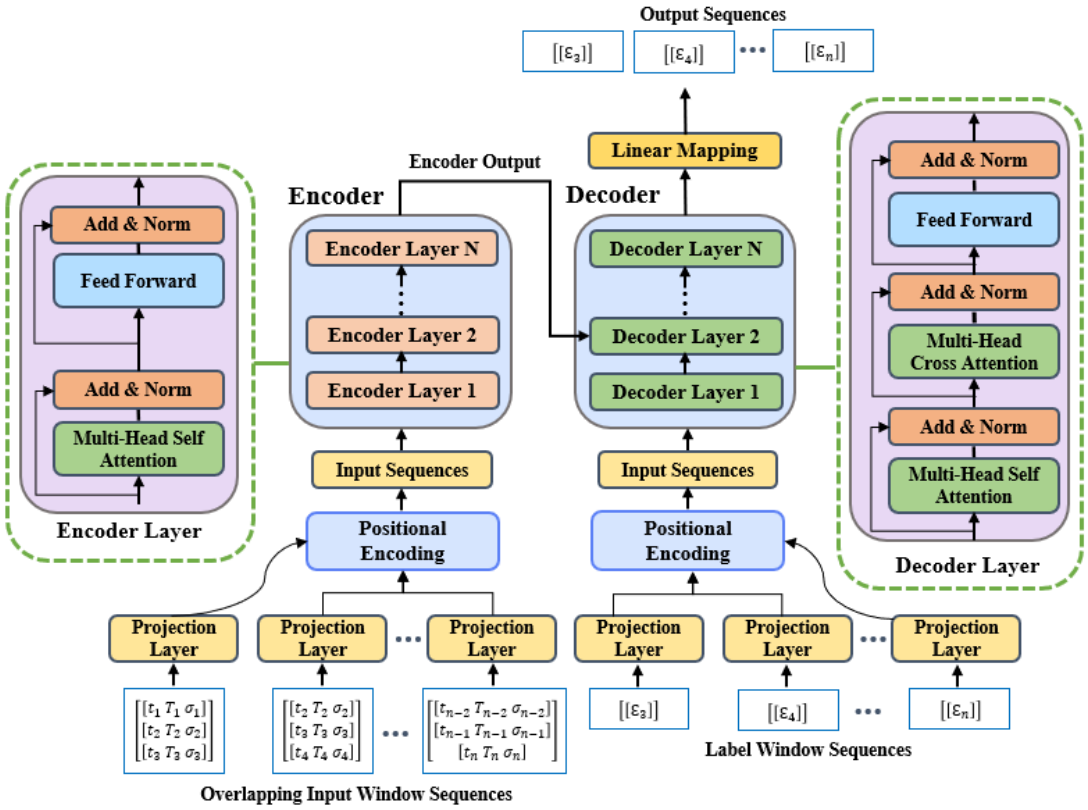


Fig. 5: Architecture of Transformer model for time series forecasting where the input consists of overlapping window sequences of input features (e.g., time, temperature, stress), while the output corresponds to the target feature at the final timestep of each input window. The model employs an encoder-decoder structure with multi-head attention and positional encodings.

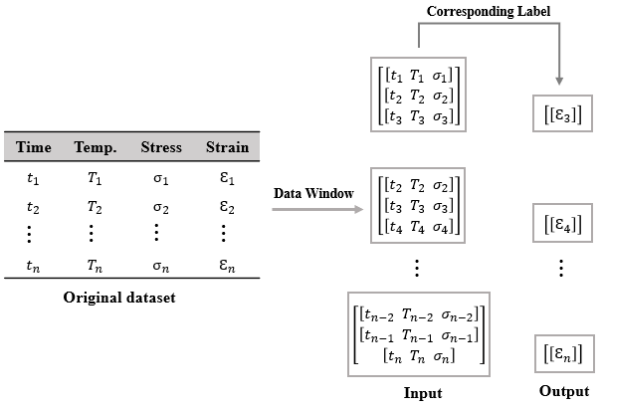


Fig. 6: The sliding window approach generates overlapping sequences of input features (e.g., time, temperature, stress), with the output corresponding to the target feature (strain) at the final timestep of each sequence.

whereas Dataset 2 contains 18,982 data points, for a total of 38,802 data points.

B. Data Preprocessing

We used normalization and sequencing as part of a preprocessing strategy to prepare the data for training. In our datasets, the input variables exhibited varying ranges and statistical properties, such as different means and standard deviations. We used z-score normalization, which scales each variable to have a mean of zero and a standard deviation of one, to normalize these variables. By ensuring that all input features were on the same scale, the models were able to assign the same weight to each feature during training. The normalization formula is as follows:

$$z_i = \frac{x_i - \mu}{\sigma}, \tag{1}$$

where z_i is the normalized value, x_i is the original value, μ is the mean, and σ is the standard deviation of the respective variable.

To capture the temporal dependencies in strain, we structured the data into sequences. Each input sequence comprised n consecutive time steps of stress and temperature values, with the corresponding output being the strain at the final timestep. For example, when $n = 3$, the input consisted of three consecutive time steps of stress and temperature values, while the output is the strain at the third timestep. This sequence generation process, illustrated in Fig. 6, used a sliding window approach to systematically extract overlapping sequences from the datasets. Larger values of n provided more historical context but reduced the number of usable data points,

as the first $n-1$ entries could not be included. For our analysis, we evaluated the DL models using two values of n: $n = 100$ and $n = 200$, as summarized in Tables I and II.

C. Evaluation Metrics

To evaluate the performance of the models, we assessed their ability to predict strain values from Datasets 1 and 2 based on the corresponding time, temperature, and stress values. Root Mean Squared Error (RMSE) and Pearson Correlation Coefficient (PCC) are two standard evaluation metrics that were used to compare the predicted strain values with the experimental data [36]. The RMSE provides an overall estimate of the prediction error taking the square root of the average squared differences between the actual and predicted values. The PCC measures the linear relationship between the predicted and actual values, and a value closer to 1 indicates a strong positive correlation. These metrics provide complementary insights into the models' accuracy and their ability to capture the relationship between input variables and strain. The formula for RMSE is as follows:

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

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and N is the total number of data points.

The formula for PCC is:

$$\text{PCC} = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2 \sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}}, \tag{3}$$

where y_i and $\hat{y}_i$ are the actual and predicted values, respectively, and $\bar{y}$ and $\bar{\hat{y}}$ are their mean values.

These metrics were computed for both datasets, and the results are detailed in Tables 2 and 3. RMSE measures the magnitude of prediction errors, while PCC evaluates how well the model captures the correlation between predicted and actual values.

D. Results and Discussion

We evaluated the performance of DL models to predict the thermomechanical behavior of SMPs using the evaluation metrics. The DL models include FNN, CNN, LSTM, and our proposed Transformer model. The models were trained on Dataset 1 with a 70:10:20 split for training, validation, and testing. To assess the generalization capability of the models, we tested the trained models from Dataset 1 on Dataset 2, normalizing Dataset 2 using the mean and standard deviation of Dataset 1.

The Transformer model achieved the lowest RMSE of 0.0406 and highest PCC of 0.9996 on Dataset 1, as shown in Table I. Fig. 7 shows the performance of the DL models to predict the strain values and, among all, the prediction of the Transformer model aligned with the actual strain values. In comparison, traditional DL models such as FNN and CNN performed moderately well but struggled to capture temporal dependencies. LSTM exhibited the highest RMSE of 0.3969 on Dataset 1 which shows that LSTM failed to learn long-term dependencies despite its recurrent structure.

TABLE I: Performance Metrics of DL Models on Dataset 1

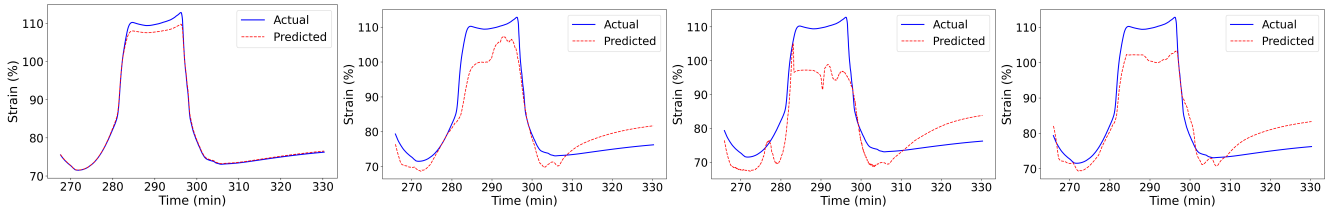
Learning Models	Window Size (n)	RMSE ↓	PCC ↑
Transformer	100	0.0655	0.9988
	200	**0.0406**	**0.9996**
FNN	100	0.2143	0.9493
	200	0.0827	0.9905
CNN	100	0.2154	0.9482
	200	0.1776	0.9867
LSTM	100	0.3051	0.8931
	200	0.3969	0.9655

TABLE II: Performance Metrics of DL Models on Dataset 2

Learning Models	Window Size (n)	RMSE ↓	PCC ↑
Transformer	**100**	**0.3365**	**0.9999**
	200	0.3802	0.9999
FNN	100	6.78	0.9704
	200	10.12	0.9202
CNN	100	17.35	0.8150
	200	22.71	0.8981
LSTM	100	13.85	0.9231
	200	19.95	0.8312

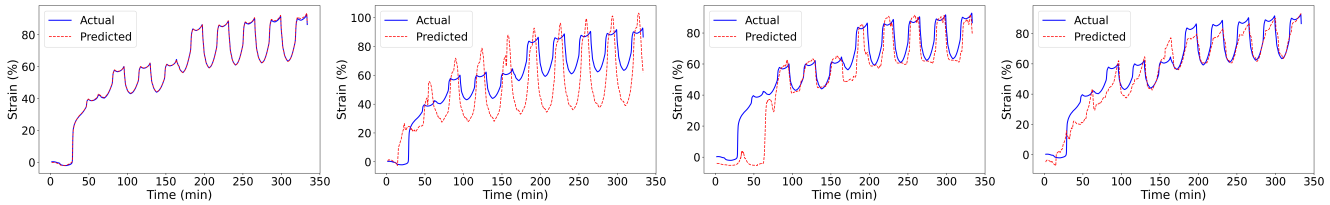
For Dataset 2 which is entirely unseen to the model trained with Dataset 1, the Transformer model shows superior performance. We observed that Transformer gives the lowest RMSE of 0.3365 and the highest PCC of 0.9999, as presented in Table II. From the results illustrated in Fig. 8, we observed that the Transformer model is capable of generalizing to the new dataset, as the model can accurately predict completely new data sets and even in different ranges of input features. In contrast, if we look into the performance of traditional DL models on Dataset 2, we can observe that FNN performed reasonably well among the traditional DL models with RMSE value of 6.78 and PCC value of 0.9704. While the prediction error using LSTM is higher than the FNN, from Fig. 8, we can see that LSTM performs better than CNN. The performance metrics of the CNN model from Table II demonstrate that CNN has the highest RMSE value of 22.71 and the lowest PCC value which shows that CNN has limitations to learn complex patterns of the input features.

From Tables I, II and Figs. 7, 8, we observed that the performance of the Transformer model is superior to the traditional DL models like FNN, CNN and LSTM. This is because the Transformer model has the ability to process all input sequences simultaneously through self-attention mechanisms. In addition, this model is capable of capturing long-range dependencies and nonlinear relationships among



(a) (b) (c) (d)

Fig. 7: Performance of deep learning models in predicting the strain curve on test portion of Dataset 1. (**a**) Transformer model. (**b**) CNN model. (**c**) LSTM model. (**d**) FNN model.



(a) (b) (c) (d)

Fig. 8: Performance of deep learning models in predicting the strain curve on Dataset 2. (**a**) Transformer model. (**b**) CNN model. (**c**) LSTM model. (**d**) FNN model.

the features more effectively than FNN, CNN and LSTM. In particular, models such as CNN and LSTM, depend on fixed receptive fields or sequential processing, whereas the Transformer model utilizes its attention mechanism to analyze the complex interactions within the data. As a result, the Transformer model demonstrates better accuracy in predicting strain, even in completely unseen data, which had different ranges of values in input features. One key challenge of using Transformer model in time series forecasting is to preserve the temporal order of observations [37]. To handle that, in our Transformer model we integrate the sliding-window approach and positional encoding that gives the model more flexibility, as a result the model can effectively preserve the temporal order of observations.

Overall, these findings in our study demonstrates the potential of Transformer model in predicting the thermomechanical behavior of SMPs with high accuracy and adaptability. This model can help to reduce reliance on extensive experimental trials to test SMPs. In addition to SMPs, the Transformer model can be used to investigate the thermomechanical behavior of other materials. This model also offers a scalable framework that can be utilized to study other dynamic material systems.

V. Conclusion

In this study, we investigated the effectiveness of DL models that includes traditional Dl models like FNN, CNN, LSTM, and our proposed Transformer model, in predicting the thermomechanical behavior of semicrystalline 2W-SMPs. The models were trained to predict the polymer's strain using stress, temperature, and time as input features. This approach provides a data-driven method for understanding the complex shape memory behavior of SMPs under thermomechanical cycles. We compare the performance of the DL models using two different performance metrics, as discussed in IV-C, on two different datasets with varying range of values into the input features. The results show that the Transformer model achieved the highest accuracy and demonstrated strong generalization capabilities among the DL models. It effectively captured complex temporal dependencies and nonlinearities in the data. While CNN and FNN models provided reasonable predictions, they struggled with unseen data, and LSTM exhibited scalability issues and higher errors. These findings highlight the potential of Transformer models for precise and adaptable predictions in dynamic material systems and significantly reduce the reliance on time-intensive experimental procedures. By replacing time-intensive experimental procedures with predictive models, the design process can be accelerated, material properties optimized, and novel applications for SMPs explored. Additionally, this approach offers a scalable framework for studying other dynamic material systems, where understanding time-dependent behavior is critical.

Future work will focus on incorporating polymer structural information into the DL framework by integrating chemical informatics. This extension would enable the model to correlate molecular structure with thermomechanical behavior and facilitate both material discovery and behavior prediction. By addressing these future directions, we aim to develop a more generalized and robust tool that can aid in the rapid evaluation of new SMP materials across diverse experimental conditions.

References

[1] H. Meng and G. Li, "A review of stimuli-responsive shape memory polymer composites," *Polymer*, vol. 54, no. 9, pp. 2199–2221, 2013.

[2] W. Sokolowski, A. Metcalfe, S. Hayashi, L. Yahia, and J. Raymond, "Medical applications of shape memory polymers," *Biomedical Materials*, vol. 2, no. 1, p. S23–S27, 2007.

[3] Y. Liu, H. Du, L. Liu, and J. Leng, "Shape memory polymers and their composites in aerospace applications: a review," *Smart Materials and Structures*, vol. 23, no. 2, 2014.

[4] C. Yan, X. Feng, C. Wick, A. Peters, and G. Li, "Machine learning assisted discovery of new thermoset shape memory polymers based on a small training dataset," *Polymer*, vol. 214, p. 123351, 2021.

[5] L. Dai, C. Tian, and R. Xiao, "Modeling the thermo-mechanical behavior and constrained recovery performance of cold-programmed amorphous shape-memory polymers," *International Journal of Plasticity*, vol. 127, p. 102654, 2020.

[6] K. Guo, Z. Yang, C.-H. Yu, and M. J. Buehler, "Artificial intelligence and machine learning in design of mechanical materials," *Mater. Horiz.*, vol. 8, pp. 1153–1172, 2021.

[7] J. Wei, X. Chu, X.-Y. Sun, K. Xu, H.-X. Deng, J. Chen, Z. Wei, and M. Lei, "Machine learning in materials science," *InfoMat*, September 2019.

[8] J. Lee, W. Park, B. Do Lee, *et al.*, "Dirty engineering data-driven inverse prediction machine learning model," *Scientific Reports*, vol. 10, p. 20443, 2020.

[9] C. Yan, X. Feng, and G. Li, "From drug molecules to thermoset shape memory polymers: A machine learning approach," *ACS Applied Materials & Interfaces*, vol. 13, no. 50, pp. 60508–60521, 2021.

[10] F. Masi, I. Stefanou, P. Vannucci, and V. Maffi-Berthier, "Thermodynamics-based artificial neural networks for constitutive modeling," *Journal of the Mechanics and Physics of Solids*, vol. 147, p. 104277, 2021.

[11] C. Yan, Q. Yang, and G. Li, "A phenomenological constitutive model for semicrystalline two-way shape memory polymers," *International Journal of Mechanical Sciences*, vol. 177, p. 105552, 2020.

[12] W. Zhao, L. Liu, J. Leng, and Y. Liu, "Thermo-mechanical behavior prediction of particulate reinforced shape memory polymer composite," *Composites Part B-engineering*, vol. 179, pp. 107455–, 2019.

[13] C. Yin, H. Zeng, J. Gu, Z. Xie, and H. Sun, "Modeling the thermomechanical behaviors of particle reinforced shape memory polymer composites," *Applied Physics A*, vol. 125, no. 6, pp. 1–10, 2019.

[14] *Coupled Electro-Thermo-Mechanical Modeling of Shape Memory Polymers*, vol. ASME 2019 Conference on Smart Materials, Adaptive Structures and Intelligent Systems of *Smart Materials, Adaptive Structures and Intelligent Systems*, 09 2019.

[15] R. W. Mailen, M. D. Dickey, J. Genzer, and M. A. Zikry, "A fully coupled thermo-viscoelastic finite element model for self-folding shape memory polymer sheets," *Journal of Polymer Science Part B*, vol. 55, no. 16, pp. 1207–1219, 2017.

[16] H. Tobushi, S. Hayashi, N. Ito, and T. Takata, "Thermomechanical constitutive model of shape memory polymer," *Transactions of the Japan Society of Mechanical Engineers. A*, vol. 66, no. 643, pp. 502–508, 1999.

[17] A. Shojaei, G. Li, and G. Li, "Thermomechanical constitutive modelling of shape memory polymer including continuum functional and mechanical damage effects," *Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 470, no. 2170, pp. 20140199–20140199, 2014.

[18] D. S. Ibarra, J. Mathews, F. Li, H. Lu, G. Li, and J. Chen, "Deep learning for predicting the thermomechanical behavior of shape memory polymers," *Polymer*, vol. 261, pp. 125395–125395, 2022.

[19] Q. Chen, W. Tu, and M. Ma, "Deep learning in heterogeneous materials: Targeting the thermo-mechanical response of unidirectional composites," *Journal of Applied Physics*, vol. 127, no. 17, pp. 175101–, 2020.

[20] H. Wen, S. Wang, J. Jin, X. Wang, X. Tang, Y. Zhang, L. Deng, P. Gong, D. Li, B. Ning, and Q. Wei, "Deep learning-based modeling of the strain rate-dependent thermomechanical processing response for a novel hiped p/m nickel-based superalloy," *Journal of Materials Processing Technology*, 2023.

[21] D. Shin, R. Alberdi, R. A. Lebensohn, and R. Dingreville, "A deep material network approach for predicting the thermomechanical response of composites," *Composites Part B-engineering*, vol. 272, pp. 111177–111177, 2024.

[22] B. Bahtiri, B. Arash, S. Scheffler, M. Jux, and R. Rolfes, "A thermodynamically consistent physics-informed deep learning material model for short fiber/polymer nanocomposites," *Computer Methods in Applied Mechanics and Engineering*, vol. 427, pp. 117038–117038, 2024.

[23] R. Dutta, D. Renshaw, C. Chen, and D. Liang, "Machine learning based approach for shape memory polymer behavioural characterization," *Array*, vol. 7, p. 100036, 2020.

[24] *Neural-Network Vibration Control of Rings With Light-Activated Shape Memory Polymer Actuators*, vol. Volume 8: 27th Conference on Mechanical Vibration and Noise of *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 08 2015.

[25] M. Deprez and E. C. Robinson, "Chapter 10 - fully connected neural networks," in *Machine Learning for Biomedical Applications* (M. Deprez and E. C. Robinson, eds.), pp. 213–231, Academic Press, 2024.

[26] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999–7019, 2022.

[27] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018.

[28] A. Ayodeji, Z. Wang, W. Wang, W. Qin, C. Yang, S. Xu, and X. Liu, "Causal augmented convnet: A temporal memory dilated convolution model for long-sequence time series prediction," *ISA Transactions*, vol. 123, pp. 200–217, 2022.

[29] P. Lara-Benítez, M. Carranza-García, and J. C. Riquelme, "An experimental review on deep learning architectures for time series forecasting," *International Journal of Neural Systems*, vol. 31, no. 03, p. 2130001, 2021.

[30] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), vol. 30, Curran Associates, Inc., 2017.

[32] N. Wu, B. Green, X. Ben, and S. O'Banion, "Deep transformer models for time series forecasting: The influenza prevalence case," 2020.

[33] L. Lu, J. Cao, and G. Li, "Giant reversible elongation upon cooling and contraction upon heating for a crosslinked cis poly(1,4-butadiene) system at temperatures below zero celsius," *Scientific Reports*, vol. 8, p. 14233, 2018.

[34] G. Scalet, S. Pandini, M. Messori, M. Toselli, and F. Auricchio, "A one-dimensional phenomenological model for the two-way shape-memory effect in semi-crystalline networks," *Polymer*, vol. 158, pp. 130–148, 2018.

[35] C. Yan, Q. Yang, and G. Li, "A phenomenological constitutive model for semicrystalline two-way shape memory polymers," *International Journal of Mechanical Sciences*, vol. 177, p. 105552, 2020.

[36] V. Plevris, G. Solorzano, N. Bakas, and M. B. Seghier, "Investigation of performance metrics in regression analysis and machine learning-based prediction models," in *Proceedings of ECCOMAS 2022*, 2022.

[37] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, pp. 11121–11128, 2023.