Advances in Reliability, Advances in Reliability, Safety and Security, Part 2 – Kołowrocki, Magryta-Mut (eds.)
© 2024 Polish Safety and Reliability Association, Gdynia, ISBN 978-83-68136-14-2 (printed), ISBN 978-83-68136-01-2 (electronic)

Advances in eliability, Safety and Security

ESREL 2024
Monograph Book Series

Analyzing Surface Quality Patterns Through Lens Of Transformer Algorithm

Candeniz Cicek, Marcin Hinz

Munich University of Applied Sciences, Munich, Germany

Abstract

This study introduces a transformative approach to assessing surface quality in cutlery manufacturing through the application of the Transformer algorithm, a cutting-edge machine learning model known for its efficiency in processing complex data sequences. By analyzing high-resolution images of finely grinded surfaces, we demonstrate the algorithm's potential to accurately predict surface roughness values, offering a novel perspective on automated quality control. Our research integrates traditional measures of surface topography with advanced image analysis techniques, aiming to bridge the gap between conventional quality assessment methods and the capabilities of modern artificial intelligence. The findings suggest that the Transformer model, with its superior data handling and analysis capabilities, can significantly enhance the precision and efficiency of surface quality evaluations in industrial settings.

This paper presents a comprehensive study on the application of Transformer algorithms in the realm of manufacturing, highlighting its implications for improving product quality, reducing waste, and streamlining production processes. Through a meticulous examination of the model's performance against established benchmarks, we offer insights into the future of manufacturing quality control, emphasizing the role of advanced computational techniques in achieving unparalleled accuracy and reliability in product assessment.

Keywords: artificial intelligence, machine learning, supervised learning, transformer algorithm, surface quality

1. Introduction

The perception and measurement of surface quality, particularly in finely grinded products, play a crucial role in determining the overall quality and customer satisfaction. The intricate process of manufacturing such high-quality surfaces involves numerous variables, including feed rate and cutting speed, which significantly affect the surface's topography and, consequently, its quality. An optimized configuration of these parameters can lead to an improvement in product durability and quality. With advancements in technology, traditional methods of surface quality assessment, which often involve manual, time-consuming, and costly techniques, are increasingly being supplemented by more efficient, automated solutions.

In recent years, the application of machine learning algorithms has provided a groundbreaking approach to analyzing and predicting surface qualities directly from image data. The Transformer algorithm, a model that utilizes self-attention mechanisms to process data in parallel rather than sequentially, has emerged as a particularly potent tool in this domain. Unlike its predecessors, such as LSTM and CNN models, the Transformer algorithm efficiently handles sequences of data, capturing complex dependencies and relationships within the data with remarkable precision. This capability makes it an ideal candidate for analyzing the nuanced patterns of lightness and texture in high-resolution images of grinded surfaces, aiming to correlate these patterns with quantifiable measures of surface roughness.

This paper explores the use of the Transformer algorithm to analyze surface quality in a novel application: the assessment of fine grinded surfaces in cutlery manufacturing. By employing a Transformer-based model to process and analyze high-resolution images of cutlery surfaces, this study aims to demonstrate the algorithm's efficacy in identifying and classifying surface quality variations. Through a comprehensive analysis of image data paired with traditional surface topography measurements, this research endeavors to offer insights into the potential of

Transformer models in revolutionizing surface quality assessment, moving towards more automated, precise, and efficient quality control processes in manufacturing.

2. Transformer - theory

The Transformer model is considered as a subset of neural networks in the field of artificial intelligence. The model is predominantly used for processing sequences. Transformer models have the significant advantage of being able to flexibly access the weights between individual input elements, which means they do not need to process data sequentially. This approach is also referred to as the attention mechanism. The attention mechanism, also called self-attention, captures complex dependencies and relationships with high efficiency.

At the heart of the Transformer model lies the architecture of the encoder and decoder. The encoder processes the input data through several layers, employing self-attention mechanisms and feed-forward networks for its operation. Self-attention mechanisms allow the encoder to evaluate the relevance of each input element relative to the entire sequence. Concurrently, the feed-forward networks serve to modify and refine the input data. Conversely, the decoder is tasked with generating the output. It takes the encoded input and incrementally constructs the target sequence. The decoder similarly uses self-attention mechanisms but focuses on the decoder input to integrate the information from the encoder. The Transformer model's notable strength resides in its capability to model complex dependencies within the data efficiently. This is achieved by bypassing the need for complicated and time-consuming sequential processing, instead concentrating on the relationships between individual elements in the input and output.

Before the Transformer algorithm is initialized, there are a few steps of data preparation that must be performed by the model. Assuming that our data have been pre-filtered, we proceed directly to the segment of data handling that is relevant to the model configuration. The dataset is initially segmented into discrete units termed 'tokens' and subsequently embedded into vectors. The embedding process can be arbitrarily initiated or adopted from a pre-trained model, as required. In addition, a positional identifier must be appended to the embedded vectors, as, in contrast to conventional RNNs or LSTMs (DiPietro et al., 2020), no information regarding the position of individual tokens within the sequence is inherently encoded. Positional encoding is implemented by the addition of position-dependent vectors to the embedded vectors by using the sine and cosine function (cf. eq. 1 and 2).

$$PE_{(pos,2i)} = sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \tag{1}$$

and

$$PE_{(pos,2i+1)} = cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \tag{2}$$

Following the data preprocessing - embedding and adorning with positional information – the data are partitioned into batches and padded. This entails the aggregation of data into coherent groups and, should the vectors vary in dimensionality, they are homogenized to equal length by padding. Subsequently, the weights for the model, including both the encoder and decoder, are initialized. This can occur randomly or may be conferred upon the model from a previously trained model, should such information be pre-existent.

After data preprocessing, the actual part of the Transformer model begins. The encoder consists of a series of n identical layers. Each layer comprises two core sublayers. The first sublayer utilizes the principle of self-attention, allowing each token in a batch to integrate information from all other tokens in the input sequence. This functionality enables the model to discern relationships within the sequence. The multi-head component allows the network to highlight various aspects of the data and to parallelly extract and combine information, which leads to an enhanced contextual understanding.

In practice, matrices are formed from the vectors of the input sequence, which constitute the base for the Query (Q), Key (K), and Value (V) matrices. The computation of the scalar product between Q and K matrices followed by the application of the softmax function yields attention weights, which are then used to weight the V matrix, as shown in eq. 3. The subsequent sublayer is a simple, fully connected feed-forward layer that complements each attention layer. Here, two linear transformations are applied, with an activation function - often ReLU or GELU (Zhang et al., 2021) - employed between them to enable the model to learn non-linear patterns. Even though the feed-forward network is applied to each position separately, the underlying parameters remain constant; this makes the transformation process independent of the exact position of the token and facilitates efficient parallel processing.

$$Attention(Q,K,V) = softmax\left(\frac{Q \times K^T}{\sqrt{d_K}}\right) \times V \tag{3}$$

Each sublayer within the encoder, e.g. the self-attention or the feed-forward layer, is accompanied by its own residual connection. This residual connection promotes the flow of information in the network by forwarding the input directly to the output and then combining it with the processed output of the respective sublayer. To ensure high-quality information, each combination of input and processed output is further embellished with a normalization layer, usually layer normalization. Layer normalization is a crucial step that helps to stabilize the distribution of activations by normalizing means and standard deviations, making training more stable and speeding up convergence. The normalization is calculated separately for each input and typically occurs after the residual connections and before entering the feed-forward layer. The integration of residual connections has proven particularly effective in allowing networks to flexibly weigh original information and prevent the vanishing gradient problem, especially in deep networks. This simplifies training and reduces the risk of learning difficulties due to vanishing or exploding gradients.

In sum, the mentioned mechanisms within the Transformer model represent innovative means to learn complex patterns while simultaneously maximizing efficiency through parallelization. Especially the activation functions ReLU and GELU play a significant role in these architectures, contributing to the necessary nonlinearity in the model.

After the encoder layer has been processed, the data is passed on to the decoder component to make predictions for the model. To establish a connection between encoder and decoder layers, the masked self-attention mechanism of the decoder must be activated. This enables the Transformer to perform autoregressive processing of the data and make predictions for previous data points in the sequence across the entire sequence without considering future information.

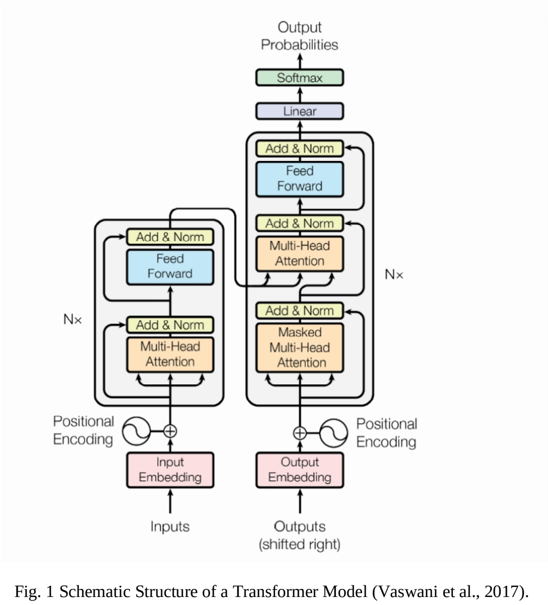

Fig. 1 Schematic Structure of a Transformer Model (Vaswani et al., 2017).

If the input data, as in our example, consists of numerical sequences, additional feed-forward networks can be integrated into the model to create more complex dependencies between data points and generate a better understanding of patterns and relationships in the data. The decoder iteratively performs these calculations until a specific criterion is met, or the sequence transitions to the next set of data. It is noteworthy that the decoder proceeds iteratively by making stepwise predictions based on the already processed sequence segments without anticipating future information.

Both main components consist of a series of sub-layers that are structurally very similar. Despite some minor differences in their functionality and alignment, they share the fundamental architecture that makes the Transformer a particularly remarkable model. A schematic representation of the Transformer model is shown in figure 1.

3. Transformer algorithms – state of the art

The Transformer model, first introduced in the paper "Attention is All You Need" by Vaswani et al. (2017), represents a groundbreaking architecture in the field of machine learning, particularly in the context of natural language processing (NLP). Since its publication, it has significantly transformed the approach to sequence-to-sequence tasks such as machine translation, text summarization, and question-answering systems (Vaswani et al., 2017).

The key innovation introduced by Transformer models is the self-attention mechanism. This mechanism allows the model to compute weighted relations between different words of a sequence regardless of their distance. Compared to earlier approaches such as RNNs (Recurrent Neural Networks) and LSTM (Long Short-Term Memory), this gives the Transformer model a remarkable ability for parallel processing and efficiency with long sequences (Hochreiter & Schmidhuber, 1997; Cho et al., 2014).

Over the years, Transformer models have experienced continuous developments and have become a central component of many current NLP systems. Examples include variations such as Google's BERT (Bidirectional Encoder Representations from Transformers), which has set new benchmarks for the performance of bidirectional models (Devlin et al., 2018), as well as OpenAI's GPT (Generative Pretrained Transformer) and their subsequent iterations, which have achieved state-of-the-art results in a variety of NLP tasks (Radford et al., 2018).

ChatGPT and Google Bard are two examples of AI-driven conversational systems based on Transformer architectures. Google Bard, as a more recent project from Google, shares similarities with ChatGPT as a large-scale language model. While few details about Google Bard are available at the time of the knowledge cutoff, ChatGPT is based on OpenAI's GPT architecture, an advancement of the original Transformer models, which can be pretrained for a broad range of NLP tasks and subsequently fine-tuned, including for use as interactive chatbot interfaces (Brown et al., 2020).

However, Transformer architectures aren't just finding favor in natural language processing; practical examples show the potential for industry. The application of the deep learning Transformer algorithm in engineering has been explored in various studies. Xu (2019) developed a Transformer image recognition system based on deep learning, achieving accurate classification of transformer images. Islam (2023) conducted a comprehensive survey of Transformer applications, highlighting its potential in computer vision, audio and speech processing, and signal processing. Zhang (2022) proposed a deep transfer learning method using a Transformer with self-attention for industrial sensor fusion tasks, demonstrating its ability to reduce data requirements and improve prediction accuracy. Cheng (2022) reviewed the development and application of Transformers, emphasizing their role in feature learning and their potential in computer vision and natural language processing. These studies collectively underscore the transformative potential of the Transformer algorithm in engineering applications.

4. Data generation

The research presented in this paper encompasses the analysis of surface imagery from a collection comprising 951 chef knives, also with 8-inch blades, as shown in fig. 2. These images were captured using a specially designed experimental setup that ensures uniform environmental conditions for each measurement. In addition to the imagery, baseline data on surface roughness characteristics were systematically recorded for all knives under examination.

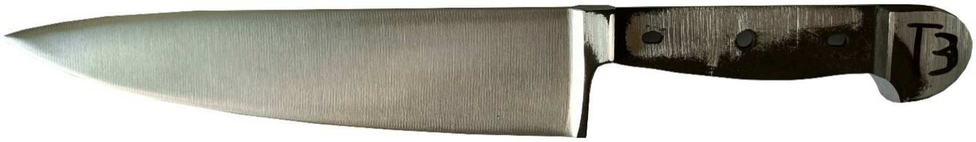

Fig. 2 Example of a chef knife with a 8-inch blade.

This study refers to traditional measurement techniques, capable of being executed with widely available standalone instruments, as "classical measurements". These methods do not require the application of additional computational steps beyond the initial data capture. While some instruments, such as the confocal system discussed herein, necessitate proprietary software for data processing, they do not demand custom programming or the application of complex mathematical models. Primary among the instruments employed in this investigation are:

- A Roughness Tester, which operates on the principle of a piezoelectric micro-probe. This device traverses a 6mm line, gauging the surface's irregularities to produce a singular value per measurement.
- An optical 3D microscope, leveraging the confocal measuring principle to stitch together 2D edge slices at varying heights into a cohesive 3D model. For a detailed discussion on this technique, refer to (Price, 2011). This method's precision surpasses that of piezoelectric devices and measures roughness across a predefined area rather than a single line.

This research utilized the MarSurf CM mobile from Mahr, equipped with a 320L lens, for confocal measurements.

Measurements were conducted 1cm below a designated 6mm line to mitigate spatial uncertainties. To ensure consistency across measurements, a custom 3D printed holder was used for both the confocal measurements and within the experimental rig, which was specifically designed for this study.

For consistent imaging across different knife types, two analogous experimental rigs were devised, each ensuring uniform lighting conditions. Essentially, a single rig was adapted for dual experimental setups, featuring white interior walls to exclude external light and diffuse internal illumination, thus preventing reflective glares on the knife surfaces. The knives were secured using a 3D printed mount for accurate positioning. Lighting was provided by two LED spotlights, aimed at the enclosure's top and bottom walls in the first setup and directly at the knives in the second, to highlight the distinctive markings left by the grinding process. The imaging equipment comprised:

- An Olympus E-520 DSLR with an Olympus Zuiko Digital 14-42mm f 1:3.5-5.6 lens for the initial experiment, capturing images at a lower resolution of 3648px by 2736px, which were then cropped to 1250px by 550px.
- A Canon EOS 77D equipped with a Canon MP-E 65mm f/2.8 1-5x macro lens for the subsequent experiment, offering higher resolution images of 6000px by 4000px without the need for cropping, thanks to the superior lens quality.

The camera was positioned normal to the knife surfaces, at a distance of 8cm, to ensure optimal imaging.

5. Model description

In this study, we conducted a comprehensive analysis of 951 samples of a specific knife type based on the Transformer algorithm to further deepen our understanding of the data and the possibilities of the application of the algorithm in the analysis of the surface quality. Each measuring object was analyzed under the same conditions and divided into ten segments, which ensured the consistency of the analysis process.

As in the previous study (cf. Hinz et al., 2023), the surface properties of the measuring objects were examined based on the brightness (gray value between 0 and 255) of each data point. The brightness values were rescaled to a standardized scale from 0.00 to 100.00, to enable a standardized evaluation of the surface quality. The images for this analysis have a high resolution of 6000 x 4000 pixels to maintain the standard set by Hinz et al. (2022) and to enable a detailed investigation of the fine patterns that can be derived from tactile roughness measurement and confocal measurement.

The analysis was carried out for the chef knives, as described in the original work. Unlike the findings there, which showed different class distributions depending on the type of knife and measurement method, all data series were used in this analysis, regardless of the unequal class distribution. This made it possible to train and interpret the data model for patterns that occur less frequently. In total, this resulted in a dataset of 57,060,000 data points that reflects the temporal development of the brightness values for each of the 951 measuring objects across the various segments, building on the work of Hinz et al. (2022). The impressive scope of the dataset provides a solid foundation for machine learning and enables fine-grained analyses, which are essential for precise classifications regarding the surface quality of the measuring objects. Classification was divided into three classes using average roughness ("Ra") in µm and subsequently binarized using the softmax function for numerical purposes.

The approach of this study is based on the realization that the middle class, which represents well-produced knives, is the most common, and this is also of great importance for the economic efficiency of the manufacturing process. Furthermore, as in Hinz et al. (2023), not only entire rows of pixels were analyzed, but they were also divided into smaller sections to maximize the amount of training data, and it was ensured that the divided pixel rows corresponded to the original marking of the entire pixel row of 1250 pixels. Through this detailed analysis and the appropriate data handling strategy, this work aims to define potential processing standards and improve production efficiency and quality assurance in knife manufacturing.

Due to the uneven distribution of classes in the dataset, it was necessary to compensate for this imbalance. By performing a separate calculation of the distribution for each target class, it was possible to determine corresponding weighting factors that should support balanced training of the model. These weighting factors reflect the inverse ratio of class frequencies and serve as multipliers for the associated class loss during training. In the course of the training process, these three calculated numerical values were used as class weights to compensate for the underrepresentation of certain classes. The use of class weights primarily helps ensure that the model suffers a higher loss when it makes mistakes with rare classes, contributing to a more balanced and fairer learning process. This strategy was implemented directly in the training code by passing these weights to the loss function of the transformer model. The effects of class weighting were evident in the improved ability of the model to recognize and accurately classify all classes.

The existing data were fed into a transformer-based model for the classification of lightness measurements. This model is distinguished by its high adaptability and performance in assessing surface qualities. The core of our approach lies in the efficient use of positional encoding and transformer encoder components for processing sequential data, which was realized without the use of a decoder. This methodological innovation brings the extracted features to the fore and highlights the innovation potential for similar industrial applications.

The deliberate choice of input dimension of 6001 ensures the processing of the complex feature landscape of our dataset and ensures that the model captures all relevant information for classification. With a model depth of four encoder layers and the same number of heads in the multi-head attention, a good balance was achieved between the model's capability and the risk of overfitting. The dropout of 0.1 improves generalizability and prevents overfitting, while the L2 regularization (weight_decay) of 0.0001 allows the model to converge quickly and stably during training.

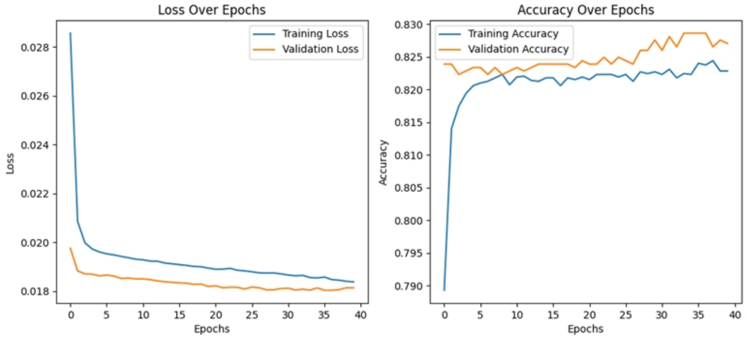

Fig. 3. Loss over epochs and accuracy over epochs.

Our consideration in the training process, reflected by the execution of 40 epochs in batches of 16 datasets each, demonstrates the pursuit of an effective learning curve and reliable results. These decisions have led to stable model accuracy, which showed a good precision during the validation phase. Of particular importance was the model's ability to detect differences in lightness values, indicating the effectiveness of the model architecture.

Nonetheless, there is a high potential of the model's optimization. Future work may include further fine-tuning of hyperparameters. Here, it is essential to consider a full hyperparameter study to ensure more robust and generalizable predictions by the model.

Model final model parameters with the best results:

- Input Dimension: input_dim = 6001
- Hidden Dimension (Embedding Dimension): hidden_dim = 64
- Number of Encoder Layers: num_layers = 4
- Number of Heads in Multi-Head Attention: num_heads = 4
- Dropout Probability: dropout_prob = 0.1

Optimization Parameters:

- Class Weights: class_weights = torch.tensor([0.458, 2.099, 2.908], dtype=torch.float32)
- Weight Decay (L2 Regularization): weight_decay = 0.0001
- Learning Rate and Optimizer: Adam with learning rate lr=0.0001

Training Settings:

- Batch Size: batch_size = 16
- Number of Epochs: num_epochs = 40

In the current study, a transformer-based model for the classification of data patterns was designed and evaluated. In contrast to earlier approaches, continuous optimizations achieved a significant increase in accuracy to approximately 83% for validation data. This is particularly noteworthy regarding classes ‘1.0’ and ‘2.0’, where the model achieved significantly high values for precision and recall. Since class ‘0.0’ is strongly underrepresented, an approach still needs to be found to provide reproducible results here. The testing and analysis of the training and validation results of the transformer-based model reveal accurate and comprehensively solid performance over various epochs. Over periods of different training epochs (30, 40, and 50), the model showed consistent accuracy on the test set with an average value of 82.60% ± 0.19%. This remarkably low standard deviation informs us about the reliability and precision of the model's behavior, indicating a low variance in the model's predictive performance with different datasets.

Through careful data analysis, it was found that the model is not only precise but also stable in its performance and delivers trustworthy results for the quality assessment of the measuring objects. This is particularly noteworthy as it shows that the model is insensitive to minor fluctuations in the training data, which underscores the reliability of the model. This low standard deviation is an indication that the model is robust enough to balance the inherent differences and inequalities in the training data. This suggests that the model should also follow a consistent performance when processing new, similar datasets. Thus, it could be a reliable tool in production and quality control by enhancing automation and efficiency in detecting and classifying knife objects.

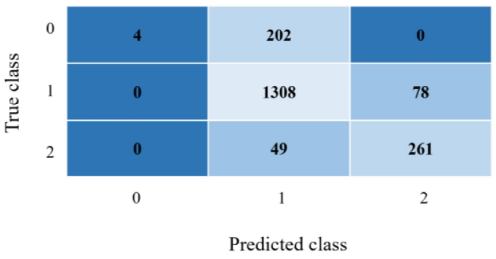

Fig. 4. Confusion Matrix.

The dynamic adaptability of the transformer was manifested in steadily decreasing loss values during the training process. This phenomenon highlights the increasing coherence of the model with the underlying structures of the datasets. The consideration of the achieved F1-scores is also decisive: The micro F1-score of 0.8270 illustrates the remarkable consistency of model predictions across the spectrum of all classes. In parallel, the Weighted F1-Score of 0.7804 reflects the model's ability to deliver reliable classification performances despite the uneven class frequencies.

The results emerging from the study demonstrate the potential of the transformer architecture when it comes to unraveling complex structural patterns in datasets and making them usable for classification purposes. They provide a resilient basis for further research work aimed at deepening the understanding of the applications of this model class and at further enhancing the precision of classification algorithms.

Table 1. Performance Metrics for the Transformer Classification Model.

	Precision	Recall	F1-Score	Support
Class 0.0	1.0	0.02	0.04	206
Class 1.0	0.84	0.94	0.89	1386
Class 2.0	0.77	0.85	0.81	310
Accuracy			0.83	

6. Comparison of Transformer and LSTM algorithm

Incorporating the given introduction, the comparison and summary of the Transformer and LSTM algorithms reveal their distinct capabilities in enhancing automated surface quality assessment in cutlery manufacturing. The Transformer algorithm, recognized for its parallel data processing and self-attention mechanisms, achieved a validation accuracy of approximately 83%, showcasing remarkable precision and recall, especially in specific classes. Its low standard deviation across various epochs indicates a stable and reliable performance, making it a promising tool for quality control.

Conversely, the LSTM algorithm, known for its sequential data handling, demonstrated an ability to adapt and improve, achieving a validation accuracy of up to 84.45% with chef knives, suggesting a robust model suitable for similar product assessments (cf. Hinz et al. 2023). Despite challenges with overfitting, strategic parameter adjustments led to significant performance enhancements.

Both algorithms exhibit potential for revolutionizing surface quality analysis, with the Transformer providing consistent, reliable predictions and the LSTM showing adaptability and improved accuracy with parameter optimization. This comparison underscores the importance of selecting the appropriate machine learning model based on specific manufacturing contexts and the nature of the data being analyzed.

7. Summary and outlook

The study's findings underscore the Transformer architecture's potential in analyzing complex structural patterns in datasets for classification purposes within the analysis of complex surfaces. It sets a foundation for further research to deepen the understanding of Transformer model applications in engineering and enhance classification algorithm precision.

The paper concludes with the potential of the Transformer algorithm in enhancing surface quality assessment through machine learning. It emphasizes the algorithm's efficiency in classifying complex data patterns and its superiority in handling data sequences.

Looking forward, the authors of this study suggest the enhancement of the model's effectiveness through deeper explorations into hyperparameter tuning, algorithmic modifications, and the integration of additional data sources. Emphasis is placed on the need for extensive hyperparameter studies to discover optimal configurations that further increase the robustness and accuracy of the Transformer model. Moreover, the authors propose expanding the application scope of the Transformer algorithm beyond the current study's context, exploring its utility in other manufacturing processes and quality control scenarios.

Acknowledgements

We would like to express our gratitude to **Mr. Mario Adam** and the entire **Mahr GmbH** company for their support in our research work. In particular, we would like to thank them for lending us their **MarSurf CM** confocal measuring system, which played a critical role in our experiments. The use of this machine allowed us to obtain valuable data that contributed significantly to our research findings. We appreciate their assistance and support throughout the duration of the project. Without their generous help, this work would not have been possible.

References

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P. 2020. Language models are few-shot learners. Advances in Neural Information Processing Systems, 33.

Cheng, B., Guo, N., Zhu, Y. 2022. Researches Advanced in the Development and Application of Transformers. Highlights in Science, Engineering and Technology 16, 155–167.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation, 1406.1078.

Devlin, J., Chang, M. W., Lee, K., Toutanova, K. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding, 1810.04805.

DiPietro, R., Hager, G.D. 2020. Deep learning: RNNs and LSTM. In: . Handbook of Medical Image Computing and Computer Assisted Intervention: Elsevier, 503–519.

Hinz, M., Pietruschka, J., Bracke, S. 2022. LSTM based Condition Monitoring of Fine Grinded Surfaces. ESREL 2022, Dublin, Ireland.

Hinz, M., Pietruschka, J., Bracke, S. 2023. Uncertainty Quantification of Different Data Sources with Regard to a LSTM Analysis of Grinded Surfaces. ESREL 2023, Southampton, England.

Hochreiter, S., & Schmidhuber, J. 1997. Long short-term memory. Neural Computation 9(8), 1735-1780.

Islam, S., Elmekki. H., Elsebai, A., Bentahar, J., Drawel, N., Rjoub, G., Pedrycz, W. 2023. A Comprehensive Survey on Applications of Transformers for Deep Learning Tasks.

Price, R. L. 2011. Basic Confocal Microscopy. New York, NY: Springer New York.

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I. 2018. Improving language understanding by generative pretraining.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N. 2017. Attention is all you need. Advances in Neural Information Processing Systems, 30.

Xu, Y., Jin, T., Xu, Y., Shi, X., Chen, S., Sun, W., Xue, Y., Wu, H. 2019. Transformer Image Recognition System Based on Deep Learning. 2019 6th International Conference on Systems and Informatics (ICSAI) IEEE, 595–599.

Zhang, X., Chang, D., Qi, W., Zhan, Z. 2021. A Study on Different Functionalities and Performances among Different Activation Functions across Different ANNs for Image Classification. Journal of Physics: Conference Series 1732 (1), 12026.

Zhang, Z., Farnsworth, M., Song, B., Tiwari, D., Tiwari, A. 2022. Deep Transfer Learning With Self-Attention for Industry Sensor Fusion Tasks. IEEE Sensors Journal 22 (15), 15235–1524.