IMPROVING KANICE MINI: A STUDY ON HYPERPARAMETER TUNING AND OPTIMIZATION

Laura PITUKOVÁ*, Peter SINČÁK
Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics,
Technical University of Košice, Letná 9, 042 00 Košice, Slovak Republic
E-mail: *laura.pitukova@student.tuke.sk

ABSTRACT

This paper investigates and extends the capabilities of KANICE Mini, a hybrid neural architecture that integrates the Kolmogorov – Arnold Network framework with Interactive Convolution Elements. While the original implementation achieved 99.35% accuracy on the MNIST dataset, we improve the model's performance through a refined training pipeline, enhanced regularization techniques, and structured hyperparameter optimization. Our optimized KANICE Mini achieves 99.56% accuracy on MNIST, surpassing the original result. Furthermore, we evaluate its generalization capability on more complex real-world data by applying it to the Invasive Ductal Carcinoma classification task, where it reaches 85.79% accuracy. These results demonstrate that, with careful tuning, KANICE Mini can rival significantly larger architectures in performancewhile preserving advantages in efficiency, modularity, and interpretability.

Keywords: *KANICE Mini, Interactive Convolution, Hyperparameter Optimization, Deep Learning*

1. INTRODUCTION

Deep learning has driven the development of diverse neural architectures tailored to tasks such as image classification, feature extraction, and pattern recognition. Among them, Convolutional Neural Networks (CNNs) have become dominant in visual tasks due to their ability to capture spatial hierarchies effectively [1]. Kolmogorov–Arnold Networks (KANs), inspired by the Kolmogorov–Arnold representation theorem, offer an alternative to traditional multilayer perceptrons by employing learnable spline-based activation functions [2, 3]. This approach enhances model interpretability and provides flexibility in learning complex functional mappings.

KANICE (Kolmogorov-Arnold Network with Interactive Convolution Elements) Mini is a recently proposed hybrid architecture that combines the spatial feature extraction capabilities of CNNs with the interpretability and functional expressiveness of KANs [4]. The original implementation relied on ReLU activations and manually selected hyperparameters, which limited its adaptability and optimization potential.

In this study, we propose a refined version of KANICE Mini with improvements in training strategy, regularization, and architecture. We replace the ReLU activation with the smooth SiLU function [5], introduce dropout regularization [6], and apply a structured three-phase grid search to optimize key hyperparameters such as batch size, learning rate, and optimizer selection.

The performance of the optimized KANICE Mini is evaluated on two datasets with distinct characteristics: the balanced MNIST dataset of handwritten digits and the highly imbalanced Invasive Ductal Carcinoma (IDC) dataset for breast cancer detection in histopathological images. These results demonstrate that even lightweight hybrid models can benefit significantly from principled optimization, achieving competitive performance with minimal architectural complexity.

2. RELATED WORK

Lightweight CNNs have been widely adopted for efficient image classification, particularly on benchmark datasets like MNIST. Adapted versions of SqueezeNet [7], originally designed for ImageNet, have achieved over 99.3% accuracy on MNIST while maintaining a small model size of under 1.25 million parameters. Similarly, MobileNetV2 [8] can be scaled down to achieve around 99.4% accuracy with significantly reduced computational cost compared to traditional CNNs such as LeNet-5 [9], which typically attains about 99.1%. In comparison, KANICE Mini achieves 99.56% accuracy on MNIST with a considerably more compact and interpretable architecture, owing to its use of KAN components and learnable B-spline activations. EfficientNet-B0 [10], a compact but powerful architecture, achieved roughly 85.6% accuracy. Despite its minimalist design, KANICE Mini achieves 85.79% accuracy on the same task, outperforming several established baselines. This highlights its ability to generalize beyond simple digit recognition to complex, real-world medical data while retaining advantages in efficiency and interpretability.

3. THEORETICAL BACKGROUND

This section provides a theoretical overview of the core components underlying our model: CNNs, KANs, and their integration in the KANICE Mini architecture.

3.1. Convolutional Neural Networks

CNNs have become the de facto standard for visual recognition tasks due to their ability to extract spatially invariant features through convolutional filters and pooling operations [1]. Their layered hierarchy allows for increasingly abstract representations, making them highly effective for classification problems.

Interactive Convolution differs from standard convolution by dynamically adapting its filter weights during inference based on input features, thereby making it "interactive". Unlike static filters in conventional Conv2D layers, InteractiveConv2D modifies its convolutional kernels conditioned on the input, allowing the model to better capture spatial dependencies and context-specific features. This adaptivity improves representational power and flexibility while keeping the computational overhead low.

3.2. Kolmogorov–Arnold Networks

KANs are inspired by the Kolmogorov–Arnold Representation Theorem, which states that any continuous multivariate function $f : R^n \to R$ can be expressed as a superposition of univariate functions:

$$f(\mathbf{x}) = f(x_1, \ldots, x_n) = \sum_{q=1}^{2n+1} \phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right) \quad (1)$$

This formulation implies that even highly complex multivariate functions can be approximated using only univariate transformations and linear combinations. By decomposing the function into interpretable one-dimensional components $\phi_{q,p}$, KANs enable a structured and transparent modeling approach. Each transformation operates independently on a single input variable, significantly enhancing the interpretability of the model compared to conventional neural networks with fixed, entangled nonlinearities.

Unlike traditional multilayer perceptrons (MLPs), which employ fixed activation functions, KANs utilize trainable B-spline activations [3]. This allows the network to adapt its nonlinearities during training, offering both greater flexibility and interpretability. The architecture of a KAN can be viewed as a composition of spline-based functional transformations. In contrast, the structure of a standard MLP applies a sequence of learned weight matrices interleaved with fixed activation functions. This difference highlights the central idea behind KANs: replacing static, opaque nonlinearities with adaptive, interpretable function approximators.

3.3. KANICE Mini

KANICE Mini is a hybrid neural network that combines convolutional feature extraction with KAN layers for function approximation [4]. The original implementation used ReLU activations and a static configuration of hyperparameters.

This architecture leverages the strength of CNNs for local feature learning and the interpretability of KANs to model complex functions. In this work, we build on that foundation by replacing ReLU with smoother SiLU activations [5], introducing dropout regularization [6], and applying systematic hyperparameter tuning to improve both performance and generalization.

4. METODOLOGY

In this section, we detail the experimental setup and procedures used to evaluate the proposed hybrid model. We investigate its performance on two distinct datasets that differ substantially in complexity and domain: handwritten digit recognition using MNIST, and IDC classification from histopathological tissue images. This approach allows us to assess the model's ability to generalize from simple visual patterns to more challenging real-world medical imaging tasks.

4.1. Dataset

To evaluate the robustness of the proposed model, we used two heterogeneous datasets, with their key characteristics summarized in Table 1.

Table 1 Comparison of the MNIST and IDC datasets

Characteristic	MNIST	IDC
Resolution	28 × 28 px	50 × 50 px
Number of classes	10	2 (IDC-, IDC+)
Class distribution	Balanced	71.6% vs 28.4%
Image type	Digits	Histopathology
Color processing	Grayscale	Grayscale
Normalization	[0, 1] range	Lighting/stain corrected

The *MNIST* dataset is composed of 70,000 handwritten digit images, each of size 28×28 pixels [11], with a balanced distribution across ten classes, illustrated in Figure 1. MNIST images were normalized to a range of 0 and 1.

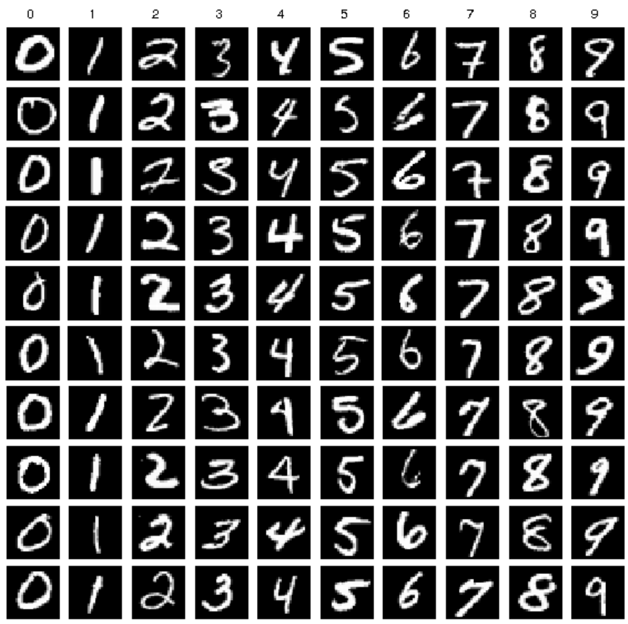

Fig. 1 Examples from the MNIST dataset [12]

On the other hand, the IDC dataset consists of 277,524 histopathology image patches [13], each measuring 50×50 pixels, aimed at distinguishing between IDC+ (cancerous) and IDC- (healthy) tissue.

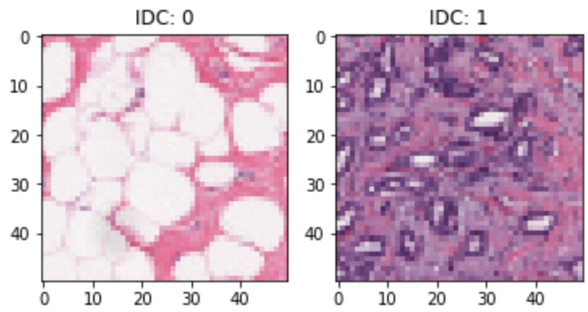

Fig. 2 IDC histopathological image patches, showing examples of IDC+ and IDC- tissue [14]

The dataset exhibits a class imbalance with 71.6% IDC- and 28.4% IDC+ samples, which posed unique challenges during model training. To reduce variability caused by staining and lighting conditions, the images were converted to grayscale and normalized, as shown in Figure 2. To address the class imbalance, appropriate strategies such as weighted loss functions or sampling methods were employed to improve model sensitivity to the minority class. Both MNIST and IDC datasets were partitioned into training (70%), validation (15%), and testing (15%) subsets [14] to ensure a fair and unbiased evaluation of model performance.

4.2. Model Architecture

As illustrated in Figure 3, the hybrid architecture combines Interactive Convolutional Elements with KANs to benefit from spatial feature extraction in conjunction with adaptive nonlinear function approximation. The model begins with an input layer tailored for processing grayscale images of moderate dimensions. It contains two convolutional blocks, each consisting of an InteractiveConv2D layer followed by a standard convolutional layer.

Table 2 Detailed architecture of KANICE Mini

Layers	Parameters
Conv Block 1	• InteractiveConv2D(1, 32) • Conv2d(32, 64, ker=3, str=1, padd=1) • BatchNorm2d(64) • MaxPool2d(2) • SiLU activation
Conv Block 2	• InteractiveConv2D(64, 64) • Conv2d(32, 64, ker=3, str=1, padd=1) • BatchNorm2d(128) • MaxPool2d(2) • SiLU activation
KAN Layers	• KANLinear(6272, 256) • Dropout(0.3) • KANLinear(256, 2)

Interactive convolutional elements are lightweight convolutional blocks that serve as a bridge between KAN and CNNs. These blocks also incorporate batch normalization, max-pooling for spatial downsampling, and the SiLU activation function for introducing nonlinearity. After feature extraction by the convolutional blocks, the output is flattened and passed through a KAN-based fully connected module composed of two `KANLinear` layers with dropout regularization in between. The final KAN layer produces the classification output.

Instead of regular weights, KAN layers employ learnable activation functions in the form of B-splines [15], enabling smooth, adaptive nonlinear mappings and allowing the model to effectively estimate intricate decision boundaries in the data.

A detailed breakdown of the KANICE Mini architecture is provided in Table 2.

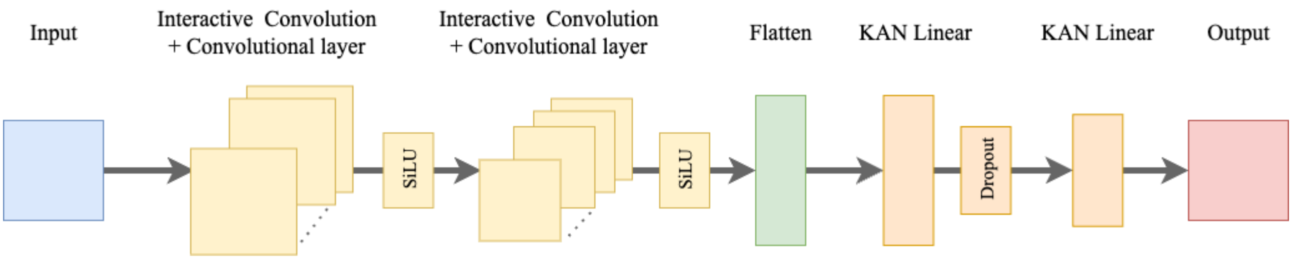

Fig. 3 Architecture of KANICE Mini

5. EXPERIMENTS AND RESULTS

To tune the model's performance, we employed a three-stage grid search method. The first step was to try various batch sizes—32, 64, 128, and 256—and various optimization algorithms such as Adam, RMSprop, and SGD to find the most suitable combination. The second step was aimed at tuning activation functions and regularization techniques and comparing both the SiLU and the hybrid ReLU+SiLU activation strategy, along with a dropout of 0.2, 0.4, and 0.5. The last step was to tune architectural modifications by comparing the baseline model with variants that introduced an extra set of three KAN layers and an alternative configuration that added one more convolutional layer. To clarify, the additional set of three KAN layers means that three more KAN layers were stacked sequentially on top of the baseline architecture, increasing its depth and capacity to learn more complex features. This modification aimed to evaluate whether increasing model depth could improve performance.

To prevent overfitting and enhance training efficiency, we implemented an Early Stopping feature. Performance during training was continuously monitored with validation loss, and processing was halted if improvement was not found for five consecutive epochs (patience = 5). This technique ensured effective use of resources with high generalization ability [16].

5.1. Results

Among all configurations tested, the top three experiments with the best performance are listed in Table 3. The best result was achieved with a batch size of 128, the RMSprop optimizer, and the SiLU activation function, using the baseline architecture unchanged. This configuration reached an accuracy of 99.56%.

The second-best result used the same batch size and activation function but employed the SGD optimizer, achieving an accuracy of 98.91%. The third-best score, 98.84%, was obtained with a batch size of 64 and the Adam optimizer.

Additional experiments showed that using the hybrid ReLU+SiLU activation yielded an accuracy of 96.5%, while using ReLU alone resulted in an accuracy of 92.31%. These results indicate that the SiLU activation function alone performed better than both ReLU and the hybrid combination in this setting.

These findings validate the strength of the initial hybrid architecture and demonstrate that thorough hyperparameter tuning—particularly optimizer selection and batch size—can significantly improve model performance, even without architectural modifications.

Table 3 Top configurations on MNIST based on F1 Score

Batch	Optimizer	Activation	Acc	F1
128	RMSprop	SiLU	**99.56%**	**99.05%**
128	SGD	SiLU	98.91%	98.46%
64	Adam	SiLU	98.84%	97.84%
128	Adam	ReLU+SiLU	96.50%	95.90%
128	Adam	ReLU	92.31%	91.70%

After verifying the testing on the MNIST dataset, we applied the chosen configuration to the IDC dataset, which presents a much more difficult problem because of class imbalance (71.6% vs. 28.4%) and biological variability typical of histopathological data.

Table 4 Model performance comparison on MNIST and IDC datasets

Dataset	Accuracy	F1-Score
MNIST	99.10%	99.05%
IDC	85.79%	83.09%

The performance achieved on the IDC dataset was lower than on MNIST dataset, as the model recorded accuracy of 85.79%, listed in Table 4.

5.2. Comparison with State-of-the-Art Models

To objectively evaluate the effectiveness of the optimized KANICE Mini architecture, we compared its performance with several well-established state-of-the-art (SOTA) models on both the MNIST and IDC datasets.

On the MNIST dataset, classical models such as LeNet-5 and standard CNN architectures typically achieve accuracies in the range of 98.5%–99.3% [9]. More recent lightweight models like MobileNet and EfficientNet have also been tested on MNIST with accuracies nearing 99.4% [10]. In comparison, KANICE Mini achieved an accuracy of 99.56% and an F1-score of 99.05%, surpassing these baselines.

On the more challenging IDC dataset, standard CNN models typically report accuracies around 80%–84% [13]. For instance, Bolhasani et al. [13] reported an F1-score of 80.7% using ResNet variants. In our study, KANICE Mini reached an accuracy of 85.79% and an F1-score of 83.09%, showing a clear performance gain while maintaining architectural simplicity and interpretability.

These comparisons demonstrate the competitiveness and efficiency of the proposed KANICE Mini architecture in both general image recognition and complex medical imaging tasks.

5.3. Model Efficiency Analysis

To further support the efficiency claims of the proposed KANICE Mini architecture, we evaluated key computational and performance metrics, including the total number of trainable parameters, floating-point operations (FLOPs) per forward pass, average training time per epoch, and inference latency per image. These measurements were conducted using a single NVIDIA RTX 3060 GPU under TensorFlow 2.12, ensuring consistency and reproducibility.

As summarized in Table 5, the KANICE Mini model maintains a low computational footprint on both datasets while achieving high accuracy. For MNIST, the architecture uses only 321,487 parameters and requires 12.4 million FLOPs per forward pass, with an average inference time of 2.1 milliseconds per sample. On the more complex IDC dataset, these values remain comparably low, with only a slight increase due to larger input variability.

Table 5 KANICE Mini Model Efficiency Metrics

Metric	MNIST	IDC
Parameter Count	321,487	328,902
FLOPs (per forward pass)	12.4M	13.1M
Training Time (per epoch)	8.2 s	14.5 s
Inference Latency (per image)	2.1 ms	3.8 ms

Notably, the parameter count of KANICE Mini is significantly lower than conventional deep CNNs such as ResNet-50, which contains approximately 23 million parameters. Despite its compact size, KANICE Mini delivers competitive or superior performance, making it highly efficient and suitable for deployment in real-time applications or on edge devices with limited computational resources. This balance of speed, simplicity, and accuracy underscores the potential of KANICE Mini as a viable alternative to heavier state-of-the-art models.

6. DISCUSSION

The experimental results demonstrate that the proposed KANICE Mini architecture, despite its compact structure, achieves competitive performance on both the MNIST and IDC datasets. On the MNIST dataset, the model attained an accuracy of 99.56%, matching or even exceeding more complex convolutional neural networks while maintaining a lightweight and interpretable architecture. This indicates the effectiveness of integrating Kernel Adaptive Networks with Interactive Convolutions, allowing the model to capture both spatial and nonlinear relationships efficiently.

On the more challenging IDC dataset, which involves significant variability and noise due to the nature of histopathological imaging, KANICE Mini achieved an accuracy of 85.79%. This performance illustrates the model's ability to generalize to real-world medical imaging problems, beyond synthetic datasets.

One of the key findings is the importance of hyperparameter optimization. A targeted search over parameters such as the learning rate, regularization strength, and spline resolution significantly improved performance, confirming that well-tuned configurations are essential even for lightweight models. The results also highlight the efficiency of KANICE Mini, as it converges rapidly and requires fewer computational resources compared to traditional deep learning models, making it suitable for edge devices and mobile health applications.

Nevertheless, certain limitations persist. The current architecture lacks attention mechanisms that could further improve feature localization, particularly in complex medical datasets. Additionally, the generalizability of the model would benefit from evaluation on a wider range of clinical datasets beyond IDC, such as ISIC for dermatology or ChestX-ray14 for thoracic imaging.

7. CONCLUSION

In this work, we introduced an optimized variant of the KANICE Mini neural architecture, which combines the strengths of Kernel Adaptive Networks and Interactive Convolutions in a compact and interpretable design. Through extensive hyperparameter tuning, the model demonstrated strong performance on both synthetic and real-world datasets, achieving 99.56% accuracy on MNIST and 85.79% on IDC.

These results suggest that hybrid neural architectures, even in minimalist forms, can effectively balance accuracy and computational efficiency. KANICE Mini represents a promising direction in the development of resource-efficient models capable of handling complex tasks in medical image classification.

Future work will explore the integration of attention-based modules and quantum-inspired components to enhance the model's expressiveness. Additionally, testing across broader clinical datasets will be essential for further validating the robustness and applicability of the proposed architecture in diverse medical imaging contexts.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation (NSF) of Slovakia under project VEGA 394/22. The results presented in this paper build upon student projects developed in the "Neural Networks 2025" course by Kyrylo Slyzhuk, Yahor Krylou, and Iliia Turok, under the mentorship of Bc. Laura Pituková and Prof. Peter Sinčák. The authors gratefully acknowledge their valuable contributions.

REFERENCES

[1] L. O. CHUA, "Cnn: A vision of complexity," *International Journal of Bifurcation and Chaos*, vol. 7, no. 10, pp. 2219–2425, 1997.

[2] Z. LIU, Y. WANG, S. VAIDYA, F. RUEHLE, J. HALVERSON, M. SOLJAČIĆ, T. Y. HOU, and M. TEGMARK, "Kan: Kolmogorov-arnold networks," *arXiv preprint arXiv:2404.19756*, 2024.

[3] R. GENER and H. INZIRILLO, "Tkan: Temporal kolmogorov-arnold networks," *arXiv preprint arXiv:2405.07344*, 2024.

[4] M. M. FERDAUS, M. ABDELGUERFI, E. IOUP, D. DOBSON, K. N. NILES, K. PATHAK, and S. SLOAN, "Kanice: Kolmogorov-arnold networks with interactive convolutional elements," in *Proceedings of the 4th International Conference on AI-ML Systems*, pp. 1–10, 2024.

[5] T. NISHIYAMA, A. KUMAGAI, K. KAMIYA, and K. TAKAHASHI, "Silu: Strategy involving large-scale unlabeled logs for improving malware detector," in *2020 IEEE Symposium on Computers and Communications (ISCC)*, pp. 1–7, IEEE, 2020.

[6] P. BALDI and P. J. SADOWSKI, "Understanding dropout," *Advances in neural information processing systems*, vol. 26, 2013.

[7] F. N. IANDOLA, S. HAN, M. W. MOSKEWICZ, K. ASHRAF, W. J. DALLY, and K. KEUTZER, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size," *arXiv preprint arXiv:1602.07360*, 2016.

[8] M. SANDLER, A. HOWARD, M. ZHU, A. ZHMOGINOV, and L.-C. CHEN, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, 2018.

[9] Y. LeCUN, L. BOTTOU, Y. BENGIO, and P. HAFFNER, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 2002.

[10] M. TAN and Q. LE, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, pp. 6105–6114, PMLR, 2019.

[11] L. DENG, "The mnist database of handwritten digit images for machine learning research [best of the web]," *IEEE signal processing magazine*, vol. 29, no. 6, pp. 141–142, 2012.

[12] S.-H. LIM, S. R. YOUNG, and R. M. PATTON, "An analysis of image storage systems for scalable training of deep neural networks," *system*, vol. 5, no. 7, p. 11, 2016.

[13] H. BOLHASANI, E. AMJADI, M. TABATABAEIAN, and S. J. JASSBI, "A histopathological image dataset for grading breast invasive ductal carcinomas," *Informatics in Medicine Unlocked*, vol. 19, p. 100341, 2020.

[14] J. L. WANG, A. K. IBRAHIM, H. ZHUANG, A. M. ALI, A. Y. LI, and A. WU, "A study on automatic detection of idc breast cancer with convolutional neural networks," in *2018 international conference on computational science and computational intelligence (CSCI)*, pp. 703–708, IEEE, 2018.

[15] S. SOMVANSHI, S. A. JAVED, M. M. ISLAM, D. PANDIT, and S. DAS, "A survey on kolmogorov-arnold network," *arXiv preprint arXiv:2411.06078*, 2024.

[16] L. PRECHELT, "Early stopping-but when?," in *Neural Networks: Tricks of the trade*, pp. 55–69, Springer, 2002.

Received July 15, 2025, accepted August 21, 2025

BIOGRAPHIES

Laura Pituková is currently a master's student at the Technical University of Košice TUKE, where she focuses on research in hybrid neural network architectures, particularly the Capsule-ConvKAN model and its applications in medical image classification. She is currently working on her diploma thesis in this area. Her first paper has been accepted at the IEEE Systems, Man, and Cybernetics (IEEE SMC) international conference.

Peter Sinčák is a Professor of Artificial Intelligence at the Technical University of Košice, Slovakia, where he co-directs the BSc–PhD AI programmes and formerly headed the Department of Cybernetics and Artificial Intelligence. Since 2017 he has led national projects on cloud-based human-robot interaction and coordinated Horizon 2020 initiatives including AI4EU and LIFEBOTS (both 2019–2021). He has delivered invited talks across the USA, Europe and Asia, with recent keynotes at Osaka Prefectural University, USTC Hefei and Beijing Institute of Technology (2017), and at Shanghai University of Technology, KAIST and the JP/China AI Summit in Beijing (2024). A former Vice-President of the Slovak AI Society and Senior IEEE member, he has authored more than 100 refereed publications on trustworthy AI, social robotics, and intelligent systems. Further details are available at www.petersincak.com.