Vision transformers transfer learning for smoking detection in public spaces with transparent AI decisions

Olexander Mazurets[1,*], Maryna Molchanova[1], Olha Zalutska[1], Ihor Kok[1], Vitaly Levashenko[2] and Abdel-Badeeh M. Salem[3]

[1] *Khmelnytskyi National University, 11, Instytuts'ka str., Khmelnytskyi, 29016, Ukraine*

[2] *Zilina University, Univerzitná 8215, 010 26 Žilina, Slovakia*

[3] *Ain Shams University, El-Khalyfa El-Mamoun Street Abbasya, Cairo, Egypt*

Abstract

The automated detection of smoking in public spaces is increasingly critical for ensuring compliance with public health regulations and mitigating the economic burden of smoking-related diseases. However, existing computer vision approaches often struggle with limited labeled data and lack the interpretability required for trustworthy automated surveillance. In this work, we propose a transfer learning method using the Vision Transformer (ViT) architecture combined with transparent AI decision-making mechanisms. Our approach employs partial parameter freezing to adapt to small datasets while preventing overfitting and utilizes Grad-CAM heat maps to provide visual explanations of the model's focus. Experimental results demonstrate that the proposed method achieves an accuracy of 0.98, surpassing the InceptionResNetV2 baseline by 0.011 and outperforming the YOLOv5-small model by 0.045 in Precision and 0.089 in Recall. The study concludes that the integration of explainable AI not only enhances detection performance but also fosters trust by verifying that the model attends to relevant semantic features, thereby contributing directly to Sustainable Development Goals #3 and #11.

Keywords

ViT, smoking detection, transparent AI decisions, transfer learning, machine learning

1. Introduction

Tobacco smoking is one of the main causes of various diseases [1], such as cancer, cardiovascular diseases and respiratory diseases, which significantly reduce the quality of life and increase the economic burden on the health care system [2]. Given restrictions on smoking in public places in many countries [3], automation of detecting violations process of such prohibitions becomes an important component for effective control over compliance with the law [4].

Despite the existence of legal restrictions on smoking in public spaces, the effectiveness of their compliance largely depends on human resources [5], which are limited and not always able to provide full-fledged monitoring in real time [6]. In this context, the use of computer vision technologies, in particular deep learning models, opens up new opportunities for creating automatic surveillance systems capable of identifying smoking facts in video or images with high accuracy and speed [7, 8]. This approach not only helps to increase the efficiency of detecting violations, but can also become an important tool in preventing the harmful effects of passive smoking on the population [9], especially in places of mass gathering of people [10]. In addition, the implementation of such solutions is a step towards the formation of smart cities [11], where technologies contribute to the implementation of state health policy [12].

The task of detecting smoking in public areas can also be contextualized within the framework of the United Nations Sustainable Development Goals, as it involves the application of innovative data

ExplAI-2025: Advanced AI in Explainability and Ethics for the Sustainable Development Goals, November 07, 2025, Khmelnytskyi, Ukraine

[*] Corresponding author.

exe.chong@gmail.com (O. Mazurets); m.o.molchanova@gmail.com (M. Molchanova); zalutsk.olha@gmail.com (O. Zalutska); kokigorek03@gmail.com (I. Kok); vitaly.levashenko@fri.uniza.sk (V. Levashenko); abmsalem@yahoo.com (A. M. Salem)

0000-0002-8900-0650 (O. Mazurets); 0000-0001-9810-936X (M. Molchanova); 0000-0003-1242-3548 (O. Zalutska); 0009-0000-1716-2207 (I. Kok); 0000-0003-1932-3603 (V. Levashenko); 0000-0001-5013-4339 (A. M. Salem)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2025-12-22

management and monitoring technologies to promote public health [13]. It aligns with SDG #3 "Good Health and Well-being", as it offers an instrument for preventing smoking-related health risks and reducing the consequences of passive smoking [14]. At the same time, it corresponds to SDG #11 "Sustainable Cities and Communities", since the use of advanced AI-based surveillance tools strengthens urban safety and supports the development of sustainable and human-centered public spaces [15]. From this perspective, the advancement of interpretable deep learning solutions for smoking detection not only enhances public health protection, but also integrates artificial intelligence into broader strategies of sustainable urban growth.

The main goal of the paper is to improve the accuracy of detecting smoking in images by applying transfer learning using the Vision Transformers architecture with transparent AI decisions. The main contribution of the paper is the method of detecting smoking using deep learning neural networks, which differs from existing ones by using partial parameter fixation and modification of training data, as well as using the explainability of neural network solutions in the form of a heat map. The proposed approach involves freezing all layers of the model, except for the final classifier, which allows adapting the model to a task with a limited amount of data and reducing the risk of overtraining.

2. Related works

The problem of automated smoking identification from digital images is studied in modern scientific works. An improved method for detecting smoking in public places based on an improved version of the YOLOv5-small algorithm is considered in [16]. The main attention is paid to increasing the accuracy and reliability of detecting cigarettes as objects in images with a complex background, small target size, and visual similarity to other objects. The proposed solution is based on the use of hybrid transformer and pyramidal structures to improve feature extraction, spatial generalization, and reduce the impact of visual interference. The authors obtained Precision 0.935 and Recall 0.891 on a specialized dataset of images with cigarettes.

The article [17] presents an approach to automated monitoring of compliance with the smoking ban within a smart city by developing a smoking detection system based on artificial intelligence. The authors propose a framework that combines the use of transfer learning with a pre-trained Inception-ResNetV2 model to classify images as containing or not containing smoking. In addition, the study contributes to the field by creating a new dataset of images in indoor and outdoor conditions, which provides a basis for further research. The proposed method demonstrates high performance on the experimental dataset, achieving Accuracy 0.969, indicating its potential suitability for real-time applications.

In [18] the potential of deep learning for detecting hidden tobacco advertising in media content is investigated. Multimodal model is proposed that combines image and text analysis using neural networks, generative methods and expert reinforcement learning mechanisms. This approach allows to detect smoking even with a limited amount of training data. The researchers [19] provide an overview of modern methods for classifying and detecting smokers using computer vision based on deep learning. The study includes a critical analysis of existing models, used datasets and their effectiveness, and also offers a systematization of approaches used in the literature. To overcome the shortage of labeled data, the authors present a new dataset CigDet, specially created for cigarette localization tasks, and evaluate its effectiveness on different variants of YOLO models, where the best results are achieved by YOLOv9 (mAP = 83.5%). The conceptual architecture of the SURRONE system is also presented, which can be used for real-world monitoring of smoking using drones, with the prospect of integration into public health systems and surveillance technologies.

In the study [20], the application of computer vision for automated detection of visual content related to e-cigarettes on youth-oriented video platforms such as TikTok was demonstrated. The YOLOv7 model was used to detect vaping devices, hands, and vapor clouds in images obtained from posts with popular hashtags. The model achieved high classification Accuracy (up to 0.929) and demonstrated reliable generalization ($F_1 = 0.81$), which demonstrates the feasibility of using such approaches for monitoring visual violations of tobacco policies in the online environment.

A method for indoor smoking detection combining infrared and visible images with the YOLOv9 architecture to improve accuracy in low-visibility conditions is proposed in [21]. By using bispectral fusion, data-level and feature-level optimization, and model training, the system achieves high average accuracy (mAP@0.5 = 0.958) on a specialized dataset.

Despite the availability of effective deep learning approaches to smoking detection, research into new methods remains relevant due to the need to expand scientific understanding and technical variability. The development of alternative architectures and data processing methods allows not only to reevaluate existing hypotheses regarding feature representation, but also to adapt known solutions to specific operating conditions.

3. Method design

3.1. Problem statement

Despite significant advances in deep learning-based smoking detection, existing methods often face limitations related to performance on small datasets and the need to adapt to various operating conditions. Many modern deep learning models, in particular those used in computer vision, require large amounts of labeled data to achieve high accuracy, which makes their application in real-world conditions difficult. Despite this, the technical potential of transfer learning remains underexplored in the context of specific tasks, such as detecting smoking in public places.

There is a need to develop strategies to improve the adaptability of deep learning models, such as freezing the parameters of all layers except the final classifier, which allows preserving the generalization ability of the model, while reducing the risk of overtraining and improving efficiency in resource-constrained conditions. The main scientific question is how transfer learning using the Vision Transformers architecture can be adapted to the task of smoking detection with limited data volumes, and whether this can lead to increased detection accuracy without the risk of overtraining. Accordingly, there is a need not only to improve the Accuracy of smoking detection, but also to verify whether the neural network correctly highlights the relevant areas of attention in images. Therefore, all AI decisions must be transparent.

3.2. Proposed method

The smoking detection in public spaces method is designed to identify smoking in public places using neural network tools (Figure 1). The input data are images for analysis and a trained neural network model (for more details about the model, see Section 3.3).

In step 1, the image is preprocessed: it is resized to 224×224 pixels, normalized using ImageNet statistics, and transformed into a tensor that meets the requirements of the input layer of the ViT model [22, 23]. After that, the image is automatically partitioned into non-overlapping patches of 16×16 pixels, each of which is prepared for further vector encoding. This partitioning is an intrinsic property of the ViT architecture and is implemented through a convolutional layer that simultaneously performs partitioning and projection. In step 2, each patch is transformed into a fixed-length vector using linear projection. These vectors form a sequence to which positional information is added. A transform encoder operates on the entire sequence of patches, which uses multi-head attention mechanisms to analyze spatial relationships between different regions of the image. The result of this stage is a representation vector for the entire image – a classification token, which contains generalized information about the image.

The resulting vector representation obtained in step 3 is passed to the final linear layer, which was trained for binary classification. The output value is interpreted as a logit, which characterizes the probability of the presence of smoking in public places. A sigmoid activation function is used to obtain a probabilistic result. Thus, the model generates a numerical estimate in range of $[0, 1]$, which is thresholded as the classes “Smoker” or “Non-Smoker”. In the study, the threshold was taken as 0.5 (if less than 0.5, the class “Non-Smoker”, if greater than or equal to – “Smoker”).

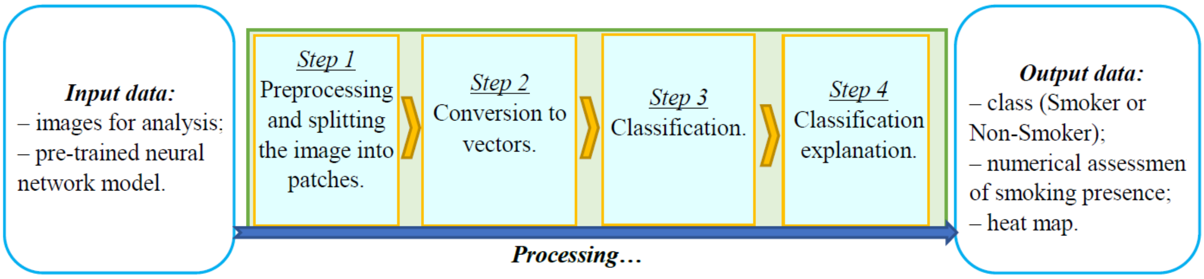

Input data:
– images for analysis;
– pre-trained neural network model.
Step 1
Preprocessing and splitting the image into patches.
Step 2
Conversion to vectors.
Step 3
Classification.
Step 4
Classification explanation.
Processing…
Output data:
– class (Smoker or Non-Smoker);
– numerical assessmen of smoking presence;
– heat map.

Figure 1: Scheme and steps of the smoking detection in public spaces method.

In order to interpret the model solution, step 4 is performed – the Grad-CAM method is applied [24]. This allows visualizing the spatial regions that most influenced the classification [25]. In the context of Vision Transformer, a projection layer of patches is used as a reference for generating an importance map. Based on the calculated gradients, a heat map is formed and superimposed on the image. It shows which areas of the image were decisive for the decision, increasing transparency and confidence in the model [26]. The output data are the class (Smoker or Non-Smoker), a numerical score of the presence of smoking (in the range from 0 to 1), and an interpretive heat map.

3.3. Neural network model

The scheme of converting the input image into the probability of identifying smoking in public places through the applied neural network architecture is shown in Figure 2. At the classification stage, the input image with a size of 224 × 224 pixels is fed to the ViT model, pre-scaled and normalized according to the statistical parameters of ImageNet. The image is divided into 196 non-overlapping patches with a size of 16 × 16 pixels, which form a regular 14 × 14 grid. Each patch is passed through a convolutional layer (patch embedding), which performs linear projection, resulting in a sequence of feature vectors of fixed dimension (768).

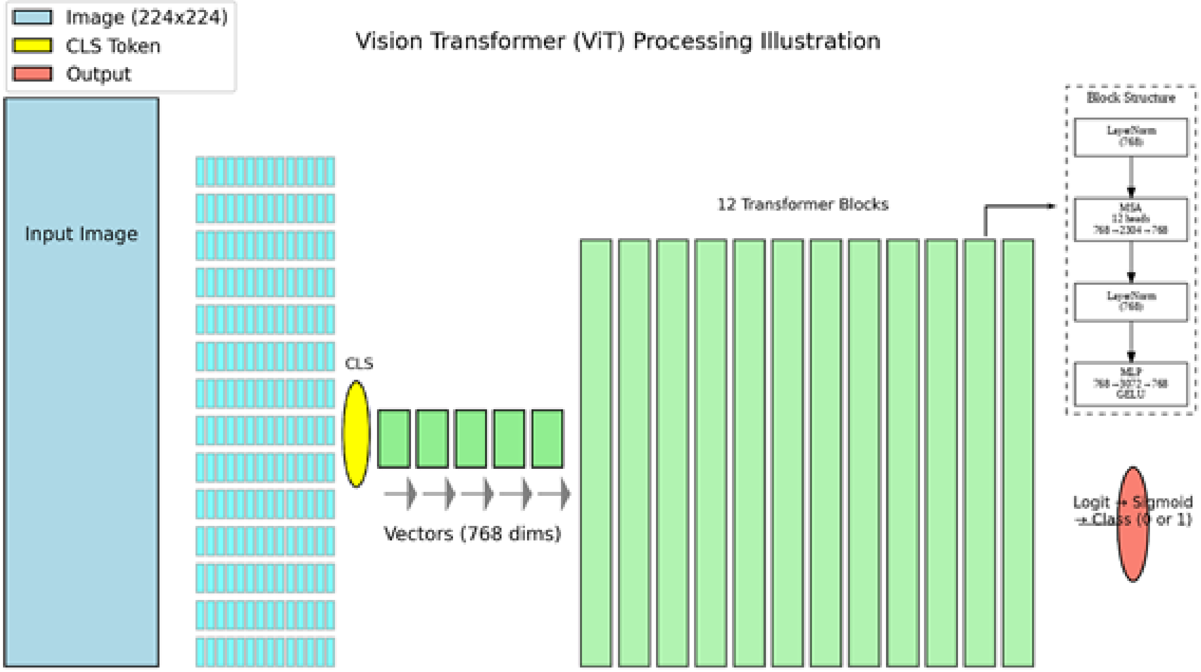


Figure 2: ViT neural network architecture.

A special classification token (CLS-token) is added to this sequence, which is designed to accumulate global information about the image. Thus, the total number of input vectors is 197. These vectors are sequentially passed to the transformer encoder, which consists of 12 serially connected blocks. Each block contains two main components: a Multi-Head Self-Attention (MHSA) mechanism and a Multi-Layer Perceptron Network (MLP), accompanied by layer normalization operations. The attention mechanism provides the model with the ability to capture dependencies between remote patches, allowing, in particular, to establish connections between semantically related regions of the image. MLP layers are responsible for nonlinear feature transformation, and normalization improves the stability of calculations.

After passing through all the transformer blocks, the updated CLS-token acts as an aggregated representation of the entire image. This token is passed to the final classification layer – a linear projection, which outputs a single scalar value (logit). A sigmoid activation function is used to transform the logit into a probabilistic interpretation. The final value is compared with a fixed threshold (0.5), which allows determining a binary class: class “1” is interpreted as smoking signs presence (Smoker), class “0” as their absence (Non-Smoker).

In order to improve the training efficiency, partial freezing of the model parameters was applied, in which only the weights of the last linear layer are optimized. This helps to speed up the training process, as the number of parameters that need to be updated is reduced, which allows to significantly reduce computational costs. The use of pre-trained weights, in particular through the use of SWAG (Stochastic Weight Averaging Gaussian) for linear layers, improves the model’s ability to generalize to new tasks, which is critically important when working with limited data sets. This approach also reduces the risk of overtraining, since the model does not “remember” the specific features of training set [27], but focuses on the general characteristics of the images. In addition, freezing most of the parameters allows to reduce the requirements for computational resources, focusing the calculations only on the optimization of the weights critical for classification, which increases the efficiency of using computational resources when training large architectures. Such a modification helps to improve the quality of classification and reduce the amount of necessary calculations, especially on small datasets.

Also, an image preprocessing procedure is applied before feeding them to the model and is aimed at ensuring structural uniformity and expanding the variability of the training data. In particular, scaling images to a single size guarantees compliance with the architectural requirements of the model, while stochastic augmentations, in particular mirroring and changing color characteristics, contribute to increasing the model’s resistance to changes in external shooting conditions. Normalization of channel intensities improves the stability and convergence speed of the optimization process by eliminating the variability of the scales of the input features. Taken together, these transformations ensure both the invariance of the model to variations in data and correct representation of information for effective training.

4. Experiment

4.1. Dataset

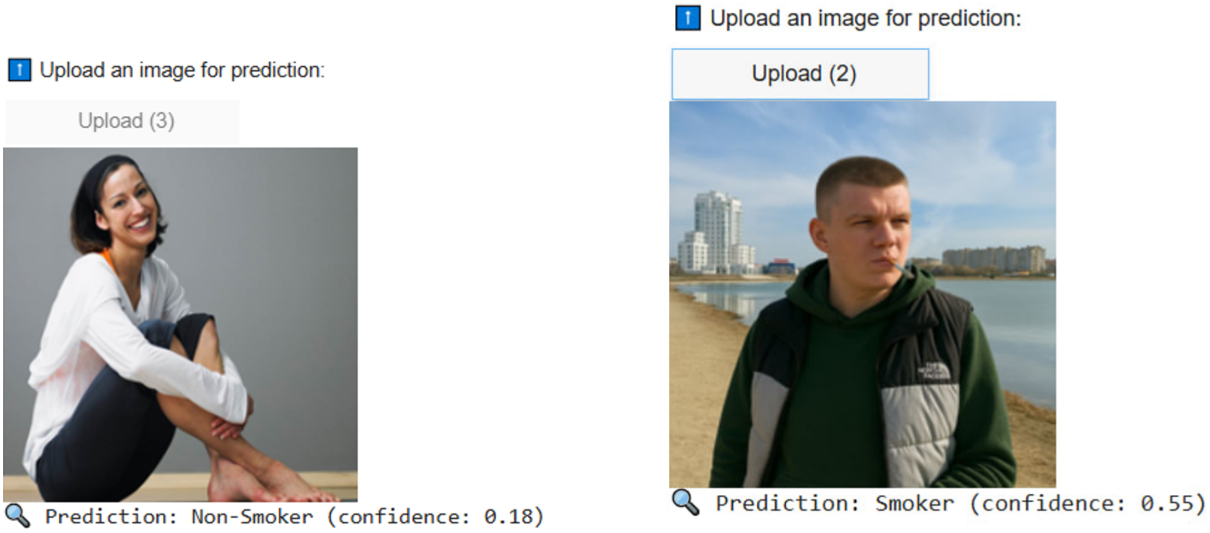
The “Smoker Detection [Image] classification Dataset” [28] represents a balanced sample of visual data consisting of two semantic categories – smokers and non-smokers. Each class includes the same number of images, which ensures symmetry in the representation of the target categories and reduces the bias of the model during training. The total number of samples in the dataset is 1120 (560 in each class).

A feature of this corpus is the presence of a high degree of interclass similarity: images in the non-smoker class were selected in such a way as to contain poses and gestures that are visually close to smoking acts, in order to complicate the classification task. Such a selection strategy increases the discriminatory power of the model and stimulates it to learn deeper contextual features. All images were unified in resolution to ensure compatibility with the architectural requirements of deep learning models.

4.2. Software implementation

To validate the proposed method of smoking detection in public spaces, a software application was implemented in the Python environment using the Google Colab cloud service [29], which provides interactive code execution with hardware acceleration support (GPU/TPU) and facilitates experimentation with image processing and deep model training. The application architecture is built on the basis of the TensorFlow library [30] and its high-level API Keras, which made it possible to implement the Vision Transformer model and organize the training process with the possibility of flexible hyperparameter tuning.

The model results were explained using the Grad-CAM technique, which was implemented by constructing a custom computational graph using automatic differentiation. The Matplotlib [31] and OpenCV [32] libraries were used to visualize heat maps and interpret classification results. An examples of image analysis by the developed software is shown in Figure 3.



```
Upload an image for prediction:
Prediction: Non-Smoker (confidence: 0.18)
```

(a) Input Image 1 (Dataset sample)

```
Upload an image for prediction:
Prediction: Smoker (confidence: 0.55)
```

(b) Input Image 2 (Author photo)

Figure 3: Examples of image analysis for detecting smoking in public spaces: (a) photo 1 from the dataset [28], (b) photo 2 shows the paper author Ihor Kok.

The results obtained and their interpretations are presented below.

5. Results and discussion

The developed software was used to conduct a study of the developed method. Part of the input data of the developed method is a transfer machine learning model. The results of the training are presented in the form of a confusion matrix and a rock curve are shown in Figure 4. Classification report is shown in Table 1. With these parameters, Accuracy value is 0.98.

The results of the experimental study indicate the effectiveness of the proposed approach based on the transfer learning of the Vision Transformer model. The achieved classification accuracy (according to the Accuracy metric = 0.98) indicates the ability of the model to generalize knowledge on new data with a minimum number of errors. High values of precision and recall (over 0.96) in combination with balanced indicators for both classes indicate the absence of a bias of the model towards one of the classes, which is critically important in tasks related to behavioral analytics. The hypothesis that transfer learning using the Vision Transformers architecture can be adapted to the task of smoking detection with limited amounts of data, which will lead to an increase in detection accuracy without the risk of overtraining, is also confirmed by the given graph of the loss function (the neural network was trained for 10 epochs, Figure 5).

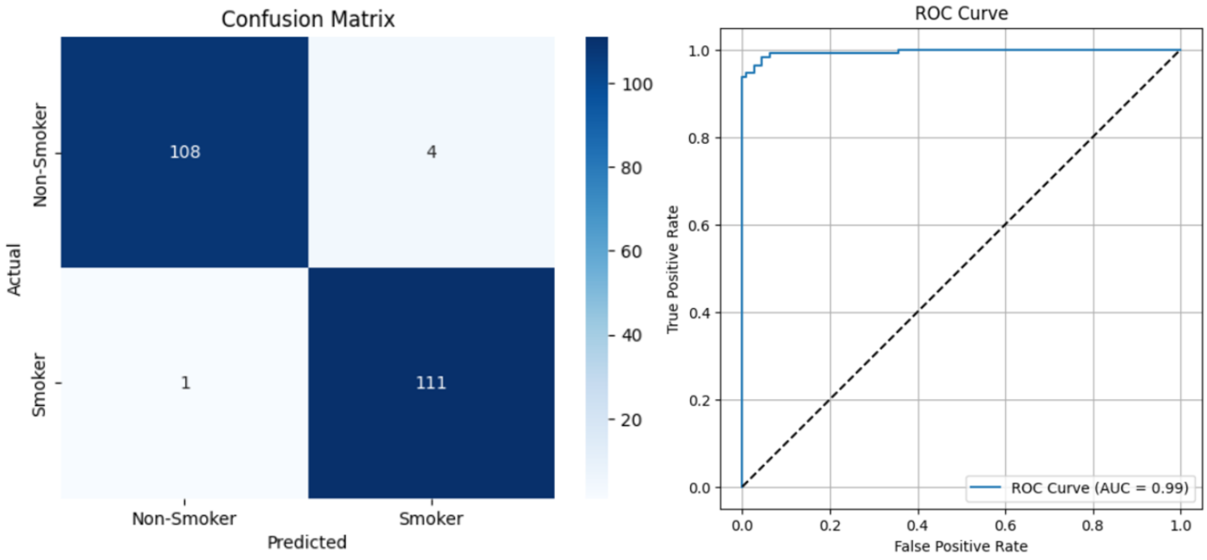
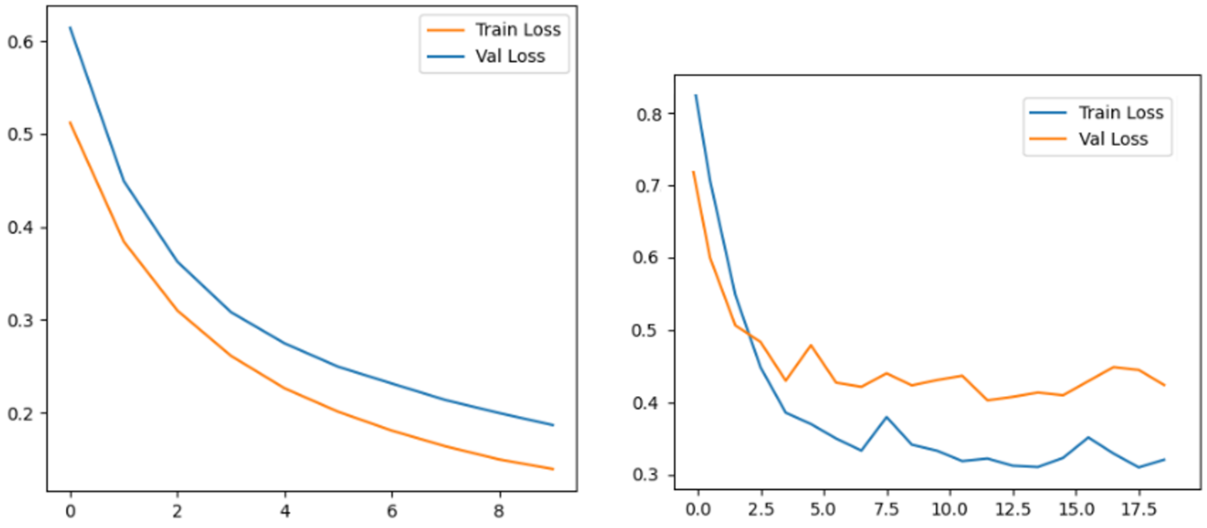


Figure 4: Confusion matrix and ROC curve of the pretrained ViT neural network.

Table 1
Classification report

Class	Precision	Recall	F_1-score	Support
Smoker	0.99	0.96	0.98	112
Non-Smoker	0.97	0.99	0.98	112
Macro metrics				
Macro avg	0.98	0.98	0.98	214
Weighted avg	0.98	0.98	0.98	214



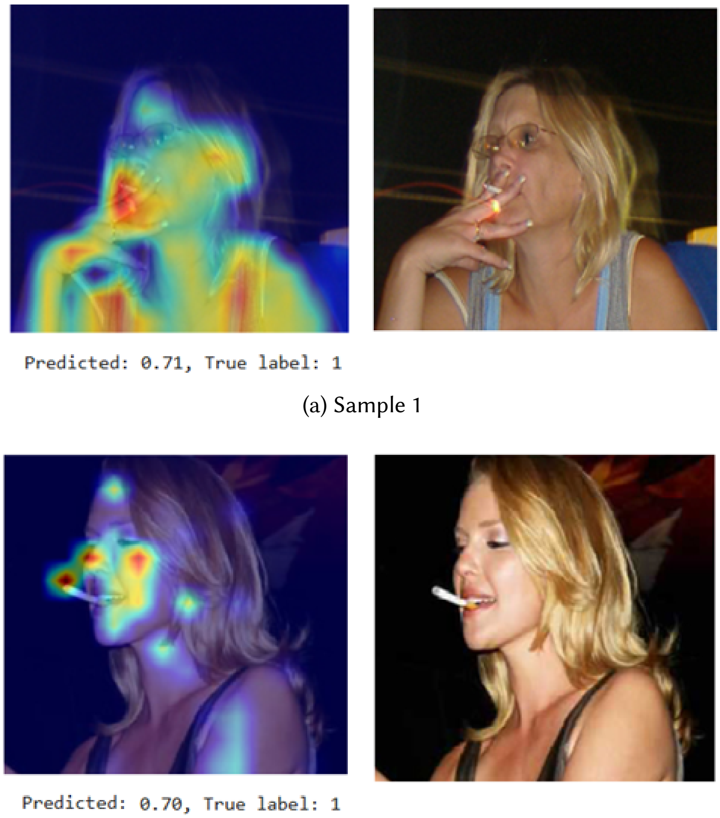
(a) Proposed Method

(b) Standard Architecture

Figure 5: Loss function graph: (a) proposed method, (b) standard architecture.

The loss function plot shows a steady decline with each epoch, while the neural network without considered adaptations in the form of training sample transformations and weight freezing demonstrated jumpy amplitudes and higher final values even with increasing number of epochs (up to 20). Analysis of the confusion matrix and ROC curve confirms the ability of the model to accurately distinguish classes, while maintaining stable sensitivity and specificity. This gives grounds to argue that the model not only learns efficiently on a limited amount of adaptation data, but also demonstrates a high level of generalization, which makes it suitable for application in real conditions.

Since the goal was not only to improve the accuracy of smoking detection, but also to check whether the neural network correctly highlights the relevant areas of attention in the images, heat maps were also implemented for the analyzed images. An example of smoking identification and the corresponding heat maps are shown in Figure 6. The results of heat map visualization confirm that the model not only demonstrates high classification accuracy, but also forms semantically based spatial attention. The active areas detected by the neural network during decision-making are mainly concentrated in areas of the image containing signs of smoking - in particular, the mouth, face and cigarette area. This indicates the ability of the model to form internal representations that correspond to meaningful visual patterns associated with the target class. This result indicates that the network is not a "black box" [33], and its behavior can be interpreted through visual attention mechanisms. The presence of a focus on relevant regions of the image increases confidence in the model and confirms that the classification decision is based on a meaningful analysis of the input signal, and not on distorted or random features.


Predicted: 0.71, True label: 1
Predicted: 0.70, True label: 1

(a) Sample 1

(b) Sample 2

Figure 6: Heatmaps for explaining neural network solutions.

The obtained results were also compared with analogues. The comparison results are presented in Table 2. The results obtained indicate a significant increase in the accuracy of recognizing the action of smoking in public places when using the developed approach based on Vision Transformer with the use of image transformations and weight freezing. The proposed model achieved an Accuracy of 0.98, which demonstrates an advantage over InceptionResNetV2, which, according to previous studies, achieved an Accuracy of 0.969. In addition, compared to the YOLOv5-small approach, for which the Precision and Recall values were 0.935 and 0.891, respectively, the developed model provides a significant improvement in all classification metrics, achieving a balanced Precision, Recall and F_1-measure value of 0.98. This indicates a better ability of the model not only to accurately classify objects, but also to reduce the number of both false positive and false negative activations. Thus, the proposed solution outperforms both traditional CNN architectures and modern object detection approaches, demonstrating the feasibility of using a transformative architecture for visual monitoring of violations in public spaces.

Table 2
Comparison results with analogues

Approach	Accuracy	Precision	Recall	F_1-score
InceptionResNetV2 [17]	0.969	–	–	–
YOLOv5-small [16]	–	0.935	0.891	–
Developed approach (ViT + transformations)	0.98	0.98	0.98	0.98

Future research will be devoted to processing streaming video, in particular in real time, which will allow expanding the application of the developed method in the context of video surveillance systems and automated behavior monitoring. Special attention is planned to be paid to adapting the model to conditions of low image quality, variability of angles and scene dynamics, which will allow increasing its stability and generalizability in real operating conditions.

6. Conclusion

The paper proposes the transfer machine learning method for smoking detection in public spaces with transparent AI decisions, which allows achieving the set goal of increasing the accuracy of detecting smoking in images. In accordance with the goal, the accuracy increased compared to known analogues by 0.011 in the Accuracy metric compared to the approach using InceptionResNetV2, as well as by 0.045 in the Precision metric and by 0.089 in the Recall metric compared to the approach based on YOLOv5-small. In addition to the main goal of achieving accuracy, visual interpretations of neural network decisions in the form of heat maps were used. Visual interpretations confirm the correctness of the detected attention zones in the images, which encourages trust in the decisions made by artificial intelligence, and also increases transparency, which is especially relevant in the field of automatic control of compliance with social norms.

The proposed method differs from existing ones by the use of partial parameter fixation (freezing all layers of the model, except for the final classifier, which allows to adapt the model to a problem with a limited amount of data and reduce the risk of overtraining and modification of training data), as well as the use of explainability of neural network solutions in the form of a heat map.

The implemented strategy of partial parameter freezing, in which only the weights of the final classifier are optimized, allowed to significantly reduce computational costs and avoid overtraining, while maintaining the model's ability to generalize. The use of stochastic averaging of weights contributed to increasing the stability of training and improving the results on validation samples. Comprehensive image preprocessing, including normalization, scaling and augmentations, ensured the model invariance to variability of shooting conditions.

In further research, it is planned to focus on real-time streaming video processing to expand the possibilities of applying the method in video surveillance systems. Particular attention will be paid to increasing the model's resilience to low image quality, changes in perspective, and scene dynamics.

Acknowledgments

The authors are grateful to Prof. Tetiana Hovorushchenko, Prof. Olexander Barmak, Prof. Iurii Krak and other Program Committee members for organizing and conducting the workshop ExplAI-2025: Advanced AI in Explainability and Ethics for the Sustainable Development Goals.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] S. Gupta, V. Kumar, P. Gupta, A comprehensive study on the harmful effects of the smoking on human beings, in: Challenges in Information, Communication and Computing Technology, CRC Press, London, 2024, pp. 577–582. doi:10.1201/9781003559085-99.

[2] T. G. Pinto, L. d. S. Avanci, A. C. M. Renno, D. C. Hipolide, J. N. d. Santos, P. R. Cury, R. A. Dedivitis, D. A. Ribeiro, The impact of genetic polymorphisms on genotoxicity (dna damage) induced by cigarette smoke in humans: A systematic review, Journal of Applied Toxicology (2025). doi:10.1002/jat.4753.

[3] J. Hart, S. Lignou, Generational smoking bans: inegalitarian without disadvantage?, Journal of Medical Ethics (2025). doi:10.1136/jme-2024-110632.

[4] A. Boretti, Rethinking selective prohibitions: the inconsistency of a generational smoking ban in a permissive society, Journal of Medical Ethics (2025). doi:10.1136/jme-2024-110577.

[5] S. Howard, G. Krishna, Do smoking bans work?, BMJ (2025). doi:10.1136/bmj.q2759.

[6] B. Saunders, A generational ban creates inequality between non-smokers, Journal of Medical Ethics (2025). doi:10.1136/jme-2024-110602.

[7] Z. Wang, L. Lei, P. Shi, Smoking behavior detection algorithm based on yolov8-mnc, Frontiers in Computational Neuroscience 17 (2023). doi:10.3389/fncom.2023.1243779.

[8] K. Bobrovnikova, S. Lysenko, B. Savenko, P. Gaj, O. Savenko, Technique for IoT malware detection based on control flow graph analysis, Radioelectronic and Computer Systems 2022 (2022) 141–153. doi:10.32620/reks.2022.1.11.

[9] M. Li, F. Liu, A novel finetuned yolov8 transfer learning model for smoking behavior detection, in: 2024 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML), IEEE, 2024, pp. 1944–1949. doi:10.1109/icicml63543.2024.10957847.

[10] Y. Fu, T. Ran, W. Xiao, L. Yuan, J. Zhao, L. He, J. Mei, Gd-yolo: An improved convolutional neural network architecture for real-time detection of smoking and phone use behaviors, Digital Signal Processing (2024). doi:10.1016/j.dsp.2024.104554.

[11] S. Dalal, U. K. Lilhore, M. Radulescu, S. Simaiya, V. Jaglan, A. Sharma, A hybrid lbp-cnn with yolov5-based fire and smoke detection model in various environmental conditions for environmental sustainability in smart city, Environmental Science and Pollution Research (2024). doi:10.1007/s11356-024-32023-8.

[12] E. Sezgin, A. B. Kocaballi, The era of generalist conversational ai to support public health communications (preprint), Journal of Medical Internet Research (2024). doi:10.2196/69007.

[13] T. Hovorushchenko, A. Moskalenko, V. Osyadlyi, Methods of medical data management based on blockchain technologies, Journal of Reliable Intelligent Environments 9 (2023) 5–16. doi:10.1007/s40860-022-00178-1.

[14] R. Gupta, G. Bhatt, S. Goel, R. Singh, Prioritizing tobacco control & its cessation under sustainable development goals with a focus on india, Indian Journal of Medical Research 157 (2023) 381. doi:10.4103/ijmr.ijmr_3030_21.

[15] I. Krak, O. Barmak, E. Manziuk, Using visual analytics to develop human and machine-centric models: a review of approaches and proposed information technology, Computational Intelligence 38 (2022) 921–946. doi:10.1111/coin.12289.

[16] Y. Li, H. Zhou, J. Feng, X. Li, X. Xu, P. Hou, X. Hu, An improved smoking behavior detection algorithm via incorporating an interference information filtering network, Engineering Applications of Artificial Intelligence 136 (2024). doi:10.1016/j.engappai.2024.109050.

[17] A. Khan, S. Khan, B. Hassan, Z. Zheng, Cnn-based smoker classification and detection in smart city application, Sensors 22 (2022) 892. doi:10.3390/s22030892.

[18] R. Lakatos, P. Pollner, A. Hajdu, T. Joó, A multimodal deep learning architecture for smoking detection with a small data approach, Frontiers in Artificial Intelligence 7 (2024). doi:10.3389/frai.2024.1326050.

[19] A. Khan, M. A. M. Elhassan, S. Khan, H. Deng, Deep learning-based smoker classification and detection: An overview and evaluation, Expert Systems with Applications 267 (2025). doi:10.1016/j.eswa.2024.126208.

[20] D. Murthy, R. R. Ouellette, T. Anand, S. Radhakrishnan, N. C. Mohan, J. Lee, G. Kong, Using computer vision to detect e-cigarette content in tiktok videos, Nicotine & Tobacco Research 26 (2024) S36–S42. doi:10.1093/ntr/ntad184.

[21] A. A. N. M. Lavu, H. Zhang, M. A. I. Jonayed, M. T. Hossain, Indoor smoking detection method based on dual spectral fusion image and yolo framework, LC International Journal of STEM 5 (2024) 13–35. doi:10.5281/zenodo.14028770.

[22] J. Maurício, I. Domingues, J. Bernardino, Comparing vision transformers and convolutional neural networks for image classification: A literature review, Applied Sciences 13 (2023) 5521. doi:10.3390/app13095521.

[23] C. Meng, W. Lin, B. Liu, H. Zhang, Z. Gan, C. Ouyang, Rts-vit: Real-time share vision transformer for image classification, IEEE Journal of Biomedical and Health Informatics (2025) 1–12. doi:10.1109/jbhi.2024.3525054.

[24] M. Jayamohan, S. Yuvaraj, A novel human action recognition using grad-cam visualization with gated recurrent units, Neural Computing and Applications (2025). doi:10.1007/s00521-025-10978-0.

[25] O. Kovalchuk, V. Slobodzian, O. Sobko, M. Molchanova, O. Mazurets, O. Barmak, I. Krak, N. Savina, Visual analytics-based method for sentiment analysis of covid-19 ukrainian tweets, in: Lecture Notes on Data Engineering and Communications Technologies, volume 149, 2023, pp. 591–607. doi:10.1007/978-3-031-16203-9_33.

[26] J. Narkhede, Comparative evaluation of post-hoc explainability methods in ai: Lime, shap, and grad-cam, in: 2024 4th International Conference on Sustainable Expert Systems (ICSES), IEEE, 2024, pp. 826–830. doi:10.1109/icses63445.2024.10762963.

[27] E. Manziuk, I. Krak, O. Barmak, O. Mazurets, V. Kuznetsov, O. Pylypiak, Structural alignment method of conceptual categories of ontology and formalized domain, in: CEUR Workshop Proceedings, volume 3003, 2021, pp. 11–22. URL: http://ceur-ws.org/Vol-3003/paper2.pdf.

[28] S. Kapadnis, Smoking dataset, 2023. URL: https://www.kaggle.com/datasets/sujaykapadnis/smoking, accessed: 2025-11-27.

[29] I. Krak, O. Sobko, M. Molchanova, I. Tymofiiev, O. Mazurets, O. Barmak, Method for neural network cyberbullying detection in text content with visual analytic, in: CEUR Workshop Proceedings, volume 3917, 2025, pp. 298–309. URL: https://ceur-ws.org/Vol-3917/paper57.pdf.

[30] M. Molchanova, V. Didur, O. Mazurets, O. Sobko, O. Zakharkevich, Method for construction and demolition waste classification using two-factor neural network image analysis, in: CEUR Workshop Proceedings, volume 3970, 2025, pp. 168–182. URL: https://ceur-ws.org/Vol-3970/PAPER14.pdf.

[31] J. D. Hunter, Matplotlib: A 2d graphics environment, Computing in Science & Engineering 9 (2007) 90–95. doi:10.1109/MCSE.2007.55.

[32] G. Bradski, The OpenCV Library, Dr. Dobb's Journal of Software Tools (2000).

[33] I. Krak, O. Zalutska, M. Molchanova, O. Mazurets, E. Manziuk, O. Barmak, Method for neural network detecting propaganda techniques by markers with visual analytic, in: CEUR Workshop Proceedings, volume 3790, 2024, pp. 158–170. URL: https://ceur-ws.org/Vol-3790/paper14.pdf.