

ViT-based Classification of Mammogram Images: Impact of Data Augmentation Techniques

Sergio Cannata¹, Giovanni Cicceri^{2*}, Giansalvo Cirrincione³,
Tiziana Currieri², Marta Lovino⁴, Carmelo Militello⁵,
Eros Pasero¹, Francesco Prinzi^{2,6}, Salvatore Vitabile²

¹Department of Electronics and Telecommunications, Politecnico di Torino, Turin,
10129, Italy.

²Department of Biomedicine, Neuroscience and Advanced Diagnostics (BiND),
University of Palermo, Palermo, 90127, Italy.

³Département Electronique - Electrotechnique - Automatique (EEA), University of
Picardie Jules Verne, Amiens, 80000, France.

⁴Department of Engineering Enzo Ferrari, University of Modena and Reggio Emilia,
Modena, 41125, Italy.

⁵Institute for High-Performance Computing and Networking (ICAR-CNR), National
Research Council, Palermo, 90146, Italy.

⁶Department of Computer Science and Technology, University of Cambridge,
Cambridge, CB2 1TN, UK.

*Corresponding author(s). E-mail(s): giovanni.cicceri@unipa.it;

Abstract

Breast cancer is a main preoccupation for women, and mammographic screening proved impressive benefits in reducing mortality risk. However, false positives and false negatives still occur due to human perception, differences in breast density, and the complexity of cancer itself. Convolutional Neural Networks (CNNs) have shown promise in medical imaging issues, but they struggle with understanding long-range spatial interactions in various image patches. Vision Transformers (ViTs) have emerged as a solution to this problem. This work used a subset of the Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM) to implement a ViT-based framework for breast cancer classification. Geometric and Diffuser-based data augmentation (DA) methods were applied and compared to evaluate the resulting performance improvement. The obtained results show how diffuser-based DA improves the performance of geometric DA. However, their combination allows for higher performance (accuracy=77.01%, sensitivity=88.89%, specificity=68.63%) and demonstrates the

feasibility and effectiveness of this approach in enhancing the model’s capabilities for breast cancer classification.

Keywords: breast cancer, mammographic image classification, Visual Transformer, data augmentation impact, Diffuser, CBIS-DDSM dataset

1 Introduction

Breast cancer is the main cause of mortality for the female population, and mammography screening shows impressive benefits in lowering the chance of death by 58.2% in 2021 [1]. However, mammography evaluation is a difficult task for radiologists. Significant false negative and positive rates are caused by various reasons, including breast density, human perception, and the particular characteristics of cancer [2, 3]. For this reason, several deep learning (DL) architectures have been developed to support the diagnostic process in mammography [4]. Convolutional Neural Networks (CNNs) are one of several DL techniques that have demonstrated promising results in addressing challenges related to medical imaging [5]. However, long-range spatial correlations within images are out of the scope of CNN-based architectures. A potential solution to this issue has emerged with the Vision Transformers (ViTs) [6, 7], which apply the Transformers’ capabilities [8] to the field of computer vision [9]. ViTs use self-attention mechanisms, enabling them to grasp associations and dependencies among image patches, leading to a deeper analysis, understanding of the image. The model exploits the Encoder Transformer block and leverages self-attention to simultaneously capture local features and gather information about the global image “context”.

In this study, a ViT-based model was implemented for breast cancer classification. To address the extensive data requirements for training, and the common issue related to small datasets in medical scenarios, geometric and Diffuser-based data augmentation (DA) methods were employed. The ViT-based architecture was trained using a subset of the Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM), to evaluate the impact of DA methods on classifier outcomes.

Our main contributions are as follows:

- flexible DA-based model for mammography classification, easily extendable to other highly heterogeneous data contexts;
- computation of attention maps obtained from the ViT model to provide an interpretation of the model findings.

2 Related Works

In [10] a deep multiple instance network was proposed for mass classification without any previous ROIs selection. In particular, the study investigates three alternative approaches. The proposed architecture’s robustness is demonstrated by experimental findings on the INbreast dataset. This approach obtains an accuracy of 74.2%. Shu *et al.* [11] proposes several pooling structures for CNNs, capable of dividing the

image into areas and selecting those with a high likelihood of malignancy as a depiction of the entire mammography image. The architectural solution proposed can be extended to many convolutional-based architectures, improving performance in the classification of mammographic images. Experimental results on INbreast and CBIS-DDSM datasets show how the pooling layers exceed the most advanced solutions. The approach obtains an accuracy 67.5%, 70.3%, and 76.2% when *i*) max pooling, *ii*) average pooling, and *iii*) region-based group-max pooling is used, respectively. In [12] is proposed a novel breast cancer mass classification model utilizing deep convolutional neural networks (DCNNs) and explored transfer learning with AlexNet and GoogleNet. Through comparative analysis of four mammogram databases, including INbreast, CBIS-DDSM, MIAS, and Egyptian National Cancer Institute, the authors identify the best-performing DCNNs, finding the adapted AlexNet model to significantly outshine GoogleNet in accuracy, robustness, and testing time. In [13] is implemented a detailed ‘*divide and conquer*’ methodology by identifying optimal DL approaches that significantly improve accuracy on the CBIS-DDSM dataset. This study provided practical insights for medical practitioners to adopt these advanced methods without extensive trial-and-error, thereby potentially increasing the efficacy of breast cancer screening and diagnosis with a gain of 5.6 % in accuracy over the baseline models.

3 Materials and Methods

3.1 Dataset Description

A subgroup of the Digital Database for Screening Mammography (DDSM), named CBIS-DDSM, was used in this work. This database contains scanned digital mammograms, including normal, malignant, and benign cases with verified pathological results [14]. In this work only malignant and benign breast lesions were considered. To better evaluate the impact of the implemented DA strategies, only a portion of the CBIS-DDSM dataset was considered. Training, validation and test sets were randomly sampled. The training subset included 500 samples, equally distributed between the two classes. The validation set comprised 233 images, with 116 malignant and 117 benign samples. Finally, the test set included 348 mammographies, containing 144 malignant and 204 benign cases.

3.2 Images Pre-processing

This section outlines the pre-processing steps, crucial for analysis and evaluation. The procedure includes manual cropping to remove artifacts, automatic thresholding to emphasize tumor areas, and automatic cropping to guarantee uniformity of image size for the CBIS-DDSM dataset. Figure 1 shows an example of the key steps involved in the pre-processing pipeline. First, the images were carefully manually cropped to remove any extraneous artifacts. We manually eliminated several artifacts to concentrate only on the regions of interest, such as radiopaque markers. To mitigate the impact of non-uniform background regions, the Otsu algorithm was used to mask the entire breast [15]. To reduce discrepancies in image sizes, speed up the training process, and reduce model complexity, the pre-processed images were automatically resized

to 640×640 . Finally, images were standardized to decrease potential variations and distortions among different image samples. Figure 2 shows the preprocessing steps performed on mammographic images.

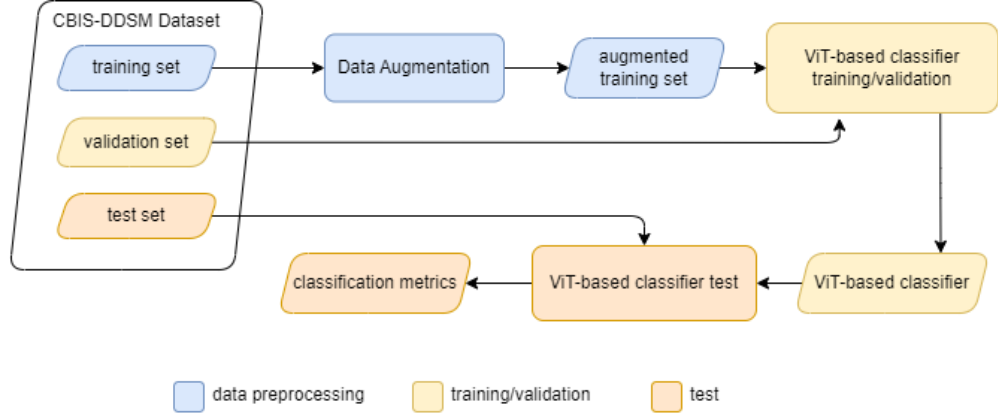


Fig. 1: Flow diagram illustrating the processing workflow.

3.3 Geometric DA

The final training set was enhanced by using geometric, or Classical DA approaches. By increasing the dataset’s diversity and variability, these strategies aimed to strengthen machine learning models’ robustness and ability for generalization. Methods used include: *i*) horizontal, vertical and diagonal flips, *ii*) exposure changes, *iii*) noise addition, *iv*) contrast enhancement, *v*) saturation enhancement, *vi*) CLAHE, and *vii*) Gaussian blurring. Starting with the initial training set of 500 images, applying these 9 transformations the training set resulted in a total of 4500 images.

Vertical and horizontal flips were used to account for variations in breast position and orientation, as well as random rotations $[-30^\circ, +30^\circ]$. To simulate different condition acquisition, exposure changes were also implemented. In particular, contrast and brightness were altered randomly, creating various exposure levels to help the models handle real-world scenarios. Noise addition helped the models differentiate between important structures and noise artifacts, thereby enhancing efficiency in noisy settings. Noise was produced using a normal distribution (mean 0 and standard deviation 1), which is a type of noise frequently seen in medical imaging equipment. Contrast improvement procedures were used to increase the distinction between structures. By tuning the pixel intensity, the increased contrast aided in detecting subtle patterns and abnormalities. Additionally, CLAHE (Contrast Limited Adaptive Histogram Equalization) was used as a local contrast improvement method. CLAHE was able to extract pertinent insights from various areas by improving the visualization of

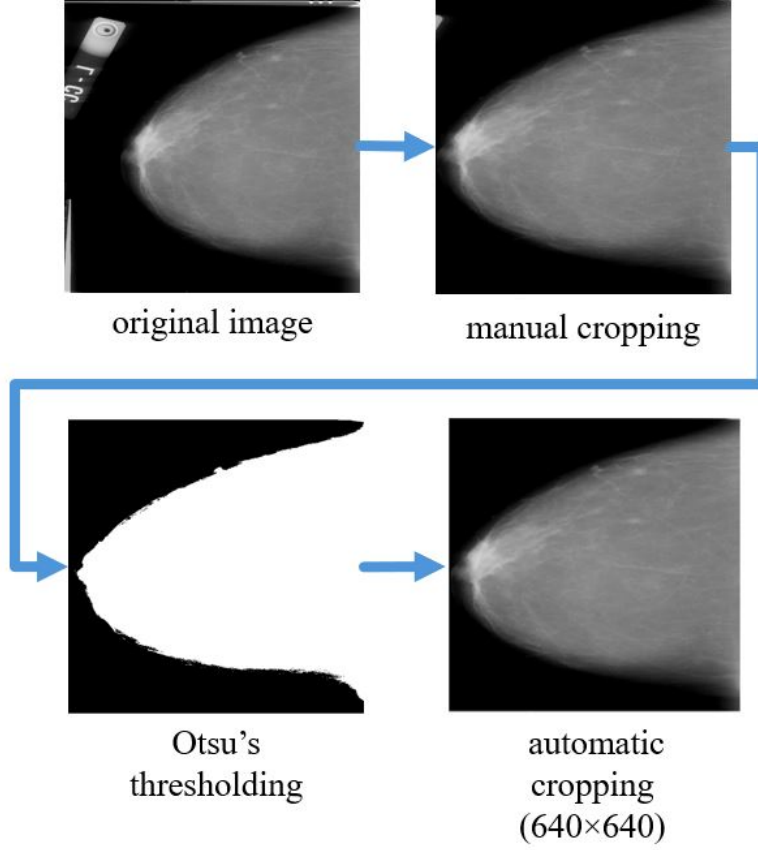


Fig. 2: Preprocessing steps performed on mammographic images.

both global and local structures. In particular, for the brightness changes the following configuration was used: $\alpha = 1.2$ (i.e. contrast increase), and $\beta = 10$ (i.e. brightness increase).

Saturation enhancement and blurring techniques were applied to mammogram images. Saturation enhancement adjusted color saturation to capture variations in color distribution, aiding classification tasks that rely on color differentiation of tissue types. Blurring reduced high-frequency noise and extraneous details, removing noise artifacts and unimportant features. This allowed to emphasize on critical structures and patterns. In particular, a Gaussian kernel with dimension 5×5 and $\sigma = 0$ was used for image filtering.

3.4 Diffuser DA

Diffusion models have seen a rapid rise in popularity for synthetic image generation, surpassing the performance of other techniques such as Generative Adversarial Networks (GAN) [16] and Variational Auto-encoders (VAE) [17]. The Denoising Diffusion

Probabilistic Model (DDPM) [18] was employed to augment the training dataset with artificial malignant and benign lesions images. DDPM training consists of gradually corrupting an image with noise via Markov chains as long as its distribution resembles that of an isotropic Gaussian. Next, a denoising process is applied to restore the image to its native form. The goal is to collect data describing the native distribution of images and use that information to create synthetic samples that follow the original distribution. DDPMs use an improved U-Net architecture [19] with skip connections and self-attention to handle different conditioning information for generating images. With more details, the architecture *'google/ddpm-cat-256'* has been used, available at [20].

Two DDPMs were trained with all the malignant and benign samples, separately, using the training set, each for 200 epochs with a learning rate of $1e^{-4}$ learning rate to produce both malignant and benign images. The two DDPMs models trained on the two classes are used to produce 2000 samples per class, with an overall amount of 4000 synthetic images.

3.5 Experiments

Experimental trials aimed to evaluate the effects of different DA methods considering geometric DA, diffuser-based DA, and their combination. In particular, the following experiments were conducted:

- After being pre-trained on the ImageNet database [21], a Vision Transformer model was trained for 100 epochs, and its results served as a benchmark for the subsequent trials;
- Then geometric and diffuser-based DA were applied in the training set, separately, and the training was performed for 50 epochs;
- Eventually, 50 epochs were employed to train the ViT using the combination of geometric and diffuser-based DA was employed.

The ViT employed in this study was implemented with 12 layers, 12 heads, and 768 as hidden sizes. Images were resized to 224x224 and 16 was used as patch size. Regarding the hyperparameters, the Adam optimizer was employed using $\epsilon = 1.0e^{-8}$ as the learning rate, 4 as gradient accumulation steps and 16 as a batch.

4 Results

Table 1 displays the confusion matrices and shows the enhanced performance achieved by using the implemented DA techniques. In particular, the use of geometric DA increases accuracy compared to the training without DA. Diffuser-based DA increases over geometric DA in terms of accuracy, specificity, and sensitivity. Eventually, the combination improve again over diffuser-based DA. From a clinical point of view, the classification performance highlights a greater capability for malignant lesion prediction (higher sensitivity), which is the most important metric for applicability in real clinical scenarios [22].

The previously mentioned aspects are illustrated in the confusion matrix in Figure 3, where the mixture of geometric and diffuser-based augmented images results in an enhancement of True Positive samples.

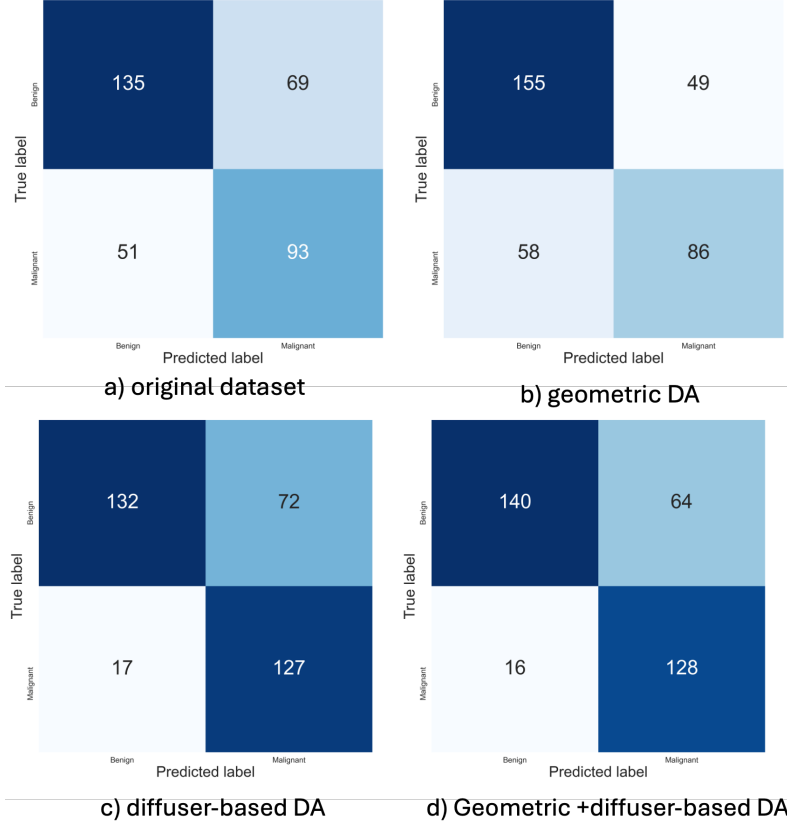


Fig. 3: Confusion matrices illustrating the test results for the proposed data augmentation methods.

Finally, Figure 4 illustrates the ViT attention maps on a malignant lesion image. These maps show the attention scores calculated by the Transformer architecture’s self-attention mechanism, indicating how information from one token influences the representation of another. This visualization offers valuable interpretability for physicians and medical practitioners [4]. The provision of this knowledge, which is seldom found in DL techniques, is a notable advantage of the proposed approach. This is particularly important as clinical diagnosis and decision-making are crucial processes that require reliable and understandable support.

Table 1: Classification results using the implemented data augmentation methods on the test set.

Experiment	Accuracy	Specificity	Sensitivity
no augmentation	65.52%	64.58%	66.18%
geometric DA	69.25%	59.72%	75.98%
diffuser-based DA	74.42%	64.70%	88.19%
geometric + diffuser-based DA	77.01%	68.63%	88.89%

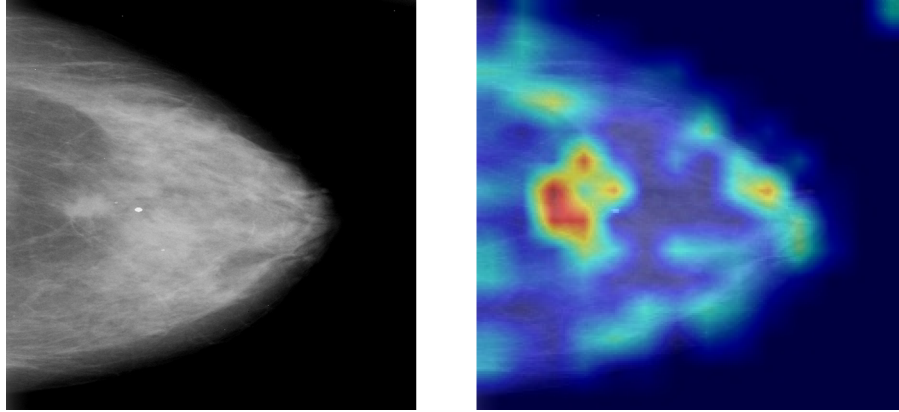


Fig. 4: An image of a malignant lesion (left) with an attention map example (right).

5 Discussion and Conclusions

This paper introduces a ViT-based approach to identify the existence of cancer in breast imaging. The database used is a subset of CBIS-DDMS, presenting low-quality and noisy images. Several DA strategies were implemented to enhance the classification capabilities of a ViT-based predictive model. From the point of view of clinical feasibility, the suggested concept may easily be adapted to other complex data sets. The results show how the combination of geometric and diffuser-based DA can enhance performance and potentially address the training on small dataset scenarios.

Acknowledgement

This research has been supported by Ministero Università e Ricerca n. 1511 del 30/09/2022; and by European Union - Next Generation EU - Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) 2022, Prot. 2022ENK9LS. Project: "EXEGETE: Explainable Generative Deep Learning Methods for Medical Image and Signal Processing" (Code: B53D23013040006).

References

- [1] Attena, F., Abagnale, L., Avitabile, A.: Online information about mammography

- screening in italy from 2014 to 2021. *BMC Women’s Health* **22**(1), 1–6 (2022)
- [2] Ekpo, E.U., Alakhras, M., Brennan, P.: Errors in mammography cannot be solved through technology alone. *Asian Pacific journal of cancer prevention: APJCP* **19**(2), 291 (2018)
 - [3] Prinzi, F., Orlando, A., Gaglio, S., Vitabile, S.: Interpretable radiomic signature for breast microcalcification detection and classification. *Journal of Imaging Informatics in Medicine*, 1–16 (2024)
 - [4] Prinzi, F., Insalaco, M., Orlando, A., Gaglio, S., Vitabile, S.: A yolo-based model for breast cancer detection in mammograms. *Cognitive Computation* **16**(1), 107–120 (2024)
 - [5] Yamashita, R., Nishio, M., Do, R.K.G., Togashi, K.: Convolutional neural networks: an overview and application in radiology. *Insights into imaging* **9**, 611–629 (2018)
 - [6] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale (2021)
 - [7] Cirrincione, G., Cannata, S., Cicceri, G., Prinzi, F., Currier, T., Lovino, M., Militello, C., Pasero, E., Vitabile, S.: Transformer-based approach to melanoma detection. *Sensors* **23**(12), 5677 (2023)
 - [8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. *Advances in neural information processing systems* **30** (2017)
 - [9] Parvaiz, A., Khalid, M.A., Zafar, R., Ameer, H., Ali, M., Fraz, M.M.: Vision transformers in medical computer vision—a contemplative retrospection. *Engineering Applications of Artificial Intelligence* **122**, 106126 (2023)
 - [10] Zhu, W., Lou, Q., Vang, Y.S., Xie, X.: Deep multi-instance networks with sparse label assignment for whole mammogram classification. In: *International Conference on Medical Image Computing and Computer-assisted Intervention*, pp. 603–611 (2017). Springer
 - [11] Shu, X., Zhang, L., Wang, Z., Lv, Q., Yi, Z.: Deep neural networks with region-based pooling structures for mammographic image classification. *IEEE Transactions on Medical Imaging* **39**(6), 2246–2255 (2020)
 - [12] Hassan, S.A., Sayed, M.S., Abdalla, M.I., Rashwan, M.A.: Breast cancer masses classification using deep convolutional neural networks and transfer learning. *Multimedia Tools and Applications* **79**(41), 30735–30768 (2020)

- [13] Jaamour, A., Myles, C., Patel, A., Chen, S.-J., McMillan, L., Harris-Birtill, D.: A divide and conquer approach to maximise deep learning mammography classification accuracies. *Plos one* **18**(5), 0280841 (2023)
- [14] Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L.: A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific data* **4**, 170177 (2017)
- [15] Otsu, N.: A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics* **9**(1), 62–66 (1979)
- [16] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative Adversarial Networks (2014)
- [17] Kingma, D.P., Welling, M.: Auto-Encoding Variational Bayes (2022)
- [18] Ho, J., Jain, A., Abbeel, P.: Denoising Diffusion Probabilistic Models (2020)
- [19] Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation (2015)
- [20] Face, H.: Denoising Diffusion Probabilistic Models (DDPM). <https://huggingface.co/google/ddpm-cat-256> [Accessed: (11-03-2024)] (2022)
- [21] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009). IEEE
- [22] Prinzi, F., Orlando, A., Gaglio, S., Vitabile, S.: Breast cancer classification through multivariate radiomic time series analysis in dce-mri sequences. *Expert Systems with Applications* **249**, 123557 (2024)