

HIGHLY EFFICIENT APPROACHES TO PROCESSING COMPLEX VISUAL DATA IN DECISION SUPPORT SYSTEMS

Oleksandr Poplavskiy¹, Sergei Pavlov², Oksana Bezsmertna², Iryna Gerasymova³, Bakhyt Yeraliyeva⁴

¹Kyiv National University of Construction and Architecture, Kyiv, Ukraine, ²Vinnitsya Technical National University, Vinnitsya, Ukraine, ³Vinnitsya Mykhailo Kotsiubynsky State Pedagogical University, Vinnitsya, Ukraine, ⁴M. Kh. Dulaty Taraz Regional University, Taraz, Kazakhstan

Abstract. Modern decision support systems (DSS) increasingly must analyze large-scale, complex, and heterogeneous data streams in real time. High-performance, AI-driven processing methods – particularly deep neural networks – offer effective solutions. This study examines the integration of contemporary architectures – YOLOv8, ResNet-50, EfficientNet-B3, and the Vision Transformer (ViT) – to enhance DSS capabilities. The models are benchmarked on a representative image-classification task using the COCO dataset for training and evaluation. Empirical results indicate that the transformer-based model (ViT) attains the highest accuracy, whereas the one-stage architecture (YOLOv8) achieves the fastest inference. EfficientNet-B3 and ResNet-50 exhibit intermediate trade-offs between accuracy and speed. Deployment considerations across DSS scenarios are outlined: YOLOv8 is appropriate for real-time, resource-constrained environments; ResNet-50 provides balanced performance; EfficientNet-B3 offers strong accuracy with moderate computational demand; and ViT delivers the best accuracy when ample data and computational resources are available. The findings are discussed in the context of DSS workflows, illustrating how the model outputs can directly inform and improve decision-making processes

Keywords: decision support systems, deep learning, image classification, real-time inference

WYSOCE WYDAJNE METODY PRZETWARZANIA ZŁOŻONYCH DANYCH WIZUALNYCH W SYSTEMACH WSPARCIA DECYZYJNEGO

Streszczenie. Nowoczesne systemy wspomagania decyzji (DSS) muszą coraz częściej analizować w czasie rzeczywistym duże, złożone i niejednorodne strumienie danych. Skutecznym rozwiązaniem są wysokowydajne metody przetwarzania oparte na sztucznej inteligencji, w szczególności głębokie sieci neuronowe. W niniejszym badaniu przeanalizowano integrację współczesnych architektur – YOLOv8, ResNet-50, EfficientNet-B3 i Vision Transformer (ViT) – w celu zwiększenia możliwości systemów DSS. Modele są porównywane w reprezentatywnym zadaniu klasyfikacji obrazów przy użyciu zbioru danych COCO do szkolenia i oceny. Wyniki empiryczne wskazują, że model oparty na transformatorze (ViT) osiąga najwyższą dokładność, podczas gdy architektura jednostopniowa (YOLOv8) zapewnia najszybsze wnioskowanie. EfficientNet-B3 i ResNet-50 wykazują pośredni kompromis między dokładnością a szybkością. Przedstawiono kwestie związane z wdrażaniem w różnych scenariuszach DSS: YOLOv8 jest odpowiedni do zastosowań w czasie rzeczywistym, z ograniczonymi zasobami; EfficientNet-B3 i ResNet-50 wykazują pośredni kompromis między dokładnością a szybkością. Przedstawiono rozważania dotyczące wdrażania w różnych scenariuszach DSS: YOLOv8 jest odpowiedni dla środowisk działających w czasie rzeczywistym, o ograniczonych zasobach; ResNet-50 zapewnia zrównoważoną wydajność; EfficientNet-B3 oferuje wysoką dokładność przy umiarkowanych wymaganiach obliczeniowych; a ViT zapewnia najlepszą dokładność, gdy dostępne są duże zasoby danych i mocy obliczeniowej. Wyniki są omawiane w kontekście przepływów pracy DSS, ilustrując, w jaki sposób wyniki modelu mogą bezpośrednio wpływać na procesy decyzyjne i je usprawniać.

Słowa kluczowe: systemy wspomagania decyzji, głębokie uczenie się, klasyfikacja obrazów, wnioskowanie w czasie rzeczywistym

Introduction

Decision Support Systems (DSS) are interactive software-hardware solutions designed to assist organizations in making informed management decisions by analysing large volumes of data. DSS have become integral to contemporary industrial and governmental operations, supporting data-driven decision processes across diverse sectors, including finance and healthcare. However, the explosive growth of data volume and complexity has stretched the capabilities of traditional DSS. Conventional approaches often struggle with real-time processing and heterogeneous data streams, limiting their effectiveness. In fact, a classic DSS may not be suited to handle streaming data or respond to abnormal events in dynamic environments. These limitations motivate the incorporation of high-performance artificial intelligence methods into DSS.

Deep learning, particularly neural networks, has emerged as a key technology to enhance DSS performance [10]. Modern DSS increasingly integrate AI and machine learning components to tackle diverse big-data challenges across industries [4, 9, 13]. By leveraging neural networks, DSS can automatically learn complex patterns from data and adapt to changing conditions in real time, improving both the accuracy and speed of decision support. For example, AI-powered DSS can rapidly analyse sensor streams [4], predict trends, and provide recommendations by considering a multitude of factors and criteria. The integration of deep learning thus enables more intelligent DSS that facilitate decision-making with greater effectiveness across domains such as agriculture, energy, logistics, and finance.

Despite their promise, deploying neural networks in DSS is not trivial. Key challenges include ensuring fast inference on large or streaming datasets, handling heterogeneous data types

(images, text, etc.), and maintaining system scalability. Recent advances in deep neural network architectures offer potential solutions. Convolutional Neural Networks (CNNs) like ResNet have proven adept at extracting features from images, while newer models like EfficientNet improve efficiency by scaling network depth and width in a balanced way [15].

Meanwhile, Vision Transformers (ViT) apply self-attention mechanisms to image patches, capturing long-range dependencies and achieving state-of-the-art accuracy when trained on sufficiently large datasets [14]. Furthermore, the YOLO (You Only Look Once) series of models exemplifies architectures optimized for real-time object detection and classification. The latest version, YOLOv8 (released 2023), introduces an anchor-free architecture and other innovations that yield high accuracy and real-time capabilities across diverse hardware [14].

This paper presents a comprehensive evaluation of high-performance neural network approaches for decision support systems (DSS), with a particular emphasis on image-based classification as a representative task. The current state of the art in applying neural networks to complex DSS-related data is reviewed, followed by the development of a methodology that leverages four contemporary architectures – YOLOv8, ResNet-50, EfficientNet-B3, and Vision Transformer (ViT) – to enhance classification performance.

Special attention is given to techniques such as parallel processing and data augmentation, which are essential for satisfying real-time operational constraints. The experimental study provides a comparative analysis of the aforementioned models based on classification accuracy, area under the ROC curve (AUC), inference speed (measured in frames per second, FPS), and model size, utilizing the COCO benchmark dataset. To maintain broad relevance across DSS domains, the widely



benchmarked COCO dataset is used instead of a single domain-specific dataset. COCO's diversity (80 object categories in varied contexts) simulates heterogeneous real-world conditions, providing a robust testbed for evaluating model performance in a general DSS setting.

The results are analysed to elucidate the advantages and limitations of each model in the context of practical DSS deployment.

1. Methodological framework for integrating high-performance neural network architectures into decision support systems

Traditional DSS implementations often rely on deterministic models, rule-based systems, or simple statistical analyses to process data and generate recommendations. While these methods work for structured data and static scenarios, they encounter severe limitations with today's Big Data characteristics. DSS must now handle data that is massive, heterogeneous, and fast-changing. For instance, a modern DSS might need to fuse structured databases with sensor streams, images, and text in real time. Classical data warehouses and OLAP tools cannot easily accommodate unstructured or real-time data. Moreover, conventional DSS lack learning capabilities; they cannot automatically improve from new data without manual reprogramming. Such systems may be incapable of processing real-time data feeds or reacting to unforeseen events outside their programmed rules [2]. These shortcomings underscore the need for more adaptive and high-performance techniques in DSS.

The current state of neural network research offers a comprehensive set of tools for enhancing data analysis within decision support systems (DSS). This work provides a concise overview of four advanced architectures, selected based on their demonstrated effectiveness in image interpretation tasks – a common requirement in many DSS applications involving visual data processing. While advanced multimodal models (e.g., vision-language OCR systems) exist, our study focuses on core image-classification architectures to keep the scope on visual data processing in DSS.

ResNet-50, introduced in 2015 [1, 3], represented a significant advancement in deep convolutional neural networks through the use of residual connections. This 50-layer architecture effectively mitigates the vanishing gradient problem by incorporating skip connections, enabling stable training of very deep models. The network achieved state-of-the-art performance on the ImageNet benchmark and continues to serve as a widely adopted backbone for computer vision tasks due to its balance of accuracy and computational efficiency. Its ability to extract multilevel features makes it particularly suitable for decision support systems (DSS) that rely on image classification and recognition modules. However, its relatively large model size (approximately 98 MB) may limit its applicability in real-time or edge scenarios without additional optimization.

EfficientNet-B3, introduced in 2019 [11], utilizes a compound scaling strategy to uniformly adjust network width, depth, and resolution with fixed coefficients. This approach allows the model to maintain high accuracy with significantly reduced computational cost. The B3 variant comprises approximately 12 million parameters and occupies around 80 MB of memory, delivering strong performance on the ImageNet dataset with a fraction of the floating-point operations required by older convolutional networks. Its high accuracy-to-computation ratio makes it well-suited for DSS applications constrained by computational resources, such as large-scale or cloud-based environments handling numerous real-time queries.

The *Vision Transformer (ViT)*, proposed in 2020 [2], marks the first successful adaptation of Transformer architectures to image classification. ViT partitions an image into fixed-size patches and processes them via a self-attention mechanism, forgoing traditional convolutional layers. This design enables ViT models to capture long-range dependencies and contextual

information across the image, leading to state-of-the-art classification accuracy when pre-trained on large datasets. In DSS scenarios, ViT is particularly valuable for complex visual tasks involving subtle global context. Nevertheless, ViT models typically demand extensive training data and computing power, with base configurations often reaching 330 MB or more in model size. These factors introduce limitations in applications requiring low-latency or on-device deployment [5, 7].

YOLOv8, released in 2023, is the latest iteration of the "You Only Look Once" family of models known for their real-time object detection capabilities. This version introduces multiple architectural upgrades over previous releases, including an anchor-free detection head, an efficient CSP (Cross Stage Partial) backbone, and improvements for streamlined deployment. Although primarily designed for object detection, YOLOv8 can also be adapted for image classification tasks. The model achieves high accuracy with low inference latency on a wide range of hardware platforms, including edge GPUs and mobile devices. With a compact footprint (approximately 50 MB in the applied configuration), YOLOv8 is especially well-suited for DSS that require rapid decision-making based on live visual input. Its support for both CPU and GPU execution, along with a developer-friendly interface, has led to widespread industry adoption [8].

A structured methodology is introduced to integrate contemporary neural network architectures into the data processing pipeline of decision support systems (DSS) for image classification tasks. The primary objective is to harness the distinct advantages of each model – such as accuracy, computational speed, and resource efficiency – while satisfying operational requirements of DSS, including scalability and real-time responsiveness. The proposed pipeline comprises dataset acquisition and preparation, model training with parallelization and hyper parameter optimization, and systematic performance evaluation using established metrics.

In the present study, the classification task is framed as the assignment of each input image to one of several predefined object categories. This task is representative of many practical applications within DSS environments, such as automated inventory recognition from warehouse images, diagnostic categorization in medical imaging, and object identification in real-time surveillance systems. The complete pipeline encompasses the following stages: data acquisition, pre-processing, neural network inference, and decision output. Although the primary focus is placed on the inference stage, a comprehensive outline of the entire workflow is provided for context.

For model training and evaluation, the Common Objects in Context (COCO) dataset is employed [6, 12]. COCO is a widely recognized benchmark dataset that comprises over 330,000 images annotated with 80 object categories captured in natural scenes. It is specifically selected for its diversity and complexity, offering multiple objects per image and varied environmental conditions that emulate real-world DSS challenges. Although the dataset provides both class labels and bounding box annotations, only class labels are used for classification tasks. A domain-specific dataset (e.g., only medical or only automotive images) was not chosen because it would narrow the applicability of results. Instead, COCO's varied images (people, animals, vehicles, etc.) ensure our evaluation covers a wide range of scenarios, mirroring the heterogeneous data a DSS might encounter.

Dataset COCO illustrating diverse object categories (people, animals, vehicles, etc.) and complex backgrounds. Such variability makes COCO a challenging benchmark for image classification models and simulates real-world data complexity in DSS.

Effective data preprocessing is a critical prerequisite for training high-performing neural networks. The adopted preprocessing pipeline in this study consists of several essential stages. First, the dataset is partitioned into training (approximately 80%), validation (10%), and testing (10%) subsets. This

stratification ensures a clear separation of data for model training, hyper parameter tuning, and final performance evaluation, thereby preventing data leakage and enabling an unbiased assessment of generalization capability.



Fig. 1. Sample images from the COCO dataset

All images are resized to a standardized input resolution – 224×224 pixels for models such as ResNet-50 and EfficientNet-B3, and 640×640 pixels for YOLOv8, adhering to the respective architectural requirements. To preserve the original aspect ratio, padding is applied as necessary. Additionally, pixel intensities are normalized channel-wise to zero mean and unit variance, which contributes to stable and efficient model convergence.

The YOLOv8 model was employed in its classification configuration. This variant uses a Cross Stage Partial (CSP) Darknet backbone and an anchor-free prediction head, producing class probabilities for entire images. Several YOLO-specific optimization techniques, including mosaic augmentation, were utilized to improve generalization. Given its architectural characteristics, the model was trained with a relatively higher learning rate, reflecting its empirically observed tolerance for more aggressive learning schedules.

ResNet-50 was implemented as a canonical 50-layer residual convolutional neural network. The final fully connected layer was modified to output predictions for 80 categories, in accordance with the COCO dataset's class taxonomy. The network was trained using a step-decay learning rate strategy, beginning with an initial learning rate of approximately 0.001, and optimized using stochastic gradient descent (SGD) with momentum (set to 0.9). Batch normalization parameters were updated during fine-tuning using a reduced learning rate to maintain stability.

The EfficientNet-B3 model was applied using the compound scaling configuration proposed in the original architecture. Given its relatively compact size and reduced parameter count, the model was trained over a larger number of epochs to ensure convergence. An Adam optimizer was selected initially, as it has demonstrated superior convergence properties with EfficientNet. Regularization methods, including weight decay and dropout (aligned with the EfficientNet design), were incorporated to mitigate overfitting risks.

The Vision Transformer (ViT-B16) model was configured to process 16×16 image patches and relies on a transformer-based encoder. A two-phase training strategy was employed: initial pretraining on the ImageNet-21k dataset, followed by fine-tuning on the COCO dataset. The AdamW optimizer was used in conjunction with a reduced learning rate, acknowledging the sensitivity of transformer models to learning rate settings. Training stabilization techniques, such as gradient clipping and warm-up scheduling, were further employed to facilitate effective convergence during fine-tuning [16].

Training for each architecture was conducted until the validation accuracy plateaued. An early stopping mechanism was applied, terminating training when no further improvement in validation performance was observed over a predefined number of epochs. This strategy mitigates the risk of overfitting while conserving computational resources.

During the training process, the choice of optimizer and loss function plays a critical role in guiding the iterative updates of model parameters. For the classification task, categorical cross-entropy is adopted as the loss function L , formally defined as:

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(\hat{y}_{i,c}) \quad (1)$$

where N is the number of training samples, C is the number of classes, $y_{i,c}$ is a binary indicator (0/1) if sample i belongs to class c , and $\hat{y}_{i,c}$ is the model's predicted probability for class c . The networks learn by minimizing this loss via backpropagation.

To optimize model parameters during training, different optimization strategies are applied based on the architecture. Specifically, the Adam optimizer is employed for YOLOv8 and Vision Transformer (ViT) due to its capacity to accelerate convergence by leveraging adaptive moment estimates. For ResNet-50, stochastic gradient descent (SGD) is used, as it is often associated with improved generalization performance in convolutional neural networks.

$$w \leftarrow w - \eta \frac{\partial L}{\partial w} \quad (2)$$

where η denotes the learning rate, and $\frac{\partial L}{\partial w}$ represents the gradient of the loss function with respect to the model weights. In the multi-GPU setup, this update step is executed in parallel, with gradients being synchronized and aggregated across all devices at each iteration to ensure consistent weight updates.

To evaluate model performance in the context of decision support systems (DSS), both predictive accuracy and computational efficiency are considered. Standard classification metrics are employed, including accuracy, precision, recall, F1-score, and the area under the ROC curve (AUC). For multi-class classification tasks, such as the 80-category COCO dataset, these metrics are computed on a per-class basis using a one-vs-all strategy and then macro-averaged to ensure that all classes are weighted equally. These measures provide a comprehensive view of model effectiveness across a variety of DSS scenarios.

In operational decision support systems (DSS), computational efficiency is as critical as predictive accuracy. To this end, several performance dimensions are evaluated. Inference time, measured in milliseconds, quantifies the average latency per image and reflects the responsiveness of a model in real-time applications. Throughput, defined as the number of images processed per second (FPS), is assessed both in single-image and batch-processing modes to capture the model's scalability under varying workloads. Model size, expressed in megabytes, indicates the memory footprint and is a key factor for deployment in resource-constrained environments such as edge devices. Additionally, resource utilization – primarily GPU memory consumption and the feasibility of CPU-based execution – is noted to assess deployment flexibility. For instance, YOLOv8 demonstrates efficient operation even on CPUs, albeit at reduced throughput, while ViT generally requires GPU acceleration due to its computational complexity. These efficiency metrics, considered alongside predictive performance, enable a holistic comparison of model suitability for DSS applications. A model offering marginally lower accuracy but significantly faster inference may be more desirable in scenarios requiring real-time decision-making.

2. Results of experimental measurements

The four selected neural network architectures – YOLOv8, ResNet-50, EfficientNet-B3, and ViT (Base) – were trained and evaluated on the COCO dataset following the methodology previously outlined. Table 1 presents a comparative summary across two key dimensions: classification effectiveness and computational efficiency. Specifically, the reported classification metrics include Accuracy, Precision, Recall, F1-score, and ROC AUC. Efficiency-related metrics comprise frames per second (FPS) on GPU, average inference time

per image (in milliseconds), and the model's memory footprint (in megabytes). This holistic evaluation facilitates an informed analysis of each model's suitability for deployment in real-world DSS environments, balancing predictive performance with resource constraints.

Table 1. Performance comparison of neural network models

Model	Accuracy	Precision	Recall	F1-score	AUC	FPS (↑)	Infer. Time (ms) (↓)	Model Size (MB)
YOLOv8	0.78	0.78	0.80	0.79	0.83	50–60	15–20	50
ResNet-50	0.76	0.75	0.78	0.76	0.80	40	25–30	98
EfficientNet-B3	0.80	0.80	0.82	0.81	0.85	30	35–40	80
ViT-B	0.82	0.82	0.84	0.83	0.88	15–20	50–55	330

The experimental results demonstrate that all four evaluated neural network models – YOLOv8, ResNet-50, EfficientNet-B3, and Vision Transformer (ViT) – achieve competitive performance on the 80-class image classification task using the COCO dataset. ViT attained the highest classification accuracy (~82%) and AUC (~0.88), followed closely by EfficientNet-B3 (~80% accuracy). YOLOv8 and ResNet-50 achieved slightly lower accuracy (~78% and ~76%, respectively), but YOLOv8 stands out for its real-time performance. Inference speed tests confirm that YOLOv8 is the fastest (~60 FPS), significantly outperforming ViT (~15–20 FPS), ResNet (~40 FPS), and EfficientNet (~30 FPS). ViT's superior accuracy is counterbalanced by its large model size (~330 MB) and slow inference, making it better suited for high-resource environments, while YOLOv8's compact size (~50 MB) and high speed make it ideal for latency-sensitive DSS applications. For example, in a live surveillance DSS, YOLOv8's fast inference would allow immediate detection of intruders, enabling the system to promptly alert security personnel. Conversely, in a medical DSS for diagnostics where accuracy is paramount, a ViT model's higher precision could directly influence a clinician's decision by providing more reliable image analysis.

A class-wise error analysis further clarifies the practical implications of these findings. All models exhibit near-perfect recognition of high-contrast, frequently occurring categories such as "person", "car", and "dog". However, rare or fine-grained classes – most notably "toaster", "hair drier", and "keyboard" – remain challenging, with recall dropping below 60% across all networks. ViT consistently outperforms the CNN-based models on these tail classes, boosting average recall by 8 percentage points, whereas YOLOv8 compensates by achieving markedly lower false-positive rates in cluttered scenes. These complementary strengths suggest that an ensemble or cascaded deployment strategy could deliver further gains in real-world DSS scenarios where both high object-recognition fidelity and rapid response times are critical.

Inference-time energy consumption was tracked as the GPU's integrated power draw (RTX A6000, TDP = 300 W) while processing 1 000 images. YOLOv8 was the most economical, using 0.45 Wh thanks to its low FLOP count and very short inference time. EfficientNet-B3 required 0.61 Wh – about 37% less than ResNet-50, which consumed 0.97 Wh. ViT was the most energy-hungry at 2.20 Wh per 1 000 images, nearly five times YOLOv8, due to its large attention workload and parameter count. Converting these figures with the average European grid factor of 253 g CO₂ per kWh, the gap between ViT and YOLOv8 translates into roughly 0.44 kg CO₂ per million images. Therefore, in scenarios where both latency and energy or carbon budgets are critical – such as battery-powered edge DSS or "green-quota" data centres – lighter models may be preferable even at the cost of modest accuracy reductions.

Model size and resource efficiency further differentiate their applicability. EfficientNet-B3 offers a strong balance between performance and model size (~80 MB), whereas ResNet-50, though slightly less accurate, remains a reliable and efficient baseline (~98 MB). YOLOv8 is optimal for real-time, resource-

constrained deployments due to its small size and high throughput. Despite ViT's accuracy advantage, its large footprint and slower processing limit its practicality in real-time scenarios. These trade-offs highlight the need to align model choice with the specific operational demands of a DSS, balancing predictive performance with latency, memory constraints, and scalability.

The comparative evaluation reveals that each neural network model offers unique advantages depending on the requirements of a decision support system (DSS). YOLOv8 is optimal for real-time applications due to its high speed, compact size, and suitability for edge deployment, despite slightly lower accuracy. ResNet-50 provides balanced performance with strong reliability and ease of integration, making it suitable for general-purpose DSS. EfficientNet-B3 delivers high accuracy with moderate resource usage, ideal for accuracy-critical yet resource-constrained environments. Vision Transformer (ViT) achieves the highest accuracy but demands significant computational resources, making it best for scenarios where precision outweighs latency and infrastructure constraints. Ultimately, model selection should align with the specific goals and operational context of the DSS, balancing accuracy, speed, and scalability.

The experiments were limited to widely used vision models (CNNs and ViT) to ensure a focused and consistent comparison. Emerging multimodal OCR models such as Mistral OCR and LightOnOCR demonstrate strong performance in document text extraction; however, they are primarily designed for tasks requiring rich text and layout understanding, which are beyond the scope of general image classification. Including such models would necessitate text-centric inputs and alternative evaluation protocols, shifting the study away from benchmarking image-focused methods applicable to typical DSS pipelines.

3. Conclusions

This study explored the integration of high-performance neural network models into the data processing pipelines of Decision Support Systems (DSS), with a focus on complex image classification tasks. The results confirm that incorporating deep learning into DSS significantly enhances their analytical capabilities by enabling rapid, adaptive, and accurate processing of large-scale heterogeneous data. Four representative architectures were evaluated – ResNet-50, EfficientNet-B3, Vision Transformer (ViT), and YOLOv8 – demonstrating that modern models offer substantial improvements in both predictive accuracy and computational efficiency over traditional techniques.

Each architecture exhibits distinct strengths that render it suitable for particular deployment scenarios. YOLOv8 achieved real-time performance (50–60 frames per second) with approximately 78% classification accuracy, making it particularly advantageous for latency-sensitive environments. ResNet-50 provided a balance between accuracy (~76%) and inference speed, reaffirming its status as a robust and widely adopted baseline in many vision tasks. EfficientNet-B3 attained higher accuracy (~80%) while maintaining a relatively compact model size, indicating its suitability for DSS operating under constrained computational resources. ViT yielded the highest classification performance (~82% accuracy and 0.88 AUC), confirming the potential of transformer-based models for capturing complex data patterns, although its significant resource demands may limit deployment to high-performance settings.

The findings underscore that selecting the optimal model for a DSS depends on a careful consideration of trade-offs between speed, memory consumption, and classification accuracy.

Avenues for future research emerge as particularly promising and warrant systematic investigation. The growing need for deploying DSS on edge devices such as mobile platforms or IoT sensors necessitates further exploration of optimization techniques like quantization, pruning, and knowledge distillation. These could render large models such as ViT and EfficientNet more suitable for low-power environments. The compact architecture of YOLOv8, combined with support for ONNX

and TensorRT, already presents promising prospects for edge deployment. The emergence of hybrid CNN-Transformer models opens new possibilities for combining the local feature extraction strengths of CNNs with the global context modelling capabilities of Transformers, potentially achieving enhanced performance at reduced computational cost. Multi-modal neural networks that integrate various data types – such as images, structured signals, and natural language – could further extend the functionality of DSS by providing more context-aware decision-making. As data distributions evolve, the inclusion of continual learning mechanisms and federated learning strategies would allow DSS to adapt without centralized data collection, thereby supporting real-time system updates and privacy-sensitive applications.

In conclusion, high-performance deep learning models present a transformative opportunity for Decision Support Systems. By aligning architectural choices with practical constraints and operational goals, it is possible to develop DSS that are not only highly accurate and data-savvy but also responsive and deployable in real-world settings. Continued progress in model efficiency, training methodology, and system-level integration will drive the evolution of next-generation DSS. These systems, empowered by artificial intelligence, are poised to become proactive and trustworthy partners in complex decision-making processes across a wide array of disciplines.

Acknowledgements

This project has received funding through the EURIZON project, which is funded by the European Union under grant agreement No. 871072.

Disclosures

The authors declare no conflicts of interest.

References

- [1] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houshy, N. (2020). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale* (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.2010.11929>
- [2] Geeksfor Geeks. (2024, August 30). *Decision Support System*. <https://www.geeksforgeeks.org/business-studies/decision-support-system/>
- [3] He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.1512.03385>
- [4] Kostopoulos, G., Davrazos, G., & Kotsiantis, S. (2024). Explainable Artificial Intelligence-Based Decision Support Systems: A Recent Review. *Electronics*, 13(14), 2842. <https://doi.org/10.3390/electronics13142842>
- [5] Lee, S. I., Koo, K., Lee, J. H., Lee, G., Jeong, S., O, S., & Kim, H. (2024). Vision transformer models for mobile/edge devices: A survey. *Multimedia Systems*, 30(2), 109. <https://doi.org/10.1007/s00530-024-01312-0>
- [6] Lin, T.-Y. (2024). *MS COCO dataset* (p. 1 kb) [JSON]. TIB. <https://doi.org/10.57702/6K0015QM>
- [7] Liu, X., Song, Y., Li, X., Sun, Y., Lan, H., Liu, Z., Jiang, L., & Li, J. (2024). *Efficient Partitioning Vision Transformer on Edge Devices for Distributed Inference* (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.2410.11650>
- [8] Medjaldi, A., Slimani, Y., & Karkar, N. (2025). Cost-Effective Real-Time Obstacle Detection and Avoidance for AGVs using YOLOv8 and RGB-D Sensors. *Engineering, Technology & Applied Science Research*, 15(2), 21738–21745. <https://doi.org/10.48084/etasr.10135>
- [9] Poplavskiy, O., Pavlov, S. V., Denysiuk, V., Belinska, S., Shvarts, I. V., Akimova, O., Kornilenko, O., Tarczynska, M., Gaweda, K., Kozbakova, A., & Smolarz, A. (2024). High-performance information technology for processing large datasets and biomedical images to improve the accuracy of computer-aided decision support systems. In R. S. Romaniuk, A. Smolarz, & W. Wójcik (Eds), *Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments 2024* (p. 25). SPIE. <https://doi.org/10.1117/12.3057444>
- [10] Taherdoost, H. (2023). Deep Learning and Neural Networks: Decision-Making Implications. *Symmetry*, 15(9), 1723. <https://doi.org/10.3390/sym15091723>
- [11] Tan, M., & Le, Q. V. (2020). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks* (arXiv:1905.11946). arXiv. <https://doi.org/10.48550/arXiv.1905.11946>
- [12] Tong, K., & Wu, Y. (2023). Rethinking PASCAL-VOC and MS-COCO dataset for small object detection. *Journal of Visual Communication and Image Representation*, 93, 103830. <https://doi.org/10.1016/j.jvcir.2023.103830>
- [13] Vassilenko, V., Poplavska, A. A., Pavlov, S. V., Kolisnyk, P. F., Poplavskiy, O., Kolisnyk, S. P., Vitrova, Y. O., & Wójcik, W. (2020). Automated features analysis of patients with spinal diseases using medical thermal images. In R. S. Romaniuk & J. Dorosz (Eds), *Optical Fibers and Their Applications 2020* (p. 20). SPIE. <https://doi.org/10.1117/12.2569780>
- [14] Wang, Y., Deng, Y., Zheng, Y., Chattopadhyay, P., & Wang, L. (2025). Vision Transformers for Image Classification: A Comparative Survey. *Technologies*, 13(1), 32. <https://doi.org/10.3390/technologies13010032>
- [15] Yaseen, M. (2024). *What is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2408.15857>
- [16] Zhai, S., Likhomanenko, T., Littwin, E., Busbridge, D., Ramapuram, J., Zhang, Y., Gu, J., Susskind, J. M., et al. (2023). Stabilizing Transformer Training by Preventing Attention Entropy Collapse. *Proceedings of the 40th International Conference on Machine Learning*, 202, 40770-40803. <https://proceedings.mlr.press/v202/zhai23a.html>

Prof. Oleksandr A. Poplavskiy
email: apoplavskiy@gmail.com

Professor of IT Department of Kyiv National University of Construction and Architecture.
Scientific directions: information technology, artificial intelligence.



<https://orcid.org/0000-0003-0465-6843>

Prof. Sergii Pavlov
e-mail: psv@vntu.edu.ua

Academician of International Applied Radioelectronic Science Academy, professor of Biomedical Engineering and Optic-Electronic Systems Department, Vinnytsia National Technical University.
Scientific directions: information technology, artificial intelligence.



<https://orcid.org/0000-0002-0051-5560>

Ph.D. Oksana Bezmernta
e-mail: bezmertnaoksana@gmail.com

Ph.D., associate professor, Department of Entrepreneurship, Logistics and Management, Faculty of Management and Information Security, Vinnytsia National Technical University, Vinnytsia, Ukraine.
Scientific directions: information technology, innovation technologies.



<https://orcid.org/0000-0003-1873-6553>

Prof. Iryna Gerasymova
e-mail: gerasimovair@gmail.com

D.Sc., prof. of Vinnytsia Mykhailo Kotsiubynskiy State Pedagogical University.
Scientific directions: information technology, pedagogic, psychology.



<https://orcid.org/0000-0002-9122-3116>
Ph.D. Bakhyt Yerallyyeva
e-mail: yerallyyevabakhyt81@gmail.com

Senior lecturer of the Information Systems Department, Faculty of Information Technology, M. Kh. Dulaty Taraz Regional University, Taraz, Kazakhstan.
Research interests: fiber optic technologies, information systems, internet of things and blockchain technologies.



<https://orcid.org/0000-0002-8680-7694>