DeepLens at ImageCLEFmedical 2026: A Continuation of CNN-Based Medical Concept Detection

Notebook for the ImageCLEFmedical Caption Lab at CLEF 2026

Bahareh Kavousi Nejad[1,*,†], Amir Hossein Salimi Rudsari[1,*,†] and Sauleh Eetemadi[1]

[1]Iran University of Science and Technology, Tehran, Iran

Abstract

The ImageCLEFmedical 2026 Concept Detection task focuses on the automatic identification of medically relevant concepts from radiological images to support semantic retrieval, annotation, and clinical decision-making. In this work, we present the participation of the DeepLens team from Iran University of Science and Technology in the standard Concept Detection subtask. Building upon our participation in ImageCLEFmedical 2025, in which the DeepLens pipeline achieved second place, we continued and adapted our CNN-based concept detection pipeline on the updated 2026 benchmark, focusing on baseline validation and zero-shot re-evaluation of our 2025 architectures on the new dataset. The EfficientNet-B0 and EfficientNet-B1 models were evaluated directly from our previous pipeline to establish robust baseline generalizability, while a DenseNet model underwent an intentional, constrained fine-tuning phase of approximately eight epochs to assess its rapid adaptability to the updated 2026 data distribution. Our approach evaluates several pretrained CNN architectures — DenseNet, EfficientNet-B0, and EfficientNet-B1 — alongside an ensemble strategy combining their predictions. Experimental evaluation on the official benchmark demonstrates that EfficientNet-B1 achieved the strongest performance among our submissions, obtaining a primary F1 score of 0.5648 and a secondary F1 score of 0.9336, while ensemble methods remained competitive. The results indicate that CNN-based transfer learning approaches continue to provide robust baselines for medical concept detection and generalize effectively across successive editions of the ImageCLEFmedical challenge.

Keywords

Medical Concept Detection, Medical Image Analysis, Convolutional Neural Networks, DenseNet, EfficientNet, Transfer Learning

1. Introduction

Medical imaging plays a fundamental role in modern healthcare by supporting diagnosis, treatment planning, and clinical decision-making across a wide range of diseases. The increasing availability of large-scale medical imaging data has created a growing demand for automated systems capable of extracting meaningful semantic information from medical images. In this context, medical concept detection aims to automatically identify clinically relevant concepts associated with an image, facilitating semantic retrieval, annotation, and downstream applications in computer-assisted diagnosis and medical education [1, 2, 3].

To advance research in medical image understanding, the ImageCLEF initiative has introduced a series of benchmark tasks focused on multimodal medical retrieval and analysis [4]. Among these, the ImageCLEFmedical 2026 challenge emphasizes automatic medical concept detection and caption generation using radiological images and synthetic data extensions [5]. The concept detection subtask addresses a challenging multi-label classification problem in which systems are required to predict multiple medically relevant concepts for each image. Such tasks are particularly difficult due to the

CLEF 2026 Working Notes, September 21–24, 2026, Jena, Germany

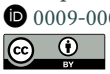
*Corresponding author.

†These authors contributed equally.

bahareh_kavousi@comp.iust.ac.ir (B. Kavousi Nejad); salimi_amirhosein@comp.iust.ac.ir (A. Salimi Rudsari); sauleh@iust.ac.ir (S. Eetemadi)

https://www.sauleh.ir/ (S. Eetemadi)

0009-0004-7958-0227 (B. Kavousi Nejad); 0009-0009-8977-2800 (A. Salimi Rudsari); 0000-0003-1376-2023 (S. Eetemadi)



high semantic variability of medical images, severe class imbalance, and the need to recognize subtle anatomical and pathological patterns.

Deep learning approaches, particularly convolutional neural networks (CNNs), have shown strong performance in medical image analysis by learning hierarchical visual representations directly from image data [2, 3]. Transfer learning from large-scale pretrained models has further improved performance in medical imaging tasks where annotated data may be limited [6, 7]. Architectures such as DenseNet and EfficientNet have demonstrated promising results in medical image classification due to their efficient feature extraction capabilities and scalability [8, 9].

In this work, we present the participation of the DeepLens team from Iran University of Science and Technology in the ImageCLEFmedical 2026 Concept Detection task. Building upon our participation in ImageCLEFmedical 2025, in which our submission achieved second place in the competition, we evaluated and adapted our previously developed CNN-based pipeline on the 2026 benchmark [10]. To assess the structural stability and multi-year cross-compatibility of deep feature extractors, this work prioritizes a direct continuation and evaluation of the 2025 pipeline components. We evaluated the DeepLens 2025 EfficientNet-B0 and EfficientNet-B1 models without secondary training on the 2026 benchmark, while executing a targeted, limited fine-tuning protocol on DenseNet to study architecture flexibility under strict epoch caps. An ensemble combination of the three architectures was also evaluated. Our findings indicate that CNN-based transfer learning approaches remain robust and competitive for medical concept detection across successive ImageCLEFmedical editions, even under tight computational constraints.

2. Related Work

2.1. Deep Learning in Medical Image Analysis

Deep learning has significantly advanced the field of medical image analysis by enabling automated feature extraction and robust visual representation learning. Convolutional neural networks (CNNs), in particular, have demonstrated strong performance across various medical imaging applications, including disease classification, segmentation, abnormality detection, and report generation [2]. Compared to traditional handcrafted feature extraction methods, deep learning models can directly learn hierarchical representations from raw image data, improving generalization across different imaging modalities and clinical settings.

Several studies have highlighted the effectiveness of deep learning in radiological image understanding. For example, CNN-based systems have achieved competitive performance in chest X-ray interpretation and disease detection tasks [11]. These advances have encouraged the broader adoption of deep learning approaches for semantic medical image understanding, including automatic annotation and concept detection.

2.2. CNN Architectures for Medical Concept Detection

Among CNN architectures, DenseNet and EfficientNet have shown promising performance in medical imaging applications due to their efficient feature propagation and scalability. DenseNet introduces dense connectivity between layers, allowing features extracted at earlier stages to be reused throughout the network and mitigating the vanishing gradient problem [8]. Such characteristics make DenseNet particularly effective in medical imaging tasks where subtle visual patterns may be clinically important.

EfficientNet proposes a compound scaling strategy that balances network depth, width, and image resolution to achieve strong predictive performance with reduced computational cost [9]. EfficientNet-based models have been successfully applied to several medical image classification problems, demonstrating robust performance under limited computational resources [12]. Due to their complementary strengths, DenseNet and EfficientNet remain widely adopted architectures for transfer learning in medical image analysis.

In the context of ImageCLEFmedical concept detection, convolutional architectures continue to provide strong baselines for multi-label prediction tasks, particularly when combined with transfer learning strategies and carefully designed preprocessing pipelines.

2.3. Transfer Learning and Multi-Label Classification

Transfer learning has become a standard practice in medical image analysis due to the limited availability of large-scale annotated medical datasets. Models pretrained on large natural image datasets can be fine-tuned for medical applications, often yielding substantial performance improvements compared to training from scratch [6, 7]. This approach enables more effective feature learning while reducing computational requirements.

Medical concept detection is commonly formulated as a multi-label classification problem, where a single image may be associated with multiple semantic concepts simultaneously [13]. Unlike traditional single-label classification, multi-label settings require models to capture co-occurring clinical findings and anatomical structures, making evaluation particularly challenging due to label imbalance and semantic overlap. Consequently, transfer learning combined with CNN-based multi-label classification has become an effective paradigm for automatic medical concept detection.

3. Dataset

3.1. Dataset Description

For the ImageCLEFmedical 2026 Concept Detection task, the development dataset is based on an updated and extended version of the Radiology Objects in COntext Version 2 (ROCOv2) dataset [14, 5]. ROCOv2 is designed to support multimodal medical image understanding tasks such as concept detection, image captioning, and medical image retrieval.

The dataset consists of radiological images extracted from biomedical articles in the PubMed Central (PMC) Open Access subset and includes associated captions and Unified Medical Language System (UMLS) concepts [14]. Compared to earlier versions, ROCOv2 introduces additional radiological images, improved concept extraction, and manually curated metadata for imaging modalities, anatomical regions, and directional information for selected modalities. The dataset contains approximately 79,789 radiological images and provides a valuable benchmark for training and evaluating medical image understanding systems.

As in previous ImageCLEFmedical editions, the test set consists of previously unseen images, requiring participating systems to generalize effectively beyond the training distribution [5]. The concept detection task is formulated as a multi-label classification problem, where each image may be associated with multiple medically relevant concepts derived from a filtered subset of the UMLS 2022 AB release. To improve feasibility and reduce noise, concepts were filtered according to semantic type and frequency, retaining only clinically meaningful and sufficiently represented labels.

3.2. Preprocessing

To ensure consistency between the 2025 training configuration and the 2026 evaluation, images from the 2026 test set were preprocessed using the same procedures applied in our previous DeepLens pipeline. Input images were resized to match the expected resolution of the pretrained architectures and normalized using standard image normalization techniques. Label representations were transformed into multi-hot encoded vectors suitable for multi-label classification.

These preprocessing steps preserved compatibility between the previously trained CNN models and the updated 2026 test data, enabling reuse of the 2025 models for evaluation on the new benchmark, including the additional fine-tuning of DenseNet described in Section 4.2.1.

4. Methodology

4.1. Overall Pipeline

The comprehensive pipeline of our multi-label medical concept detection framework is illustrated in Figure 1. Our approach to medical concept detection follows a convolutional neural network (CNN)-based multi-label classification framework developed in our previous participation in ImageCLEFmedical 2025 [10]. For the 2026 edition, the pipeline was largely reused: the EfficientNet-B0 and EfficientNet-B1 models were applied directly to the 2026 benchmark, while DenseNet was further adapted through limited additional fine-tuning as described in Section 4.2.1. Given an input medical image, the system predicts multiple medically relevant concepts simultaneously using CNN-based feature extraction followed by a multi-label classification head. The final predictions are obtained through sigmoid activation, enabling independent probability estimation for each concept label. To investigate model robustness and generalization, we evaluated multiple CNN architectures individually and in an ensemble setting.

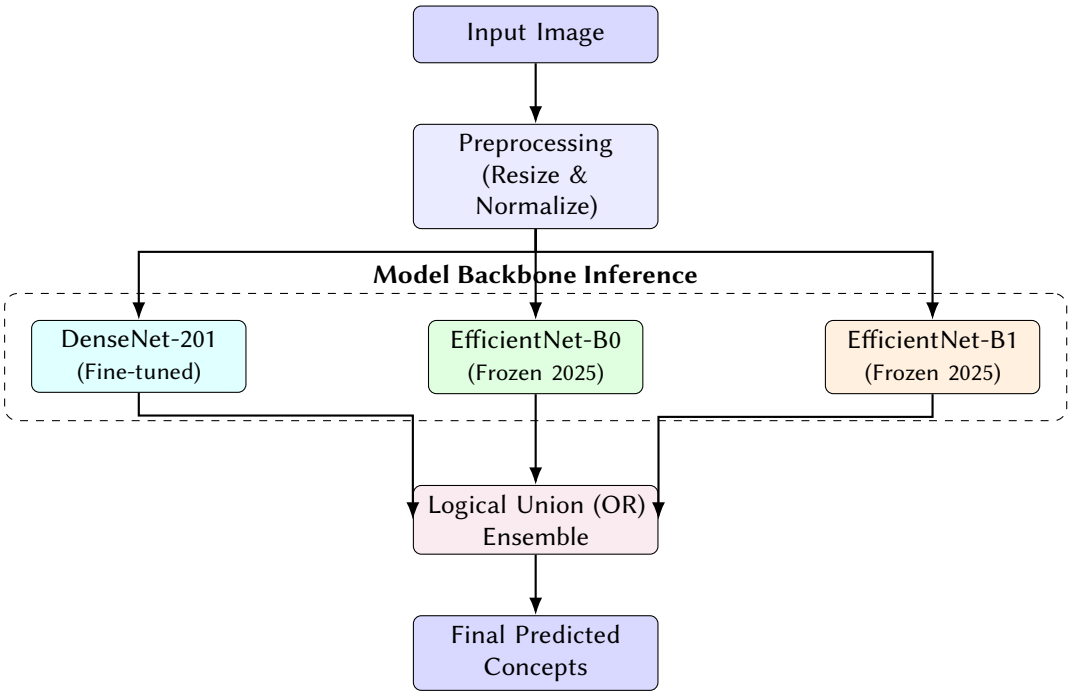


Figure 1: Overall DeepLens pipeline architecture for the ImageCLEFmedical 2026 Concept Detection task. Preprocessed images are simultaneously passed through three distinct deep learning backbones. Their independent label sets are subsequently integrated using a logical union (OR) aggregation strategy to yield the final predictions.

4.2. CNN-Based Architectures

4.2.1. DenseNet

DenseNet was employed as one of the primary backbone architectures due to its ability to preserve low-level and high-level visual information through dense feature connectivity [8]. Dense connections between layers improve gradient propagation and encourage feature reuse, making DenseNet particularly suitable for medical imaging tasks where subtle visual patterns may be diagnostically relevant.

In our 2025 pipeline, a pretrained DenseNet model was adapted for multi-label concept detection by replacing the final classification layer with a task-specific output head corresponding to the target concept space, with sigmoid activation applied to produce independent probabilities for each label. For

the 2026 submission, limited additional fine-tuning of DenseNet was performed for approximately eight epochs to assess its adaptability under the restricted experimental conditions of this year's participation.

4.2.2. EfficientNet Models

We further evaluated EfficientNet-B0 and EfficientNet-B1 architectures, which were selected due to their favorable trade-off between computational efficiency and predictive performance [9]. EfficientNet employs compound scaling of network depth, width, and resolution, allowing efficient feature extraction while maintaining strong representational capacity.

In our original 2025 pipeline development [10], these models were initialized with ImageNet-pretrained weights and fine-tuned on the multi-label medical dataset. To handle the co-occurrence of multiple labels, both backbones were trained using Binary Cross-Entropy (BCE) loss and optimized via the Adam optimizer with an initial learning rate of 10^{-3} and a mini-batch size of 32. Input image dimensions were set to 224×224 pixels for EfficientNet-B0 and 240×240 pixels for EfficientNet-B1. A learning rate scheduler (`ReduceLROnPlateau`) was applied to decrease the learning rate by a factor of 0.1 when the validation loss plateaued.

Similar to DenseNet, these pretrained EfficientNet backbones were adapted for multi-label classification by replacing the final fully connected classification layer to match the exact number of target medical concepts from the 2025 challenge space, utilizing a sigmoid activation function for independent probability estimation per label. For the 2026 submission, these specific trained EfficientNet models were frozen and reused directly from our previous DeepLens repository without further modification or secondary training, allowing us to evaluate their baseline generalizability on the updated 2026 ImageCLEFmedical dataset distribution.

4.3. Ensemble Strategy

To aggregate multi-label predictions across our standalone configurations (DenseNet, EfficientNet-B0, and EfficientNet-B1), we applied a logical union (OR) aggregation strategy. Mathematically, let $P_m \in \{0,1\}^C$ denote the binary prediction vector generated by model $m \in M$ for a given image, where C represents the total number of medical concepts in the vocabulary. The ensemble's final prediction vector P_{ens} is formulated element-wise as:

$$P_{\text{ens},c} = \bigvee_{m \in M} P_{m,c} \quad \forall c \in \{1, \ldots, C\} \tag{1}$$

Under this paradigm, a medical concept code is assigned to a target image if at least one model in the pool predicts its presence with a confidence score exceeding its individual classification threshold.

The primary theoretical justification for adopting a logical union is the maximization of prediction recall. In complex multi-label clinical settings where a single image may contain a sparse distribution of multiple overlapping anatomical or pathological labels, individual networks often suffer from high false-negative rates due to variations in architectural feature representation or localized gradient behaviors. By accumulating the distinct conceptual insights discovered across diverse backbones, the logical union attempts to construct an all-inclusive label set.

However, a critical limitation of this strategy is its high susceptibility to inflating false positives. Because a logical union relies entirely on a permissive inclusion rule, a single noisy or over-confident prediction from any individual network automatically forces a true positive claim in the final vector. This risk is particularly severe when individual models primarily converge and agree on highly predictable, high-frequency concepts, which destroys prediction diversity and leaves the system vulnerable to background feature noise on rare classes.

Our evaluations confirm this theoretical flaw. The ensemble strategy achieved a Primary F1 score of 0.5632, failing to outperform the standalone EfficientNet-B1 baseline (0.5648). This drop in marginal utility suggests that the minor gains achieved in macro-level recall were ultimately offset by a disproportionate inflation of false positives across shared, high-frequency concepts. This outcome demonstrates

that while a logical union can serve as a simple multi-model baseline, more sophisticated intersection criteria or learnable gating mechanisms (such as majority voting or stacking) are necessary to preserve precision as the underlying distribution shifts.

4.4. Training Configuration

The EfficientNet-B0 and EfficientNet-B1 configurations were imported directly from our 2025 DeepLens pipeline and evaluated on the 2026 test collection without any further architectural modifications or weight updates. Conversely, the DenseNet-201 model variant was fine-tuned for a targeted duration of approximately 8 epochs. This network was optimized using the Adam optimizer with an initial learning rate (η_0) set to 0.001 and a weight decay coefficient of 0.0.

To govern parameter steps during fine-tuning, a `ReduceLROnPlateau` learning rate schedule was implemented. The scheduler monitored validation loss under a minimization mode with an adjustment factor of 0.2, an epoch patience threshold of 3, an improvement threshold of 10^{-3}, and a cooldown period of 2 epochs, terminating at a lower boundary learning rate of 10^{-6}. All fine-tuning steps for DenseNet were executed with a mini-batch size of 32, and input images were uniformly resized to 128×128 pixels before being passed to the network backbones.

5. Experimental Results

5.1. Experimental Setup

All validation and evaluation experiments were executed within an environment powered by an NVIDIA T4 GPU with 16 GB of VRAM. The system pipeline was implemented using the PyTorch deep learning framework. All models were evaluated on the official 2026 test split using the same framework.

5.2. Evaluation Metric

Performance was assessed using the official evaluation framework of the ImageCLEFmedical 2026 Concept Detection task [5]. Following the challenge protocol, system runs are evaluated by computing the F_1 score between the predicted and ground-truth concept sets for each individual image using the default binary averaging implementation of the `scikit-learn` library. For each target image, binary indicator arrays (y_{pred} and y_{true}) are constructed across the active vocabulary, and the instance-level F_1 score is computed. The final challenge score is derived by summing these individual instance scores and averaging them over the total number of elements in the test collection.

To make our evaluation completely self-contained, we explicitly define the operational and structural differences between the two reported metrics tracks:

- **Primary F1-Score:** This track evaluates comprehensive concept coverage across the entire test collection. It considers any and all predicted or ground-truth medical concepts belonging to the filtered subset of the UMLS 2022 AB release established for this year's challenge.
- **Secondary F1-Score:** This track introduces a focused sub-evaluation that isolates high-confidence terms. Before executing the F_1 scoring calculation, both the system's predicted concept sets and the official ground-truth sets are strictly filtered down to a refined subset containing exclusively the manually annotated concepts.

Tracking both metrics allows us to evaluate how our models perform on the full automatically-extracted concept distribution versus the highly accurate, human-curated annotation subset.

5.3. Results and Discussion

The comparative evaluation of our baseline replications, fine-tuned variants, and ensemble strategy is presented in Table 1. To provide context within the broader challenge cohort, the final official

leaderboard standings for the top-performing participant runs on the ImageCLEFmedical 2026 Concept Detection track are summarized in Table 2.

Our primary submission—utilizing the standalone frozen EfficientNet-B1 baseline architecture—achieved a Primary F1-score of 0.5648 and a Secondary F1-score of 0.9336. This outcome demonstrates that our frozen 2025 pipeline achieves competitive performance on the updated 2026 benchmark without any retraining on the new data.

Table 1
Performance of our submitted configurations on the ImageCLEFmedical 2026 Concept Detection task.

Model	Primary F1	Secondary F1
DenseNet	0.5542	0.9276
EfficientNet-B0	0.5632	0.9323
EfficientNet-B1	**0.5648**	**0.9336**
Ensemble	0.5632	0.9323

Table 2
Official ImageCLEFmedical 2026 Concept Detection Task Leaderboard Standings.

Rank	Team / Owner	Primary F1-Score	Secondary F1-Score
1	dsgt_caption	0.5790	0.9657
2	archimedes	0.5781	0.9591
3	UIT_Worker	0.5725	0.9533
4	basharderar	0.5725	0.9419
5	AUEB/DMST	0.5721	0.9503
6	nhanbui	0.5713	0.9553
7	**DeepLens**	**0.5648**	**0.9336**
8	dingglebellw	0.5606	0.9428
9	NUSTPB	0.5594	0.9465
10	krishnatewari	0.5234	0.8403
11	UIT_HighDefinition	0.2739	0.9349

Among our internal configurations, the standalone EfficientNet-B1 achieved the strongest overall performance. Interestingly, the ensemble configuration (logical union) did not yield an architectural advantage, matching the performance metrics of the individual EfficientNet-B0 run at a Primary F1-score of 0.5632. This exact match suggests that EfficientNet-B0's predictions were a near-superset of DenseNet's activated labels, causing the logical union to contribute no additional correct predictions beyond those already captured by EfficientNet-B0 alone. This indicates a high label prediction overlap between the baseline structures, which limited the marginal utility of a logical union across largely overlapping model outputs.

Furthermore, the fine-tuned DenseNet-201 variant underperformed relative to the frozen 2025 baselines, securing a Primary F1-score of 0.5542. As discussed in Section 6, this performance deficit stems from rapid validation divergence during its short training span rather than structural capacity constraints. Nevertheless, the fact that our completely un-retrained 2025 pipeline achieves a competitive 7th-place finish out of 11 teams highlights the exceptional durability of well-optimized CNN feature extractors across evolving challenge partitions.

6. Discussion and Limitations

The experimental results on the ImageCLEFmedical 2026 benchmark reveal notable insights regarding model performance across consecutive challenge editions [5]. Specifically, our frozen 2025 EfficientNet-B1 baseline achieved a Primary F1 score of 0.5648, outperforming the newly fine-tuned DenseNet variant (0.5542) despite receiving no additional training on the 2026 data.

(a) Loss Progression

(b) F1-Score Evaluation

Figure 2: DenseNet-201 optimization metrics across fine-tuning epochs.

An analysis of the DenseNet training progression (illustrated in Figure 2) indicates that this performance difference is primarily due to early overfitting. The DenseNet model reached its peak validation capability at Epoch 6, achieving its lowest validation loss of 0.0060 and its highest validation F1 score of 0.5370. However, at Epoch 7, a clear divergence occurred: while the training loss continued to decrease to 0.0048, the validation loss rose to 0.0062 and the validation F1 score dropped to 0.5254. This indicates that optimizing with a standard learning rate (10^{-3}) for a limited number of epochs caused the network to overfit the training partition, reducing its ability to generalize effectively to the unseen 2026 test collection.

Additionally, a structural limitation applies to our direct reuse of the 2025 pipeline. Because the classification heads for EfficientNet-B0 and EfficientNet-B1 were kept frozen to assess multi-year baseline durability, their final output layers map exclusively to the 2025 medical concept vocabulary [10]. Consequently, the system encounters a major coverage issue when evaluating the updated 2026 benchmark: the models are incapable of predicting any entirely new medical concepts introduced in this year's updated ROCOv2 dataset [14]. Any novel anatomical terms or clinical concepts automatically result in false negatives, explaining why the standalone 2025 models encountered a performance ceiling. This highlights the necessity of dynamically updating classification heads as benchmark vocabularies expand.

7. Conclusion

In this work, we presented the participation of the DeepLens team from Iran University of Science and Technology in the ImageCLEFmedical 2026 Concept Detection task. Building upon our 2025 participation, where our pipeline achieved second place, we assessed the cross-year generalization capabilities of our established CNN-based framework on the updated 2026 benchmark [10]. To establish a clear evaluation baseline, EfficientNet-B0 and EfficientNet-B1 were evaluated via direct transfer from our 2025 repository without parameter updates, while DenseNet was fine-tuned under a fixed, strict budget of approximately eight epochs to test immediate domain adaptability. An ensemble strategy combining the three configurations via a logical union was also assessed to chart the utility of frozen model aggregations.

Experimental evaluation demonstrated that EfficientNet-B1 achieved the strongest performance among the submitted models, obtaining a primary F1 score of 0.5648 and a secondary F1 score of 0.9336. While ensemble methods remained competitive, the strongest results were obtained using an individual EfficientNet-based architecture.

Overall, the results indicate that CNN-based transfer learning approaches continue to provide robust and competitive baselines for medical concept detection across successive ImageCLEFmedical editions. Future work will focus on more extensive retraining, systematic optimization, and the exploration of multimodal and transformer-based approaches to further improve concept detection performance.

8. Code Availability

The implementation scripts, notebooks detailing the original training steps from our 2025 pipeline, multi-label serialization workflows, and the ensemble union configurations used to construct the submissions for this year's challenge are fully open-sourced and available on the project repository at: https://github.com/DeepLensIUST/ImageCLEF2026-Concept-Detection.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT (OpenAI) for the following purposes: grammar and spelling checking, and writing style improvement. The authors have reviewed and edited all AI-generated content and take full responsibility for the content of this publication.

References

[1] H. Müller, N. Michoux, D. Bandon, A. Geissbuhler, A review of content-based image retrieval systems in medical applications—clinical benefits and future directions, International Journal of Medical Informatics 73 (2004) 1–23. doi:10.1016/j.ijmedinf.2003.11.024.

[2] G. Litjens, T. Kooi, B. E. C. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, C. I. Sáanchez, A survey on deep learning in medical image analysis, Medical Image Analysis 42 (2017) 60–88. doi:10.1016/j.media.2017.07.005.

[3] J. Wang, H. Zhu, S.-H. Wang, Y.-D. Zhang, A review of deep learning on medical image analysis, Mobile Networks and Applications 26 (2021) 351–380. doi:10.1007/s11036-020-01672-7.

[4] B. Ionescu, H. Müller, D. Stanciu, A. Radu, R. Bolborici, M. Negru, A. Ene, V. Vasilescu, A.-A. Nicolae, L. Ştefan, M. Constantin, M. Dogariu, A. Andrei, H. Damm, T. M. G. Pakull, A. Ben Abacha, A. García Seco de Herrera, C. M. Friedrich, R. Brüngel, L. Reinartz, H. Schäfer, C. S. Schmidt, B. Bracke, P. Nath, B. Eryılmaz, M. Hjuler, D. Fabre, C. Lemaire, B. Lecouteux, D. Schwab, D. Dimitrov, M. S. Hee, M. Ahsan, S. Ahmad, D. Zlatkova, G. Pachov, Z. Xie, P. Nakov, I. Koychev, J. E. Heras Rivera, D. K. Low, W. Yim, J. Ruzevick, D. Child, M. Kurt, Z. Sun, F. Xia, M. Yetisgen, A. Radzhabov, Y. Prokopchuk, V. Kovalev, D. Karpenka, S. A. Hicks, S. Gautam, M. A. Riegler, V. Thambawita, P. Halvorsen, M. El Sakka, J. Mothe, A. Băicoianu, C. Florea, M. Ivanovici, Overview of ImageCLEF 2026: Multimodal challenges in medicine, science, agritech, and security, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction, Proceedings of the Seventeenth International Conference of the CLEF Association (CLEF 2026), Springer Lecture Notes in Computer Science LNCS, Jena, Germany, 2026.

[5] H. Damm, T. M. G. Pakull, L. Reinartz, B. Bracke, B. Eryılmaz, P. Nath, R. Brüngel, C. S. Schmidt, H. Schäfer, A. Ben Abacha, A. García Seco de Herrera, H. Müller, C. M. Friedrich, Overview of ImageCLEFmedical 2026 – medical concept detection and caption generation with synthetic data extensions, in: CLEF2026 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org, Jena, Germany, 2026. To appear.

[6] P. Kora, C. P. Ooi, O. Faust, et al., Transfer learning techniques for medical image analysis: A review, Biocybernetics and Biomedical Engineering 42 (2022) 79–107. doi:10.1016/j.bbe.2021.11.004.

[7] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks?, in: Advances in Neural Information Processing Systems (NeurIPS), 2014.

[8] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261–2269. doi:10.1109/CVPR.2017.243.

[9] M. Tan, Q. V. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, in: Proceedings of the 36th International Conference on Machine Learning (ICML), 2019, pp. 6105–6114.

[10] A. Salimi Rudsari, B. Kavousi Nejad, M. Hajihosseini, S. Eetemadi, Detecting concepts for medical images: Contributions of the DeepLens team at IUST to ImageCLEFmedical caption, in: CLEF 2025 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org, Bologna, Italy, 2025. URL: https://ceur-ws.org/Vol-4038/paper_204.pdf.

[11] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, A. Y. Ng, CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning, 2017. URL: https://arxiv.org/abs/1711.05225. arXiv:1711.05225.

[12] M. Shvetsova, D. Romanov, D. Dylov, EfficientNet-based convolutional neural network for automatic COVID-19 detection from chest X-ray images, Frontiers in Medicine 8 (2021) 798055.

[13] G. Tsoumakas, I. Katakis, Multi-label classification: An overview, International Journal of Data Warehousing and Mining (IJDWM) 3 (2007) 1–13. URL: https://doi.org/10.4018/jdwm.2007070101. doi:10.4018/jdwm.2007070101.

[14] J. Rückert, L. Bloch, R. Brüngel, A. Idrissi-Yaghir, H. Schäfer, C. S. Schmidt, S. Koitka, O. Pelka, A. B. Abacha, A. G. S. de Herrera, H. Müller, P. Horn, F. Nensa, C. M. Friedrich, ROCOv2: Radiology objects in context version 2, an updated multimodal image dataset, Scientific Data 11 (2024). doi:10.1038/s41597-024-03496-6.