ISSN 2007-9737

Detecting Non-CXR Inputs via Validation-Calibrated Thresholds in a Three-Level Convolutional Autoencoder for Chest Radiographs

Rodolfo Salgado-Rivera[1], Jorge Emmanuel Zamora-Zamora[1], Cornelio Yáñez-Márquez[1], Oscar Camacho-Nieto[2,*], Antonio Alarcón-Paredes[1]

[1] Instituto Politécnico Nacional, Centro de Investigación en Computación, Mexico

[2] Instituto Politécnico Nacional, Centro de Innovación y Desarrollo Tecnológico en Cómputo, Mexico City, Mexico

rsalgador2024@cic.ipn.mx, jzamoraz2024@cic.ipn.mx, ocamacho@ipn.mx, cyanez@cic.ipn.mx, aalarcon@cic.ipn.mx

Abstract. Ensuring that images submitted to web-based diagnostic systems are genuine chest radiographs (CXR) is essential to prevent errors in automated disease detection workflows. This work proposes a three-level convolutional autoencoder trained exclusively on CXR images to identify non-CXR inputs through reconstruction-error analysis. A threshold calibrated on the 95th percentile of the validation MSE distribution (0.000960) was used to determine class membership. The model achieved balanced accuracies above 0.89 across multiple test domains, including medical (BUSI, Dental OPG, MildDemented) and non-medical datasets (ImageNet, Baggage X-ray). The results confirm the model's capability to act as a lightweight, unsupervised gate that filters off-target images before diagnostic inference. This approach supports data integrity, reliability, and operational safety in CXR-based disease detection systems.

Keywords. Autoencoder, chest X-ray validation, reconstruction error, out-of-distribution detection, web-based medical imaging.

1 Introduction

Chest radiography (CXR) is among the most frequently requested imaging studies in clinical practice and supports a wide spectrum of cardiothoracic assessments, from routine screening to emergency decision making. In modern, high-throughput workflows, ensuring that inputs are truly CXRs and exhibit consistent modality characteristics is increasingly recognized as a prerequisite for reliable downstream analysis and triage in computer-aided diagnosis (CAD) systems [28, 32, 35]. Early, image-centric validation stages that safeguard dataset integrity can mitigate failure modes arising from heterogeneous inputs, missing or inconsistent metadata, and distributional drift across acquisition sites [18, 6, 19].

The relevance of robust CXR pipelines is underscored by reports showing that both model performance and calibration depend critically on dataset curation and modality consistency [35, 18]. As multi-institutional repositories and real-world deployments expand, sources of covariate shift (e.g., equipment, positioning, export formats) and label noise can compound, degrading generalization if not addressed by principled data hygiene and front-end verification [6, 19]. Recent work also highlights operational motivations: screening services and tele-radiology settings benefit from automated guards that reject non-target inputs prior to inferencing, preventing propagation of spurious cases through disease classifiers and prioritization queues [28, 32, 9].

Traditionally, pipelines have relied on accession-level checks, DICOM metadata heuristics, or

rule-based prefilters (e.g., view tags) to constrain inputs. While useful, these approaches are brittle under mixed provenance, legacy PACS exports, and inconsistent tagging, and they offer limited protection when metadata are incomplete or incorrect [35, 18].

Machine learning has transformed CXR analysis through supervised convolutional neural networks (CNNs) and transfer learning, enabling accurate screening and triage and spurring interest in modular pipelines that integrate pre-analysis quality control [28, 32, 25]. In parallel, unsupervised and one-class formulations—particularly autoencoders that learn the manifold of in-distribution medical images—offer a principled way to detect off-target or out-of-distribution (OOD) inputs without enumerating exhaustive negative classes [29, 26, 1, 24].

However, several gaps remain. First, dedicated, unsupervised modules for modality verification (e.g., "is this image a CXR?") are seldom treated as first-class, calibrated components in CXR pipelines; when present, they often appear as ancillary checks with limited reporting on operating-point selection and validation procedures [29, 35]. Second, thresholding strategies for reconstruction-error detectors are inconsistently documented, sometimes tuned heuristically or on test data, which hinders reproducibility and deployment [26, 1]. Third, recent evidence from distribution-shift and robustness studies suggests that input-space irregularities (including off-target views or non-CXR content) contribute materially to downstream performance variance, reinforcing the need for an explicit front-end gate [18, 6, 19].

Although these challenges are specific to radiography, insights from other computer-vision fields demonstrate how domain-specific validation can be achieved through homogeneous training data and threshold-based decision strategies. Recent advances in computer vision have enabled automatic validation and classification systems capable of achieving high accuracy in diverse visual domains. For instance, convolutional neural networks have been successfully applied to the early detection of plant diseases using web-based deep learning systems, reaching accuracies above 98 % [5]. Similarly, lightweight convolutional architectures such as MobileNetV2 have demonstrated efficiency in real-time visual classification tasks like sign-language recognition [10].

These results illustrate the robustness and versatility of convolutional approaches and motivate their adaptation to the validation of medical radiographic inputs in clinical workflows.

The purpose of this study is to formalize and evaluate an autoencoder-based input-validation stage for CXR pipelines. We propose and train a three-level convolutional autoencoder exclusively trained on chest radiographs and derive a decision threshold from the reconstruction-error distribution on an independent validation set to reject non-CXR images while preserving true CXRs for downstream analysis. This design contributes a transparent, dataset-grounded calibration protocol and quantifies the effect of a front-end gate on data integrity in pre-diagnostic workflows [29, 28, 32, 19].

2 Related Work

Deep learning has proven effective in various visual domains, including agriculture, gesture recognition, and medical imaging, demonstrating the adaptability of convolutional models across contexts [5, 10].

Other approaches have focused on computer-vision-based analysis for plant classification or pH estimation using supervised neural networks [7].

Research on automated CXR analysis has been led by supervised CNNs trained on large public corpora, where transfer learning and attention mechanisms have delivered strong normal–abnormal discrimination and promising triage behavior [28, 25, 12]. Hybrid pipelines that combine representation learning with reconstruction objectives have also been explored for screening support; for example, chest-screening systems integrating transfer learning with convolutional sparse denoising autoencoders (CSDAE) aggregate classifier outputs and reconstruction signals to stratify studies and assist workflow decisions [32, 9].

In addition, meta-heuristic optimization methods like the Whale Optimization Algorithm combined with Support Vector Machines (WOA–SVM) have been proposed for feature selection in 2D image data, improving classification performance through

parameter tuning [22]. These complementary findings reinforce the value of compact and optimization-driven designs, a principle also applied in our lightweight autoencoder for radiographic validation.

Complementary efforts examine data-centric reliability, documenting how heterogeneity and label issues in real-world datasets affect model calibration and generalization, and advocating for better curation and validation practices [35, 18, 6].

Unsupervised anomaly detection in radiographic imaging leverages manifold learning to separate in-distribution anatomy from atypical inputs without relying on exhaustive negative labels. In high-dimensional X-ray contexts, convolutional autoencoders (often paired with density models such as Gaussian mixtures) have been used to produce anomaly scores that support dataset construction and input verification [29, 26]. Variants including variational and denoising autoencoders (e.g., EVAe-Net, CSDAE) have been proposed to improve robustness under distribution shift and noise, with encouraging results in both thoracic and non-thoracic settings [1, 32, 30]. Within CXR, recent studies explore self-supervised and one-class formulations to enhance representation quality for OOD recognition and anomaly scoring, further motivating dedicated front-end validation modules [19, 24, 17].

Beyond the thorax, radiographic pipelines in other anatomical domains provide additional evidence for autoencoder-based validation. For instance, in gastric X-ray examinations, deep autoencoding combined with mixture modeling has successfully excluded off-target esophageal images during organ-specific dataset assembly, illustrating the utility of reconstruction-based criteria for large-scale curation [29]. Related lines examine attention-driven and cross-view learning strategies that, while primarily supervised, point to the value of robust feature formation that can complement unsupervised validation at ingestion [25, 34, 20]. Taken together, these works suggest a coherent path: calibrate unsupervised, image-centric validators on held-out data, integrate them as first-line gates for modality verification, and combine them—when appropriate—with supervised modules for downstream diagnosis and triage [35, 4, 15, 3, 21, 23].

3 Materials and Methods

This section describes the design, training, and evaluation process of the proposed model for validating image membership in the chest X-ray class. The model's main objective is to learn a compact representation of the images belonging to a base class (chest X-rays) and use the reconstruction error as an indicator of membership or deviation from the training domain.

The model was evaluated using five datasets with distinct visual characteristics to analyze its generalization capabilities.

3.1 Architecture

The proposed architecture follows the principle of lightweight convolutional designs to minimize computational cost while maintaining discriminative capacity. Such compact models have proven efficient in similar visual tasks [10, 22], supporting their adoption in medical validation contexts.

The encoder-decoder structure is adapted to perform data validation rather than categorical classification, with thresholds calibrated from the validation set to separate in-domain (CXR) and out-of-domain (non-CXR) inputs.

The proposed model is a symmetric convolutional autoencoder created using Keras [8], designed to reconstruct RGB images of size 224×224×3. The architecture is composed of an encoder, which progressively reduces the spatial dimensionality, and a decoder, which reconstructs the original image from the compressed latent representation.

In total, the network contains 25,731 parameters, of which 25,619 are trainable and 112 are non-trainable, corresponding mainly to Batch Normalization statistics. Image resizing and normalization were performed using the Rescaling layer from framework.
Table 1 shows the main operations of the proposed autoencoder.

Table 1. Main operations

Stage	Main operations
Encoder	Conv2D + MaxPooling (×3)
Decoder	Conv2DTranspose + Conv2D

Table 2. Output size and Filters

Stage	Output size	Filters
Encoder	(28×28×8)	32 → 16 → 8
Decoder	(224×224×3)	8 → 16 → 32

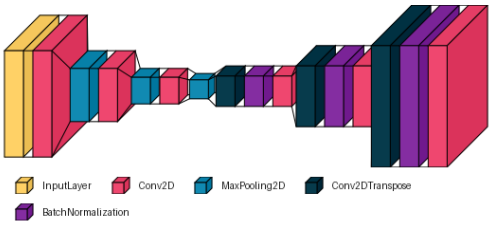

Fig. 1. Proposed architecture

3.1.1 Encoder

The encoder consists of three convolutional blocks interleaved with max-pooling operations. In the first block, a Conv2D layer with 32 filters and a 3×3 kernel extracts low-level visual features, followed by a MaxPooling2D layer that reduces the spatial resolution from 224×224 to 112×112.

The second block applies a convolution with 16 filters and another pooling operation, further decreasing the resolution to 56×56. The third block contains a Conv2D layer with 8 filters, followed by a final pooling operation that reduces the output to 28×28×8, defining the latent space of the model.

This descending configuration (32 → 16 → 8 filters) enables the network to efficiently compress information, retaining the most relevant structural and spatial features while discarding redundancies.

3.1.2 Decoder

The decoder performs the inverse reconstruction process using transposed convolutions and standard convolutions.
Initially, a Conv2DTranspose layer expands the latent representation to 56×56×8, followed by a BatchNormalization layer to stabilize training.

A subsequent Conv2D layer with 16 filters refines the reconstructed features. Two additional upsampling blocks (Conv2DTranspose) increase the resolution to 112×112×16 and finally to 224×224×32, each followed by normalization and convolution operations (Conv2D) to smooth the reconstruction.

The final output layer is a Conv2D with three filters and a sigmoid activation, generating the reconstructed image with the same color space and dimensions as the original input.
Table 2 shows the size of the outputs and the number of layers of the proposed autoencoder.

3.1.3 Regularization and Stability Mechanisms

To improve training stability and accelerate convergence, Batch Normalization layers were included after each upsampling step.

No dropout or L2 regularization was applied, since the model was optimized to minimize the Mean Squared Error (MSE) loss function (equation 1), prioritizing high reconstruction fidelity between input and output images:

$$MSE(y, \hat{y}) = \frac{\sum_{i=0}^{N-1}(y_i - \hat{y}_i)^2}{N}. \qquad (1)$$

3.1.4 Summary and Rationale

This architecture provides a compact yet expressive representation of image data, balancing structural simplicity and representational power.

The design allows the model to reconstruct images with high fidelity while effectively distinguishing between in-domain and out-of-domain samples based on reconstruction error statistics.

The proposed architecture is shown in Fig. 1, the image was made with visualkeras [14].

3.2 Test datasets

For the positive class, the NIH Chest X-rays TFRecords dataset [31] was used, consisting of chest radiograph images.

The dataset was split into three independent partitions: 78,484 images for training, 28,030 for validation, and 5,606 for testing.
To assess the model's performance against data from different domains (negative class), five additional datasets were employed:

— Breast Ultrasound Images Dataset (BUSI), breast ultrasound scans [2].

— Dental OPG X-ray Dataset, panoramic dental radiographs [16].

— Baggage X-ray Threats, baggage scans containing threat and non-threat items [33].

— ImageNet, general-purpose natural images [11].

— MildDemented, brain MRI images [27].

For the evaluation, the BUSI and Baggage X-ray Threats datasets were used in their entirety, as their sizes are smaller than the NIH test partition.

For the remaining datasets (Dental OPG X-ray, ImageNet, and MildDemented), a random sample of 5,606 images was selected to match the size of the test partition, ensuring a balanced and comparable evaluation across domains.

This configuration allowed analyzing the model's ability to discriminate between in-domain and out-of-domain data based on reconstruction error distributions.

3.3 Experimental procedure

The model was trained using the training and validation partitions of the NIH Chest X-rays TFRecords dataset, as previously described. The Adam optimizer and Mean Squared Error (MSE) loss function were employed to minimize the reconstruction error between the input and the output images generated by the autoencoder.

Three callbacks were integrated into the training process to enhance model stability and performance:

1. **EarlyStopping**: configured with `monitor='val_loss'`, a patience of 5 epochs, and `restore_best_weights=True`, to prevent overfitting by stopping training early when the validation loss did not improve.
2. **ModelCheckpoint**: set to `save_best_only=True` to save the best-performing model based on `monitor='val_loss'` in `.keras` format within the output directory.
3. **ReduceLROnPlateau**: configured to halve the learning rate (`factor=0.5`) after two epochs without improvement in validation loss, with a minimum learning rate of 1×10^{-6}.

Training was executed on a workstation with the following hardware configuration:

— **Processor:** Intel® Core™ i7-14700K (20 cores, 28 threads, base clock 800 MHz, boost up to 5.6 GHz).

— **Architecture:** x86_64 with AVX2, FMA y AVX-VNNI instruction support.

— **Cache memory:** L1 = 1.8 MB, L2 = 28 MB, L3 = 33 MB.

— **Virtualization:** VT-x enabled.

— **Operating System:** Ubuntu 24.04.3 LTS (*noble*).

The total training time was approximately 65 minutes, using CPU-based computation. This procedure yielded the best-performing model on validation data, ensuring stable and reproducible results across subsequent evaluations. Fig. 2 shows the performance of the autoencoder training.

After training, the validation set was used to determine the **membership threshold** for the positive class. For each validation image, the reconstruction **Mean Squared Error (MSE)** was computed and its distribution analyzed.

The **95th percentile** value (0.000960) was selected as the threshold, where images with MSE below or equal to this limit were considered in-domain, and those above it were classified as out-of-domain samples.

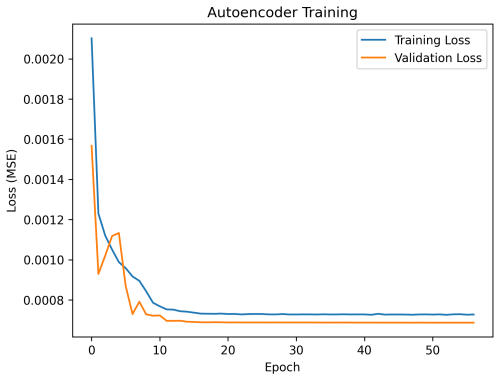

Fig. 2. Autoencoder training performance with the training and validation sets

Table 3. Summary of model performance across test datasets

Dataset	I.R.	B.A.
Breast Ultrasound (BUSI)	7.19	0.9141
Dental OPG X-ray	10.84	0.9230
Baggage X-ray Threats	1.00	0.7526
ImageNet	1.00	0.9209
MildDemented	1.00	0.8906

4 Results

The model's performance was evaluated across the five test datasets described previously.
For each dataset, the Imbalance Ratio (I.R.), confusion matrix components (TP, FN, FP, TN), and derived metrics —Sensitivity, Specificity, Precision, F1 score, Balanced Accuracy (B.A.), and Matthews Correlation Coefficient (MCC)— were computed.
The results are summarized as follows:

— **Breast Ultrasound Images Dataset (BUSI)**:
I.R. = 7.19; TP = 4743, FN = 863, FP = 14, TN = 766.
Sensitivity = 0.8461, Specificity = 0.9821, Precision = 0.9971, F1 = 0.9154, Balanced Accuracy = 0.9141, MCC = 0.6221.

— **Dental OPG X-ray Dataset**:
I.R. = 10.84; TP = 4743, FN = 863, FP = 0, TN = 517.
Sensitivity = 0.8461, Specificity = 1.0000, Precision = 1.0000, F1 = 0.9166, Balanced Accuracy = 0.9230, MCC = 0.5630.

— **Baggage X-ray Threats**:
I.R. = 1.00; TP = 4743, FN = 863, FP = 1911, TN = 3695.
Sensitivity = 0.8461, Specificity = 0.6591, Precision = 0.7128, F1 = 0.7738, Balanced Accuracy = 0.7526, MCC = 0.5142.

— **ImageNet**:
I.R. = 1.00; TP = 4743, FN = 863, FP = 24, TN = 5582.
Sensitivity = 0.8461, Specificity = 0.9957, Precision = 0.9950, F1 = 0.9145, Balanced Accuracy = 0.9209, MCC = 0.8514.

— **MildDemented**:
I.R. = 1.00; TP = 4743, FN = 863, FP = 363, TN = 5243.
Sensitivity = 0.8461, Specificity = 0.9352, Precision = 0.9289, F1 = 0.8856, Balanced Accuracy = 0.8906, MCC = 0.7844.

The performance of the model was evaluated across five test datasets, as described in the methodology section.
For clarity and space optimization in the two-column format, only the **Imbalance Ratio (I.R.)** and **Balanced Accuracy** are presented in Table 3.
These two indicators effectively summarize the relationship between class distribution and the overall balance between sensitivity and specificity, providing a concise overview of the model's discriminative performance across domains.

As shown, the model achieved the highest balanced accuracy on the **Dental OPG** and **ImageNet** datasets, with values above 0.92, indicating strong generalization to both medical and non-medical domains. The lowest performance was observed on the **Baggage X-ray Threats** dataset (0.75), reflecting the visual and structural dissimilarity of this data compared to the chest X-ray training set. Overall, the model maintained balanced accuracy above 0.89 on most datasets, confirming the robustness of the proposed threshold calibration strategy.

5 Conclusion and Future Work

The autoencoder architecture achieved high discrimination between valid and non-valid medical images, demonstrating its potential as a front-end validation stage for diagnostic systems. This generalization across diverse datasets is consistent with observations in deep learning applications to plant-disease detection [5] and color-based classification in agricultural imaging [7], where visual variability did not hinder robust model performance. These results suggest that threshold-calibrated convolutional autoencoders can effectively act as anomaly detectors across visual domains.

The proposed convolutional autoencoder demonstrated that reconstruction-based thresholds derived from validation data can reliably differentiate chest radiographs from non-CXR inputs. The model achieved consistent balanced accuracy across all non-CXR datasets (mean B.A. = 0.88). This stability supports its reliability as a validation layer preceding automated diagnostic models, reducing the risk of propagating spurious or unrelated images.

Future work will integrate the validation module within a web-based diagnostic platform for real-time image verification and explore threshold adaptation to varying acquisition conditions, multi-view (PA/LAT) validation, and hybrid unsupervised–supervised architectures. Furthermore, incorporating multimodal validation layers—combining image and textual medical data—may improve dataset integrity, as shown by feature-based extraction of medical entities in Spanish [13].

Acknowledgments

The authors gratefully acknowledge the support of the Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECTIHI), the Laboratorio de Cómputo Inteligente of the Centro de Investigación en Computación, Instituto Politécnico Nacional (CIC-IPN), and the Centro de Nanociencias y Micro y Nano Tecnologías (CNMN-IPN), whose assistance, resources, and expertise were instrumental in the successful completion of this research.

References

1. **Addo, D., Zhou, S., Jackson, J. K., Nneji, G. U., Monday, H. N., Sarpong, K., Patamia, R. A., Ekong, F., Owusu-Agyei, C. A. (2022).** EVAE-Net: An Ensemble Variational Autoencoder Deep Learning Network for COVID-19 Classification Based on Chest X-ray Images. Diagnostics, Vol. 12, No. 11, pp. 2569. DOI: 10.3390/diagnostics12112569.
2. **Aki, S. (2023).** Breast Ultrasound Images Dataset (BUSI). https://www.kaggle.com/datasets/sabahesaraki/breast-ultrasound-images-dataset.
3. **Alloqmani, A., B., Y., Irshad, A., Alsolami, F. (2021).** Deep Learning based Anomaly Detection in Images: Insights, Challenges and Recommendations. IJACSA, Vol. 12, No. 4. DOI: 10.14569/IJACSA.2021.0120428.
4. **Anai, S., Hisasue, J., Takaki, Y., Hara, N. (2022).** Deep Learning Models to Predict Fatal Pneumonia Using Chest X-Ray Images. Canadian Respiratory Journal, Vol. 2022, pp. 1–12. DOI: 10.1155/2022/8026580.
5. **Arredondo-Basurto, E. R., Serrano-Talamantes, J. F., Olguín-Carbajal, M., Sandoval-Gutiérrez, J., Herrera-Lozada, J. C., Rivera-Zárate, I., Hernández-Bolaños, M. (2025).** Detección de enfermedades en hojas de tomate usando aprendizaje profundo. Computación y Sistemas, Vol. 29, No. 3. DOI: 10.13053/cys-29-3-5518.
6. **Baugh, M., Tan, J., Müller, J. P., Dombrowski, M., Batten, J., Kainz, B. (2023).** Many Tasks Make Light Work: Learning to Localise Medical Anomalies from Multiple Synthetic Tasks. In **Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R.**, editors, Medical Image Computing and Computer Assisted Intervention – MICCAI 2023, Vol. 14220. Springer Nature Switzerland, Cham, pp. 162–172. DOI: 10.1007/978-3-031-43907-0_16.
7. **Brizuela-Ramírez, J. E., García-Díaz, N., García-Virgen, J., Gutiérrez-Magaña, S. M.,**

Rico-Bautista, D. (2025). Classification of pH Status in Stevia Plants Using Supervised Neural Networks and Computer Vision. Computación y Sistemas, Vol. 29, No. 3. DOI: 10.13053/cys-29-3-5902.

8. **Chollet, F., others (2015).** Keras. https://keras.io.
9. **Costa, M., Pereira, S. C., Pedrosa, J., Mendonça, A. M., Campilho, A. (2023).** Deep Feature-Based Automated Chest Radiography Compliance Assessment. 2023 IEEE 7th Portuguese Meeting on Bioengineering (ENBENG), IEEE, Porto, Portugal, pp. 64–67. DOI: 10.1109/ENBENG58165.2023.10175341.
10. **Dash, S. K., Padhi, A., Sahu, A. K., Sahu, M. (2025).** Indian Sign Language Recognition using MobileNetV2 Fine-Tuned by Transfer Learning. Computación y Sistemas, Vol. 29, No. 3. DOI: 10.13053/cys-29-3-5894.
11. **Dimensi0n (2023).** ImageNet 256. https://www.kaggle.com/datasets/dimensi0n/imagenet-256.
12. **Faisal, M., Darmawan, J. T., Bachroin, N., Avian, C., Leu, J. S., Tsai, C.-T. (2023).** CheXViT: CheXNet and Vision Transformer to Multi-Label Chest X-Ray Image Classification. 2023 IEEE International Symposium on Medical Measurements and Applications (MeMeA), IEEE, Jeju, Korea, Republic of, pp. 1–6. DOI: 10.1109/MeMeA57477.2023.10171855.
13. **García-Robledo, G. A., Cuevas-Rasgado, A. D., Bravo, M., Reyes-Ortiz, J. A. (2025).** Generation of Feature Vectors for Identifying Medical Entities in Spanish. Computación y Sistemas, Vol. 29, No. 3. DOI: 10.13053/cys-29-3-5002.
14. **Gavrikov, P., Patapati, S. (2020).** visualkeras. https://github.com/paulgavrikov/visualkeras. Publication Title: GitHub repository.
15. **Gholamipoor, R., Rafiee, N., Kollmann, M. (2022).** Pneumonia Detection With Semantic Similarity Scores. 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), IEEE, Kolkata, India, pp. 1–5. DOI: 10.1109/ISBI52829.2022.9761494.
16. **IMT Kaggle Team (2024).** Dental OPG X-ray Dataset. https://www.kaggle.com/datasets/imtkaggleteam/dental-opg-xray-dataset.
17. **Jahan, N., Hasan, M. A. M. (2022).** Autoencoder-based Unsupervised Anomaly Detection for Covid-19 Screening on Chest X-Ray Images. 2022 19th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), IEEE, Mexico City, Mexico, pp. 1–6. DOI: 10.1109/CCE56709.2022.9975962.
18. **Kim, H., Lee, S., Shim, W. J., Choi, M.-S., Cho, S. (2023).** Homogenization of multi-institutional chest x-ray images in various data transformation schemes. J. Med. Imag., Vol. 10, No. 6. DOI: 10.1117/1.JMI.10.6.061103.
19. **Liu, X., Alvén, J., Häggström, I., Zach, C. (2024).** Deep Nearest Neighbors for Anomaly Detection in Chest X-Rays. In **Cao, X., Xu, X., Rekik, I., Cui, Z., Ouyang, X.**, editors, Machine Learning in Medical Imaging, Vol. 14349. Springer Nature Switzerland, Cham, pp. 293–302. DOI: 10.1007/978-3-031-45676-3_30.
20. **Luo, L., Chen, H., Xiao, Y., Zhou, Y., Wang, X., Vardhanabhuti, V., Wu, M., Han, C., Liu, Z., Fang, X. H. B., Tsougenis, E., Lin, H., Heng, P.-A. (2022).** Rethinking Annotation Granularity for Overcoming Shortcuts in Deep Learning–based Radiograph Diagnosis: A Multicenter Study. Radiology: Artificial Intelligence, Vol. 4, No. 5, pp. e210299. DOI: 10.1148/ryai.210299.
21. **Mao, Y., Xue, F.-F., Wang, R., Zhang, J., Zheng, W.-S., Liu, H. (2020).** Abnormality Detection in Chest X-Ray Images Using Uncertainty Prediction Autoencoders. In **Martel, A. L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M. A., Zhou, S. K., Racoceanu, D., Joskowicz, L.**, editors, Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, Vol. 12266. Springer International Publishing, Cham, pp. 529–538. DOI: 10.1007/978-3-030-59725-2_51.

22. **Medjahed, S. A., Boukhatem, F. (2025).** WOA-SVM: Whale Optimization Algorithm and Support Vector Machine for Hyperspectral Band Selection and 2D Images Feature Selection. Computación y Sistemas, Vol. 29, No. 3. DOI: 10.13053/cys-29-3-5102.

23. **Ohshima, H., Mishima, K. (2023).** Oral biosciences: The annual review 2022. Journal of Oral Biosciences, Vol. 65, No. 1, pp. 1–12. DOI: 10.1016/j.job.2023.01.008.

24. **Rafiee, N., Gholamipoor, R., Kollmann, M. (2023).** Abnormality Detection for Medical Images Using Self-Supervision and Negative Samples. DOI: 10.1101/2023.05.29.542748.

25. **Rocha, J., Pereira, S. C., Pedrosa, J., Campilho, A., Mendonca, A. M. (2022).** Attention-driven Spatial Transformer Network for Abnormality Detection in Chest X-Ray Images. 2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS), IEEE, Shenzen, China, pp. 252–257. DOI: 10.1109/CBMS55023.2022.00051.

26. **Siddalingappa, R., Kanagaraj, S. (2021).** Anomaly Detection on Medical Images using Autoencoder and Convolutional Neural Network. IJACSA, Vol. 12, No. 7. DOI: 10.14569/IJACSA.2021.0120717.

27. **Singhal, A. (2024).** Alzheimer's Multiclass Dataset (MildDemented). https://www.kaggle.com/datasets/aryansinghal10/alzheimers-multiclass-dataset-equal-and-augmented.

28. **Tang, Y.-X., Tang, Y.-B., Peng, Y., Yan, K., Bagheri, M., Redd, B. A., Brandon, C. J., Lu, Z., Han, M., Xiao, J., Summers, R. M. (2020).** Automated abnormality classification of chest radiographs using deep convolutional neural networks. npj Digit. Med., Vol. 3, No. 1, pp. 70. DOI: 10.1038/s41746-020-0273-z.

29. **Togo, R., Watanabe, H., Ogawa, T., Haseyama, M. (2020).** Deep convolutional neural network-based anomaly detection for organ classification in gastric X-ray examination. Computers in Biology and Medicine, Vol. 123, pp. 103903. DOI: 10.1016/j.compbiomed.2020.103903.

30. **Tsuneki, M. (2022).** Deep learning models in medical image analysis. Journal of Oral Biosciences, Vol. 64, No. 3, pp. 312–320. DOI: 10.1016/j.job.2022.03.003.

31. **Uzmenkov, N. (2021).** NIH Chest X-rays TFRecords. www.kaggle.com/datasets/nickuzmenkov/nih-chest-xrays-tfrecords.

32. **Wang, C., Elazab, A., Jia, F., Wu, J., Hu, Q. (2018).** Automated chest screening based on a hybrid model of transfer learning and convolutional sparse denoising autoencoder. BioMed Eng OnLine, Vol. 17, No. 1, pp. 63. DOI: 10.1186/s12938-018-0496-2.

33. **Zakria, M. (2025).** Baggage X-ray Threats. https://www.kaggle.com/datasets/muhammadzakria2001/baggage-xray-threats.

34. **Zhu, Z., Chen, Q., Yu, L., Wang, L., Zhang, D., Magnier, B., Wang, L. (2023).** Cross-View Deformable Transformer for Non-displaced Hip Fracture Classification from Frontal-Lateral X-Ray Pair. In **Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R.**, editors, Medical Image Computing and Computer Assisted Intervention – MICCAI 2023, Vol. 14225. Springer Nature Switzerland, Cham, pp. 444–453. DOI: 10.1007/978-3-031-43987-2_43.

35. **Çallı, E., Sogancioglu, E., Van Ginneken, B., Van Leeuwen, K. G., Murphy, K. (2021).** Deep learning for chest X-ray analysis: A survey. Medical Image Analysis, Vol. 72, pp. 102125. DOI: 10.1016/j.media.2021.102125.

Article received on 28/08/2025; accepted on 10/12/2025.
**Corresponding author is Oscar Camacho-Nieto.*