# DEEP LEARNING IN ACTION

## Deep Learning Application: Automated Medical Image Diagnostics

Murthy Kanuri

Intelligent Agents

University of Essex

## Overview of the Technology (What it Does)

This technology uses deep learning—typically convolutional neural networks (CNNs)—to analyse medical images such as X-rays, MRIs, CT scans and pathology slides (Esteva et al., 2017; Ardila et al., 2019). The aim is to detect, classify and quantify disease features that may be subtle or hard for the human eye to identify consistently. Applications include early cancer detection (e.g., lung nodules, skin lesions) and screening for diabetic retinopathy (Gulshan et al., 2016), with potential to improve reporting consistency and throughput, and—if evaluated across diverse populations—to support health equity (WHO, 2021).

## Brief Synopsis of How it Works

Deep models require large, labelled datasets—typically thousands of expert-annotated images (Esteva et al., 2017).

- **Training:** The CNN learns spatial patterns linked to disease (e.g., texture/morphology separating malignant tumours from benign tissue).
- **Feature extraction:** Deep nets learn hierarchical features directly from data, avoiding manual feature engineering (LeCun, Bengio & Hinton, 2015).
- **Prediction:** In deployment, the model processes a new **unlabelled** scan and outputs a probability (e.g., likelihood of malignancy) to assist the clinician; many pipelines add **explanations** (e.g., Grad-CAM heatmaps) for transparency (Selvaraju et al., 2017).

## Potential Socio-Technical Impacts (Ethics, Privacy, Fairness)

Using deep learning in high-stakes industries like health raises several important ethical and social issues.

*A. Bias and Fairness (Social Justice)*

- Risk: If the training data underrepresent certain populations, performance disparities may be aggravated (Seyyed-Kalantari et al., 2021).
- Mitigation: Curate **representative** datasets; report performance disaggregated by sex, age, ethnicity and site/scanner; and conduct external validation across multiple hospitals and imaging systems (Gulshan et al., 2016).

*B. Transparency and Accountability (Ethics)*

- **Risk:** CNNs are often viewed as **'black boxes,'** making clinical trust difficult and complicating the assignment of responsibility when errors occur.
- **Mitigation**: We must **maintain human oversight** throughout the process; integrate explainability methods like **Grad-CAM** *after* the prediction; and clearly **document the model's limits and governance** (Selvaraju et al., 2017; WHO, 2021).

*C. Data Privacy and Consent*

- *Risk:* Training at scale uses sensitive health data; de-identification does not eliminate re-identification risk in all cases (Rocher, Hendrickx & de Montjoye, 2019).
- Mitigation: The methodology must be framed with the principles of data minimisation and purpose limitation in the UK GDPR (ICO, 2024), achieve an appropriate legal basis or consent and apply solid technical and organisational measures to ensure protection (ICO, 2024). Researchers should also consider privacy preservation policies for sensitive data, such as access control policy, audit policy and federated learning if feasible.

# References


- **Ardila, D.**, Kiraly, A.P., Bharadwaj, S. **et al.** (2019) 'End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest CT', *Nature Medicine*, 25(6), pp. 954–961. https://doi.org/10.1038/s41591-019-0447-x
- **Esteva, A.**, Kuprel, B., Novoa, R.A. **et al.** (2017) 'Dermatologist-level classification of skin cancer with deep neural networks', *Nature*, 542(7639), pp. 115–118. https://doi.org/10.1038/nature21056
- **Gulshan, V.**, Peng, L., Coram, M. **et al.** (2016) 'Development and validation of a deep learning algorithm for detection of diabetic

retinopathy in retinal fundus photographs', *JAMA*, 316(22), pp. 2402–2410. https://doi.org/10.1001/jama.2016.17216

- **Information Commissioner's Office (ICO)** (2024) *UK GDPR: Principles – data minimisation*. Available at: https://ico.org.uk/ (Accessed: 13 October 2025).
- **LeCun, Y.**, Bengio, Y. and Hinton, G. (2015) 'Deep learning', *Nature*, 521(7553), pp. 436–444. https://doi.org/10.1038/nature14539
- **Rocher, L.**, Hendrickx, J.M. and de Montjoye, Y.-A. (2019) 'Estimating the success of re-identifications in incomplete datasets using generative models', *Nature Communications*, 10, 3069. https://doi.org/10.1038/s41467-019-10933-3
- **Selvaraju, R.R.**, Cogswell, M., Das, A. **et al.** (2017) 'Grad-CAM: Visual explanations from deep networks via gradient-based localization', in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 618–626. https://doi.org/10.1109/ICCV.2017.74
- **Seyyed-Kalantari, L.**, Zhang, H., McDermott, M.B.A., Chen, I.Y. and Ghassemi, M. (2021) 'Underdiagnosis bias of AI algorithms applied to chest X-rays in under-served patient populations', *Nature Medicine*, 27, pp. 2176–2182. https://doi.org/10.1038/s41591-021-01595-0
- **World Health Organization (WHO)** (2021) *Ethics and governance of artificial intelligence for health*. Geneva: WHO. Available at: https://www.who.int/publications/i/item/9789240029200 (Accessed: 13 October 2025).