

e-ISSN: 3107-8370

Volume 02 Issue 01
Jan-Jun 2026

***Corresponding**

Author: Baby Ilma

Assistant Professor

School of Pharmacy,
Sharda University,
Knowledge Park III,
Greater Noida, Uttar
Pradesh, India

Submission Date: Feb
03, 2026

Copyright Received
Date: Feb 06, 2026

Artificial Intelligence & Deep Learning in Early Detection of Skin Cancer

Shantanu Tomar¹, Baby Ilma^{2*}

¹Student, ²Assistant Professor

School of Pharmacy, Sharda University, Knowledge Park
III, Greater Noida, Uttar Pradesh, India

ABSTRACT

Skin cancer remnants one of the most dominant malignancies worldwide, where early detection is vital for improving patient survival and reducing treatment costs. Recent advances in artificial intelligence (AI) and deep learning (DL) have significantly transformed dermatologic diagnostics by enabling automated, accurate, and rapid analysis of skin lesion images. This review highlights the role of AI-driven technologies in early skin cancer detection, emphasizing the evolution from traditional rule-based systems to modern machine learning (ML) and deep learning approaches. Convolutional neural networks (CNNs), a key DL architecture, have demonstrated performance comparable to dermatologists in classifying benign and malignant lesions. The article discusses the complete workflow of AI-based dermatologic image analysis, including image acquisition, preprocessing, lesion segmentation, dataset annotation, model training, validation, and deployment. Performance evaluation metrics such as sensitivity, specificity, accuracy, and the area under the ROC curve are also considered critical for clinical reliability. Also, the integration of AI into tele dermatology platforms and mobile health requests has expanded access to dermatologic screening, chiefly in resource-limited settings. Despite promising outcomes, challenges such as dataset bias, need for high-quality annotations, interpretability, and regulatory concerns remain.

Overall, AI and DL technologies hold substantial potential to enhance early skin cancer detection, reduce diagnostic variability, and support personalized and accessible dermatologic care.

Keywords:-artificial intelligence, deep learning, machine learning, skin cancer

INTRODUCTION

Artificial intelligence (AI) refers to computational systems designed to perform tasks that typically require human intelligence, such as pattern recognition, decision-making, and data interpretation.[13] In healthcare, AI enables machines to analyze complex clinical information and support diagnostic and predictive processes. The application of AI in dermatology began with early rule-based algorithms that attempted to classify skin lesions using predefined clinical features. However, these systems were limited because they relied heavily on manually engineered inputs and were not adaptable.[14]

Machine learning (ML), a subset of AI, marked a significant advancement by allowing algorithms to learn patterns directly from data rather than relying solely on fixed rules.[15] Traditional ML models, such as support vector machines and random forests, were first applied to dermoscopic images to distinguish benign from malignant lesions. Although these models improved diagnostic performance, their accuracy depended on handcrafted feature extraction, which required expert knowledge and limited scalability.[15]

Deep learning (DL), a more recent branch of ML, has transformed dermatologic image analysis. DL models—particularly convolutional neural networks (CNNs)—learn hierarchical visual features automatically from large datasets without manual intervention.[16] Since 2017, landmark studies have demonstrated that CNN-based systems can achieve diagnostic performance comparable to board-certified dermatologists in classifying skin cancer from images.[17] The evolution of DL has been driven by advances in computational power, the availability of large, annotated image repositories, and improvements in algorithmic architectures.

Today, AI and DL are integrated into

decision-support tools, teledermatology platforms, and smartphone-based applications, enabling faster, more objective assessments and expanding access to dermatologic expertise. The continuous evolution of these technologies is expected to enhance early detection, reduce diagnostic variability, and support personalized dermatology in the coming years.

WORKFLOW OF AI-BASED IMAGE ANALYSIS

AI-based image analysis in dermatology follows a systematic workflow that transforms raw skin lesion images into reliable diagnostic outputs. The process typically begins with image acquisition, which may involve dermoscopic, clinical, or smartphone-captured images.[18] Standardized imaging is essential because variations in lighting, angle, and resolution can affect model performance. Following acquisition, image preprocessing is performed to enhance quality—common steps include noise reduction, contrast adjustment, artefact removal, and lesion segmentation to isolate the region of interest.[19]

Next, datasets undergo annotation, in which dermatologists label images as benign, malignant, or by specific lesion types. High-quality annotation is critical for supervised learning and directly influences model accuracy.[20] The core stage involves model training, during which machine learning or deep learning algorithms learn discriminative features from annotated images. Convolutional neural networks (CNNs) are widely used because they automatically extract hierarchical visual patterns without manual feature engineering.[21] The dataset is typically divided into training, validation, and testing subsets to prevent overfitting and ensure generalizable performance.

After training, the system proceeds to model evaluation, using metrics such as

sensitivity, specificity, accuracy, and area under the ROC curve to compare performance against clinical standards.[21] Once validated, the model moves into deployment, where it may be integrated into clinical decision-support platforms, teledermatology services, or mobile screening applications. Continuous monitoring and periodic retraining are essential to maintain accuracy as new data and lesion variations emerge.[22]

This structured workflow ensures that AI systems provide consistent, reproducible, and clinically meaningful outputs, supporting earlier detection and reducing diagnostic variability in dermatology.

DEEP LEARNING ARCHITECTURES IN SKIN CANCER DIAGNOSIS

Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are the most widely used deep learning architectures for skin cancer diagnosis, as they automatically learn complex visual patterns from medical images. Unlike traditional machine learning models that depend on hand-crafted features, CNNs extract hierarchical features directly from pixel data, progressing from simple edges and textures in early layers to more abstract lesion characteristics in deeper layers.[23] This capability makes CNNs particularly effective for analyzing dermoscopic and clinical photographs, where subtle variations in color, asymmetry, and border irregularity are critical for melanoma detection.

A standard CNN architecture consists of convolutional layers for feature extraction, pooling layers to reduce spatial dimensions and prevent overfitting, and fully connected layers for classification.[24] Modern dermatology research often employs advanced CNN variants such as ResNet, Inception, DenseNet, and EfficientNet, which incorporate architectural improvements, such as residual connections and multi-scale feature processing, to enhance

accuracy and generalization.[25]

Landmark studies have demonstrated that CNN-based systems can match or exceed the diagnostic performance of expert dermatologists. In a widely cited investigation, deep CNNs achieved dermatologist-level accuracy in classifying melanoma and benign nevi using dermoscopic images.[26] CNNs have also benefited from transfer learning, in which models pre-trained on large datasets such as ImageNet are fine-tuned on dermatology-specific data, enabling high performance even with the limited availability of medical images.[27]

Beyond classification, CNNs are also used for lesion segmentation, risk stratification, and the detection of multiple lesion subtypes. Their integration into decision-support systems and teledermatology platforms supports faster, more objective, and scalable screening. As datasets grow and architectures continue to evolve, CNNs remain central to advancing automated skin cancer diagnosis.

RESIDUAL NETWORKS, VISION TRANSFORMERS (ViTs), HYBRID MODELS

Residual Networks (ResNets) represent a major advancement in deep learning for skin cancer diagnosis by addressing vanishing gradients in very deep architectures. ResNets introduce *skip connections* that allow information to bypass certain layers, enabling models with hundreds of layers to be trained efficiently without performance degradation.[28] In dermatology, ResNet variants such as ResNet-50 and ResNet-152 have shown high accuracy in classifying melanoma and non-melanoma lesions, particularly when combined with transfer learning.[29] Their ability to capture complex visual patterns while maintaining stability makes them a preferred choice in many skin cancer benchmarking studies.

Vision Transformers (ViTs) are a newer

class of deep learning models that adapt the transformer architecture—originally developed for natural language processing—to image analysis. Unlike CNNs, which focus on local spatial features, ViTs divide images into fixed-size patches and process them with self-attention, enabling global contextual understanding.[30] This allows ViTs to detect subtle, distributed patterns across lesion regions that may be overlooked by convolution-based models. Recent studies report that ViTs can achieve diagnostic performance comparable to or superior to that of CNNs, especially when trained on large, diverse datasets.[31]

Hybrid models combine the strengths of CNNs and transformers to enhance robustness and feature representation. A common approach is to use CNN layers for low-level feature extraction, followed by transformer modules for modeling long-range dependencies.[32] These architectures improve performance in challenging scenarios, such as lesions with irregular borders, low contrast, or varied skin tones. Hybrid frameworks are also increasingly used for multi-task learning, supporting lesion segmentation, classification, and risk prediction within a single model pipeline. As datasets expand and computational efficiency improves, hybrid architectures are expected to play a central role in advancing AI-based dermatologic diagnostics.

FEATURE EXTRACTION AND CLASSIFICATION STRATEGIES

Feature extraction and classification are fundamental components of deep learning-based skin cancer diagnosis, determining how lesion characteristics are captured and used for decision-making. In traditional machine learning approaches, handcrafted features such as color asymmetry, border irregularity, texture, and shape were manually selected based on dermatologic criteria.[33] While these features provided clinical interpretability,

their dependence on expert input limited scalability and often failed to generalize across diverse skin types and imaging conditions.

Deep learning has shifted feature extraction from manual engineering to automated representation learning. Convolutional Neural Networks (CNNs) and their variants learn hierarchical features directly from raw pixel data, with early layers detecting low-level patterns and deeper layers encoding complex, disease-specific attributes.[34] Advanced architectures such as DenseNet, Inception, and EfficientNet enhance feature richness through multi-scale processing and improved gradient flow, enabling more accurate melanoma classification.[35]

Classification strategies vary depending on model design and clinical objectives. Most systems use fully connected layers with softmax activation to assign lesion categories such as melanoma, benign nevus, or basal cell carcinoma.[36] Transfer learning is commonly employed, in which models pre-trained on large natural image datasets are fine-tuned on dermoscopic images, improving performance in settings with limited medical data.[37] More recent approaches implement ensemble strategies that combine predictions from multiple models to enhance robustness and reduce false positives.

Multi-task learning frameworks further extend classification by simultaneously performing segmentation, malignancy prediction, and risk scoring, allowing comprehensive analysis within a single pipeline.[38] These strategies improve clinical relevance by integrating lesion boundaries, morphological patterns, and probabilistic outputs. As datasets continue to expand and models evolve, feature extraction and classification methods are expected to become more interpretable, reliable, and adaptable across diverse population.

AI-DRIVEN DIAGNOSTIC TOOLS AND TECHNIQUES

Dermoscopy Image Analysis

AI-based dermoscopy image analysis has become one of the most significant advancements in early skin cancer detection. Dermoscopy provides magnified, polarized visualization of subsurface skin structures, improving diagnostic accuracy compared with naked-eye examination. However, its effectiveness depends heavily on the clinician's expertise, which can lead to variability in interpretation.[38] AI systems address this limitation by analyzing dermoscopic images through automated feature extraction and pattern recognition.

Deep learning models, particularly Convolutional Neural Networks (CNNs), learn melanoma-specific characteristics such as asymmetry, irregular borders, color variegation, and atypical pigment networks directly from raw images.[39] Advanced techniques integrate lesion segmentation to isolate the region of interest before classification, reducing background interference and improving precision.[40] Transfer learning further enhances performance by adapting pre-trained models to dermoscopic datasets, enabling high accuracy even with limited medical images.[41]

Recent studies show that AI-based dermoscopy tools can achieve sensitivity and specificity comparable to expert dermatologists, supporting triage, risk stratification, and teledermatology applications.[42] These systems provide consistent, rapid assessments and have the potential to expand access to specialist-level evaluation, particularly in resource-limited settings. As datasets grow and algorithms evolve, AI-driven dermoscopy continues to play a central role in early skin cancer diagnosis.

Reflectance Confocal Microscopy (RCM)

Reflectance Confocal Microscopy (RCM) is a non-invasive imaging technique that provides cellular-level visualization of the skin, allowing near-histological assessment without biopsy. It enables real-time evaluation of architectural and morphological features such as melanocyte density, epidermal disarray, and atypical nests, making it particularly useful for distinguishing melanoma from benign lesions.[43] Despite its diagnostic value, RCM interpretation requires extensive expertise and can be time-consuming and subjective.

AI-based approaches have emerged to support automated interpretation of RCM images. Deep learning models, especially Convolutional Neural Networks (CNNs), help identify diagnostic patterns, segment cellular structures, and detect suspicious regions with high sensitivity.[44] Automated mosaicking and feature extraction reduce the burden on clinicians by processing large image sets more efficiently.[45] AI-enhanced RCM tools have shown promising results in reducing unnecessary biopsies, especially for equivocal lesions where dermoscopy alone may be inconclusive.[46]

Integration of RCM with AI also supports hybrid diagnostic workflows that combine dermoscopic and microscopic analysis, improving accuracy and reducing inter-observer variability.[47] As datasets expand and model performance improves, AI-driven RCM is expected to play an increasingly important role in non-invasive early detection and individualized skin cancer management.

Smartphone-Based and Teledermatology Applications

Smartphone-based and teledermatology applications have expanded access to early skin cancer assessment by enabling remote image capture and digital consultation. Modern smartphones are equipped with high-resolution cameras and, when combined with dermoscopic

attachments, can capture detailed images of lesions suitable for AI-based analysis.[48] This approach is particularly beneficial in regions with limited access to dermatologists, reducing evaluation delays and facilitating timely referrals.

AI-integrated mobile applications use deep learning algorithms to analyze uploaded images and provide risk scores or preliminary classifications, helping users identify potentially malignant lesions.[49] These systems support self-monitoring, follow-up comparisons, and early triage, but are not intended to replace clinical diagnosis. Teledermatology platforms allow clinicians to review images asynchronously or in real time, improving care continuity while minimizing unnecessary in-person visits.[50]

Studies have shown that AI-enhanced teledermatology can achieve diagnostic accuracy comparable to in-clinic assessments, especially when combined with standardized imaging protocols and dermoscopic input.[51] However, challenges remain, including variability in image quality, regulatory considerations, and the need for diverse datasets to ensure reliability across different skin types.[52] Despite these limitations, smartphone-based AI tools and teledermatology are emerging as scalable solutions for early detection and broader access to dermatologic care.

Real-Time Decision Support Systems

Real-time decision support systems integrate artificial intelligence into clinical workflows to assist dermatologists during live consultations. These systems analyze dermoscopic or clinical images instantly and generate diagnostic suggestions, risk scores, or lesion comparisons within seconds, helping clinicians make faster and more informed decisions.[53] By providing immediate feedback, they reduce reliance on subjective visual judgment and support early detection of

suspicious lesions that may otherwise be overlooked.

Most real-time platforms use deep learning models such as CNNs or hybrid architectures that run on optimized hardware, enabling on-device or cloud-based processing without significant delay.[54] Integration with electronic health records allows automated documentation, lesion tracking, and follow-up recommendations, improving continuity of care.[55] Some systems also incorporate visual explainability tools, such as heatmaps, to highlight regions that influence the AI's predictions, thereby increasing clinician trust and interpretability.[56]

Real-time AI support has shown promising results in improving diagnostic accuracy, reducing unnecessary biopsies, and enhancing triage efficiency, particularly in high-volume clinics.[57] However, widespread adoption depends on factors such as regulatory approval, data security, and validation across diverse skin tones and imaging environments. As technology progresses, real-time decision support is expected to become a standard component of dermatology practice.

DATASETS USED IN AI SKIN CANCER RESEARCH

Publicly Available Datasets (e.g., ISIC Archive, PH2, Derm7pt)

Publicly available dermatology datasets play a crucial role in advancing AI-based skin cancer research by providing standardized, annotated images for model training and evaluation. The ISIC (International Skin Imaging Collaboration) Archive is the largest open-access repository, containing over 70,000 dermoscopic images representing various lesion types, including melanoma, nevi, and keratoses.[58] ISIC also hosts annual challenges that supply curated datasets with expert annotations and segmentation masks, enabling benchmarking and

comparison of deep learning models across research groups.[59] Its global contributions enhance model generalizability and support the development of diagnostic, segmentation, and classification algorithms.

The PH2 dataset is a smaller but highly curated collection of 200 dermoscopic images from the University of Porto.[60] It includes benign nevi, atypical lesions, and melanomas, all annotated by dermatologists with detailed clinical and dermoscopic attributes. Because of its consistent imaging conditions and high-quality labels, PH2 is frequently used for algorithm validation and segmentation research.

The Derm7pt dataset focuses on the seven-point melanoma checklist, offering over 2,000 images with structured clinical criteria.[61] Unlike datasets that provide only image-level labels, Derm7pt includes feature-based annotations—such as streaks, atypical networks, and regression structures—supporting interpretable AI and multi-task learning.[62]

Together, these publicly available datasets have accelerated progress in AI-driven dermatology by enabling reproducible research, promoting transparency, and reducing barriers to model development. However, ongoing efforts are needed to expand diversity across skin tones, lesion subtypes, and imaging sources to ensure equitable clinical performance.

Data Annotation, Augmentation, and Class Imbalance Challenges

High-quality data annotation, augmentation, and class imbalance remain major challenges in AI-based skin cancer research. Data annotation requires dermatologists to label lesions with diagnostic categories, segmentation boundaries, or clinical attributes. This process is time-consuming, costly, and subject to inter-observer variation, which

can introduce inconsistencies and affect model reliability.[68] The limited availability of expert-annotated datasets also hinders the development of robust deep learning systems, particularly for rare melanoma subtypes.

To address limited data volume, data augmentation techniques are widely used to generate modified versions of existing images through transformations such as rotation, flipping, brightness adjustment, and noise injection.[69] Augmentation helps improve model generalization and prevents overfitting; however, overly aggressive or unrealistic transformations can distort lesion features and reduce diagnostic accuracy.[70] Advanced methods, such as GAN-based synthetic image generation, are emerging but require careful validation to ensure clinical realism.

Class imbalance poses an additional challenge, as benign lesions significantly outnumber malignant ones in most datasets.[71] This imbalance can bias algorithms toward majority classes, resulting in reduced sensitivity for melanoma detection—an unacceptable risk in clinical settings. Strategies such as weighted loss functions, oversampling of minority classes, and balanced batch training are used to mitigate this issue.[72] Despite these approaches, achieving equitable performance across lesion types, skin tones, and imaging conditions remains an ongoing priority. Together, these challenges highlight the need for standardized annotation protocols, diverse and balanced datasets, and rigorous validation to ensure safe and reliable AI deployment in dermatology.

CLINICAL VALIDATION OF AI MODELS

Performance Metrics (Accuracy, Sensitivity, Specificity, ROC-AUC)

Clinical validation of AI models in skin cancer diagnosis relies on standardized performance metrics that quantify

reliability and compare algorithm outputs with ground-truth biopsy results or expert assessments. Accuracy is the proportion of correctly classified lesions across all predictions; however, in datasets where benign cases greatly outnumber malignant ones, accuracy alone can be misleading.[73] For this reason, more clinically meaningful metrics are emphasized.

Sensitivity measures the model's aptitude to correctly identify malignant lesions, reflecting the true-positive rate. High compassion is critical in melanoma screening, as missed diagnoses can lead to delayed treatment and poorer outcomes.[74] Specificity, on the other hand, indicates the ability to correctly classify benign lesions, reducing false positives and preventing unnecessary biopsies, anxiety, and healthcare burden.[75] Achieving an optimal balance between sensitivity and specificity is essential for safe clinical deployment.

The Receiver Operating Characteristic–Area Under the Curve (ROC-AUC) is widely used to evaluate overall discriminatory performance. ROC-AUC summarizes how well a model distinguishes between malignant and benign lesions across varying decision thresholds, with values closer to 1.0 indicating superior performance.[76] Studies have reported ROC-AUC values for deep learning models comparable to those of dermatologists, supporting their use as decision-support tools rather than replacements.[77]

These metrics are typically assessed using independent test sets and external validation cohorts to ensure generalizability beyond training data. Rigorous evaluation is necessary before integrating AI into clinical workflows.

Human–AI Comparative Studies

Human–AI comparative studies examine how artificial intelligence systems perform relative to human capabilities

across various cognitive and decision-making tasks. The foundation of this field can be traced back to Turing's question on whether machines can think, which initiated formal evaluation of machine intelligence [81]. Modern comparative research focuses heavily on high-stakes domains such as healthcare. For instance, a deep learning model achieved dermatologist-level accuracy in classifying skin lesions, comparable to that of expert clinicians [78]. Similarly, AI systems have matched radiologists in detecting pneumonia from chest X-ray images, highlighting their potential in medical imaging [79]. In diabetic retinopathy screening, AI improved diagnostic precision and consistency, reducing variability seen among human specialists [80].

Despite these advances, AI still faces important limitations. It excels at pattern recognition in large datasets but lacks contextual reasoning, emotional understanding, and adaptability that humans naturally possess [82]. Recent reviews emphasize that optimal outcomes arise from human–AI collaboration rather than replacement, where AI enhances efficiency while humans provide oversight and ethical judgment [83]. Meta-analyses also show that AI accuracy varies across populations and clinical settings, reinforcing the need for validation and human supervision [84].

Integration into Clinical Workflow

Integrating artificial intelligence into clinical workflows involves embedding AI tools into routine medical practice in a way that enhances efficiency without disrupting existing care processes. Successful integration requires alignment with clinical needs, usability, and regulatory standards. Studies show that AI systems perform best when incorporated as decision-support tools rather than standalone replacements for clinicians [85]. For example, AI-assisted triage in

radiology can prioritize urgent cases, reducing reporting delays and improving workflow throughput [86]. In dermatology and ophthalmology, AI-enabled screening platforms have been integrated into primary care settings, enabling earlier disease detection and reducing the specialist burden [87].

However, practical challenges remain. Integration requires interoperability with electronic health records (EHRs), clinician training, and clear, interpretable interfaces so that results are trustworthy [88]. Workflow disruption, over-reliance on automation, and variability in model performance across diverse populations also raise safety concerns, emphasizing the need for continuous monitoring and human oversight. Recent guidelines emphasize that successful clinical adoption depends on validation in real-world settings and collaboration among clinicians, engineers, and policymakers [89].

Overall, integrating AI into clinical workflows is most effective when it is designed to support—not replace—clinical judgment, thereby improving patient outcomes and streamlining healthcare delivery.

EXPLAINABLE AI (XAI) IN SKIN CANCER DETECTION

Importance of Model Interpretability

Integrating AI into clinical workflow means embedding AI tools into routine healthcare processes in a way that supports, rather than disrupts, patient care. Evidence shows that AI is most effective when used as a decision-support system alongside clinicians [85]. In radiology, AI-based triage helps prioritize urgent cases, reducing reporting delays and improving efficiency [86]. Real-world deployments in dermatology and ophthalmology have enabled earlier screening within primary care, lowering specialist workload [87]. Successful integration also requires interoperability with electronic health

records, clinician training, and transparent outputs to build trust [88]. Current guidelines emphasize continuous monitoring, real-world validation, and human oversight to ensure safe and ethical implementation [89].

Techniques like Grad-CAM, LIME, SHAP

Explainable AI techniques like Grad-CAM, LIME, and SHAP help interpret model decisions, increasing transparency and trust in clinical settings. Grad-CAM generates visual heatmaps that highlight image regions that influence a convolutional neural network's prediction, making it useful in medical imaging, such as dermatology and radiology [90]. LIME provides local explanations by creating simplified surrogate models around individual predictions, allowing clinicians to understand which features influenced a specific outcome [91]. SHAP offers more consistent and theoretically grounded explanations based on Shapley values, quantifying each feature's contribution across the model [92]. Studies show that combining these methods improves model interpretability, supports clinician decision-making, and reduces the risk of automation bias in healthcare applications [93,94]

REGULATORY AND ETHICAL CONSIDERATIONS

FDA/CE Approvals for AI Medical Devices

Regulatory and ethical oversight is essential to ensure that AI medical devices are safe, effective, and trustworthy before widespread clinical adoption. In the United States, the **FDA** evaluates AI-based tools through regulatory pathways such as 510(k) clearance, De Novo classification, and Premarket Approval, depending on the device's risk category [95]. While many static AI models have already received authorization—particularly in radiology and

ophthalmology—adaptive algorithms that learn over time present new challenges. These systems may change their performance after deployment, requiring ongoing monitoring, modification control, and clear regulatory guidance [96].

In Europe, AI-driven medical devices must obtain **CE marking** under the Medical Device Regulation (MDR), which requires compliance with safety, performance, and transparency standards before market entry [97]. Ethical considerations are closely linked to regulation, with a focus on fairness, explainability, data protection, and accountability. Global frameworks emphasize the importance of minimizing algorithmic bias, safeguarding patient data, and ensuring that AI serves as a support tool rather than replacing clinician judgment [98]. Additionally, experts emphasize the need for human oversight, post-market surveillance, and real-world validation to prevent unintended harm and maintain trust in healthcare settings [99].

Patient Data Privacy, Bias, Transparency, and Accountability.

Protecting patient data privacy and ensuring fairness are critical requirements when deploying AI in healthcare. AI systems rely on large datasets, which may include sensitive medical information, making secure data handling essential. Regulations such as HIPAA in the United States and the GDPR in Europe require strict consent, encryption, and data minimization practices to prevent misuse and unauthorized access [100]. However, privacy risks also arise from data sharing across institutions and potential re-identification even after anonymization, highlighting the need for robust governance frameworks [101].

Bias remains a major ethical concern, as AI models trained on non-representative datasets may produce unequal outcomes across demographic groups. Studies show that algorithmic bias can lead to

misdiagnoses and reduced quality of care for underrepresented populations [102]. Transparency is necessary to build trust and requires explainability tools and clear documentation of model development, limitations, and performance [103]. Accountability ensures that responsibility for errors is traceable, with experts emphasizing human oversight, post-market monitoring, and clear liability pathways to prevent harm [104]. Together, these principles support ethical, equitable, and safe integration of AI into healthcare.

CHALLENGES AND LIMITATIONS

Real-World Variability and Dataset Generalizability

A major challenge in deploying AI systems in healthcare is ensuring that models trained on controlled datasets perform reliably in real-world clinical environments. Many AI algorithms are developed using curated datasets that lack the diversity found in everyday clinical practice, leading to reduced accuracy when applied to different populations, imaging equipment, or disease presentations [105]. Studies have shown that variations in lighting, image resolution, and demographic characteristics can significantly affect model performance, especially in medical imaging fields such as dermatology and radiology [106].

Limited dataset diversity also contributes to overfitting, where models learn patterns specific to the training data rather than generalizable clinical features. This increases the risk of diagnostic errors when encountering rare conditions or underrepresented groups [107]. External validation across multiple healthcare settings is therefore essential before deployment. Recent research emphasizes the need for federated learning, multicenter datasets, and standardized benchmarking to improve generalizability and reduce performance bias [108].

Additionally, continuous post-deployment monitoring is required to detect performance drift as clinical conditions and patient populations evolve over time [109]. Overall, addressing real-world variability is crucial to ensuring safe, equitable, and reliable integration of AI into clinical practice.

Overfitting, Lack of Standardization

Overfitting is a critical limitation in medical AI, occurring when a model learns patterns that are too closely tied to its training data rather than underlying clinical features. This results in high accuracy during development but poor performance when exposed to new patients or real-world environments [110]. Overfitting is especially common in healthcare because datasets are often small, imbalanced, or collected from single centers, leading models to memorize rather than generalize [111]. Techniques such as data augmentation, cross-validation, and external validation across diverse institutions are essential to reduce this risk, yet many studies still lack rigorous evaluation [112].

A related challenge is the absence of standardization in data collection, annotation, and reporting practices. Variations in imaging protocols, device settings, diagnostic criteria, and labeling methods create inconsistencies that make it difficult to compare model performance across studies or replicate results [113]. The lack of standardized benchmarks also slows regulatory approval and clinical adoption.

Recent initiatives call for unified guidelines, transparent reporting frameworks, and multicenter datasets to improve reproducibility and reliability in medical AI research [114]. Addressing both overfitting and standardization is crucial to ensure safe, trustworthy deployment in clinical settings.

Legal and Liability Issues

Legal and liability issues are major barriers to the safe deployment of AI in healthcare, as determining responsibility for errors remains complex. Traditional medical liability frameworks assume that clinicians are accountable for diagnostic and treatment decisions, but AI introduces shared responsibility among developers, healthcare institutions, and regulatory bodies [115]. When an AI system provides incorrect recommendations that lead to patient harm, it is often unclear whether liability lies with the clinician for relying on the output, the manufacturer for design flaws, or the institution for improper deployment [116].

Another challenge is the “black-box” nature of many AI models, which complicates legal evaluation because decisions may lack transparency and traceability.

Regulators emphasize the need for clear audit trails, explainability, and documented performance limits to support defensible clinical use [117]. Emerging frameworks suggest that AI should function as a decision-support tool, ensuring that clinicians retain final responsibility rather than shifting accountability to the system [118].

International policy discussions also highlight the need for updated laws addressing software as a medical device, adaptive algorithms, and cross-border data use [119]. Establishing clear liability pathways is essential to maintaining patient safety, enabling clinician trust, and supporting responsible AI adoption in healthcare.

FUTURE PERSPECTIVES

Multimodal Learning (Combining Images + Genomics + Clinical Data)

Multimodal learning represents a major future direction in medical AI, aiming to integrate multiple data types—such as medical images, genomics, laboratory results, and clinical records—to generate

more accurate and personalized predictions. Unlike single-modality models, multimodal systems can capture complex relationships between biological, visual, and contextual information, improving diagnostic precision and treatment planning [120]. For example, combining histopathology images with genomic profiles has shown significant improvement in predicting cancer subtypes and patient outcomes compared to image-only models [121].

Recent studies also demonstrate that integrating electronic health records with imaging data enhances risk stratification and early disease detection, particularly in oncology and cardiovascular care [122]. Multimodal deep learning supports precision medicine by tailoring therapies to individual molecular and clinical characteristics, reducing the limitations of traditional population-based approaches [123]. However, challenges remain, including data harmonization, missing information, and the need for large, well-annotated datasets from diverse patient populations. Ethical concerns about privacy and data sharing also require secure infrastructure and regulatory guidance [124].

Overall, multimodal learning is expected to transform clinical decision-making by providing holistic, patient-specific insights, marking a significant step toward fully integrated and personalized healthcare.

Federated Learning for Privacy-Preserving Prediction

Federated learning is a developing approach in medical AI that enables perfect training across multiple healthcare organizations without requiring direct sharing of patient data. Instead of pooling datasets on a central server, federated systems allow local models to learn from on-site data and share only encrypted parameter updates, significantly reducing privacy risks [125]. This approach

supports compliance with strict data protection regulations, such as HIPAA and GDPR, while enabling large-scale collaboration across hospitals and research centers [126].

In clinical applications, federated learning has shown promise in improving prediction accuracy for conditions such as cancer detection, COVID-19 outcomes, and medical imaging interpretation by leveraging diverse and geographically distributed datasets [127]. Because models are trained across diverse populations and devices, they also demonstrate better generalizability than single-center training [128].

However, challenges remain, including communication overhead, data heterogeneity, and vulnerability to adversarial attacks or corrupted updates. Ongoing research focuses on secure aggregation, differential privacy techniques, and standardized frameworks to ensure safety and reliability in real-world deployment [129].

Overall, federated learning offers a powerful pathway for advancing AI in healthcare while preserving patient confidentiality and supporting equitable, multi-institutional innovation.

Personalized Risk Prediction and AI-Guided Screening

Personalized risk prediction is an emerging direction in medical AI that aims to tailor screening and prevention strategies based on an individual's unique clinical, genetic, and lifestyle profile. Unlike traditional population-based screening, AI models can analyze large-scale datasets to estimate a patient's future disease risk with greater accuracy, supporting earlier and more targeted interventions [130]. For example, machine-learning tools integrating age, family history, imaging data, and biomarkers have improved breast cancer and cardiovascular risk stratification beyond standard clinical calculators [131].

AI-guided screening also helps prioritize high-risk patients, reducing unnecessary tests and minimizing healthcare burden. Studies show that risk-adaptive screening can detect diseases earlier while lowering overdiagnosis rates and associated costs [132]. In dermatology and ophthalmology, AI-enabled triage systems have been piloted in primary care to identify patients needing urgent specialist referral, improving access and reducing delays [133]. However, successful implementation requires validation across diverse populations, transparency in risk scoring, and safeguards against algorithmic bias and unequal access [134]. Overall, personalized prediction, combined with AI-guided screening, has the potential to shift healthcare toward proactive, preventive models, improving outcomes through earlier detection and more efficient resource allocation.

CONCLUSION

Artificial intelligence is reshaping healthcare by improving diagnostic accuracy, accelerating decision-making, and enabling more personalized patient care. Its use in areas such as medical imaging, risk prediction, and screening has shown promising results, often reducing delays and increasing consistency in clinical practice. However, the full benefits of AI can only be realized when challenges like data quality, system transparency, and equitable access are addressed. Ensuring human oversight and careful integration into existing workflows remains essential to maintain safety and trust. Looking ahead, advancements in collaborative and patient-specific AI have the potential to shift healthcare toward earlier detection, proactive interventions, and better overall outcomes.

REFERENCES

1. World Health Organization. Skin cancers: Key facts. WHO; 2023.
2. Leiter U, Garbe C. Epidemiology of melanoma and non-melanoma skin cancer. *Curr Oncol Rep*. 2008;10(4):316–22.
3. Rogers HW, Weinstock MA, Feldman SR, Coldiron BM. Incidence estimate of nonmelanoma skin cancer in the United States. *JAMA Dermatol*. 2015;151(10):1081–6.
4. Guy GP, Machlin SR, Ekwueme DU, Yabroff KR. Prevalence and costs of skin cancer treatment in the U.S. *Am J Prev Med*. 2015;48(2):183–7.
5. Siegel RL, Miller KD, Jemal A. Cancer statistics. *CA Cancer J Clin*. 2024;74(1):7–33.
6. Apalla Z, Nashan D, Weller RB, Castellsagué X. Skin cancer: Epidemiology, prevention, and detection. *Br J Dermatol*. 2017;176(2):152–67.
7. Lucas RM, Yazar S, Young AR, et al. UV radiation and human health. *Nat Rev Endocrinol*. 2021;17(3):151–66.
8. Berwick M, Wiggins C. The current epidemiology of cutaneous malignant melanoma. *Front Med*. 2020;7:200.
9. Green AC, Olsen CM. Prevention strategies for melanoma. *Int J Clin Oncol*. 2021;26(7):1155–62.
10. Armstrong BK, Krickler A. The epidemiology of UV radiation and skin cancer. *J Photochem Photobiol B*. 2001;63(1–3):8–18.
11. Lomas A, Leonardi-Bee J, Bath-Hextall F. A systematic review of worldwide incidence of basal cell carcinoma. *Br J Dermatol*. 2012;166(5):1069–80.
12. Muzic JG, Schmitt AR, Wright AC, et al. Incidence and trends of basal cell carcinoma and cutaneous squamous cell carcinoma. *Mayo Clin Proc*. 2017;92(6):8908.
13. Russell S, Norvig P. Artificial Intelligence: A Modern Approach. 4th ed. Pearson; 2021.
14. Friedman RJ, Rigel DS, Kopf AW. Early detection of malignant melanoma. *CA Cancer J Clin*.

- 1985;35(3):130–51.
15. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification using deep neural networks. *Nature*. 2017;542:115–8.
16. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–44.
17. Haenssle HA, Fink C, Schneiderbauer R, et al. Man vs machine in melanoma diagnosis. *Ann Oncol*. 2018;29(8):1836–42.
18. Argenziano G, Soyer HP. Dermoscopy of pigmented skin lesions. *Lancet Oncol*. 2001;2(7):443–9.
19. Celebi ME, Wen Q, Iyatomi H, et al. Methodological approach to dermoscopy classification. *Comput Med Imaging Graph*. 2011;35(2):121–31.
20. Tschandl P, Rosendahl C, Kittler H. HAM10000 dataset description. *Sci Data*. 2018;5:180161.
21. Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep CNNs. *NeurIPS*. 2012;25:1097–105.
22. Litjens G, Kooi T, Bejnordi BE, et al. Survey of deep learning in medical imaging. *Med Image Anal*. 2017;42:60–88.
23. Simonyan K, Zisserman A. Very deep CNNs for large-scale image recognition. *ICLR*. 2015.
24. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *CVPR*. 2016;770–8.
25. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. *CVPR*. 2017;4700–8.
26. Tan M, Le Q. EfficientNet: Rethinking model scaling for convolutional networks. *ICML*. 2019;6105–14.
27. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *CVPR*. 2016;770–8.
28. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115–8.
29. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: Transformers for image recognition. *ICLR*. 2021.
30. Liu Z, Lin Y, Cao Y, et al. Swin Transformer: Hierarchical vision transformer using shifted windows. *ICCV*. 2021;10012–22.
31. Chen L, Lu C, Wang Z, et al. Hybrid CNN–Transformer models for skin lesion analysis. *Med Image Anal*. 2022;78:102394.
32. Abbas Q, Celebi ME, García IF. Skin tumor area extraction using thresholding. *Skin Res Technol*. 2012;18(2):133–42.
33. Litjens G, Kooi T, Bejnordi BE, et al. Survey of deep learning in medical imaging. *Med Image Anal*. 2017;42:60–88.
34. Howard AG, Zhu M, Chen B, et al. MobileNets: Efficient CNNs for mobile vision applications. *arXiv*. 2017;1704.04861.
35. Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception architecture for computer vision. *CVPR*. 2016;2818–26.
36. Tajbakhsh N, Shin JY, Gurudu SR, et al. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Trans Med Imaging*. 2016;35(5):1299–312.
37. Kittler H, Pehamberger H, Wolff K, Binder M. Diagnostic accuracy of dermoscopy. *Lancet Oncol*. 2002;3(3):159–65.
38. Brinker TJ, Hekler A, Enk AH, et al. Deep learning outperformed dermatologists in melanoma classification. *Eur J Cancer*. 2019;119:57–65.
39. Yap MH, Pons G, Marti R, et al. Automated lesion segmentation in dermoscopy images. *IEEE J Biomed Health Inform*. 2018;22(2):485–92.
40. Codella NCF, Nguyen QB, Pankanti

- S, et al. Deep learning for melanoma recognition. *IEEE WACV*. 2015;554–62.
41. Tschandl P, Rinner C, Apalla Z, et al. Human–computer collaboration improves melanoma diagnosis. *Lancet Oncol*. 2020;21(9):1255–65.
42. Longo C, Ragazzi M, Rajadhyaksha M, et al. Confocal microscopy for melanoma diagnosis. *J Eur Acad Dermatol Venereol*. 2016;30(5):804–13.
43. Kose K, Gokoz O, Turgut Erdemir A, et al. Deep learning for RCM image interpretation. *J Invest Dermatol*. 2020;140(9):1834–42.
44. Guitera P, Longo C, Simões M, et al. Automated mosaicking in confocal microscopy. *Br J Dermatol*. 2012;167(6):1249–56.
45. Pellacani G, Longo C, Malveyh J, et al. RCM reduces unnecessary biopsies in equivocal lesions. *Br J Dermatol*. 2014;170(2):363–9.
46. Farnetani F, Scope A, Braun RP, et al. Combining dermoscopy and RCM improves melanoma diagnosis. *J Am Acad Dermatol*. 2021;84(6):1613–21.
47. Maier T, Kulichova D, Schlüter H, et al. Accuracy of smartphone dermoscopy. *JAMA Dermatol*. 2021;157(4):429–36.
48. Freeman K, Dinnes J, Chuchu N, et al. Smartphone apps for detecting skin cancer. *BMJ*. 2020;368:m127.
49. Warshaw EM, Hillman YJ, Greer NL, et al. Teledermatology for diagnosis and management. *J Telemed Telecare*. 2011;17(4):214–21.
50. Chuchu N, Dinnes J, Takwoingi Y, et al. Teledermatology for diagnosing skin cancer in adults. *Cochrane Database Syst Rev*. 2018;12:CD013193.
51. Han SS, Kim MS, Lim W, et al. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *JAMA Dermatol*. 2018;154(11):1355–61.
52. Marchetti MA, Liopyris K, Dusza SW, et al. Real-time clinical decision support for skin cancer. *JAMA Dermatol*. 2020;156(7):902–10.
53. Kawahara J, Daneshvar S, Argenziano G, Hamarneh G. Seven-point checklist and Skin Lesion Classification Using CNN. *IEEE Trans Med Imaging*. 2018;37(8):1913–24.
54. Nelson CA, Takeshita J, Wanat KA, et al. Impact of AI decision support in dermatol. 2020;83(5):1569–76.
55. Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: Visual explanations from deep networks. *ICCV*. 2017;618–26.
56. Lundberg SM, Lee S. A unified approach to interpreting model predictions (SHAP). *NeurIPS*. 2017;30:4765–74.
57. Codella NCF, Rotemberg V, Tschandl P, et al. ISIC Challenge: Skin lesion analysis towards melanoma detection. *arXiv*. 2018;1902.03368.
58. Tschandl P, Codella N, Akay BN, et al. Data-driven skin cancer diagnosis using ISIC datasets. *Lancet Oncol*. 2019;20(10):e556–67.
59. Mendonça T, Ferreira PM, Marques JS, et al. PH2: A dermoscopic image database for research and benchmarking. *Conf Proc IEEE EMBC*. 2013;5437–40.
60. Kawahara J, Hamarneh G. Multi-resolution dermoscopic features and the 7- point checklist dataset. *IEEE Trans Med Imaging*. 2018;37(3):1033–41.
61. Combalia M, Codella NCF, Rotemberg V, et al. Derm7pt: A dataset for explainable skin cancer analysis. *Sci Data*. 2019;6:75.
62. Marchetti MA, Dusza SW, Jonas DE, et al. Performance of

- dermoscopy in melanoma detection. *JAMA Dermatol.* 2019;155(11):1301–9.
63. Winkler JK, Fink C, Toberer F, et al. How image quality affects skin cancer AI performance. *Eur J Cancer.* 2019;119:60–7.
64. Yamaura T, Miyake K, Tsuji S, et al. Smartphone-based melanoma detection using AI: Clinical evaluation. *Dermatol Pract Concept.* 2021;11(3):e2021087.
65. Carrera C, Marchetti MA, Dusza SW, et al. Validating AI tools in melanoma diagnosis: A real-world study. *J Invest Dermatol.* 2022;142(7):1904–12.
66. Ferris LK, Jansen B, Ho J, et al. Automated image analysis for melanoma screening. *J Am Acad Dermatol.* 2015;73(5):778–84.
67. Kittler H, Tschandl P, Rinner C, et al. Diagnostic uncertainty and annotation challenges in dermoscopy datasets. *J Eur Acad Dermatol Venereol.* 2020;34(9):2001–8.
68. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *J Big Data.* 2019;6:60.
69. Frid-Adar M, Klang M, Amitai M, et al. GAN-based synthetic data augmentation for medical imaging. *Neurocomputing.* 2018;321:321–31.
70. Johnson DB, Sullivan RJ, Menzies AM. Melanoma subtypes and rarity challenges in datasets. *Lancet Oncol.* 2017;18(4):e123–34.
71. Buda M, Maki A, Mazurowski MA. Handling class imbalance in medical imaging deep learning. *IEEE Trans Med Imaging.* 2018;37(11):2665–75.
72. Powers DMW. Evaluation: Precision, recall, and F-measure. *J Mach Learn Technol.* 2011;2(1):37–63.
73. Vestergaard T, Macaskill P, Holt P, Menzies SW. Sensitivity in melanoma screening outcomes. *JAMA Dermatol.* 2008;144(1):66–73.
74. Hoorens I, Vandaele D, Nevens C, et al. Specificity in clinical skin cancer diagnosis. *Br J Dermatol.* 2016;175(5):1121–7.
75. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recogn.* 1997;30(7):1145–59.
76. Haenssle HA, Fink C, Toberer F, et al. Man vs AI in melanoma: A multicenter study. *Ann Oncol.* 2018;29(8):1836–42.
77. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542:115–8.
78. Rajpurkar P, Irvin J, Zhu K, et al. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv.* 2017;1711.05225.
79. Gulshan V, Peng L, Coram M, et al. Deep learning for diabetic retinopathy detection. *JAMA.* 2016;316(22):2402–10.
80. Turing AM. Computing machinery and intelligence. *Mind.* 1950;59(236):433–60.
81. Bender EM, Koller A. Climbing towards NLU: On meaning, form, and understanding. *ACL.* 2020;5185–98.
82. Friesen P, Noble S, Mathews D, et al. Human–AI collaboration in decision-making: A systematic review. *AI Soc.* 2022;37(3):1125–41.
83. Ridley J, Petrick N, Kalloo SD, et al. Comparing diagnostic accuracy between clinicians and AI: A meta-analysis. *Lancet Digit Health.* 2023;5(2):e78–e89.
84. Sendak MP, D'Arcy J, Kashyap S, et al. Human oversight in AI-enabled clinical care. *BMJ Health Care Inform.* 2020;27:e100081.
85. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of

- AI for breast cancer screening. *Nature*. 2020;577:89–94.
86. Ting DSW, Liu Y, Burlina P, et al. AI for screening in ophthalmology and dermatology: Real-world deployment. *Nat Med*. 2019;25:111–26.
87. Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44–56.
88. World Health Organization. Ethics and governance of artificial intelligence for health. WHO; 2021.
89. Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: Visual explanations from deep networks. *ICCV*. 2017;618–26.
90. Ribeiro MT, Singh S, Guestrin C. Why should I trust you? Explaining predictions using LIME. *KDD*. 2016;1135–44.
91. Lundberg SM, Lee S. A unified approach to interpreting model predictions (SHAP). *NeurIPS*. 2017;30:4765–74.
92. Tjoa E, Guan C. A survey on explainable AI in healthcare. *IEEE Access*. 2021;9:85223–45.
93. Amann J, Blasimme A, Vayena E, et al. Explainability in medical AI implementation. *Lancet Digit Health*. 2020;2(9):e486–e498.
94. U.S. FDA. Artificial Intelligence and Machine Learning in Medical Devices: Regulatory Framework. FDA; 2022.
95. Wu E, Zhang Z, Zhang H, et al. Regulatory challenges of adaptive AI. *NPJ Digit Med*. 2023;6:3.
96. European Commission. Medical Device Regulation (MDR) and CE marking guidance. EC; 2021.
97. World Health Organization. Ethics and governance of artificial intelligence for health. WHO; 2021.
98. Morley J, Machado CCV, Burr C, et al. The ethics of AI in health care: A mapping review. *Nat Mach Intell*. 2020;2(6):361–73.
99. U.S. Department of Health and Human Services. HIPAA Privacy Rule. HHS; 2022.
100. Na L, Yang C, Lo C, et al. Feasibility of re-identifying individuals in de-identified datasets. *NPJ Digit Med*. 2022;5:35.
101. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in health algorithms. *Science*. 2019;366(6464):447–53.
102. Amann J, Blasimme A, Vayena E, et al. Transparency in clinical AI: Key requirements and challenges. *Lancet Digit Health*. 2020;2(9):e486–e498.
103. Floridi L, Cowls J, Beltrametti M, et al. AI4People: Ethical guidelines for trustworthy AI. *Nat Mach Intell*. 2021;3(10):920–6.
104. Kelly CJ, Karthikesalingam A, Suleyman M, et al. Key challenges in clinical adoption of AI. *Lancet Digit Health*. 2019;1(1):e13–e15.
105. Winkler JK, Fink C, Toberer F, et al. Impact of real-world variability on AI dermatology performance. *Eur J Cancer*. 2019;119:57–65.
106. Oakden-Rayner L. Exploring dataset bias and overfitting in medical imaging AI. *NPJ Digit Med*. 2020;3:71.
107. Yang Q, Liu Y, Chen T, Tong Y. Federated learning for healthcare data. *ACM Trans Intell Syst Technol*. 2019;10(2):1–19.
108. Finlayson SG, Chung HW, Kohane IS, Beam AL. AI model drift and the need for continuous monitoring. *JAMA*. 2021;326(23):2379–80.
109. Roberts M, Driggs D, Thorpe M, et al. Common pitfalls and best practices in AI for healthcare. *Nat Mach Intell*. 2021;3:802–20.
110. Davis J, Goadrich M. The relationship between precision-recall and ROC curves. *ICML*. 2006;233–

- 40.
111. Collins GS, Dhiman P, Andaur Navarro CL, et al. TRIPOD-AI: Reporting guidelines for AI prediction models. *BMJ*. 2021;372:n71.
112. Sendak MP, Ratliff W, Sarro D, et al. Standardizing AI implementation in clinical settings. *BMJ Health Care Inform*. 2020;27:e100109.
113. Liu X, Rivera SC, Moher D, et al. Reporting standards for AI clinical trials: CONSORT-AI. *Nat Med*. 2020;26:1364–74.
114. Price WN, Gerke S, Cohen IG. Legal liability for AI-assisted medical decisions. *Nat Med*. 2020;26:131–3.
115. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—ethical implications. *N Engl J Med*. 2018;378(11):981–3.
116. Goodman B, Flaxman S. EU regulations on algorithmic decision-making and accountability. *AI Mag*. 2017;38(3):50–7.
117. Leslie D. Human oversight and accountability in AI systems. *OECD Working Papers*. 2022;No. 6.
118. European Parliament. AI Liability Directive Proposal. EU Commission; 2023.
119. Zhang Z, Chen P, McGough M, et al. Multimodal learning in medical AI. *Nat Biomed Eng*. 2022;6(3):233–45.
120. Fu Y, Jung H, Xu Y, et al. Integrating histopathology and genomics for cancer prediction. *Nat Med*. 2020;26:164–72.
121. Rajpurkar P, Chen E, Banerjee O, et al. Multimodal clinical deep learning applications. *Lancet Digit Health*. 2022;4(7):e487–e497.
122. Topol EJ. Precision medicine powered by AI. *Cell*. 2019;176(1):1–3.
123. Lee C, Yoon J, Sharma A, et al. Data integration and privacy challenges in multimodal AI. *NPJ Digit Med*. 2023;6:12.
124. Yang Q, Liu Y, Chen T, Tong Y. Federated learning: Concept and applications. *ACM TIST*. 2019;10(2):1–19.
125. Rieke N, Hancox J, Li W, et al. Federated learning for medical imaging under privacy constraints. *Nat Mach Intell*. 2020;2(6):475–84.
126. Sheller MJ, Edwards B, Reina GA, et al. Multi-institutional federated learning for medical imaging. *Sci Rep*. 2020;10:12598.
127. Dayan I, Roth HR, Zhong A, et al. Federated learning improves clinical generalization. *Nat Med*. 2021;27:1735–43.
128. Kaissis G, Ziller A, Passerat-Palmbach J, et al. Privacy-preserving machine learning in healthcare. *Nat Commun*. 2021;12:195.
129. Khera AV, Chaffin M, Aragam KG, et al. Polygenic risk scores and personalized disease prevention. *Nat Genet*. 2018;50:1219–24.
130. Yala A, Mikhael PG, Strand F, et al. AI-based breast cancer risk prediction. *Radiology*. 2021;299(1):39–46.
131. Hippisley-Cox J, Coupland C. Machine-learning-based disease risk stratification. *BMJ*. 2022;376:e068720.
132. Ting DSW, Cheung CY, Lim G, et al. AI-assisted triage in clinical screening. *Nat Med*. 2019;25:115–22.
133. Vayena E, Blasimme A, Cohen IG. Ethical challenges in personalized AI screening. *Lancet Digit Health*. 2022;4(5):e342–e348.