ROYAL SOCIETY OF CHEMISTRY

CRITICAL REVIEW

Check for updates

Artificial intelligence (AI) in healthcare diagnosis: evidence-based recent advances and clinical implications

Jay Bhatt,† Sweny Jain† and Dhiraj Devidas Bhatia (iD) *

Artificial intelligence (AI) is increasingly shaping modern healthcare by improving the accuracy and efficiency of disease diagnosis. This review summarises the modern advancements in AI-driven diagnostic technologies, with a focus on machine learning (ML) and deep learning (DL) applications for the detection and characterization of cancer, cardiovascular diseases, diabetes, neurodegenerative disorders, and bone diseases. AI models, particularly those employing convolutional neural networks, have demonstrated expert-level performances in interpreting medical images, genomic profiles, and electronic health records, often surpassing traditional diagnostic methods in terms of sensitivity, specificity, and overall accuracy. Using advanced methods like machine learning and deep learning, AI systems can analyze large and complex medical datasets—including images, electronic health records, and laboratory results—to detect patterns linked to various diseases. While integration of AI into clinical practice has shown significant benefits, challenges remain in ensuring the reliability, interpretability, and broad adoption of these systems. Thus, continued research and careful implementation are needed to maximize the potential of AI in transforming diagnostic processes and improving patient outcomes.

1. Introduction

Artificial intelligence (AI) has become a groundbreaking technology in medical diagnostics, providing enhanced analysis of multifaceted clinical data and assisting in precise and effective decision-making for the diagnosis of various diseases. Artificial intelligence (AI) has progressively become a more useful and reliable tool for multiple applications, particularly in healthcare. By facilitating improved efficiency and organization, it has the potential to enhance clinical practice, thus improving patient care and outcomes. John McCarthy defined the term artificial intelligence (AI) as "the science and engineering of making intelligent machines". AI started as a simple series of "if, then rules" and has expanded and developed over the years to include more complex algorithms that achieve a similar performance to the human brain.[1] In the field of healthcare, the diagnosis of a disease plays an important part, where any source or circumstance that leads to pain, dysfunction, illness or, eventually, the death of a human being is called a disease. Disease diagnosis could be easy or tricky and complicated depending on the area of disease. For

numerous diseases, conventional diagnostic methods are often manual and susceptible to human error. Incorporating AI enables automated diagnosis, potentially improving accuracy and minimizing mistakes when compared with traditional human-based assessments. It has gained attention due to its low cost, minimal need for manpower, and limited infrastructure and equipment requirements. Although extensive datasets are available, there is a lack of effective tools capable of accurately identifying patterns and making reliable predictions.[2] Advanced algorithms are used by AI in a number of healthcare domains, such as diagnosis support, treatment planning, patient profiling, and disease prediction.[3]

Artificial Intelligence (AI) is basically a computational system designed to carry out tasks that require human intelligence, such as reasoning, learning, and decision-making, often autonomously, while machine learning (ML) is a subfield of AI, where algorithms learn patterns from data to make predictions without any explicit programming for each task. Deep learning (DL) is a branch of ML using neural networks with multiple layers to model complex data patterns, which is particularly effective for large-scale and high-dimensional datasets such as medical images. ML requires preprocessing of the input data to determine results and avoid false predictions, whereas DL requires dealing with large data sets and development of a deep data structure with

Department of Biological Sciences and Engineering, Indian Institute of Technology Gandhinagar, Palaj, Gujarat, 382355, India. E-mail: dhiraj.bhatia@iitgn.ac.in
† Both the authors have contributed equally.

multiple processing layers. Similar to the neural networks and connections in the human brain, "deep learning" is skilled to automatically mine and learn features from "big data" of healthcare (*i.e.*, genetics, imaging, healthcare records, and most "-omics" data) through the use of a multi-layered architecture known as convolutional neural networks (CNNs).[4] CNNs are a type of deep learning algorithm used in image processing, which have been designed to replicate the functioning of biological neural connections in the brain. A CNN consists of several layers that examine an input image to identify patterns and create specific filters.[15] Deep convolutional neural networks (CNNs) show potential for highly variable tasks across many object categories.[6–11] AI analyzes huge datasets and recognises patterns that would be tough for humans to detect, which has led to advances in various horizons of healthcare.[12–14] In CNN for diagnosis, the input data is usually multiple images for the characterization of specific diseases such as MRI, endoscopy, and sonography. The final results are produced by the combination of all features by the fully connected layers. Fig. 1 illustrates the key advancements in field of

AI and their roles. These advancements play a significant role in the development of a model/system for the diagnosis of specific diseases.

One of the pioneering AI-based detection systems was created using patient data provided by physicians and built on a knowledge base consisting of approximately 600 rules, *i.e.*, the "backward chaining" AI system called MYCIN designed in the early 1970s.[15,16] MYCIN generated a list of potential pathogens and recommended antibiotic treatments tailored to the patient's body weight. Its rule-based framework later served as the foundation for developing other systems, such as EMYCIN. INTERNIST-1, a larger medical base to help primary care physicians in diagnosis, was later developed using the same framework as EMYCIN.[1,17] Various AI techniques, which include machine learning and deep learning, are widely used in healthcare for tasks such as disease diagnosis, drug discovery, and identifying patient risk. To achieve accurate disease diagnosis using AI, a variety of medical data sources are essential, including ultrasound, magnetic resonance imaging (MRI), mammography, genomics, and computed tomography (CT) scans.

Convolutional Neural Network (CNN)	• A deep learning architecture designed for spatial data analysis, using convolutional layers to detect hierarchical features, especially effective in image-based diagnostics.
Large Language Model (LLM)	• Advanced AI models trained on extensive text corpora to understand, generate, and manipulate human language, supporting tasks like summarization and report generation.
Generative AI	• AI systems developed for generating new content, such as text or images, by learning the underlying data distribution, including models like GANs and LLMs.
Natural Language Processing (NLP)	• AI techniques for automated understanding and generation of human language, applied to clinical notes and medical literature analysis.
Medical Imaging AI	• Application of AI algorithms to analyze and interpret medical images for diagnostic support, often achieving high accuracy in tasks like abnormality detection.
Multi-modal Learning	• An approach that integrates multiple data types (e.g., images, text, lab results) to enhance analysis and diagnostic accuracy.
Agentic AI	• Autonomous AI systems capable of independent data interpretation and decision-making, reducing the need for direct human oversight in clinical workflows.
Artificial General Intelligence (AGI)	• Hypothetical AI with the ability to comprehend and perform a various types of tasks at or above human-level competence.
Artificial Immune System (AIS)	• Computational models inspired by biological immune mechanisms, used for anomaly detection and pattern recognition in biomedical informatics.

Fig. 1 Overview of key AI advancements in healthcare, including models and techniques like CNNs, LLMs, NLP, and medical imaging using AI, with their core applications and roles in diagnostics and clinical workflows.

2. Framework for AI in disease detection modelling

2.1 Building prediction model

Artificial intelligence (AI) refers to the ability of machines to learn in ways like humans, such as recognizing images and patterns in complex situations. In healthcare, AI is transforming how patient data is collected, processed, analysed, and used to improve care.[18]

System design is the core conceptual structure of any system. It includes how the system is organized, how it operates, and how it responds under different conditions. Understanding system design helps users recognize its capabilities and limitations. Before applying any algorithm, real-world data must go through a preparation process to ensure its quality.[19] This is necessary because real-world data often contains errors that must be addressed. Data pre-processing involves cleaning, correcting, and organizing raw data for better analysis. The data pre-processing includes several steps. In data cleaning, techniques are used to fill missing values or remove unwanted symbols. In data integration, information is combined from various sources and corrected for errors before use.[20,21] Data transformation involves adjusting the data format or scale based on the algorithm requirements. Normalization is often used here to make the data consistent.[22] This step is crucial for many data mining techniques. After cleaning and transforming, the data is refined and optimized, along with the symptoms of the patient. Data reduction aims to shrink the dataset to a manageable size without losing valuable information. Once prepared, the dataset is divided into training and testing sets. The training data is used to identify patterns, while the testing data evaluates the performance of the model, as shown in Fig. 2.[23] These sets usually come from the same dataset. After data preparation, the next step is to assess the accuracy of the model. Analytical models are then applied to evaluate the likelihood of specific outcomes based on input factors. These models are useful for predicting diseases by analysing symptoms and past medical history.[24,25]

2.2 Neural network for detection

A neural network functions by processing complex data patterns, such as medical images and patient records, through interconnected layers of artificial neurons. In disease diagnosis, this enables the network to automatically extract meaningful features and learn associations between input data (e.g., X-rays, MRIs, and lab results) and disease states, often detecting subtle abnormalities that may be missed by traditional methods (Fig. 3).

A neural network for disease diagnosis is structured in layers. The initial input layers analyze basic features from medical images or data, such as edges and simple shapes, and the results are called feature maps, which represent where and how strongly certain features appear in the image. As the data moves through deeper layers, the network detects more complex patterns and combinations of features. The pooling layer decreases the size of these feature maps by
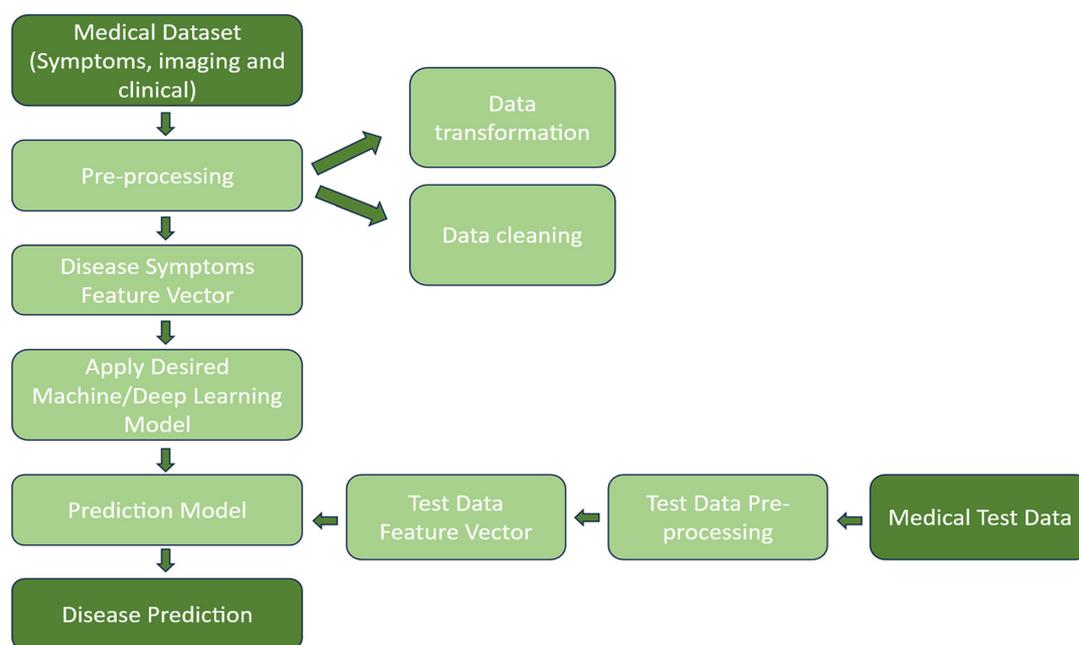
Fig. 2 Framework to understand the development of diagnostic systems/models. Visual model for disease detection using machine learning and deep learning methods, which shows the framework of the disease detection system. Before training the model, the data should be pre-processed and filtered accordingly with symptoms. After that, model training begins to evaluate the test data. The test data also must be pre-processed and filtered before the evaluation. Once the model starts prediction, the results are analyzed and compared with the validated data to determine the sensitivity, specificity, and other evaluation parameters, which are discussed in the upcoming sections.
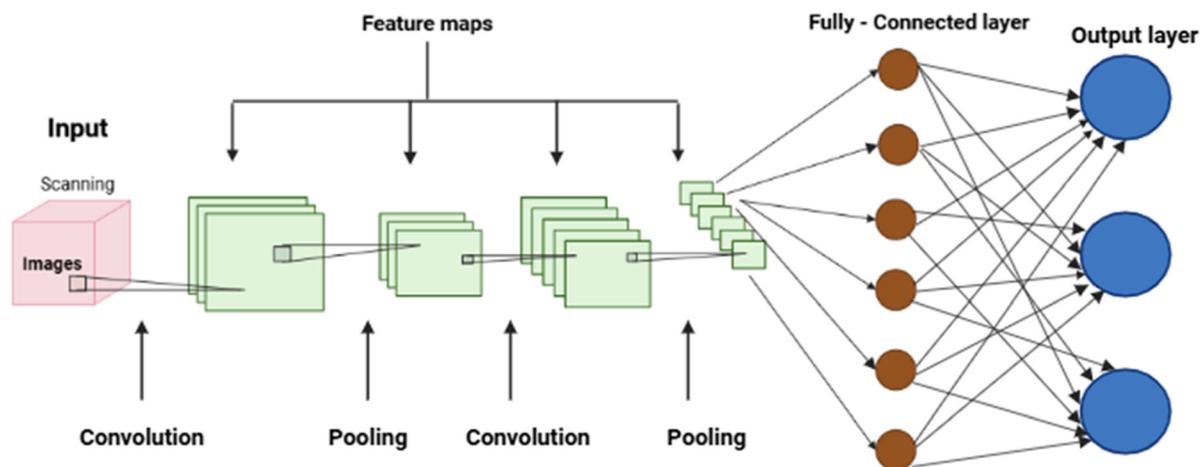
Fig. 3 Basic functioning of neural networks for disease diagnosis.

summarizing the information in small regions. The final layer interprets these patterns to help determine if a disease is present and, if so, which one. This layered approach enables the network to process and interpret medical information efficiently, supporting accurate and rapid diagnosis.

2.3 Model evaluation parameters

For every diagnostic system, there are certain parameters commonly used in the evaluation and interpretation of artificial intelligence (AI) models for medical diagnosis, enabling readers to better understand the performance and reliability of the presented diagnostic system. Clarifying these concepts is essential for transparently communicating the strengths, limitations, and clinical relevance of AI-based

diagnostic tools. They are used to explain the results of their scientific findings, along with their explanation and examples (Table 1).

Apart from them, some other parameters that have been used in neural network architectures for medical diagnosis are convolutional layers, which extract features from imaging data using filters, and activation functions such as ReLU (rectified linear unit) that introduce non-linearity by outputting positive values unchanged and zeroing negatives. Pooling layers, such as average pooling, help shrink the spatial size of feature maps, while conserving significant information. Dropout layers help prevent overfitting by randomly disengaging neurons during training. Collectively, these metrics and neural network components enable systems such as GRAIDS and TumorDetNet to achieve expert-level diagnostic accuracy, streamline clinical workflows, and

Table 1 Common metrics in AI diagnostic results

Term	Explanation	Example
Sensitivity (true positive rate)	Calculates the percentage of real positives that the model correctly identified. High sensitivity reduces the chance of missed diagnoses	8428 endoscopic images achieved sensitivity of 98%, which means 98% of people with the disease will be correctly identified as positive[26]
Specificity (true negative rate)	Calculates the percentage of real negatives that were correctly identified. High specificity reduces false positives	OsteoSight achieved specificity of 0.852 for detection, which means it is 85.2% accurate at identifying healthy individuals[27]
Accuracy	Percentage of accurate predictions the model made, including both true positives and true negatives	Random forest achieves 93% accuracy in predicting stroke recurrence risk, which means 93% of all test results (both positive and negative) are correct[28]
AUC-ROC (area under the receiver operating characteristic curve)	Checks model performance across all classification thresholds; the higher the value, the better the distinction between classes	X-rays, achieved an AUROC of 0.834 (95% CI: 0.789–0.880) compared to DXA[27]
Positive predictive value (PPV)	The likelihood that individuals who test positive actually have the disease	PPV of 0.814 for GRAIDS *vs.* 0.974 for competent endoscopists, stating 81.4% chance that the person truly has the disease[29]
Negative predictive value (NPV)	The likelihood that individuals who test negative actually do not have the disease	GRAIDS achieved a high negative predictive value (0.978), comparable to experts (0.980), meaning there is 97.8% chance the person truly does not have the disease[29]
Cohen's kappa	A statistical measure of inter-rater agreement adjusted for chance; values >0.6 indicate moderate to substantial agreement	Cohen's kappa of 0.62, comparable to the range of 23 expert uropathologists of (0.60–0.73), where <0: poor, 0.01–0.20: slight, 0.21–0.40: fair, 0.41–0.60: moderate, 0.61–0.80: substantial, 0.81–1: almost perfect[30]

enhance outcomes in oncology. Understanding these terms will help in understanding the results and conclusions drawn by researchers in their diagnostic studies.

3. Diseases diagnosis

This section presents recent and notable advancements in the diagnosis of several major diseases through machine and deep learning approaches. It explores the prediction methods and their outcomes for conditions such as cancer, diabetes, heart disease, neurodegenerative disease, and bone disease detailing the diagnostic methods enabled by these technologies.

3.1 Diagnosis of cancer

Cancer is one of the most complicated diseases humankind has faced. Recent developments in ML and DL have taught AI systems to analyse medical images, genomic data, clinical reports and electronic health records with high sensitivity and specificity. Early diagnosis can prevent the further development of cancerous cells if proper treatment is provided on time. Among researchers, the diagnosis of cancer has been one of the biggest obstacles. There are several AI models that help in the prediction or recurrence of cancer. One of them is METACANS, a multimodal artificial intelligence (AI) model that integrates whole screen images with clinicopathological features to predict axillary lymph node (ALN) metastasis. METACANS was developed using data from 1991 cases and validated externally on five different cohorts comprising 2166 cases. It recorded an area under the curve (AUC) of 0.733 (95% CI, 0.711–0.755), negative predictive value of 0.846, sensitivity of 0.820, specificity of 0.504, and balanced accuracy of 0.662. Even without extra labelling or annotation, METACANS can detect features such as micropapillary growth, infiltrative patterns, and tissue necrosis through pathological imaging linked to metastatic behaviour. This highlights its potential in supporting preoperative axillary evaluation in breast cancer.[30] Apart from breast cancer and its metastasis, another study leveraged deep neural networks to automate the detection and grading of prostate cancer in needle biopsy samples of a group of people with the same characteristics. This was a population-based STHLM3 diagnostic study of 6682 slides from 976 participants and external validation datasets. The AI system demonstrated an exceptional distinguishing performance, achieving an AUC-ROC of 0.997 (95% CI 0.994–0.999) for identifying benign from malignant cores in internal testing and 0.986 (0.972–0.996) on external validation. Tumor extent predictions showed strong correlation with pathologist measurements ($r = 0.96$ internal, $r = 0.87$ external) and or Gleason grading, and AI achieved a mean pairwise Cohen's kappa of 0.62, which is comparable to the range of 23 expert uropathologists (0.60–0.73). These results suggest that the system could reduce pathology workloads and provide expert-level grading consistency.[31,32]

A wide range of images can be used to train a model for diagnosis, for example, using 1 036 496 endoscopy images from 84 424 individuals across six Chinese hospitals, the gastrointestinal artificial intelligence diagnostic system (GRAIDS) was developed and validated for detecting upper gastrointestinal cancers. GRAIDS showed high diagnostic accuracy across validation sets, as follows: 0.955 (95% CI 0.952–0.957) in internal validation, 0.927 (0.925–0.929) in prospective testing, and 0.915–0.977 in external validation. Compared to experts and endoscopists, GRAIDS matched the expert-level sensitivity (0.942 *vs.* 0.945; $p = 0.692$) and beat competent (0.858; $p < 0.0001$) and trainee (0.722; $p < 0.0001$) endoscopists. Although its positive predictive value (0.814) lagged competent endoscopists (0.974), GRAIDS achieved a high NPV (0.978), which is comparable to experts (0.980).[29] To determine whether the tumour is malignant or benign, a system was developed to detect upper gastrointestinal subepithelial lesions (SELs) on endoscopic ultrasonography (EUS) images, prioritizing the differentiation of gastrointestinal stromal tumors (GISTs) from benign lesions. They trained the model with 16 110 EUS images from 631 pathologically confirmed cases. The system achieved an accuracy of 86.1%, outperforming all endoscopists. In distinguishing GISTs from non-GISTs, it demonstrated 98.8% sensitivity, 67.6% specificity, and 89.3% accuracy, surpassing everyone.[33] These models can help physicians and healthcare providers in treatment.

Apart from that, a system was trained to detect tumor-infiltrating lymphocytes (TILs) in testicular germ cell tumors for prognosis. Manual TIL annotations from 259 regions across 28 hematoxylin–eosin-stained whole-slide images (WSIs) were used to train the algorithm, and then it was tested, which was subsequently applied to WSIs from 89 patients. This AI showed potential for better early prognosis.[34] The use of transfer learning and deep learning in an IoT (internet of things) system has helped to assist doctors in the diagnosis of melanoma or skin cancer. The CNN models used included neural architecture search network (NASNet), dense convolutional network (DenseNet), visual geometry group (VGG), MobileNet, inception, residual networks (ResNet), inception-ResNet, extreme inception (Xception), the Bayes, random forest (RF), support vector machines (SVM), K-nearest neighbors (KNN) and perceptron multilayer (MLP) for the classification of injuries. Among the combinations, the DenseNet201 extraction model, combined with the KNN classifier achieved an accuracy of 96.805% for the ISBI-ISIC (examined lesions between nevi and melanomas) dataset and 93.167% for the PH2 (based on lesions of common nevus, atypical nevi, and melanomas).[35] An AI model was developed using a Single Shot Multibox Detector (SSD) and 523 OSCC images and tested on 66 OSCC, 49 leukoplakia, and 405 other oral disease images to detect oral squamous cell carcinoma (OSCC) and dysplastic leukoplakia from 1043 clinical images captured with a single-lens reflex camera. For OSCC-only detection, it achieved 93.9% sensitivity, 81.2% specificity, and 98.8% negative

predictive value (NPV). When simultaneously detecting both OSCC and leukoplakia, its performance remained robust with 83.7% sensitivity, 94.5% NPV, and maintained 81.2% specificity. The proposed method could diagnose oral cancer and leukoplakia with higher accuracy and reliability and can help in avoiding unnecessary biopsies (Fig. 4).[36]

3.2 Diagnosis of diabetes

Diabetes is a chronic disease where the body cannot properly regulate blood sugar levels either due to insufficient insulin production or the inability of the body to use insulin effectively. Currently, diabetes is most often diagnosed using a blood test; however, the early diagnosis of diabetes is crucial and it significantly reduces the risk of severe complications such as heart disease, kidney failure, nerve damage, and vision loss, and can even lower rates of hospitalization and death. One study explored how nonlinear heart rate variability (HRV) parameters can be used to predict diabetes employing artificial neural networks (ANN) and support vector machines (SVM). Electrocardiogram (ECG) signals from two groups of male Wistar rats, *i.e.*, healthy controls and those made diabetic through streptozotocin administration, were collected by researchers. Each group had five rats, aged between 10 to 12 weeks and weighing around 200 grams. According to the ECG recordings, a total of 526 data samples was generated, and thirteen nonlinear

HRV features were extracted from them. Then, these features were used to train and test an ANN model with a structure of 13 input neurons, 7 hidden neurons, and 1 output neuron. They achieved a classification accuracy of 86.3% at a learning rate of 0.01. When the same features were applied to an SVM classifier, it improved with accuracy of 90.5%. The findings indicate that diabetes causes noticeable changes in nonlinear HRV patterns, and it can help us in early prognosis.[37]

Another approach for classifying heart rate variability (HRV) signals to distinguish between diabetic and healthy individuals was done using a deep learning-based system. Utilizing models such as long short-term memory (LSTM), CNN with a hybrid CNN-LSTM architecture, this method was developed to capture the complex temporal patterns present in HRV data. After feature extraction, these representations are provided into a support vector machine (SVM) for the final classification task. The integration of SVM led to slight performance improvements of 0.03% for CNN and 0.06% for the CNN-LSTM combination compared to previous models without SVM. The system achieved a high classification accuracy of 95.7%, indicating its strong potential as a diagnostic aid for detecting diabetes using ECG-derived HRV signals.[38] To predict the type 2 diabetes mellitus (T2DM) risk, a system was developed by a machine learning framework using six classification algorithms. Evaluated on two datasets, a custom 18-question survey dataset and the standard PIMA Indian Diabetes Database, the models were compared using
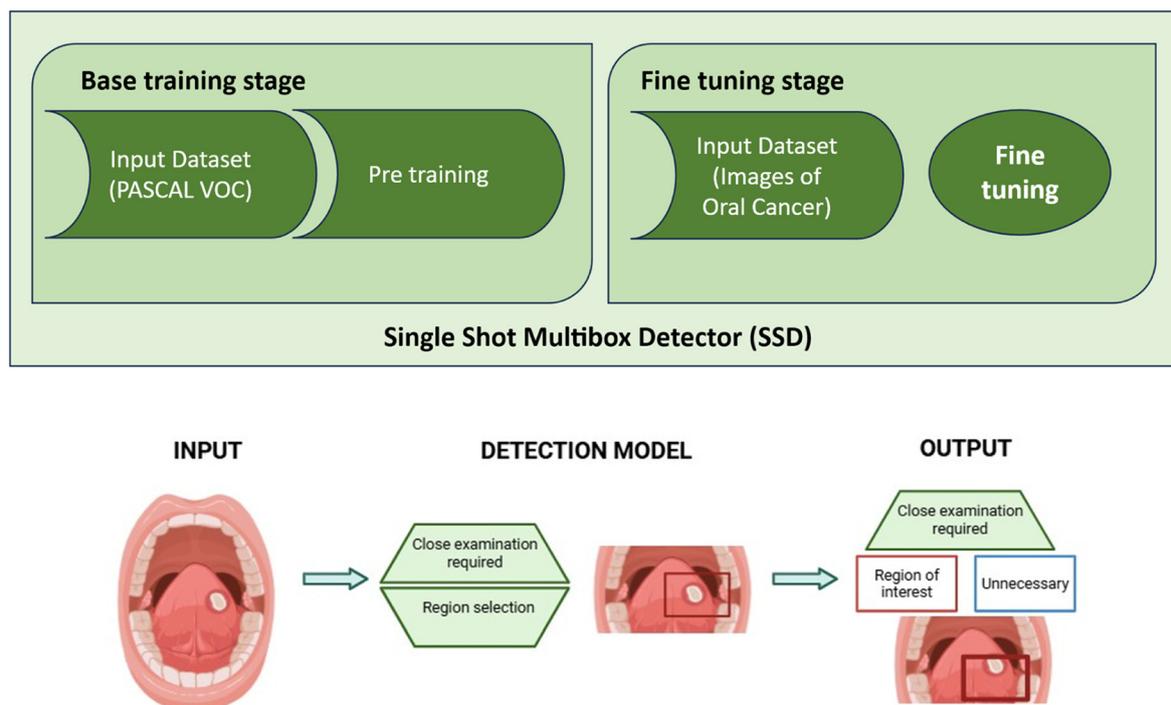
Fig. 4 SSID detector for the diagnosis of oral cancer. It outlines the detection model construction. The deep neural network is pretrained using the large-scale PASCAL-VOC 20128 dataset. Later, the model is fine-tuned with images of oral cancer, and the model is constructed using a deep learning method. The model is trained to detect targets of regions of oral cancer and leukoplakia in each image (red frame) drawn by an oral surgeon. The red boxes in the images represent the location of the lesions as annotated by an oncologist. This model can be used to identify the presence and location of oral diseases that require close examination by inputting the oral image.

accuracy, precision, recall, and sensitivity. Random forest, an ensemble learning method that generates multiple decision trees, and merging their outputs to improve the prediction accuracy and reduce overfitting, making it especially effective for complex classification problems such as diabetes risk prediction outperformed others. It achieved 94.10% accuracy on the custom dataset and the highest accuracy on the PIMA dataset.[39]

3.3 Diagnosis of heart-related diseases

Heart diseases are conditions that affect the heart and blood vessels, including problems such as blockage in arteries, irregular rhythm of heartbeats, and weakness in heart muscle. These diseases are the leading cause of death worldwide and can develop silently over time or present suddenly with symptoms such as chest pain and shortness of breath. Currently, heart-related diseases are mainly diagnosed using tests such as electrocardiograms (ECG/EKG), echocardiograms, and blood tests to check for markers of heart damage or risk factors. With a sensitivity of 94.2%, accuracy of 99.37% and specificity of 99.66%, a 1D-convolutional neural network model (1D-CNN) was developed for the detection of arrhythmia, a heart disease. This AI is fast, accurate and simple to use.[40] To understand and diagnose hypertension, a study analyzed health examination data (2005–2016) from 18 258 individuals, focusing on health parameters recorded at hypertension diagnosis [Year (0)] and during two preceding annual visits [Year (−1) and Year (−2)]. Machine learning models (XGBoost, ensemble) and traditional logistic regression were applied to predict the outcomes. XGBoost is an advanced gradient boosting algorithm designed for speed and performance, while ensemble methods combine multiple models to improve the predictive accuracy and robustness. The data were randomly partitioned into a derivation cohort (75%, $n = 13\,694$) for model training and a validation cohort (25%, $n = 4564$) for performance evaluation. This ML-based approach achieved a strong predictive performance, with AUC-ROC scores of 0.877 (XGBoost) and 0.881 (ensemble model) in validation. These results surpassed traditional statistical methods such as logistic regression, highlighting the advantage of advanced algorithms in identifying subtle risk patterns.[41]

Between 2008 and 2016, researchers compared 709 patients with idiopathic pulmonary arterial hypertension (IPAH) to over 2.8 million similar patients without IPAH. They built a prediction model using information such as how often patients saw specialists, other diagnoses, and age. When tested, this model was very specific (99.99%) but not highly sensitive (14.10%), meaning it rarely detects people incorrectly but missed many true cases.[42] An AI system was developed with the aim to reduce stroke recurrence risk through a real-time patient monitoring framework. By tracking critical biomarkers (e.g., blood pressure, heart rate variability, and neurological indicators) via wearable sensors, the system employs machine learning classification algorithms to detect deviations from baseline parameters and provide automated alerts. Tree-based ensemble methods such as random forest achieved 93% accuracy in predicting risk thresholds.[28] Nowadays, there are many wearable technologies that monitor heart rate, pulse rate, and blood oxygen level using AI. Many AI systems developed for heart-related diagnosis can be very efficient given that these diseases have silent symptoms.

3.4 Diagnosis of bone-related diseases

Bone diseases are conditions that weaken the structure or function of bones, such as osteoporosis, osteoarthritis, Paget's disease, and osteogenesis imperfecta, making them more prone to pain, deformity, and fractures. These disorders can result from genetic factors, injury, poor nutrition, aging, or underlying medical conditions and often progress silently until considerable damage occurs. Thus, the early diagnosis of bone diseases is essential because it improve treatment outcomes, maintain mobility and independence and can prevent further damage. One recent study demonstrated that OsteoSight™, an AI tool for detecting low bone mineral density (BMD) from routine hip/pelvic X-rays, achieved an AUROC of 0.834 (95% CI: 0.789–0.880) compared to DXA, with the specificity of 0.852 (minimizing false positives) and sensitivity of 0.628 (moderate true positive detection). These results are significant given that OsteoSight™ can address critical gaps in osteoporosis screening by enabling opportunistic detection during routine imaging. Early identification of at-risk patients may reduce fracture-related morbidity and healthcare costs.[27] One of the studies explored the application of deep learning to orthopedic radiographs. They used 256 000 wrist, hand, and ankle images from a hospital database. Later, they categorized images into four classes of fractures, laterality (left/right), body part, and exam view. Five publicly available deep learning models were adapted and trained on this dataset, with the highest-performing network evaluated against a gold standard for fracture detection, and the performance of the model was compared to assessments by two experienced orthopedic surgeons, who analyzed the images at the same resolution as the algorithm. The results showed that all the networks achieved ≥90% accuracy in recognizing laterality, body part, and exam view. The top model demonstrated 83% accuracy in fracture detection, matching the diagnostic performance of experienced surgeons under standardized conditions. The interobserver agreement between surgeons, measured via Cohen's kappa, was 0.76, indicating substantial consistency. This system outperformed the specialist, proving to be an efficient and reliable detection system.[43]

A team of researchers developed a high-throughput bone-on-a-chip system that replicates the natural bone environment for evaluating osteoporosis treatments. This platform integrates mouse osteocytes and osteoblasts co-cultured within a 3D osteoblast-derived decellularized extracellular matrix (OB-dECM), offering a more biologically

accurate model than standard collagen-based systems. The engineered microenvironment significantly improved the cell survival, osteocyte development, and expression of osteogenic markers. To validate its utility, the system was used to assess the therapeutic effect of an anti-SOST antibody, a drug relevant to osteoporosis treatment, by measuring β-catenin nuclear translocation in osteoblasts. The results showed a notable increase in both β-catenin intensity and nuclear presence in the treated samples.[44]

3.5 Diagnosis of neurodegenerative diseases

Neurodegenerative diseases are conditions where nerve cells in the brain or nervous system gradually lose function and die, leading to problems with movement, memory, or thinking. These diseases, such as Alzheimer's, Parkinson's, Huntington's, and ALS, typically worsen over time and are more common as people age. Currently, neurodegenerative diseases are diagnosed using a combination of clinical assessments, neuropsychological tests, and advanced tests such as brain imaging (MRI and PET) or analysis of biomarkers in blood or cerebrospinal fluid. Early diagnosis is crucial because it enables timely interventions that can slow disease progression, help maintain independence longer, and improve the quality of life. In the case of Alzheimer's disease, a study evaluated various automated classification techniques for distinguishing Alzheimer's disease (AD) patients from healthy individuals using FDG-PET imaging. The methods assessed include the general linear model, scaled subprofile modeling, and support vector machines (SVM). Among them, the SVM combined with the Iterative Single Data Algorithm achieved the highest performance, with a sensitivity of 0.84 and specificity of 0.95, validated through 10-fold cross-validation.[45] One study explored the electroencephalography (qEEG) as a biomarker for detecting functional brain changes associated with Huntington's disease (HD), even before noticeable motor or cognitive symptoms arise. Researchers aimed to automatically differentiate between HD gene carriers and healthy individuals using qEEG data and identify EEG features that align with clinical indicators of disease progression. The study involved 26 individuals carrying the HD gene (average age 49.7 years) and 25 healthy controls (average age 52.7 years), with EEG signals recorded for three minutes while subjects were at rest. A statistical pattern recognition approach was applied to a wide range of EEG features to create an EEG-based classification index, which was validated using 10-fold cross-validation. This index ranged from 0 (normal) to 1 (indicative of HD), providing a continuous measure of the disease state. The classifier achieved 83% sensitivity, 83% specificity, and 83% overall accuracy, with an area under the ROC curve (AUC) of 0.9. Employing AI and qEEG data, this can serve as a useful, non-invasive biomarker for HD diagnosis and monitoring.[46] Diffusion tensor imaging (DTI) and diffusion kurtosis imaging (DKI) are advanced MRI techniques used to identify microstructural brain changes in conditions such as Alzheimer's disease. A group of researchers investigated whether combining diffusivity and kurtosis parameters enhances the ability of DKI to detect Alzheimer's-related abnormalities more effectively than using either measure alone. Using SVM, rigorous validation was performed to ensure classifier reliability. Using the optimized classifiers, brain abnormalities were identified in a sample of 53 individuals, including 27 diagnosed with Alzheimer's disease. The combined approach achieved a high classification accuracy of 96.23% and was more effective in identifying abnormal brain regions than using diffusivity or kurtosis alone despite the complementary nature of the two measures.[47]

In patients with amyotrophic lateral sclerosis (ALS), a system was developed using clinical data with MRI-based imaging through deep learning. A total of 135 ALS patients was included, all of whom had MRI scans during their initial outpatient visit. The patients were then closely followed over time, and their survival durations were recorded. Based on the survival time from the disease onset, participants were categorized into short, medium, or long survival groups. For the deep learning analysis, the dataset was divided into training (83 patients), validation (20 patients), and testing (32 patients) sets. The models trained solely on clinical features achieved a prediction accuracy of 68.8%, while models using MRI-based structural connectivity or brain morphology each reached 62.5%. Importantly, when all three data sources, *i.e.*, clinical features, structural connectivity, and brain morphology, were combined, the prediction accuracy significantly improved to 84.4%. These results highlight that the predictive value of MRI data in ALS prognosis can be efficient.[48] Although AI can detect based on smile and facial features, a recent study demonstrated that combining facial and speech features collected during natural conversations with a chatbot, analyzed using machine learning, can accurately distinguish individuals with Alzheimer's disease or mild cognitive impairment from healthy controls; using 8 facial and 21 sound features, this system achieved a high diagnostic accuracy (AUC = 0.94).[49]

4. Comparative analysis

The comparative analysis illustrated in Table 2 shows detailed information such as type of AI tool, disease type, sample type, and the results of the work done by researchers on different diseases.

5. Conclusions

Accuracy is vital in diagnosing diseases, given that it plays a key role in treatment planning and ensuring patient health and safety. Artificial intelligence (AI) is a broad and evolving field made up of data, algorithms, deep learning, neural networks, and analytical tools, which continues to adapt to the growing needs of healthcare. A sew studies highlight the importance of AI in identifying various illnesses and

Table 2 Comparative study for multiple disease detection

Sr. no.	AI tool	Disease	Sample type	Result	References
1	DNNs	Prostate cancer	Slides from needle core biopsies	AUC = 0.997 (95% CI 0.994–0.999) on independent test dataset AUC = 0.986 (0.972–0.996) on external validation	50
2	Deep convolutional neural network (CNN)	Periodontal disease	Panoramic dental radiographs	73.04% in detecting an alveolar bone loss total accuracy for the multi-classification was 59.42%	51
3	PyCaret 2.3.10	Headache disorders	Patients' questionnaire sheets	Micro-average accuracy, sensitivity, specificity, precision, and F-values for the test dataset were 93.7%, 84.2%, 84.2%, 96.1%, and 84.2%, respectively	52
4	EALAI-CFDNBD and BOOA	Cognitive fatigue	Neurophysiological bio signal data	Accuracy value of 97.59%	53
5	LIDGAX	Xanthogranulomatous cholecystitis (XGC) and gallbladder cancer (GBC)	Clinical, imaging, and laboratory data	AUC 0.95 and accuracy 0.92	54
6	Deep neural network	Neurodegenerative disease	Alpha-synuclein (aSyn) oligomers and fibrils at various known ratios using immunoassay-coupled nanoplasmonic infrared metasurface sensor	Overall accuracy score of 94.66%	55
7	Ensemble technique	Type-II diabetes	Questionnaire based data collection	Accuracy of 97.34%	56
8	Single shot multibox detector (SSD)-deep learning	Oral cancer and dysplastic leukoplakia	Oral lesion images	Sensitivity of 93.9% versus 83.7%, a negative predictive value of 98.8% versus 94.5%, and a specificity of 81.2% versus 81.2%	36
9	JustNN tool-C 4.5, random forest, CART, random tree, and REP tree classification method	Liver disease	Indian liver patient dataset-583 instances based on ten different biological parameters	99% accuracy	57
10	Decision tree, random forest, classification, regression tree	Thyroid disease	Thyroid disease dataset	Decision tree: 98%, random forest: 99%	58
11	Artificial neural network (ANN)	Oral cancer (OC)	SERS spectra of exhaled breath	Accuracy of 99% (AUC) of 0.996	59

demonstrate how machine learning and deep learning are applied in diagnosing different diseases, while others take a closer look at how AI is being used to support the diagnosis of wide range of health conditions, ranging from Alzheimer's and cancer to diabetes, heart disease, stroke, and skin and liver disorders. By reviewing different techniques and real-world applications, we tried to understand not just how these systems work, but where they show the most promise. Along the way, we highlighted some of the common challenges in diagnosing these diseases and how AI might help overcome them. We also compared various approaches using performance measures such as accuracy, sensitivity, specificity, AUC, and F-score to get a clearer picture of what is working best. In the end, this study points to a future where AI can become a valuable partner in healthcare, not replacing doctors, but helping them make faster, more accurate decisions, especially in complex or early-stage cases. Despite the progress made, accurate medical diagnosis still faces challenges. These issues must be addressed to keep up with the discovery of new diseases and to improve treatment. Although AI holds great promise, many healthcare providers

remain cautious and do not fully rely on AI systems due to doubts about their reliability in detecting diseases and interpreting symptoms. Therefore, it is still necessary to improve and train AI models to enhance their accuracy in disease prediction. AI can never be a medical professional but it can assist them in delivering accuracy and save their efforts. To assure how AI models works against real-life cases, many researchers have created an extensive dataset with patient profiles, each containing details such as age, gender, risk factors, and specific symptoms. Using this data, they trained and tested several models including decision tree, random forest, naive Bayes, logistic regression, and K-nearest neighbors. Each model was fine-tuned and validated through 10-fold cross-validation, confusion matrices, ROC-AUC, and precision-recall curves, ensuring that they could handle both well-balanced and tricky, imbalanced data scenarios. To see how the models work with real-world cases, the teams tested them using clinical vignettes, which are realistic medical scenarios designed to simulate everyday hospital challenges and most of them provide outstanding accurate results. These AI solutions with accurate efficiency assist

professionals to deliver early diagnosis and are making a real difference in the way everyday healthcare is delivered. Looking ahead, AI research should focus on solving these shortcomings to strengthen the collaboration between AI tools and healthcare professionals. A major obstacle for AI in healthcare is the limited interpretability of complex models, which often function as "black boxes" and offer little understanding of how decisions are made. This lack of clarity reduces the confidence of clinicians and makes it difficult to justify results and interpretation in real clinical settings. To tackle both transparency and privacy concerns, federated learning (FL) has gained attention given that it allows multiple healthcare institutions to collaboratively train AI models without sharing sensitive patient data. Kumar *et al.* and team showed that combining explainable AI (XAI) techniques such as SHAP with FL enable interpretable and privacy-preserving prediction of Parkinson's disease. Similarly, Li *et al.* introduced federated neural additive models (FedNAMs), which break down neural networks into feature-specific components, helping clinicians visualize how each input contributes to a diagnosis, while maintaining data security. Together, these studies show that integrating interpretability into FL frameworks can enhance the trust, transparency, and responsible use of AI in clinical practice. Moreover, using decentralized federated learning models can allow different medical centres to build shared AI training systems using local data, helping with early disease detection even in remote areas. Some challenges such as the quality and quantity of training data may affect the precision of AI-based diagnostic tools. Applying AI in healthcare also raises ethical questions due to potential biases in algorithms and the possibility of losing jobs for healthcare professionals.[60–66] AI relies heavily on the quality and completeness of the data it is trained on. Consequently, if the training data lacks sufficient information or contains errors, the system may produce inaccurate disease predictions. This can lead to serious consequences for patients, given that AI cannot always guarantee the reliability of its diagnostic output. Despite growing interest and research in this field, the integration of AI into routine clinical practice is still limited, with many solutions remaining in the development or prototype phase.[67–74]

Conflicts of interest

The authors declare no conflict of interest.

Data availability

No new data were created or generated for this manuscript. It is a review article.

Acknowledgements

References

1 V. Kaul, S. Enslin and S. A. Gross, History of artificial intelligence in medicine, *Gastrointest. Endosc.*, 2020, **92**(4), 807–812, Available from: https://www.sciencedirect.com/science/article/pii/S0016510720344667.

2 S. Kaur, J. Singla, L. Nkenyereye, S. Jha, D. Prashar and G. P. Joshi, *et al.*, Medical Diagnostic Systems Using Artificial Intelligence (AI) Algorithms: Principles and Perspectives, *IEEE Access*, 2020, **8**, 228049–228069.

3 N. Ghaffar Nia, E. Kaplanoglu and A. Nasab, Evaluation of artificial intelligence techniques in disease diagnosis and prediction, *Discov. Artif. Intell.*, 2023, **3**(1), 5, DOI: 10.1007/s44163-023-00049-5.

4 D. G. Vinsard, Y. Mori, M. Misawa, S. e. Kudo, A. Rastogi and U. Bagci, *et al.*, Quality assurance of computer-aided detection and diagnosis in colonoscopy, *Gastrointest. Endosc.*, 2019, **90**(1), 55–63, Available from: https://www.sciencedirect.com/science/article/pii/S001651071930210X.

5 S. A. Hoogenboom, U. Bagci and M. B. Wallace, Artificial intelligence in gastroenterology. The current state of play and the potential. How will it affect our practice and when?, *Tech. Innov. Gastrointest. Endosc.*, 2020, **22**(2), 42–47, Available from: https://www.sciencedirect.com/science/article/pii/S1096288319300737.

6 S. Ioffe and C. Szegedy, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, in *Proceedings of the 32nd International Conference on Machine Learning [Internet]*, ed. F. Bach and D. Blei, Proceedings of Machine Learning Research, Lille, France, 2015, vol. 37, pp. 448–456, Available from: https://proceedings.mlr.press/v37/ioffe15.html.

7 C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed and D. Anguelov, *et al.*, Going deeper with convolutions, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.

8 A. Krizhevsky, I. Sutskever and G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, in *Advances in Neural Information Processing Systems [Internet]*, ed. F. Pereira, C. J. Burges, L. Bottou and K. Q. Weinberger, Curran Associates, Inc., 2012, Available from: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

9 C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, Rethinking the Inception Architecture for Computer Vision, *arXiv*, 2015, preprint, arXiv:1512.00567, DOI: 10.48550/arXiv.1512.00567.

10 K. He, X. Zhang, S. Ren and J. Sun, Deep Residual Learning for Image Recognition, *arXiv*, 2015, preprint, arXiv:1512.03385, DOI: 10.48550/arXiv.1512.03385.

11 Y. LeCun, Y. Bengio and G. Hinton, Deep learning, *Nature*, 2015, **521**(7553), 436–444, DOI: 10.1038/nature14539.

12 M. I. Jordan and T. M. Mitchell, Machine learning: Trends, perspectives, and prospects, *Science*, 2015, **349**(6245), 255–260, DOI: **10.1126/science.aaa8415**.

13 K. VanLEHN, The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems, *Educ. Psychol.*, 2011, **46**(4), 197–221, DOI: **10.1080/00461520.2011.611369**.

14 A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter and H. M. Blau, *et al.*, Dermatologist-level classification of skin cancer with deep neural networks, *Nature*, 2017, **542**(7639), 115–118, DOI: **10.1038/nature21056**.

15 E. Shortliffe, Mycin: A Knowledge-Based Computer Program Applied to Infectious Diseases*, Proceedings the *Annual Symposium on Computer Application [sic] in Medical Care Symposium on Computer Applications in Medical Care*, 1977 Jun.

16 E. Shortliffe, Mycin: A Knowledge-Based Computer Program Applied to Infectious Diseases*, Proceedings the *Annual Symposium on Computer Application [sic] in Medical Care Symposium on Computer Applications in Medical Care*, 1977 Jun.

17 C. A. Kulikowski, Beginnings of Artificial Intelligence in Medicine (AIM): Computational Artifice Assisting Scientific Inquiry and Clinical Art – with Reflections on Present AIM Challenges, *Yearb. Med. Inform.*, 2019, **28**(01), 249–256, DOI: **10.1055/s-0039-1677895**.

18 M. A. J. Tengnah, R. Sooklall and S. D. Nagowah, in *Telemedicine Technologies [Internet]*, ed. D. H. Jude and V. E. Balas, Academic Press, 2019, ch. 9 – A Predictive Model for Hypertension Diagnosis Using Machine Learning Techniques, p. 139–52, Available from: **https://www.sciencedirect.com/science/article/pii/B978012816948300009X**.

19 T. Jo, K. Nho and A. J. Saykin, Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data, *Front. Aging Neurosci.*, 2019, **11**, DOI: **10.3389/fnagi.2019.00220**.

20 J. Chen, D. Remulla, J. H. Nguyen, A. Dua, Y. Liu and P. Dasgupta, *et al.*, Current status of artificial intelligence applications in urology and their potential to influence clinical practice, *BJU Int.*, 2019, **124**(4), 567–577, DOI: **10.1111/bju.14852**.

21 P. H. C. Chen, K. Gadepalli, R. MacDonald, Y. Liu, S. Kadowaki and K. Nagpal, *et al.*, An augmented reality microscope with real-time artificial intelligence integration for cancer diagnosis, *Nat. Med.*, 2019, **25**(9), 1453–1457, DOI: **10.1038/s41591-019-0539-7**.

22 I. M. Nasser and S. S. Abu-Naser, Predicting Tumor Category Using Artificial Neural Networks, *Int. J. Acad. Health Med. Res.*, 2019, **3**, 1–7, Available from: **www.ijeais.org/ijahmr**.

23 V. Sarao, D. Veritti and P. Lanzetta, Automated diabetic retinopathy detection with two different retinal imaging devices using artificial intelligence: a comparison study, *Graefe's Arch. Clin. Exp. Ophthalmol.*, 2020, **258**(12), 2647–2654, DOI: **10.1007/s00417-020-04853-y**.

24 T. D. L. Keenan, T. E. Clemons, A. Domalpally, M. J. Elman, M. Havilio and E. Agrón, *et al.*, Retinal Specialist versus Artificial Intelligence Detection of Retinal Fluid from OCT: Age-Related Eye Disease Study 2: 10-Year Follow-On Study, *Ophthalmology*, 2021, **128**(1), 100–109, Available from: **https://www.sciencedirect.com/science/article/pii/S0161642020305807**.

25 R. Rajalakshmi, R. Subashini, R. M. Anjana and V. Mohan, Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence, *Eye*, 2018, **32**(6), 1138–1144, DOI: **10.1038/s41433-018-0064-9**.

26 Y. Horie, T. Yoshio, K. Aoyama, S. Yoshimizu, Y. Horiuchi and A. Ishiyama, *et al.*, Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks, *Gastrointest. Endosc.*, 2019, **89**(1), 25–32, Available from: **https://www.sciencedirect.com/science/article/pii/S0016510718329262**.

27 R. J. Pignolo, J. J. Connell, W. Briggs, C. J. Kelly, C. Tromans and N. Sultana, *et al.*, Opportunistic assessment of osteoporosis using hip and pelvic X-rays with OsteoSight™: validation of an AI-based tool in a US population, *Osteoporosis Int.*, 2025, **36**(6), 1053–1060, DOI: **10.1007/s00198-025-07487-0**.

28 R. Ani, S. Krishna, N. Anju, M. S. Aslam and O. S. Deepa, Iot based patient monitoring and diagnostic prediction tool using ensemble classifier, in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2017, pp. 1588–1593.

29 H. Luo, G. Xu, C. Li, L. He, L. Luo and Z. Wang, *et al.*, Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study, *Lancet Oncol.*, 2019, **20**(12), 1645–1654, DOI: **10.1016/S1470-2045(19)30637-0**.

30 D. Park, Y. M. Lee, T. Eo, H. J. An, H. Kang and E. Park, *et al.*, Multimodal AI model for preoperative prediction of axillary lymph node metastasis in breast cancer using whole slide images, *npj Precis. Oncol.*, 2025, **9**(1), 131, DOI: **10.1038/s41698-025-00914-9**.

31 P. Ström, K. Kartasalo, H. Olsson, L. Solorzano, B. Delahunt and D. M. Berney, *et al.*, Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study, *Lancet Oncol.*, 2020, **21**(2), 222–232, DOI: **10.1016/S1470-2045(19)30738-7**.

32 W. Bulten, K. Kartasalo, P. H. C. Chen, P. Ström, H. Pinckaers and K. Nagpal, *et al.*, Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge, *Nat. Med.*, 2022, **28**(1), 154–163, DOI: **10.1038/s41591-021-01620-2**.

33 K. Hirai, T. Kuwahara, K. Furukawa, N. Kakushima, S. Furune and H. Yamamoto, *et al.*, Artificial intelligence-based diagnosis of upper gastrointestinal subepithelial lesions on endoscopic ultrasonography images, *Gastric Cancer*, 2022, **25**(2), 382–391, DOI: **10.1007/s10120-021-01261-x**.

34 N. Linder, J. C. Taylor, R. Colling, R. Pell, E. Alveyn and J. Joseph, *et al.*, Deep learning for detecting tumour-infiltrating

lymphocytes in testicular germ cell tumours, *J. Clin. Pathol.*, 2019, **72**(2), 157–164, Available from: **https://jcp.bmj.com/content/72/2/157**.

35 D. d. A. Rodrigues, R. F. Ivo, S. C. Satapathy, S. Wang, J. Hemanth and P. P. R. Filho, A new approach for classification skin lesion based on transfer learning, deep learning, and IoT system, *Pattern Recognit. Lett.*, 2020, **136**, 8–15, Available from: **https://www.sciencedirect.com/science/article/pii/S0167865520301987**.

36 A. Kouketsu, C. Doi, H. Tanaka, T. Araki, R. Nakayama and T. Toyooka, *et al.*, Detection of oral cancer and oral potentially malignant disorders using artificial intelligence-based image analysis, *Head Neck*, 2024, **46**(9), 2253–2260, DOI: **10.1002/hed.27843**.

37 Y. Aggarwal, J. Das, P. M. Mazumder, R. Kumar and R. K. Sinha, Heart rate variability features from nonlinear cardiac dynamics in identification of diabetes using artificial neural network and support vector machine, *Biocybern. Biomed. Eng.*, 2020, **40**(3), 1002–1009, Available from: **https://www.sciencedirect.com/science/article/pii/S0208521620300590**.

38 G. Swapna, R. Vinayakumar and K. P. Soman, Diabetes detection using deep learning algorithms, *ICT Express*, 2018, **4**(4), 243–246, Available from: **https://www.sciencedirect.com/science/article/pii/S2405959518304624**.

39 N. P. Tigga and S. Garg, Prediction of Type 2 Diabetes using Machine Learning Classification Methods, *Procedia Comput. Sci.*, 2020, **167**, 706–716, Available from: **https://www.sciencedirect.com/science/article/pii/S1877050920308024**.

40 Ö. Yıldırım, P. Pławiak, R. S. Tan and U. R. Acharya, Arrhythmia detection using deep convolutional neural network with long duration ECG signals, *Comput. Biol. Med.*, 2018, **102**, 411–420, Available from: **https://www.sciencedirect.com/science/article/pii/S0010482518302713**.

41 H. Kanegae, K. Suzuki, K. Fukatani, T. Ito, N. Harada and K. Kario, Highly precise risk prediction model for new-onset hypertension using artificial intelligence techniques, *J. Clin. Hypertens.*, 2020, **22**(3), 445–450, Available from: **https://onlinelibrary.wiley.com/doi/abs/10.1111/jch.13759**.

42 D. G. Kiely, O. Doyle, E. Drage, H. Jenner, V. Salvatelli and F. A. Daniels, *et al.*, Utilising artificial intelligence to determine patients at risk of a rare disease: idiopathic pulmonary arterial hypertension, *Pulm. Circ.*, 2019, **9**(4), 2045894019890549, DOI: **10.1177/2045894019890549**.

43 J. Olczak, N. Fahlberg, A. Maki, A. S. Razavian, A. Jilert and A. Stark, *et al.*, Artificial intelligence for analyzing orthopedic trauma radiographs, *Acta Orthop.*, 2017, **88**(6), 581–586, DOI: **10.1080/17453674.2017.1344459**.

44 K. Paek, S. Kim, S. Tak, M. K. Kim, J. Park and S. Chung, *et al.*, A high-throughput biomimetic bone-on-a-chip platform with artificial intelligence-assisted image analysis for osteoporosis drug testing, *Bioeng. Transl. Med.*, 2023, **8**(1), e10313, DOI: **10.1002/btm2.10313**.

45 A. Katako, P. Shelton, A. L. Goertzen, D. Levin, B. Bybel and M. Aljuaid, *et al.*, Machine learning identified an Alzheimer's disease-related FDG-PET pattern which is also expressed in Lewy body dementia and Parkinson's disease

dementia, *Sci. Rep.*, 2018, **8**(1), 13236, DOI: **10.1038/s41598-018-31653-6**.

46 O. F. F. Odish, K. Johnsen, P. van Someren, R. A. C. Roos and J. G. van Dijk, EEG may serve as a biomarker in Huntington's disease using machine learning automatic classification, *Sci. Rep.*, 2018, **8**(1), 16090, DOI: **10.1038/s41598-018-34269-y**.

47 Y. Chen, M. Sha, X. Zhao, J. Ma, H. Ni and W. Gao, *et al.*, Automated detection of pathologic white matter alterations in Alzheimer's disease using combined diffusivity and kurtosis method, *Psychiatry Res., Neuroimaging*, 2017, **264**, 35–45, Available from: **https://www.sciencedirect.com/science/article/pii/S092549271630186X**.

48 H. K. van der Burgh, R. Schmidt, H. J. Westeneng, M. A. de Reus, L. H. van den Berg and M. P. van den Heuvel, Deep learning predictions of survival based on MRI in amyotrophic lateral sclerosis, *NeuroImage Clin.*, 2017, **13**, 361–369Available from: **https://www.sciencedirect.com/science/article/pii/S2213158216301899**.

49 H. Takeshige-Amano, G. Oyama, M. Ogawa, K. Fusegi, T. Kambe and K. Shiina, *et al.*, Digital detection of Alzheimer's disease using smiles and conversations with a chatbot, *Sci. Rep.*, 2024, **14**(1), 26309, DOI: **10.1038/s41598-024-77220-0**.

50 P. Ström, K. Kartasalo, H. Olsson, L. Solorzano, B. Delahunt and D. M. Berney, *et al.*, Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study, *Lancet Oncol.*, 2020, **21**(2), 222–232, DOI: **10.1016/S1470-2045(19)30738-7**.

51 G. Alotaibi, M. Awawdeh, F. F. Farook, M. Aljohani, R. M. Aldhafiri and M. Aldhoayan, Artificial intelligence (AI) diagnostic tools: utilizing a convolutional neural network (CNN) to assess periodontal bone level radiographically—a retrospective study, *BMC Oral Health*, 2022, **22**(1), 399, DOI: **10.1186/s12903-022-02436-3**.

52 M. Katsuki, Y. Matsumori, S. Kawamura, K. Kashiwagi, A. Koh and S. Tachikawa, *et al.*, Developing an artificial intelligence–based diagnostic model of headaches from a dataset of clinic patients' records, *Headache*, 2023, **63**(8), 1097–1108, DOI: **10.1111/head.14611**.

53 S. Nooh, M. Ragab, R. Aboalela, A. A. M. AL-Ghamdi, O. A. Abdulkader and G. Alghamdi, An exploratory analysis of longitudinal artificial intelligence for cognitive fatigue detection using neurophysiological based biosignal data, *Sci. Rep.*, 2025, **15**(1), 15736, DOI: **10.1038/s41598-025-96816-8**.

54 K. Zhang, J. He and W. Ji, *et al.*, Machine learning model for differentiating xanthogranulomatous cholecystitis and gallbladder cancer in multicenter largescale study, *NPJ Digit. Med.*, 2025, **8**, 590, DOI: **10.1038/s41746-025-01991-7**.

55 D. Kavungal, P. Magalhães, S. T. Kumar, R. Kolla, H. A. Lashuel and H. Altug, Artificial intelligence–coupled plasmonic infrared sensor for detection of structural protein biomarkers in neurodegenerative diseases, *Sci. Adv.*, 2023, **9**(28), eadg9644, DOI: **10.1126/sciadv.adg9644**.

56 A. Sarwar, M. Ali, J. Manhas and V. Sharma, Diagnosis of diabetes type-II using hybrid machine learning based

ensemble model, *Int. J. Inf. Technol.*, 2020, **12**(2), 419–428, DOI: **10.1007/s41870-018-0270-5**.

57 M. M. Musleh, E. Alajrami, A. J. Khalil, B. S. Abu-Nasser, A. M. Barhoom and S. S. Abu-Naser, Predicting Liver Patients using Artificial Neural Network, *Int. J. Acad. Inf. Syst. Res.*, 2019, **3**, 1–11, Available from: **www.ijeais.org/ijaisr**.

58 D. C. Yadav and S. Pal, Prediction of thyroid disease using decision tree ensemble method, *Human-Intelligent Systems Integration*, 2020, **2**(1), 89–95, DOI: **10.1007/s42454-020-00006-y**.

59 X. Xie, W. Yu, Z. Chen, L. Wang, J. Yang and S. Liu, *et al.*, Early-stage oral cancer diagnosis by artificial intelligence-based SERS using Ag NWs@ZIF core-shell nanochains, *Nanoscale*, 2023, **15**(32), 13466–13472.

60 B. Khan, H. Fatima, A. Qureshi, S. Kumar, A. Hanan and J. Hussain, *et al.*, Drawbacks of Artificial Intelligence and Their Potential Solutions in the Healthcare Sector, *Biomed. Mater. & Devices*, 2023, **1**(2), 731–738, DOI: **10.1007/s44174-023-00063-2**.

61 F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li and S. Ma, *et al.*, Artificial intelligence in healthcare: past, present and future, *Stroke Vasc. Neurol.*, 2017, **2**(4), DOI: **10.1136/svn-2017-000101**, Available from: **https://svnsite-bmj.vercel.app/content/2/4/230**.

62 N. Naik, B. M. Z. Hameed, D. K. Shetty, D. Swain, M. Shah and R. Paul, *et al.*, Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility?, *Front. Surg.*, 2022, **9**, DOI: **10.3389/fsurg.2022.862322**.

63 S. Jayakumar, V. Sounderajah, P. Normahani, L. Harling, S. R. Markar and H. Ashrafian, *et al.*, Quality assessment standards in artificial intelligence diagnostic accuracy systematic reviews: a meta-research study, *NPJ Digit. Med.*, 2022, **5**(1), 11, DOI: **10.1038/s41746-021-00544-y**.

64 A. Vaid, S. Jaladanki, J. Xu, S. Teng, A. Kumar, S. Lee, S. Somani, I. Paranjpe, J. De Freitas, T. Wanyan, K. Johnson, M. Bicak, E. Klang, Y. Kwon, A. Costa, S. Zhao, R. Miotto, A. Charney, E. Böttinger, Z. Fayad, G. Nadkarni, F. Wang and B. Glicksberg, Federated Learning of Electronic Health Records to Improve Mortality Prediction in Hospitalized Patients With COVID-19: Machine Learning Approach, *JMIR Med. Inform.*, 2021, **9**(1), e24207, DOI: **10.2196/24207**, **https://medinform.jmir.org/2021/1/e24207**.

65 L. Aissaoui Ferhi, *et al.*, Enhancing diagnostic accuracy in symptom-based health checkers: a comprehensive machine learning approach with clinical vignettes and benchmarking, *Front. Artif. Intell.*, 2024, **7**, 1397388, DOI: **10.3389/frai.2024.1397388**.

66 A. Amato and D. Branco, SemFedXAI: A Semantic Framework for Explainable Federated Learning in Healthcare, *Information*, 2025, **16**, 435, DOI: **10.3390/info16060435**.

67 J. Bajwa, U. Munir, A. Nori and B. Williams, Artificial intelligence in healthcare: transforming the practice of medicine, *Future Healthc. J.*, 2021, **8**(2), e188–e194, Available from: **https://www.sciencedirect.com/science/article/pii/S2514664524005277**.

68 C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado and D. King, Key challenges for delivering clinical impact with artificial intelligence, *BMC Med.*, 2019, **17**(1), 195, DOI: **10.1186/s12916-019-1426-2**.

69 J. Lawrence, *et al.*, Topological Design and Synthesis of High-Spin Aza-triangulenes without Jahn-Teller Distortions, *ACS Nano*, 2023, **17**(20), 20237–20245, DOI: **10.1021/acsnano.3c05974**.

70 J. Chen, Y. Li, Y. Jiang, L. Mao, M. Lai, L. Jiang, H. Liu and Z. Nie, TiO_2/MXene-Assisted LDI-MS for Urine Metabolic Profiling in Urinary Disease, *Adv. Funct. Mater.*, 2021, **31**, 2106743, DOI: **10.1002/adfm.202106743**.

71 H. Jin, *et al.*, Robust Multifunctional Ultrathin 2 Nanometer Organic Nanofibers, *ACS Nano*, 2024, **18**(32), 21576–21584, DOI: **10.1021/acsnano.4c08229**.

72 M. Song, Y. Li, W. Ma, J. Chen and Z. Nie, Hand-Held Nanoelectrospray Ionization with Frequency and Amplitude Tunability for Metabolomics of Saline Biosamples, *Anal. Chem.*, 2025, **97**(33), 18327–18334, DOI: **10.1021/acs.analchem.5c03797**.

73 A. Curioni, Artificial intelligence: Why we must get it right, *Informatik-Spektrum*, 2018, **41**(1), 7–14, DOI: **10.1007/s00287-018-1087-0**.

74 E. J. Topol, High-performance medicine: the convergence of human and artificial intelligence, *Nat. Med.*, 2019, **25**(1), 44–56, DOI: **10.1038/s41591-018-0300-7**.