International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(5s) 2026 pp. 985-995
https://doi.org/10.70917/ijcisim-2026-2846
Article

AI and Big Data Convergence in Predictive Analytics for Early Disease Detection and Personalized Treatment

Ashraful Islam[1], Tajul Islam Rafi[2], Zerin Akter Tanni[3] Anika Anwar Shoshi[4], Md Abu Kawsar Prodhan Hemal[5], Subha Shamarukh[6]

[1]Washington University of Science and Technology, Alexandria, VA, USA, ashralam.student@wust.edu, https://orcid.org/0009-0001-8067-7331
[2]Pacific states university, Los Angeles, CA 90010, USA, p25840@psuca.edu, https://orcid.org/0009-0004-9277-0607
[3]St. Francis College, Brooklyn, New York, USA, zerinaktertanni@gmail.com, https://orcid.org/0009-0002-7143-2874
[4]Dr. Sirajul Islam Medical College & Hospital Ltd, anikashoshi3@gmail.com, https://orcid.org/0009-0007-2185-6448
[5]Pacific States University, Los Angeles, CA 90010, USA, p26181@psuca.edu, https://orcid.org/0009-0002-3376-5192
[6]University of Rochester, Rochester, New York, USA, shamarukhsubha@gmail.com, https://orcid.org/0009-0000-2170-1541
***Corresponding Author:** Subha Shamarukh, shamarukhsubha@gmail.com

Abstract: Healthcare is being reshaped by the pairing of artificial intelligence with large-scale data analytics, and one of the clearest effects of that pairing is earlier, more accurate disease detection alongside treatment plans built around the individual patient rather than the average one. As these two technologies have matured together, they have made it possible to build predictive models that can make sense of the kind of messy, multi-source data hospitals that generate electronic health records, medical images, genomic sequences, and streams from wearable sensors. This paper looks at recent progress in AI-driven predictive analytics across several clinical areas, including oncology, infectious disease, neurology, cardiometabolic conditions, mental health, neurodevelopmental disorders, and precision medicine. Across these areas, a common toolkit keeps reappearing: machine learning, deep learning, transformer architectures, autoencoders, automated machine learning, explainable AI, and generative AI, each applied to some combination of diagnosis, risk prediction, treatment planning, and decision support. Taken together, the studies reviewed here suggest that combining heterogeneous health data with these newer analytical methods genuinely improves diagnostic accuracy, opens the door to earlier intervention, and supports care that is tailored to the individual. That said, the field is not without friction. Data heterogeneity, weak generalizability across populations, privacy and security concerns, algorithmic bias, and a persistent lack of model transparency all remain real obstacles. Explainable AI has become one of the more important responses to this last problem, since clinicians are understandably reluctant to act on predictions they cannot interrogate. Overall, the convergence of AI and big data gives precision medicine a solid foundation to build on, but only if that progress is paired with rigorous validation, sound ethical governance, and models clinicians can trust.
Keywords: Artificial Intelligence; Big Data Analytics; Predictive Analytics; Early Disease Detection; Personalized Medicine; Machine Learning; Deep Learning; Explainable Artificial Intelligence; Generative Artificial Intelligence; Precision Healthcare

1. Introduction

Every day, health systems around the world produce staggering amounts of data, electronic health records, diagnostic scans, genomic sequences, continuous streams from wearable devices, far more than traditional statistical methods were ever built to handle. For a long time, the gap between how much data existed and how much of it could be turned into timely, individualized clinical insight kept widening. That gap has started to close over roughly the last two or three years, as artificial intelligence and big data infrastructure have matured into genuinely complementary tools. AI brings the algorithmic muscle needed to pick out complex, non-linear patterns in the data; big data infrastructure supplies the scale required to train and test those algorithms across populations diverse enough to matter clinically.

This convergence shows up most clearly in three connected areas. The first is diagnostic imaging and screening, where deep learning models, convolutional neural networks and, increasingly, transformer-based architectures, have proven capable of catching disease markers that are easy for a human reader to miss or that sit right at the edge of what the eye can reliably detect. The second is precision and genomic medicine, where machine learning models trained on multi-modal data are being used to pin down biomarkers, stratify patients by risk, and adjust treatment to the individual rather than the population. The third is population-level surveillance and chronic disease management, where predictive analytics built from longitudinal clinical and wearable data are helping clinicians' step in before an acute event happens rather than after.

This review deliberately keeps its scope narrow, focusing on twenty-nine studies published between 2024 and 2026 in journals and conference proceedings tied to established publishers such as MDPI, IEEE, Springer Nature, and Frontiers Media. The studies span tuberculosis screening; breast, pancreatic, prostate, and skin oncology; Alzheimer's disease and broader neurodegenerative screening; type 2 diabetes; sepsis; antimicrobial resistance; autism spectrum diagnosis; care for hospitalized psychiatric patients; generative AI applied to drug discovery; and bibliometric work mapping explainable AI in medicine. Keeping the window this tight and this recent is a deliberate choice, it means the trends pulled out of this review reflect where the field stands right now, not techniques that have already been superseded.

The rest of the paper is organized as follows. Section 2 lays out the review methodology. Sections 3 through 8 works through the literature by clinical domain, diagnostic imaging, oncology and precision medicine, neurology and cognitive disorders, infectious disease and antimicrobial resistance, mental health and neurodevelopmental care, and cardiometabolic disease alongside drug discovery. Section 9 turns to explainability and clinical trust. Sections 10 and 11 pull the reviewed studies together in tabular form and compare the underlying AI/ML techniques side by side. Section 12 discusses the challenges that cut across all these domains, Section 13 looks ahead to where the field might go next, and Section 14 concludes.

2. Review Methodology

This paper takes a narrative review approach rather than a formal systematic review with quantitative meta-analysis, largely because the goal here is to draw out thematic patterns across a body of recent, topically related work rather than to pool effect sizes. The twenty-nine studies included were chosen against four criteria: they had to be published between January 2024 and 2026; they had to bear directly on the use of AI or big data analytics in disease detection, screening, or personalized treatment; they had to appear in a peer-reviewed journal or peer-reviewed conference proceeding tied to a recognized academic publisher; and they had to include enough methodological detail, in either the abstract or the full text, to let us characterize the data modality, the algorithmic approach, and the clinical application.

For each study included, we pulled out the clinical or biological domain, the data modality used, the general class of AI or machine learning technique, and the stated clinical objective. Those four attributes form the backbone of the comparison tables in Sections 10 and 11. Rather than organizing the studies chronologically, we grouped them thematically, since what we were after was convergent methodological trends across disease areas within this tight, current window rather than a history of how the field arrived here. Figure 1 shows how the twenty-nine reviewed studies are distributed by publication year within that window.

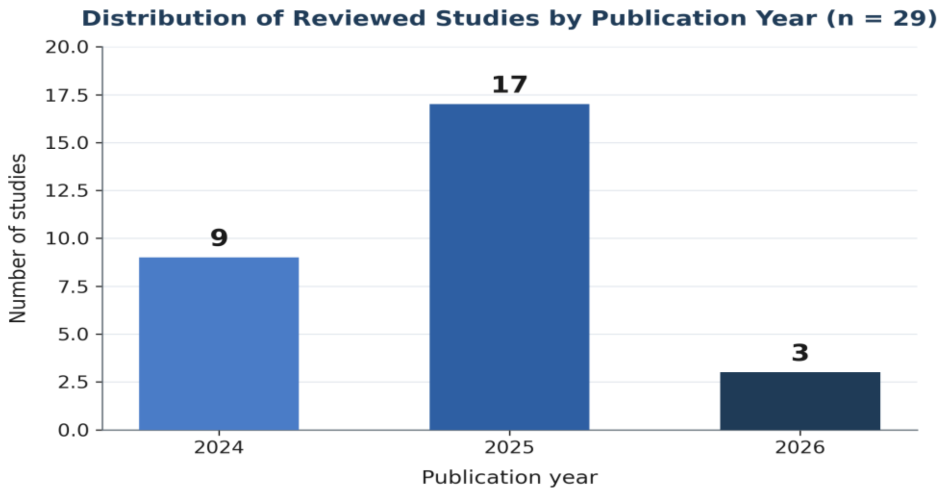

Figure 1. Distribution of the twenty-nine reviewed studies by publication year (2024-2026).

3. AI-Driven Diagnostic Imaging and Early Disease Screening

Medical imaging is still one of the busiest corners of healthcare AI, and the most recent work in this space reads less like a break from earlier methods and more like a steady refinement of them. Sikder et al. (2026), for instance, built a hybrid model that pairs convolutional neural networks with Swin transformer components for tuberculosis screening from chest radiographs, and built explainability directly into the architecture so that a given prediction could be traced back to the specific regions of the image driving it. Kotei and Thirunavukarasu (2024) took a similar tack, combining transformer and convolutional components for tuberculosis detection from chest X-rays, and found that the hybrid setup beat either architecture family working alone.

Oncological imaging tells a similar story. Khair et al. (2025) built a deep neural network-based system for identifying pancreatic tumors, a domain where catching disease early matters enormously, given how poor the prognosis tends to be once pancreatic cancer reaches a late stage. Breast cancer imaging, in particular, has drawn an unusual amount of recent attention. Carriero et al. (2024) reviewed the current state of deep learning across mammography, ultrasound, and MRI for breast cancer detection, and noted that radiomic methods are increasingly being pushed past diagnosis and into prognosis. Abdullah et al. (2025) went further with a systematic review and meta-analysis of deep learning applied specifically to breast MRI, pooling diagnostic performance estimates from forty studies and flagging how uneven the reporting quality still is across that literature. In prostate oncology, Horasan and Güneş (2024) proposed an ensemble deep learning framework that combines several pretrained architectures through soft voting to make MRI-based prostate cancer detection more reliable. Dermatological screening has followed much the same path: Hussain and Toscano (2024) reviewed machine learning and deep neural network approaches to skin cancer recognition and pointed to dataset imbalance and weak interpretability as the two obstacles still standing between these models and real clinical use.

4. Big Data and Predictive Analytics in Oncology and Precision Medicine

Cancer care remains one of the strongest proving grounds for AI-driven personalization, largely because tumors vary so much from one patient to the next that population-level guidelines can only take treatment planning so far. Ahmed et al. (2025) reviewed how big data analytics are being used to personalize cancer treatment, pulling together how integrated clinical, molecular, and treatment-response data can inform which therapy a given patient should receive. Their analysis makes a point worth underlining it is usually data integration, not any single clever algorithm, that ends up being the bottleneck when trying to put precision oncology into practice at scale.

Related work has tackled that integration problem from a different angle, at the level of representation learning. Manik (2025) used autoencoder neural networks to bring together heterogeneous cancer data types, applying unsupervised representation learning to compress high-dimensional, multi-modal cancer data into a lower-dimensional feature space that downstream predictive models can work with. That approach matters most in the very common situation where clinical, imaging, and molecular data show up in incompatible formats and resolutions, a headache familiar to anyone who has tried to combine real-world oncology datasets.

5. Neurology and Cognitive Disorders

Neurodegenerative and cognitive conditions pose a distinct problem for AI: the signals that matter clinically tend to be subtle, slow to develop, and scattered across several different types of data at once. Manik et al. (2026) introduced a data-driven framework for detecting Alzheimer's disease and broader cognitive decline, arguing that structured modeling drawing on multiple clinical sources can flag at-risk individuals earlier than standard neuropsychological testing manages on its own. Set against the wider neurodegenerative literature, Yousefi et al. (2024) carried out a systematic review of machine learning applications for early detection and screening in this space and found that how well a model performs depends heavily on what kind of input data it was trained on.

Sensor-based screening has moved forward too. Cardinali et al. (2025) validated a machine learning-driven microwave sensing system for early Alzheimer's detection, showing that a low-cost, non-invasive conformal antenna array could pick up disease-related electromagnetic signatures in cerebrospinal fluid without needing expensive imaging equipment. Figure 2 pulls these different threads together, showing how imaging, sensing, and structured clinical variables converge in a shared analytical pipeline underlying both the oncology work discussed above and the neurology work discussed here.

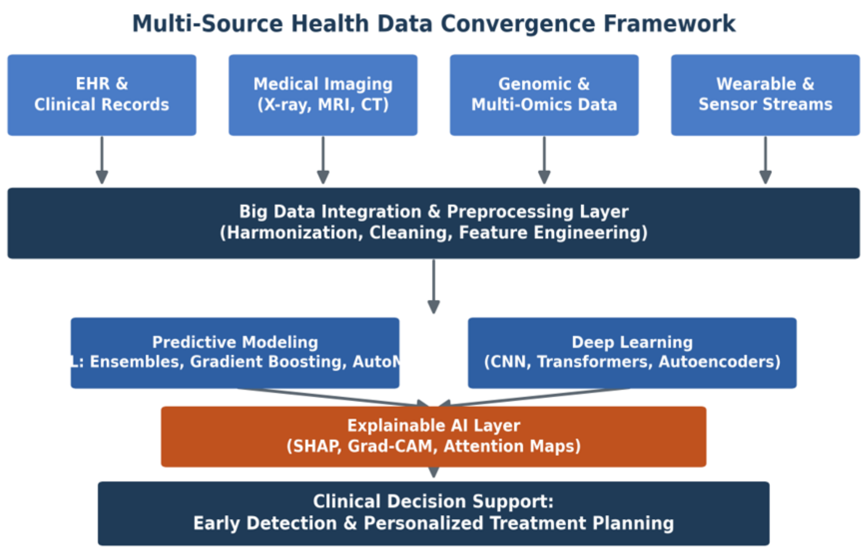

Figure 2. Conceptual framework illustrates the convergence of multi-source health data through big data integration, machine learning and deep learning modeling, and explainable AI layers to support clinical decision-making.

6. AI in Infectious Disease Screening and Antimicrobial Resistance Surveillance

AI contributes to infectious disease control in two related but distinct ways: speeding up diagnosis for the individual patient and supporting surveillance of emerging threats like antimicrobial resistance at the population level. The tuberculosis screening work from Sikder et al. (2026) and Kotei and Thirunavukarasu (2024), already discussed in Section 3, is a good example of the first using chest radiograph classification to flag likely cases for confirmatory testing in settings where lab capacity is limited.

At the population level, tracking antimicrobial resistance is fundamentally a data-integration problem, since resistance patterns emerge from the interplay of prescribing habits, pathogen genomics, and geographic spread, no single data source captures all of that on its own. Broschat et al. (2024) framed this challenge in an editorial synthesizing machine learning approaches to antimicrobial discovery and resistance, arguing that predictive models trained on genomic and surveillance data are becoming central to antimicrobial stewardship. These two levels of effort complement each other well: earlier individual diagnosis cuts down on unnecessary antimicrobial use, while surveillance analytics give clinicians and public health authorities a way to see resistance trends coming before they spread widely.

7. AI-Assisted Mental Health and Neurodevelopmental Care

Mental health and neurodevelopmental conditions are a harder problem for AI-based detection because diagnosis so often rests on behavioral and contextual cues rather than clean biological markers. Samiun et al. (2025) conducted a scoping review of how AI is being used to manage hospitalized patients with mental illness, and found that these applications span risk prediction, monitoring treatment response, and supporting clinical workflows, though the quality and consistency of the evidence base still varies a good deal from study to study.

On the neurodevelopmental side, Manik et al. (2026) looked at AI-driven approaches to improving early autism diagnosis and personalizing interventions, paying particular attention to what this means for special education policy and practice in the United States. Ehsan et al. (2025) applied automated machine learning to streamline autism spectrum disorder diagnosis and found that an automated pipeline cut down substantially on the manual work needed to land on a well-performing model, without giving up diagnostic accuracy in the process. Temiz et al. (2025) came at the problem from a more biological angle, applying explainable machine learning to metagenomic data both to predict autism spectrum disorder and to identify candidate gut-microbiome biomarkers, a nice example of how interpretability tools can support diagnosis and hypothesis generation at the same time.

8. Cardiometabolic Disease Prediction and AI-Enabled Drug Discovery

Type 2 diabetes prediction has turned into one of the busiest subfields in this literature. Manik et al. (2025) applied AI-driven machine learning and big data analytics to early detection of type 2 diabetes, showing how predictive modeling for a chronic metabolic condition can support both earlier diagnosis and more individualized drug management. Liu et al. (2025) ran a ten-year longitudinal study using machine learning to predict incident type 2 diabetes among relatively healthy adults and found that gradient-boosted models gave the strongest discrimination among the algorithms they tested. Kiran et al. (2025) added a broader lens with a thirty-three-year bibliometric and literature analysis of machine learning and AI in type 2 diabetes prediction and found a clear acceleration in both publication volume and methodological sophistication in the most recent years. Caballero-María et al. (2025) took a multi-omics approach, building a deep learning model that draws on clinical, biochemical, and gut microbiota profiles to predict type 2 diabetes, while Li et al. (2025) showed that adding dietary indicators such as glycemic index and load meaningfully improved models predicting the progression from prediabetes to full type 2 diabetes.

Predictive analytics for structured clinical data reach beyond diabetes and into acute care as well. Ben Khalfallah et al. (2025) applied machine learning to estimate length of hospital stay and mortality risk in sepsis and found that ensemble models produced risk stratification clinicians could actually act on when allocating resources. Predictive analytics has also found its way into drug development, where generative AI is increasingly used to shorten what has traditionally been a very long discovery timeline. Manik et al. (2025) looked at how machine learning and big data analytics can support genomics-based drug discovery within a precision medicine framework, connecting molecular target identification to downstream treatment personalization. Nilima et al. (2024) examined this same push from artificial intelligence and machine learning more broadly, reinforcing that predictive modeling is now being used not just to identify promising compounds but also to optimize trial design and anticipate therapeutic response. Fahad et al. (2025) reviewed the broader landscape of generative AI in clinical applications between 2020 and 2025 and concluded that molecular design and drug discovery are among the fastest-growing application areas, right alongside medical imaging and electronic health record analysis.

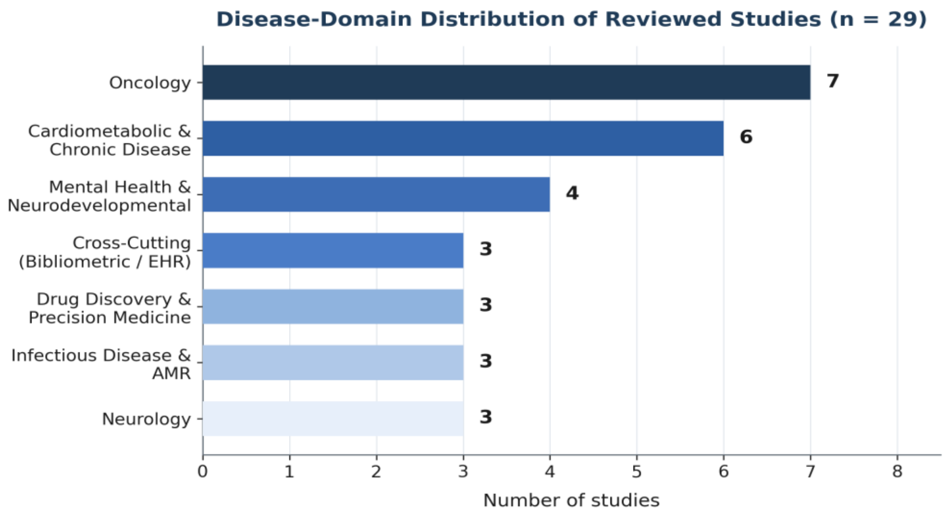

Figure 3. Disease-domain distribution of the twenty-nine reviewed studies, showing concentration in oncology, followed by cardiometabolic and chronic disease, mental health and neurodevelopmental care, and neurology.

9. Explainability, Trust, and Clinical Translation

One theme keeps coming back across this literature: model transparency isn't optional if these tools are ever going to be adopted in real clinical settings. A model can perform beautifully and still get ignored by clinicians if it can't explain itself, accuracy alone doesn't buy trust. Sikder et al. (2026) tackled this head-on by building explainability directly into their hybrid tuberculosis screening architecture, so that predictions could be traced visually back to specific regions of the radiograph rather than handed to clinicians as an unexplained output.

Zooming out, Frasca et al. (2024) carried out a systematic bibliometric review of explainability and interpretability research in medicine, working through 448 articles and documenting exponential growth in this subfield over the past decade. Lin et al. (2025) ran a similar bibliometric analysis of the broader advance of AI in medicine and identified imaging and diagnostics as the most mature application clusters, with explainability standing out as one of the fastest-growing emerging themes. Read together, these two bibliometric studies suggest explainability research has stopped being a side concern and has become a central, fast-expanding pillar of applied medical AI.

More broadly, the explainability methods that show up across this literature, attention-based visualization, feature attribution, saliency mapping, all serve the same basic function: turning a model's high-dimensional internals into something a clinician can actually check against their own expertise. As Figure 4 illustrates, the explainability layer sits deliberately between the modeling layer and the clinical decision-support output, which reflects its role as a translation step rather than an optional add-on tacked on at the end.

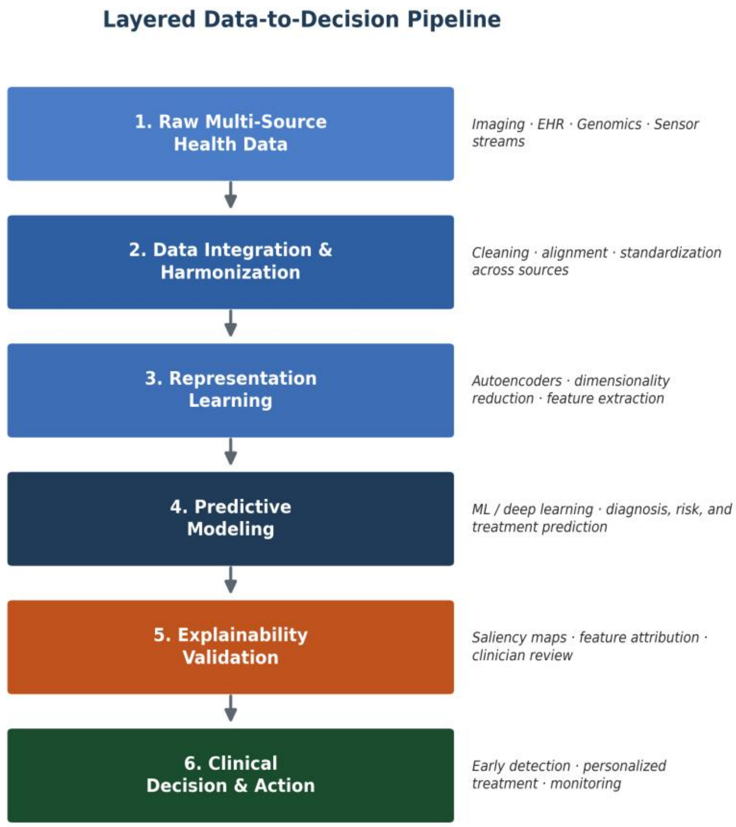

Figure 4. Layered data-to-decision pipeline illustrating the progression from raw multi-source health data through integration, representation learning, predictive modeling, and explainability validation prior to clinical use.

10. Summary of Reviewed Studies

Table 1 pulls together the twenty-nine reviewed studies, all published between 2024 and 2026, organized by disease domain, primary data modality, analytical technique, and publication venue.

Table 1. Summary of reviewed AI and big-data studies in disease detection and personalized treatment (2024-2026).

Study	Disease/Domain	Data Modality	AI/ML Technique	Publication Venue
Sikder et al. (2026)	Infectious disease (TB)	Chest X-ray imaging	Hybrid CNN-Swin transformer	Discover Artificial Intelligence
Kotei & Thirunavukarasu (2024)	Infectious disease (TB)	Chest X-ray imaging	Transformer-CNN hybrid	IEEE Access
Khair et al. (2025)	Oncology (pancreatic)	Medical imaging	Deep neural network	Intelligent Decision Technologies
Carriero et al. (2024)	Oncology (breast)	Multi-modal imaging	Deep learning review	Diagnostics (MDPI)
Abdullah et al. (2025)	Oncology (breast)	MRI	Deep learning meta-analysis	European Radiology (Springer)
Horasan & Güneş (2024)	Oncology (prostate)	MRI	Ensemble deep learning	Diagnostics (MDPI)
Hussain & Toscano (2024)	Oncology (skin)	Dermoscopic imaging	ML/DNN review	Symmetry (MDPI)
Ahmed et al. (2025)	Oncology (general)	Multi-modal clinical review	Big data analytics	Journal of Engineering

Manik (2025)	Oncology (general)	Heterogeneous cancer data	Autoencoder neural network	J. Info. Systems Eng. & Mgmt.
Manik et al. (2026)	Neurology (Alzheimer's)	Multi-source clinical	Data-driven detection	IEEE ICECET Proceedings
Yousefi et al. (2024)	Neurology (neurodegenerative)	Multi-modal	ML systematic review	Frontiers in Neurology
Cardinali et al. (2025)	Neurology (Alzheimer's)	Microwave sensing	ML classification	Sensors (MDPI)
Broschat et al. (2024)	Infectious disease (AMR)	Genomic/surveillance data	ML editorial synthesis	Frontiers in Bioinformatics
Samiun et al. (2025)	Mental health (hospitalized)	Clinical/behavioral	Scoping review	Discover Public Health
Manik et al. (2026)	Neurodevelopmental (autism)	Clinical/behavioral	AI-driven diagnostic approach	IEEE ICECET Proceedings
Ehsan et al. (2025)	Neurodevelopmental (autism)	Clinical/behavioral	Automated machine learning	Diagnostics (MDPI)
Temiz et al. (2025)	Neurodevelopmental (autism)	Metagenomic data	Explainable ML	Applied Sciences (MDPI)
Manik et al. (2025)	Cardiometabolic (T2 diabetes)	Structured clinical data	AI/ML + big data analytics	Springer AIR 2025 Proceedings
Liu et al. (2025)	Cardiometabolic (T2 diabetes)	Longitudinal clinical data	ML incidence prediction	Diagnostics (MDPI)
Kiran et al. (2025)	Cardiometabolic (T2 diabetes)	N/A (bibliometric)	Bibliometric analysis	Frontiers in Digital Health
Caballero-María et al. (2025)	Cardiometabolic (T2 diabetes)	Clinical + gut microbiota	Deep learning	Applied Sciences (MDPI)
Li et al. (2025)	Cardiometabolic (prediabetes)	Clinical + dietary data	ML with dietary indicators	Nutrients (MDPI)
Ben Khalfallah et al. (2025)	Acute care (sepsis)	ICU clinical data	Ensemble ML	Computation (MDPI)
Manik et al. (2025)	Precision medicine/genomics	Genomic data	ML + big data analytics	Journal of Posthumanism
Nilima et al. (2024)	Drug discovery	Molecular/clinical data	AI/ML modeling	IEEE COMPAS Proceedings
Fahad et al. (2025)	Drug discovery/cross-cutting	N/A (review)	Generative AI review	Frontiers in Digital Health
Frasca et al. (2024)	Cross-cutting (XAI)	N/A (bibliometric)	Bibliometric review	Discover Artificial Intelligence
Lin et al. (2025)	Cross-cutting (AI in medicine)	N/A (bibliometric)	Bibliometric analysis	Frontiers in Medicine
Swinckels et al. (2024)	Cross-cutting (EHR)	Longitudinal EHR	ML/DL scoping review	J. Medical Internet Research

11. Comparative Overview of AI/ML Techniques

Table 2 lines up the main classes of AI and machine learning technique found across the reviewed 2024-2026 literature, along with their typical clinical applications, relative strengths, and known limitations.

Table 2. Comparative summary of AI/ML technique classes identified across the reviewed studies.

Technique Class	Representative Application(s)	Key Strengths	Key Limitations
Convolutional Neural Networks (CNN)	Chest X-ray and breast/prostate imaging screening	Strong local feature extraction from images; well-established tooling	Limited long-range context; large labeled datasets typically required

Transformer-based architectures	Hybrid tuberculosis imaging classification	Captures long-range spatial dependencies; pairs well with CNN backbones	Computationally intensive; interpretability requires additional mechanisms
Autoencoders	Multi-modal cancer data integration	Effective unsupervised dimensionality reduction for heterogeneous data	Latent representations can be difficult to interpret clinically
Ensemble/gradient-based ML	Chronic disease risk prediction, diabetes and sepsis screening	Performs well on structured/tabular clinical data; relatively interpretable	May underperform deep learning on unstructured data such as images
Automated machine learning (AutoML)	Autism spectrum disorder diagnosis	Reduces manual model-selection effort; accessible to non-specialists	Can obscure understanding of why a given pipeline was selected
Sensor-based ML (microwave, wearable)	Alzheimer's disease screening	Enables low-cost, non-invasive, continuous monitoring outside the clinic	Signal noise and hardware variability affect reliability
Explainable AI techniques	Cross-cutting clinical decision support	Improves clinician trust and supports regulatory review	No consensus evaluation standard; explanations can be misleading if misapplied
Generative AI models	Drug discovery, molecular design	Accelerates candidate generation beyond traditional screening	Requires rigorous experimental validation of generated candidates

12. Challenges and Limitations

Even with all this recent progress, several challenges keep cutting across the literature and limiting how much it reaches the clinic. The first is data heterogeneity, which hasn't gone away. Imaging, genomic, clinical, and sensor data differ enormously in format, resolution, and quality, and pulling them together as the autoencoder-based work in Section 4 attempts, takes a lot of preprocessing effort that published papers tend to underreport.

The second is model generalizability, which is often limited by how small or geographically narrow the training cohorts are. A number of the studies reviewed here describe models validated on narrow datasets, which raises real questions about how those models would perform in populations with different demographics, environments, or healthcare systems than the ones they were trained on. Swinckels et al. (2024) reached a similar conclusion in their scoping review of machine learning applied to longitudinal electronic health records, noting that most studies report engineering performance metrics while offering comparatively little evidence of actual clinical outcomes.

Third, interpretability and clinician trust remain live issues even where explainability tools have already been built in, as with Sikder et al.'s (2026) hybrid tuberculosis model. Saliency maps and feature attribution genuinely help with transparency, but they don't fully resolve the underlying tension between how complex a model is and how interpretable it can realistically be, especially for the strongest-performing deep learning architectures.

Fourth, ethical and regulatory questions, data privacy, informed consent for secondary use of data, algorithmic bias, carry particular weight in sensitive areas like mental health and neurodevelopmental diagnosis, where a mislabeled or biased prediction could follow someone for a long time, including through educational and social service systems.

13. Future Directions

A few methodological shifts would likely move this field forward. More external, multi-site validation would help settle whether models trained in the specific settings described across this review generalize to broader and more varied patient populations. Standardized reporting for AI-healthcare studies, covering data provenance, cohort characteristics, and validation methodology, would also make it easier to compare across what is currently a fragmented body of recent work.

Federated learning, which lets institutions train models together without pooling sensitive patient data in one place, looks like a promising way to address both the heterogeneity and privacy concerns raised in Section 12. In the same vein, continued work on explainability methods built specifically for high-dimensional imaging and genomic data would help close the interpretability gap that currently limits how much clinicians trust complex model output. Finally, tighter integration between diagnostic and therapeutic AI pipelines, building on the shared analytical foundations underlying both the drug discovery and disease-detection work covered in Sections 4 through 8 could help turn predictive insight into treatment recommendations clinicians can act on.

14. Conclusion

The convergence of artificial intelligence and big data analytics is reshaping how disease gets detected and how treatment gets personalized, across a wide range of clinical areas, tuberculosis and cancer screening, neurodegenerative disease, chronic metabolic conditions, mental health, and drug discovery among them. The twenty-nine studies reviewed here, all published within the past three years, point to a shared underlying architecture: data gets acquired from heterogeneous sources, integrated and reduced through representation learning, run through predictive models drawing on an increasingly varied toolkit of machine learning and deep learning techniques, and, where it's been built in, passed through explainability mechanisms meant to earn clinical trust. Progress has been fast and substantial, but the field still has real work to do on data heterogeneity, generalizability, interpretability, and ethical governance. Getting there will mean standardizing how these studies validate their models, learning on privacy-preserving collaborative learning where it fits, and designing explainability with clinicians in mind from the start the surest path to turning the predictive potential documented in this literature into consistent, equitable gains for patients.

References

1. Abdullah, K. A., Marziali, S., Nanaa, M., Escudero Sánchez, L., Payne, N. R., & Gilbert, F. J. (2025). Deep learning-based breast cancer diagnosis in breast MRI: Systematic review and meta-analysis. European Radiology. https://doi.org/10.1007/s00330-025-11406-6
2. Ahmed, M. K., Rozario, E., Mohonta, S. C., Mou, J. F., Saimon, A. S. M., Moniruzzaman, M., Manik, M. M. T. G., & Hasan, R. (2025). Leveraging big data analytics for personalized cancer treatment: An overview of current approaches and future directions. Journal of Engineering. https://doi.org/10.1155/je/9928467
3. Ben Khalfallah, H., Jelassi, M., Demongeot, J., & Bellamine Ben Saoud, N. (2025). Advancements in predictive analytics: Machine learning approaches to estimating length of stay and mortality in sepsis. Computation, 13(1), 8. https://doi.org/10.3390/computation13010008
4. Broschat, S. L., Siu, S. W. I., & de la Fuente-Nunez, C. (2024). Editorial: Machine learning approaches to antimicrobials: Discovery and resistance. Frontiers in Bioinformatics, 4, 1458237. https://doi.org/10.3389/fbinf.2024.1458237
5. Caballero-María, P., Caballero-Villarraso, J., Arenas-Montes, J., Díaz-Cáceres, A., Castañeda-Nieto, S., Alcalá-Díaz, J. F., Delgado-Lista, J., Rodríguez-Cantalejo, F., Pérez-Martínez, P., López-Miranda, J., & Camargo, A. (2025). Deep learning model approach to predict diabetes type 2 based on clinical, biochemical, and gut microbiota profiles. Applied Sciences, 15(4), 2228. https://doi.org/10.3390/app15042228
6. Cardinali, L., Mariano, V., Rodriguez-Duarte, D. O., Tobón Vasquez, J. A., Scapaticci, R., Crocco, L., & Vipiana, F. (2025). Early detection of Alzheimer's disease via machine learning-based microwave sensing: An experimental validation. Sensors, 25(9), 2718. https://doi.org/10.3390/s25092718
7. Carriero, A., Groenhoff, L., Vologina, E., Basile, P., & Albera, M. (2024). Deep learning in breast cancer imaging: State of the art and recent advancements in early 2024. Diagnostics, 14, 848. https://doi.org/10.3390/diagnostics14080848
8. Ehsan, K., Sultan, K., Fatima, A., Sheraz, M., & Chuah, T. C. (2025). Early detection of autism spectrum disorder through automated machine learning. Diagnostics, 15(15), 1859. https://doi.org/10.3390/diagnostics15151859
9. Fahad, N., Rabbi, R. I., Hasan, S. B., Prity, F. S., Ahmed, R., Ahmed, F., Hossen, M. J., Liew, T. H., Sayeed, M. S., & Goh, K. O. M. (2025). Generative AI in clinical (2020-2025): A mini-review of applications, emerging trends, and clinical challenges. Frontiers in Digital Health, 7, 1653369. https://doi.org/10.3389/fdgth.2025.1653369
10. Frasca, M., La Torre, D., Pravettoni, G., & Cutica, I. (2024). Explainable and interpretable artificial intelligence in medicine: A systematic bibliometric review. Discover Artificial Intelligence, 4, 15. https://doi.org/10.1007/s44163-024-00114-7
11. Horasan, A., & Güneş, A. (2024). Advancing prostate cancer diagnosis: A deep learning approach for enhanced detection in MRI images. Diagnostics, 14, 1871. https://doi.org/10.3390/diagnostics14171871
12. Hussain, S. I., & Toscano, E. (2024). An extensive investigation into the use of machine learning tools and deep neural networks for the recognition of skin cancer: Challenges, future directions, and a comprehensive review. Symmetry, 16, 366. https://doi.org/10.3390/sym16030366

13. Khair, F. B., Saimon, A. S. M., Hossain, S., Manik, M. M. T. G., Ahmed, M. K., Islam, M. S., & Moniruzzaman, M. (2025). Deep neural network-based imaging system for efficient pancreatic tumor identification. Intelligent Decision Technologies. https://doi.org/10.1177/18724981251381581
14. Kiran, M., Xie, Y., Anjum, N., Ball, G., Pierscionek, B., & Russell, D. (2025). Machine learning and artificial intelligence in type 2 diabetes prediction: A comprehensive 33-year bibliometric and literature analysis. Frontiers in Digital Health, 7, 1557467. https://doi.org/10.3389/fdgth.2025.1557467
15. Kotei, E., & Thirunavukarasu, R. (2024). Tuberculosis detection from chest X-ray image modalities based on transformer and convolutional neural network. IEEE Access, 12, 97417-97427. https://doi.org/10.1109/ACCESS.2024.3428446
16. Li, Z., Li, Y., Mao, Z., Wang, C., Hou, J., Zhao, J., Wang, J., Tian, Y., & Li, L. (2025). Machine learning models integrating dietary indicators improve the prediction of progression from prediabetes to type 2 diabetes mellitus. Nutrients, 17(6), 947. https://doi.org/10.3390/nu17060947
17. Lin, M., Lin, L., Lin, L., Lin, Z., & Yan, X. (2025). A bibliometric analysis of the advance of artificial intelligence in medicine. Frontiers in Medicine, 12, 1504428. https://doi.org/10.3389/fmed.2025.1504428
18. Liu, Y.-Q., Chang, T.-W., Lee, L.-C., Chen, C.-Y., Hsu, P.-S., Tsan, Y.-T., Yang, C.-T., & Chu, W.-M. (2025). Use of machine learning to predict the incidence of type 2 diabetes among relatively healthy adults: A 10-year longitudinal study in Taiwan. Diagnostics, 15(1), 72. https://doi.org/10.3390/diagnostics15010072
19. Manik, M. M. T. G. (2025). Integrative analysis of heterogeneous cancer data using autoencoder neural networks. Journal of Information Systems Engineering and Management, 10(3s). https://doi.org/10.52783/jisem.v10i3s.4746
20. Manik, M. M. T. G., Khan, R., Das, S., Tasnim, A. F., Yeasmin, S., Semi, M. M. A., Sobur, A., Rob, M. A., & Samiun, M. (2026). Enhancing early autism diagnosis and personalized interventions in the US: AI-driven approaches with implications for special education policy and practice. In Proceedings of the 2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET). IEEE. https://doi.org/10.1109/ICECET63943.2025.11472491
21. Manik, M. M. T. G., Mohonta, S. C., Karim, F., Miah, M. A., Islam, M. S., Chy, M. A. R., Adnan, M., & Saimon, A. S. M. (2025). AI-driven precision medicine leveraging machine learning and big data analytics for genomics-based drug discovery. Journal of Posthumanism, 5(1). https://doi.org/10.63332/joph.v5i1.1993
22. Manik, M. M. T. G., Rahman, R., Tasnim, A. F., Nilima, S. I., Yeasmin, S., Das, S., & Akhter, F. (2026). RBXRT++: A new data-driven approach for detecting Alzheimer's and cognitive decline. In Proceedings of the 2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET). IEEE. https://doi.org/10.1109/ICECET63943.2025.11472184
23. Manik, M. M. T. G., Saimon, A. S. M., Ahmed, M. K., Hossain, S., Moniruzzaman, M., & Islam, M. S. (2025). Predictive modelling for early detection of type 2 diabetes using AI-driven machine learning algorithms and big data analytics. In International Conference on AI and Robotics (AIR 2025). Springer. https://doi.org/10.1007/978-3-032-05548-4_33
24. Nilima, S. I., Hossain, M. A., Sharmin, S., Rahman, R., Esa, H., & Manik, M. M. T. G. (2024). Advancement of drug discovery using artificial intelligence and machine learning. In 2024 IEEE International Conference on Computing, Applications and Systems (COMPAS). IEEE. https://doi.org/10.1109/COMPAS60761.2024.10796748
25. Samiun, M., Rony, M. K. K., Yeasmin, S., Manik, M. M. T. G., Debnath, A., Aziz, M. B., Tasnim, A. F., & Nilima, S. I. (2025). The role of artificial intelligence in managing hospitalized patients with mental illness: A scoping review. Discover Public Health, 22. https://doi.org/10.1186/s12982-025-00814-0
26. Sikder, T. R., Chy, M. A. R., Hossain, M. J., Uddin, S. M. M., Faruk, M. I., Manik, M. M. T. G., & Adnan, M. (2026). XACT-TB: An explainable hybrid CNN-Swin transformer framework for tuberculosis screening from chest X-ray images. Discover Artificial Intelligence. https://doi.org/10.1007/s44163-026-01509-4
27. Swinckels, L., Bennis, F. C., Ziesemer, K. A., Scheerman, J. F. M., Bijwaard, H., de Keijzer, A., & Bruers, J. J. (2024). The use of deep learning and machine learning on longitudinal electronic health records for the early detection and prevention of diseases: Scoping review. Journal of Medical Internet Research, 26, e48320. https://doi.org/10.2196/48320
28. Temiz, M., Bakir-Gungor, B., Ersoz, N. S., & Yousef, M. (2025). Machine learning-based prediction of autism spectrum disorder and discovery of related metagenomic biomarkers with explainable AI. Applied Sciences, 15(16), 9214. https://doi.org/10.3390/app15169214
29. Yousefi, M., Akhbari, M., Mohamadi, Z., Karami, S., Dasoomi, H., Atabi, A., Sarkeshikian, S. A., Abdoullahi Dehaki, M., Bayati, H., Mashayekhi, N., Varmazyar, S., Rahimian, Z., Asadi Anar, M., Shafiei, D., & Mohebbi, A. (2024). Machine learning based algorithms for virtual early detection and screening of neurodegenerative and neurocognitive disorders: A systematic review. Frontiers in Neurology, 15, 1413071. https://doi.org/10.3389/fneur.2024.1413071