

REVIEW

Open Access



Unlocking precision medicine: clinical applications of integrating health records, genetics, and immunology through artificial intelligence

Yi-Ming Chen^{1,2,3,4,5,6†}, Tzu-Hung Hsiao^{3,7,8†}, Ching-Heng Lin^{3,7,9,10*} and Yang C. Fann^{11*} 

Abstract

Artificial intelligence (AI) has emerged as a transformative force in precision medicine, revolutionizing the integration and analysis of health records, genetics, and immunology data. This comprehensive review explores the clinical applications of AI-driven analytics in unlocking personalized insights for patients with autoimmune rheumatic diseases. Through the synergistic approach of integrating AI across diverse data sets, clinicians gain a holistic view of patient health and potential risks. Machine learning models excel at identifying high-risk patients, predicting disease activity, and optimizing therapeutic strategies based on clinical, genomic, and immunological profiles. Deep learning techniques have significantly advanced variant calling, pathogenicity prediction, splicing analysis, and MHC-peptide binding predictions in genetics. AI-enabled immunology data analysis, including dimensionality reduction, cell population identification, and sample classification, provides unprecedented insights into complex immune responses. The review highlights real-world examples of AI-driven precision medicine platforms and clinical decision support tools in rheumatology. Evaluation of outcomes demonstrates the clinical benefits and impact of these approaches in revolutionizing patient care. However, challenges such as data quality, privacy, and clinician trust must be navigated for successful implementation. The future of precision medicine lies in the continued research, development, and clinical integration of AI-driven strategies to unlock personalized patient care and drive innovation in rheumatology.

Keywords Artificial intelligence, Precision medicine, Electronic health records, Genetics, Immunology, Autoimmune

[†]Yi-Ming Chen and Tzu-Hung Hsiao have contributed equally.

*Correspondence:

Ching-Heng Lin
epid@vghtc.gov.tw
Yang C. Fann
fann@ninds.nih.gov

¹ Division of Allergy, Immunology and Rheumatology, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung 40705, Taiwan

² School of Medicine, National Yang Ming Chiao Tung University, Taipei 11221, Taiwan

³ Department of Medical Research, Taichung Veterans General Hospital, Taichung 40705, Taiwan

⁴ Department of Post-Baccalaureate Medicine, College of Medicine, National Chung Hsing University, Taipei 112304, Taiwan

⁵ Graduate Institute of Clinical Medicine, College of Medicine, National Chung Hsing University, Taichung 402202, Taiwan

⁶ Precision Medicine Research Center, College of Medicine, National Chung Hsing University, Taichung 402202, Taiwan

⁷ Department of Public Health, College of Medicine, Fu Jen Catholic University, New Taipei City 242062, Taiwan

⁸ Institute of Genomics and Bioinformatics, National Chung Hsing University, Taichung 402202, Taiwan

⁹ Department of Industrial Engineering and Enterprise Information, Tunghai University, Taichung 407224, Taiwan

¹⁰ Institute of Public Health and Community Medicine Research Center, National Yang Ming Chiao Tung University, Taipei 11221, Taiwan

¹¹ Division of Intramural Research, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA



This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Precision medicine marks a pivotal shift in healthcare, steering away from the conventional 'one-size-fits-all' approach to a more personalized strategy tailored to individual patient profiles with immune disorders [1, 2]. This approach is deeply rooted in the integration of multi-dimensional data sources, including genetic information, immunological profiles, and extensive health records. The importance of integrating these diverse data types is magnified by the application of artificial intelligence (AI), which significantly enhances our ability to interpret complex biological data and translate it into clinically actionable insights [1].

AI, particularly through its subsets of machine learning (ML) and deep learning, is instrumental in dissecting the multi-omics and multifaceted datasets typical of modern healthcare systems [3]. These technologies excel at recognizing patterns and anomalies within large datasets, facilitating more precise and predictive healthcare. In genetics, AI algorithms are crucial for analyzing genetic markers quickly and accurately, offering insights into patient susceptibilities and potential responses to treatments [4]. This capability not only accelerates the process of genetic screening but also enhances the specificity of therapeutic interventions tailored to individual genetic profiles.

In the domain of immunology, AI's impact is equally transformative. AI models are adept at simulating immune system dynamics and predicting how it might respond to various treatments. This is particularly valuable in the design of personalized immunotherapies, which aim to optimize efficacy while minimizing adverse effects, thereby significantly advancing patient care in conditions such as autoimmune diseases and cancer immunotherapies [3, 5–7].

Moreover, the integration of AI with electronic health records (EHRs) provides a comprehensive overview of patient health histories, enhancing diagnostic accuracy and treatment efficacy [8]. AI-driven analysis of EHRs can identify hidden patterns across patient populations, predict disease progression, and suggest preventative measures tailored to individual health profiles [9]. This holistic approach not only improves individual patient outcomes but also contributes to broader public health strategies.

However, while the integration of AI into precision medicine offers substantial benefits, it also brings challenges such as ensuring data privacy, securing patient consent, and managing the ethical complexities associated with AI-driven decisions. The success of AI in healthcare hinges on navigating these challenges carefully, maintaining the integrity of patient data, and

ensuring that AI-driven interventions are transparent and equitable [10].

The aim of this comprehensive review is to explore the clinical applications of integrating health records, genetics, and immunology in precision medicine, facilitated by the advancements in AI, for patients with autoimmune rheumatic diseases (Fig. 1). This review will examine how AI-driven analytics can transform the landscape of precision medicine by offering more accurate prognostics and targeted therapeutic interventions.

AI-driven analysis of health records

AI has made significant strides in healthcare, offering innovative solutions to enhance patient care and streamline medical processes. We can see numerous successful applications of AI in healthcare, all built upon vast Electronic Health Records (EHR). Especially, deep learning has gained major attention due to its remarkable success in tasks like face recognition, image classification, object detection, and segmentation [11]. These advancements stem from deep learning models' ability to automatically learn intricate patterns in extensive datasets, enabling highly accurate and complex visual recognition. This success has spurred AI adoption across diverse fields. For example, Aidoc's AI platform transforms radiology departments by providing continuous, intelligent support, allowing radiologists to make faster and more accurate decisions. It analyzes CT scans to identify critical conditions like strokes, pulmonary embolisms, and brain hemorrhages, facilitating timely interventions [12]. Similarly, Google's AI system has achieved a breakthrough in diabetic retinopathy screening, examining retinal photos with accuracy comparable to ophthalmologists, enabling early detection and treatment to potentially prevent vision loss [13]. Beyond image recognition, AI requires real EHR data to support its judgments and remove noise; without this, AI cannot learn effectively.

In addition to imaging applications, AI also shows promise in drug discovery and epidemiological forecasting. For example, Atomwise's platform, AtomNet, uses deep learning to analyze molecular structures and predict effective molecules for drug development, potentially providing treatments for diseases like Ebola and multiple sclerosis [14]. BlueDot's AI platform, meanwhile, monitors global health trends to predict the spread of infectious diseases, offering insights that support proactive healthcare measures [15]. These models rely on collecting various types of EHR data and repeated long-term measurements.

The success of AI in healthcare can be attributed to its ability to process vast amounts of data swiftly and accurately. Before modeling, the application of AI in EHR requires several steps (Table 1):

AI-Driven Precision Medicine Integration Framework

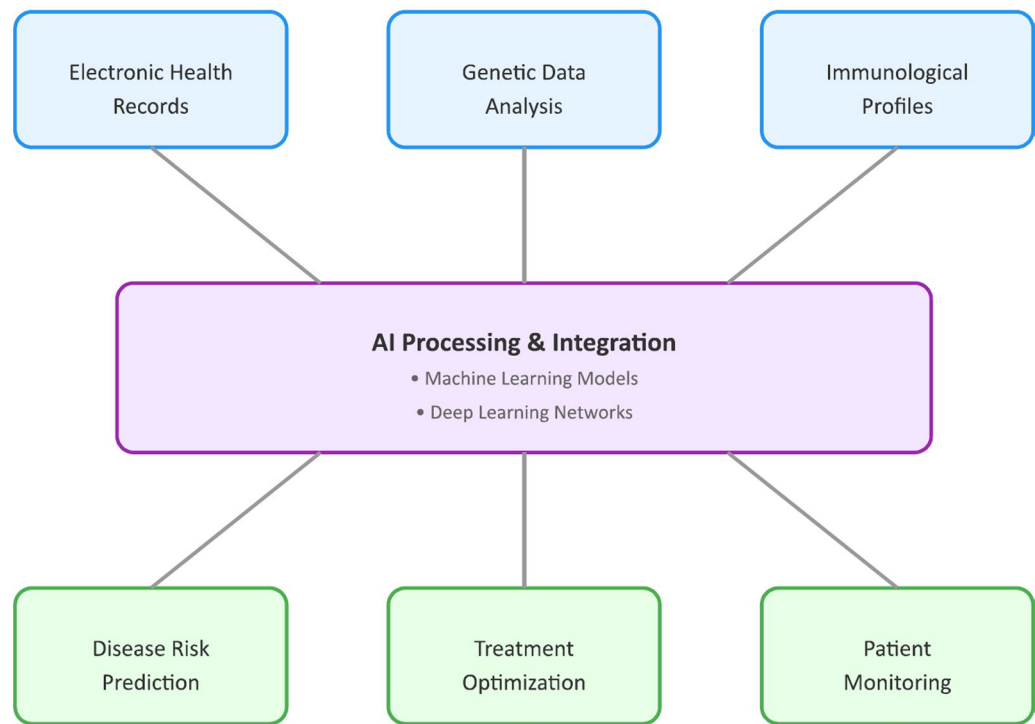


Fig. 1 Framework for AI-Driven Integration of Multi-modal Data in Precision Medicine

Table 1 The Steps for Applying AI to Process EHR

Step	Aim	Role of AI
Step 1: Data collection	Gather comprehensive data from multiple sources	ML algorithms extract information from unstructured EHR data. NLP models identify key terms
Step 2: Data cleaning	Address missing values, remove duplicates, and fix inconsistencies	AI corrects errors
Step 3: Normalization and standardization	Adjust data to a standard scale for comparability	AI models normalize and standardize quantitative data
Step 4: Data preservation	Maintain integrity and confidentiality of data	AI helps monitor and secure data throughout the process

Step 1: data collection

The process begins with data collection, where EHRs are aggregated from sources like hospitals, clinics, and healthcare providers. Leveraging AI, particularly ML, greatly enhances data collection from EHRs. ML algorithms can extract vital information from unstructured EHR data, such as clinical notes, lab reports, and radiology images. Natural Language Processing (NLP) models can identify key terms, diagnoses, medications, and patient demographics. This automation streamlines the data collection process and reduces the need for manual effort [16, 17].

Step 2: data cleaning

Following data collection, data cleaning is essential. This phase involves meticulously addressing missing values, eliminating duplicates, and rectifying any inconsistencies. AI can significantly enhance data cleaning in EHRs by correcting errors like misspelled units through fuzzy search and unit conversion, as well as using clinical context to clean numeric values and detect outliers. The goal is to achieve high automation while maintaining data quality and ensuring alignment with data governance policies [18].

Step 3: normalization and standardization

Normalization and standardization are subsequent steps that adjust numerical data to a common scale. This is especially important for lab results and other quantitative measures, ensuring that all data points are comparable and interpretable by AI models [19].

Step 4: data preservation

Maintaining the integrity and confidentiality of patient data is essential throughout this process.

Each of these steps plays a vital role in preparing data for AI processing. Together, they ensure that the data fed into AI algorithms is high-quality, structured appropriately, and contains the most relevant information for analysis.

Potential liability and reimbursement

The implementation of AI in EHR systems raises critical considerations regarding liability and reimbursement. From a liability perspective, the use of AI-driven clinical decision support tools creates complex questions about responsibility when adverse events occur. Healthcare providers must navigate situations where AI recommendations conflict with clinical judgment, requiring clear protocols for documentation and decision-making processes. Medical institutions need robust frameworks to determine liability distribution among healthcare providers, AI system developers, and healthcare organizations. This includes establishing clear audit trails for AI-assisted decisions and maintaining transparency in algorithmic decision-making processes. Regarding reimbursement, the integration of AI tools into clinical practice faces several challenges. Current healthcare payment systems are not fully equipped to handle AI-assisted services, with limited Current Procedural Terminology (CPT) codes specifically designed for AI applications. While some insurers have begun covering AI-assisted diagnostics, particularly in medical imaging, reimbursement policies for AI applications in clinical decision support remain inconsistent. Healthcare organizations must carefully evaluate the cost-effectiveness of AI implementations, considering both the initial investment and ongoing operational costs against potential reimbursement rates. Recent studies suggest that value-based care models may better accommodate AI-enabled services compared to traditional fee-for-service systems. The Centers for Medicare and Medicaid Services (CMS) has started developing frameworks for AI reimbursement, but standardization across different payers remains a significant challenge. These considerations underscore the need for healthcare organizations to develop comprehensive strategies addressing both liability protection and financial sustainability when implementing AI-driven EHR systems.

AI-powered genetics analysis

Genetics have undergone a revolutionary transformation in recent years, driven by rapid advancements in high-throughput sequencing technologies and the exponential growth of genomic data [20]. As those national-wide database, such as UK Biobank and All of US, demonstrated, the vast amount of information generated by these techniques has created an unprecedented opportunity for researchers to unravel the intricate relationships between genetic variations phenotypic traits and diseases, paving the way for personalized medicine and precision healthcare [21, 22]. However, the sheer volume and complexity of genomic data have also posed significant challenges in data analysis and interpretation [23, 24].

AI, a powerful computational paradigm that has shown remarkable prowess in extracting insights from large and complex datasets [25, 26]. AI techniques, such as machine learning and deep learning, have recently emerged as invaluable tools for navigating the genomic landscape, enabling researchers to identify patterns, make predictions, and uncover hidden relationships within the vast expanse of genetic information [4]. For instance, AI can help in analyzing genomic sequences to predict the impact of genetic variations on protein structure and pathogenicity, and pinpointing potential active pathway. Additionally, AI-driven tools are improving the efficiency and accuracy of genome techniques such as CRISPR and neoantigen identification, accelerating the pace of genetic research and the development of novel treatments for genetic disorders. In this review, we provide a comprehensive overview of the current state-of-the-art AI methodologies in variant calling and MHC-Peptide binding, highlighting the potential benefits of these approaches (Table 2).

Variant calling

Over the past two decades, Illumina sequencing technologies have dominated variant calling studies with their production of short reads. Numerous tools have been developed specifically for analyzing these short reads, including the Genome Analysis Toolkit (GATK) pipeline. GATK integrates a series of best practice workflows that begin with initial data processing, including quality control and alignment, and continue through variant calling, filtering, and annotation. This comprehensive approach ensures accurate and reliable analysis of genomic data.

Utilizing deep learning techniques, e.g. DeepVariant, excel in accurately identifying genetic variants from next-generation sequencing data. Unlike traditional methods, DeepVariant interprets sequencing data akin to images through convolutional neural networks (CNNs), allowing it to detect subtle genetic variations with exceptional precision. This tool is pivotal in enhancing

Table 2 Summary of AI-powered Genetics Analysis

Section	Description	Key AI Tools	Applications
Variant Calling	Tools such as DeepVariant use CNNs for high accuracy in detecting genetic variants in short and long reads, aiding fields like medical genetics and evolutionary biology	DeepVariant, DNAscope	Medical genetics, Evolutionary biology
Identification of Pathogenicity	AI models like AlphaMissense and PrimateAI-3D leverage protein structure for variant classification, improving pathogenicity prediction and variant reclassification in diagnostics	AlphaMissense, PrimateAI-3D, RENOVO, StrVCTVRE	Genetic diagnostics, Clinical variant interpretation
Splicing and Copy Number Variation	Non-coding variants such as splicing and CNVs are detected by tools like SpliceAI, which aid in genetic disorder diagnostics and personalized medicine by analyzing functional consequences	SpliceAI, DL-CNV, SVcnn	Genetic disorder diagnostics, Disease mechanism understanding
MHC-Peptide Binding Predictions	AI models like NetMHCpan-4.0 and BigMHC enhance predictions of peptide-MHC binding, crucial for immunotherapy and vaccine development, by using neural networks for improved accuracy	NetMHCpan-4.0, BigMHC, MHC-flurry-2.0	Immunotherapy, Vaccine development

variant calling accuracy for both single nucleotide polymorphisms (SNPs) and insertions/deletions (indels), thus significantly impacting fields such as medical genetics, personalized medicine, and evolutionary biology. It have showed outperform with other statistical based methods in the short reads [27].

Pacific Biosciences (PacBio) introduced a single-molecule real-time (SMRT) sequencing platform that can generate high-fidelity (HiFi) long reads with an average length of 13.5 kilobases (kb) using a Circular Consensus Sequence (CCS) approach. Oxford Nanopore Technologies (ONT) has changed the sequencing paradigm by introducing sequencers that are portable with real-time data delivery and are able to generate ultra-long reads. Again, DeepVariant is available for those long read variant calling [28, 29]. Also other AI-based variant callers such as DNAscope, which combines GATK’s Haplotype-Caller with a machine-learned genotyping model was developed for short reads and long reads [30].

Identification of pathogenicity

Identifying a pathogenic variant is a critical process in genetic diagnostics and involves a systematic approach combining multiple sources of evidence. The American College of Medical Genetics and Genomics (ACMG) provides a set of guidelines for variant classification, which includes evaluating population data, computational predictions, functional studies, segregation analysis, and de novo occurrences. Additionally, databases like ClinVar

are used to compare findings with existing classifications and supporting evidence [31, 32]. Still lots of variants are classified as a variant of uncertain significance (VUS) when there is insufficient scientific evidence to determine whether the genetic change is benign (harmless) or pathogenic (disease-causing). This uncertainty can arise due to factors such as the rarity of the variant and limited research studies.

AI models like AlphaMissense and PrimateAI-3D have demonstrated remarkable capabilities in predicting the pathogenicity of genetic variants based on structure and function of proteins and evolutionary conservation [33, 34]. AlphaMissense is a deep learning model that leverages protein structural information and evolutionary constraints from related sequences to accurately classify genetic variants as benign or pathogenic [33]. Similarly, PrimateAI-3D is a semi-supervised 3D convolutional neural network that operates on protein structures, enabling it to distinguish between benign and disease-causing variants [34]. SIGMA, also a deep learning model based on protein structural information have been proposed [35]. RENOVO and StrVCTVRE are machine learning tools that reclassify VUS [36, 37]. RENOVO uses random forests to classify variants as pathogenic or benign based on publicly available data, reclassifying 67% of ClinVar VUSs. StrVCTVRE distinguishes pathogenic from benign structural variants overlapping exons, allowing clinicians to prioritize novel SVs and resolve undiagnosed cases.

Those methods have demonstrated remarkable capabilities in predicting the pathogenicity of variants based on structural, evolutionary, and other informative features, enabling the reclassification of many VUS and aiding in the resolution of undiagnosed cases.

Splicing and copy number variation

Non-coding variants play a pivotal role in the development and progression of various diseases. Splicing variants, which disrupt the normal process of gene expression, can lead to the production of abnormal or non-functional proteins, contributing to genetic disorders and complex diseases. Copy number variations (CNVs), involving deletions or duplications of large DNA segments, can encompass entire genes or regulatory regions, causing gene dosage imbalances and altered expression levels. These changes can directly contribute to genetic disorders or modulate disease susceptibility and severity in complex conditions like autism, schizophrenia, and cancers. Unraveling the impacts of non-coding variants is crucial for understanding disease mechanisms, improving diagnostics, and developing targeted therapies in the era of personalized medicine. SpliceAI, a ML-based tool developed at the Broad Institute, is instrumental in this endeavor [38]. It predicts the effects of genetic variants on splicing, the process of removing non-coding regions from pre-messenger RNA. By analyzing genomic sequences and splicing data, SpliceAI identifies variants likely to disrupt or create new splice sites, leading to aberrant splicing and potential protein dysfunction. Considering genomic context, splice site sequences, and regulatory elements, it provides scores and predictions on the variant's impact on exon skipping, intron retention, or new splice site creation. This computational approach aids in interpreting non-coding and splice site variants, unveiling their functional consequences and potential involvement in genetic disorders and complex diseases, ultimately contributing to targeted therapies and personalized medicine strategies.

Several novel computational methods for detecting CNVs and structural variations (SVs) from next-generation sequencing data were developed. These include ifCNV, which uses isolation forests for CNV detection without a reference dataset [39]; DL-CNV, an alignment-free deep learning approach for targeted CNV detection [40]; and SVcnn, an accurate deep learning method that improves multi-allelic SV detection from long-read sequencing data [41]. With improved detection of genetic variants like CNVs and SVs, these computational methods facilitate genetic diagnosis, personalized medicine approaches, and understanding disease mechanisms. They overcome limitations of traditional methods, enabling reliable large-scale genomic studies to uncover

genetic contributors to diseases and advance precision medicine strategies.

Predictions MHC-peptide binding

The accurate prediction of peptide binding to MHC molecules is crucial for understanding the immune response and developing effective vaccines and immunotherapies. Traditional methods for MHC-peptide binding prediction, such as position-specific scoring matrices (PSSMs) and quantitative structure–activity relationship (QSAR) models, have limitations in capturing complex patterns and nonlinear interactions [42, 43]. Deep learning techniques, particularly neural networks, have emerged as powerful tools for addressing these challenges and improving the accuracy of MHC-peptide binding predictions. For example, NetMHCpan-4.0 is a widely used computational tool for predicting binding of peptides to any MHC class I molecule based on artificial neural networks [44]. BigMHC comprises an ensemble of seven deep neural networks trained on mass spectrometry data of peptides presented by MHC-I molecules. It also performed transfer learning on data from assays of antigen-specific immune response to further improve the prediction [45]. Other deep learning methods include MHCflurry-2.0 (an ensemble of neural networks) [46], MHCnuggets (many allele-specific LSTM networks for variable length peptides) [47], and MixMHCpred (using positional weight matrices to extract epitope motifs) [48]. The emergence of these advanced machine learning techniques has allowed for significant accuracy gains in MHC-peptide binding prediction.

Artificial intelligence has catalyzed a revolution in genetics through the implementation of machine learning and deep learning techniques, enabling highly accurate variant calling, pathogenicity prediction, splicing analysis, and MHC-peptide binding predictions. The applications of these novel AI methodologies have extended beyond these domains, encompassing the predictions of T cell receptor (TCR) and B cell receptor (BCR)-epitope interactions [49, 50], outperforming human experts in answering genetics-related questions through large language models [51], and facilitating tumor risk prediction and interpretation via multi-omics data integration [52]. Future research directions in this field include the integration of multi-modal data, the development of interpretable models, the implementation of privacy-preserving AI algorithms, and the exploration of generative models. The contributions of AI in genetics span a wide range of applications, including personalized medicine, drug discovery and development, disease prevention, evolutionary biology, and genome editing, driving the advancement of precision healthcare and enhancing

our comprehension of the intricate relationships between genotypes and phenotypes.

Medical utilization

Recent advancements in AI for genetics have profound implications for biomedicine, particularly in precision healthcare. The identification of genetic variants, which are instrumental in understanding disease susceptibility and patient-specific treatment response, has been enhanced through AI-driven techniques. By accurately identifying pathogenic variants, tools such as DeepVariant, AlphaMissense, and PrimateAI-3D enable clinicians and researchers to predict the functional impact of genetic mutations, facilitating targeted treatment approaches for genetic disorders. Similarly, AI's role in variant reclassification contributes to resolving VUS, allowing for clearer diagnostic interpretations that can impact clinical decision-making. In addition, AI-based splicing and CNV analysis models, including SpliceAI and DL-CNV, support the identification of structural and non-coding mutations, which are key to diagnosing complex diseases like cancer and neurodegenerative disorders. Furthermore, advances in MHC-peptide binding prediction, vital for immunotherapy and vaccine development, allow for a more nuanced understanding of immune responses. By integrating biological and clinical data, AI aids in uncovering novel therapeutic targets and improving the accuracy of disease models, ultimately propelling a shift towards personalized medicine that addresses unique patient genetic profiles.

AI-enabled immunology data analysis

AI has emerged as a pivotal tool in analyzing immunological data, significantly enhancing our understanding of complex immune responses. The use of AI, particularly in the processing and interpretation of high-dimensional data such as that obtained from flow cytometry and CyTOF (Cytometry by Time-Of-Flight), allows researchers to decipher intricate immune cell interactions and functionalities at an unprecedented scale and depth [53, 54]. ML significantly enhances the analysis of immunological data through various advanced techniques.

Dimensionality reduction

ML utilizes dimensionality reduction techniques, including Principle Component Analysis (PCA), t-Distributed Stochastic Neighbor Embedding (tSNE), UMAP to simplify the complex datasets generated by flow cytometry and CyTOF [55–57]. By reducing the number of variables while retaining essential information, ML algorithms can more effectively identify meaningful patterns within the data, facilitating a better understanding of immune cell behavior and interactions [53].

Cell population identification

ML methods facilitate in identifying distinct cell populations within high-dimensional immunological data based on their phenotypic and functional characteristics, allowing for precise mapping of immune cell heterogeneity [53]. Unsupervised and supervised machine learning methods are pivotal for cell-type identification in immunological studies. Unsupervised learning, such as K-means clustering and hierarchical clustering, groups cells based on their inherent similarities in multi-dimensional data sets without prior labels, revealing novel cell populations and states [58]. In contrast, supervised learning methods, including linear discriminant analysis (LDA) classifier and other algorithms of DGCyTOF and DeepCyTOF using neural network models for cell annotation, rely on labeled training data to classify cells into predefined types, improving the accuracy and robustness of cell-type identification [59]. Integrating both approaches enable a comprehensive understanding of cellular heterogeneity and dynamics, crucial for advancing immunological research and therapeutic development.

Sample classification

Sample classification in immunological research often leverages cell subset information and predictive modeling using single-cell data to enhance the precision of diagnostics and treatment strategies [53]. By utilizing cell subset information, ML-driven methods, including CITRUS and FloReMi, can categorize samples based on distinct immunological profiles, facilitating a deeper understanding of immune responses across different conditions, such as immune responses to cancer immunotherapy [60]. Predictive modeling of CellCNN and Deep CNN, using single-cell data, allows for the identification of unique cellular signatures that correlate with specific disease states or therapeutic responses [53, 61]. This approach improves the predictive accuracy for patient outcomes, enabling the development of personalized treatment plans that are tailored to individual immunological profiles.

AI for immunofluorescence image analysis

ML has significantly advanced the analysis of immunofluorescence images, offering enhanced accuracy and efficiency in identifying and classifying cellular patterns.

Our group developed an AI-based models to recognize competent-level antinuclear antibody (ANA) patterns and mixed patterns on human epithelial (HEp-2) cell images according to the International Consensus on ANA Patterns (ICAP) standards [62]. A large dataset of 51,694 HEp-2 cell images with patterns assigned by experienced technologists was used to train six deep convolutional neural network (CNN) architectures. The

InceptionResNetV2 model achieved the highest F1 score of 0.86 and kappa of 0.82 on a testing set, demonstrating excellent agreement with experienced human readers for recognizing 11 ICAP competent-level patterns as well as mixed patterns containing up to four overlapping patterns. This highlights the potential of applying transfer learning and fine-tuning publicly available pre-trained CNN models on large clinically-relevant datasets to develop robust automated systems for immunofluorescence image analysis and ANA pattern recognition.

Cytokine, chemokine and protein analysis

AI's applications in immunology extend to identifying immune system cell patterns, biomarkers, and therapeutic targets. For instance, ML models are increasingly used to analyze cytokine, chemokine profiles and protein microarray data, which are critical mediators of immunotherapeutic outcome [63]. Our group has demonstrated how ML algorithms can identify patterns and correlations within protein expression profiles that may not be evident through traditional analysis methods [64]. These algorithms, including support vector machines and random forests, can classify proteins based on their expression levels, predict potential biomarkers, and facilitate the understanding of immunogenicity for anti-tumor necrosis factor (TNF)- α biological therapy in patients with rheumatoid arthritis. By integrating AI with cytokine and chemokine data, researchers can now predict how different individuals' immune systems might react to specific diseases or treatments, facilitating the development of targeted therapies.

Overall, AI-driven analysis not only enhances the discovery of new biomarkers and therapeutic targets but also improves the accuracy and efficiency of immunological research, paving the way for advanced personalized treatments in immunology.

Integration of AI in precision medicine

Synergistic approach to integrating AI analysis of health records, genetics, and immunology data

The integration of AI into precision medicine is not just about leveraging AI in isolated clinical applications but about creating a synergistic approach that encompasses health records, genetics, and immunology data. By applying AI algorithms across these diverse data sets, clinicians can gain a holistic view of a patient's health status and potential risks. ML models are proving invaluable in identifying high-risk patients in need of rheumatology testing by analyzing EHRs. According to the study published in *Nature Communications* by Forrest et al., an ML model was developed to predict the necessity for autoimmune disease testing [65]. The model demonstrated high accuracy in identifying patients who should undergo

rheumatological evaluation by analyzing longitudinal EHR data from over 161,584 individuals. This approach allowed for earlier detection of the need for autoantibody testing and rheumatology encounters, identifying at-risk patients up to 5 years before traditional clinical assessments would typically do so, thereby accelerating diagnosis and treatment of autoimmune conditions.

Moreover, ML models can significantly enhance the diagnosis of systemic lupus erythematosus (SLE) by integrating EHRs, genomic data, and immunofluorescence image datasets. From our study published in *Bio-Data Mining*, ML approaches such as Random Forest and Extreme Gradient Boosting were utilized to analyze data from patients with positive ANA [66]. These models effectively identified SLE patients by evaluating clinical features, genomic variations, and specific patterns in immunofluorescence images. Furthermore, the study identified genetic variations associated with lupus in patients with high and low titer ANA, providing deeper mechanistic insights into the disease. The integration of these diverse data sources enables more accurate and timely diagnosis of SLE, facilitating better patient management and treatment outcomes. This AI-supported approach enables the development of highly personalized treatment plans that are predictive, preventive, personalized, and participatory (P4 medicine). Such multi-modal integration not only emphasizes the diagnostic process but also enhances the accuracy of treatment outcomes, making it a cornerstone of modern healthcare strategies [67].

Examples of AI-driven precision medicine platforms and tools for clinical decision support in rheumatology

Several AI-driven platforms and tools have been developed to assist healthcare providers in clinical decision-making [68, 69]. One notable example is IBM Watson Health, which analyzes the meaning and context of structured and unstructured data in clinical notes and reports to provide personalized treatment recommendations [70]. Another is Google's DeepMind Health, which has developed AI systems for analyzing medical images and electronic health records to support clinical decisions [71].

In the field of rheumatology, AI is being applied to enable more precise diagnosis and treatment selection for patients with rheumatic diseases. One example is an AI-driven precision medicine platform designed to quantify radiographic damage and optimize therapeutic strategies for RA patients [72]. It has demonstrated significant improvements in identifying effective treatments for individual patients, thereby enhancing personalized care. Researchers at Stanford have developed an AI platform called REVAMP that integrates multi-omics data

like genomics and proteomics to stratify patients with autoimmune diseases and match them to appropriate therapies [73]. Guan et al. discovered that ML algorithms could predict disease progression and response to therapy in RA patients taking anti-TNF- α inhibitors based on clinical and genomic data. By using AI to analyze complex biomedical information, these platforms clearly facilitate precision rheumatology and optimize care for patients based on their specific disease characteristics and molecular profiles.

Evaluation of real-world outcomes and clinical benefits of AI-driven precision medicine approaches

Evaluating the real-world outcomes and clinical benefits of AI-driven precision medicine approaches is crucial for their successful implementation in rheumatology practice. A study by Lezcano-Valverde et al. demonstrated that an ML model based on Random Survival Forests (RSF) could accurately predict the risk of mortality in patients with RA using clinical and serological data [74]. Another study by Myasoedova et al. showed that a supervised ML method could predict the response to methotrexate in patients with RA based on clinical and pharmacogenomic data, enabling clinicians to make informed decisions about treatment initiation and optimization [75]. These studies provide evidence for the clinical benefits and real-world impact of AI-driven precision medicine approaches in rheumatology with potential to revolutionize future patient care and outcomes.

Challenges and considerations

During the implementation of AI based on real world EHRs, several key challenges were encountered. Firstly, EHR data is often messy, inconsistent, and incomplete, making it crucial to ensure high-quality, standardized data for AI training. This requires significant preprocessing and cleaning efforts. For examples, the preprocessing steps to extract simple Blood Pressure (BP) values from clinical notes, structured fields, or imaging reports include standardizing all values to a consistent format like “120/80 mmHg” with uniform units. Outlier detection identifies extreme values (e.g., BP > 200/120 mmHg) for review, while missing BP readings are addressed by either imputing from patient history or neighboring values when appropriate. Quality assurance involves validating BP values against accepted ranges (e.g., systolic < 140 mmHg, diastolic < 90 mmHg), checking for temporal consistency (e.g., sudden BP spikes or drops), and detecting anomalies (e.g., negative BP values). Once cleaned and standardized, BP data can then be used for AI model training, enabling ML models to learn patterns, predict hypertension risk, and monitor patient health. This preprocessing approach

applies to various EHR data elements like labs, medications, and diagnoses, ensuring reliable AI training for better predictability and better patient outcomes.

Rigorous testing and validation are necessary to ensure AI algorithms perform accurately and safely in clinical settings, involving extensive training on diverse datasets and continuous monitoring for refinement. Best practices include using a diverse and representative dataset that mirrors real-world scenarios and includes varied demographics, contexts, and edge cases (e.g. under represented population). Continuous monitoring is necessary to track the AI model’s performance post-deployment, detecting any drift or degradation in accuracy over time. Additionally, bias detection and mitigation are essential; analyzing data for biases such as racial, gender, or socioeconomic disparities and implementing strategies to minimize their impact on model predictions is critical. Robust testing should simulate diverse scenarios and edge cases to verify that the model performs well across different situations. These practices help maintain high performance and reliability as well as trustworthy in AI algorithms [76, 77].

Secondly, balancing AI insights with patient privacy is essential. Protecting sensitive patient information while leveraging data for AI necessitates strict adherence to data security protocols and ethical guidelines, ensuring patient trust. Ensuring compliance with privacy regulations is vital for organizations. Best practices include staying informed by continuously monitoring changes in laws and regulations and educating your team about privacy requirements. Hiring data privacy professionals who understand regulations can effectively guide compliance efforts. Implementing controls such as regularly reviewing policies and procedures to ensure alignment with laws and being transparent about data handling practices is essential. Investing in robust security systems to protect sensitive information and limiting access to authorized personnel enhance data security. Establishing a process for reporting non-compliance and defining clear escalation paths for addressing violations ensure prompt and effective responses to any breaches. These practices collectively help maintain compliance with privacy regulations [78].

Thirdly, convincing healthcare professionals to trust and adopt AI recommendations is challenging. Developing user-friendly interfaces and clearly demonstrating the benefits of AI, such as improved diagnostic accuracy and workflow efficiency, are crucial for gaining clinician acceptance and integration into their daily practice. Addressing these challenges is essential for the successful implementation and utilization of AI in EHR systems, ultimately enhancing healthcare delivery and patient outcomes [79].

Finally, while AI holds great promise in clinical decision support, it's essential to address potential risks and ethical concerns. Recent studies have highlighted several critical considerations [80, 81]: (1) The risk of AI systems perpetuating or amplifying existing healthcare disparities if trained on biased datasets, (2) The potential for over-reliance on AI recommendations leading to decreased clinical autonomy and judgment, (3) The challenge of maintaining transparency in AI decision-making processes, particularly with complex 'black box' algorithms, and (4) The need for clear accountability frameworks when AI systems contribute to clinical decisions. These concerns necessitate careful implementation strategies that prioritize patient safety, maintain physician autonomy, and ensure equitable care delivery. Healthcare organizations must establish clear protocols for AI system validation, regular performance monitoring, and mechanisms to detect and address potential biases or errors in AI recommendations. Additionally, clinicians should be trained to appropriately interpret and critically evaluate AI-generated suggestions within the context of their clinical expertise and patient-specific factors.

Conclusion

In conclusion, this review highlights the transformative potential of AI-driven precision medicine, integrating health records, genetics, and immunology data. By analyzing vast and complex datasets, AI enables unprecedented insights into disease mechanisms and patient-specific responses, facilitating accurate diagnoses and personalized treatments. To fully utilize AI's potential in revolutionizing healthcare, continued research, development, and clinical integration are crucial. Investing in and adopting AI-driven precision medicine strategies will lead to a new era of personalized patient care and medical breakthroughs, ultimately improving outcomes and driving innovation in rheumatology.

Acknowledgements

We apologize to the many investigators whose important studies could not be cited here because of space limitations. YCF is supported by the Office of Intramural Research, Office of the Director and Division of Intramural Research of National Institute of Neurological Disorders and Stroke, National Institutes of Health, USA.

Author contributions

Y.M.C, T.H.H, C.H.L wrote and drafted the manuscript. Y.C.F. reviewed and revised the manuscript. T.H.H and C.H.L retrieved literature. Y.M.C designed this manuscript.

Funding

This research was supported by funding from Academia Sinica 40-05-GMM and AS-GC-110-MD02, National Science and Technology Council, Taiwan [NSTC -111-2634-F-A49-014, NSTC-111-2218-E-039-001, and NSTC-111-2314-B-075A-003-MY3], and Taichung Veterans General Hospital, Taiwan [TCVGH-1137310C, TCVGH-1137319C, TCVGH-1137302D, TCVGH-1127304B and TCVGH-1137302B].

Availability of data and materials

The datasets during and/or analysed during the current study available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 28 June 2024 Accepted: 2 December 2024

Published online: 07 February 2025

References

- Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med*. 2015;372(9):793–5.
- Momtazmanesh S, Nowroozi A, Rezaei N. Artificial Intelligence in rheumatoid arthritis: current status and future perspectives: a state-of-the-art review. *Rheumatol Ther*. 2022;9(5):1249–304.
- Prelaj A, Miskovic V, Zanitti M, et al. Artificial intelligence for predictive biomarker discovery in immuno-oncology: a systematic review. *Ann Oncol*. 2024;35(1):29–65.
- Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. *Nat Rev Genet*. 2015;16(6):321–32.
- Kingsmore KM, Puglisi CE, Grammer AC, Lipsky PE. An introduction to machine learning and analysis of its use in rheumatic diseases. *Nat Rev Rheumatol*. 2021;17(12):710–30.
- Fasano S, Milone A, Nicoletti GF, Isenberg DA, Ciccia F. Precision medicine in systemic lupus erythematosus. *Nat Rev Rheumatol*. 2023;19(6):331–42.
- Lin CMA, Cooles FAH, Isaacs JD. Precision medicine: the precision gap in rheumatic disease. *Nat Rev Rheumatol*. 2022;18(12):725–33.
- Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med*. 2018;1:18. <https://doi.org/10.1038/s41746-018-0029-1> Published 2018 May 8.
- Steinfeldt J, Wild B, Buerger T, et al. Medical history predicts phenome-wide disease onset and enables the rapid response to emerging health threats. *Nat Commun*. 2024;15(1):4257. <https://doi.org/10.1038/s41467-024-48568-8> Published 2024 May 20.
- Char DS, Shah NH, Magnus D. Implementing machine learning in health care-addressing ethical challenges. *N Engl J Med*. 2018;378(11):981–3.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–44.
- Aidoc's AI platform. <https://www.aidoc.com/platform/>.
- Yim J, Chopra R, Spitz T, et al. Predicting conversion to wet age-related macular degeneration using deep learning. *Nat Med*. 2020;26(6):892–9.
- Atomwise's platform. <https://www.atomwise.com/>.
- BlueDot's AI platform. <https://bluedot.global/>.
- Dashtban A, Mizani MA, Pasea L, et al. Identifying subtypes of chronic kidney disease with machine learning: development, internal validation and prognostic validation using linked electronic health records in 350,067 individuals. *EBioMedicine*. 2023;89: 104489.
- Zech J, Pain M, Titano J, Badgeley M, Schefflein J, Su A, et al. Natural language-based machine learning models for the annotation of clinical radiology reports. *Radiology*. 2018;287:570–80.
- Shi X, Prins C, Van Pottelbergh G, Mamouris P, Vaes B, De Moor B. An automated data cleaning method for Electronic Health Records by incorporating clinical knowledge. *BMC Med Inform Decis Mak*. 2021;21(1):267.
- Reich C, Ostropelets A, Ryan P, et al. OHDSI Standardized Vocabularies—a large-scale centralized reference ontology for international data harmonization. *J Am Med Inform Assoc*. 2024;31(3):583–590.
- Mardis ER. DNA sequencing technologies: 2006–2016. *Nat Protoc*. 2017;12(2):213–8.

21. All of Us Research Program Genomics Investigators. Genomic data in the All of Us Research Program. *Nature*. 2024;627(8003):340–346.
22. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203–9.
23. Gómez-López G, Dopazo J, Cigudosa JC, Valencia A, Al-Shahrour F. Precision medicine needs pioneering clinical bioinformaticians. *Brief Bioinform*. 2019;20(3):752–66.
24. Duffy DJ. Problems, challenges and promises: perspectives on precision medicine. *Brief Bioinform*. 2016;17(3):494–504.
25. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med*. 2022;28(1):31–8.
26. Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface*. 2018;15(141):20170387.
27. Poplin R, Chang PC, Alexander D, et al. A universal SNP and small-indel variant caller using deep neural networks. *Nat Biotechnol*. 2018;36(10):983–7.
28. Abdelwahab O, Belzile F, Torkamaneh D. Performance analysis of conventional and AI-based variant callers using short and long reads. *BMC Bioinformatics*. 2023;24(1):472.
29. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol*. 2019;20(1):129.
30. Freed D, Pan R, Chen H, Li Z, Hu J, Aldana R. DNAScope: High accuracy small variant calling using machine learning. *bioRxiv* 2022.2022.2005.2020.492556.
31. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. *Genet Med*. 2015;17(5):405–24.
32. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*. 2016;44(D1):D862–8.
33. Cheng J, Novati G, Pan J, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science*. 2023;381(6664):eadg7492.
34. Gao H, Hamp T, Ede J, Schraiber JG, McRae J, Singer-Berk M, Yang Y, Dietrich ASD, Fizev PP, Kuderna LFK et al. The landscape of tolerated genetic variation in humans and primates. *Science* 2023, 380(6648):eabn8153.
35. Zhao H, Du H, Zhao S, Chen Z, Li Y, Xu K, Liu B, Cheng X, Wen W, Li G, et al. SIGMA leverages protein structural information to predict the pathogenicity of missense variants. *Cell Rep Methods*. 2024;4(1): 100687.
36. Favalli V, Tini G, Bonetti E, Vozza G, Guida A, Gandini S, Pelicci PG, Mazarella L. Machine learning-based reclassification of germline variants of unknown significance: the RENOVO algorithm. *Am J Human Genet*. 2021;108(4):682–95.
37. Sharo AG, Hu Z, Sunyaev SR, Brenner SE. StrVCTVRE: a supervised learning method to predict the pathogenicity of human genome structural variants. *Am J Human Genet*. 2022;109(2):195–209.
38. Jaganathan K, Panagiotopoulou SK, McRae JF, Darbandi SF, Knowles D, Li YI, Kosmicki JA, Arbelaez J, Cui W, Schwartz GB. Predicting splicing from primary sequence with deep learning. *Cell* 2019, 176(3):535–548. e524.
39. Cabello-Aguilar S, Vendrell JA, Van Goethem C, Brousse M, Gozé C, Frantz L, Solassol J. ifCNV: a novel isolation-forest-based package to detect copy-number variations from various targeted NGS datasets. *Molecular Therapy-Nucleic Acids*. 2022;30:174–83.
40. Zhang YX, Jin LC, Wang B, Hu D, Wang L, Li P, Zhang J, Han K, Tian G, Yuan D. DL-CNV: a deep learning method for identifying copy number variations based on next generation target sequencing. *Math Biosci Eng*. 2020;17:202–15.
41. Zheng Y, Shang X. SVcnn: an accurate deep learning-based method for detecting structural variation based on long-read data. *BMC Bioinformatics*. 2023;24(1):213.
42. Liao WW, Arthur JW. Predicting peptide binding to major histocompatibility complex molecules. *Autoimmun Rev*. 2011;10(8):469–73.
43. Tian F, Yang L, Lv F, Yang Q, Zhou P. In silico quantitative prediction of peptides binding affinity to human MHC molecule: an intuitive quantitative structure-activity relationship approach. *Amino Acids*. 2009;36(3):535–54.
44. Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M: NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol*. 2017;199(9):3360–3368.
45. Albert BA, Yang Y, Shao XM, Singh D, Smit KN, Anagnostou V, Karchin R. Deep neural networks predict class I major histocompatibility complex epitope presentation and transfer learn neoepitope immunogenicity. *Nat Mach Intell*. 2023;5(8):861–72.
46. O'Donnell TJ, Rubinsteyn A, Laserson U: MHCflurry 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing. *Cell Syst*. 2020;11(4):418–419.
47. Shao XM, Bhattacharya R, Huang J, Sivakumar IKA, Tokheim C, Zheng L, Hirsch D, Kaminow B, Omdahl A, Bonsack M, et al. High-Throughput Prediction of MHC Class I and II Neoantigens with MHCnuggets. *Cancer Immunol Res*. 2020;8(3):396–408.
48. Gfeller D, Schmidt J, Croce G, Guillaume P, Bobisse S, Genolet R, Queiroz L, Cesbron J, Racle J, Harari A: Improved predictions of antigen presentation and TCR recognition with MixMHCpred2.2 and PRIME2.0 reveal potent SARS-CoV-2 CD8(+) T-cell epitopes. *Cell Syst*. 2023;14(1):72–83 e75.
49. Fast E, Dhar M, Chen B: TAPIR: a T-cell receptor language model for predicting rare and novel targets. *bioRxiv* 2023.
50. Wang M, Patsenker J, Li H, Kluger Y, Kleinstein SH. Language model-based B cell receptor sequence embeddings can effectively encode receptor specificity. *Nucleic Acids Res*. 2024;52(2):548–57.
51. Duong D, Solomon BD. Analysis of large-language model versus human performance for genetics questions. *Eur J Hum Genet*. 2024;32(4):466–8.
52. Yan H, Weng D, Li D, Gu Y, Ma W, Liu Q. Prior knowledge-guided multilevel graph neural network for tumor risk prediction and interpretation via multi-omics data integration. *Brief Bioinform*. 2024;25(3): bbae184.
53. Hu Z, Bhattacharya S, Butte AJ. Application of machine learning for cytometry data. *Front Immunol*. 2022;12: 787574.
54. Greene E, Finak G, D'Amico LA, et al. New interpretable machine-learning method for single-cell data reveals correlates of clinical response to cancer immunotherapy. *Patterns (N Y)*. 2021;2(12): 100372.
55. Becht E, McInnes L, Healy J, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*. Published online December 3, 2018. <https://doi.org/10.1038/nbt.4314>
56. Ding J, Condon A, Shah SP. Interpretable dimensionality reduction of single cell transcriptome data with deep generative models. *Nat Commun*. 2018;9(1):2002.
57. Van Der Maaten L. Accelerating T-SNE using tree-based algorithms. *J Mach Learn Res*. 2014;15:3221–45.
58. Cheung M, Campbell JJ, Whitby L, Thomas RJ, Braybrook J, Petzing J. Current trends in flow cytometry automated data analysis software. *Cytometry A*. 2021;99(10):1007–21. <https://doi.org/10.1002/cyto.a.24320>.
59. Li H, Shaham U, Stanton KP, Yao Y, Montgomery RR, Kluger Y. Gating mass cytometry data by deep learning. *Bioinformatics*. 2017;33:3423–30.
60. Spitzer MH, Carmi Y, Reticker-Flynn NE, et al. Systemic immunity is required for effective cancer immunotherapy. *Cell*. 2017;168(3):487–502. e15.
61. Arvaniti E, Claassen M. Sensitive detection of rare disease-associated cell subsets via representation learning. *Nat Commun*. 2017;8:14825.
62. Wu YD, Sheu RK, Chung CW, et al. Application of supervised machine learning to recognize competent level and mixed antinuclear antibody patterns based on ICAP international consensus. *Diagnostics (Basel)*. 2021;11(4):642.
63. Wei F, Azuma K, Nakahara Y, et al. Machine learning for prediction of immunotherapeutic outcome in non-small-cell lung cancer based on circulating cytokine signatures. *J Immunother Cancer*. 2023;11(7): e006788.
64. Chen PK, Lan JL, Chen YM, et al. Anti-TROVE2 antibody determined by immune-related array may serve as a predictive marker for adalimumab immunogenicity and effectiveness in RA. *J Immunol Res*. 2021;2021:6656121.
65. Forrest IS, Petrazzini BO, Duffy Á, et al. A machine learning model identifies patients in need of autoimmune disease testing using electronic health records. *Nat Commun*. 2023;14(1):2385.
66. Chung CW, Chou SC, Hsiao TH, Zhang GJ, Chung YF, Chen YM. Machine learning approaches to identify systemic lupus erythematosus in anti-nuclear antibody-positive patients using genomic data and electronic health records. *BioData Min*. 2024;17(1):1.
67. Martínez-García M, Hernández-Lemus E. Data integration challenges for machine learning in precision medicine. *Front Med (Lausanne)*. 2022;8: 784455.
68. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44–56.

69. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med*. 2018;1:18.
70. Xu J, Yang P, Xue S, et al. Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges and future perspectives. *Hum Genet*. 2019;138(2):109–24.
71. Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. *Health Technol (Berl)*. 2017;7(4):351–67.
72. Sun D, Nguyen TM, Allaway RJ, et al. A crowdsourcing approach to develop machine learning models to quantify radiographic joint damage in rheumatoid arthritis. *JAMA Netw Open*. 2022;5(8): e2227423.
73. Chen B, Khodadoust MS, Olsson N, et al. Predicting HLA class II antigen presentation through integrated deep learning. *Nat Biotechnol*. 2019;37(11):1332–43.
74. Lezcano-Valverde JM, Salazar F, León L, et al. Development and validation of a multivariate predictive model for rheumatoid arthritis mortality using a machine learning approach. *Sci Rep*. 2017;7(1):10189.
75. Myasoedova E, Athreya AP, Crowson CS, et al. Toward individualized prediction of response to methotrexate in early rheumatoid arthritis: a pharmacogenomics-driven machine learning approach. *Arthritis Care Res (Hoboken)*. 2022;74(6):879–88.
76. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in health-care: transforming the practice of medicine. *Future Healthc J*. 2021;8(2):e188–94.
77. Bideault G, Scaccia A, Zahel T, et al. Verification and validation of computational models used in biopharmaceutical manufacturing: potential application of the ASME verification and validation 40 standard and FDA Proposed AI/ML Model Life Cycle Management Framework. *J Pharm Sci*. 2021;110(4):1540–4.
78. Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Med Ethics*. 2021;22:122.
79. Isaacks DB, Borkowski AA. Implementing trustworthy AI in VA high reliability health care organizations. *Fed Pract*. 2024;41(2):40–3.
80. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. *N Engl J Med*. 2023;388(13):1201–8.
81. Bhinder B, Gilvary C, Madhukar NS, Elemento O. Artificial intelligence in cancer research and precision medicine. *Cancer Discov*. 2021;11(4):900–15.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.