

105682_Auto_Edited.docx

WORD COUNT

2171

TIME SUBMITTED

10-APR-2025 11:09AM

PAPER ID

115615419

Name of Journal: *Artificial Intelligence in Gastroenterology*

Manuscript NO: 105682

Manuscript Type: MINIREVIEWS

Precision Medicine in Celiac Disease: A step ahead !

Precision Medicine in Celiac Disease

Hakim Rahmoune, Nada Boutrid, Isra Benchoufi

Abstract

Celiac disease (CD) is a common autoimmune disorder where gluten ingestion triggers an immune response, damaging the small intestine in genetically predisposed individuals. Affecting around 1% of the global population, CD presents with diverse symptoms, including gastrointestinal issues like diarrhea and extraintestinal conditions such as anemia and osteoporosis, often complicating diagnosis. Advances in serology, histology, and genetic testing, such as HLA-DQ2/DQ8 analysis, have improved diagnostic accuracy.

Precision medicine is transforming CD management by integrating genetic, clinical, and lifestyle data to enable risk prediction, personalized therapies, and improved outcomes. Tools like machine learning (ML) enhance early diagnosis, dietary management, and drug discovery, while electronic medical records (EMRs) support comprehensive patient profiling and disease monitoring. These technologies facilitate personalized healthcare delivery tailored to individual patient profiles.

Key Words: Artificial Intelligence; Precision Medicine; Machine Learning; Electronic Medical Records; Celiac Disease

Rahmoune H, Boutrid N, Benchoufi I. Precision Medicine in Celiac Disease: a step ahead!. *Artif Intell Gastroenterol* 2025; In press

Core Tip: Advances in precision medicine, powered by artificial intelligence (AI), machine learning (ML) and deep learning (DL), are transforming the diagnosis and management of celiac disease (CD). By integrating histopathologic features, genetic data, and electronic medical records (EMRs), these technologies enable accurate disease classification, risk prediction, and personalized treatment strategies. ML and DL models have demonstrated high accuracy in predicting CD severity, while EMR-based approaches facilitate tailored dietary and therapeutic interventions. These innovations

underscore the potential of AI-driven methodologies to enhance patient outcomes and revolutionize the clinical approach to CD.

³ INTRODUCTION

Celiac disease (CD) is a chronic immune-mediated enteropathy triggered by the ingestion of gluten—a protein found in wheat, barley, and rye—in genetically predisposed individuals (1). It is estimated to affect approximately 1% of the global population, making it one of the most common autoimmune disorders worldwide (2,3).

The clinical manifestations of CD are highly variable and can be classified into gastrointestinal and extraintestinal symptoms. Gastrointestinal symptoms include diarrhea, abdominal pain, and bloating, while extraintestinal manifestations encompass conditions like anemia, osteoporosis, infertility, and neurological disorders such as ataxia and peripheral neuropathy. This wide spectrum of symptoms differ not only in type (digestive vs non-digestive) but also in severity, ranging from fatigue and mild discomfort to severe malabsorption syndromes ; and these presentations of CD often overlaps with other gastrointestinal and systemic disorders, further complicating timely diagnosis (4,5).

The diagnosis of CD relies on a combination of clinical assessment, serological testing for antibodies (e.g., anti-tissue transglutaminase and anti-endomysial antibodies), and histological examination of duodenal biopsies, which typically reveal villous atrophy, crypt hyperplasia, and intraepithelial lymphocytosis (6,7). Advances in non-invasive diagnostic tools, including genetic testing for human leucocyte antigen (HLA) alleles HLA-DQ2 and HLA-DQ8, have further improved the accuracy of CD diagnosis. A meta-analysis with systematic review was published including 24 studies on the association between HLA-DQB102 gene doses and the clinical and histological characteristics of CD (including clinical presentation, histology, age at diagnosis, and comorbidities). Classical CD (digestive form) was more prevalent in patients with a double dose of HLA-DQB102 vs a single allele (Odds Ratio, OR = 1.758) (8).

The prevalence of CD shows significant geographic variability. For instance, the Saharawi refugee population in Algeria has the highest recorded prevalence, affecting nearly 6% of the population (9). In contrast, in Asia, the seroprevalence and prevalence are estimated to be 1.6% and 0.5%, respectively(10) . Similarly, in Latin America, the prevalence ranges between 0.46% and 0.64%, with Colombians showing the lowest rates(11).

These differences may be attributed to genetic, environmental, and dietary factors, as well as variations in diagnostic practices and awareness levels.

The recent advent of artificial intelligence (AI) is revolutionizing CD management by enabling more accurate diagnosis, risk prediction, and personalized treatment strategies. These advanced technologies represent a paradigm shift in how we approach this complex autoimmune disorder (12-16): AI-driven models, leveraging machine learning (ML) and deep learning (DL), can analyze large-scale genomic, serological, and clinical data to improve diagnostic accuracy and identify at-risk individuals even before symptom onset.

2. Precision Medicine in Celiac Disease

AI on precision medicine, whereby to facilitate the gene and environment determinants of gluten intolerance identification during digestive diagnosis and predictive algorithms for improved disease subtype (*i.e.* genotypes) classifications or treatments outcomes forecasts to personalize recommendations (17).

Precision medicine refers to a broad approach for tailoring healthcare based on individual variability in genes, environment, and lifestyle factors. Recent technological developments have greatly improved the precision medicine approaches for CD: Next-generation sequencing is establishing comprehensive genomic profiles; proteomic technologies enables high-throughput biomarker discovery; and multiplex serological assays eases simultaneous assessment of various autoantibodies providing a more nuanced understanding of the immune responses in CD (18-20).

Actually, precision medicine would optimize the diagnosis, treatment, and prevention of diseases like CD (14–16) through three main goals:

1. Risk Prediction: Identifying individuals at high risk of developing CD through genetic and serological screening.

2. Outcome Prediction: Predicting disease progression and potential complications using advanced modeling techniques.

3. Personalized Therapy: Developing tailored management plans based on individual clinical profiles and genetic information.

Recent advances in technology, including the use of machine learning (ML) and electronic medical records (EMRs), would facilitate the application of precision medicine in CD by integrating diverse data types such as genetic markers, clinical records, and lifestyle factors(21).

Beyond these topics related to precision medicine and machine learning, artificial intelligence has also been used in CD to utilize different data modalities, including genomic data, EMR, and textual data, to provide personalized therapy (15).

3. EMR and Precision Medicine in CD

Electronic medical records (EMRs) are digital systems that store patient health information, including medical history, laboratory results, imaging studies, and treatment outcomes (22) .

Integrating AI with EMRs can deliver real-time, data-driven recommendations and provide a rich resource for advancing precision medicine in CD by enabling data integration, risk stratification and patients monitoring:

- Data Integration: Combining genetic data, biomarkers, and clinical information to create a comprehensive patient profile. This integration supports more accurate diagnosis and personalized management (23) .

- Risk Stratification: Identifying individuals at high risk for CD through analysis of EMR data (24) , including abnormal laboratory results (*i.e.* genetics or autoimmunity) or patient's patterns (age, sex, clinical symptoms, associated autoimmunity...etc).

- Follow-up Monitoring: Tracking disease progression and response to therapy over time (e.g., gluten-free diet, micro- and macro-nutrient assessments, and supplementation) and guiding adjustments in management plans (25).

5 Indeed, a pioneer study from the Mayo clinic (Rochester, Minnesota, USA) showed that computerized EMR-based algorithms can help identify patients at high risk of CD, and identified 6 individuals at high risk of CD and those who would benefit from CD screening (26).

Finally, machine learning models can process vast amounts of EMR data to predict disease onset, detect early symptoms, and recommend targeted interventions. This computational approach transforms routine clinical data into actionable insights that support precision medicine initiatives. The integration of EMR systems with ML algorithms creates a powerful framework for precision medicine in CD. EMRs serve as the foundation by providing structured clinical data, while ML techniques analyze these datasets to identify patterns and make predictions. For example, natural language processing applications can extract relevant information from clinical notes, radiology reports, and pathology findings to create comprehensive patient profiles that inform personalized treatment decisions (27).

4 ML and Precision Medicine in CD

4 Machine learning (ML) is a subset of artificial intelligence that uses algorithms to analyze complex datasets and make predictions or decisions without explicit programming (15,26) .

In the specific context of CD, evidence has proven that ML is efficient in several ways:

1. Early Diagnosis and Risk Assessment: ML models can analyze clinical, genetic, and environmental data to predict the likelihood of developing CD, facilitating earlier diagnosis and intervention (28). Recent studies demonstrate that ML models achieve up to 90% accuracy in predicting CD development in high-risk populations, compared to 64% accuracy with traditional risk assessment methods (29). A recent convolutional neural network applied to biopsy images showed an accuracy of predicting celiac disease at 99.97% (30)

2. Genetic Analysis: ML algorithms can identify genetic markers associated with CD, such as HLA-DQ2 and HLA-DQ8, and predict an individual's genetic susceptibility (31).

3. Subtype Identification: CD presents in various forms, including classical, atypical, and potential CD. ML techniques can analyze patient data to differentiate subtypes, enabling tailored treatment plans (15) .

4. Predicting Disease Progression: By analyzing longitudinal datasets, ML models can predict disease progression, guiding clinicians in monitoring and managing patients effectively (32).

5. Dietary Management: For CD patients, adhering to a gluten-free diet is critical, and ML-powered tools can easily identify hidden gluten in food products, recommend safe alternatives, and provide personalized dietary advice (33) . **Advanced ML techniques such as computer vision and natural language processing are being deployed in smartphone applications that can analyze food labels or restaurant menu descriptions to detect potential gluten content. These applications use deep learning models trained on extensive food databases to identify ingredients that might contain hidden gluten with over 90% accuracy, providing real-time guidance for patients managing strict gluten-free diets (34,35).**

6. Patient Stratification: ML can stratify patients based on disease severity, response to treatment, and risk factors, optimizing resource allocation and therapeutic strategies (36) .

7. Drug Discovery: ML can analyze big datasets to identify new drug targets, predict responses to existing treatments, and accelerate the development of novel therapies for CD (37) .

8. Patient Support and Education: ML-driven applications, such as chatbots or mobile apps, offer CD patients personalized information, emotional support, and dietary tips, improving their quality of life (38).

Table 1: Summary of ML roles and precision medicine in CD

ML role in CD

Description

1. Early Diagnosis

ML models predict CD likelihood using clinical and genetic data, enabling earlier intervention.

2. Genetic Analysis

Algorithms identify genetic markers (HLA-DQ2/DQ8) to assess susceptibility to CD.

3. Subtype Identification

ML differentiates between CD subtypes for personalized treatment plans.

4. Disease Progression Prediction

Analyzing datasets helps predict disease progression and guide patient management.

5. Dietary Management

Identifying gluten in foods and providing personalized dietary recommendations.

6. Patient Stratification

ML stratifies patients by severity and treatment response, optimizing care strategies.

7. Drug Discovery

Analyzing data to identify new drug targets and predict treatment responses for CD therapies.

8. Patient Support

ML applications offer personalized information and support, enhancing patient quality of life.

The success of ML models depends on the availability of high-quality, diverse datasets that reduce biases and ensure accuracy. Effective clinical integration of ML also

requires rigorous validation, ongoing monitoring, and robust oversight to ensure models deliver safe, reliable, and equitable care (12,38) .

The application of ML in CD leverages various algorithmic approaches tailored to specific clinical questions:

Supervised learning methods, particularly support vector machines and random forests, have demonstrated particular utility in classifying CD subtypes based on clinical features. Meanwhile, deep learning neural networks excel at analyzing complex histopathological images to detect subtle tissue changes indicative of CD (30,32).

Unsupervised learning techniques, including clustering algorithms, have successfully identified novel patient subgroups with distinct clinical trajectories and treatment responses (23) .

5. Future Perspectives and Challenges

Precision medicine does hold great promise for advancing the diagnosis and management of CD, but several challenges are still remaining (29, 30):

- **Data Quality:** to ensure the accuracy, consistency, and completeness of datasets used for ML training and EMR analyses is mere critical.

- **Ethical Considerations:** Ethical considerations extend beyond basic privacy concerns into complex issues of informed consent, algorithmic transparency, and equitable access. Healthcare AI systems must comply with regulations such as the General Data Protection Regulation (GDPR) in Europe and the Health Insurance Portability and Accountability Act (HIPAA) in the United States. Furthermore, ensuring that patients fully understand how their data will be used in ML models presents unique challenges in the digital health era (39,40) .

The potential for algorithmic bias—where ML models may perform differently across demographic groups—requires rigorous validation across diverse populations to ensure that precision medicine approaches benefit all CD patients equitably (41).

· **Clinical Integration:** The successful integration of such advanced technologies into routine clinical practice requires a continuous training of healthcare practitioners and a robust validation of these tools in diverse ethnies and populations (42).

Addressing these challenges would pave the way for broader implementation of precision medicine for celiac patients.

CONCLUSION

Precision medicine is revolutionizing CD management through the integration of EMRs and ML technologies, enabling earlier detection, personalized therapies, and enhanced disease monitoring. These advanced computational approaches leverage diverse data sources to optimize diagnostic accuracy, treatment selection, and outcome prediction. Future research should focus on external validation of ML models across diverse populations, implementation strategies for clinical integration, and development of accessible tools that bridge the technological divide. By addressing these challenges, we can realize the full potential of precision medicine to transform CD care and ultimately improve patient outcomes worldwide.

6%

SIMILARITY INDEX

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet	24 words — 1%
2	casadesante.com Internet	23 words — 1%
3	es.readkong.com Internet	23 words — 1%
4	www.mdpi.com Internet	16 words — 1%
5	www.science.gov Internet	14 words — 1%
6	academic.oup.com Internet	13 words — 1%
7	0-www-mdpi-com.brum.beds.ac.uk Internet	12 words — 1%

EXCLUDE QUOTES ON

EXCLUDE SOURCES < 12 WORDS

EXCLUDE BIBLIOGRAPHY ON

EXCLUDE MATCHES < 12 WORDS