www.jcimcr.org

JCIMCR OPEN ACCESS Journal of Clinical Images and Medical Case Reports

ISSN 2766-7820

Research Article

Open Access, Volume 5

Early detection of celiac disease through its common symptoms using machine learning algorithms

Mahshid Mohsenzadeh Hedesh[1]; Zahra Maharat[2]; Amirreza Khalaji[3,4]; Leila Pazouki[5]; Shiva Khalili Dooraki[6]*

[1]Department of Natural Sciences, Bonn-Rhein-Sieg University of Applied Sciences, Rheinbach, Germany.

[2]Department of Cell and Molecular Biology and Microbiology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran.

[3]Immunology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran.

[4]Connective Tissue Diseases Research Center, Tabriz University of Medical Sciences, Tabriz, Iran.

[5]Department of Biology, University of Louisville, Louisville, KY, USA.

[6]Department of Biology, Tabriz Science and Research Branch, Islamic Azad University, Tabriz, Iran.

***Corresponding Author: Shiva Khalili Dooraki**
Department of Biology, Tabriz Science and Research Branch, Islamic Azad University, Tabriz, Iran.
Email: shivakhalilidooraki@gmail.com

Received: Feb 16, 2024
Accepted: Mar 05, 2024
Published: Mar 12, 2024
Archived: www.jcimcr.org

DOI: www.doi.org/10.52768/2766-7820/2915

Keywords: Celiac disease; Machine learning; Classification; K-nearest neighbors; Support vector machine; Naive bayes; Decision tree; Random forest.

Abstract

Celiac disease is a common systemic immune-mediated disease caused by an abnormal immune response to gluten proteins, a protein found in grains such as wheat, barley, and rye. The only effective treatment for celiac disease is a lifelong gluten-free diet. This disease has spread worldwide, and its prevalence in the general population is estimated at 1% worldwide. Celiac disease is highly heritable, and its pathogenesis involves gluten antigens presented on the surface of HLA complexes, mainly haplotypes DQ2 and DQ8. However, even if the genetic predisposition shown by these haplotypes is known to be obligatory for celiac disease, it is not sufficient to explain the overall predisposition to the disease. The first step to diagnosing the disease is usually based on serological tests and small bowel biopsy, but due to non-standard serological tests and inappropriate biopsies, the diagnosis of celiac disease is difficult. In addition, the onset of celiac disease includes a wide range of symptoms, which makes early diagnosis of celiac disease very important and vital to prevent long-term complications of these annoying symptoms. For this reason, considering the importance of early diagnosis of this disease, our goal in this study was to apply several machine learning algorithms to train several models and test their performance in predicting celiac disease based on common features and symptoms. This study was conducted on 50 suspected celiac disease samples with an average age of 32 years. 70% of the samples were positive for the disease, and the remaining 30% were negative. The 10-fold cross-validation method was used for training the model. Finally, by using a metaclassifier and the majority vote of all 5 models, including K-Nearest Neighbor, Support Vector Machine, Naive Bayes, Decision Tree, and Random Forest, we were able to achieve an accuracy of 0.8, recall of 0.88, precision of 1, and f-measure of 0.88. The most important features were identified to optimize the prediction performance. The 5 most important features were age, gluten sensitivity, chronic diarrhea, abdominal pain, and lactose intolerance.

Citation: Hedesh MM, Maharat Z, Khalaji A, Pazouki L, Dooraki SK. Early detection of celiac disease through its common symptoms using machine learning algorithms. J Clin Images Med Case Rep. 2024; 5(3): 2915.

Introduction

Celiac disease is a chronic autoimmune disorder affecting the digestive system. It is caused by a reaction to gluten, a protein found in wheat, barley, and rye [1]. When a patient with celiac disease consumes gluten, their immune system attacks the small intestine, damaging the lining (Figure 1) and causing a range of symptoms. Celiac disease was first identified by Samuel Gee in 1888, and the role of gluten in the root of its pathology became clearer in 1953 [2].

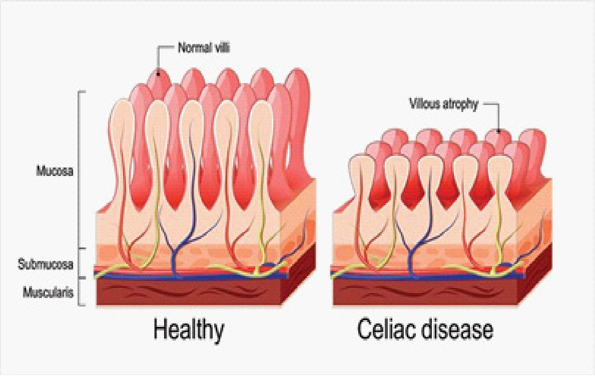

Figure 1: Comparison of small intestinal villi in healthy people and people with celiac disease.

The pathogenesis of CD includes gluten antigens presented on the surface of HLA complexes, mainly of haplotypes DQ2 or DQ8 [3]. In particular, it has been observed that 90-95% of CD patients express HLA-DQ2, and the remaining 5-10% express HLA-DQ8 [4]. DQ2 is present in the white population of Western Europe, Northern and Western Africa, the Middle East, and Central Asia, while DQ8 is widespread in people from Latin America and Northern Europe [5]. Studies carried out in Middle Eastern countries showed that compared to Western countries, the spread of celiac disease is higher even in individuals who are not at risk. Its prevalence varies depending on geographical and ethnic variations. The highest prevalence is in Europe 0.8% and Oceania 0.8%, while the lowest prevalence is in South America 0.4%. Celiac disease prevalence was 1.5 times higher in women than men, and approximately two times higher in children than in adults. Breast milk, mode of delivery, and the age of gluten intake in infants are risk factors for developing celiac disease and may affect its occurrence [6].

Patients with celiac disease have a higher risk of concomitant autoimmune disorders, while patients with autoimmune diseases, mainly those with diabetes or thyroid disease, sometimes develop celiac disease. It has been reported that there is a connection between CD and several rheumatic disorders. Juvenile Idiopathic Arthritis (JIA) is known as chronic arthritis with an autoimmune etiology, and CD is connected with susceptibility to JIA [7]. CD prevalence has also been observed to be higher in patients with autoimmune liver disorders [8]. The occurrence of Autoimmune Thyroiditis (AT) in Celiac Disease (CD) is well documented in adults but less so in children.

Celiac disease affects approximately 1% of the population worldwide. Contrary to increased knowledge about celiac disease, up to 95% of celiac patients still remain undiagnosed. Although a number of patients have significant clinical signs of celiac disease, there are many undiagnosed cases even in developed countries. It is estimated that 2.5 million Americans with CD are undiagnosed. The diagnosis of celiac disease is difficult because the clinical presentation of CD varies and can differ depending on age. Celiac disease can occur at any age from early childhood to old age. It can occur after gluten intake within the first 2 years of life or be seen in the second or third decade of life and can have a wide range of symptoms, including the main and common manifestations of the Gastrointestinal System (GIS) of celiac disease are nausea, abdominal distention, chronic diarrhea, vomiting, and frequent abdominal pain. In addition, common extra-intestinal manifestations include lack of growth, osteopenia, short stature, osteoporosis, chronic anemia, increased liver enzymes, delayed puberty, irritability, arthritis chronic fatigue, neuropathy, arthralgia, amenorrhea, and tooth enamel defects. GIS signs of celiac disease such as diarrhea are seen in approximately 50% of patients [9]. Children with celiac disease who are diagnosed at younger ages have had fewer symptoms over the past 20 years. Generally, symptoms present at around 6-18 months; however, in recent years, generally have been presenting symptoms at a later age different from classical symptoms [2].

Diagnosis of CD relies on clinical, serological, and histological evidence, and the increase in rates of diagnosis can be partly attributed to the use of sensitive serology testing [10]. Before performing a serological test for celiac disease, one should pay attention to whether they are gluten-free in their diet or not. In this case, the result of the serological tests may be negative and the diagnosis of celiac disease may be difficult.

One of the indicators of celiac disease diagnosis is intestinal biopsy, but intestinal biopsy is not compatible with celiac disease in many cases. The reason for a negative intestinal biopsy may be due to the involvement of the small intestine mucosa, low gluten consumption, and inappropriate biopsy. In the diagnosis of CD by histopathology, upper endoscopy is performed with a biopsy of the duodenum (beyond the duodenal bulb) or the jejunum to obtain multiple (four to eight) samples of the duodenum. It is known that not all areas may be affected equally; for example, if biopsies are taken from healthy intestinal tissue, the result will be a false negative. Even in the same bioptic fragment, the presence of different degrees of damage may appear. Most people with celiac disease emerge from a normal-looking small intestine on endoscopy before biopsies are examined in the lab. Lactose intolerance also occurs as a consequence of small bowel injury due to conditions such as viral gastroenteritis, giardiasis, celiac disease, or Crohn's disease. The typical symptoms of lactose intolerance include abdominal pain, bloating, flatus, diarrhea, borborygmi, and less frequently, nausea and vomiting.

The gold standards in diagnosing CD are bowel biopsy and positive serological markers such as TTG-IgA and EMA-IgA. European guidelines recommend that in adolescents and children with symptoms compatible with celiac disease, the diagnosis can be made without the need for an intestinal biopsy if TTG antibody titers are 10 times higher than the normal range [11].

A Gluten-Free Diet (GFD) is the recommended treatment for CD. This can be challenging, as gluten is found in many common foods. It is important to be aware of celiac disease and its symptoms, as undiagnosed and untreated celiac disease can lead to serious complications including osteoporosis, infertility, and malignancies such as T-cell lymphoma. This can explain the increased mortality rate among patients with CD [12]. However, people with celiac disease can live a long life with proper diagnosis, following a GFD, and having a healthy lifestyle. Life-long adherence to a GFD can be complex and costly and will require considerable changes in eating habits that can be challenging for an individual. There are significant challenges associated with adherence to a GFD, including cost, availability of GF products, and psychological barriers. Such limitations, especially restrictions in social situations, can lead to dietary nonadherence resulting in a poorer quality of life [13].

In recent years, medical data used for clinical practice has been readily available in electronic form. This data informs the decision-making of clinicians in patient care, with the ultimate goal of tailoring each patient's assessment and plan to the individual. Indeed, this is the goal of "precision medicine": utilizing different data modalities, including genomic data, Electronic Medical Records (EMR), textual data (e.g. unstructured patient notes), and image data (e.g. CT scans, MRI, endoscopy), to optimize and personalize the treatment for a patient [14].

Healthcare can be widely augmented using Machine Learning (ML) technology. One of the basic requirements of any ML-based model is to be able to incorporate a large amount of data. Electronic Medical Records (EMR) are huge and ever-growing databases that can be used to assist physicians with diagnosis, management, and tailored recommendations [15]. In the medical field, this obviously requires training datasets selected by specialized clinicians. Machine learning is an emerging technique in healthcare that helps predict and diagnose various diseases based on the symptoms that patients have. After data collection and preprocessing, the data is prepared for training machine learning algorithms. After training, we can predict the disease for the input symptoms by combining the predictions of all algorithms. Machine learning algorithms are useful in celiac disease and similar diseases that require extensive and continuous testing to diagnose and treat patients. After the machine learning model is learned, different samples are given to the model, which, depending on the type of model used, bring binary values between 0 and 1.

Early diagnosis and management of celiac disease are crucial for preventing long-term complications and improving the quality of life for those affected. However, celiac disease can be difficult to diagnose, as symptoms can vary widely and may overlap with other digestive disorders.

Bioinformatics is a recently developed science that uses information technology to understand biological phenomena. Bioinformatics is used for in silico analyses of genome sequence data [16,21], protein engineering [22], investigating cancer cell lines [2328], DNA computing [29], metagenomics [30,32]. It also applies Artificial Intelligence (AI) in healthcare, and recent studies have explored its potential in predicting and diagnosing different cancers and diseases. By analyzing large amounts of patient data and identifying patterns and risk factors, AI algorithms can help identify individuals who are at high risk of developing celiac disease [33], even before they show symptoms.

In this article, we will explore the current state of research on AI and celiac disease prediction, including the use of machine learning algorithms and other AI techniques.

Materials and methods

Statistical society: The research included 50 celiac-diagnosed patients or individuals with disease symptoms referred from Tehran (Iran) and other parts of the country to the Clinical and Specialty Laboratory, as a referral lab. A questionnaire was filled out by every individual regarding their personal and clinical characteristics.

Data collection: The studied individuals are people with symptoms and a high suspicion of celiac disease, as well as people at risk, including first-degree relatives of the patient. These individuals were referred to the gastroenterology department and had either not been recently diagnosed or treated by a gastroenterologist were in the recurrence phase of the disease or had not responded to treatment. After undergoing serological and endoscopic tests by a gastroenterologist, they were evaluated for the final investigation and confirmation of celiac disease.

First, personal characteristics such as age, gender, weight, reason for visiting the doctor, and bothersome symptoms (diarrhea, abdominal pain, bloating, constipation, weight loss, unexplained anemia, bone pains, muscle spasms, fatigue, growth retardation, joint pain, convulsions, tingling in the legs, painful mouth sores, painful skin lesions including Herpetiformis, delay in teeth development), family history, history of allergies to specific foods and drugs, history of certain diseases (diabetes, thyroid issues, rheumatism), and the results of serology tests were recorded in a special form.

Statistical analysis: The number of patients in this study was selected based on the following statistical method:

$$n=\frac{\frac{Z^2pq}{d^2}}{1+\frac{1}{N}\left(\frac{Z^2pq}{d^2}-1\right)}$$

In this formula, considering the alpha coefficient of 5%, the confidence interval is 95%, and the confidence level is 1.96, which was obtained from similar studies in geographical areas near Iran, the sample size was estimated to be 50 people.

The present study was performed on 50 individuals including 16 males and 34 females. The average age was 34.75 (min: 6 and max: 75) in males and 31.79 (min: 2 and max: 64) in females.

The importance of the studied features: Celiac Disease (CD) is an immune-mediated disease of the small bowel attributable to gluten sensitivity in susceptible patients [34]. The CD is diagnosed by the presence of clinical symptoms, serological markers, and histological examination of intestinal biopsies [35]. Histological evaluation typically shows a spectrum of disease, ranging from intraepithelial lymphocytosis to total mucosal damage characterized by atrophy and loss of villi, hyperplasia of the crypts, and increased apoptosis of the epithelium [3]. The pathogenesis of CD includes gluten antigens presented on the surface of HLA complexes, mainly of haplotypes DQ2 or DQ8 [36].

Celiac disease can manifest with a diversity of signs and symptoms, both specific (including gastrointestinal signs, abdominal pain, flatulence, weight loss, malnutrition, malabsorption, chronic diarrhea, and failure of children to grow normally), which begins regularly between six months and two years of age, and nonspecific are more common, especially in people older than 2 years (such as fatigue, iron deficiency anemia, dermatitis herpetiformis, low bone mineral density, and oral manifestation) [37]. It is also associated with autoimmune diseases, such as type 1 diabetes, Hashimoto's, and thyroiditis [38].

Celiac disease has been reported in about 1% of the population, but it is often underdiagnosed because numerous patients report either no symptoms or very few symptoms. Among these symptoms, the most common historically, are diarrhea and weight loss. Currently, Iron Deficiency Anemia (IDA) is often the presenting feature at diagnosis, being reported in over half of CD patients (including subclinical CD patients), with a higher prevalence in adults than in children [39]. Its prevalence in children is 1%-8.3% and the sex distribution is similar [40]. CD results from a reaction with gluten, which is a group of different proteins found in wheat and other grains such as barley and rye. Moderate amounts of oats are regularly tolerated, as long as they are free from contamination with other gluten-containing grains. The incidence of harm may depend on the type of oats. The CD appears in people with a genetic predisposition. When exposed to gluten, the abnormal immune response may result in the production of many different auto-antibodies that can affect a number of distinct organs. In the small intestine, this causes an inflammatory reaction and may lead to villous atrophy. This affects the absorption of nutrients, often leading to anemia [38,41]. The associated histological alterations (i.e., atrophy of the duodenal mucosa) are responsible for malabsorption and multiple micronutrient deficiencies (e.g., iron, vitamin B12, and folic acid), which might be involved in the pathogenesis and morphologic features of anemia. However, nutritional deficiencies alone cannot explain this phenomenon in all cases [42].

Lactose Malabsorption (LM) is caused by the incomplete hydrolysis of lactose due to lactase deficiency, resulting in reduced expression of the lactase enzyme in the small intestine. LM may occur as a primary or secondary disorder due to other intestinal diseases. LM leads to Lactose Intolerance (LI), which is the occurrence of gastrointestinal symptoms after ingesting lactose. A lactose-restricted diet is typically recommended for symptom relief, although it may lead to nutritional disadvantages with reduced calcium and vitamin intake. The frequency of LI varies according to ethnicity and has been reported as high as almost 100% in Southeast Asia, approximately 80% in Southern Europe, and less than 5% in Northern Europe [43].

The proliferation and overstimulation of autoreactive lymphocytes lead to diarrhea, abdominal pain, and decreased absorption of nutrients such as calcium and iron, due to the loss of intestinal microvilli. This mucosal damage is the main factor for impaired lactase production, resulting in shared symptoms with individuals who have Lactose Intolerance (LI). As a result, CD is often misdiagnosed and relapses until a correct diagnosis and treatment initiation are made. After correct treatment with a strict Gluten-Free Diet (GFD), LI- and CD-associated symptoms usually improve. However, if CD is not controlled, the function of other organs may be affected, resulting in symptoms characteristic of other diseases, including salivary gland issues, pancreas problems (insulin-dependent diabetes mellitus comorbidity), irritating dermatological blisters known as Dermatitis Herpetiformis (DH), severe anemia, migraines, Thyroid Impairment (TI), bone mass loss, and cancer (thyroid cancer, small intestine cancer, and lymphoma) [44].

Microscopic Colitis (MC) is an inflammatory condition in which patients suffer from chronic diarrhea with evidence of chronic inflammation under the microscope but show normal colonic morphology macroscopically [45].

Common oral and dental manifestations of CD include mouth ulcers, recurrent aphthous stomatitis (RAS), and ulcers. As first reported by Aine [46], dental enamel defects include delayed tooth eruption, angular cheilitis, atrophic glossitis, and burning tongue. Dental enamel hypoplasia has a reported prevalence ranging from 10% to 97% [47] and appears to be more prevalent in children, compared with adults with CD, and in patients with CD compared to the general population. Furthermore, it is thought to be secondary to nutritional deficiencies and immune disturbances during the period of enamel formation in the first seven years [12].

Internationally, celiac disease has an effect between 1 in 100 and 1 in 170 people [38,48]. However, rates vary between different regions of the world from 1 in 300 to 1 in 40 [48]. It was also found that 1 out of every 105 blood donors carries IgA TG in their blood. Because of the variable signs and symptoms, about 85% of sufferers are believed to go undiagnosed [41]. It was also found that the percentage of people with a clinically diagnosed disease (symptoms trigger a diagnostic test) is 0.05-0.27% in most studies [38,41].

Methodology

Overview of methodology: Identifying and diagnosing diseases, which are the most important factors for the treatment of any disease, are very difficult in themselves. On the other hand, many signs and symptoms are non-specific, and this also makes the diagnosis more difficult. The use of machine learning can predict disease diagnosis based on the creation of a model in which the symptoms of each patient can be entered and provide a model of a specific disease.

Machine learning is a branch of computer science that has been successful in early diagnosis in various fields of medicine using computational methods.

The best way to reduce the death rate from any disease is early diagnosis and treatment. Therefore, to predict diseases, medical science is turning to new prediction model technologies based on machine learning algorithms.

There are different types of machine learning techniques that this paper focuses on, such as Naive Bayes (NB), Decision Tree (DT), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Random Forest (RF) for celiac disease detection.

Machine learning algorithms:

K Nearest Neighbors algorithm (KNN): KNN is one of the simplest and most widely used machine learning algorithms used to solve classification and regression problems. This algorithm gets neighbors among data using Euclidean distance between points of data [49]. It is a data classification algorithm that detects a new item by calculating the nearest neighbor with the same characteristics as the item in a defined area. The value of K (which is fixed and defined by the user) identifies all items with similar existing features to the new item and surrounds all cases to find the new case for the same category.

Therefore, the value of k should be chosen carefully because if the sample size is small, it can greatly affect the selection of the optimal neighborhood size K, and the decrease in classification performance is easily caused by the sensitivity of K selection [50].

After obtaining the distances, we need to sort them and determine which one is closest to the new sample. Then, we find the optimal k value and make a prediction.

Support Vector Machine algorithm (SVM): The SVM is a popular machine-learning tool that provides solutions for problems with classification and regression. This algorithm performs classification based on labeled data in the best way. SVM finds the best hyperplane by finding the points on the edge of class descriptors and divides the dataset into two distinct classes. The distance between the classes is known as the margin. The better accuracy is achieved when there is a higher margin. The far margin is measured between the extreme surface and the nearest data point from each set of the classified dataset. The data points that lie on the boundary are called support vectors. SVM is not only able to deal with two-class or binary classification problems but it is also developed to solve multilayer problems using a group of hyperplanes [51]. In other words, in this algorithm, we will have dimensions for as many features under consideration.

Naive Bayes algorithm (NB): NB algorithms are considered one of the most popular machine learning algorithms because they allow each feature to participate in the final prediction independently of other features. This method has many advantages, such as small training data, simple computing, and ease of implementation. In addition, it can handle big data and incomplete data (missing values), and it is not sensitive to irrelevant features and data noise [52]. Another use of this algorithm is to classify documents and filter spam emails. The important and key components of Naive Bayes include prior, posterior, and class conditional probability.

If the probability value is discontinuous, it is called likelihood, but if its value is continuous (decimal numbers) it is called probability. Whichever group in the study has a higher probability number, the probability of its occurrence is higher, and in other words, it is the maximum likelihood and is predicted as the answer.

Decision Tree algorithm (DT): A decision tree is one of the earliest and most outstanding machine learning algorithms used to solve regression and classification issues by repeatedly dividing data depending on a particular variable. The data is divided into nodes, and the tree's leaf represents the final decision.

The nodes of a decision tree have multiple levels, where the first and top-most node is called the root node. All internal nodes show tests on input variables. Depending on the test result, the classification algorithm branches toward the appropriate node where the process of test and branching repeats until it reaches the leaf node. Finally, the leaf or terminal nodes correspond to the prediction outcomes. Decision trees are very easy to interpret and learn, and they are a common part of many medical diagnostic protocols [53].

Random Forest classification (RF): The random forest is a collection of decision trees. To enable this model, ensemble classification is used, which starts by identifying a set of key features to grow each decision tree. Based on the feature chosen as an internal node, the shape of the DT will be different, and their size can be small or large, and the number of their branches is few or many. All of these are based on the features that have been selected. As its name suggests, RF selects variables randomly, so different trees are created. Then each DT is checked, and prediction is done for each DT. Then a majority voting is taken from all the predictions obtained, and the final prediction is made [54].

Importance features selection: High-dimensional data analysis is a challenge for researchers in the fields of machine learning and data mining. Importance feature selection provides an effective way to solve this problem by removing irrelevant and redundant data, which can reduce computation time, improve learning accuracy, and facilitate a better understanding of the learning model or data. Feature selection refers to the process of obtaining a subset from an original feature set according to a certain feature selection criterion, which selects the relevant features of the dataset.

In this study, after deriving a set of features, we next used ML models to evaluate whether these features hold relevant information to classify individuals as celiac disease patients or healthy. It is likely that most of the repertoire features are not associated with autoimmune diseases, so a process of importance feature selection is critical for the classification task. To identify the features that optimize the prediction performance, we used a feature selection step. To select the strongest predictor features, the number of features was determined using the scikit-learn Python library. This function performs cross-validated selection of the optimal number of features by removing 0 to N features using recursive feature elimination and then selecting the best subset based on the cross-validation of the model. To assess model performance, we performed 20% sample holdout cross-validation, where the model was trained on the remaining 80% of the data and then scored based on the holdout samples after selecting the best stratification features.

Results

The features described earlier, including Skin Manifestations, Dental Problems, Oral Ulcers, Vomit, Weight Loss, Bloat, Abdominal Pain, Lactose Intolerance, Anemia, Muscle Weakness, Continuous Constipation, Chronic Diarrhea, Boredom, Gluten Sensitivity, TTG IgG, TTG IgA, Sex, and Age, were measured in 50 patients, and the values of these features were shown in Figure 2. The values of the features were normalized between 0 and 1.

The 10-fold cross-validation method was used for modeling, where in each step, 10% of the data were selected for testing, and modeling was done with the remaining 90% of the data.

In order to measure the ability of each of the K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Naive Bayes

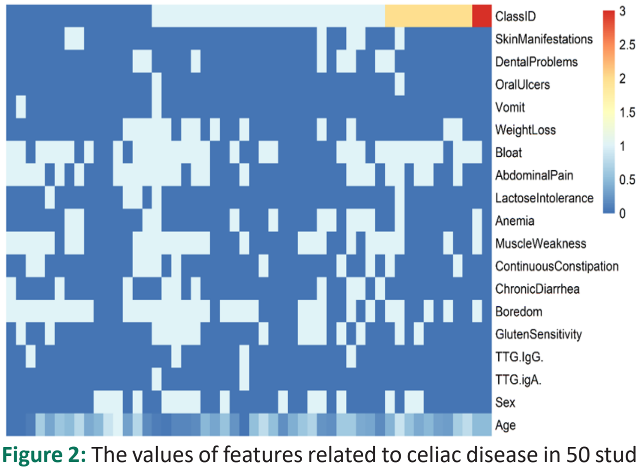

Figure 2: The values of features related to celiac disease in 50 studied patients.

(NB), Decision Tree (DT), and Random Forest (RF) classifiers, which were fully explained in the method section, the following four evaluation measures have been used.

Accuracy: The ability of an instrument to measure the accurate value is known as accuracy. In other words, it is the closeness of the measured value to a standard or true value.

Accuracy is also used as a statistical measure of how well a binary classification test correctly identifies or excludes a condition. That is, the accuracy is the proportion of correct predictions (both true positives and true negatives) among the total number of cases examined.

$$\text{Accuracy}=\frac{\text{Number of correct predictions}}{\text{Total number of predictions}}=\frac{TP+TN}{TP+TN+FP+FN}$$

TP, FP, TN, and FN stand for True Positive, False Positive, True Negative, and False Negative, respectively.

Accuracy was obtained by the methods of KNN: 0.6, SVM: 0.7, NB: 0.8, DT: 0.6, and RF: 0.6.

Precision: Precision is how close measurement values are to each other and, basically, how many decimal places are at the end of a given measurement. High precision coincides with a low sample standard deviation. Standard deviation is a measurement of how widely spread a data set is. A sample standard deviation specifically represents the spread of data from a particular sample and may not accurately represent the true spread of data from the population. Precision is defined as follows:

$$\text{Precision}=\frac{TP}{TP+FP}$$

Precision was obtained by the method of KNN: 0.72, SVM: 1, NB: 0.86, DT: 0.57, and RF: 0.86.

Recall: Recall is a metric that quantifies the number of correct positive predictions made out of all positive predictions that could have been made.

Unlike precision, which only comments on the correct positive predictions out of all positive predictions, recall provides an indication of missed positive predictions. In this way, recall provides some notion of the coverage of the positive class.

$$\text{Recall}=\frac{TP}{TP+FN}$$

Recall was obtained by the method of KNN: 0.72, SVM: 0.7, NB: 0.86, DT: 0.67, and RF: 0.67.

F-Measure: In statistical analysis of binary classification, the F-measure or F-score is a measure of a test's accuracy. It is calculated from the precision and recall of the test. The F-measure is a way of combining the precision and recall of the model, and it is defined as the harmonic mean of the model's precision and recall.

$$\text{F-Measure}= 2\ \frac{\text{Precision x Recall}}{\text{Precision + Recall}}$$

F-measure was obtained by the method of KNN: 0.72, SVM: 0.82, NB: 0.86, DT: 0.62, and RF: 0.75.

Finally, by using a Meta-Classifier and the majority voting of all 5 models (KNN, SVM, NB, DT, and RF), we were able to achieve higher accuracies and make final predictions. With the Meta-Classifier, the accuracy is 0.8, recall is 0.88, precision is 1, and the f-measure is 0.88. As shown in Figure 3, the Meta-Classifier has the highest accuracy among all classifiers.

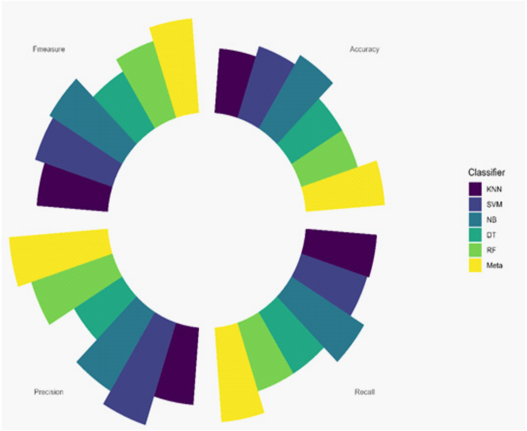

Figure 3: Prediction performance of the best classifiers. The length and color of the bars indicate the percentage of evaluation measures and types of classifiers respectively

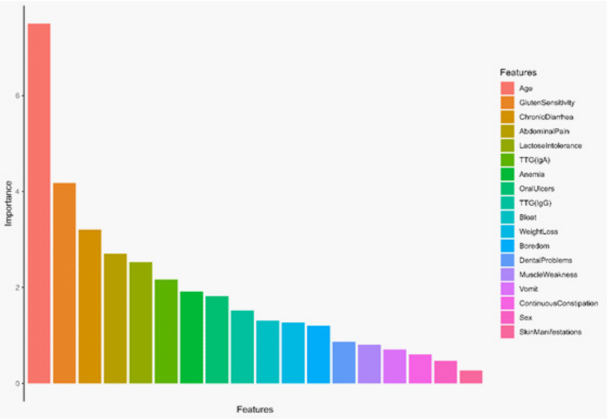

Figure 4: Predicting the importance of features. The length and color of the bars indicate the importance and features respectively.

By collecting a questionnaire from the study subjects, which included common characteristics abundantly seen in celiac patients based on past studies, we experimented to create an ML-based model to distinguish celiac disease patients from healthy individuals. We found that age, gluten sensitivity, chronic diarrhea, abdominal pain, and lactose intolerance play a more important role in the diagnosis and classification of celiac patients than healthy individuals (Figure 4).

Discussion

Our study presents initial findings that suggest machine learning can be used as an alternative, non-invasive method for screening patients with suspected celiac disease. This approach focuses on diagnosing general and common symptoms associated with the disease.

The aim of this research area is to assist in clinical decision-making for the treatment of individuals with potential celiac disease during their initial clinical and laboratory evaluation. It is worth noting that most individuals with potential celiac disease do not progress to a fully developed disease within an 8-year follow-up period.

Meijer et al. discovered that the risk of developing CD in children with affected First-Degree Relatives (FDR) is much higher during the first 10 years of life than previously believed. While it was previously assumed to be around 5-10%, their data revealed that by the age of 8, this rate actually reaches 17%. This highlights the need for effective recommendations for early screening. Additionally, their research confirmed that CD tends to develop in children with FDR at a very young age, with the average age of diagnosis in their study being 4 years. As a result, it is important to advise high-risk children to begin CD screening earlier and more frequently than children in other high-risk groups. It is also worth noting that the risk of developing CD varies depending on the current age of the child [55].

In a study conducted by Majsiak et al. 796 patients with confirmed CD diagnoses in Poland were analyzed. Of these patients, 224(28.1%) were children and 572(71.9%) were adults. The study revealed that the average duration of symptoms before the CD diagnosis was significantly shorter in children (3.1 years) compared to adults (9 years). The most common symptoms before the CD diagnosis were abdominal pain and bloating in children (70.4%) and chronic fatigue in adults (74.5%). These findings highlight the delayed diagnosis of CD, particularly in adults, emphasizing the need for doctors to be aware of the various ways CD can present itself [56].

Lemos et al. also discovered a similar profile between CD and LI, possibly due to changes in gut microbiota. They found that one-third of CD participants experienced symptoms induced by lactose. These symptoms may be linked to intestinal damage and extraintestinal symptoms, warranting further investigation. Out of the CD patients, 20 (58.8%) were diagnosed with LI. Considering the majority of CD patients also had LI, the researchers analyzed the CD and LI groups together [57].

Nimri et al. conducted a study that revealed how Microscopic Colitis (MC) can result in chronic diarrhea. The diagnosis of MC is made through histopathology, which shows a high number of intraepithelial lymphocytes, specifically more than 20 lymphocytes per high-power field. As a result, chronic diarrhea is strongly linked to CD [58]. In another study by Wang et al. the main clinical symptoms observed in CD patients in Xinjiang were chronic diarrhea, severe malnutrition, osteoporosis, anemia, fatigue, and decreased BMI. BMI serves as a crucial indicator for assessing and predicting CD, and diarrhea is a common symptom associated with the disease. The immune response triggered by gluten consumption in susceptible individuals leads to impaired intestinal absorption and osmotic diarrhea. Among the 21 CD patients examined in their study, the predominant presentation was abundant watery and fatty diarrhea. Due to a lack of understanding and limited diagnostic criteria for CD, diarrhea often becomes chronic, making the disease more challenging to manage. Consequently, most patients experience significant weight loss, along with anemia, iron and vitamin D deficiencies, and other forms of malnutrition [59].

The Ritter et al. study involved 101 participants, with an average age of 6.5 years (2.8). Of these participants, 51% were women. The study included 38 patients with CD, 18 patients with abdominal pain, and 45 healthy individuals. The sleep disturbance scale scores for children were 37.4(8.7), 41.3(11.3), and 45.4(13.7) in the healthy control, CD, and abdominal pain groups, respectively (P=0.024). The study found a significant difference in arousal domain disorders (P=0.044). Furthermore, a trend towards improvement in the sleep disturbance scale for children was observed in children with CD who experienced abdominal pain after following a gluten-free diet for 6 months (P=0.07) [60]. In the study conducted by Jabeen et al. it was reported that 22% of cases experienced severe abdominal pain, while 9.2% experienced nausea. Among patients with CD, 70% suffered from varying degrees of diarrhea (mild, moderate, and severe) and reported severe abdominal pain in 22% of cases, while nausea in 9.2%. 70% of patients with CD suffered from various degrees of diarrhea (mild, moderate, and severe) [61]. When examining the literature, it was observed that patients with CD frequently experienced abdominal distension. Additionally, Thapa et al. found that over 70% of CD patients experienced diarrhea [62].

Tavakoli et al. conducted a study in South Khorasan (Iran) with 110 individuals diagnosed with CD, who had an average age of 28.38±15.25 years. Among the participants, 78(70.9%) were men and 32(29.1%) were women. The most prevalent gastrointestinal symptoms reported were abdominal pain in 70(63.6%) individuals, diarrhea in 44(40%), constipation in 43(39.1%), and nausea in 35(31.8%) [63].

Comparing these findings to our own study, we also identified a significant percentage of our target population who experienced these common symptoms. According to Figure 4, some of these symptoms, in order of importance and the highest frequency among the studied samples, included gluten sensitivity, chronic diarrhea, abdominal pain, lactose intolerance, TTG (IgA), and anemia. This study was conducted on 50 samples suspected of celiac disease. The average age of the subjects studied was 32 years. The minimum age was 2 years, and the maximum age was 75 years. 68% of the studied subjects were women, and 32% were men. Finally, 70% of the studied samples were positive for celiac disease, and 30% were negative.

Also, based on previous studies, age is considered a very important factor because early diagnosis of celiac disease in asymptomatic patients leads to a better quality of life and better adherence. In our study, as shown in Figure 4, the age indicator had the highest importance. If the disease is diagnosed in time, and the diet is followed, less damage is done to the intestinal mucosa, which ultimately prevents the occurrence of other common and annoying symptoms. As reported, the only effective treatment is a lifelong gluten-free diet.

Therefore, due to the very high importance of early diagnosis of celiac disease, the aims of this study were to develop and verify a model based on machine learning for the early diagnosis and identification of suspected celiac disease samples, with the common features and symptoms and finally its timely prevention. Machine learning can classify or predict the input data through statistical methods and algorithms, such as recognizing and specifically classifying the patient as celiac or non-celiac.

Piccialli et al. developed a traditional multivariate approach to predict the natural history of potential celiac disease in a sample of 340 children. They used a discriminant analysis model to analyze clinical data collected at the time of diagnosis (time 0). In addition to an existing follow-up dataset for Potential Celiac patients (PCD), the researchers proposed Machine Learning (ML) methods to expand the analysis and identify influential features that can predict outcomes. The ML methods, including Random Forests, Extremely Randomized Trees, Boosted Trees, and Logistic Regression, were used to select the most important features for predicting the outcome. This feature selection process effectively reduced the total number of features from 85 to 19. The ML methods produced results with high accuracy, with specificity scores consistently above 75% and two methods achieving over 98%. The best-performing sensitivity was 60%.

The optimized spanning trees model was able to accurately classify PCD patients using the 19 selected features, with an accuracy of 0.80, sensitivity of 0.58, and specificity of 0.84. This study successfully categorized PCD patients who are most likely to develop overt CD using machine learning techniques [64].

A review study conducted by Grossi et al. revealed that the use of Artificial Intelligence (AI) techniques for diagnosing and predicting gastrointestinal conditions is potentially more effective than traditional statistical methods. This finding is particularly significant considering the complexity of gastrointestinal diagnoses and the need for invasive tests. The CD group exhibited several common symptoms including irritability, edema, diarrhea, thin subcutaneous tissue, and weight loss. The initial phase of the study, which tested AI techniques, demonstrated that Bayesian classifiers and k-nearest neighbors could accurately identify potential CD diagnoses with good sensitivity. Moreover, these techniques exhibited reliable specificity in indicating negative CD diagnoses. Therefore, these AI techniques hold promise as clinical decision-support tools and warrant further investigation in future studies [65].

The measurements indicate that the selected algorithm is reliable for identifying patients with CD. In our study, like the previous studies that were mentioned earlier, the models learned by machine learning algorithms were able to distinguish with an accuracy of over 80%, which is comparable to the gold standard and the physician's impression. However, additional research is needed to minimize false positive and false negative outcomes. It is important to note that a potential diagnosis of CD that is not confirmed should be seen as an opportunity for further investigation rather than an error. Overall, this study offers a dependable tool for analyzing the clinical manifestations of CD.

The objective of this study like previous studies is to showcase the potential of machine learning technology in distinguishing between individuals suspected of having celiac disease and those who are healthy. The benefits of this approach include its portability, ability to provide fast real-time results, and cost-effectiveness.

Conclusion

Since 2008, when a computer-aided diagnosis of celiac disease was first published, many advances have been made toward a fully automated system. Despite the important step that has been taken in this direction, there are still many challenges that need to be addressed before a computer-aided diagnosis system can be used in clinical practice. If public datasets are readily available to researchers, research progress and interest in computer-aided celiac disease diagnosis will grow. Artificial intelligence is widely used in the prediction of various diseases and provides various models for the prevention of various types of diseases. It is very helpful for researchers in developing effective health care policies and early detection of diseases and can reduce the risk factors.

In this study, we used 5 classifiers: K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Naive Bayes (NB), Decision Tree (DT), and Random Forest (RF). The ability of each classifier was measured using 4 evaluation measures: Accuracy, Precision, Recall, and F-Measure. Using a Meta-Classifier and the majority voting of all 5 classifiers, we were able to achieve higher accuracy and final prediction. The results of the meta-classifier included an accuracy of 0.8, recall of 0.88, precision of 1, and f-measure of 0.88.

In addition, using machine learning, we proposed a model that can predict the new sample by using the common and predisposing features of celiac disease. To select the strongest predictor features, the number of features was determined using the scikit-learn Python library, and the 5 most important features were age, gluten sensitivity, chronic diarrhea, abdominal pain, and lactose intolerance.

References

1. Gujral N, Freeman HJ, Thomson AB. Celiac disease: prevalence, diagnosis, pathogenesis and treatment. World journal of gastroenterology: WJG. 2012; 18(42): 6036.

2. Parzanese I, Qehajaj D, Patrinicola F, Aralica M, Chiriva-Internati M, Stifter S, et al. Celiac disease: From pathophysiology to treatment. World journal of gastrointestinal pathophysiology. 2017; 8(2): 27.

3. Fernández-Bañares F, Esteve M, Farré C, Salas A, Alsina M, Casalots J, et al. Predisposing HLA-DQ2 and HLA-DQ8 haplotypes of coeliac disease and associated enteropathy in microscopic colitis. European journal of gastroenterology & hepatology. 2005; 17(12): 1333-8.

4. Schuppan D, Junker Y, Barisani D. Celiac disease: from pathogenesis to novel therapies. Gastroenterology. 2009; 137(6): 1912-33.

5. Durham J, Temples HS. Celiac Disease in the pediatric population. Journal of Pediatric Health Care. 2018; 32(6): 627-31.

6. Singh P, Arora A, Strand TA, Leffler DA, Catassi C, Green PH, et al. Global prevalence of celiac disease: systematic review and meta-analysis. Clinical gastroenterology and hepatology. 2018; 16(6): 823-36. e2.

7. Sahin Y, Sahin S, Barut K, Cokugras FC, Erkan T, Adrovic A, et al. Serological screening for coeliac disease in patients with juvenile idiopathic arthritis. Arab Journal of Gastroenterology. 2019; 20(2): 95-8.

8. Roumeliotis N, Hosking M, Guttman O. Celiac disease and cardiomyopathy in an adolescent with occult cirrhosis. Paediatrics & child health. 2012; 17(8): 437-9.

9. Sahin Y. Clinical evaluation of children with celiac disease: a single-center experience. Archives of Clinical Gastroenterology. 2020; 6(1): 026-30.

10. Al-Toma A, Volta U, Auricchio R, Castillejo G, Sanders DS, Cellier C, et al. European Society for the Study of Coeliac Disease (ESsCD) guideline for coeliac disease and other gluten-related disorders. United European gastroenterology journal. 2019; 7(5): 583-613.

11. Al-dossary OAI, Ahmed RA, Al-Moyed KAA, Al-Ankoshy AAM, Al-Najhi MMA, Al-Shamahy HA. Celiac disease among gastrointestinal patients in Yemen: its prevalence, symptoms and accompanying signs, and its association with age and gender. Universal Journal of Pharmaceutical Research. 2021; 6(5): 1-6.

12. Lebwohl B, Ludvigsson JF, Green PH. Celiac disease and non-celiac gluten sensitivity. Bmj. 2015;351.

13. Zarkadas M, Dubois S, MacIsaac K, Cantin I, Rashid M, Roberts K, et al. Living with coeliac disease and a gluten-free diet: a C anadian perspective. Journal of Human Nutrition and Dietetics. 2013; 26(1): 10-23.

14. Bray MS, Loos RJ, McCaffery JM, Ling C, Franks PW, Weinstock GM, et al. NIH working group report—using genomic information to guide weight management: From universal to precision treatment. Obesity. 2016; 24(1): 14-22.

15. Golden JA. Deep learning algorithms for detection of lymph node metastases from breast cancer: helping artificial intelligence be seen. Jama. 2017; 318(22): 2184-6.

16. Khorsand B, Khammari A, Shirvanizadeh N, Zahiri J, Arab SS. OligoCOOL: A mobile application for nucleotide sequence analysis. Biochemistry and Molecular Biology Education. 2019; 47(2): 201-6.

17. Soltanyzadeh M, Khorsand B, Baneh AA, Houri H. Clarifying differences in gene expression profile of umbilical cord vein and bone marrow-derived mesenchymal stem cells; a comparative in silico study. Informatics in Medicine Unlocked. 2022; 33: 101072.

18. Zahiri J, Khorsand B, Yousefi AA, Kargar M, Shirali Hossein Zade R, Mahdevar G. AntAngioCOOL: computational detection of anti-angiogenic peptides. J Transl Med. 2019; 17(1): 71.

19. Khorsand B, Savadi A, Naghibzadeh M. SARS-CoV-2-human protein-protein interaction network. Inform Med Unlocked. 2020; 20: 100413.

20. Khorsand B, Savadi A, Naghibzadeh M. Comprehensive host-pathogen protein-protein interaction network analysis. BMC bioinformatics. 2020; 21: 1-22.

21. Khorsand B, Savadi A, Zahiri J, Naghibzadeh M. Alpha influenza virus infiltration prediction using virus-human protein-protein interaction network. Math Biosci Eng. 2020; 17(4): 3109-29.

22. Sahlolbei M, Azangou-Khyavy M, Khanali J, Khorsand B, Shiralipour A, Ahmadbeigi N, et al. Engineering chimeric autoantibody receptor T cells for targeted B cell depletion in multiple sclerosis model: An in-vitro study. Heliyon. 2023; 9(9).

23. Samandari-Bahraseman MR, Khorsand B, Zareei S, Amanlou M, Rostamabadi H. Various concentrations of hesperetin induce different types of programmed cell death in human breast cancerous and normal cell lines in a ROS-dependent manner. Chemico-Biological Interactions. 2023; 382: 110642.

24. Samandari Bahraseman MR, Khorsand B, Esmaeilzadeh-Salestani K, Sarhadi S, Hatami N, Khaleghdoust B, et al. The use of integrated text mining and protein-protein interaction approach to evaluate the effects of combined chemotherapeutic and chemopreventive agents in cancer therapy. Plos one. 2022; 17(11): 0276458.

25. Shiralipour A, Khorsand B, Jafari L, Salehi M, Kazemi M, Zahiri J, et al. Identifying Key Lysosome-Related Genes Associated with Drug-Resistant Breast Cancer Using Computational and Systems Biology Approach. Iranian Journal of Pharmaceutical Research. 2022; 21(1).

26. Janfaza S, Banan Nojavani M, Khorsand B, Nikkhah M, Zahiri J. Cancer Odor Database (COD): a critical databank for cancer diagnosis research. Database. 2017; 2017: 055.

27. Janfaza S, Khorsand B, Nikkhah M, Zahiri J. Digging deeper into volatile organic compounds associated with cancer. Biology Methods and Protocols. 2019; 4(1): 014.

28. Sadeghnezhad E, Sharifi M, Zare-maivan H, Khorsand B, Zahiri J. Cross talk between energy cost and expression of Methyl Jasmonate-regulated genes: from DNA to protein. Journal of Plant Biochemistry and Biotechnology. 2019; 28: 230-43.

29. Khorsand B, Savadi A, Naghibzadeh M. Parallelizing Assignment Problem with DNA Strands. Iranian Journal of Biotechnology. 2020; 18(1): 2547.

30. Kharaghani AA, Harzandi N, Khorsand B, Rajabnia M, Kharaghani AA, Houri H. High prevalence of Mucosa-Associated extended-spectrum β-Lactamase-producing Escherichia coli and Klebsiella pneumoniae among Iranain patients with inflammatory bowel disease (IBD). Annals of Clinical Microbiology and Antimicrobials. 2023; 22(1): 86.

31. Houri H, Aghdaei HA, Firuzabadi S, Khorsand B, Soltanpoor F, Rafieepoor M, et al. High Prevalence Rate of Microbial Contamination in Patient-Ready Gastrointestinal Endoscopes in Tehran, Iran: an Alarming Sign for the Occurrence of Severe Outbreaks. Microbiology Spectrum. 2022; 10(5): 01897-22.

32. Khorsand B, Asadzadeh Aghdaei H, Nazemalhosseini-Mojarad E, Nadalian B, Nadalian B, Houri H. Overrepresentation of Enterobacteriaceae and Escherichia coli is the major gut microbiome signature in Crohn's disease and ulcerative colitis; a comprehensive metagenomic analysis of IBDMDB datasets. Frontiers in Cellular and Infection Microbiology. 2022: 1498.

33. Rostami-Nejad M, Asri N, Olfatifar M, Khorsand B, Houri H, Rostami K. Systematic review and dose-response meta-analysis on the Relationship between different gluten doses and risk of coeliac disease relapse. Nutrients. 2023; 15(6): 1390.

34. Shannahan S, Leffler DA. Diagnosis and updates in celiac disease. Gastrointestinal Endoscopy Clinics. 2017; 27(1): 79-92.

35. Grodzinsky E, Hed J, Skogh T. IgA antiendomysium antibodies have a high positive predictive value for celiac disease in asymptomatic patients. Allergy. 1994; 49(8): 593-7.

36. Hill ID, Dirks MH, Liptak GS, Colletti RB, Fasano A, Guandalini S, et al. Guideline for the diagnosis and treatment of celiac disease in children: recommendations of the North American Society for Pediatric Gastroenterology, Hepatology and Nutrition. Journal of pediatric gastroenterology and nutrition. 2005; 40(1): 1-19.

37. Rubin JE, Crowe SE. Celiac disease. Annals of internal medicine. 2020; 172(1): ITC1-ITC16.

38. Fasano A, Catassi C. Celiac disease. New England Journal of Medicine. 2012; 367(25): 2419-26.

39. Stefanelli G, Viscido A, Longo S, Magistroni M, Latella G. Persistent iron deficiency anemia in patients with celiac disease despite a gluten-free diet. Nutrients. 2020; 12(8): 2176.

40. Garnier-Lenglné H, Cerf-Bensussan N, Ruemmele FM. Celiac disease in children. Clinics and research in hepatology and gastroenterology. 2015; 39(5): 544-51.

41. Tovoli F, Masi C, Guidetti E, Negrini G, Paterini P, Bolondi L. Clinical and diagnostic aspects of gluten related disorders. World Journal of Clinical Cases: WJCC. 2015; 3(3): 275.

42. Martín-Masot R, Nestares MT, Diaz-Castro J, López-Aliaga I, Alférez MJM, Moreno-Fernandez J, et al. Multifactorial etiology of anemia in celiac disease and effect of gluten-free diet: A comprehensive review. Nutrients. 2019; 11(11): 2557.

43. Usai-Satta P, Scarpa M, Oppia F, Cabras F. Lactose malabsorption and intolerance: What should be the best clinical management? World journal of gastrointestinal pharmacology and therapeutics. 2012; 3(3): 29.

44. Theethira TG, Dennis M, Leffler DA. Nutritional consequences of celiac disease and the gluten-free diet. Expert review of gastroenterology & hepatology. 2014; 8(2): 123-9.

45. Lebwohl B, Sanders DS, Green PH. Coeliac disease. The Lancet. 2018; 391(10115): 70-81.

46. Ballinger A, Hughes C, Kumar P, Hutchinson I, Clark M. Dental enamel defects in coeliac disease. The Lancet. 1994; 343(8891): 230-1.

47. Rivera E, Assiri A, Guandalini S. Celiac disease. Oral Diseases. 2013; 19(7): 635-41.

48. Guandalini S, Assiri A. Celiac disease: a review. JAMA pediatrics. 2014; 168(3): 272-8.

49. Abdulqader DM, Abdulazeez AM, Zeebaree DQ. Machine learning supervised algorithms of gene selection: A review. Machine Learning. 2020; 62(03): 233-44.

50. Geler Z, Kurbalija V, Ivanović M, Radovanović M. Weighted kNN and constrained elastic distances for time-series classification. Expert Systems with Applications. 2020; 162: 113829.

51. Ibrahim I, Abdulazeez A. The role of machine learning algorithms for diagnosing diseases. Journal of Applied Science and Technology Trends. 2021; 2(01): 10-9.

52. Muhamad H, Prasojo CA, Sugianto NA, Surtiningsih L, Cholissodin I. Optimasi naïve bayes classifier dengan menggunakan particle swarm optimization pada data iris. J Teknol Inf dan Ilmu Komput. 2017; 4(3): 180.

53. Chauhan YJ. Cardiovascular disease prediction using classification algorithms of machine learning. International Journal of Science and Research. 2020; 9(5): 194-200.

54. Lakshmi SV, Meena M, Kiruthika N. Diagnosis of chronic kidney disease using random forest algorithms. International Journal of Research in Engineering, Science and Management. 2019; 2(3): 559-62.

55. Meijer CR, Auricchio R, Putter H, Castillejo G, Crespo P, Gyimesi J, et al. Prediction models for celiac disease development in children from high-risk families: Data from the PreventCD cohort. Gastroenterology. 2022; 163(2): 426-36.

56. Majsiak E, Choina M, Gray AM, Wysokiński M, Cukrowska B. Clinical Manifestation and Diagnostic Process of Celiac Disease in Poland—Comparison of Pediatric and Adult Patients in Retrospective Study. Nutrients. 2022; 14(3): 491.

57. Lemos DS, Beckert HC, Oliveira LC, Berti FC, Ozawa PM, Souza IL, et al. Extracellular vesicle microRNAs in celiac disease patients under a gluten-free diet, and in lactose intolerant individuals. BBA advances. 2022;2:100053.

58. Nimri FM, Muhanna A, Almomani Z, Khazaaleh S, Alomari M, Almomani L, et al. The association between microscopic colitis and celiac disease: a systematic review and meta-analysis. Annals of Gastroenterology. 2022; 35(3): 281.

59. Wang M, Kong W-J, Feng Y, Lu J-J, Hui W-J, Liu W-D, et al. Epidemiological, clinical, and histological presentation of celiac disease in Northwest China. World Journal of Gastroenterology. 2022; 28(12): 1272.

60. Reiter J, Abuelhija H, Slae M, Millman P, Davidovics Z, Chaimov E, et al. Sleep disorders in children with celiac disease: a prospective study. Journal of Clinical Sleep Medicine. 2023; 19(3): 591-4.

61. Jabeen S, Khan AU, Ahmed W, Ahmad M-u-D, Jafri SA, Bacha U, et al. Disease specific symptoms indices in patients with celiac disease—a hardly recognised entity. Frontiers in Nutrition. 2022; 9: 944449.

62. Thapa B. Celiac disease: Indian experience. Nutrition in Children in Developing Countries. 2004: 355.

63. Tavakoli T, Salmani F, Fard MS. Evaluation of the Epidemiological, Clinical, and Laboratory Characteristics of Patients with Celiac Disease in South Khorasan (Iran). Modern Care Journal. 2023.

64. Piccialli F, Calabrò F, Crisci D, Cuomo S, Prezioso E, Mandile R, et al. Precision medicine and machine learning towards the prediction of the outcome of potential celiac disease. Scientific Reports. 2021; 11(1): 5683.

65. Grossi E, Mancini A, Buscema M. International experience on the use of artificial neural networks in gastroenterology. Digestive and liver disease. 2007; 39(3): 278-85.