CeliPert—A Medical Expert System for the Knowledge-Based Interpretation of Test Results for Celiac Disease

Gerhard Granditsch[1], Andrea Rappelsberger[2], Wolf-Dietrich Huber[1], Swetlana Cirillo[2], Dieter Kopecky[2], and Klaus-Peter Adlassnig[2]

[1]Department of General Pediatrics, University Children's Hospital
Währinger Gürtel 18–20, A-1090 Vienna, Austria
e-mail: {gerhard.granditsch, wolf-dietrich.huber}@univie.ac.at
[2]Department of Medical Computer Sciences, Section on Medical Expert and Knowledge-Based Systems
University of Vienna Medical School, Spitalgasse 23, A-1090 Vienna, Austria
e-mail: {andrea.rappelsberger, dieter.kopecky, kpa}@akh-wien.ac.at

Abstract. *Celiac disease is a small intestinal enteropathy induced by the oral intake of gluten in genetically predisposed individuals. On a gluten-free diet, the structure and function of the small bowel mucosa are normalized, and signs of malabsorption, clinical signs, and immunological markers disappear. In many laboratories, the sera of patients are investigated for antigliadin antibodies and endomysial antibodies. This is considered an important step in the work-up of patients suspected to have celiac disease. At the gastroenterology laboratory of the University Children's Hospital of Vienna, the test results of every serum sample are supplemented by an expert's comments, which assesses the diagnostic value of the findings and issues recommendations to the clinician. CeliPert, a medical expert system based on accumulated knowledge and experience concerning celiac disease, was developed with a view to reducing workload. Its primary task is to provide interpretation. This automated diagnosis permits rapid evaluations of a large number of sera, the majority of which only have to be signed by the expert. By employing the underlying knowledge base designed as a decision graph, CeliPert interprets all available results of a patient's blood and biopsy samples with respect to his age and current dietary condition, and arrives at an assessment regarding celiac disease. CeliPert is designed as a flexible system: It contains a graphics-based knowledge acquisition system that enables modification of its knowledge base according to the most recent developments in this field of medicine.*

1. Background for the Development of the Expert System

At the gastroenterology laboratory of the University Children's Hospital of Vienna, examinations of patients' sera for antigliadin antibodies and endomysial antibodies are an important step in the work-up of patients suspected to have celiac disease. The test results of every serum sample are supplemented by an expert's comments regarding the diagnostic value of the results and, if necessary, recommendations for further diagnostic and therapeutic procedures. The steadily

increasing number of serum samples due for work-up made it necessary to devise measures that would reduce the workload without compromising on quality. The solution to the problem was a computer-aided procedure in the form of a medical expert system known as CeliPert [1].

2. What is Celiac Disease?

Celiac disease or gluten-sensitive enteropathy is a small intestinal enteropathy induced by the oral intake of gluten in genetically predisposed individuals. It is a lifelong disorder that may present in children or adults. Removal of gluten from the diet leads to full remission [2].

Cereals, a basic food for humans, mainly contain starch and (up to 15%) proteins. The water-insoluble protein residue remaining after starch has been extracted from the dough made from flour is known as gluten. The alcohol-soluble fraction of wheat gluten is termed gliadin. In clinical practice, gliadin and gluten are frequently used as synonyms. We usually ingest gliadin in the form of bread, pasta, cereals, or sweets. Protein fractions of rye, barley, and possibly oats may cause the same damage. Similar proteins from rice, maize, millet, and buckwheat are considered non-toxic.

In celiac patients the oral intake of gluten leads to T-cell induced damage of the small bowel mucosa. Samples of this tissue are gained by an oral biopsy: The villi, finger- or leaflike protrusions of the inner surface of the small bowel are stunted or absent (villous atrophy, "flat mucosa") and the crypts, short, tubular recesses are elongated (crypt hyperplasia). The density of plasma cells in the mucosa is increased, as is the number of lymphocytes in the epithelium. The absorbing surface of the mucosa is markedly reduced, resulting in malabsorption. Depending on the length of small bowel involved, this may lead, on the one hand, to deficiency states and clinical signs such as abnormal stool, undernutrition, and failure to thrive, i.e., symptomatic celiac disease, or, on the other hand, the patient may have no symptoms, i.e., silent celiac disease.

The cornerstone of the treatment of celiac disease is strict and lifelong avoidance of toxic proteins in the diet, i.e., a diet devoid of wheat, rye, barley, and oats or a gluten-free diet.

3. What is Done in the Laboratory?

Shortly after gliadin had been demonstrated as the pathogenic exogenous agent in celiac disease, antigliadin antibodies were found in the serum of celiac patients [3]. Subsequently, the estimation of immunoglobulin G antigliadin antibodies (AGA-IgG) and later of immunoglobulin A (IgA) antigliadin antibodies (AGA-IgA) [4] became an important tool for the diagnosis of celiac disease, especially in children [5]. The ELISA technique, performed in microtiter plates, is the generally accepted procedure. Antibodies of the IgA class are of greater specificity while antibodies of the IgG class are of greater sensitivity and can even be found in IgA-deficient patients. In adults, antigliadin antibodies are less reliable due to their lower sensitivity in this age group [6].

This diagnostic gap was filled by the detection of so-called endomysial antibodies (EMA) in indirect immunofluorescence on sections of monkey esophagus [7]. EMA are mainly of the IgA class, and can be interpreted as auto-antibodies produced in dependence of the oral intake of gliadin. In adults, once an IgA deficiency has been excluded, the sensitivity seems to approach 100% [8]. In children, especially in those under two years of age, the sensitivity is somewhat lower. Thus, false negative results are occasionally obtained [9]. On the other hand, EMA are sometimes found in healthy individuals with a normal small bowel mucosa, particularly when screening family members of known celiac patients or other populations at risk such as patients with type I diabetes mellitus. Since villous atrophy was demonstrated years later in several EMA positive patients [10], the term "potential celiac disease" is used for this constellation [11]. Taking into account this additional group of patients, the specificity of EMA seems to be very close to 100%.

Tissue transglutaminase is possibly the only or at least the major antigen, reacting with the serum antibodies shown to be endomysium antibodies [12]. The discovery of a defined antigen made available the use of the ELISA technique, apt for automation, instead of the more tedious indirect immunofluorescence procedure. The first assays using guinea pig transglutaminase resulted in more than 90% corresponding results [13].

At the gastroenterology laboratory of the University Children's Hospital of Vienna, about 7,000 patients' sera per year are tested for antigliadin antibodies of class IgG and IgA, as well as endomysial antibodies of class IgA. Transglutaminase antibodies are not yet included in the routine procedure. For every serum, the test results are supplemented by an expert's comment.

4. What Should the Expert System Accomplish?

A computer-aided system was developed to assist in the complex procedure of making a diagnosis and providing suggestions for therapy. CeliPert is a medical expert system based on accumulated specific knowledge and experience on celiac disease. The input of the results of serological examinations and biopsy samples combined with the patient's age, current dietary state, and its duration, finally results in the output of a diagnostic report. The report contains the respective medical diagnosis as well as suggestions for further examinations and behavior, which will be comprehensible even to physicians who have not been specifically trained in this field. This automated diagnostic procedure permits rapid evaluations of a large number of serum samples, which only have to be signed by the expert. Rare or unusual constellations of findings need to be supplemented by the expert's personal comment.

5. The Structure and Environment of CeliPert

CeliPert was developed with the object-oriented programming language Java. It uses the database of MedFrame [14,15], an expert system shell for developing medical expert systems that is currently implemented at the Section of Medical Expert and Knowledge-Based Systems of the Department of Medical Computer Sciences of the University of Vienna, which is located in the Vienna General Hospital. Poet 6.1, an object-oriented database management system, was selected to provide permanent storage capabilities for CeliPert.

CeliPert is based on a client/server architecture consisting of three components—CeliServer, CeliApplet, and CeliBuilder—and can be executed either within a World Wide Web (WWW) browser or from an autonomous program.

CeliServer forms the server module of CeliPert and is used by both, CeliBuilder and CeliApplet, to perform their tasks. CeliServer can only be executed as a stand-alone application and is responsible for database access, communication management, and performing inferences on patient's data.

CeliApplet provides an easily usable means of interpreting the results of a patient's blood and biopsy samples by triggering the inference mechanism of CeliServer. CeliApplet can be used only within WWW browsers.

CeliBuilder may be executed as a stand-alone application or within WWW browsers. It contains a graphics-based knowledge acquisition system that permits modification of CeliPert's knowledge base. Besides, CeliBuilder can also execute CeliApplet inside of its frames and thus enables the knowledge engineer to immediately test a new knowledge base.

The user interface of CeliApplet was designed to fulfill the required functions in a simple fashion. The patient's personal data, dietary states, results of biopsies, and serological test results are specified in the respective input masks. Once the inference is started, a detailed inference log with the derived interpretation is displayed on the screen.

6. Knowledge Acquisition for CeliPert

The accumulated knowledge and experience on celiac disease and serological diagnostics had to be reorganized in a kind of engineering process in order to permit an inference process. The knowledge acquisition process comprised the determination of interpretations, so that all possible cases are covered, as well as the identification of serological and clinical preconditions indicating a particular interpretation.

According to the current dietary state of a patient, three interpretation groups were defined: Normal food (NF), gluten-free diet (GFD), subdivided in strict and inaccurate diet, and gluten loading (GL).

The following list contains the basic rules of the diagnostic process:

- EMA positive means high suspicion of celiac disease, even without any clinical or biochemical signs. A small bowel biopsy has to be performed. If a flat mucosa is documented, the suspected celiac disease is confirmed. Even if the biopsy reveals completely normal mucosa, celiac disease is not excluded. The patient with EMA in his serum and a normal small bowel mucosa has to be estimated as a case of "potential celiac disease" and must be carefully followed under normal diet.

- In contrast, positive results in IgA-AGA and especially in IgG-AGA cannot be directly equated with celiac disease. The terms 'positive' and 'negative' are in general use, and are expected by the clinician. However, lower and higher numbers form a continuum with lesser or greater probability of celiac disease. Thus, it would be better not to use 'positive' and 'negative' at all when estimating antigliadin antibodies.

- Depending on the age of the patient, history, clinical picture, signs of malabsorption, and duration of gliadin intake, the individual level of antigliadin antibodies indicates a greater or lesser probability of celiac disease.

- In younger children with untreated celiac disease, usually EMA, IgG-AGA, and IgA-AGA are highly positive. In some of these EMA may be falsely negative.

- In adults (and older children) with celiac disease, IgG-AGA is occasionally positive in the lower range or even negative. All EMA, IgG-AGA, and IgA-AGA estimations are more frequently positive.

- A patient with a selective IgA deficiency (when the concentration in serum of total IgA is less than 5.0 mg/100 ml) cannot produce any EMA-IgA or AGA-IgA, even when suffering from untreated celiac disease. Routine estimation of total IgA is the optimal procedure. At least in patients with "EMA-negative, IgA-AGA-negative, or IgG-AGA-positive" test results as well as in those with strong clinical signs of celiac disease and negative results of antibody testing, a selective IgA deficiency must be ruled out by this procedure.

- If EMA are found negative and IgG-AGA and IgA-AGA positive, with values in the medium range, the recommendation for serum controls or small bowel biopsy will depend on the clinical presentation and the presence or absence of signs of malabsorption.

- In patients with negative EMA, IgG-AGA, and IgA-AGA estimations, celiac disease can be excluded with high probability.

- Celiac patients on a gluten-free diet may be expected to show a continuous reduction in the mentioned antibody concentrations, more rapidly in IgA antibodies and more slowly in IgG antibodies.

- Gluten loading will, in many cases, lead to positive antibody readings after a variable length of time. Dietary transgressions may or may not result in positive antibody estimations, especially of EMA.

Knowledge acquisition was done by a knowledge engineer, both by directly eliciting knowledge from the medical expert in several interviews and by reviewing about 6,900 cases in the respective database. The elicited knowledge was encoded for the knowledge base and entered with the help of the graphical knowledge acquisition system. Finally, the knowledge base was refined in close cooperation with the medical expert in order to achieve acceptable performance.

7. Knowledge Representation and Inference Mechanism of CeliPert

Knowledge representation in CeliPert is done in a rule-based fashion by means of a so-called decision graph or knowledge graph that may be interpreted as a finite state automaton consisting of states (represented by nodes in the graph) and transitions (corresponding to edges). Each state corresponds to a medical interpretation—consisting of a diagnosis and suggestions for further examinations and behavior—and belongs to one of three interpretative categories: (a) normal food (NF), (b) gluten-free diet (GFD), and (c) gluten loading (GL). The decision criteria that are equivalent to the serological and clinical preconditions for a particular interpretation are attached to the respective states as entry conditions. All states have to be connected directly or through other states with the initial state, where the inference mechanism starts.

The elements of the graph can be used to formulate if-then sentences: If premise, then conclusion. For instance:

If (EMA $\geq$ 10) and (AGA-IgG $\geq$ 35) and (AGA-IgA $\geq$ 25) and (dietary state = NF),
then interpretation 1.24 : This finding implies untreated celiac disease.

The inference engine is realized by deductive reasoning within the decision graph. A short description of an example inference will explain it easily:

Table 1 contains a sequence of a patient's findings, arranged by increasing age, as used for automatic interpretation in CeliPert.

Inference in CeliPert is done step by step in the decision graph for each finding of a patient— composed of the results of serological examinations (EMA, AGA-IgA, AGA-IgG), the patient's age, the current dietary state and its duration, and the result of the biopsy sample, if available. For each finding the possible transitions are evaluated. As there are no conditions attached to the

transitions, a finding corresponds to a transition when it satisfies the entry condition of the state to which the transition is leading. Only one transition from the current state will be fulfilled by the current finding since the knowledge engineer had to ensure that the conditions of all transitions that leave a state not only cover all possible findings but also are mutually exclusive. The transition that corresponds to the current finding is used. The state to which the transition leads becomes the new current state with the corresponding interpretation. Thus, the inference for the first finding of a 1.4-year-old child on a gluten-containing diet since 12 months, with EMA and AGAs strongly positive, will lead to the interpretation "this finding implies untreated celiac disease."

Table 1: Exemplary sequence of findings.

Finding	EMA	AGA-IgG	AGA-IgA	Age (y)	Diet	Time (m)	Biopsy
1	++++ (40)	> 200.0	> 200.0	1.41	NF	12.00	
2	++ (20)	78.9	56.2	1.58	GFD	2.49	flat mucosa
3	+ (10)	45.1	27.2	1.83	GFD	5.52	
4	-	40.0	12.3	2.50	GFD	13.59	

As long as further findings are available, the process is repeated for the next one. When no further findings are available, the interpretative process ceases and the interpretation of the most current state becomes the result of the deduction. Thus, every state in the knowledge graph is a potential final point of an inference.

The inference for finding 4 in the example reference will, under consideration of findings 1 to 3 known from earlier investigations, finally lead to the interpretation "This finding implies that the patient is on a gluten-free diet."

The decision graph of CeliPert comprises the start node and 80 nodes (interpretations) that are interconnected with 3,789 edges. The 80 nodes are allocated to the three interpretation groups in the following way: normal diet: 41, gluten-free diet: 24, and gluten loading: 15.

The decision graph is mathematically complete: All possible situations that may occur in the search for and course of celiac disease are covered, and a large number of findings can be processed.

8. Practical Use of CeliPert in Laboratory Routine

After a test run with some thousand sera, a slightly modified version of CeliPert will be integrated into the laboratory system CeliLab—a part of which is already in use—that affords extensive functions to support the entire sequence of operations performed in the process of examination. Finally, the process will start with the input of the patients' data, printing of self-adhesive bar code labels for the serum tubes, and of working lists for usual laboratory work-up. Optical density values

of the ELISA procedure in microtiter plates (AGA) enter the system online while the results of readings in the fluorescence microscope (EMA) have to be typed. The patient's data, all single test results of the serum under investigation, and the final result of the inference mechanism, i.e., CeliPert's proposal of interpretation, will be shown on the screen. The gastroenterologist in charge may accept and sign, after which the final text will be printed. Alternatively, he may alter or remove parts of the text, add prepared sentences from a "thesaurus", or formulate a new text. If the evaluation of a serum sample is not yet completed, the sample is automatically put on the working list of the next day.

9. Conclusion

CeliPert is a knowledge-based system that helps the physician to interpret the results of serological tests made in patients suspected to have celiac disease or patients being followed up for celiac disease. By categorizing most potential cases automatically, it relieves the laboratory staff of tedious routine work and enables them to concentrate on contradictory and more challenging cases that need the expert's personal comment. CeliPert is part of the fully integrated laboratory system CeliLab, that provides extensive functions to partly automate the daily laboratory operations processes.

A graphical knowledge acquisition system serves to extend and modify the underlying knowledge base. Since the collected data of a patient need not to be stored in the database to perform an inference, CeliApplet provides the possibility to simply experiment with a great variety of combinations of serological test results, biopsy results, and dietary states. The ongoing application of CeliPert will amplify experience on celiac disease in a stepwise fashion. The signing gastroenterologist may alter the interpretation by punctual modification of the knowledge graph. CeliPert is designed as a flexible system that can be adapted not only to changes in general dietary habits and in patients' reaction to the noxious agent, but also to the introduction of new tests in the laboratory and thus to most recent scientific and clinical discoveries in this field of medicine.

In order to provide a means to access the benefits of CeliPert from WWW browsers not capable of executing Java Applets, an HTML-based interface to the inference mechanism was implemented. Once anonymous test results are entered into standard HTML forms, the corresponding interpretation is presented in a traditional HTML page. This implementation can be accessed at http://medexpert.imc.akh-wien.ac.at/CeliWWW. By providing knowledge about interpretation of serological tests for celiac disease in the WWW, its information is made available to clinics worldwide and even to external users.

References

[1] Cirillo, S. (2001) *CeliPert: A Medical Expert System for the Knowledge-Based Interpretation of Test Results for Coeliac Disease*. Master Thesis, University of Vienna, Austria.

[2] Walker-Smith, J.A. (2000) Coeliac Disease. In Walker, W.A., Durie, P.R., Hamilton, J.R., Walker-Smith, J.A., and Watkins, J.B. (eds.) *Pediatric Gastrointestinal Disease*. B.C.Decker Inc., Hamilton, Ontario, 727–746.

[3] Berger, E. (1958) Zur allergischen Pathogenese der Cöliakie. *Bibliotheca Paediatrica* 67, 5–55.

[4] Savilahti, E., Viander, M., Perkkio, M., Vainio, E., Kalimo, K., and Reunala,T. (1983) IgA Antigliadin Antibodies: A Marker of Mucosal Damage in Childhood Coeliac Disease. *Lancet* i, 320–322.

[5] Bürgin-Wolff, A. Berger, R. Gaze,H., Huber,H., Lentze,M.J., and Nussle, D. (1989) IgG, IgA and IgE Gliadin Antibody Determinations as Screening Test for Untreated Coeliac Disease in Children, a Multicentre Study. *European Journal of Pediatrics* 148, 496–502.

[6] Troncone, R. and Ferguson, A. (1991) Anti-Gliadin Antibodies. *Journal of Pediatric Gastroenterology and Nutrition* 12, 150–158.

[7] Chorzelski, T.P., Sulej, J., Tchorzewska, H., Jablonska, S., Beutner, E.H., and Kumar, V. (1983) IgA Class Endomysium Antibodies in Dermatitis Herpetiformis and Coeliac Disease. *Annals of the New York Academy of Sciences* 420, 324–325.

[8] Kumar, V., Lerner, A., Valeski, J.E., Beutner, E.H., Chorzelski, T.P., and Rossi,T. (1989) Endomysial Antibodies in the Diagnosis of Coeliac Disease and the Effect of Gluten on Antibody Titers. *Immunological Investigations* 18, 533–544.

[9] Bürgin-Wolff, A. et al. (1991) Antigliadin and Antiendomysium Antibody Determination for Coeliac Disease. *Archives of Diseases in Childhood* 66, 942–947.

[10] Högberg, L., Stenhammar, L., Falth-Magnusson, K., and Grodzinsky, E. (1997) Ani-Endomysium and Antigliadin Antibodies as Serological Markers for a Very Late Mucosal relapse in a Coeliac Girl. *Acta Paediatrica* 82, 887–889.

[11] Ferguson, A., Arranz, E., and O'Mahony, S. (1993) Clinical and Pathological Spectrum of Coeliac Disease – Active, Silent, Latent, Potential. *Gut* 34, 150–151.

[12] Dieterich,W., Ehnis, T., Bauer, M., Donner,P., Volta, V., Riecken, E.O., and Schuppan, D. (1997) Identification of Tissue Transglutaminase as the Autoantigen of Coeliac Disease. *Nature Medicine* 3, 797–801.

[13] Troncone, R. et al. (1999) IgA Antibodies to Tissue Transglutaminase: An Effective Diagnostic Test for Coeliac Disease. *Journal of Pediatrics* 134, 166–171.

[14] Kolousek, G. (1997) *The System Architecture of an Integrated Medical Consultation System and Its Implementation Based on Fuzzy Technology*. PhD Thesis, Vienna University of Technology, Vienna, Austria.

[15] Kopecky, D. (2001) An Overview of the Current State of the MedFrame/CADIAG-IV Project. In Adlassnig, K.-P. (ed.) *Intelligent Systems in Patient Care*. Österreichische Computer Gesellschaft, Wien, 89–96.