RESEARCH ARTICLE

A Fuzzy-Based Clinical Decision Support System for Coeliac Disease

M. E. TABACCHI [1], **D. TEGOLO** [1], **D. CASCIO** [2], **C. VALENTI** [1], **S. SORCE** [3], **V. GENTILE** [2],
V. TAORMINA [1], **I. BRUSCA** [4], **G. MAGAZZÙ** [5], **A. GIULIANO** [6], **AND G. RASO** [2]

[1]Dipartimento di Matematica e Informatica, Università degli Studi di Palermo, 90123 Palermo, Italy
[2]Dipartimento di Fisica e Chimica, Università degli Studi di Palermo, 90123 Palermo, Italy
[3]Facoltà di Ingegneria e Architettura, Università degli Studi di Enna "Kore," 94100 Enna, Italy
[4]Ospedale Fatebenefratelli, Buccheri La Ferla, 90123 Palermo, Italy
[5]Dipartimento di Patologia Umana dell'adulto e dell'età evolutiva, Università di Messina, 98121 Messina, Italy
[6]AcrossLimits Ltd., 1520 Valletta, Malta

Corresponding author: D. Cascio (donato.cascio@unipa.it)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was
granted by the Local Ethics Committees (Messina and Malta).

ABSTRACT Coeliac disease (CD) is a permanent inflammatory disease of the small intestine characterized by the destruction of the mucous membrane of this intestinal tract. Coeliac disease represents the most frequent food intolerance and affects about 1% of the population, but it is severely underdiagnosed. Currently available guidelines require CD-specific serology and atrophic histology in duodenal biopsy samples to diagnose CD in adults. In paediatric CD, but recently in adults also, non-invasive diagnostic strategies have become increasingly popular. In order to increase the rates of correct diagnosis of the disease without the use of biopsy, researchers have recently been using approaches based on artificial intelligence techniques. In this work, we present a Clinical Decision Support System (CDSS)system for supporting CD diagnosis, developed in the context of the Italy-Malta cross-border project ITAMA. The implemented CDSS has been based on a neural-network-based fuzzy classifier. The system was developed and tested using a Virtual Database and a Real Database acquired during the ITAMA project. Analysis on 10,000 virtual patients shows that the system achieved an accuracy of 99% and a sensitivity of 99%. On 19,415 real patients, of which 109 with a confirmed diagnosis of coeliac disease, the system achieved 99.6% accuracy, 85.7% sensitivity, 99.6% specificity and 96% precision. Such results show that the developed system can be used effectively to support the diagnosis of the CD by reducing the appeal to invasive techniques such as biopsy.

INDEX TERMS Coeliac disease, computer aided diagnosis, artificial intelligence, endoscopy, neural network, fuzzy classifier, CDSS.

I. INTRODUCTION

Coeliac disease (CD) is a rapidly expanding disorder both in terms of prevalence in the world and in terms of a more significant number of diagnosed patients; it is an autoimmune disease that can occur at all stages of life. Advances in understanding the pathogenetic and genetic factors that

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

influence risk have led to the development and refinement of diagnostic tools. It is a chronic disease of the small intestine characterized by an abnormal immune response; the latter is due to exposure to gluten present in the diet in genetically predisposed subjects. The "environmental" factor triggering coeliac disease is represented by gluten, a protein complex contained in some cereals (wheat, barley, rye) [1]. Coeliac disease is an autoimmune disorder induced by dietary gluten in genetically predisposed subjects. CD has

a prevalence of ~ 1% in many populations around the world, and the breadth of established clinical presentations continues to increase, making the disorder a significant relevance in the medical field [2], [3]. The CD has been analysed in its many aspects, and although pathogenesis and pathophysiology remain unknown, it is assumed that the disease is strictly connected to genetic interactions, environmental and immunological factors.

Like other underdiagnosed disorders, CD is depicted as an iceberg of which the most considerable part is submerged [4]. There are no particular manifestations in the silent form of coeliac disease, and for this reason, it is difficult to diagnose. The latent form instead characterizes those subjects who, despite having a predisposition to coeliac disease (positivity of AGA anti-gliadin antibodies and anti-endomysial EMA antibodies), currently have a normal intestinal mucosa that does not present atrophy of the villi. However, atrophy will appear after some time, and therefore periodic monitoring is necessary. With a major awareness of the disease an increasing number of patients are diagnosed and therefore, as with many other autoimmune disorders, the real incidence in the population seems to have increased [5].

In diagnosing coeliac disease, serology is usually the first step in diagnosing or ruling out the disease in symptomatic patients or for screening. The biopsy is essential for the definitive diagnosis of the pathology. The serological markers of coeliac disease are: IgA against tTG, Endomysial antibodies (IgA), IgG against DGD, IgA versus deamidated gliadin peptide, IgG versus tTG. A small number of coeliac disease patients have had negative serological test results. Therefore, biopsies should be performed if there is a high clinical suspicion of coeliac disease, regardless of these findings. For asymptomatic patients, especially children, who have slight increases in serological markers of the disease, biopsy analysis can be delayed. Level gastroscopy duodenojejunal with intestinal biopsy is undoubtedly helpful to confirm the diagnosis and to ascertain the degree, being an invasive examination, it is desirable, especially in children, that it is carried out only when necessary.

The European Society for Paediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) guidelines [6] report the possibility of avoiding biopsy to diagnose coeliac disease in genetically susceptible children with high titres of tTG_IgA antibodies. Unfortunately, it is only possible in a few patients and is influenced by the lack of standardization of anti tTG_IgA kits.

In order to satisfy the heterogeneity of data and their complexity, a Decision Support System (DSS) has to be considered for their manipulation. DSSs continue to be increasingly requested in the clinical setting, and there are still many open problems even in the field of interoperability. Due to the multidisciplinary nature of a DSS, certain precautions to meet the needs of interoperability must be taken into consideration in the design phase during the construction, use and maintenance of the DSS. Some of these considerations were addressed by Sutton *et al.* in [7] as DSS evolve in complexity

(Artificial Intelligence), interoperability (multidisciplinarity) and data sources (Cloud, open data,...). Thus, decision support systems become a relevant part of the tools that use artificial intelligence (AI). They have the task of solving open questions to deepen the understanding of the correlations between the data representing the events. The DSS, supported by multidisciplinary approaches, revisits how data are treated and analysed and generates virtually and globally cognitive pathways that highlight unconventional solutions and considerations. This approach improves how data are analysed and understood, their knowledge and correlation.

knowledge-based, data-driven, or lacking a priori knowledge. The strategy of the former is based on rules that are not necessarily deterministic; it recovers data from information systems (i.e.: databases) or in real-time from Biometric systems and evaluates the rules involved. Finally, it produces an output event (Alarm, screening, diagnostic pathway,...). Non-knowledge-based DSSs are data-driven, and the output events result from modelling applications on machine learning with no specific medical knowledge needs to take into account to set up the model. Such a model without knowledge, adopted for the creation of the DSS, is currently being studied in the scientific community; they are incredibly complex to implement, and they leave no room for understanding the results, whether they are correct or incorrect, even when they have a high degree of sensitivity and specificity.

The inherent imprecision of medical data, as well as the fact that a patient enters the diagnostic pathways from different medical sources (as an outpatient, an inpatient, referred by a physician, after blood tests and other unrelated diagnostic tools have been administered) makes standard classification methods less easy to adapt to the CDSS backend. While well-known classifiers such as SVM and NN produce precise results on binary classification problems, the diagnosis of a coeliac patient requires a number of steps, and a CDSS should offer prioritisation advice on each of these steps, regardless the completion of the whole diagnostic pathway. Fuzzy classifiers allow for taking into account this inherent dynamicity and imprecision.

In this work, we present a fuzzy-based Clinical Diagnostic Support System developed within the ITAMA project (henceforth ITAMACDSS). ITAMA (ICT Tools for the diagnosis of Autoimmune diseases in the Mediterranean Area, [8]) is a cross-border project between Italy and Malta funded by the European Regional Development Fund within the INTERREG V-A Italia - Malta Cooperation Programme, in which the common territorial challenge is to improve the quality of life and well-being of the population affected by autoimmune diseases, containing the costs of health systems through a strategic commissioning demand towards the world of research. In Sicily and Malta, autoimmune diseases present a high incidence, probably due to the high consumption of starchy foods. In ITAMA, a mass screening was carried out on more than 20,000 Maltese children. The screening was based on a Medical History Questionnaire (MHQ) and a Point-of-Care Test (PoCT). Children tested positive based on the

result of the MHQ, of the PoCT, or both, were invited for further investigation, and in particular for the anti-Actin IgA to verify the possibility of avoiding biopsy in a large number of patients.

The implemented CDSS is based on a fuzzy classifier using neural networks. The system was developed and tested using a Virtual Database and a Real Database acquired during the ITAMA project.

Since the objective was to minimize the length and impact of the diagnostic pathway for a correct diagnosis of celiac disease while maximizing the effectiveness, it was necessary to validate the CDSS with numerous pilot tests both on simulated data generated in collaboration with medical experts and on real data acquired in real contexts (hospitals). The validation using 'virtual patients' (i.e. artificially generated data based on current knowledge of the disease and its symptoms/related conditions) was a preparatory step in order to reach a starting point of the system before completion of the data collection on real patients, and is reported for comparison purposes.

Despite a large amount of data available offered by the acquisition in the clinical field, the developed Framework intends to propose an architecture that can also be used in those sectors in which the data are not numerous and therefore cannot take advantage of methods based on the DL. Our Framework, whose intelligence is based on Fuzzy rules and therefore can be analysed and verifiable, acquires anamnestic data validated by doctors or specific health personnel, provides both suggestions on personalized diagnostic paths and interpretations of the data to lead to a faster diagnosis of celiac disease.

This paper is organized as follows. Section II discusses existing literature. In Section III the databases – a virtual one (III.A) generated using knowledge from the diagnostic state of the art in coeliac disease, and a real one (III.B) obtained through data collected by the ITAMA project during a mass screening – on which the CDSS Fuzzy classifier has been trained are presented. Section IV details data cleaning and extraction procedure, the proposed approach for designing a fuzzy-based CDSS for coeliac disease, its architectural implementation and the communication protocol with ITAMA DB. In Section V results from the classification procedure are presented and discussed, and hints at the complete implementation of the CDSS system are given. Discussion on the results as well as further considerations are given in Section IV.

II. RELATED WORKS

In general, the DSSs based on Deep Learning (DL), Machine Learning (ML), and Neural Networks (NN) are very efficient if applied to training with large data sets; but large data sets are not always available in the medical domain [9].

Qatawneh *et al.* present in [10] a clinical decision support system based on Artificial Neural Networks (ANN) to predict the risk of developing Venous Thromboembolism (VTE). A dataset with 150 medical records was used for training and testing, and the system was trained using a resilient backpropagation algorithm with a ten-fold cross-validation scheme to assess the generalization of the system thus, the results show an accuracy of the system is 81%.

The introduction of DSSs for diagnosing CD could improve diagnostic work-up, allowing cost, time and labour savings and improving the procedure's safety, avoiding biopsy sampling and prolonged sedation associated with the multiple biopsy protocol. In particular, DSS based on Fuzzy Logic are enjoying growing research interest in solving classification problems in a wide range of application fields [11], especially in medicine, where the possibility of presenting classification results together with a measurement of the association is very tempting [12].

The interest of the scientific community in the development of DSS systems, also thanks to new performing machine learning techniques, is certainly growing [13], [14]. However, the problem of developing CD diagnosis support systems is still poorly explored, perhaps due to the difficulty of the problem but undoubtedly also due to the lack of public databases. For example, in a recent review work [15], after proper research, the authors have identified only 41 publications consisting of original work describing techniques for computer-aided CD diagnosis.

Gadermayr *et al.* [16] summarize recent trends in computer-aided coeliac disease diagnosis based on upper endoscopy and proposed pipelines for fully-automated patient-wise diagnosis and for integrating expert knowledge into the automated decision process.

In [17], the authors presented a feature descriptor for the classification of video capsule endoscopy images. In addition, they introduced a system for small intestine motility characterization based on deep CNN for individual motility events. Experimental results showed a mean classification accuracy of 96% for six intestinal motility events.

In the last few years, deep learning methods have also been used to classify endoscopic images. In this context, the best known convolutional neural networks, i.e. AlexNet [18], [19], GoogLeNet [20], VGGf net [21], and VGG16 net [22], have been used for this purpose.

Wang *et al.* in [23] propose a deep-learning-based methodology to recalibrate the module to identify images with regions significant for celiac disease from healthy ones. The developed module determines the most salient feature on the features' map and is hooked to a Support Vector Machine and a k-nearest neighbour module to perform a linear discriminant analysis. Their method reports 95.94%, 97.20% and 95.63% for accuracy, sensitivity and specificity with the 10-time 10-fold cross-validation strategy.

Amirkhani *et al.* [24] developed a method based on a fuzzy cognitive map (FCM) and a possibilistic fuzzy C-means clustering algorithm (PFCM) for the categorization of CD. The research goal was to develop an expert system for classifying patients with CD into three grades A, B1, and B2, which is the latest grading method available. Three experts have extracted seven key defining features of CDs that were considered FCM

concepts. For the three analysis classes, the authors obtained 88%, 90% and 91% accuracy, respectively.

The authors of [25] proposed a fuzzy logic-based method to predict coeliac disease by entering sharp values of various symptoms. The analysis was conducted on 700 individuals. The system, which was based on the Mamdani model, shows an accuracy of 96.11%.

However, to our knowledge, to date, there are no works in which clinical support systems for coeliac disease screening have been developed, i.e. designed to work from non-invasive diagnostic tests (and to limit the use of biopsy). This is probably due to the paucity of screening data as well as the complexity of the diagnostic problem.

III. THE DATABASES

To develop and test the CDSS, we first had to create a Virtual Patient Database, the rationale of which is explained in section III.A. Basically, this was due to the Sars-CoV-2 pandemic, which delayed the actual data collection within the project. The Virtual DB schema is compatible one-to-one with the Real DB one, so that as the real data was acquired, they can be appropriately transferred into the complete DB (virtual + real), and in this sense, they can be used by the CDSS without significant changes. The Real DB and the actual collected data are described in section III.B. As a result, one million virtual patients have been generated.

A. VIRTUAL DATABASE

In order to optimize and speed up the CDSS development, a way has been devised to massively generate data from "virtual patients". Such data has been generated by assessing the scientific literature to derive known incidences of coeliac disease factors in the general population and, where appropriate, by treating such variables as an independent. The data was discussed and approved with the ITAMA medical staff of experts in coeliac disease. Furthermore, the variables that have been included mirror precisely the original variables from the ITAMA DB, and as such, have been modified during the project. The final list of variables, shown in Table 4, is coherent with the list from real data acquisition, allowing the virtual model to be fully interchangeable with the real model, for comparison purposes. The ITAMA group has made a public version of the virtual database available to the scientific community.[1]

The virtual DB was created using statistical information on the pathology. Specifically, since the prevalence of the celiac disease in the general population is 1:100 [26], a set of patient batches was generated, each consisting of 1 positive patient and a random number of negative patients extracted from a Gaussian distribution with a mean of 100 and variance of two standard deviations. Virtual data is generated as follows.

[1] https://app.itamaproject.eu/virtualdb

1) FOR NEGATIVE PATIENTS

Questionnaires have been generated so that epidemiological data reflects the statistical prevalence of diseases (where known) in the general population: Anaemia (1:4), Osteopenia (1:3), Chronic Diarrhoea (1:20), Failure to Grow (1:140), Genetic Disorders (1:1000), Coeliac Mother (1:100). In the computation of the cumulative number of positive questions, the resulting value is scaled for consistency with the corresponding data for the real patients.

The Point-of-Care Test (PoCT) has negative or inconclusive outcome (highly unlikely positive PoCT), maintaining the distribution of inconclusive (1:600), and considering a number of defective tests equal to 1:1200, which is consistent with literature. In the case of negative PoCT and negative Questionnaire, the Blood Test has missing values. Otherwise (positive PoCT or positive MHQ), the logic detailed below is followed. First, a value for the total IgA is generated, following the PoCT result: if PoCT is negative, IgA generated value is higher than the threshold with mean 7 and variance 2 stdev; if PoCT is inconclusive, IgA has lower random values with random distribution between 0 and 0.25. In the case of a deficit of the total IgA, a value is generated for the tTG_IgG with mean 2 and variance 2 stdev and the value of the tTG_IgA is missing. In the other cases (i.e. if the total IgAs are sufficient) a value for the tTG_IgA is generated from a Gaussian distribution with mean 4.5 and variance 2 stdev, and the value for the tTG_IgG will be missing.

If the blood test is positive, the Biopsy will obviously have a negative result (class 1 or 2 – the patient being generated is CD-negative), otherwise it has a missing value.

2) FOR POSITIVE PATIENTS

The distribution of positive responses in the Questionnaire is reviewed considering the known prevalence, compared to that used for the questionnaire of negative cases: Anaemia (1:2), Osteopenia (2:5), Chronic Diarrhoea (1:3), Failure to Grow (1:5), Genetic Disorders (1:20), Coeliac Mother (1:18).

The PoCT is positive (599:600) or inconclusive (1:600). Blood Tests follow a logic similar to that for negative cases but will always be positive: first a value for the total IgA is generated. In the case of inconclusive PoCT, with mean 0.125 and variance 1 stdev, otherwise mean 8 and variance 2 stdev. In the case of inconclusive PoCT, a value is generated for the tTG_IgG with mean 14 and variance 2 stdev and the value of the tTG_IgA will remain missing. Values in the negative range are discarded. In the case of a positive PoCT, a value for the tTG_IgA is generated from a Gaussian distribution with mean 24 and long tail on the right, and the value for the tTG_IgG will remain missing. Values in the negative range are discarded

The Biopsy has assigned a random positive evaluation with uniform 1:3 distribution among classes 3a, 3b, 3c.

B. REAL DATABASE

The real database was designed with the primary goal to support the ITAMA project to reach its objectives. Therefore, the data relating to the subjects have been entered after they have given informed consent. The resulting schema considers all known constraints and requirements both from ethical, technical and functional points of view, and at the same time, it is ready to be extended for further possible functions to be supported. Other general guidelines and best practices were followed for its design, such as data isolation and interoperability.

As far as data isolation is concerned, the DB schema is provides for the following features:

- **minimal data redundancy**: each table represents a single piece of information so that it can be easily identified as a table storing the related data;
- **privacy**: any possible personal data can be stored in a single table, thus allowing for controlled access to it and separation from the remaining data, with no impact on the information extraction and data aggregation;
- **reusability**: the schema can be re-used, as is or even partially, in different same or similar contexts;
- **safety**: there are no "critical" tables, so losing someone will have little impact on the stored information.

As far as interoperability is concerned, we used some well-known best practices: all table names start with a lowercase 't', all field names do not contain non-ASCII characters, and they are composed (as much as possible) of meaningful groups of characters, making it easier to understand the data they represent and the table they belong to.

The current schema of the database is composed of twenty-nine tables, based on which it is possible to build all the queries needed to achieve the project's goals, in terms of suitably structured data storage, and information extraction for statistical analysis and support for the early diagnosis process. Concerning the current instance of the DB, i.e. the actual stored data, it consists of ∼189k rows in total. In more detail, it stores 20,454 patients' basic information, spread on 4 tables and 103,513 rows:

- Demographics (age, gender, ethnicity);
- Medical history (answers to 29 multiple-choice questions);
- Point-of-Care (PoCT) data (pictures, results).

The results of second-level blood tests for 875 patients are spread among 12 tables and 2573 rows storing values for tTG_IgA, tTG_IgG, Total_IgA, EMA, AntiActine AAC, and DPG_IgG exams.

Two more tables contain third-level (endoscopy) exams results for 165 patients, namely biopsy (evaluation based on Marsh index) and mucosal deposits (pictures, evaluation).

The final evaluation and the related details are spread on two tables and 39,807 rows, storing for each candidate the complete diagnostic pathway (doctors' decisions on the diagnostic steps for each participant), and the final diagnosis (Coeliac/Non-coeliac).
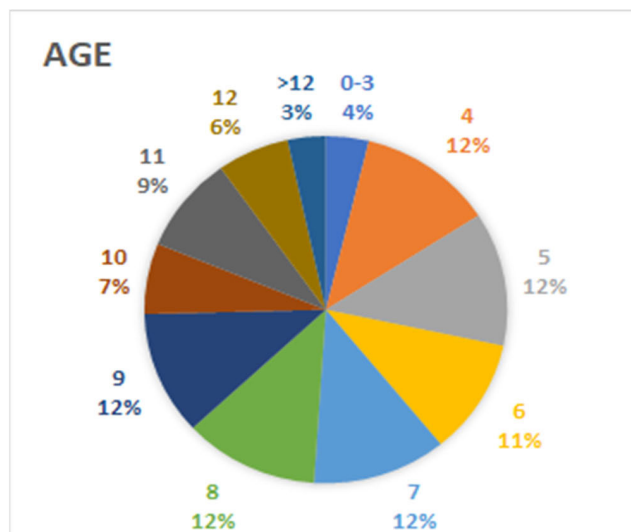

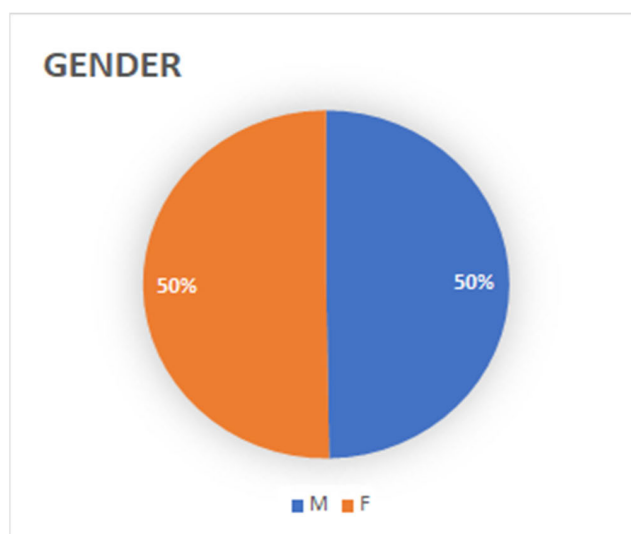**FIGURE 1.** Age distribution in the real database.

FIGURE 2. Gender distribution in the real database.

The database also keeps track of all the collected data, also those not directly usable for the project's goals but still useful for side statistics (e.g.: defective PoCTs or incomplete personal information).

The amount of stored data is 95.8 GiB, of which 31.4 MiB corresponds to internal data, and the remaining corresponds to indexed images of PoCTs and mucosal deposits. The following figures 1-3 show the distribution of participants by age, gender, and ethnicity.

The following tables and figures show the distribution by age, gender, and ethnicity of patients who tested positive at the PoCT (Table 1, Fig. 4), patients who declared five or more symptoms at the MHQ (Table 2, Fig. 5), and patients diagnosed as celiac (Table 3, Fig. 6).

Fig. 7 shows an overview of the diagnostic pathway followed by all the patients based on the result of the PoCT and of the following diagnostic steps (blood and/or endoscopy exams).
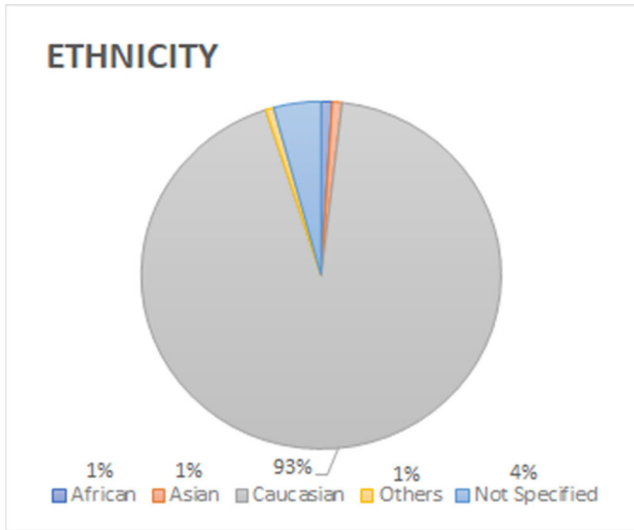
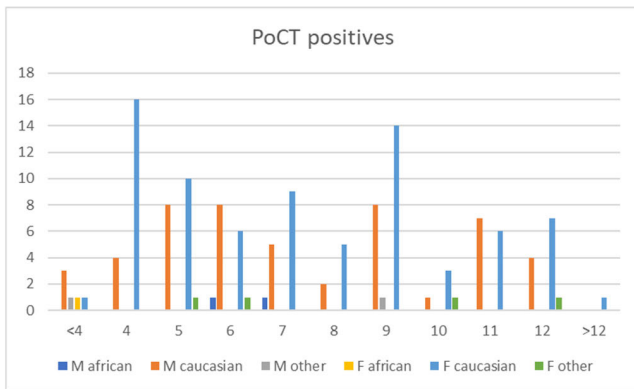**FIGURE 3.** Ethnic distribution in the real database.

FIGURE 4. Distribution of PoCT positives by age, gender, and ethnicity in the real database.
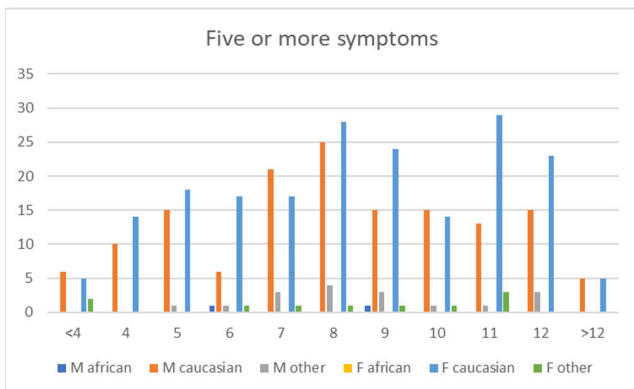

FIGURE 5. Distribution of patients with five or more symptoms declared at MHQ by age, gender, and ethnicity in the real database.

FIGURE 6. Distribution of patients diagnosed as celiac by age, gender, and ethnicity in the real database.

TABLE 1. PoCT positives by age, gender, and ethnicity.

age	M			F			total
	african	caucasian	other	african	caucasian	other	
<4	0	3	1	1	1	0	6
4	0	4	0	0	16	0	20
5	0	8	0	0	10	1	19
6	1	8	0	0	6	1	16
7	1	5	0	0	9	0	15
8	0	2	0	0	5	0	7
9	0	8	1	0	14	0	23
10	0	1	0	0	3	1	5
11	0	7	0	0	6	0	13
12	0	4	0	0	7	1	12
>12	0	0	0	0	1	0	1

TABLE 2. People with >4 symptoms on MHQ by age, gender, and ethnicity.

age	M			F			total
	african	caucasian	other	african	caucasian	other	
<4	0	6	0	0	5	2	13
4	0	10	0	0	14	0	24
5	0	15	1	0	18	0	34
6	1	6	1	0	17	1	26
7	0	21	3	0	17	1	42
8	0	25	4	0	28	1	58
9	1	15	3	0	24	1	44
10	0	15	1	0	14	1	31
11	0	13	1	0	29	3	46
12	0	15	3	0	23	0	41
>12	0	5	0	0	5	0	10

IV. METHOD

A. RATIONALE

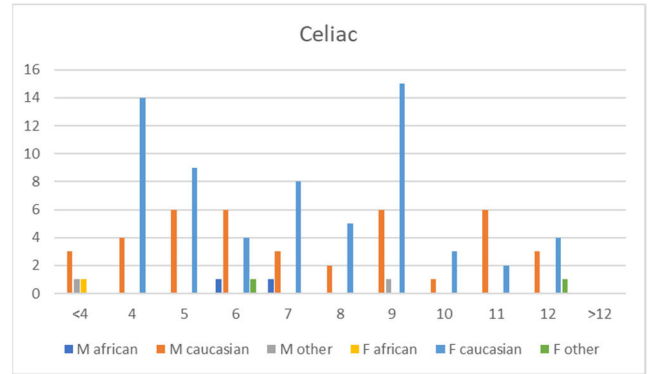A CDSS assists physicians and decision-makers in the healthcare sector by providing further insights on data, employing methodologies commonly used in Artificial Intelligence. The purpose of ITAMACDSS is to allow decision-makers in the coeliac disease diagnostic process to evaluate better the status of a subject that has entered the diagnostic pathway by prioritizing the subjects for which a positive diagnosis is more plausible, at the same time reducing diagnosis costs and optimising the use of costly and uncomfortable medical procedures. ITAMACDSS conforms to standard practices and general philosophy in the discipline: a CDSS supports and supplements physicians, it does not replace them. Furthermore,

TABLE 3. Celiac by age, gender, and ethnicity.

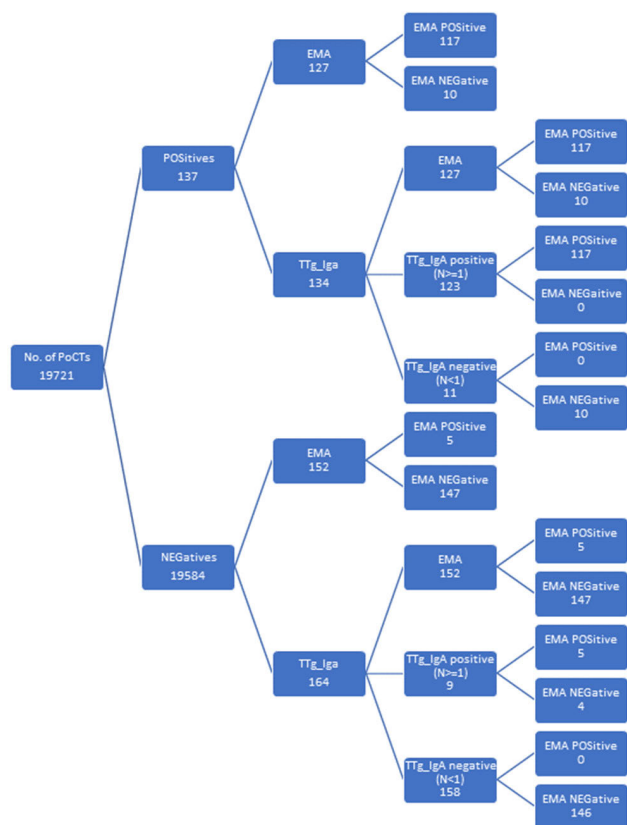| age	M			F			total
	african	caucasian	other	african	caucasian	other	
<4	0	3	1	1	0	0	5
4	0	4	0	0	14	0	18
5	0	6	0	0	9	0	15
6	1	6	0	0	4	1	12
7	1	3	0	0	8	0	12
8	0	2	0	0	5	0	7
9	0	6	1	0	15	0	22
10	0	1	0	0	3	0	4
11	0	6	0	0	2	0	8
12	0	3	0	0	4	1	8
>12	0	0	0	0	0	0	0

FIGURE 7. Overview of the diagnostic pathway of patients in the real database.

it does not replace a clinician's expertise or is intended as a fully autonomous diagnostic system. Any suggestion offered by the system has to be weighed by an expert to consider any other relevant factor that can influence a correct decision on the diagnostic pathway.

In order to obtain the results described above, the CDSS should take in account some factors: contrary to the classic case-use of classifiers, the problem with assisting in a diagnostic path is that data arrives bit by bit following the diagnostic protocol steps, and the decisions on the prosecution of the diagnostic pathway have to be taken each time. As well, unless based on a specific diagnostic mean, such as biopsy, the final classification often has a degree of uncertainty.

B. CLASSIFICATION

The ITAMACDSS team has decided to base the classifier portion of the system, which is described in the present article, on a fuzzy paradigm. Fuzziness in input and output provides a more natural expression, which lends itself better to introspection and conforms to the anthropocentric principles that medicine should adhere to. A fuzzy-based classifier can also deal with multiple explicantia for a single explicandum, i.e. the same semantic data input is expressed using different syntactic models [27], [28]. This feature is extremely important with data of the kind we have dealt with in the project, as attested by the different semantics that are present in the database. Fuzziness offers a more natural treatment for missing data, which can be considered to belong to each fuzzy set with shallow confidence; as the diagnostic pathways go on, ownership will increase in the correct class and decrease elsewhere. The same results are more difficult and less natural to obtain with classical architectures, as missing data is usually swept under the rug of neurons, and treated as an obstacle and not as a natural part of the phenomena that are analysed. As well, since in coeliac disease diagnosis, the ground truth is only obtained through an invasive procedure, the fact that the system will inherently generate approximate, imprecise results should also encourage the clinician to trust the system and to gain complete control of the diagnostic decisions, incorporating suggestion from the CDSS in a less invasive instance.

The implemented CDSS is based on a fuzzy classifier using neural networks, feed-forward backpropagation with momentum. Development took place in Python 3.6 language using the appropriate ML libraries. The system was trained and tested on real data databases acquired in the ITAMA Project and on the Virtual DB previously described. The classifier uses a five-class response (MIN, LOW, MED, HIG, MAX); input variables in ground truth belong to MIN with maximal value when the subject is confirmed as negative and to MAX if the subject is confirmed positive. A classification is counted as True Positive (TP) (respectively True Negative, TN) when a ground truth MAX is classified as MAX (MIN classified as MIN). A classification is counted as a False Positive (FP) (respectively False Negative, FN) when a ground truth MIN is classified as MAX (MAX classified as MIN). Other combinations of classifications are not counted toward the final computation of the confusion matrix.

C. DATA CLEANING AND FEATURES EXTRACTION

The first process of data cleaning was applied to the bulk data. First, each subject data consisting only of 0 or missing values was excluded. Then, the optimization and tuning of the system were carried out according to the scheme presented in Fig. 8, based on a simplified version of the optimization procedure proposed by Pota *et al.* in [8].

In detail:

- **Step A:** several user settings, including the kind and measurements for the neural architecture, are defined.
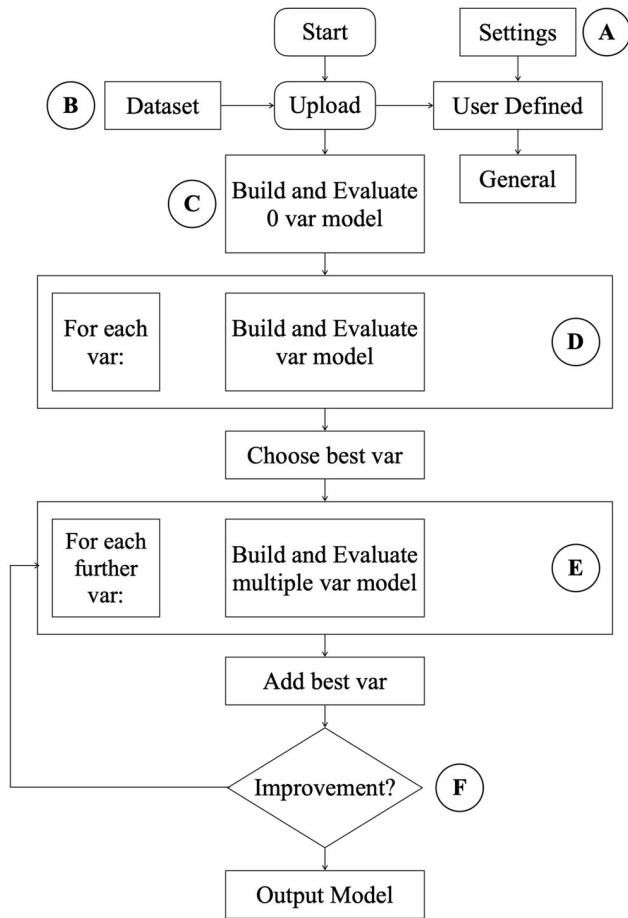
FIGURE 8. General Schematics for optimization. Preparation of settings from general knowledge and choice of architecture (A); preparation and cleaning of the data, input into the system (B); building of a 'zero variable' model (C); choice of the first variable to use as input (D); choice of the further variables to use as input (E); application of halting criteria (F).

TABLE 4. Optimised parameters.

	Parameter
No. of neurons per layer (Neural architecture)	7, 5, 3, 1
Max Epochs	200
Rate	0.5
Momentum	0.5
Epsilon	0.05
Variables from ITAMA DB (Neural input)	• Total 'Yes' answers (Questionnaire) • Chronic Diarrhoea and Malabsorption • Irregular Bowel Habits • PoCT result • AAC • IgA_Tot • tTG_IgA

- **Step F:** the process stops when adding other best variables does not significantly improve classification, as defined by a threshold selected by previous knowledge of the problem As the original number of variables did not impede the classification speed, no a-priori maximum number of variables was forced on the optimisation process.

Following the variables optimisation process, an additional parameters' choice was applied to the fuzzy neural network by applying a simple gradient with momentum descent to the number of neurons in each level and the number of levels. The complete optimisation process has selected the parameters reported in Table 4.

D. COMMUNICATING WITH ITAMADB

Once the classifier is completed, ITAMACDSS is interfaced with the ITAMADB system to offer diagnostic support to clinicians and health care personnel, and the system can accept new patients' data as input. ITAMACDSS communicates with ITAMADB System via a public API, using a microservice architecture, once again shielded from the final user. In the following, a brief description of the public API is given. All the API commands are available using a common endpoint. The following commands describe the API.

1) STATUS = ITAMACDSS_ENROLL (ID, DATA[])

When a new patient is enrolled in the system, ITAMADB issues this command, passing the patient id along a list of acquired data on the patient itself. ITAMACDSS provides an entry in its private DB for the patient.

Query Parameters:
Response Codes:

This choice is based on standard parameters and previous knowledge.

- **Step B:** the dataset is extracted from ITAMA DB, cleaned and verified. All meaningless data is stripped, and the rows that contain only null or zero values are excluded.
- **Step C:** a zero variables model, which amounts to just a choice of network and fuzzy sets to use in input and output features of the classifier, is built.
- **Step D:** the model is trained variable by variable, with repeated subsets of the data, taking the mean of the repeated tests as the final result. Each variable contained in the ITAMA DB is tested as the only input of the classifier, and the target parameter was the accuracy of the classifier on the training data. The best variable (i.e. the single variable that guarantees the best classification results) is then selected and included in the list of the variables to use in the final classification.
- **Step E:** the process is iterated using all the remaining training variables. Each time the remaining best variable is found and is added to the classifier.

name	type	description
id	integer (follows DB)	id associated with a specific patient in the system
data[]	list	data associated with the patient identified by id

status	description
ok	the entry has been correctly created
update	the id already exists, no new entry is created and the patient data id updated accordingly

2) STATUS = ITAMACDSS_UPDATE (ID, DATA[])

Each time new data is gathered about a patient (e.g. the results from a questionnaire, or a lab analysis), ITAMADB issues this command, passing the patient id along with a list of the newly gathered data on the patient itself. ITA-MACDSS updates the data on the patient accordingly and generates a risk factor index and a confidence index for the patient.

Query Parameters:

name	type	description
id	integer (follows DB)	id associated with a specific patient in the system
data[]	list	new data gathered about the patient identified by id

Response Codes:

status	description
ok	the entry has been correctly created
no	no patient associated with the id passed. No update is carried on

Response Fields:

3) STATUS = ITAMACDSS_DIAGNOSIS (ID, POSITIVE)

When a patient has been diagnosed, ITAMADB issues this command, passing the patient id along with the diagnosis. Then, ITAMACDSS deletes the entry in its private DB for the patient, allowing for a further enrolment of the same patient, and moves the data to a different private DB for further analysis and updated model training.

Query Parameters:
Response Codes:

E. RISK AND CONFIDENCE FACTORS

After the classifier has been trained, each time a new patient data set is passed to the system, a Risk Factor (RF) and a Confidence Factor (CF) are computed. RF is the classification result, considered as the best class that fits the data (MIN, LOW, MED, HIG, MAX). CF is a parameter that addresses the idea that information about a patient is partial and increases over time until a specific diagnosis is reached. The confidence factor is computed according to the following: each stage of the diagnostic pathway is associated to a weight factor w_i, according to clinical experience and previous knowledge; in the simplest version, a linear scale in the [0, 1] interval that mirrors the stage in the diagnostic pathway. In the first version of the application, $w_1 = 0.30$, $w_2 = 0.55$, $w_3 = 0.70$. Each weight is then multiplied by a function of the level of belonging to the fuzzy set selected by the algorithm. In the first version of the application, the value is $w_i \cdot 2(fs_i)$. The resulting value is then thresholded in the [0, 1] range. Such value can be displayed next to the patients' data directly or using any eidetic device (such as colour, intensity, shapes).

V. RESULTS

As for the analysis conducted on the virtual database, training of the already optimized model has been carried out using 10 batches of 10K virtual patients each, split 70/30 (training set/test set), following the procedure detailed in the previous section, that has been previously submitted to the system, using custom-made Python code.

Due to the significant unbalance between the cardinality of positive and negative virtual patients (negative ≫ positive), the positive/negative ratio has been fixed in both training and test sets in order to avoid having sets with only negative patients. Table 5 shows the confusion matrix obtained by averaging the results of 10 rounds of 10,000 virtual patients each (repeated validation of random subsampling) split between train and test sets according to the configuration described in the previous section. Patients with no useful data (i.e. 0 or NaN in all columns) have been omitted.

name	type	description
RFI	float [0,1]	Risk Factor Index. Lower values correspond to less risk of being positive. RFI should be reduced to a three values indicator (e.g. red, yellow, green) associated with the patient's risk of being positive by the DB interface.
CI	float [0,1]	Confidence Index for the clinical suggestion. Lower Confidence correspond to less data gathered on the patient. CI should be thresholded by the DB interface.

name	type	description
id	integer (follows DB)	id associated with a specific patient in the system
positive	boolean	True for a positive diagnosis, False otherwise (negative/inconclusive)

status	description
ok	the entry has been correctly deleted, the patient moved to the further training private DB
no	no patient associated with the id passed. No update is carried on

TABLE 5. Confusion matrix virtual database.

	Ground Truth True	Ground Truth False
Predicted True	29	2
Predicted False	1	1583

The sensitivity and specificity of the method can be obtained from the confusion matrix in Table 5:

- the sensitivity represents the number of correctly recognized positive, true positives (TP), on the total number of positives, obtained by adding the true positives and the false negatives (FN), i.e. as in equation (1):

$$\text{sensitivity} = \frac{TP}{TP + FN} \qquad (1)$$

- the specificity is the number of recognized negatives, true negatives (TN), on the total of negatives, obtained by adding the true negatives and the false positives (FP), i.e. as in equation (2):

$$\text{specificity} = \frac{TN}{TN + FP} \qquad (2)$$

The values obtained for these two parameters are:
- sensitivity= 96.7%
- specificity= 99.9%.

The accuracy value, i.e. the fraction of correct classifications, and the precision (also called positive predictive value), i.e. the fraction of positive classifications that resulted true, were also calculated, obtaining the following results in equations (3) and (4):

$$\text{accuracy} = \frac{TN + TP}{TN + TP + FN + FP} = 99.8\% \qquad (3)$$

$$\text{precision} = \frac{TP}{TP + FP} = 93.5\% \qquad (4)$$

TABLE 6. Confusion matrix real database.

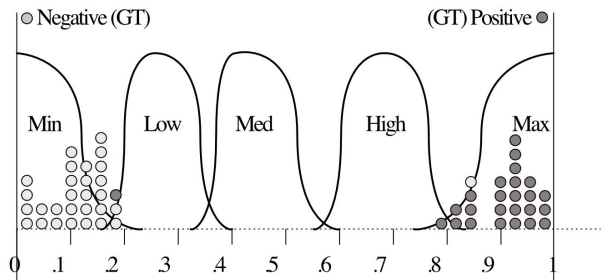|	Ground Truth True	Ground Truth False
Predicted True	24	1
Predicted False	4	1399

FIGURE 9. Typical Membership function from the classification procedure (single run). Scatterplot of mean values for center points of fuzzy numbers representing test values are included for clarity (Negative (GT): each bullet = 50 cases. Positive (GT): each bullet = 1 case).

For the study on the real database, a total of 19,415 patients, of which 109 diagnosed with coeliac disease, were analysed. Also for this database, the same parameters used for the virtual database and already presented in Table 4 were used. While the positive/negative ratio is lower (0.006 for real data, 0.01 for virtual data), such values are in the same order of magnitude, and allow a direct comparison of results from the two models. Training for the real data model has been carried out using the following parameters: split 80/20 (training set/test set), with fixed positive/negative ratio. Patients with no useful data (i.e. 0 or NaN in all columns) have been omitted. Averaged and rounded results from 1000 split batches using Monte Carlo repeated sub-sampling validation are as in Table 6 (TP mean = 23.78, stdev = 1.41; FP mean = 1.00, stdev = 0.17; FN mean = 3.67, stdev = 1.36; TN mean = 1399.76, stdev = 0.65). An example of the typical membership function obtained through the classification process is given in Fig. 9.

By means of the confusion matrix in Table 6, the following performance values were obtained:

- sensitivity = 85.7%;
- specificity = 99.6%;
- accuracy = 99.6%;
- precision = 96.0%.

Computation time for the training phase is O(n); each batch took approximately 200 +/− 14 sec. to compute on an M1 processor. Running time on a single subject is O(1).

In order to highlight the effectiveness of the proposed method, an analysis was also carried out using the best known and most used classifiers for medical imaging; SVM, kNN, neural network. The comparison of the results obtained by the various classifiers on the virtual database is shown in Table 7,

TABLE 7. Confusion matrix virtual database.

Classifier	sensitivity	specificity	accuracy	precision
SVM	100%	100%	100%	100%
kNN	100%	100%	100%	100%
Neural Network	100%	99.8%	99.8%	100%
Fuzzy	96.7%	99.9%	99.8%	93.5%

TABLE 8. Confusion matrix real database.

Classifier	sensitivity	specificity	accuracy	precision
SVM	92.3%	97.2%	96.6%	82.7%
kNN	84.6%	98.3%	96.6%	88.0%
Neural Network	100%	96.6%	97%	81.2%
Fuzzy	85.7%	99.6%	99.6%	96.0%

while the comparison of results on the real database is shown in Table 8.

As for the other classifiers analysed:

- the implemented neural network showed the best results with 7-5-1 architecture;
- the implemented SVM has a linear kernel;
- the KNN showed the best results for K = 5.

VI. DISCUSSION AND CONCLUSION

This paper presents results from ITAMACDSS, a clinical decision and support system for coeliac disease diagnosis. The system is based on fuzzy rules, particularly dedicated to data-based knowledge extraction for medical data classification. All the degrees of freedom that characterize the modelling process were analysed to identify the developer's correct choices, obtain a confidence-weighted classifier, and improve the classification performance and interpretability of the system. Some parameters have been introduced to define the required trade-off between performance and interpretability. As suggested by the procedure, these parameters can be used to identify the associated design choices. Finally, the remaining degrees of freedom can be found using the preferred optimization method. In order to give greater concreteness to the results obtained and to better evaluate the effectiveness of the proposed method, traditional classification methods (KNN, SVM, neural networks) have been implemented and evaluated. Results obtained from the fuzzy classifier employed by the project compare favourably with traditional ones, which turned out to suffer from overfitting. The fuzzy classification additionally allows to calculate risk and confidence factors, that can be usefully employed in evaluating priorities in the diagnostic pathway of the patients. Furthermore, fuzzy approaches allow for suggestions on the next steps to follow during the diagnostic pathway, instead of at the end of it, as natural for other methods.

The data acquired and hosted in the DB highlights various medical conditions associated with specific symptoms and signs. The CDSS helps assessing the physical health of a person by providing both a tool to support the diagnosis of coeliac disease and a device capable of verifying the correctness of the progress of investigation during the identification of the disorders. Furthermore, the CDSS includes new mathematical methodologies relating to the area of Artificial Intelligence. These models are used to determine both the functions of belonging to the various classes (Min to Max) and the relative values to predict the onset of pathology.

A virtual DB was created which allowed for the tuning of the proposed method. The performance of the CDSS on the real and virtual database were comparable, thus confirming the goodness of the implemented virtual DB. The results obtained by classifying the virtual patients' data are much better than what is usually obtained by classifiers applied to clinical diagnosis. This can be explained by the fact that since data is artificially created from the application of known distributions (albeit through a random generation of parameters), the classifier had the chance to understand the overlying distribution in an optimal way.

It can be observed that the ITAMACDSS classifier, when used on real patients' data and in conjunction with a confidence factor assessment mechanism, represents a good predictor of coeliac disease, and that by using the system as a diagnostic support it is possible to support the clinician in assessing the coeliac status of a patient with high accuracy and precision by looking at blood tests and PoCT results, reducing the number of costly and uncomfortable procedures such as biopsy.

Further work remains to be done on a better correlation between virtual and real data, in order to obtain a virtual model of the coeliac parameters that can be helpful in further revising clinical guidelines. Another area that can be considered an active research topic lies in the presentation of the tool's suggestions to the users of the systems, clinicians and health personnel alike.

The evaluation process confirms the system's robustness in the presence of a large amount of data (~22K subjects) and the adequacy of the results compared to international statistics. However, the system can move towards more advanced intelligent systems to support medical diagnostics. The improved computational performance, the identification of new diagnostic paths derived from data analysis and the re-edition of the DB containing the data will mitigate the necessary energy consumption. Furthermore, these aspects will extend the field of action of the CDSS to other branches of medicine in which a decision-making system based on Artificial Intelligence finds huge interest and diversified applications.

ETHICS

This study received approval from the local ethics committees (Messina and Malta).

INFORMED CONSENT STATEMENT

Informed consent was obtained from all subjects involved in the study.

REFERENCES

[1] R. Troncone and B. Jabri, "Coeliac disease and gluten sensitivity," *J. Internal Med.*, vol. 269, no. 6, pp. 582–590, Jun. 2011.

[2] J. F. Ludvigsson, T. R. Card, K. Kaukinen, J. Bai, F. Zingone, D. S. Sanders, and J. A. Murray, "Screening for celiac disease in the general population and in high-risk groups," *United Eur. Gastroenterol. J.*, vol. 3, no. 2, pp. 106–120, Apr. 2015, doi: 10.1177/205064061456 1668.

[3] A. Rostom, J. A. Murray, and M. F. Kagnoff, "American gastroenterological association (AGA) institute technical review on the diagnosis and management of celiac disease," *Gastroenterology*, vol. 131, no. 6, pp. 1981–2002, Dec. 2006.

[4] C. Catassi, I. M. Rätsch, E. Fabiani, M. Rossini, F. Bordicchia, F. Candela, G. V. Coppa, and P. L. Giorgi, "Coeliac disease in the year 2000: Exploring the iceberg," *Lancet*, vol. 343, no. 8891, pp. 200–203, 1994, doi: 10.1016/S0140-6736(94)90989-X.

[5] B. Lebwohl, J. F. Ludvigsson, and P. H. R. Green, "Celiac disease and non-celiac gluten sensitivity," *BMJ*, vol. 351, p. h4347, Oct. 2015.

[6] S. Husby, S. Koletzko, I. R. Korponay-Szabó, M. L. Mearin, A. Phillips, R. Shamir, R. Troncone, K. Giersiepen, D. Branski, C. Catassi, M. Lelgeman, M. Mäki, C. Ribes-Koninckx, A. Ventura, and K. P. Zimmer, "European society for pediatric gastroenterology, hepatology, and nutrition guidelines for the diagnosis of coeliac disease," *J. Pediatric Gastroenterol. Nutrition*, vol. 54, no. 1, pp. 136–160, Jan. 2012, doi: 10.1097/MPG.0B013E31821A23D0.

[7] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, "An overview of clinical decision support systems: Benefits, risks, and strategies for success," *npj Digit. Med.*, vol. 3, no. 1, pp. 1–10, Dec. 2020.

[8] *ICT Tools for the Diagnosis of Autoimmune Diseases in the Mediterranean Area*. Accessed: Jul. 18, 2022. [Online]. Available: https://www.itamaproject.eu/

[9] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153–1159, May 2016.

[10] Z. Qatawneh, M. Alshraideh, N. Almasri, L. Tahat, and A. Awidi, "Clinical decision support system for venous thromboembolism risk classification," *Appl. Comput. Informat.*, vol. 15, no. 1, pp. 12–18, Jan. 2019, doi: 10.1016/J.ACI.2017.09.003.

[11] E. I. Papageorgiou, "A new methodology for decisions in medical informatics using fuzzy cognitive maps based on fuzzy rule-extraction techniques," *Appl. Soft Comput.*, vol. 11, no. 1, pp. 500–513, Jan. 2011, doi: 10.1016/j.asoc.2009.12.010.

[12] M. Pota, M. Esposito, and G. De Pietro, "Designing rule-based fuzzy systems for classification in medicine," *Knowl.-Based Syst.*, vol. 124, pp. 105–132, May 2017.

[13] M. L. E. Asmar, K. I. Dharmayat, A. J. Vallejo-Vaz, R. Irwin, and N. Mastellos, "Effect of computerised, knowledge-based, clinical decision support systems on patient-reported and clinical outcomes of patients with chronic disease managed in primary care settings: A systematic review," *BMJ Open*, vol. 11, no. 12, Dec. 2021, Art. no. e054659.

[14] G. L. Masala, B. Golosio, P. Oliva, D. Cascio, F. Fauci, S. Tangaro, M. Quarta, S. C. Cheran, and E. L. Torres, "Classifiers trained on dissimilarity representation of medical pattern: A comparative study," *Nuovo Cimento Della Societa Italiana Fisica C*, vol. 28, no. 6, pp. 905–912, 2005.

[15] A. Molder, D. V. Balaban, M. Jinga, and C.-C. Molder, "Current evidence on computer-aided diagnosis of celiac disease: Systematic review," *Frontiers Pharmacol.*, vol. 11, p. 341, Apr. 2020.

[16] M. Gadermayr, G. Wimmer, H. Kogler, A. Vécsei, D. Merhof, and A. Uhl, "Automated classification of celiac disease during upper endoscopy: Status quo and quo vadis," *Comput. Biol. Med.*, vol. 102, pp. 221–226, Nov. 2018, doi: 10.1016/J.COMPBIOMED.2018.04.020.

[17] S. Segui, M. Drozdzal, G. Pascual, P. Radeva, C. Malagelada, F. Azpiroz, and J. Vitrià, "Generic feature learning for wireless capsule endoscopy analysis," *Comput. Biol. Med.*, vol. 79, pp. 163–172, Dec. 2016, doi: 10.1016/J.COMPBIOMED.2016.10.011.

[18] G. Wimmer, A. Vecsei, and A. Uhl, "CNN transfer learning for the automated diagnosis of celiac disease," in *Proc. 6th Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Dec. 2016, pp. 1–6, doi: 10.1109/IPTA.2016.7821020.

[19] Y. J. Yang and C. S. Bang, "Application of artificial intelligence in gastroenterology," *World J. Gastroenterol.*, vol. 25, no. 14, pp. 1666–1683, Apr. 2019, doi: 10.3748/WJG.V25.I14.1666.

[20] T. Zhou, G. Han, B. N. Li, Z. Lin, E. J. Ciaccio, P. H. Green, and J. Qin, "Quantitative analysis of patients with celiac disease by video capsule endoscopy: A deep learning method," *Comput. Biol. Med.*, vol. 85, pp. 1–6, Jun. 2017, doi: 10.1016/J.COMPBIOMED.2017.03.031.

[21] M. Gadermayr, G. Wimmer, A. Uhl, H. Kogler, A. Vécsei, and D. Merhof, "Fully-automated CNN-based computer aided celiac disease diagnosis," in *Proc. ICIAP*. Cham, Switzerland: Springer, 2017, pp. 467-478, doi: 10.1007/978-3-319-68548-9.

[22] G. Wimmer, A. Vécsei, M. Häfner, and A. Uhl, "Fisher encoding of convolutional neural network features for endoscopic image classification," *J. Med. Imag.*, vol. 5, no. 3, p. 1, Sep. 2018, doi: 10.1117/1.JMI.5.3.034504.

[23] X. Wang, H. Qian, E. J. Ciaccio, S. K. Lewis, G. Bhagat, P. H. Green, S. Xu, L. Huang, R. Gao, and Y. Liu, "Celiac disease diagnosis from videocapsule endoscopy images with residual learning and deep feature extraction," *Comput. Methods Programs Biomed.*, vol. 187, Apr. 2020, Art. no. 105236, doi: 10.1016/J.CMPB.2019.105236.

[24] A. Amirkhani, M. R. Mosavi, K. Mohammadi, and E. I. Papageorgiou, "A novel hybrid method based on fuzzy cognitive maps and fuzzy clustering algorithms for grading celiac disease," *Neural Comput. Appl.*, vol. 30, no. 5, pp. 1573–1588, Sep. 2018.

[25] S. Thukral and J. S. Bal, "Fuzzy logic: An easiest technique to predict celiac disease," *Sci. Technol. J.*, vol. 7, no. 2, pp. 89–94, Jul. 2019.

[26] G. Magazzù *et al.*, "Recognizing the emergent and submerged iceberg of the celiac disease: ITAMA project—Global strategy protocol," *Pediatric Rep.*, vol. 14, no. 2, pp. 293–311, Jun. 2022, doi: 10.3390/PEDIATRIC14020037.

[27] H. Ahmadi, M. Gholamzadeh, L. Shahmoradi, M. Nilashi, and P. Rashvand, "Diseases diagnosis using fuzzy logic methods: A systematic and meta-analysis review," *Comput. Methods Programs Biomed.*, vol. 161, pp. 145–172, Jul. 2018, doi: 10.1016/J.CMPB.2018.04.013.

[28] A. Minutolo, M. Esposito, and G. De Pietro, "A fuzzy framework for encoding uncertainty in clinical decision-making," *Knowl.-Based Syst.*, vol. 98, pp. 95–116, Apr. 2016.

M. E. TABACCHI received the Ph.D. degree in physics from Université Paris-Sud XI (now Université Paris-Saclay, France). He has been a Visiting Academician at UniSurrey and Imperial College (U.K.). He is currently an Assistant Professor (RTD) in mathematical logic at the Dipartimento di Matematica ed Informatica, Università degli Studi di Palermo, Italy, and the Scientific Director and the CTO at the Istituto Nazionale di Ricerche Demopolis, Italy. He is also a Fuzzy Logician, an Aquatint Printmaker, and a Transdisciplinary Researcher, whose scientific interests lie at the border between logic, data analysis, and cognitive science. He co-directs *Archives for Soft Computing* journal. He is a member of the Direction Board of *Lettera Matematica* and an Ambassador at FMSquare Foundation, Switzerland. He has been a member of the Board of Associazione Italiana di Scienze Cognitive for two terms. He is the author or coauthor of more than 120 scientific papers, and the Co-Editor for volumes *Designing Cognitive Cities* (Springer) and *Fuzziness and Medicine: Philosophical Reflections and Application Systems in Health Care* (Springer). He holds two national patents on fuzzy software methodologies. He has served as the chair and a reviewer for a number of national and international conferences, among them many editions of WILF, IFSA-EUSFLAT, and AISC.

D. TEGOLO received the M.Sc. degree in mathematics and applications from the Università degli studi di Palermo, Palermo, Italy, in 1985. He is currently an Associate Professor in computational vision and programming with the Department of Mathematics and Informatics, University of Palermo, where he was also the Co-Director. He is also an Associate Researcher at the Institute of Biophysics, National Research Council (CNR-IBF). He is with the Human Brain Project Team. He is also an Associate Member at the Innovation Value Institute, Maynooth University. Moreover, he was also an Associate Researcher at the National Institute for Astrophysics and the Italian Institute for Nuclear Physics, Palermo, and the ITALTEL Telecommunication SpA. He is the author or coauthor of several research papers in various journals and refereed conferences on computer vision and applications. His current research interests include mathematical models for complex systems, biomedical image analysis, features detection in 3-D space, and parallel and serial architectures for image analysis.

D. CASCIO received the M.Sc. degree in Physics from University of Palermo, Italy, in 2004. He obtained a PhD in Applied Physics in 2008 and in 2016 he won a position as Researcher at the University of Palermo. Since 2019, he has been an Associate Professor at the Department of Physics and Chemistry, Palermo. His research interests include machine learning and medical imaging. In particular, in recent years he has studied and implemented systems based on deep learning aimed essentially at the analysis of biomedical images. He has been associated with the National Institute of Nuclear Physics and has participated in numerous international and national research projects. He is the author or coauthor of two patents and numerous research papers in prestigious international journals. He is a Referee of numerous *Applied Physics* and *Machine Learning* journals-Based an Editor of the *Applied Sciences* journal (Switzerland).

C. VALENTI received the M.Sc. and Ph.D. degrees in mathematics from the University of Palermo, Italy, in 1997 and 2003, respectively. He is currently an Associate Professor in computer engineering with the Department of Mathematics and Computer Science, University of Palermo. His research interests include the theoretical study and implementations on parallel and heterogeneous machines with high performances for discrete geometry and medical imaging, particularly regarding discrete tomography and digital pathology as automatic evaluation of blood cancers, analysis of retinal fundus images, videocapillaroscopy of the oral mucosa, and automatic measurement of the nuchal translucency. He is an editor and a referee of international journals in the field of digital image analysis.

S. SORCE received the M.S. and Ph.D. degrees in computer science and engineering from the University of Palermo, in 2001 and 2006, respectively. He is currently an Associate Professor with the Faculty of Engineering and Architecture, Università degli Studi di Enna "Kore," Italy. He is also a Visiting Researcher at Brunel University London, U.K., and the Department of Mechanical and Aerospace Engineering, College of Engineering, Design and Physical Sciences. His research and teaching interests include novel interfaces and novel human–computer interaction paradigms, pervasive displays (technologies, techniques, deployment, evaluation), ubiquitous computing, and pervasive systems. He has published over 65 scientific papers. He is a Senior Member of the ACM.

V. GENTILE received the M.Sc. degree in computer engineering and the Ph.D. degree in technological innovation engineering from the University of Palermo, Italy, in 2013 and 2017, respectively. He is currently a Senior Data Scientist at synbrAIn. He has been in academia for several years, with more than three years of postdoctoral activities in (University of Palermo) and (University of Hertfordshire, U.K.). His research interests include the field of human–computer interaction and applied artificial intelligence, and he has visited and collaborated with many research institute in Italy and Europe. He has authored more than 30 scientific publications, and served as a conference organizer and a peer reviewer for various international journals. He has also been an Adjunct Lecturer with the University of Palermo. Prior to, and in parallel with his academic career, he worked also as a web and mobile software developer, and a tech writer and a content manager for several webzines.

V. TAORMINA received the M.Sc. degree in computer engineering from the University of Palermo, Italy, in 2012, and the Ph.D. degree in information and communication technologies, in 2021. He is currently an Assistant Professor (RTDa) in computer science with the Department of Mathematics and Computer Science, University of Palermo. His current research activity is focused on explainable artificial intelligence (XAI) in medical imaging. He is the author or coauthor of numerous research papers in international journals. His research interests include pattern recognition, image processing, and convolutional neural networks.

I. BRUSCA received the M.D. degree. He is currently the Medical Director of the Clinical Patology Unit, Buccheri La Ferla Hospital, Palermo, Italy, and the Sicily Referral Center of Immunoallergogy. He is also the Coordinator of the Protein Study Group of Italian Society of Pathology and the Laboratory Medicine (SIPMeL) He also serves as a member of the SIPMeL Study Group on the Diagnosis of Autoimmune Diseases and a member of the SIPMeL Study Group on Allergogy. His research interest includes the diagnostic role of the autoantibodies in autoimmune diseases. He also has a great expertise on molecular allergology.

A. GIULIANO is currently a Maltese Businesswoman in the technology and ICT sectors with a strong background in international research and development projects. She has been the CEO and the Managing Director at AcrossLimits, since 2001, and an Expert Evaluator, a Reviewer, the National Innovation Champion and an Ambassador for the European Innovation Council, and a jury member for the European commission and several national governments in Europe. Her expertise focuses on making the best use of European opportunities for organizational and business growth. She is passionate about female empowerment and is the Past President of the Foundation for Women Entrepreneurs in Malta.

G. MAGAZZÙ received the Medical degree from the University of Messina, Italy, in 1973.

Since 2002, he has been a Full Professor in paediatrics. From 1999 to 2009, he was the Director of the Paediatric Gastroenterology and Cystic Fibrosis Unit, University Hospital Policlinico Gaetano Martino, Messina, Italy. He has held various national and international position, such as the Secretary (1995–1998), the President of Italian Society of Paediatric Gastroenterology Hepatology and Nutrition (1998–2001), the President of the 35th Annual ESPGHAN Meeting, in 2002, was responsible of Italian Society for the study of Cystic Fibrosis Education Committee (2004–2007), the Vice-President of Italian Society for the study of Cystic Fibrosis (2008–2010), and the President of Italian Society for the study of Cystic Fibrosis (2011–2013). He is the author or coauthor of more than 130 research papers in various journals and refereed conferences, in the field of celiac disease and cystic fibrosis, IBD, functional gastrointestinal disorders, and nutrition.

Prof. Magazzù is a member of European Society for Paediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN), European Cystic Fibrosis Society (ECFS), Italian Society of Paediatric Gastroenterology Hepatology and Nutrition (SIGENP), Italian Society for the study of Cystic Fibrosis (SIFC), and Italian Society of Paediatrics (SIP).

G. RASO graduated in physics from the University of Catania, in 1981, with a thesis in nuclear physics. In December 1982, he began his research activity at the INFN section of Pisa within international collaboration at CERN on particle physics. From 1984 to 1987, he carried out his research activity at the L. A. L. of Orsay, France, with a contract from CNRS, in 1984, and IN2P3 (1985–1987). In this period, he participated in the experiment ALEPH, one of the four experiments at LEP, the accelerator site at CERN. In January 1987, he continued his research activity in the Italian component of ALEPH, as an Initial Researcher at the INFN, Bari Section. In 1998 and 1999, he was also a part of the CMS collaboration, one of the two experiments on particle physics that will come into operation, in 2008, at LHC (CERN), dealing with central tracking with silicon detectors. In November 1999, he was appointed as an Associate Professor in applied physics, as a winner of the national competition, at the Faculty of Engineering, University of Palermo. Since, he has been carrying out his research activity in applied physics at the Department of Physics and Chemistry, University of Palermo. Since January 2005, he has been a Full Professor in applied physics.

• • •