An End-to-End Semantic Platform for Nutritional Diseases Management

Ivan Donadello and Mauro Dragoni(✉)

Fondazione Bruno Kessler, Via Sommarive 18, 38123 Trento, Italy
{donadello,dragoni}@fbk.eu

Abstract. The self-management of nutritional diseases requires a system that combines food tracking with the potential risks of food categories on people's health based on their personal health records (PHRs). The challenges range from the design of an effective food image classification strategy to the development of a full-fledged knowledge-based system. This maps the results of the classification strategy into semantic information that can be exploited for reasoning. However, current works mainly address the single challenges separately without their integration into a whole pipeline. In this paper, we propose a new end-to-end semantic platform where: (i) the classification strategy aims to extract food categories from food pictures; (ii) an ontology is used for detecting the risk factors of food categories for specific diseases; (iii) the Linked Open Data (LOD) Cloud is queried for extracting information concerning related diseases and comorbidities; and, (iv) information from the users' PHRs are exploited for generating proper personal feedback. Experiments are conducted on a new publicly released dataset. Quantitative and qualitative evaluations, from two living labs, demonstrate the effectiveness and the suitability of the proposed approach.

1 Introduction

Nutritional diseases can lead to heart diseases, cancer, or type-2 diabetes and are responsible for approximately 678,000 annual deaths in the U.S. They also have a huge impact on the healthcare spending[1]: the annual cost of diabetes associated with diet and inactivity in the U.S. is 245 billions of dollars. Prevention would help people to stay healthy, to lead productive lives, to avoid/delay the onset of diseases and keep diseases far from becoming worse or debilitating. It would also reduce the costs of public health.

Dietary tracking is fundamental for the self-management of nutritional diseases. A common modality for tracking eaten food is to keep a diary of food pictures. This opens the challenge of recognizing all the taken food from users' pictures. However, for an effective management of nutritional diseases, the dietary tracking should be coupled with a reasoning system that (i) checks if the user diet is compliant with some dietary restrictions or with his/her clinical history

[1] https://cspinet.org/eating-healthy/why-good-nutrition-important.

C. Ghidini et al. (Eds.): ISWC 2019, LNCS 11779, pp. 363–381, 2019.
https://doi.org/10.1007/978-3-030-30796-7_23

and (ii) eventually provides useful feedback [16]. This integration requires the mapping of the visual food categories (e.g., cold cuts) into diseases to pay attention (e.g., cardiovascular diseases). However, current approaches are limited to the single image food detection [7,18] or to the nutritional diseases management with logical rules [17]. In addition, image food detection approaches classify meal images according to the whole recipe. Hence, they do not infer the contained food categories. The detection of these categories is fundamental for people affected by particular diseases, such as, diabetes, hypertension, or obesity.

In this paper, we propose an end-to-end semantic platform that supports the management of nutritional diseases. The system covers the whole pipeline from data acquisition (meal pictures taken with a smartphone) to tailored messages to users in order to correct wrong dietary habits within a behavior change context. Here, we focus on the following aspects originally presented in this contribution:

– A multi-label classification of food pictures according to the food categories contained in a specific food recipe of the Mediterranean diet. The classification is performed with a Convolutional Neural Network (CNN).
– An extension of a state-of-the-art ontology (i.e., the HeLiS ontology [9]) about the dietary and physical activity domains with knowledge about the risk level of food categories with respect to a set of diseases.
– A strategy for navigating over the Linked Open Data (LOD) Cloud to infer matches between the user clinical history and the potential risks of diseases and comorbidities induced by an excessive intake of some food categories.
– A new dataset of food pictures, the classification models and the source code of the classification tool. These are released in order to support the reproducibility of the results and to foster further research in this direction.

The significance of our work relies on the integration of deep learning in a Semantic Web (SW) platform for healthcare. Indeed, Computer Vision (CV) methods have no mapping in the semantic space of an ontology, thus they are rarely used as input providers for reasoning systems. SW systems (for healthcare) instead deal with a clean input. This can be time consuming and could affect the scalability. The proposed SW platform allows us to investigate the right balance between effort and effectiveness. We evaluated the proposed platform from three perspectives: (i) the effectiveness of the food categories classification, (ii) the usability of the mobile application adopted by users, and (iii) the effectiveness of the generated messages. In all cases, the obtained results confirm the soundness of the proposed end-to-end semantic platform.

2 Related Work

The end-to-end platform proposed in this paper conjugates two research areas: the classification of food images and the effective navigation of the LOD Cloud for gathering and exploiting knowledge for the realization of intelligent systems.

The recognition of foods from images is the first step for dietary tracking. This task has been studied by the Computer Vision community with techniques of image classification/segmentation and volume estimation. The first works rely on the extraction of visual features from the images and the consequent use of classifiers. The main features used are local/global features and local binary patterns [3,14]. The classifiers are k-NN classifiers, Support Vector or Kernel Machines. Successively, CNNs became the standard technique for food classification [18], thus avoiding the use of engineered features. The Food524DB dataset is used in [7] for food recognition with CNNs and gathers the Food50, Food-101, UEC FOOD-256 and VIREO Food-172 datasets.

Other works estimate the quantity of food in the dish and thus the intake calories. The semantic segmentation of the food dish is performed, then techniques of volume estimation compute the food quantity. However, this requires a database of foods with relatives densities [6,8]. Other works exploit a reference object (e.g., a thumb [23] or a wallet [22]) for the volume computation. Im2Calories [20] uses a CNN to predict a depth map of the image that is used to build the 3D model of the meal. Quantity estimation can be addressed with multi-task learning by training CNNs that learns both the food classification and the relative calories/volume. However, this technique requires a dataset with the annotated calories [11] or the depth information in the images [15].

Few works among those mentioned above predict food categories and match them with some nutritional facts in a database [8,11]. They predict only one food category (e.g., pasta) for each detected food and this can be inaccurate. Indeed, a pasta dish should be avoided by a person suffering of diabetes. However, a pasta dish might have carbonara sauce, containing cold cuts that are not suitable for people suffer from cardiovascular diseases. Therefore, it is important to perform a multi-label classification of the several food categories in the dish.

The promotion of healthy lifestyle through dietary counseling is a recent topic with few available working systems. Nevertheless, some SW-based approaches have been previously proposed. The Medical Decision Support System in [2] supports (i) the collection of patients' relevant information via a mobile application prompting questions related to the patient's medical background, and (ii) the creation of customized advices based on the information collected and on the changes of patient's lifestyle.

In [19] the authors present an approach for designing a semantic reasoning engine to support coaching profiles. This system uses a web-based interface for collecting user data and an ontology for analyzing and processing them. This way, created profiles can be used to optimize the coaching activities of professionals. The work presented in [5] aims to integrate multiple knowledge sources for the development of a dietary consultation system for chronic kidney diseases. The system demonstrates how a knowledge-based approach can achieve sound problem solving modeling and effective knowledge inference. The evaluation involved 84 case patients about recommending appropriate food serving amounts from different food groups for balanced key nutrient ingestion. Finally, in [10] the authors discussed the use of SW technologies to build a system for supporting and motivating people in following healthy lifestyles. SW technologies are used for modeling domain knowledge, and for performing reasoning activities by combining real-time user-generated data and domain expert knowledge.

To the best of our knowledge, our platform innovates the state-of-the-art as it integrates multiple modalities (images, reasoning, LOD Cloud and Personal Health Records) of managing information. Indeed, current CV approaches classify food images according to their recipe label with very poor reasoning about the food intake and related diseases. On the other hand, SW systems do not deal with a noisy input. Our full-fledged solution supports the transformation of food images content into semantic information. This is used for gathering from the LOD Cloud the knowledge of the nutritional diseases associated with the detected food categories. This knowledge is exploited in a fine-grained reasoning process for generating proper personalized feedback.

3 Architecture

Here, we discuss the pipeline modules developed (or reused from existing platforms) for supporting the detection and processing of food categories from users' pictures, see Fig. 1. Such food categories are exploited for (i) detecting the risk level with respect to specific diseases; (ii) navigating the LOD cloud for extracting related diseases and possible comorbidities; and, (iii) linking all collected information with the user's Personal Health Record (PHRs) in order to generate proper feedback.

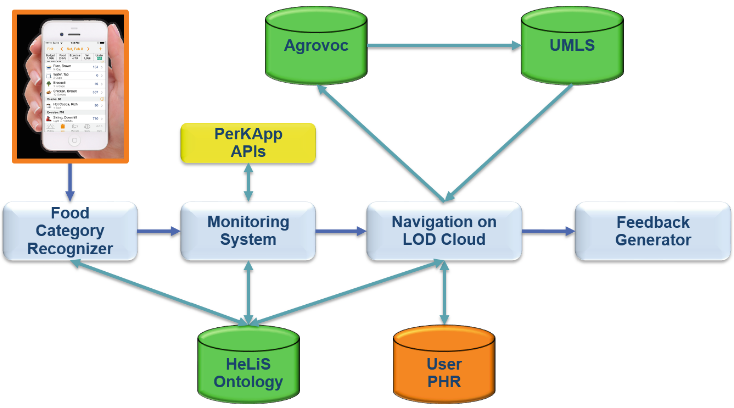

Fig. 1. Architecture of the end-to-end system. Green boxes are external resources, i.e., ontologies. Orange boxes are input data (pictures or PHRs). Light blue boxes are the modules of the system. (Color figure online)

The input module of the pipeline is a mobile application allowing users of taking pictures of consumed food. This kind of input represents the best trade-off between efficiency and effectiveness. On the one hand, the efficiency is supported by the low effort required for providing data. Taking a picture requires less time than providing the complete list of the consumed food. Hence, the reduced effort implies a lower abandonment rate. On the other hand, the effectiveness is given by the fact that it is unfeasible to have a knowledge base with the description of all possible variants for a recipe. Thus, a recipe-based classification system could fail in recognizing all the eaten food categories whereas a food-category-based classification system can be more accurate, see Sect. 8. A correct detection of food categories impacts the consequent reasoning over medical knowledge bases, the inference of risk levels for specific diseases, the alignment of such diseases with users' PHRs and the generation of personalized feedback.

Before detailing the modules of our end-to-end semantic platform, we describe the adopted state-of-the-art components: the HeLiS ontology [9] and the PerKApp platform [16]. The HeLiS ontology is the most updated ontology covering the dietary and physical activity domains. It also defines a model for describing the Mediterranean diet rules that can be associated with user profiles. We extended HeLiS with the risk level of each food category with respect to some diseases, see Sect. 4. The PerKApp platform is a behavior change persuasive platform designed for generating persuasive messages to support people in healthy lifestyles adoption. PerKApp exposes APIs that give access to a subset of its reasoning facilities. This allows the development of applications that monitor users dietary behaviors. Our system exploits the PerKApp APIs to reason about the consumed food categories and trigger the navigation of the LOD Cloud.

The task of recognizing food categories is performed by the **Food Category Recognizer** module (Sect. 5). Such a module uses a CNN trained with recipe images annotated with the contained food categories. During the classification the CNN receives as input the picture taken by the user and it predicts the detected food categories.

Once consumed food categories have been recognized, they are passed to the **Monitoring System** module (i.e., an interface for the PerKApp platform). As mentioned above, this module verifies, through reasoning operations, if the user violated one of the assigned rules defined within the HeLiS ontology.

In case an undesired behavior is detected, information about the involved food categories are sent to the **Navigation on the LOD Cloud** module. This module acquires from nutritional and medical knowledge bases (available in the LOD Cloud) disease information linked with the received food categories. This process is performed through the following activities:

1. The module looks up into the HeLiS extension for the risk level of the detected food categories with respect to the modeled diseases. Such information have been provided by domain experts only for a subset of possible nutritional diseases. Currently the HeLiS extension contains knowledge for five nutritional diseases, see Sect. 4. The rationale behind the limited number is: (i) we want to limit the effort of the domain experts in providing all the knowledge, and (ii) missing information (other nutritional diseases of the literature) are acquired through the second step.

2. The HeLiS ontology is connected to the LOD Cloud due to the alignments with AGROVOC[2], see the *equivalentClass* annotation property in HeLiS. In this step, the module exploits the diseases modeled in HeLiS for accessing to the related nutritional diseases defined within AGROVOC (i.e., children and sibling diseases).
3. PHRs have a very specific medical terminology and they contain detailed information that do not match with the AGROVOC diseases. Hence, the module navigates the LOD Cloud from AGROVOC to the UMLS Knowledge Base[3] to collect information about comorbities associated with the diseases extracted from AGROVOC. Indeed, comorbidities are not directly associated with food categories, thus only the navigation of the LOD Cloud enables the finding of the ones that a user already had in his/her PHR. The UMLS is a medical knowledge base containing both a taxonomy of diseases and properties concerning associated diseases, comorbidities, recidivity degree, etc. Such low-level information increases the likelihood to find an alignment between the content of a PHR and the knowledge collected from the LOD Cloud. For reaching UMLS from AGROVOC, the module exploits the path AGROVOC → Bio2RDF → LinkedCT → Pubmed → UMLS as described in [26].
4. The last step consists in matching all the information extracted from both AGROVOC and UMLS with the information contained in the PHR of the user. The result of this match is provided to the last module of the pipeline.

To perform the LOD Cloud navigation we use the LOD-a-lot [12] service, i.e., a dump of the LOD Cloud that can be queried by using a single SPARQL endpoint for all the involved resources. This prevents us from the possible unreliability of some SW resources, e.g., a fault in the path from AGROVOC to UMLS.

Finally, once the system has computed (i) the intake food categories, (ii) the risk levels of associated diseases, (iii) the related diseases and possible comorbidities, and (iv) the alignments with the user's PHR, it generates a proper feedback that is returned to the user mobile application. The **Feedback Generator** module relies on a template-based engine where the structured information of above is realized into natural language sentences. More details on how templates are populated are in [16].

4 Background Knowledge

The role of background knowledge in our platform is two-fold. First, background knowledge allows our semantic platform to go beyond the sole classification of food images. Indeed, background knowledge enables the possibility of exploiting logic relationships and inference capabilities for reusing the food classification results to support users for more complex goals. For example, the prediction of

[2] http://aims.fao.org/vest-registry/vocabularies/agrovoc.
[3] https://www.nlm.nih.gov/research/umls/.

some food categories might represent a warning for people affected by specific diseases, e.g., pasta for people affected by diabetes. Moreover, background knowledge can contains conceptual models about specific dietary patterns that can be used to improve users' lifestyle, avoiding the rise or sharpening of chronic diseases, and to support a behavioral changing. Second, the exploitation of knowledge resources enables the access to the LOD Cloud. This focuses the modelling only on extending HeLiS since all other semantic information exploited by the system are available through the LOD Cloud.

The background knowledge exploited here is HeLiS [9]: a state-of-the-art ontology for supporting healthy lifestyles. It defines the dietary and physical activity domains together with entities that model concepts concerning users' profiles and the monitoring of their activities. Details about the conceptual model and the methodology for building it are in [9]. The HeLiS ontology has been extended by adding, to the dietary domain, information concerning the risk level of food categories with respect to specific diseases[4]. We discuss the main concepts involved into the food category classification together with the ones modeled within the HeLiS ontology extension, see Fig. 2.

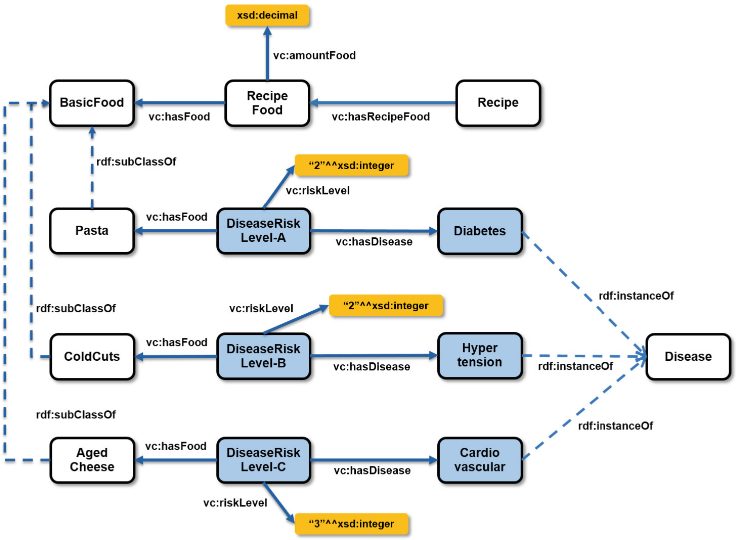

Fig. 2. Excerpt of the HeLiS ontology including the main concepts (white boxes) and instances (blue boxes) exploited by our semantic platform. Solid lines are object properties, dashed lines are RDF core properties (Color figure online)

Instances of the *BasicFood* concept describe foods for which micro-information of nutrients (carbohydrates, lipids, proteins, minerals, and vitamins)

[4] The HeLiS extension is available on the HeLiS website http://w3id.org/helis.

are available. Moreover, these instances belong also to subclasses of the *Basic-Food* concept, such as *Pasta*, *Aged Cheese*, *Eggs*, *Cold Cuts* and *Vegetal Oils*. On the other hand, instances of the *Recipe* concept, describe the composition of complex dishes (such as *Pasta with Carbonara Sauce*) by expressing them as a list of instances of the *RecipeFood* concepts. This concept reifies the relationships between each *Recipe* individual, the list of *BasicFood* it contains and the amount of each *BasicFood*. Besides this dual classification, instances of both *BasicFood* and *Recipe* concepts are categorized under a more fine-grained structure. With respect to the number of individuals, currently, HeLiS contains 986 individuals of type *BasicFood* and 4408 individuals of type *Recipe*.

The *Disease* concept defines the diseases supported by the system such that information about the risk level relationship with specific *BasicFood* is available. Currently, we instantiate the *Disease* concept for diabetes, kidney diseases, cardiovascular diseases, hypertension and obesity. Diseases are defined as single individuals instead of concepts for avoiding the creation of a new individual for each specific disease for each user. Instances of the *DiseaseRiskLevel* concept reifies the relationships between each *Disease* and *BasicFood* individuals and with the risk level of a *BasicFood* for that *Disease*. The risk level is represented by a value ranging from 0 (no risk) to 3 (high risk). For readability we report in Fig. 2 only some instances of the *DiseaseRiskLevel* concept, e.g., *DiseaseRiskLevel-A*, *DiseaseRiskLevel-B*, and *DiseasesRiskLevel-C*.

The HeLiS ontology is used by the **Food Category Recognizer module** for getting the list of available food categories, by the **Monitoring System** for supporting the reasoning process, and by the **Navigation on LOD Cloud** as starting point for getting the list of diseases associated with the detected food categories.

5 Multi-label Food Category Classification

Our goal is to assign every food image with a set of food category labels. These categories compose the food recipe in the image and are provided by HeLiS. We address this problem as a multi-label image classification task where $\mathcal{X} \in \mathbb{R}^d$ is the input domain of our images and the subclasses of *BasicFood* are the possible food category labels. Given an image $\boldsymbol{x} \in \mathcal{X}$, we need to predict a vector $\boldsymbol{y} = \{y_1, y_2, \dots, y_K\} \subseteq$ *BasicFood* where y_i is the i-th food category label associated to $\boldsymbol{x}$. State-of-the-art methods in food image recognition mostly classify images according to only one single label taken from *Recipe* without multi-label classification. Here, we exploit two strategies for food-categories classification: (i) a direct multi-label classification of the food categories with a CNN and (ii) a single-label image classification of the food recipes (e.g., *Pasta with Carbonara Sauce*) with a CNN and then the logical inference of all its food categories (i.e., *Pasta*, *Eggs*, etc) through the *RecipeFood* concept.

5.1 Methods

Current methods in image classification use supervised deep learning techniques based on CNNs [13]. These are able to learn the salient features of an image in order to classify it according to some training examples. CNNs exploit several combinations of the hidden layers (convolutions, poolings, activations) in order to improve their performance. In both methods (i) and (ii) we separately train (on the dataset in Sect. 8.1) one of the most performing CNN, the Inception-V3 [25]. This network presents convolutional filters of multiple sizes operating at the same level. This makes the network "wider" and able to better detect the salient parts of an image. Finally, the network has a standard fully-connected layer for predicting the classes. Moreover, this networks does not present some redundant connections between neurons that affect the efficiency of the other CNNs. Further details and performance results can be found in [25].

Direct Multi-label Classification. We train the Inception-V3 for directly learning the vector $\boldsymbol{y}$ of the food categories in *BasicFood*. We use a sigmoid as activation function of the last fully-connected layer and binary cross entropy as loss function. This is a standard setting for multi-label classification.

Single-Label Classification and Inference. Another method to classify food categories consists in: firstly, to classify an input image with a CNN according to the food label (in *Recipe*) it contains (e.g., *Pasta with Carbonara Sauce*). This is the standard multiclass classification where one image is classified with only one food label among many classes. Secondly, to infer the food category labels from the food label by using the *RecipeFood* reification. The detection of *Pasta with Carbonara Sauce* implies the presence of the food categories: *Pasta*, *Eggs*, *Aged cheese*, *Vegetal Oils* and *Cold cuts*. Let CNN be an Inception-V3 trained to multiclassify food labels in *Recipe*. We use a softmax as activation function of the last fully-connected layer and categorical cross entropy as loss function. Thus $CNN(\boldsymbol{x}) = \langle s_1, s_2, \ldots, s_n \rangle$ with $s_i \in \mathbb{R}$ is the classification score of the network for the label $l_i \in \mathit{Recipe}$. Let $l^* \in \mathit{Recipe}$ be the label with highest score in $CNN(\boldsymbol{x})$, then the food category labels vector $\boldsymbol{y}$ is defined as:

$$\boldsymbol{y} = \{y_i \in \mathit{BasicFood} \mid \exists w \in \mathit{RecipeFood} : \mathit{hasFood}(w, y_i) \wedge \mathit{hasRecipeFood}(l^*, w)\}$$

6 From Image Classification to LOD Cloud Navigation

Here we show how the system works through a concrete example. Let us consider a user suffered from anomalies of blood pressure and with a nasal polyps surgery five years before the use of the platform. Her electronic PHR contains all these information related to her clinical history. Then, let us assume that she is going to eat a pasta with carbonara sauce and she sends to the system the meal picture taken with her smartphone. The **Food Category Recognizer** detects the presence of these food categories: *Pasta*, *Eggs*, *AgedCheese*, *VegetalOils* and *ColdCuts*. These are sent to the **Monitoring System**.

As first action, this module adds into the dietary diary of the user, represented as a set of individuals within HeLiS, the consumed food categories. Then, through logical reasoning, the **Monitoring System** checks if the intake food categories follow the rules associated with the user's profile encoded in HeLiS. According with the user's dietary diary and the rules of his profile, the system detects an undesired behavior: in the last 7 days the user has consumed the *ColdCuts* food category four times while the associated rules limits the consumption of *ColdCuts* three times per week.

These undesired food categories are passed to the **Navigation on LOD Cloud** module and trigger the retrieval of possible diseases to pay attention if the user exceeds with *ColdCuts* food consumption. The module queries the HeLiS extension for all instances of type *DiseaseRiskLevel* having an *hasFood* object property instantiated with the concept *ColdCuts*. By looking in Fig. 2, the module finds the *DiseaseRiskLevel-2* individual and from it, retrieves the individual *Hypertension* of type *Disease*. From the *DiseaseRiskLevel-2* individual, the module looks for the *riskLevel* data property for retrieving the risk level associated with the pair $<Hypertension, ColdCuts>$. If the value is greater than 1, the module starts to navigate through the LOD Cloud for finding all related information. Indeed, HeLiS mainly focuses on healthy lifestyles and it is not a medical ontology. Hence, the acquisition of further medical information concerning the diseases associated with the consumed food categories has to be performed from the LOD Cloud. The navigation starts from the alignment between HeLiS and AGROVOC. Here, the system retrieves the children and sibling diseases of *Hypertension* provided by the diseases taxonomy of AGROVOC. Examples of children diseases of *Hypertension* are *Embolism*, *HeartAttack* and *Phlebitis*. However, the specific medical terminology in PHRs do not always match with the diseases in AGROVOC. Hence, the system continues the navigation through the LOD Cloud to refine the list of AGROVOC diseases by extracting information from the UMLS Knowledge Base. In our example, from the *Hypertension* concept, extracted from HeLiS, the platform reaches the *BloodPressureAnomalies* associated disease and *NasalPolyps* possible comorbidity within the UMLS Knowledge Base. The latter contains also the *recidivity* attribute. Every retrieved disease and their attributes are searched in the electronic PHR to provide a more accurate user feedback. In our case, the module finds alignments with *BloodPressureAnomalies* and with *NasalPolyps* that are two diseases that the patient suffered from.

The gathered information (*ColdCuts*, *Hypertension*, *NasalPolyps*, *BloodPressureAnomalies*, *riskLevel(BloodPressureAnomalies, High)*, and *hasAttribute(NasalPolyps, Recidivity)*) are processed by the **Feedback Generation Module**. Its language generation engine fills message templates to realize tailored motivational messages. Concerning our scenario, a sample message is the following: *"This week you have eaten too much cold cuts. Do yo know that an excessive intake of cold cuts could cause the recidivity of nasal polyps and significantly increases the probability of having anomalies in your blood pressure? Next time you can try a meal with some fresh fish."*.

7 Use Cases: The *Key to Health* and *Salute Plus* Living Labs

As specific case studies, we validated our platform within two living labs: *Key To Health* and *Salute Plus*. The *Key To Health* living lab promotes healthy lifestyles in workplaces with the aim of preventing the onset of chronic diseases through organizational interventions directed to workers. Actions can concern the promotion of a correct diet, physical activity, social and individual well-being, or the discouragement of bad habits, such as smoking and alcohol consumption. Within the *Key To Health* living lab, our platform has been used by 120 workers of our institution (both researchers and employers) as a tool to persuade them to follow healthy recommendations. The *Salute Plus* living lab is part of *Trentino Salute 4.0*[5], a digital health initiative promoted by the local healthcare department. Such an initiative aims at proving innovative technological solutions to citizens to promote healthy lifestyles. Table 1 shows the main demographic information concerning the users involved in the two living labs. Whereas Table 2 provides statistics about the usage of the platform. Even if the *Salute Plus* living lab runs from a longer period (it is still active), we consider for the evaluation the data acquired during the first 49 days in order to provide a fair comparison with the *Key To Health* living lab.

Table 1. Demographic information of the users involved in the evaluation campaign.

Dimension	Property	Value	
		Key to health	Salute plus
Gender	Male	57%	48%
	Female	43%	52%
Age	25–35	12%	27%
	36–45	58%	45%
	46–55	30%	28%
Education	High school	0%	56%
	Master degree	42%	43 %
	Ph.D. degree	58%	1 %
Occupation	Ph.D. student	8%	n.a
	Administration	28%	n.a
	Researcher	64%	n.a

The *Key To Health* and *Salute Plus* use cases allowed us to deploy our platform into two different scenarios. The former is a controlled environment in which we performed a complete evaluation both from the quantitative and qualitative perspectives. Whereas, the latter is a real-world environment that allowed us to

[5] http://www.trentinosalutedigitale.it/#primo.

Table 2. Usage statistics during the living labs. We report the number of users involved, the number of days since each living lab started, the number of meals introduced by the users (each meal can be composed by several pictures), and the number of RDF triples currently stored.

Living lab	# Users	Days running	Meals provided	Triples
Key To Health	120	49	18,816	470,400
Salute Plus	2,870	112	902,944	16,704,464

observe (i) the engineering suitability of our platform, and (ii) the effectiveness of our solution within a more open context for comparing the results obtained in the controlled environment. For each living lab, users were equipped with a mobile application[6] based on the services provided by our platform. We analyzed the usage of the mobile application for seven weeks by monitoring the users' input and the associated undesired behaviors (hereafter called "violations"). Results and discussions are in Sect. 8.

8 Experiments

Within the living labs, we validated our platform from both quantitative and qualitative perspectives. The former focuses on the performance of the food category recognizer (Sect. 8.1). The latter regards the whole platform: (i) the user experience with the mobile application and (ii) the effectiveness of the generated messages (Sect. 8.2). Concerning the second point, we compared the impact of the messages generated by using only the reasoning results (a.k.a. the *control group*) with the messages generated by exploiting the knowledge extracted from the LOD Cloud combined with the information in the PHRs. Finally, lessons learnt from this experience are provided (Sect. 8.3).

8.1 Quantitative Evaluation

Good performance of the food category recognizer are important as the misclassification of a meal could generate wrong messages or even no message. In the example of Sect. 6, we noticed that the single-label classification and inference method could wrongly classifies some Carbonara images as *Tomato and Ricotta Cheese Pasta*, thus containing *FreshCheese* instead of *Eggs* and *TomatoSauces* instead of *ColdCuts*. In this case no message will be generated and the user could

[6] The mobile applications are available on the stores and they are compliant, as the whole platform, with the GDPR rules. However, since PHRs from the Trentino Healthcare Department are used, the mobile applications cannot be used by people living outside our province. For informative purposes, here the Google Play Store links: https://play.google.com/store/apps/details?id=eu.fbk.trec.saluteplus https://play.google.com/store/apps/details?id=eu.fbk.trec.lifestyle.

consume another meal with *ColdCuts* next time. Here, we compare the multi-label method against the (more standard) single-label classification of the food recipe and the inference of the food categories. Our claim is that a classification error in a single food recipe affects the majority of the inferred food categories leading to inaccurate results.

The Food and Food Categories (FFoCat) Dataset.[7] We leverage the food and food category concepts in HeLiS extension for the multi-label classification. However, current food image datasets are not built with these concepts as labels, so it is necessary to build a new dataset (named FFoCat) with these concepts. We start by sampling some of the most common recipes in *Recipe* and use them as food labels. The food categories are then automatically retrieved from *BasicFood* with a SPARQL query. Examples of food labels are *Pasta with Carbonara Sauce* and *Baked Sea Bream*. Their associated food categories are *Pasta, AgedCheese, VegetalOils, Eggs, ColdCuts* and *FreshFish, VegetalOils*, respectively. We collect 156 labels for foods (*Recipe* concept) and 51 for food categories (*BasicFood* concept). We scrape the Web, using Google Images as search engine, to automatically download all the images related to the food labels. Then, we manually clean the dataset by checking if the images are compliant with the related labels. This results in 58,962 images with 47,108 images for the training set and 11,854 images for the test set (80-20 ratio of splitting). Then, by leveraging HeLiS properties, we enrich the image annotations with the corresponding food category labels to perform multi-label classification. The dataset is affected by some natural imbalance, indeed the food categories present a long-tail distribution: only few food categories labels have the majority of the examples. On the contrary, many food categories labels have few examples. This makes the food classification challenging.

Experimental Settings and Metrics. For both multi and single-label classification we separately train the Inception-V3 network from scratch on the FFoCat training set to find the best set of weights. The fine tuning using pre-trained ImageNet weights did not perform sufficiently. We run 100 epochs of training with a batch size of 16 and a learning rate of 10^{-6}. At each epoch images are resized to 299×299 pixels and are augmented by using rotations, width and height shifts, shearing, zooming and horizontal flipping. This results in a training set 100 times bigger than the initial one. We use early stopping to prevent overfitting. The training has been performed with the Keras framework (TensorFlow as backend) on a PC equipped with a NVIDIA GeForce GTX 1080.

As performance metric we use the mean average precision (MAP) that summarizes the classifier precision-recall curve: $MAP = \sum_{i=1}^{n}(R_n - R_{n-1})P_n$, i.e., the weighted mean of precision P_n achieved at each threshold level n. The weight is the increase of the recall in the previous threshold: $R_n - R_{n-1}$. The macro AP is the average of the AP over the classes, the micro instead considers each entry of the predictions as a label. We prefer MAP instead of accuracy as the latter for sparse vectors can give misleading results: high results for output vectors with all zeros.

[7] The dataset, its comparison and the code are available at https://bit.ly/2Y7zSWZ.

Results of the Multi-label Classification. Given an (set of) input image(s) $\boldsymbol{x}$, the computing of the precision-recall curve requires the predicted vector(s) $\boldsymbol{y}$ of food category labels and a score associated to each label in $\boldsymbol{y}$. In the multi-label method this score is directly returned by the Inception-V3 network. In the single-label and inference method this score needs to be computed. We test two strategies: (i) we perform *exact inference* of the food categories from HeLiS and assign the value 1 to the scores of each $y_i \in \boldsymbol{y}$; (ii) the food categories labels inherit the *uncertainty* returned the CNN: the score of each y_i is the value s_i returned by $CNN(\boldsymbol{x})$. Results are in Table 3. The direct multi-label has very good performance (both in micro and macro AP) in comparison with the single-label models. The micro-AP is always better than the macro-AP as it is sensible to the mentioned imbalance of the data. This confirms our claim that errors in the single recipe classification propagate to the majority of the food categories the recipe contains. That is, the inferred food categories will be wrong because the recipe classification is wrong. On the other hand, errors in the direct multi-label classification will affect only few food categories. With these good results, we use the direct multi-label classification method in our platform. We also performed a qualitative analysis. The single-label method misclassifies an image with *Backed Potatoes* as *Backed Pumpkin* thus missing the category of *FreshStarchyVegetables*. Another image contains a *Vegetable Pie* but the single-label method infers the wrong category of *PizzaBread*. In another image, this method mistakes *Pasta with Garlic, Oil and Chili Peppers* with *Pasta with Carbonara Sauce*, thus inferring wrong *Eggs* and *ColdCuts*. Here the multi-label method classifies all the categories correctly. Therefore, the multi-label method allows a more fine grained classification of the food categories w.r.t. the single-label method. The latter has better results if the score returned by the CNN is propagated to the food categories labels w.r.t. the exact inference. Good performance on food categories classification are important as they reduce the noise for the following modules of the platform.

Table 3. The multi-label classification of food categories outperforms in average precision (AP) the methods based on single-label classification and logical inference.

Method	Micro-AP (%)	Macro-AP (%)
Multi-label	**76.24**	**50.12**
Single-class exact	50.53	31.79
Single-class uncert	60.21	42.51

8.2 Qualitative Evaluation of the System

We present here the validation performed by involving users from the living labs concerning (i) the overall usability of the mobile application and (ii) the effectiveness of the generated messages, i.e. how the number of detected violations changed through time.

Usability Evaluation. The usability of the mobile application has been evaluated through the System Usability Scale (SUS), analyzing the intuitiveness and simplicity of the system. Only the users involved in the *Key To Health* living lab participated to this validation. The evaluation protocol consists in multiple use sessions and follows these steps:

1. Training meetings with the 120 involved users for an introductory explanation of the functionalities available in the mobile application.
2. Four days of usage of the mobile application by the users.
3. Meetings with the users for collecting questions about functionalities. Release of a new version of the mobile application integrating bug fixes reported by the users.
4. Four days of usage of the mobile application by the users.
5. Final meetings with the users and distribution of evaluation questionnaires.

According to the usability test requirements provided by [21], the number of users involved in the test granted the discovery of 100% of the usability problems. The average score obtained from the SUS was 83.1 out of 100, that, according to the adjective rating scale proposed by [1], corresponds to "excellent". Further interviews were conducted to evaluate the impact of the mobile application in their daily life at the end of the seven weeks of pilot study. Users appreciated the system and considered the mobile application a useful tool, especially for increasing the awareness about their eating habits.

Effectiveness of Generated Messages. The last validation we performed concerned the analysis of explanations effectiveness on the users involved within the *Key to Health* and *Salute Plus* living labs. Our goal was to measure the effectiveness of the explanations generated by our platform by observing the evolution of the number of detected violations. The *Key To Health* living lab allowed to plan a more rigorous evaluation thanks to the exploitation of a close environment. The 120 users involved were split in two groups. The first one (92 users) received messages generated by using the whole system: from the results of the reasoning process to the navigation of the LOD Cloud with exploitation of PHRs. Whereas the second group (28 users working as control group) received feedback, as canned text messages, exploiting only the reasoner's results. We expect to find a higher decrease in the number of violations through the time by the users receiving persuasive messages. Concerning, the *Salute Plus* living lab, we did not have the control group since the agreement with our Local Healthcare Department foreseen that all citizens were able to access the complete set of services of the platform. However, we could check if results on both living labs converged or not. Results concerning the evolution of the violation numbers are presented in Fig. 3. We considered two different kinds of rules: (i) DAY-Rules: these rules define the maximum (or minimum) number of portions of a specific food category that can be consumed during a day, and (ii) WEEK-Rules: these rules define the maximum (or minimum) number of portions of a specific food category that can be consumed during a week. DAY-Rules are verified at the end of each day, while WEEK-Rules are verified at the end of each week. The blue and the purple lines represent the average number of violations observed for the *Key To Health* and *Salute Plus* users, respectively. The red and the azure lines are the standard deviations. Observations related to the control group are reported by the green (average number of violations) and the orange line (standard deviation). The increasing trend of the gap between the blue/purple and green lines (for both the DAY and WEEK-Rules) demonstrates the positive impact of the messages sent by the whole platform. In particular, concerning DAY-Rules, the average number of violations per user at the end of the observed period is acceptable as it drops of about 67%. For the WEEK-Rules, however, the drop remained limited. Notice that for both living labs we have a confident decrease of detected violations. Hence, we can conclude that the whole platform was effective within both living labs. The standard deviation is higher for the *Salute Plus* living lab: this is due to the high number of involved people that, unavoidable, led to a marked variance of their behaviors. Notice that both standard deviation lines remain contained within low bounds. In addition, we did not detect the presence of outliers.

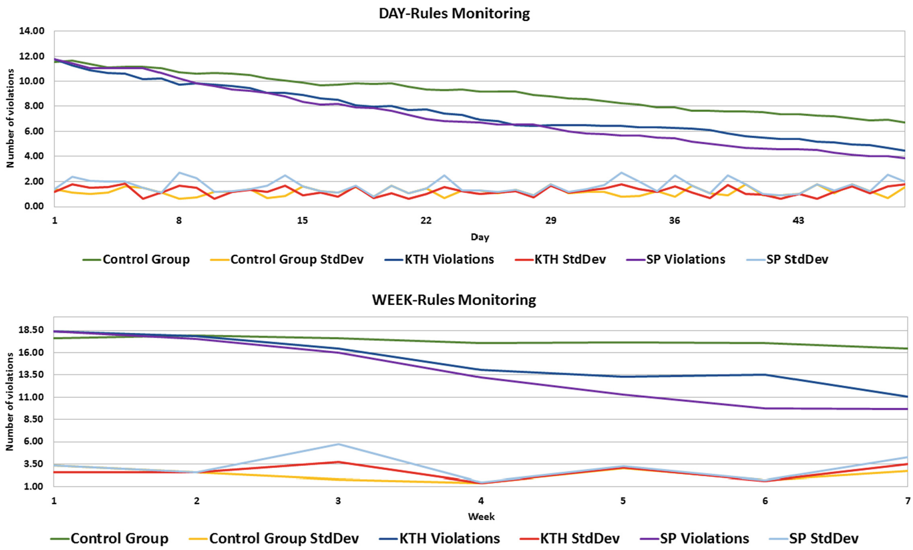

Fig. 3. Evolution of the average number of detected violations (per user), for the DAY and WEEK-Rules, during the *Key To Health* and *Salute Plus* observation period. (Color figure online)

8.3 Lessons Learnt

Both the *Key to Health* and *Salute Plus* experiences allowed us to collect some lessons that will improve the effectiveness of our platform and the design of future living labs.

Real-Time Suitability. The proposed system aims to be deployed into a real-time context. Personalized feedback and recommendations have to be provided timely to users based on the evolution of their behaviors and of the surrounding environment. Hence, we observed the performance of the whole reasoning process implemented into our platform. Therefore, we focus on the optimization process brought us to the deployment of a solution able to support an efficient real-time generation of personalized messages. Our results derived from the optimization of rules design and rules evaluation schedule. In a first stage, we designed few complex rules for covering all possible monitoring activities. On the one hand, we were able to cover several constraints with one rule. On the other hand, the computational time required for evaluating these rules was too high leading to a personalized tracking of users' behavior that was neither effective nor efficient. Hence, in a second stage, we split the rules in simpler ones and schedule their evaluation depending on their timing property (DAY and WEEK). This strategy improved of the reasoning performance by making the platform deployable within a real-time environment and allowed us to have an easier control on the overall reasoning process. A future improvement of personal tracking capabilities will be the investigation of stream reasoning for monitoring a continuous flow of information as well as to exploit learning strategies for suggesting new rules or adaptations of existing ones. An example in the health domain is the real-time monitoring of the glycemic index.

User Perception About Personalization. We consider the actual perception that the users had about the personalization capabilities of the proposed platform. During the focus group organized at the end of the *Key to Health* use case, we collected feedback about such perception by asking to users when the system can be improved concerning personalized interactions. Overall, the users appreciated the system responsiveness and message tailoring capabilities when data about food consumption were provided. However, a common request was related to the possibility of exploiting the geographical information that can be acquired through the smartphones. This information was relevant for motivating people in changing habits within some real-life situations, e.g., to not stop at a vendor machine during a walk. Suggested examples include the possibility of sending alerts, based on the current user location, about close healthy nutrition shops, restaurants cooking recipes that are compliant with users goals and users' habits. These suggestions will lead the next version of the personalization component of our platform in order to improve the perception that the system is providing a real-time support.

9 Conclusions

This paper discusses an end-to-end semantic platform that maps food categories detected from meal images into semantic information of an ontology. The goal is alerting people about the potential risks of food with respect to their PHRs. The platform integrates (i) deep learning for classifying food categories from images; (ii) an ontology associating food categories with possible nutritional diseases; (iii) the navigation of the LOD Cloud for extracting further diseases' knowledge; (iv) the use of PHRs for the generation of proper feedback. We provided a new dataset of annotated images useful for fostering the research. Concerning the image classification, the multi-label food classification outperforms a more standard method based on single-image classification and inference of the food categories. Regarding the feedback generation, the user-based evaluation demonstrated the efficacy of our semantic platform into real-world scenarios.

Future work will focus on exploiting the combination of deep learning with ontologies (in a multi-task learning setting) by using constraints-based methods, such as, Logic Tensor Networks [24], already applied to image classification tasks. This direction will be tested on bigger and standard image datasets, such as, VIREO FOOD-172 [4]. On the semantic part, the HeLiS ontology will be extended with further diseases in order to improve the overall capability of the system. Stream reasoning algorithms will be studied to support the generation of feedback by considering the wider dietary behavior of a user instead of a single recipe. Finally, the proposed semantic platform opens the possibility of an integration into intelligent systems implementing behavior change policies for supporting users in adopting healthy lifestyles.

References

1. Bangor, A., Kortum, P., Miller, J.: Determining what individual SUS scores mean: adding an adjective rating scale. J. Usability Stud. **4**, 114–123 (2009)
2. Benmimoune, L., Hajjam, A., Ghodous, P., Andres, E., Talha, S., Hajjam, M.: Ontology-based medical decision support system to enhance chronic patients' lifestyle within e-care telemonitoring platform. Stud. Health Technol. Inform. **213**, 279–282 (2015)
3. Bossard, L., Guillaumin, M., Van Gool, L.: Food-101 – mining discriminative components with random forests. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8694, pp. 446–461. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10599-4_29
4. Chen, J., Ngo, C.: Deep-based ingredient recognition for cooking recipe retrieval. In: ACM Multimedia, pp. 32–41. ACM (2016)
5. Chi, Y., Chen, T., Tsai, W.: A chronic disease dietary consultation system using OWL-based ontologies and semantic rules. J. Biomed. Inform. **53**, 208–219 (2015)
6. Ciocca, G., Napoletano, P., Schettini, R.: Food recognition: a new dataset, experiments, and results. IEEE J. Biomed. Health Inform. **21**(3), 588–598 (2017)
7. Ciocca, G., Napoletano, P., Schettini, R.: Learning CNN-based features for retrieval of food images. In: Battiato, S., Farinella, G.M., Leo, M., Gallo, G. (eds.) ICIAP 2017. LNCS, vol. 10590, pp. 426–434. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70742-6_41
8. Dehais, J., Anthimopoulos, M., Shevchik, S., Mougiakakou, S.G.: Two-view 3D reconstruction for food volume estimation. IEEE Trans. Multimed. **19**(5), 1090–1099 (2017)
9. Dragoni, M., Bailoni, T., Maimone, R., Eccher, C.: HeLiS: an ontology for supporting healthy lifestyles. In: Vrandečić, D., et al. (eds.) ISWC 2018. LNCS, vol. 11137, pp. 53–69. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00668-6_4

10. Dragoni, M., Rospocher, M., Bailoni, T., Maimone, R., Eccher, C.: Semantic technologies for healthy lifestyle monitoring. In: Vrandečić, D., et al. (eds.) ISWC 2018. LNCS, vol. 11137, pp. 307–324. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00668-6_19
11. Ege, T., Yanai, K.: Image-based food calorie estimation using knowledge on food categories, ingredients and cooking directions. In: ACM Multimedia (Thematic Workshops), pp. 367–375. ACM (2017)
12. Fernández, J.D., Beek, W., Martínez-Prieto, M.A., Arias, M.: LOD-a-lot - a queryable dump of the LOD cloud. In: d'Amato, C., et al. (eds.) ISWC 2017. LNCS, vol. 10588, pp. 75–83. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68204-4_7
13. Kawano, Y., Yanai, K.: Food image recognition with deep convolutional features. In: UbiComp Adjunct, pp. 589–593. ACM (2014)
14. Kawano, Y., Yanai, K.: Foodcam-256: a large-scale real-time mobile food recognition system employing high-dimensional features and compression of classifier weights. In: ACM Multimedia, pp. 761–762. ACM (2014)
15. Lu, Y., Allegra, D., Anthimopoulos, M., Stanco, F., Farinella, G.M., Mougiakakou, S.G.: A multi-task learning approach for meal assessment. In: MADiMa@IJCAI, pp. 46–52. ACM (2018)
16. Maimone, R., Guerini, M., Dragoni, M., Bailoni, T., Eccher, C.: Perkapp: a general purpose persuasion architecture for healthy lifestyles. J. Biomed. Inform. **82**, 70–87 (2018)
17. Mamykina, L., Levine, M.E., Davidson, P.G., Smaldone, A.M., Elhadad, N., Albers, D.J.: Data-driven health management: reasoning about personally generated data in diabetes with information technologies. JAMIA **23**(3), 526–531 (2016)
18. Mezgec, S., Koroušić Seljak, B.: Nutrinet: a deep learning food and drink image recognition system for dietary assessment. Nutrients **9**(7), 657 (2017)
19. Mikolajczak, S., Ruette, T., Tsiporkova, E., Angelova, M., Boeva, V.: A semantic reasoning engine for lifestyle profiling in support of personalised coaching. In: International Conference on Global Health Challenges, pp. 79–83 (2015)
20. Myers, A., et al.: Im2Calories: towards an automated mobile vision food diary. In: ICCV, pp. 1233–1241. IEEE Computer Society (2015)
21. Nielsen, J., Landauer, T.K.: A mathematical model of the finding of usability problems. In: Proceedings of ACM INTERCHI 1993 Conference, pp. 24–29 (1993)
22. Okamoto, K., Yanai, K.: An automatic calorie estimation system of food images on a smartphone. In: MADiMa @ ACM Multimedia, pp. 63–70. ACM (2016)
23. Pouladzadeh, P., Shirmohammadi, S., Almaghrabi, R.: Measuring calorie and nutrition from food image. IEEE Trans. Instrum. Measur. **63**(8), 1947–1956 (2014)
24. Serafini, L., Donadello, I., d'Avila Garcez, A.S.: Learning and reasoning in logic tensor networks: theory and application to semantic image interpretation. In: SAC, pp. 125–130. ACM (2017)
25. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: CVPR, pp. 2818–2826. IEEE Computer Society (2016)
26. Tilahun, B., Kauppinen, T., Keßler, C., Fritz, F.: Design and developmentof a linked open data-based health information representation andvisualization system: potentials and preliminary evaluation. JMIR Med. Inform. **2**(2), e31 (2014)