THE DEVELOPMENT OF AI-BASED FUNCTIONALITIES TO SUPPORT A MOBILE APPLICATION IN PERSONALISED NUTRITION AND PHYSICAL ACTIVITY

A White Paper

Abstract

This paper presents the design and development of three specialized AI-based functionalities that are part of the backend of the PROTEIN mobile application and aim to facilitate the development of personalized nutrition and physical activity plans, as well as the recognition of food categories and meal ingredients and the estimation of a user's eating rate. These functionalities are considered essential for users to be guided towards personalized and healthy diet and lifestyle.

Introduction / Background

Food is of utmost importance to human physical and mental health as it does not only provide the necessary energy for the functioning of the human body, but it is also a vehicle that promotes joy, friendship and social intercourse. However, not all types of food are equally beneficial to the overall health, as their accompanying nutrients are processed differently by the human body. Thus, maintaining a well-balanced and healthy diet is crucial for the health and nutrition of humans, especially now that food globalization, lifestyle changes, economic, and socio-cultural factors pose significant challenges to the adherence of people to healthy diets [1]. This is why food intake monitoring systems can provide substantial benefits to their users by suggesting ways to adapt to and maintain a healthy diet. However, to keep users engaged to a healthy diet and to achieve optimized results and benefits, personalized nutrition is the key to success.

Food recommender systems are often categorized based on the method used to recommend healthy diets to content-based (CB), collaborative filtering-based (CF) and hybrid approaches [2, 3, 4]. Content-based approaches rely on users' individual tastes, activities and profile to tailor recommendations through the collection of user scores who rate their preference on specific foods or recipes [5, 6]. On the other hand, collaborative filtering-based approaches attempt to find similarities between user profiles and as a result they make recommendations that would be appropriate for similar users [5, 7]. Finally, hybrid approaches combine CB and CF techniques based on the notion that balancing user preferences and nutritional needs is the optimal strategy

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 817732

[8, 9]. Such techniques employ dietary constraints based on the users' health and nutritional expectations, which are then combined with the user preferences to identify optimal meals.

Food intake identification plays an important role in the development of automated nutrition recommender systems, as traditional approaches for food intake monitoring, such as questionnaires, are effective only up to a certain point. Although food intake assessment is not such a demanding task when dealing with packaged food products since the nutritional content is usually provided by the manufacturers, it can be really challenging in the case of meals prepared at home or restaurants. In this case, food images can be employed to achieve up to a certain degree food identification, considering the existence of related publicly available databases [10, 11] and well documented deep-learning-based approaches [12, 13].

On the other hand, eating rate analysis is very important in personalized nutrition studies since eating behavior can be correlated with obesity and eating disorders [14]. Eating behavior is usually estimated through self-reporting measures, despite their limitations in reliability, based on ease of collection and analysis [15]. A better and widely used alternative is the objective analysis of eating during meals based on human annotations of in-meal behavioral events (e.g., bites). However, this methodology is time-consuming and often affected by human error, limiting its scalability and cost-effectiveness for large-scale research [16]. Thus, it is imperative to research new methodologies for the automatic estimation of eating-related events.

Proposed Solution(s)

A. Food and Activity Recommender System (FARS)

This system is responsible for generating and recommending personalized nutrition and physical activity (NAP) plans to PROTEIN users. It makes recommendations by combining two different layers of personalization: a) dietary rules defined by nutrition experts and b) user preferences. An overview of the NAP generation process is shown in Figure 1.

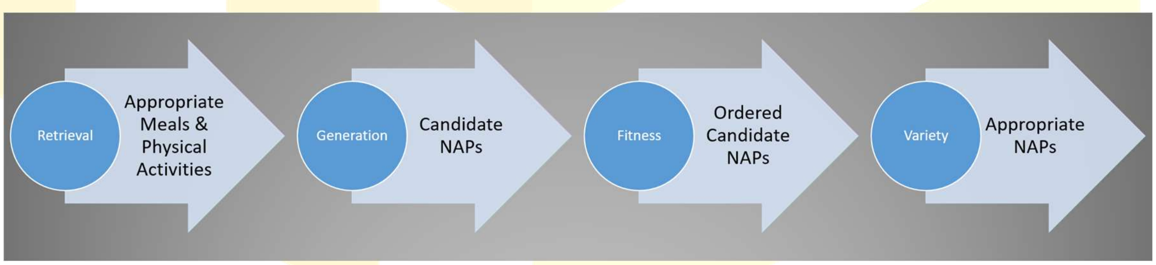

Figure 1: NAP generation process

Initially, a knowledge-based expert system, called Reasoning-based decision support system (RDSS) is employed to propose appropriate meals and physical activities through logical reasoning over experts' knowledge, user profiles and the available meal options stored in a database. At its core, RDSS is comprised of a powerful fuzzy inferencing engine

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 817732

that allows for semantic matchmaking of candidate meal options against user information (i.e., medical conditions, user-declared food preferences, sensor measurements, etc.) and expert's knowledge on the nutritional content of foods, food restrictions and specific nutritional needs relevant to particular user conditions.

Then, candidate daily NAPs are generated by taking into account the list of appropriate meals and physical activities returned from the RDSS system, as well as the personalized nutritional and physical activity requirements and goals of the PROTEIN users in terms of micronutrients, macronutrients and energy intake formulated based on the experts' knowledge and the user profiles. The daily NAPs are generated by randomly placing and combining meal options for the different meals of the day (i.e., breakfast, morning snack, lunch, afternoon snack, dinner and supper), as well as for morning and afternoon physical exercises.

The generated daily NAPs are then ordered based on the user preferences to form weekly NAPs, relying on the notion that plans desired by users are more easily to be followed and adopted. Finally, a set of rules are employed to add variety to the generated NAPs by removing repetitive meal options and repetitive sequences of meals in weekly NAPs. In this way, the proposed FARS system is capable of recommending nutritional and physical activity plans tailored to the medical conditions and dietary needs of users, while also providing meal options that are both desirable and of large variety so that users are more inclined to follow and maintain to improve their health and well-being.

FARS was integrated with the mobile application and deployed in the PROTEIN cloud, allowing the generation of NAPs to be made on the PROTEIN server, while frontend and backend services to depict the personalized NAPs and other important information and allow users to state preferences being readily available in the PROTEIN mobile application. Indicative snapshots of generated NAPs for PROTEIN users are presented in Figure 2.

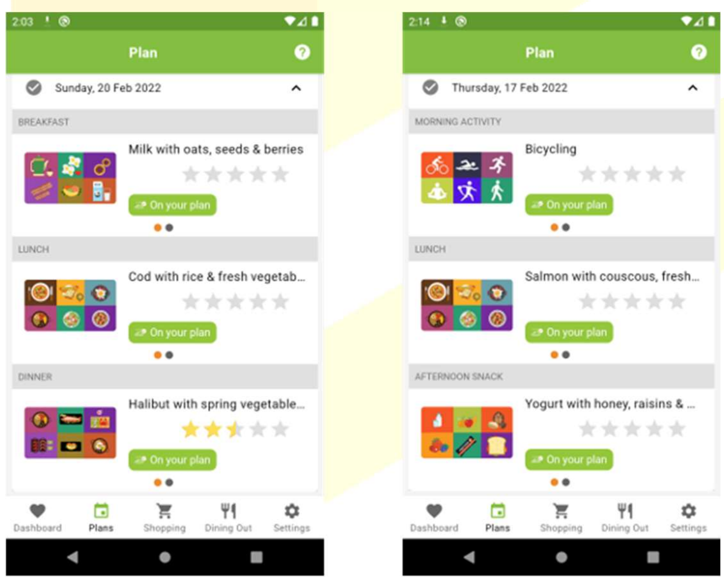

Figure 2: Snapshots of NAPs generated by the proposed FARS system.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 817732

B. Food category recognition algorithm

Leveraging the growing and popular habit of smartphone users taking pictures of their meals, a novel algorithm for the identification of food categories and ingredients from single images was developed and integrated in the PROTEIN mobile application as a backend service with a frontend screen [17]. To overcome the need for annotation of new images every time it is required to expand the recognition capability of the model to new food categories, the proposed zero-shot category recognition method (shown in Figure 3) based on transfer learning was proposed.

More specifically, a deep network, pre-trained on Recipe1M [18] dataset is employed to transform the input image to the associated ingredients (~1,000 different options) (1). Afterwards, each recognized ingredient is transformed to a word-vector using the word2vec algorithm [19, 20], which is a neural network that aims to learn word associations in order to detect synonymous words or suggest additional words for forming a partial sentence (2). In the case of food identification, the word2vec transformation allows for exploitation of the semantic relationship between the ingredients recognized in the first step, which serve as input to a new neural network (3) that infers the word vector of the food class (4). The neural network is pre-trained on the Recipe1M ingredient vectors (input) and corresponding food categories vectors (label), obtained using the word2vec algorithm. Finally, the output food class vector is passed through the word2vec and Annoy[1] library that yield its nearest neighbors, i.e., the most relevant food categories (5).

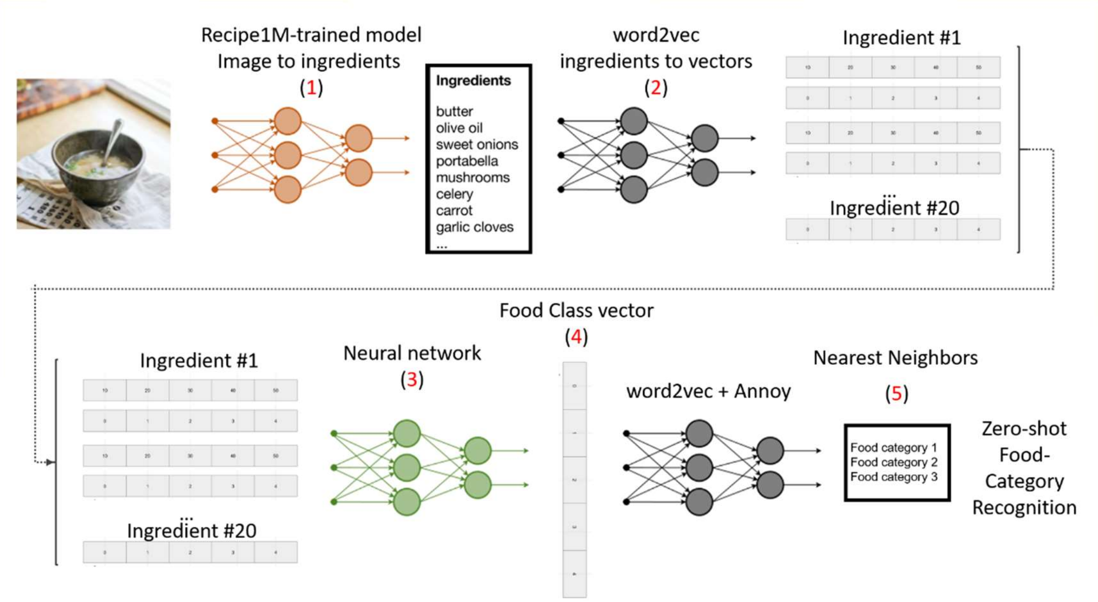

Figure 3: The proposed methodology for zero-shot food category recognition based on the word2vec algorithm.

[1] https://radimrehurek.com/gensim/auto_examples/tutorials/run_annoy.html

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 817732

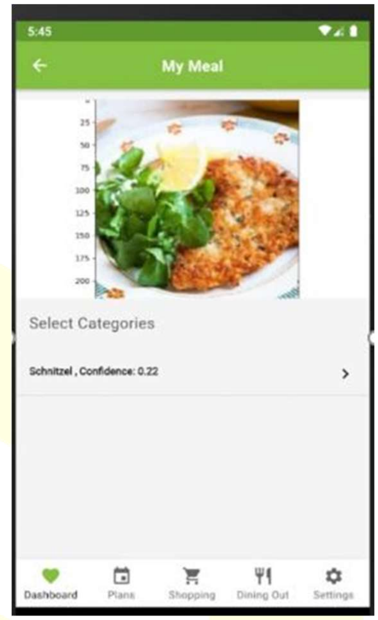

Figure 4: Snapshot of the food category recognition algorithm on the PROTEIN mobile application.

The proposed food category recognition algorithm runs on the PROTEIN mobile application and receives as input an image either from the user's photo library or from the phone's camera directly. The output is a list of predicted food categories, ranked by their confidence score. An indicative snapshot of an output of the module running on the PROTEIN application can be seen in Figure 4.

C. Eating rate analysis algorithm

To accurately estimate a user's eating rate in an unobtrusive way (i.e., without employing wearable sensors that can inhibit user's actions), a novel algorithm that receives as input video sequences from the smartphone camera was developed and integrated in the PROTEIN mobile application as a backend service with a frontend screen [21, 22]. More specifically, the proposed eating rate analysis algorithm is presented in Figure 5 and consists of two main components or networks. The first network is a well-known skeletal feature extraction network, called OpenPose [23, 24] that is responsible for processing the input video and extracting 2D points that correspond to joints of the body and face. Then, only a small portion of these joints and more specifically the joints of hands, head, neck and mouth are selected as input to a second deep network that is concerned with the modelling of their time variations to identify movements that correspond to bite instances. From these bite instances, the eating rate is computed as bites performed in a time interval of one minute.

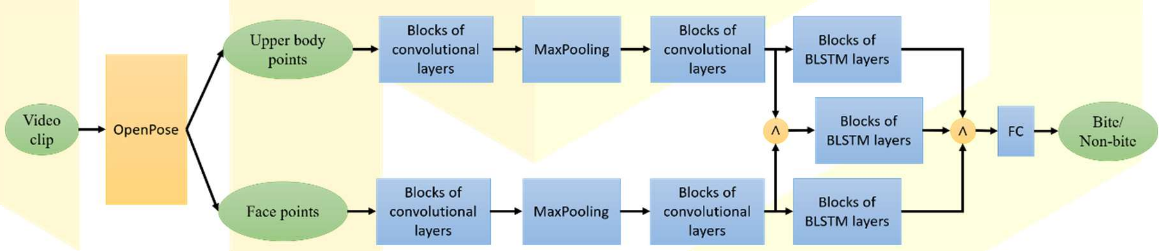

Figure 5: Pipeline of the proposed eating rate analysis algorithm.

In the PROTEIN mobile application, a user can press a button to start and stop a meal video recording, while he/she can also see what the camera of his/her smartphone is showing to adjust his/her position and distance from the camera so that both hands and face are visible during the recording. The proposed bite detection algorithm is executed in near real-time on the background by grabbing and processing video frames on the fly, while avoiding storing images in the device or uploading them anywhere, thus respecting the privacy of the user. Moreover, audio-visual feedback is provided to the user to inform them on their eating rate and how far is this eating rate from the normal levels. In the current

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 817732

implementation, a normal eating rate of 8 bite instances per minute is considered. Figure 6 presents a few snapshots of the eating rate analysis algorithm running on the PROTEIN mobile application, along with visual feedback regarding the user's eating rate. A colored bar represents the user's eating rate, with orange color denoting a low eating rate, green color denoting a normal eating rate and red color denoting a high eating rate. A high eating rate is also accompanied by auditory feedback in the form of a 'beep' sound that informs the user to reduce his/her eating rate.

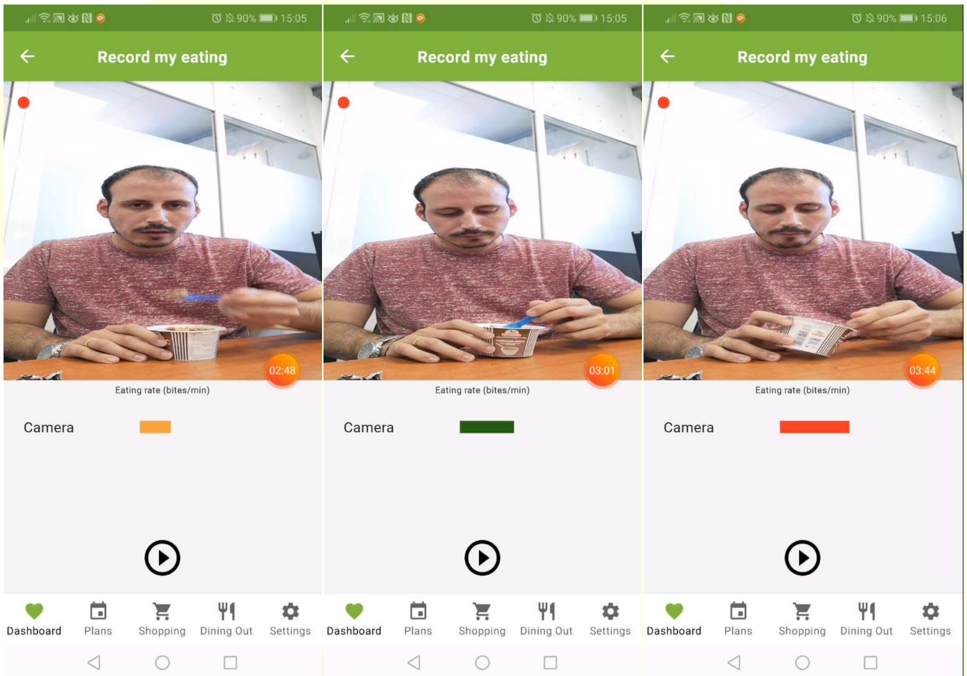

Figure 6: Snapshot of the eating rate analysis algorithm in the PROTEIN mobile application. Note: The participant shown in the picture provided his consent to appear in this study.

Future Direction / Long-Term Focus

Regarding, the FARS system the long-term focus is on its extension towards taking into consideration additional nutrients and optimizing its plan selection algorithm. Moreover, the replacement of the FARS rule-based components with deep networks will also be investigated. On the other hand, regarding the food category recognition algorithm, the future direction involves the use of more sophisticated deep network architectures and the exploitation of new large publicly available food datasets for training that can lead to significant improvements in the accuracy of the algorithm. Finally, regarding the eating rate analysis algorithm, the long-term focus is on the use of less computationally demanding deep networks and improved network pruning techniques that can significantly improve the speed of the algorithm on mobile devices. Additionally, the training of the algorithm with additional data can enhance its accuracy and robustness in the task of eating rate analysis.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 817732

Results / Conclusion

This white paper presents three specialized AI-based functionalities that are responsible for a) providing personalized nutrition and physical activity plans, b) recognizing food categories and meal ingredients and c) estimating a user's eating rate. The FARS system was validated with 3000 virtual users and its generated plans achieved accuracies of 93.3%, 92.7% and 85.9% in terms of energy intake (i.e., calories), macronutrients and micronutrients, respectively, with respect to the proposed by the nutritional experts quantities. On the other hand, the food category recognition algorithm was tested in Recipe1M food dataset, achieving a state-of-the-art performance of 0.5 F1-score in ingredient recognition, while the accuracy of the algorithm in the final 38 food categories was 84.6%. In a similar fashion, the eating rate analysis algorithm was tested in the EBEPA+ dataset [25, 26], consists of 37 videos depicting healthy individuals eating different types of meals and 39 videos of individuals suffering from Parkinson's disease. The results showed the ability of the proposed algorithm to robustly recognize bite instances with an F1-score of 0.93, while a comparison with manual annotations revealed a total correlation coefficient for the number of bites equal to 0.94 and a total correlation coefficient for the meal duration equal to 0.99. Thus, the proposed algorithm is capable of performing a detailed behavioral analysis of meals without any loss of information and fidelity, in a fraction of the time (and effort) that would be required if the meal annotation was performed by trained human annotators. These results demonstrate the accuracy of the PROTEIN's AI-based functionalities in their respective tasks and their suitability for a system that can analyze users' dietary behavior to provide personalized and healthy nutritional and physical activity plans tailored to the users' needs and preferences.

Appendices

Appendix A – Authors

Konstantinidis Dimitrios – Primary author

Stefanidis Kiriakos – Contributor

Tsatsou Dorothea - Contributor

Theodoridis Thomas - Contributor

Gymnopoulos Lazaros - Contributor

Dimitropoulos Kosmas - Contributor

Appendix B – References

[1] Dernini, S., & Berry, E. M. (2015). Mediterranean diet: from a healthy diet to a sustainable dietary pattern. Frontiers in nutrition, 2, 15.

[2] Trattner, C., & Elsweiler, D. (2017). Food recommender systems: important contributions, challenges and future research directions. arXiv preprint arXiv:1711.02760.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 817732

[3] Theodoridis, T., Solachidis, V., Dimitropoulos, K., Gymnopoulos, L., & Daras, P. (2019, June). A survey on AI nutrition recommender systems. In Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments (pp. 540-546).

[4] De Croon, R., Van Houdt, L., Htun, N. N., Štiglic, G., Abeele, V. V., & Verbert, K. (2021). Health Recommender Systems: Systematic Review. Journal of Medical Internet Research, 23(6), e18035.

[5] Freyne, J. & Berkovsky, S. (2010). Intelligent food planning: Personalized recipe recommendation. In Proceedings of the 15th international conference on intelligent user interfaces (pp. 321–324). New York, NY, USA: ACM.

[6] Freyne, J., Berkovsky, S. & Smith, G. (2011). Recipe recommendation: Accuracy and reasoning. In International conference on user modeling, adaptation, and personalization (pp. 99–110).

[7] Ge, M., Elahi, M., Fernaández-Tobías, I., Ricci, F. & Massimo, D. (2015). Using tags and latent factors in a food recommender system. In Proceedings of the 5th international conference on digital health 2015 (pp. 105–112). New York, NY, USA: ACM.

[8] Tran, T. N. T., Atas, M., Felfernig, A., & Stettinger, M. (2018). An overview of recommender systems in the healthy food domain. Journal of Intelligent Information Systems, 50(3), 501-526.

[9] Trattner, C., & Elsweiler, D. (2019). Food Recommendations. In Collaborative recommendations: Algorithms, practical challenges and applications (pp. 653-685).

[10] Min, W., Jiang, S., Sang, J., Wang, H., Liu, X., & Herranz, L. (2016). Being a supercook: Joint food attributes and multimodal content modeling for recipe retrieval and exploration. IEEE Transactions on Multimedia, 19(5), 1100-1113.

[11] Salvador, A., Hynes, N., Aytar, Y., Marin, J., Ofli, F., Weber, I., & Torralba, A. (2017). Learning cross-modal embeddings for cooking recipes and food images. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3020-3028).

[12] Salvador, A., Drozdzal, M., Giro-i-Nieto, X., & Romero, A. (2019). Inverse cooking: Recipe generation from food images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 10453-10462).

[13] Attokaren, D. J., Fernandes, I. G., Sriram, A., Murthy, Y. S., & Koolagudi, S. G. (2017, November). Food classification from images using convolutional neural networks. In TENCON 2017-2017 IEEE Region 10 Conference (pp. 2801-2806). IEEE.

[14] Almiron-Roig, E.; Tsiountsioura, M.; Lewis, H.B.; Wu, J.; Solis-Trapala, I.; Jebb, S.A. Large portion sizes increase bite size and eating rate in overweight women. Physiol. Behav. 2015, 139, 297–302.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 817732

[15] Hufford, M.R. Special methodological challenges and opportunities in ecological momentary assessment. In The Science of Real-Time Data Capture: Self-Reports in Health Research; Oxford University Press: Oxford, UK, 2007; pp. 54–75.

[16] Robinson, E.; Hardman, C.A.; Halford, J.C.; Jones, A. Eating under observation: A systematic review and meta-analysis of the effect that heightened awareness of observation has on laboratory measured energy intake. Am. J. Clin. Nutr. 2015, 102, 324–337.

[17] Theodoridis, T., Solachidis, V., Dimitropoulos, K., Daras, P., A Cross-Modal Variational Framework for Food Image Analysis, in International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, October 25-28, 2020.

[18] Salvador, A., Hynes, N., Aytar, Y., Marin, J., Ofli, F., Weber, I., & Torralba, A. (2017). Learning cross-modal embeddings for cooking recipes and food images. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3020-3028).

[19] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

[20] Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. arXiv preprint arXiv:1310.4546.

[21] D. Konstantinidis, K. Dimitropoulos, I. Ioakimidis, B. Langlet, and P. Daras, "A Deep Network for Automatic Video-Based Food Bite Detection," in 12th International Conference on Computer Vision Systems, 2019, pp. 586–595.

[22] D. Konstantinidis, K. Dimitropoulos, B. Langlet, P. Daras, and I. Ioakimidis, "Validation of a Deep Learning System for the Full Automation of Bite and Meal Duration Analysis of Experimental Meal Videos," Nutrients, vol. 12, no. 1, p. 209, 2020.

[23] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," in 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017, pp. 7291–7299.

[24] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, "Hand keypoint detection in single images using multiview bootstrapping," in 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017, pp. 4645–4653.

[25] P. Fagerberg et al., "Lower Energy Intake among Advanced vs. Early Parkinson's Disease Patients and Healthy Controls in a Clinical Lunch Setting: A Cross-Sectional Study," Nutrients, vol. 12, no. 7, p. 2109, 2020.

[26] K. Kyritsis et al., "Assessment of real-life eating difficulties in Parkinson's disease patients by measuring plate to mouth movement elongation with inertial sensors," Sci. Rep., vol. 11, no. 1, p. 1632, 2021.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 817732