An Evolutionary Habit Recommender to reduce the risk of Overweight and Obesity

Omitted for blind review

No Institute Given

Abstract. This paper presents the design of a healthy habits recommender system based on evolutionary computing. The evolutionary algorithm is tasked with finding values for a set of variables, namely modifiable habits, to reduce the risk of obesity, taking into account a selection of fixed values such as age or gender.
In order to evaluate the different combinations of values generated by the evolutionary algorithm, we employed a series of models obtained through Machine Learning techniques. The data used to train the models was collected as part of the Genobia project, involving over 1100 individuals who completed a survey on lifestyle habits and social-economic status and a genetic analysis.
The objective of the work presented in this document is to provide a practical tool to assist individuals in making changes to their habits and routines in order to prevent or reduce the risk of obesity.

Keywords: Evolutionary computation · Obesity · Fitness · Habit Recommender · Chromosome · Genetic Algorithm

1 Introduction

Overweight and Obesity is a significant global public health issue, linked to numerous chronic diseases such as diabetes, heart disease, and certain cancers ([6]). As the incidence of obesity continues to rise, the need for effective preventive measures becomes increasingly critical. Traditional approaches to obesity prevention often fail to consider the personalized nature of risk factors, which vary greatly depending on age, lifestyle, and other individual characteristics.

In recent years, evolutionary computation ([5]) has emerged as a powerful tool for solving complex optimization problems across various domains. Inspired by natural evolutionary processes, this method has shown considerable promise in addressing health-related challenges by optimizing multiple variables simultaneously. The application of evolutionary algorithms to personalized health recommendations offers a novel approach to mitigating obesity risk, allowing for tailored advice that adapts to the unique needs of each individual.

The motivation behind this work stems from the potential of evolutionary computation to create a more effective, personalized habit recommender system. By optimizing lifestyle factors such as diet and physical activity, the system aims to provide individuals with actionable guidance that can significantly reduce their risk of obesity.

The keys to design and implement a habit recommender system are:

- **Training** of logical models of the risk of being overweight or obese to evaluate the individuals and quantify how good they are.
- **Optimization** of the values for the selected habit's variables.
- **Development** of a friendly graphical user interface.

The main contributions of this work are:

–

The rest of this paper is structured as follows: Section ...

2 State of the art

Throughout the years, health has been a constant topic of discussion, with numerous theories and recommendations emerging about which habits contribute to a healthier lifestyle. From dietary guidelines to the risks (or benefits, as some studies point to) of alcohol and tobacco consumption, the pursuit of well-being has led to a diverse array of approaches, each claiming to offer the best path to longevity and quality of life.

However, the advancements of digital technology have revolutionized how we understand and manage our health, offering personalized insights that were previously unfathomable.

Our objective is to take a innovative approach, making a program that is able to give customised advice depending on the user's habits by leveraging evolutionary computing. Our application takes the user's habits into account and makes changes relative to the previous ones so that the user does not feel overwhelmed by them.

Recommenders are everywhere in our day to day life, from the ones used on e-commerce to show us related items to the ones we have already bought, to more hidden ones, such as the ones social media use to show us whatever content is related to what we have liked before.

While there are plenty of websites that offer health advice, these only provide generic recommendations that are mostly common knowledge, such as 'Aim for eight hours of sleep a night' or 'Eat a nutritious breakfast every morning', which are not bad suggestions, but they are too standard and non-specific as to make a difference in someone's life if followed one by one, and too overwhelming to implement all at once.

On the other hand, there are other applications that take a more specific approach into a single topic with an emphasis on the user achieving these goals.

Some of these applications are:

- **MyFitnessPal**

This MyFitnessPal mobile application keeps track of the calories consumed everyday by maintaining a meal diary and informing the user of their nutritional values and components. It also has a more fitness oriented section wherein workout routines are suggested for the user (though it seems more of a secondary function, as it is more hidden). When compared to our application it is a similar approach, yet with a more narrow set of habits, which allows it to focus more on the specifics of the suggestions provided. [3]
- **Holly Health Habit Coach**
  Holly Health Habit Coach is more oriented towards the creation of healthy habits, divided between fitness related habits (exercising and eating), sleeping habits and mental health habits. Upon login, an AI takes your information and suggests one or two habits so that you can get accustomed to them progressively. According to our tests this application provides less personalized advice, with more focus towards the upkeeping of these habits. In studies like "Small steps, long-term outcomes", "How does Holly Health impact health and well-being" and "Holly Health impacts pain, mood, and self-management skills", the efficacy of this application is shown in terms of habit creation.
- **SmokeFree and Drinkaware**
  SmokeFree and Drinkaware are not mobile apps, but rather websites that provide free advice on how to stop respectively smoking and drinking habits, information of their effects on the user's health, various replacement treatments, and live chats with agents to talk about their day to day problems resulting from these habits. Smoking and drinking are widely known for being noxious towards the user's health, as seen in [2] and [1].
  These applications, instead of giving plain advice, aim to make the user aware of their problem and try to be more helpful towards getting them to quit it by providing temporary replacements and support plans.

With limitations as lack of customization, being limited to a certain scope, being either too insistent with their reminders or users forgetting about them too easily present on the most similar programs, we present a different approach by giving more concrete advice and providing more possibilities to improve health related habits by gathering a broader spectrum of data.

3 Genobia Dataset

The datasets we employed in the development of the habit recommender system for models training and testing are from the GenObIA Project.

The GenObIA Project focuses on the development of predictive algorithms using artificial intelligence to identify individuals at risk of developing overweight, obesity, and associated pathologies.

The Consortium GenObIA is constituted of Research Groups and Associated Groups. One of the Research Groups is the Complutense University of Madrid, specifically the Faculty of Medicine and the Faculty of Computer Science.

4 Habit Recommender system

4.1 Ensembled Model for the prediction of Obesity

Grammatical Evolution is a Genetic Programming technique whose objective is the same as all other Evolutionary Algorithms: to find a function with the best fitness possible for a given objective. These functions take the form of series of integers that are mapped to a program through the use of a grammar. We apply genetic operators to select, mutate and cross the functions in order to get the best solution we can find.

The models themselves were trained through the use of the Grammatical Model Tool, a versatile program developed by the AbSys Group that uses Grammatical Evolution to train solutions which can be later used in other programs, or to analyze correlation between different variables of a problem. For said training, the Grammatical Model Tool program receives a file with the parameters that will be used, containing the number of times the program will ran, the number of generations it will run for each time, how many models there will be in each iteration, and some other settings that the program will use.

In our case, we ran the program 3000 times with 500 generations for each run, and 500 individuals in each of these generations. For the evolutionary algorithm, we used a 70% crossover rate, a 20% mutation probability, and tournament selection, with 3 models for each selected individual.

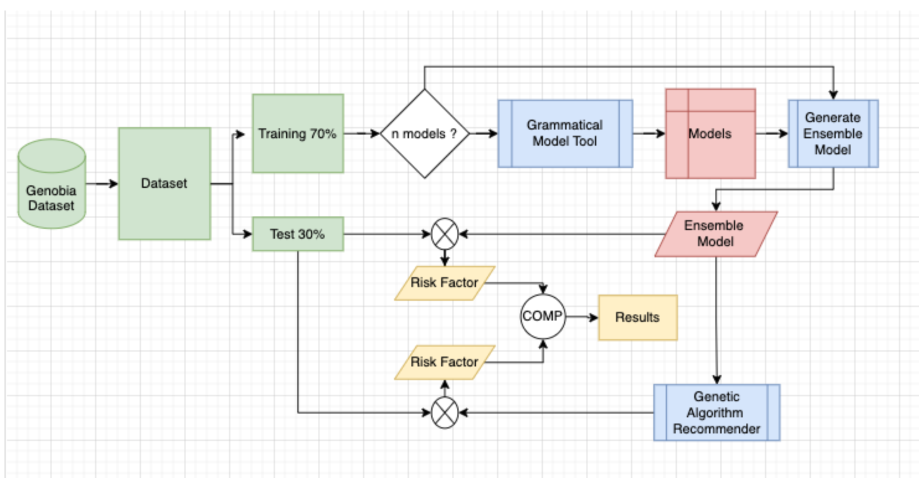

Fig. 1: Model assembly

Once the program finished, we performed validation on the produced models (we used around 70% of the GenObIA database selected randomly for the training, and left the remaining 30% for the validation), and selected the top 100 that brought us the best results, in order to use them as the core of our application.

Finally, we take these 100 models and combine them into a single, much bigger model, which we use to evaluate each of the individuals our program generates, to find out the user's risk of suffering obesity.

4.2 Recommender system

Explicar el esquema

Before getting into the development of the habit recommender system, we will show the different parts of the application in Fig 2 .

First of all, we had to train the models. To do so, we used the datasets provided by GenObIA Project and BNF grammars ([5]) in PMT. Then, from the resulting models, we select the best ones to use as the evaluation system in the algorithm.

After that, we implemented a functionality to initialize the population based on the initial features the user submits and the different phases of the Genetic Algorithm.

The way the recommender system works is that it receives the users features as input and initializes the population based on that input. Then it evaluates the new population and starts the cycle of the Genetic Algorithm. The population goes through the different phases of the algorithm in every iteration, from the generation of the elite, the selection, the crossover, the mutation, the evaluation and finally the introduction of the elite. When the looping ends, the best individual from the population is selected and outputted as list of recommendations.

4.3 Fitness evaluation

The models are logical expressions that vary in length and number of variables. To evaluate the fitness of the individuals, we replace the variables in the models with the values of the respective variables in the individual and evaluate the expressions one by one. If the expression is true, it adds one to the fitness, otherwise, we go to the next model (Fig 3).

4.4 Chromosome

The variables we used in our program changed through the development of the habit recommender system but from the start we divided the variables in 2 different groups: the variables that will not change during the execution of the algorithm and the ones that will. The reason is that even though every variable will be taken into account in the evaluation, some elements can be changed but others can not.

After testing and analyzing results, we ended up with 49 variables, its distribution is shown in Fig 4 and the final list of variables is listed below:

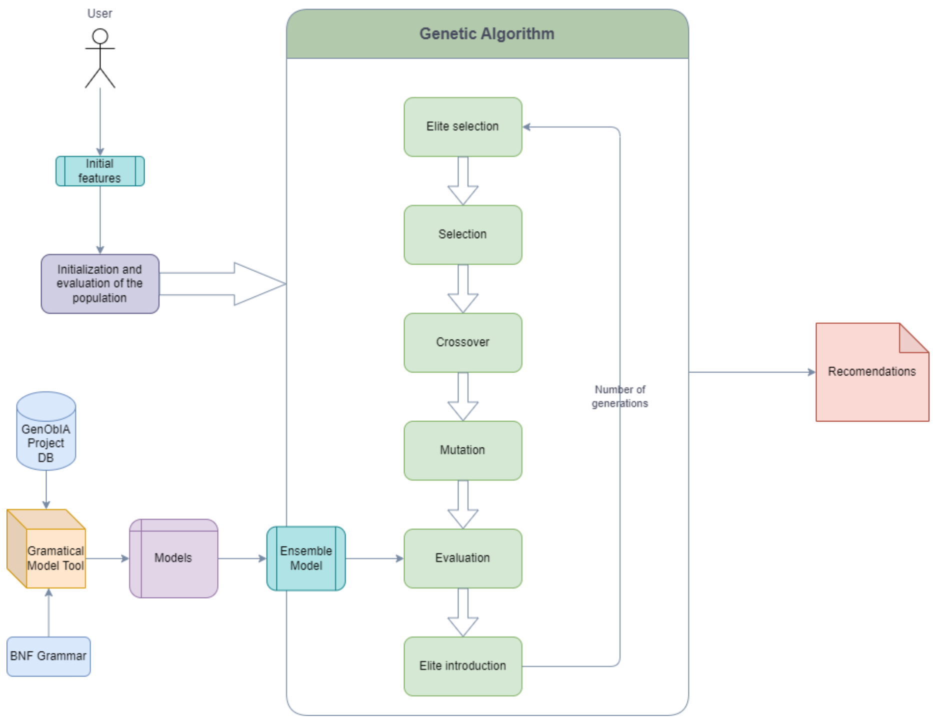

Fig. 2: Diagram of the program

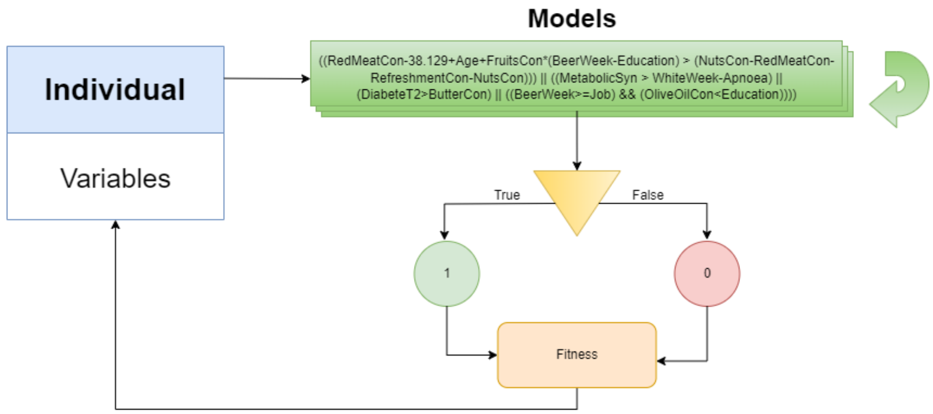

Fig. 3: Diagram of the fitness evaluation

1. **Sex**, whether the individual is male or female at birth.
2. **Age**.
3. **Population**, the number of people that lives in the individual's locality, the ranges goes from less than 2.500, between 2.500 and 20.000, between 20.000 and 50.000 and more than 50.000.
4. **Education**, the level of education of the individual, it can be no education, elementary education, secondary education or university education.
5. **Earning**, the economic situation of the individual, the ranges are less than 1.000€ per month, between 1.000€ and 2.000€ per month and more than 2.000€ per month.
6. **Job**, the type of job the individual has, from a list of 16 type of jobs.
7. **Stress**, whether the individual is stressed or not.
8. **Sleep 8**, whether the individual sleeps at least 8 hours.
9. **Spirit week**, the number of glasses of spirit per week.
10. **Beer week**, the number of glasses of beer per week.
11. **Wine week**, the number of glasses of red wine per week.
12. **White week**, the number of glasses of white wine per week.
13. **Pink week**, the number of glasses of pink wine per week.
14. **Number of smokes**, number of the cigarettes per week
15. **Pipe**, number of pipes or pipe charges smoked daily by the user.
16. **Cigar**, number of cigars smoked daily by the user.
17. **Ex-smoker years**, represents the years that the individual stopped smoking.
18. **Ex-smoker unknown**, the individual stopped smoking, but didn't know for how much time.
19. **Cancer mam**, breast cancer.
20. **Cancer col**, colon cancer.
21. **Cancer pros**, prostate cancer.
22. **Cancer lung**, lung cancer.

Constants	
Sex	Cancer_col
Age	Cancer_pros
Population	Cancer_lung
Education	Cancer_other
Earning	Cancer_mam
Job	Heart_attack
Ex-smoker Year	Heart_angina
Ex-smoker unk	Heart_failure
Diabetes	Metabolic_syn
COPD	Apnea
Stress	Asthma

Variables	
Spirit week	Refreshment cons
Beer week	Legume cons
Wine week	Fish cons
White week	Industrial bakery cons
Pink week	Nuts cons
Num smokes	White meat cons
Pipe	Sauté cons
Cigar	Dairy cons
Sleep 8	Skimmed cons
Olive oil cons	Intense exer min week
Vegetables cons	Moderate exer min week
Fruits cons	Min seated
Red meat cons	Walking min week
Butter cons	

Fig. 4: Distribution of the variables 3

23. **Cancer other**, other type of cancer.
24. **Heart attack**, whether the individual suffers from myocardial infarction.
25. **Heart angina**, whether the individual suffers from angina pectoris.
26. **Heart failure**, whether the individual suffers from heart failure.
27. **Diabetes**, whether the individual suffers from type 2 diabetes.
28. **Metabolic syn**, whether the individual suffers from metabolic syndrome.
29. **Apnoea**, whether the individual suffers from apnoea.
30. **Asthma**, whether the individual suffers from asthma.
31. **COPD**, whether the individual suffers from chronic obstructive pulmonary disease.
32. **Olive oil consumption**, tablespoons of olive oil consumed per day.
33. **Vegetables consumption**, portions of vegetables eaten per day. Garnish counts as half of a portion.
34. **Fruits consumption**, pieces of fruit eaten per day, including natural fruit juice.
35. **Red meat consumption**, portions of beef or pork, including burgers, sausages, and cold meats, per day. Each portion would be the equivalent of 100 to 150 grams.
36. **Butter consumption**, portions of butter or cream per day. Each portion would be equivalent to 12 grams.
37. **Refreshment consumption**, glasses of carbonated or sweet drinks per day.
38. **Legume consumption**, portions of legumes per week. Each portion would be equivalent to 150g.
39. **Fish consumption**, portions of fish or seafood per week. Each portion would be equivalent of 100 to 150 grams of fish or 4 to 5 pieces of seafood.
40. **Industrial bakery consumption**, times the individual consumes industrial bakery per week.

41. **Nuts consumption**, portions of nuts per week. Each portion would be equivalent to 30 grams.
42. **White meat consumption**, whether the individual prefers eating white meat such as chicken, turkey or rabbit instead of red meat.
43. **Sauté consumption**, times the individual consumes sauté as accompaniment of pasta, rice or other dishes per week.
44. **Dairy consumption**, times the individual consumes dairy products per day.
45. **Skimmed consumption**, whether the dairy products are skimmed or not.
46. **EIMS**, minutes per week of intense exercise.
47. **EMMS**, minutes per week of moderate exercise.
48. **ECMS**, minutes per week of walking.
49. **Minutes seated**, the average time the individual has spent sit in the last week

5 Individual design and implementation

The Individual Class contains the chromosome with the information of the individuals and evaluates the quality with the fitness evaluation system.

5.1 Chromosome

Each allele of the chromosome of the individuals would represent a element of the variable list. As some elements can not be modified due to their nature, we implemented an auxiliary boolean array to go through the chromosome, accessing only to the variable ones.

5.2 Initialization

One of the most challenging parts of our program is the randomization of the alleles. Each variable represents a different feature so we group them based on their magnitude. Every time the dataset changes, we also modified the groups as we understood better the variables. Even then, the categorization and the range limitations we made are mostly based on what we considered more logical after reviewing the results as we lack medical knowledge. The function that returns a random value based on the variable is called getLimitedRandom and contains the following groups:

1. The value is binary. This group includes **Sleep 8** and **White meat consumption**. Both are yes no questions with no strings attached.
2. The value is binary, it can only be higher or equal than the original. This group only includes **Skimmed consumption**. If the individual takes skimmed products, means that they probably can not eat lactose, so if the original value is yes, that is 1, it should stay that way.

3. The value can only be equal or lower than the original. This group includes **Spirit week**, **Beer week**, **Wine week**, **Pink week**, **Number of smokes**, **Pipe** and **Cigar**. We considered that the values higher than the original could not contribute positively to the fitness because these variables are considered bad habits, so we cut them off. We also added a condition that is if the original value of the yes or no question related to these variables is no, that is 0, these variables' value is automatically set to 0 as well.
4. The value can only be equal or lower than the original. This group includes **Butter consumption**, **Refreshment consumption**, **Industrial bakery consumption** and **Minutes seated** and shares the same logic than the previous group, but without the extra condition.
5. The value can change up to 2, either increasing or decreasing, from the original. This group includes **Olive oil consumption**, **Vegetables consumption**, **Fruits consumption**, **Red meat consumption**, **Diary consumption**, **Fish consumption** and **Nuts consumption**. We considered that it is difficult to suddenly change the diet so we decided on a smaller range in this group.
6. This group includes **EIMS**, **EMMS** and **ECMS**. The values can change up to 30, 120 and 270, respectively, either increasing or decreasing, from the original. The variation increases as the exercises' intensity decreases.

5.3 Evaluation system

In order to implement the evaluation system in our application, we used Apache Commons JEXL to transform the models we had into logical operations. It creates a JexlExpression with a string as an input and evaluates it as a logical operation. The last piece left to finish the parser was a regular expression that replaces the variables of the models with the actual value of our individual.

6 Algorithm design and implementation

The evolutionary algorithm we implemented in our app is the standard version. We chose the simpler approach because we wanted to lessen its burden on the efficiency.

We decided to implement a deterministic tournament ([4]) with a tournament size of 3 as the selection method. We chose the uniform crossover with 80% crossover rate, because each allele of the chromosome represents a different feature of the individual and it does not make sense to mix values from different features. We implemented the basic mutation with 5% mutation rate and we also added 2% of elitism in our algorithm.

Regarding the population size and the number of generations, we recorded the time the program took with different values in both population size and number of generations as shown Fig 5 . We utilized the same seed in these four executions and the same input, 100 individuals and 100 models. the time each execution took, showed in milliseconds, differs but the total improvement are very similar, thus we decided to set the population size as 100 and the number of generations as 100.

A noteworthy detail is that the limitations of hardware affects the time the program takes to execute. The executions made to extract the data were run on the ABSysGroup's server.

We also noticed that the crossover takes around the 48% of the total time and the evaluation takes around 50% of the total time. The reason being that the crossover function has to go over the population and each allele of the chromosome and the evaluation function has to sort the population.

pop100_gen100						
Total Time	Original Fitness total	New Fitness Total	Time/Person	Minutes	Fitness Improv/Person	Total improvement
12182960	3520	5719	121829,6	2,03049333	21,99	2199

pop100_gen150						
Total Time	Original Fitness total	New Fitness Total	Time/Person	Minutes	Fitness Improv/Person	Total improvement
18507911	3520	5729	185079,11	3,08465183	22,09	2209

pop150_gen100						
Total Time	Original Fitness total	New Fitness Total	Time/Person	Minutes	Fitness Improv/Person	Total improvement
18653127	3520	5727	186531,27	3,1088545	22,07	2207

pop150_gen150						
Total Time	Original Fitness total	New Fitness Total	Time/Person	Minutes	Fitness Improv/Person	Total improvement
27707884	3520	5729	277078,84	4,61798067	22,09	2209

Fig. 5: Time and Performance Analysis

7 Results

8 Conclusions

Taking everything into account, the project turn out very interesting. We managed to implement the habit recommender system we had in mind while leveraging evolutionary computation.

This project could pave the way for a myriad of possibilities in the medical field. There are countless options, both for investigation and for practical applications that can take advantage of evolutionary computation.

Some of the limitations we encountered are the following: the data was biased, almost half of the individuals were young people between 18 and 30 years old, due to the lack of medical knowledge on nutrition, the changes we made on the variables were based on common sense and our limited knowledge and our lack of knowledge in the statistic field limited what we could do regarding data collection and display.

We hope this project can create a gateway for others to try and explore other medical fields leveraging evolutionary computation.

9 Improvements and Future Work

There is always room for improvement. In this section we will list some improvements that we have considered but could not make.

9.1 Medical and statistic knowledge

As we said before, the limit and conditions we set on the variables on generating random values are based on common sense and our limited knowledge on nutrition, so it would be nice to have feedback and recommendations from an expert on the field. Because something that seems good to us may be actually something bad, or vice versa.

About statistics, if we knew more, we could have extracted more interesting information from the data we produced and draw better conclusions from them.

9.2 Inclusion of genes

Regarding the dataset we were provided by GenObIA, there are genetic information as well but we did not make use of them. It would be a nice addition to the algorithm because it can add another depth to the variables but it would be optional because the genes are not information everyone has at hand.

The genes being optional variables implies having different set of models trained with the genes variables and a different chromosome as well. There would be a element similar to a checkbox to notify the algorithm which models and chromosome use.

9.3 Friendlier graphical user interface

The GUI we made only has the basic features. It would be great to add more to facilitate the user to fill the form and in general offer a better experience to the user.

A more sequential approach could be taken to the gathering of the data, such as showing different windows (one after the other) for each of the information panels that we have, providing a more interactive experience.

9.4 Mobile phone and web application

As for us, we made the program in Java because we were very familiar with it but it would be great to port the program to mobile or web platforms so it could reach and help more people.

10 Acknowledgments

References

1. Barnett, N.P., Clerkin, E.M., Wood, M., Monti, P.M., O'Leary Tevyaw, T., Corriveau, D., Fingeret, A., Kahler, C.W.: Description and predictors of positive and negative alcohol-related consequences in the first year of college. Journal of Studies on Alcohol and Drugs **75**(1), 103–114 (2014). https://doi.org/10.15288/jsad.2014.75.103
2. Dai, X., Gil, G.F., Reitsma, M.B., Ahmad, N.S., Anderson, J.A., Bisignano, C., Carr, S., Feldman, R., Hay, S.I., He, J., Iannucci, V., Lawlor, H.R., Malloy, M.J., Marczak, L.B., McLaughlin, S.A., Morikawa, L., Mullany, E.C., Nicholson, S.I., O'Connell, E.M., Okereke, C., Sorensen, R.J.D., Whisnant, J., Aravkin, A.Y., Zheng, P., Murray, C.J.L., Gakidou, E.: Health effects associated with smoking: a burden of proof study. Nature Medicine **28**(10), 2045–2055 (2022). https://doi.org/10.1038/s41591-022-01978-x, epub 2022 Oct 10
3. Laing, B.Y., Mangione, C.M., Tseng, C.H., Leng, M., Vaisberg, E., Mahida, M., Bholat, M., Glazier, E., Morisky, D.E., Bell, D.S.: Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Annals of Internal Medicine **161**(10 Suppl), S5–S12 (2014). https://doi.org/10.7326/M13-3005
4. Miller, B.L., Goldberg, D.E.: Genetic algorithms, tournament selection, and the effects of noise. Complex Syst. **9**(3) (1995), `http://www.complex-systems.com/abstracts/v09_i03_a02.html`
5. Ryan, C., O'Neill, M., Collins, J.: Introduction to 20 years of grammatical evolution. Handbook of grammatical evolution pp. 1–21 (2018)
6. World Health Organization, W.: Obesity and overweight (march 2024), `https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight`, accessed: 2024-09-12