A fuzzy inference model for cancer risk prediction

María Camila Vásquez Correa
Mathematics department
Universidad EAFIT
Medellín, Colombia
mvasqu49@eafit.edu.co

Abstract—In this article is presented a model of fuzzy inference in cancer risk, with three variables of support: habits, heritage and age. The model is tested in various scenarios with different methods for logical operations and defuzzification.

Index Terms—Fuzzy, Inference, FIS, cancer, risk.

I. Introduction

Cancer is the leading life-threatening disease for people in today's world. Although cancer formation is different and doctors often cannot explain why one person develops cancer and another does not, research shows that certain risk factors increase the chance that a person will develop cancer [1]. Cancer prevention through consequent screening programs, early discovery and timely, improved and diversified means of treatment are usually the most successful ways to reduce mortality.

Where uncertainty exists such as in the medical field, fuzzy logic could play an important role in making decisions. Fuzzy logic is the science of reasoning, thinking, and inference that recognizes and uses the real world phenomenon that everything is a matter of degree. In the simplest terms, fuzzy logic theory is an extension of binary theory that does not use crisp definitions and distinctions. Instead of assuming everything must be defined crisply into black and white (binary view), fuzzy logic is a method that captures and uses the concept of fuzziness in a computationally effective manner [2].

The objective of this work is to design and implement a fuzzy inference model and to analyze its results in predicting the risk of cancer in several cases. We suppose then that three variables are somehow enough to describe a possibility of developing cancer and we test the model in different scenarios.

The rest of the paper is organized as follows: first, there is a review of the current work regarding the topic. Second, there is an explanation of the methodology that is going to be used. Then the results of the model are presented and finally some conclusions are derived from the work.

II. Literature review

Many authors have addressed the problem of an early diagnose of cancer using fuzzy models that incorporate characteristics of the patients. And although these models share a common core, they are very different and approach the problem from different perspectives.

Most of the existing models try to model the risk of developing cancer in a specific organ. Such models include: breast cancer ([3], [4], [5], [6], [7], [8], [9], [10], [11], [12]), prostate cancer ([13], [14]), oral cancer ([15]), gastric cancer ([16]), colon cancer ([17]) and lung cancer ([18]). However, there exists some other models that try to asses the risk of developing any kind of cancer, such as [19] and [20].

They vary also in the type of fuzzy model that is being used. Most of them are a fuzzy inference model of type Mandani ([3], [20], [17], [5], [6], [7], [16], [12]) or Sugeno ([19], [9]). Though neural fuzzy models ([13], [15], [14], [18]), fuzzy cognitive maps ([4], [8], [10]) and fuzzy probabilistic models ([11]) have been used in this field.

The way of modelling the characteristics of the patients is very similar in some models. They include hereditary characteristics, like the genetic history; lifestyle characteristics, such as the nutrition, the smoking habits, the body mass index and physical activity; biological factors like hormones, pre-existing tumors and cell conditions. Finally, they consider socio-demographic factors, such as age and gender.

In this work we aim to use three of the main groups of characteristics, socio-demographic, lifestyle and hereditary, to propose a Mandani fuzzy inference model for predicting the risk of developing any type of cancer.

III. Methodology

The model used to carry out this work is described in detail below:

A. Variables

- u_1: Quality of the habits (lifestyle).
- u_2: Quality of the genetic heritage. A good genetic heritage implies rare or non history of cancer.
- u_3: Age of the person.
- y: Risk of developing cancer.

B. Speech Universe

- $u_1 \in \mathbb{Z} \cap [0, 10]$: Being 0 the worst habits and 10 the best.
- $u_2 \in \mathbb{Z} \cap [0, 10]$: Being 0 the worst heritage and 10 the best.
- $u_3 \in \mathbb{Z} \cap [0, 90]$.
- $y \in \mathbb{R} \cap [0, 100]$.

C. Linguistic categories

- u_1: Poor, medium and high quality.
- u_2: Poor, medium and high quality.
- u_3: Young, medium and old.
- y: High, medium and low.

D. Membership functions

In the Figure 1 are shown the respective membership functions for each one of the linguistic categories. They are defined as:

- u_1:
  - *Poor*: $trapezoid(x; 0, 0, 2, 4)$
  - *Medium*: $gauss(x; 5, 1)$
  - *High*: $trapezoid(x; 6, 8, 10, 10)$
- u_2:
  - *Poor*: $sigmoid(x; 3, -2.64)$
  - *Medium*: $gauss(x; 5, 1)$
  - *Low*: $sigmoid(x; 7, 2.64)$
- u_3:
  - *Poor*: $sigmoid(x; 30, -0.53)$
  - *Medium*: $gauss(x; 50, 15.\bar{33})$
  - *Low*: $sigmoid(x; 70, 0.264)$
- y:
  - *Poor*: $sigmoid(x; 30, -0.264)$
  - *Medium*: $gauss(x; 50, 10)$
  - *Low*: $sigmoid(x; 70, 0.264)$

E. Rule base

- If [u_1 Poor] and [u_2 Poor] and [u_3 Low] then [y Medium]
- If [u_1 Poor] and [u_2 Poor] and [u_3 Medium] then [y Medium]
- If [u_1 Poor] and [u_2 Poor] and [u_3 High] then [y High]
- If [u_1 Poor] and [u_2 Medium] and [u_3 Low] then [y Low]
- If [u_1 Poor] and [u_2 Medium] and [u_3 Medium] then [y Medium]
- If [u_1 Poor] and [u_2 Medium] and [u_3 High] then [y High]
- If [u_1 Poor] and [u_2 High] and [u_3 Low] then [y Low]
- If [u_1 Poor] and [u_2 High] and [u_3 Medium] then [y Low]
- If [u_1 Poor] and [u_2 High] and [u_3 High] then [y Medium]
- If [u_1 Medium] and [u_2 Poor] and [u_3 Low] then [y Medium]
- If [u_1 Medium] and [u_2 Poor] and [u_3 Medium] then [y High]
- If [u_1 Medium] and [u_2 Poor] and [u_3 High] then [y High]
- If [u_1 Medium] and [u_2 Medium] and [u_3 Low] then [y Low]
- If [u_1 Medium] and [u_2 Medium] and [u_3 Medium] then [y Medium]
- If [u_1 Medium] and [u_2 Medium] and [u_3 High] then [y High]
- If [u_1 Medium] and [u_2 High] and [u_3 Low] then [y Low]
- If [u_1 Medium] and [u_2 High] and [u_3 Medium] then [y Low]
- If [u_1 Medium] and [u_2 High] and [u_3 High] then [y Medium]
- If [u_1 High] and [u_2 Poor] and [u_3 Low] then [y Medium]
- If [u_1 High] and [u_2 Poor] and [u_3 Medium] then [y Medium]
- If [u_1 High] and [u_2 Poor] and [u_3 High] then [y Medium]
- If [u_1 High] and [u_2 Medium] and [u_3 Low] then [y Low]
- If [u_1 High] and [u_2 Medium] and [u_3 Medium] then [y Low]
- If [u_1 High] and [u_2 Medium] and [u_3 High] then [y Medium]
- If [u_1 High] and [u_2 High] and [u_3 Low] then [y Low]
- If [u_1 High] and [u_2 High] and [u_3 Medium] then [y Low]
- If [u_1 High] and [u_2 High] and [u_3 High] then [y Low]

IV. Results

The rule base was implemented using a Mamdani and the aggregation method max min. Additionally, three methods of deffuzification methods (Centroid of the area, bisector of the area and mean of maximum) and two T-norms used for the activation of the rules (Minimum and product) with their respective S-rule (Maximum and sum - product). The implementation was made on Python using the module scikit-fuzzy.

A. Response surfaces

In this section, are described and shown some of the response surfaces of the model. In Figure 2 are shown 9 response surfaces, each of them corresponding to the variation of two input variables and the fixed value of one of them, this last one being low, medium and high.

In Figures 2a-2c are shown the results for the risk considering the heritage and the age, with fixed habits (low, medium and high). These three surfaces show that really good habits tend to have lower general risk, also depending on the age it goes higher. Low habits tend to have the highest risk, as expected. However, for medium habits, the slope changes compared to the low habits, giving more importance to the heritage and the age, and we can see that, good heritage gives usually a low risk, as well as a low age, as expected. And with high habits, the surface is completely different, given the fact that the model considers the three variables as equally important and good habits can make the risk go down.

In Figures 2d-2f are shown the surfaces for the risk of cancer considering the heritage and the habits, with fixed age (young, medium and old). Here we can see several
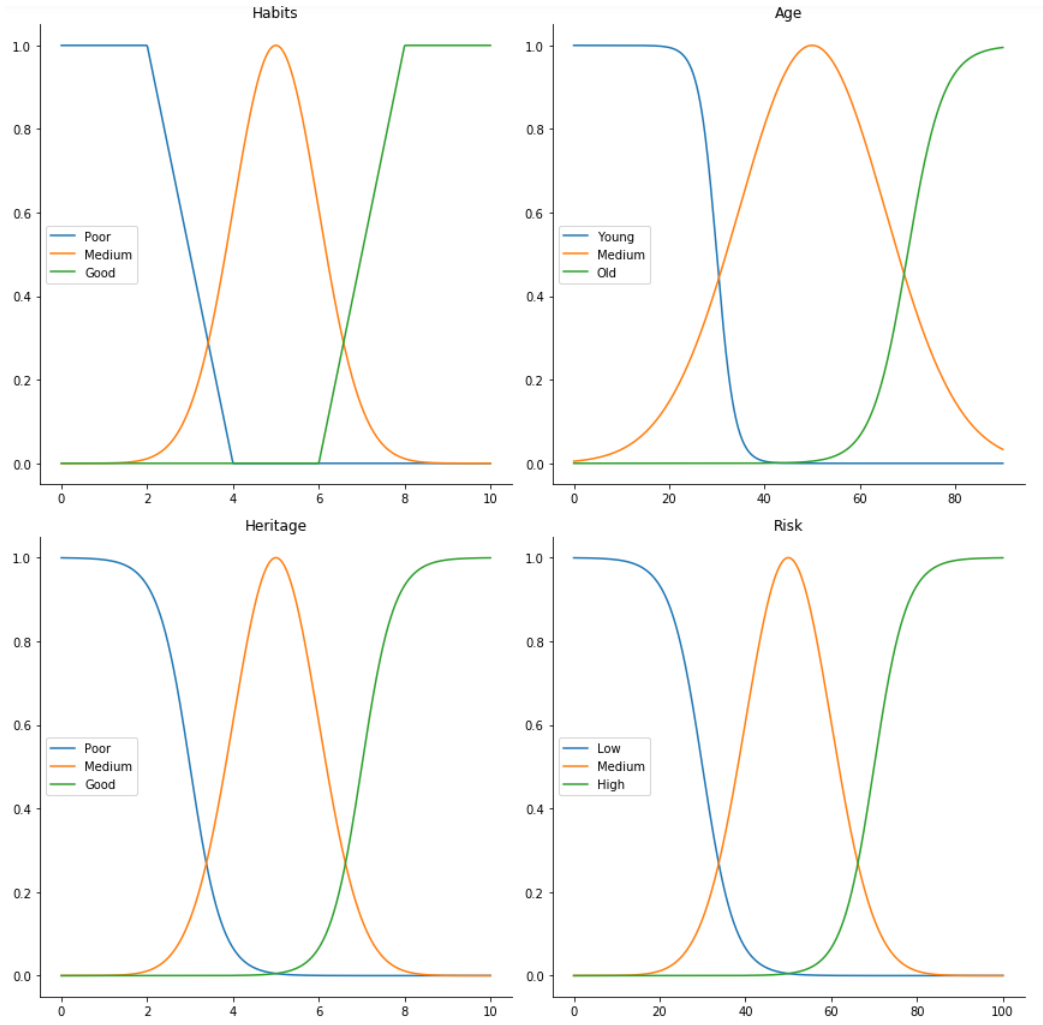
Fig. 1: Membership functions

variations of the surfaces, and the final risk reaches a higher value considering the age. In the young person, the risk is never above 50, in the medium age the risk goes up to 60 in some configurations and in the old age the risk goes the highest, with a maximum value of 90. This actually makes sense, one would expect to have a general higher risk for a higher age, considering the aging facts affecting the body. Also, the heritage and habits can change drastically this risk, shown in the peaks of the surfaces.

In Figures 2g-2i are shown the surfaces for the risk of cancer considering the age and the habits, with fixed heritage (low, medium and high). Two of these three surfaces do not have a smooth transition between stages, probably due to a rule that produces an abrupt change. The risk in general is not very high, as seen in the first triplet of surfaces. However, it changes among surfaces, and we see that, for a low heritage, the risk is generally higher than in the other two cases. This is an expected result, considering the fact that history of cancer in one's family can lead to the people developing cancer with

the same characteristics.

B. Test cases

Then, three test cases were evaluated with each one of the T-norms and the defuzzification methods:

- A healthy person, medium age, with good heritage. (habits = 7, age = 20, heritage = 8)
- A non-healthy young person with poor heritage. (habits = 4 age = 10 heritage = 3)
- A healthy old person with poor heritage. (habits = 10 age = 75 heritage = 3)

1) Test case 1: In the Figures 3 and 4 are shown the results for the first case, an ideal case, in which a person is not expected to develop cancer soon. The final risk of cancer is 15, 18 or 19, depending on the defuzzification method and the norm used. However, these values are not far apart from each other and indicate that the person is not in immediate danger, a reasonable conclusion taking into account the fact that they have healthy habits and a good genetic heritage. We see that, in this case, the minimum norm produces a higher output than
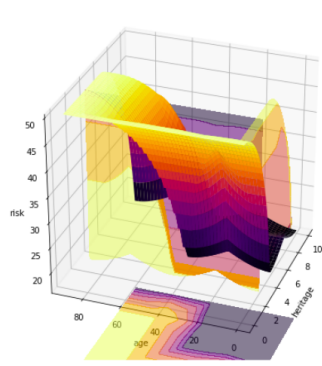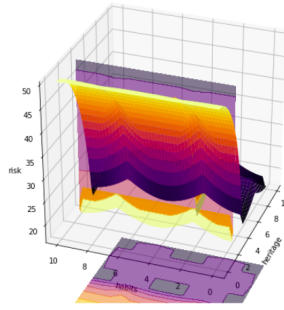
(a) Surface for the risk, considering the heritage and the age, with high habits (10).
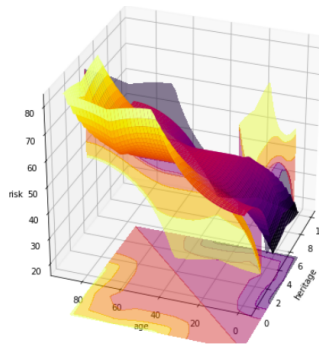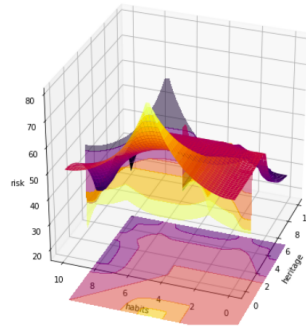
(b) Surface for the risk, considering the heritage and the age, with medium habits (5).
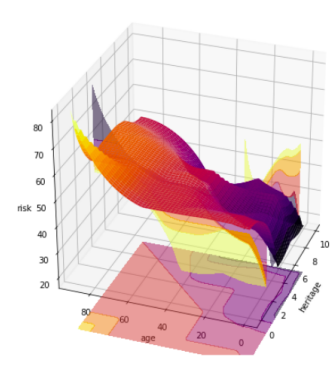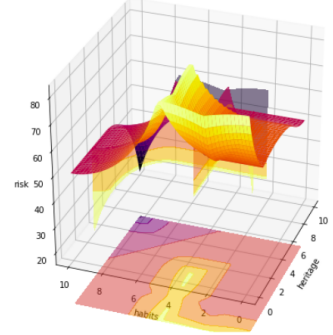
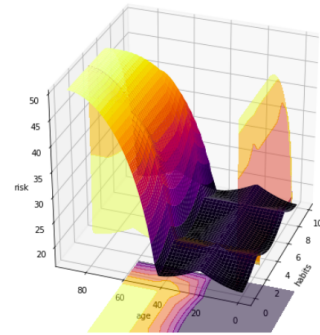(c) Surface for the risk, considering the heritage and the age, with low habits (1).

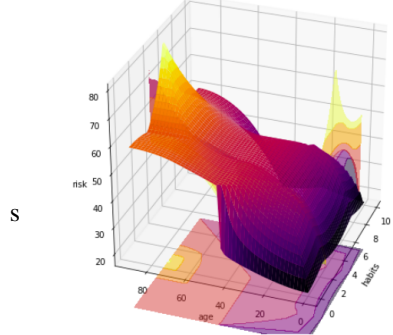(d) Surface for the risk, considering the heritage and the habits, with a low age (4).

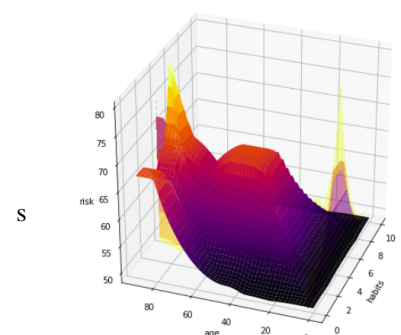(e) Surface for the risk, considering the heritage and the habits, with a medium age (50).

(f) Surface for the risk, considering the heritage and the habits, with a high age (90).

(g) Surface for the risk, considering the age and the habits, with a high heritage (10).

(h) Surface for the risk, considering the age and the habits, with a medium heritage (5).

(i) Surface for the risk, considering the age and the habits, with a low heritage (1).

Fig. 2: Response surfaces for the risk of developing cancer.

the product norm, and the mean of maximum defuzzification method always produces the lowest output for the risk. The results match the original assumption of a healthy person in no immediate danger holds.

2) Test case 2: In the Figures 5 and 6 are shown the results for the second test case, in which a person is really young but has a bad heritage, that is, maybe they have some history of cancer in the family. Also, they do not have really good habits. We can see that the model takes all into account and for different rules has medium, low or high risk, each one for their habits, age and genetic heritage. The final risk is 45, 47

or 50, depending on the defuzzification method, but they do not vary with the norm used for the rule base. In this case, the mean of maximum produces the highest output of the risk. This is a case in which we identify that, even though the habits and the heritage have a bad quality, the fact that the patient is really young makes the risk medium. One can argue that the intuitive probability of this happening is really low, but cases show that really young people can develop cancer due to genetic factors.

3) Test case 3: In the Figures 7 and 8 are shown the results for the final test case. This patient has an advanced

age, however, their habits are really good, and even when their heritage is not the best, the risk of developing cancer is medium, varying between 44 and 50. In this case, the mean of maximum defuzzification produces again the highest output, and the product norm is higher than the minimum norm. Here, we identify the fact that, for the model, the habits have the same importance as the heritage, and can compensate a high age and a bad heritage. Also, the risk is medium, as one will expect from an old person having really healthy habits.

V. CONCLUSIONS

The constructed model for predicting the risk of developing cancer in one person considering their age, quality of their habits and quality of the genetic heritage has really interesting results. The fact that the risk can ignore something like a heritage of cancer just by considering good habits, is a sign that the three of the variables are equally important, and one can compensate another. These good habits can also include: exercise, good nutrition, good sleep cycle and frequent doctor visits, that we will expect to have a really good influence in our health, and in our risk of developing cancer. However, the risk is not low or none, but one can expect to be healthy when old and having good habits.

Other test cases shown that our assumptions for young people and people that are not that young but have really bad habits and heritage match the results of the model. The risk does not imply an immediate sickness for people of the second kind, nor zero risk for the first kind, but it reflects what we would expect to be true. Also, in the surfaces we observed that the age is the most important factor for the model, and is also what we observe generally in the real world.

So, is the model good enough to tell someone somehow accurately which can be their risk of developing cancer? Probably not. The factors considered in this model are really general, and do not consider the fact that, for example, some conditions of the habits, like smoking or medical check, can have a bigger influence than, for example, sleep cycle. Also, they do not consider some biological factors, like blood pressure, that can have an impact in the risk. The heritage is also really wide, and does not consider the degree of consanguinity with the people with history of cancer. Finally, the model predicts a general risk, but can be refined by adding an specific organ in which one can develop cancer, due to the fact that some factors (like smoking, for lung cancer) can be more determinant in one or other part of the body.

REFERENCES

[1] D. Coggon and C. Martyn, "Time and chance: the stochastic nature of disease causation," *The Lancet*, vol. 365, no. 9468, pp. 1434 – 1437, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140673605663805

[2] L. A. Zadeh, "Fuzzy sets and information granularity," *Advances in Fuzzy Systems-Applications and Theory*, vol. 26, no. 4, pp. 551–556, 1996.

[3] V. Balanică, I. Dumitrache, M. Caramihai, W. Rae, and C. Herbst, "Evaluation of breast cancer risk by using fuzzy logic," *UPB Scientific Bulletin, Series C: Electrical Engineering*, vol. 73, no. 1, pp. 53–64, 2011.

[4] A. Büyükavcu, Y. E. Albayrak, and N. Göker, "A fuzzy information-based approach for breast cancer risk factors assessment," *Applied Soft Computing Journal*, vol. 38, pp. 437–452, 2016.

[5] B. M. Gayathri and C. P. Sumathi, "Mamdani fuzzy inference system for breast cancer risk detection," *2015 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2015*, 2016.

[6] R. Khezri, R. Hosseini, and M. Mazinani, "A fuzzy rule-based expert system for the prognosis of the risk of development of the breast cancer," *International Journal of Engineering, Transactions A: Basics*, vol. 27, no. 10, pp. 1557–1564, 2014.

[7] K. Latha, "Visualization of Risk in Breast Cancer Using Fuzzy Logic in MATLAB Environment," *International Journal of Computational Intelligence Techniques*, vol. 4, no. 1, pp. 114–117, 2013.

[8] E. I. Papageorgiou, J. Subramanian, A. Karmegam, and N. Papandrianos, "A risk management model for familial breast cancer: A new application using Fuzzy Cognitive Map method," *Computer Methods and Programs in Biomedicine*, vol. 122, no. 2, pp. 123–135, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.cmpb.2015.07.003

[9] A. A. Shleeg and I. M. Ellabib, "Comparison of Mamdani and Sugeno Fuzzy Interference Systems for the Breast Cancer Risk," *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 7, no. 10, pp. 1343–1347, 2013.

[10] J. Subramanian, A. Karmegam, E. Papageorgiou, N. Papandrianos, and A. Vasukie, "An integrated breast cancer risk assessment and management model based on fuzzy cognitive maps," *Computer Methods and Programs in Biomedicine*, vol. 118, no. 3, pp. 280–297, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.cmpb.2015.01.001

[11] F. Tatari, M. R. Akbarzadeh-T, and A. Sabahi, "Fuzzy-probabilistic multi agent system for breast cancer risk assessment and insurance premium assignment," *Journal of Biomedical Informatics*, vol. 45, no. 6, pp. 1021–1034, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.jbi.2012.05.004

[12] A. Yilmaz and K. Ayan, "Risk analysis in breast cancer disease by using fuzzy logic and effects of stress level on cancer risk," *Scientific Research and Essays*, vol. 6, no. 24, pp. 5179–5191, 2011.

[13] L. Benecchi, "Neuro-fuzzy system for prostate cancer diagnosis," *Urology*, vol. 68, no. 2, pp. 357–361, 2006.

[14] G. Cosma, G. Acampora, D. Brown, R. C. Rees, M. Khan, and A. G. Pockley, "Prediction of pathological stage in patients with prostate cancer: A neuro-fuzzy model," *PLoS ONE*, vol. 11, no. 6, pp. 1–27, 2016.

[15] R. M. Dom, B. Abidin, S. A. Kareem, S. M. Ismail, and N. M. Daud, "Determining the critical success factors of oral cancer susceptibility prediction in Malaysia using fuzzy models," *Sains Malaysiana*, vol. 41, no. 5, pp. 633–640, 2012.

[16] R. Safdari, H. K. Arpanahi, M. Langarizadeh, M. Ghazisaiedi, H. Dargahi, and K. Zendehdel, "Design a fuzzy rule-based expert system to aid earlier diagnosis of gastric cancer," *Acta Informatica Medica*, vol. 26, no. 1, pp. 19–23, 2018.

[17] R. M. Brand, D. D. Jones, H. T. Lynch, R. E. Brand, D. Watson, R. Ashwathnayaran, and H. K. Roy, "Risk colon cancer in hereditary non-polyposis colorectal cancer patients as predicted by fuzzy modeling: Influence of smoking," *World Journal of Gastroenterology*, vol. 12, no. 28, pp. 4485–4491, 2006.

[18] A. Yılmaz, S. Arı, and Ü. Kocabıçak, "Risk analysis of lung cancer and effects of stress level on cancer risk through neuro-fuzzy model," *Computer Methods and Programs in Biomedicine*, vol. 137, pp. 35–46, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.cmpb.2016.09.002

[19] Y. Atınccedil and A. Kuuml rşat, "Comparison with Sugeno Model and Measurement Of Cancer Risk Analysis By New Fuzzy Logic Approach," *African Journal of Biotechnology*, vol. 11, no. 4, pp. 979–991, 2012.

[20] G. Dudek, A. Strzelewicz, M. Krasowska, A. Rybak, and R. Turczyn, "Fuzzy analysis of the cancer risk factor," *Acta Physica Polonica B*, vol. 43, no. 5, pp. 947–959, 2012.

[21] L. A. Zadeh, "Probability measures of fuzzy events," pp. 421–427, 1968.

[22] L. A. Zadeh, "Fuzzy sets," p. 4, 1965.

[23] O. L. Quintero, *Machine Intelligence for Human Decision Making*. Universidad EAFIT, November 2018.

[24] E. Kentel and M. M. Aral, "Probabilistic-fuzzy health risk modeling," *Stochastic Environmental Research and Risk Assessment*, vol. 18, no. 5, pp. 324–338, 2004.
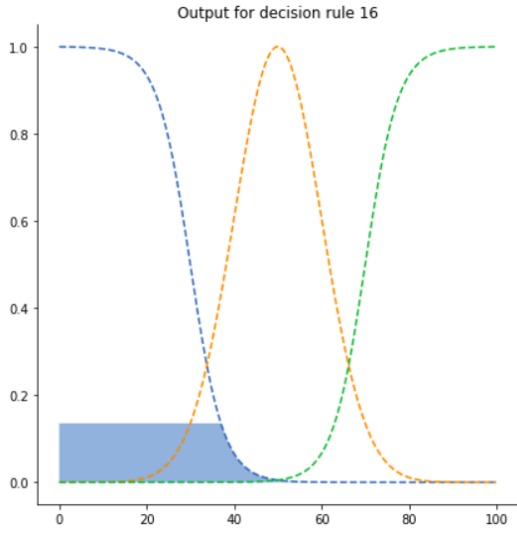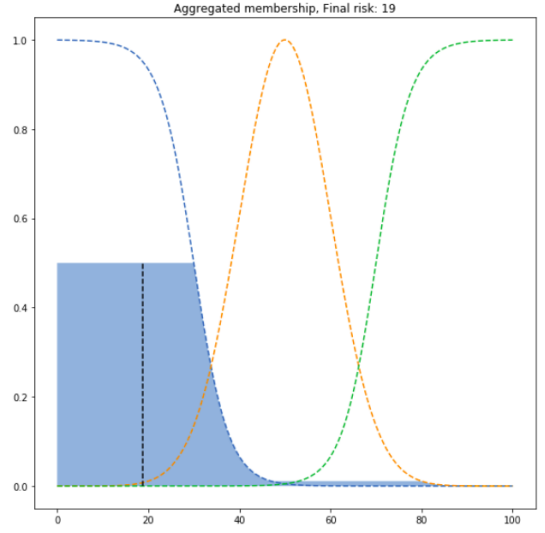
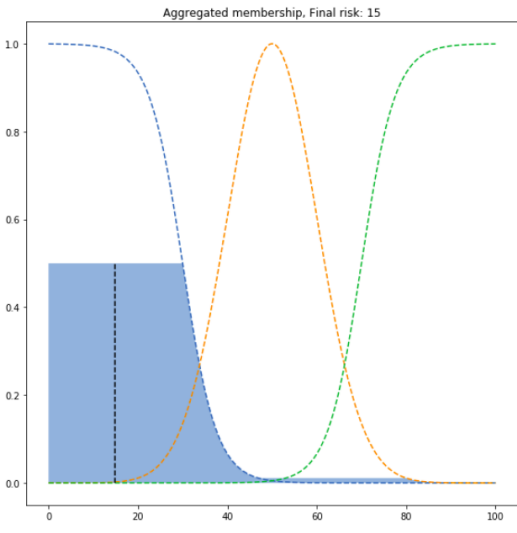(a) Output for decision rule 16 with the T-norm minimum.
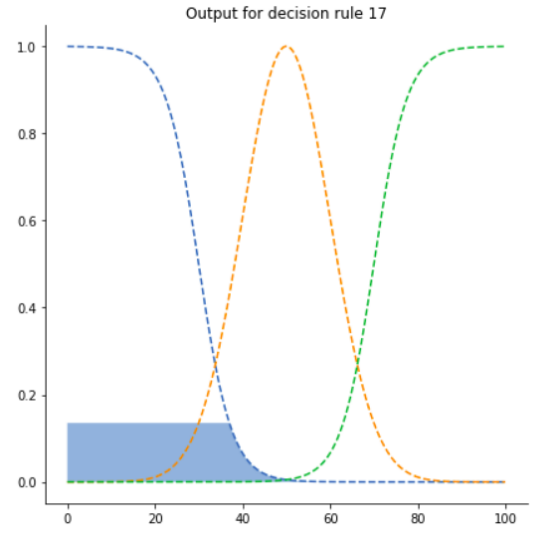
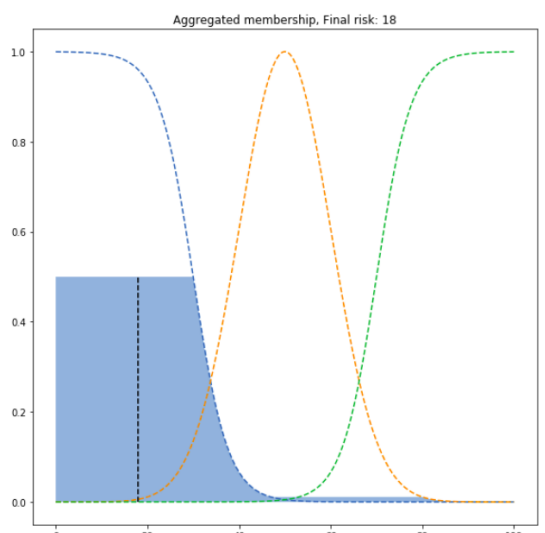(b) Output for decision rule 17 with the T-norm minimum.
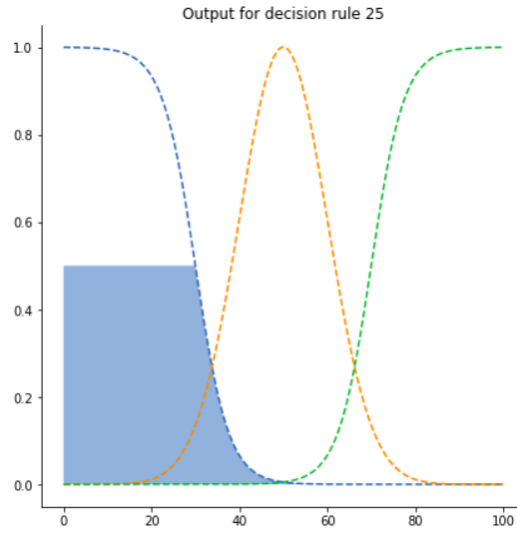
(c) Output for decision rule 25 with the T-norm minimum.

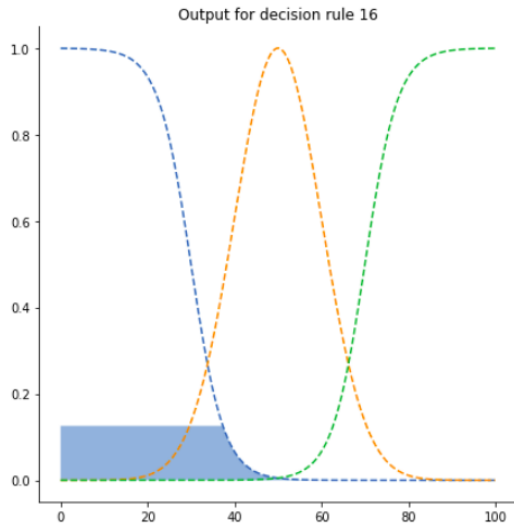(d) Final aggregation result with the centroid method for defuzzification.

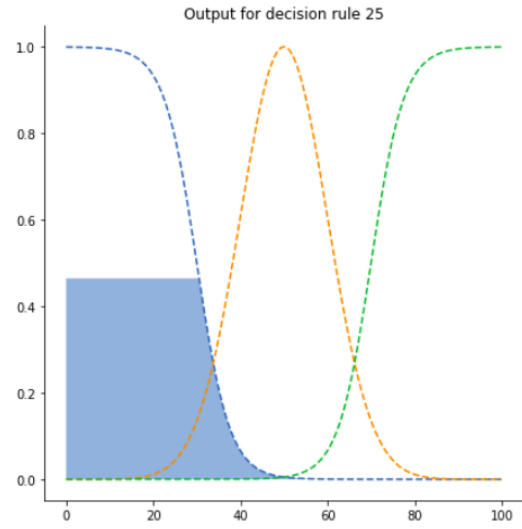(e) Final aggregation result with the mean of maximum method for defuzzification.

(f) Final aggregation result with the bisector method for defuzzification.
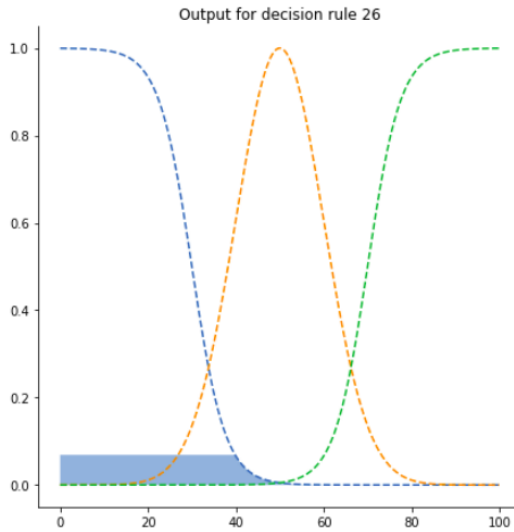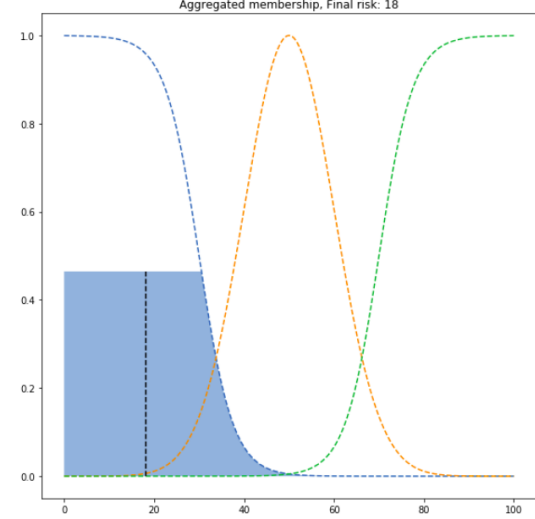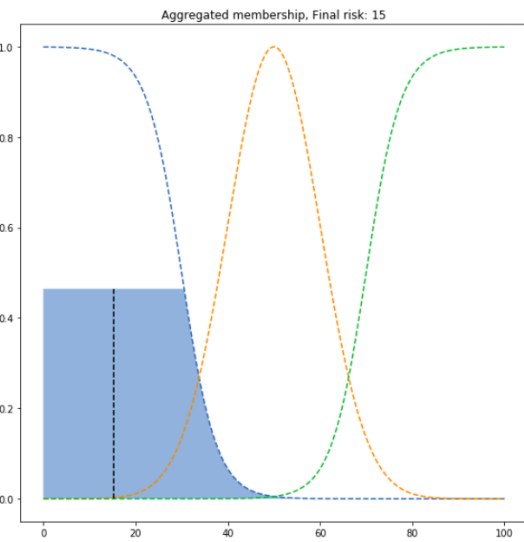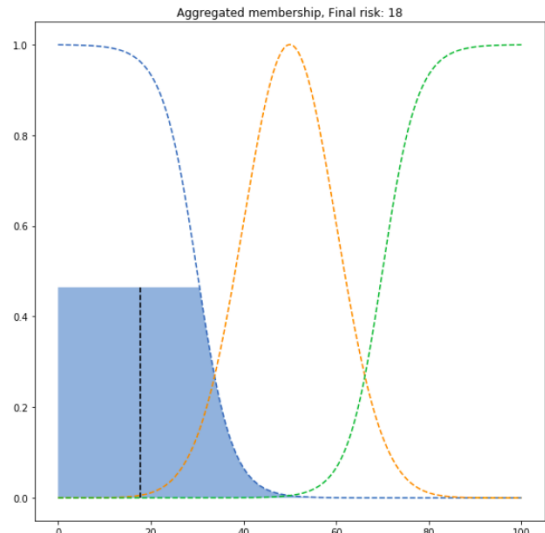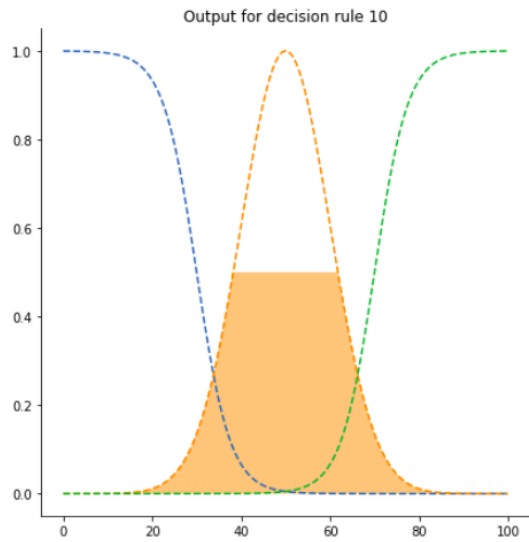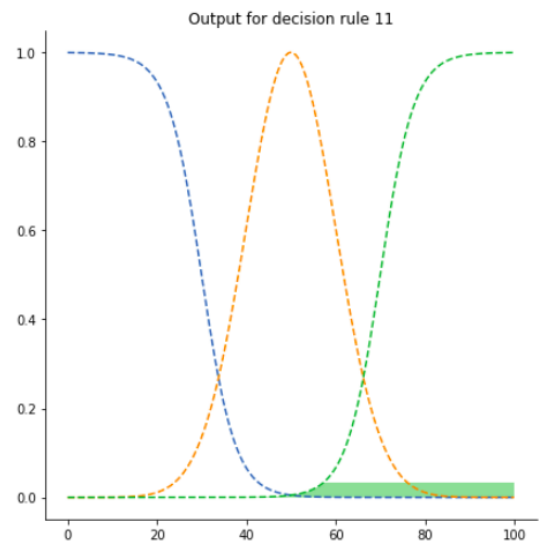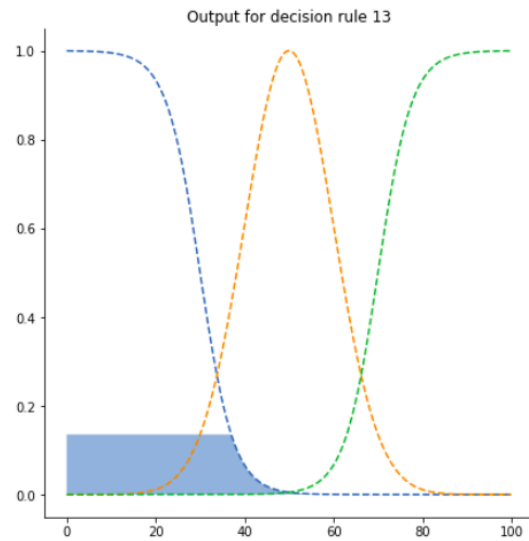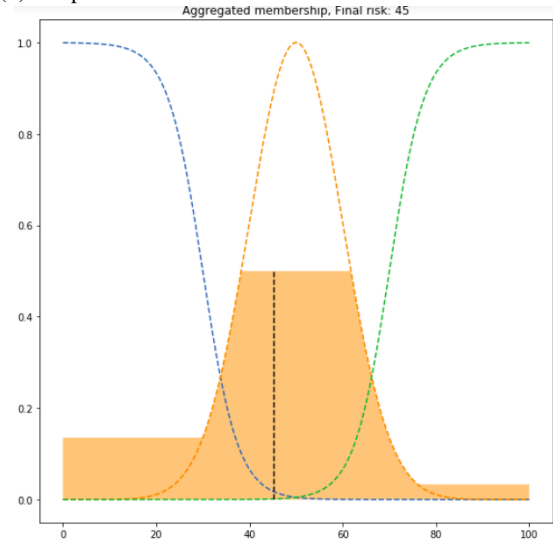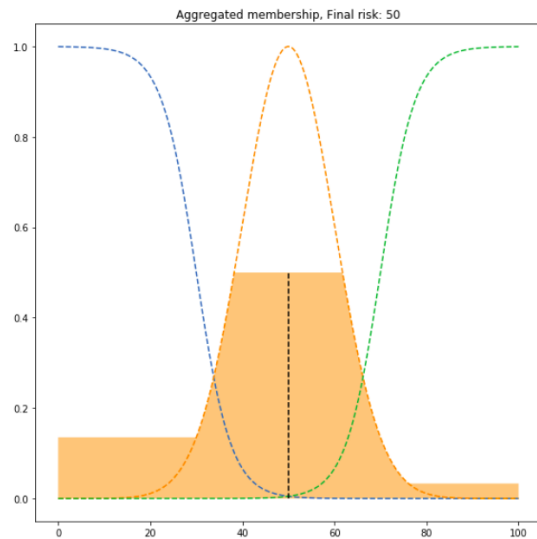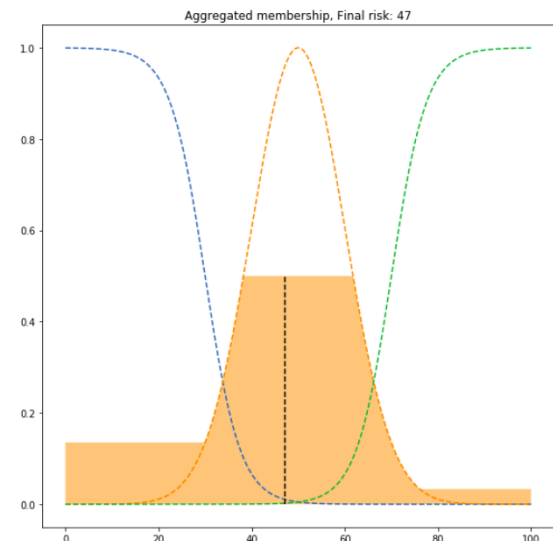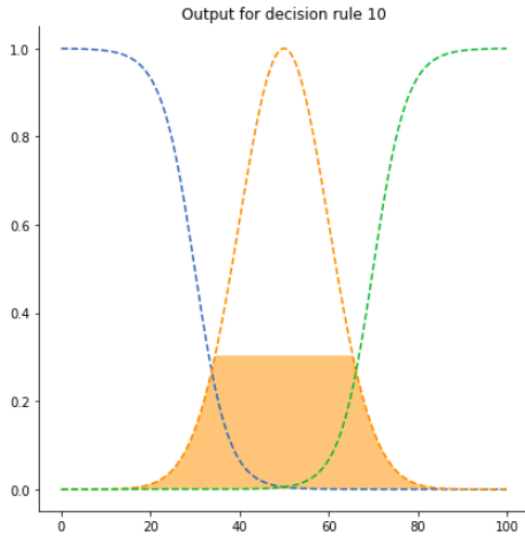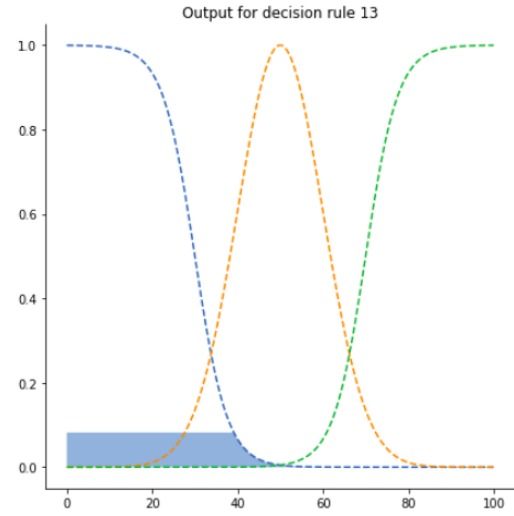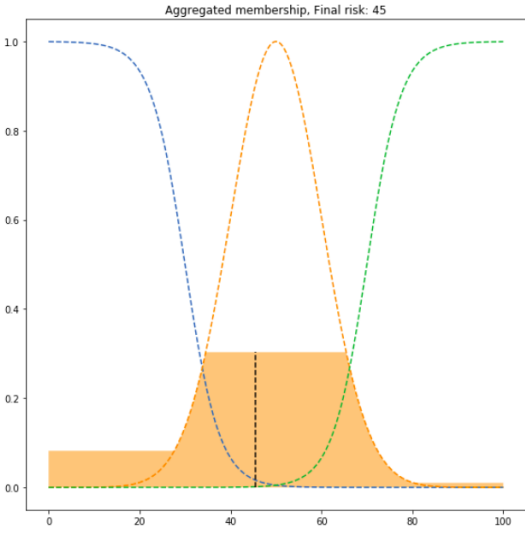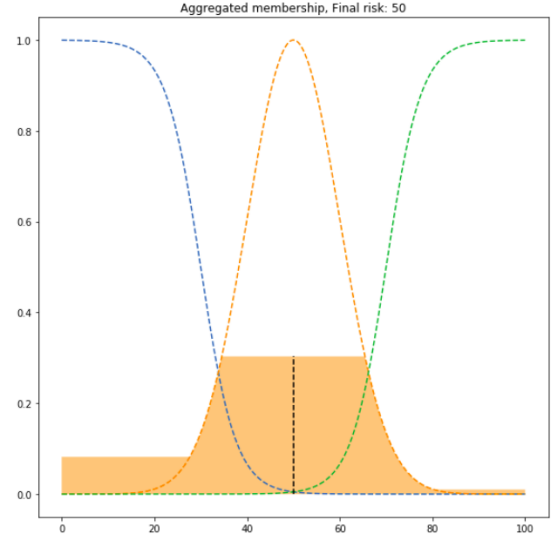
Fig. 3: Results for test case 1 using the T-norm minimum.

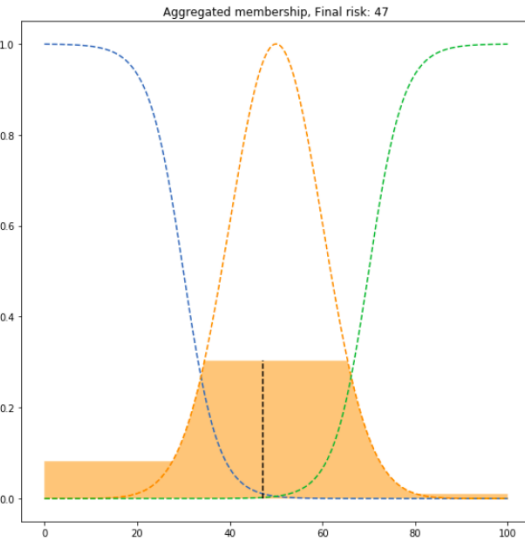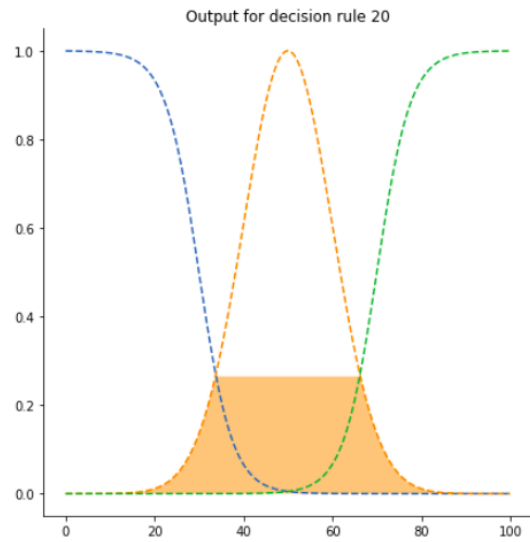(a) Output for decision rule 16 with the T-norm product.

(b) Output for decision rule 17 with the T-norm product.

(c) Output for decision rule 25 with the T-norm product.

(d) Final aggregation result with the centroid method for defuzzification.

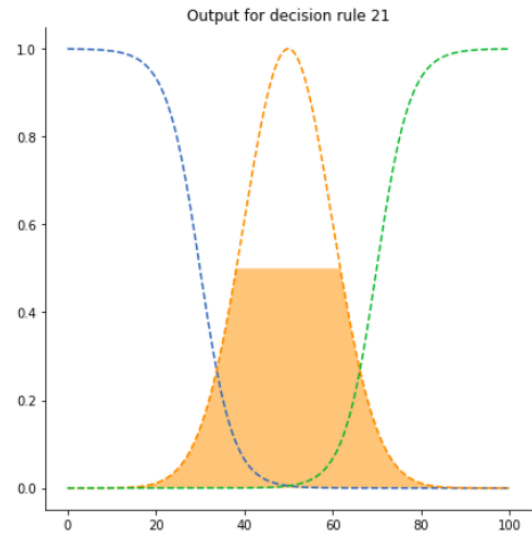(e) Final aggregation result with the mean of maximum method for defuzzification.

(f) Final aggregation result with the bisector method for defuzzification.
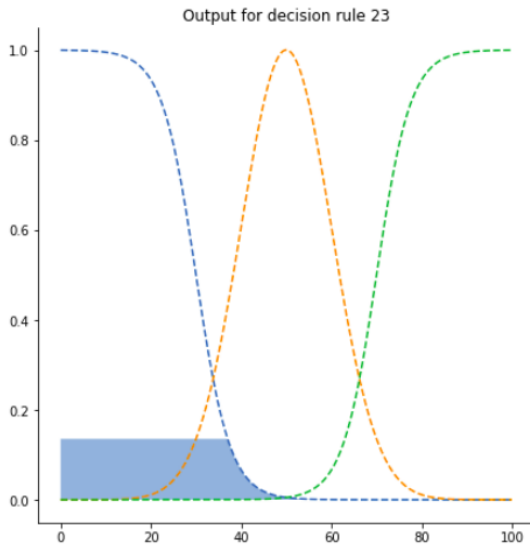
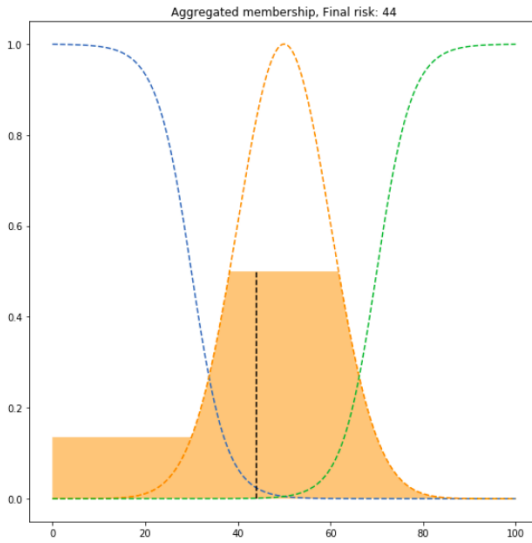Fig. 4: Results for test case 1 using the T-norm product.

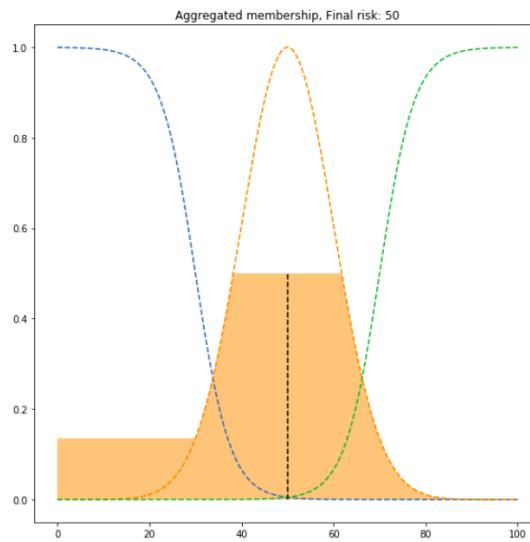(a) Output for decision rule 10 with the T-norm minimum.

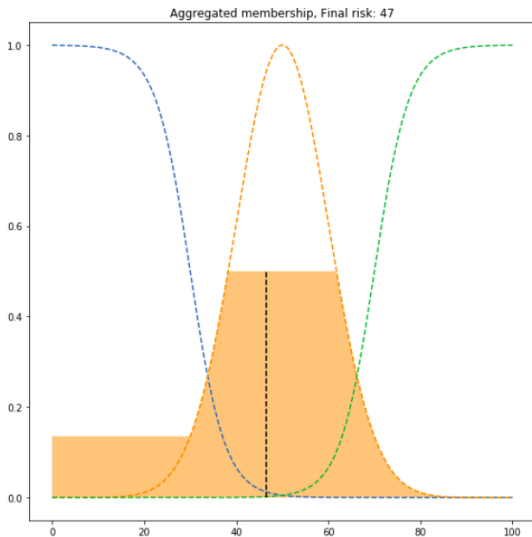(b) Output for decision rule 11 with the T-norm minimum.

(c) Output for decision rule 13 with the T-norm minimum.

(d) Final aggregation result with the centroid method for defuzzification.

(e) Final aggregation result with the mean of maximum method for defuzzification.

(f) Final aggregation result with the bisector method for defuzzification.

Fig. 5: Results for test case 2 using the T-norm minimum.

(a) Output for decision rule 10 with the T-norm product.

(b) Output for decision rule 13 with the T-norm product.

(c) Final aggregation result with the centroid method for defuzzification.

(d) Final aggregation result with the mean of maximum method for defuzzification.

(e) Final aggregation result with the bisector method for defuzzification.

Fig. 6: Results for test case 2 using the T-norm product.

(a) Output for decision rule 20 with the T-norm minimum.

(b) Output for decision rule 21 with the T-norm minimum.

(c) Output for decision rule 23 with the T-norm minimum.

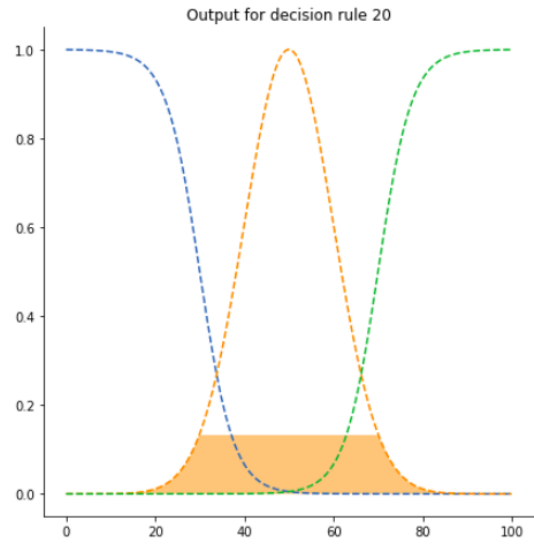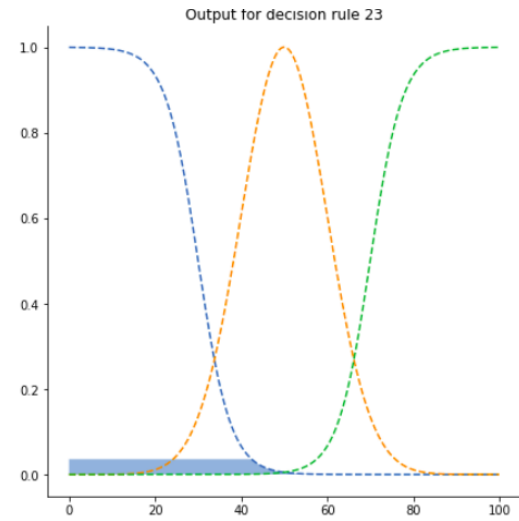(d) Final aggregation result with the centroid method for defuzzification.

(e) Final aggregation result with the mean of maximum method for defuzzification.

(f) Final aggregation result with the bisector method for defuzzification.
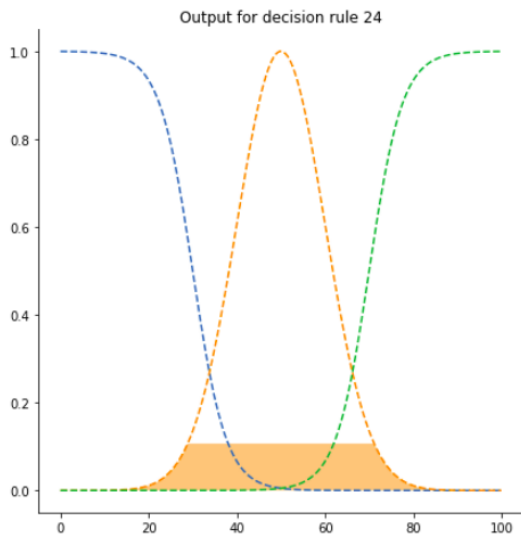
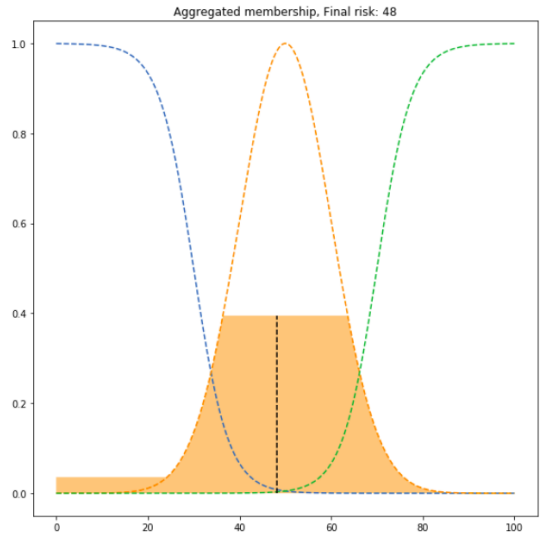Fig. 7: Results for test case 3 using the T-norm minimum.

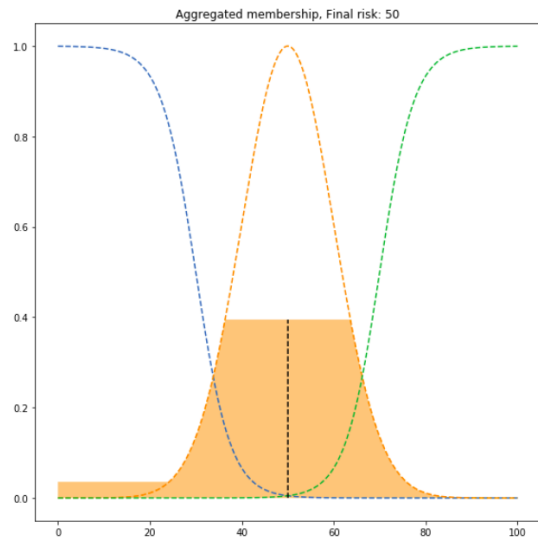(a) Output for decision rule 20 with the T-norm product.

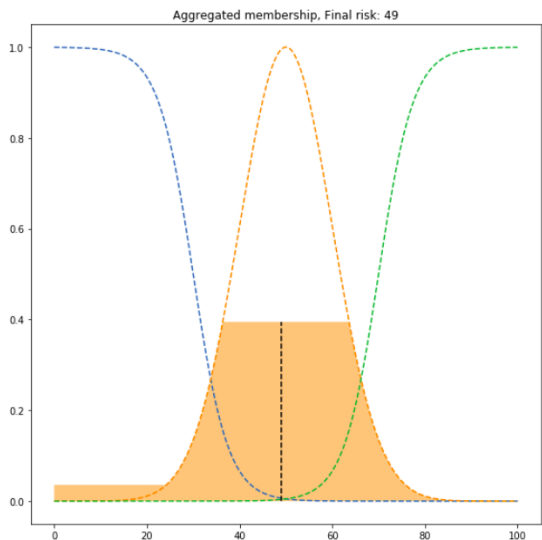(b) Output for decision rule 23 with the T-norm product.

(c) Output for decision rule 24 with the T-norm product.

(d) Final aggregation result with the centroid method for defuzzification.

(e) Final aggregation result with the mean of maximum method for defuzzification.

(f) Final aggregation result with the bisector method for defuzzification.

Fig. 8: Results for test case 3 using the T-norm product.