A General Framework for Context-Aware Fuzzification of Four Ordered Categories: A Case Study on BMI Categories

ALIJANI Zahra and DAŇKOVÁ Martina

Institute for Research and Applications of Fuzzy Modeling,
University of Ostrava, Centre of Excellence IT4Innovations,
30. dubna 22, 70200 Ostrava, Czech Republic
E-mail: Zahra.Alijani@osu.cz, Martina.Dankova@osu.cz

Abstract

This paper presents a general methodological framework for constructing context-aware fuzzy partitions that extend conventional crisp categorizations. The approach is based on Novák's theory of fuzzy contexts and is implemented using the R package `lfl`. It enables smooth and interpretable transitions between adjacent classes while preserving the original categorical structure. To illustrate the procedure, we apply it to derive fitness-specific fuzzy partitions of Body Mass Index, where the conventional four categories (underweight, normal weight, overweight, obese) are adapted according to individual levels of cardiorespiratory fitness.

1 Methodological Framework

We propose a **methodological, two-stage procedure** for transforming a crisp four-class categorization into a fuzzy partition using Novák's *context-based approach*, implemented in the R language package `lfl`. The framework is general and can be applied to any ordered system of four crisp categories. For illustration, we use the conventional Body Mass Index (BMI) classes (*underweight*, *normal*, *overweight*, *obese*) and demonstrate how their boundaries can be contextually adapted based on cardiorespiratory fitness (CRF) derived from $\dot{V}O_{2\,\max}$. In more detail, we propose a data-driven methodology for defining the context of the BMI linguistic terms using Novák's theory [1] of trichotomous evaluative linguistic expressions (CTX3) that are *explicitly conditioned* on cardiorespiratory fitness (CRF) expressed by $\dot{V}O_{2\,\max}$.

Novelty. To the best of our knowledge, no previous work has learned *fitness-conditioned* fuzzy BMI partitions with VO_2-based fuzzy weights merely by modifying the given context center.

To ensure comparability of the conditioning variable across individuals, we first define a normalized, dimensionless fitness index v derived from the maximal oxygen uptake $\dot{V}O_{2\,\max}$:

$$v = \frac{\dot{V}O_{2\,\max} - \dot{V}O_{2\,\max}^{\min}}{\dot{V}O_{2\,\max}^{\max} - \dot{V}O_{2\,\max}^{\min}}, \qquad v \in [0,1],$$

where $\dot{V}O_{2\,\max}^{\min}$ and $\dot{V}O_{2\,\max}^{\max}$ denote the minimal and maximal observed (or expected) values of $\dot{V}O_{2\,\max}$ in the population. This index could later be refined to account for age and sex effects.

Acknowledgement The contribution has been funded from the project "Research of Excellence on Digital Technologies and Wellbeing CZ.02.01.01/00/22_008/0004583", which is co-financed by the European Union.

https://www.doi.org/10.15452/978-80-7599-515-5.2026.02

- $v = 0$ corresponds to very low cardiorespiratory fitness,
- $v = 1$ corresponds to excellent cardiorespiratory fitness.

This normalization establishes a *context variable* that allows for fuzzy modification of the originally crisp category boundaries. In the second stage, each crisp boundary midpoint c is replaced by a *context-dependent center* $c'(v)$, producing a $\dot{V}O_{2\,\max}$-driven fuzzy partition of the BMI domain:

$$c'(v) = c + \alpha\,(v - 0.5),$$

where α is a sensitivity parameter that controls how strongly the boundary shifts with fitness. Typical values of α range between 4 and 6, although its calibration should rely on empirical data or expert domain knowledge.

For instance, with $\alpha = 6$ and an original threshold $c = 25$, we obtain

$$c'(0) = 22, \qquad c'(1) = 28,$$

which is consistent with epidemiological evidence indicating that individuals with high $\dot{V}O_{2\,\max}$ ("fit but overweight") exhibit health risks comparable to those of normal-weight individuals.

This example illustrates how Novák's fuzzy context modeling provides a principled way to overcome the limitations of crisp classification by allowing smooth and interpretable transitions between categories while preserving the original four-class structure. Full algorithmic details and implementation are available in the `lfl` package; due to space constraints, we summarize only the main ideas here.

2 Introduction to BMI

The classical BMI is defined as BMI $= \frac{m}{h^2}$, where m denotes body mass (kg) and h denotes height (m). The standard categorization of individuals based on the World Health Organization (WHO) BMI standards is the following:

1. **Underweight (BMI $<$ 18.5)** High risk for nutritional deficiencies.
2. **Normal weight (BMI 18.5-24.9)** Optimal health range.
3. **Overweight (BMI 25-29.9)** Moderate risk.
4. **Obesity (BMI $\geq$ 30)** Severe risk of cardiovascular complications.

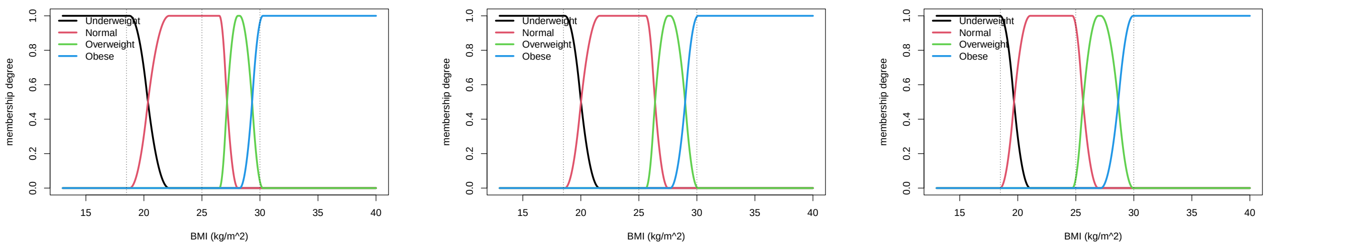
The distribution of these categories across the population provides a snapshot of overall health status.

Despite its simplicity and wide use, BMI faces many conceptual limitations that motivate ongoing research, e.g. BMI does not distinguish between fat and lean mass, optimal BMI thresholds differ across a population, health risk is not a monotonic function of BMI, and two individuals with the same BMI (e.g., athlete vs. sedentary) may have very different health statuses.

Fuzzy BMI approaches aim to overcome this limitation, but current models often rely on fixed partitions. Determining the *optimal number, shape, and overlap* of fuzzy sets that best reflect real-world health risks remains an open and critical research problem; see, e.g., fuzzy BMI indices that combine BMI with body-fat percentage for smoother classifications [2]. In more detail, fuzzy approaches replace crisp BMI thresholds with fuzzy sets and linguistic categories. In our case, we define using the preset linguistic expressions, their modifiers, and logical operations from the `lfl` package the following fuzzified classes:

under_weighted(x) = extremely_small(x),
normal(x) = not_extremely_small(x) & not_extremely_very_rougly_big(x),
over_weighted = not(normal(x)) $\vee$ extremely_small(x)) & not_extremely_big(x),
obese = ex_bi(x),

where $\vee$ and & stand for the Łukasiewicz disjunction and conjunction, respectively. In Figure 1, you can observe three different settings of the midpoint values of CTX3.

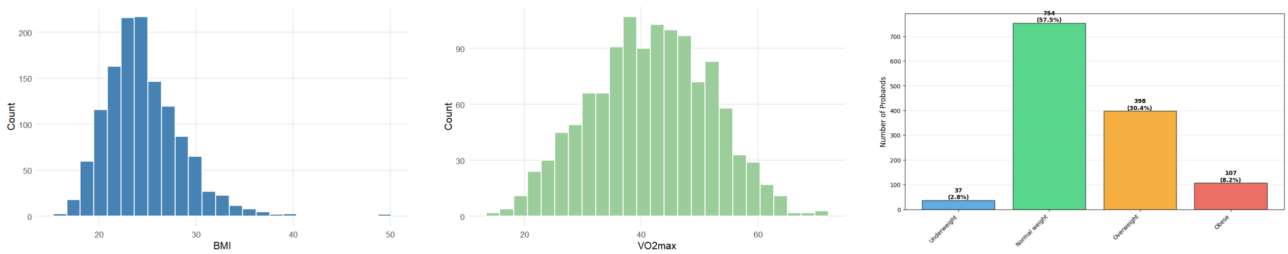


(a) $c = 26$ (b) $c = 25$ (c) $c = 24$

Figure 1: BMI fuzzy classes — comparison of three breakpoint settings with the fixed minimum context value 17.5 and maximum set to 31.

3 Comparison with the Standard WHO Classes

For our experiment, we use data from the Healthy Aging in Industrial Environment (4HAIE) cohort, which originally included 1,314 asymptomatic individuals. In this dataset, we observe the following distribution, see Figure 2. A Random Forest classifier is employed to classify individuals into one of four BMI categories: underweight, normal weight, overweight, and obese.



(a) Histogram of BMI (b) Histogram of VO_2max (c) Standard BMI classes

Figure 2: Histograms for 4HAIE data set.

Next, we used the mean decrease in impurity for model feature evaluation.

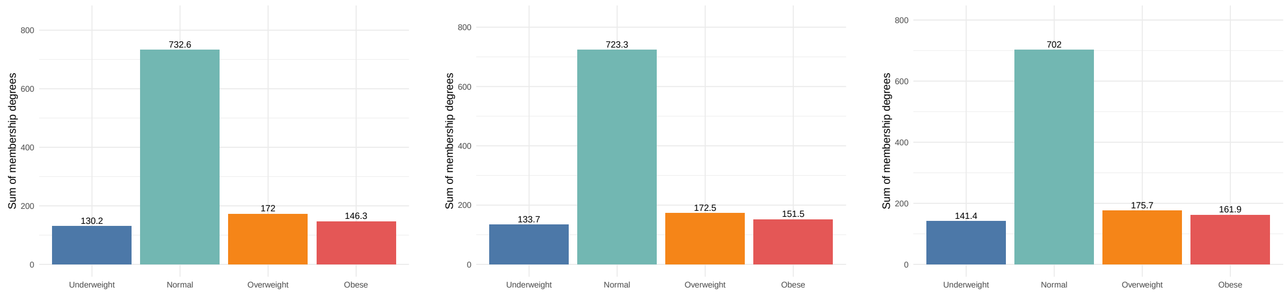
1. **Weight (36.07%)** Obviously, body weight in kilograms is the most influential feature in the model as a direct component of the BMI formula.

2. $\dot{V}O_{2\,\max}$ **(12.85%)** Maximal oxygen consumption ($\dot{V}O_{2\,\max}$) is a key indicator of cardiorespiratory fitness. Higher $\dot{V}O_{2\,\max}$ values typically correlate with better physical conditioning and lower fat mass. This feature helps distinguish between individuals with

All the data are available upon request from `https://haie-lerco.cz/data/`.

similar BMI but different fitness levels. hich are crucial for distinguishing between similar anthropometric profiles.

The remaining top five features are as follows: Age (9.85%), height (8.30%), and normalized minutes of physical activity (4.02%) significantly contribute to BMI prediction, reflecting the combined effects of age-related metabolic changes, height's moderating role in the BMI formula, and lifestyle activity patterns on body composition. Hence, we have confirmed that $\dot{V}O_{2\,\max}$ is the most influential non-defining parameter of BMI and should, therefore, be used as the principal input for individual-level BMI modification. Figure 3 illustrates three different transformations corresponding to distinct values of α.



(a) $\alpha = 4$ (b) $\alpha = 6$ (c) $\alpha = 10$

Figure 3: BMI fuzzy cardinalities for 4HAIE data with VO_2max-driven context.

4 Conclusion

The presented method is lightweight, transparent, and potentially suitable for deployment in population screening and clinical decision-support systems. It should be noted that there remains substantial room for improvement in several aspects—particularly in the choice of transformation—since $\dot{V}O_{2\,\max}$ is inherently nonlinear, and the applied transformation should adequately capture this property. Further refinement of this aspect will be the subject of future research.

References

[1] Novák, V.: *A comprehensive theory of trichotomous evaluative linguistic expressions.* Fuzzy Sets and Systems **159**(22) (2008) 2939–2969.

[2] Miyahira, S.A., Araujo, E.S.: *Fuzzy obesity index (MAFOI) for obesity evaluation and bariatric surgery indication.* Journal of Translational Medicine **9** (2011) 134.