APPLICATION OF FUZZY METRICS IN CLUSTERING PROBLEMS OF AGRICULTURAL CROP VARIETIES

Andrijana Stamenković[1], Nataša Milosavljević[2], Nebojša M. Ralević[3]
**Corresponding author E-mail: nralevic@uns.ac.rs*

A R T I C L E I N F O

Original Article

Received: *22 January 2024*

Accepted: *20 February 2024*

doi:10.59267/ekoPolj2401121S

UDC 519:633.1

Keywords:

agricultural crop varieties, fuzzy metrics, mathematical modeling, machine learning, clustering, variable environment method

JEL: *C02, C38*

A B S T R A C T

The problem of image-based detection of the variety of beans, using artificial intelligence, is currently dealt with by scientists of various profiles. The idea of this paper is to show the possibility of applying different types of distances, primarily those that are fuzzy metrics, in clustering models in order to improve existing models and obtain more accurate results. The paper presents the method of variable neighborhood search, which uses both standard and fuzzy t-metrics and dual fuzzy s-metrics characterized by appropriate parameters. By varying those parameters of the fuzzy metric as well as the parameters of the metaheuristic used, we have shown how it is possible to improve the clustering results. The obtained results were compared with existing ones from the literature. The criterion function used in clustering is a fuzzy metric, which is proven in the paper.

Introduction

When it comes to agricultural products, the dry bean (Phaseolus vulgaris L.) is the most important and the most produced (Fabaceae - Leguminosae) legume worldwide. The United Nations General Assembly declared 2016 the International Year of Brewing with the role "Nutritious grains for a secure future" considering that this group of crops is the

1 Andrijana Stamenković, MSc, Research associate, University of Novi Sad, Faculty of Technical Science, Department of Power, Electronic and Telecommunication Engineering, Trg Dositeja Obradovića 6, 21000 Novi Sad, Republic of Serbia. Phone: +381677152636, E-mail: stamenkovic.dt5.2011@uns.ac.rs, ORCID ID (https://orcid.org/0009-0003-4452-6370)

2 Nataša S. Milosavljević, Ph.D., Assistant Professor, University of Belgrade, Faculty of Agriculture, Department of Mathematics and Physics, Nemanjina 6, 11080 Zemun, Republic of Serbia, Phone: +381641271879, E-mail: natasam@agrif.bg.ac.rs, ORCID ID (https://orcid.org/0000-0003-4056-089X)

3 Nebojša M. Ralević Ph.D., Full Professor, University of Novi Sad, Faculty of Technical Science, Department of Fundamental Sciences, Trg Dositeja Obradovića 6, 21000 Novi Sad, Republic of Serbia. Phone: +381631028130, E-mail: nralevic@uns.ac.rs, ORCID ID (https://orcid.org/0000-0002-3825-9822)

source protein and high energy food. The most widespread legumes or grain legumes in the world are beans, dry peas and lentils, and in Serbia beans are indisputable (Vasić et al., 2019). The taste of beans and dishes made from them are so popular that many peoples consider them to be its national dish. In terms of its nutrition, it is one of the richest sources vegetable proteins, and often the only one for the poor population of the planet. Except tall their protein content is high and their quality is high, and their composition is also high usability in the human body can be a substitute for meat (Tepić et al., 2007). Starchy food is important, and thus energy food, which is also significant for the poor segment of the population. For a modern, urban man oriented towards life with little physical activity, it is especially important as source of dietary fibers, ballast substances, also necessary in the diet (Costa et al., 2006), as well as antioxidant substances (Karadžić Banjac et al., 2019). In addition to the chemical composition, an important role is played in assessing the quality of bean grains there is also a cooking time. Apart from food, beans are also used for other purposes, such as example for wastewater treatment (Šćiban et al., 2010).

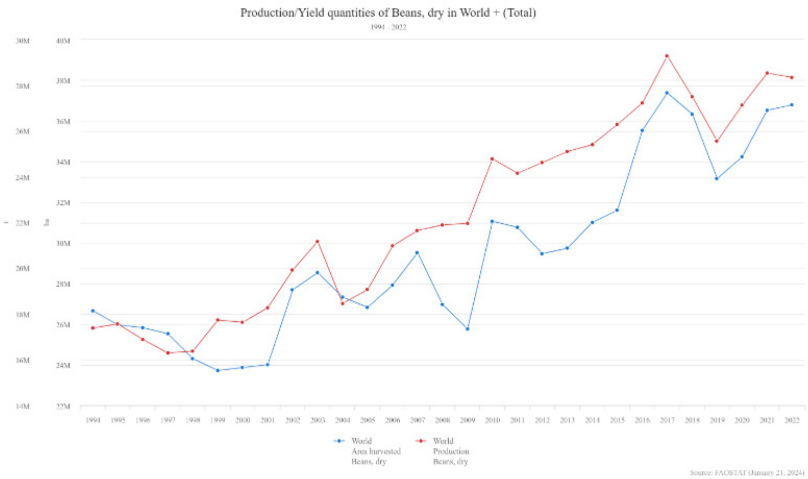
In modern agriculture where special attention is paid to sustainability and economics of leguminous production, and thus beans, occupy an important a valued place primarily because of its ability to interact with bacteria of the Rhizobium sp. they fix nitrogen from the air and thereby increase soil fertility. That's right additional food is also provided to plants in combined crops with legumes.

Beans can be successfully grown together with other crops, as they tolerate it well shade, and the microclimate that prevails during the gathering is very pleasing to him. Apart from that the advantage is the short growing season of modern varieties, which in our climatic conditions, enables the cultivation of beans as a second crop. But the short growing season of beans, as well as being in the field in the hottest time of the year makes the production of beans very intensive and risky. All agrotechnical measures must be not only adequate but also timely because every day of delay means a lot. This could be mitigated by providing better conditions for the growth and development of bean crops. Adequate nutrition, achieving the optimal composition of plants, irrigation and crop protection should be highlighted here. Unfortunately, protection against predominant diseases and pests in our country is usually insufficient effective, late, inadequate and too late, both due to inadequate choice of variety, establishment and care of crops, i.e. prevention, as well as due to poor information about the occurrence of the disease and pests and how to protect them.

It occupies a special role in the agriculture. It is sensitive to climate change and this can be mitigated by breeding the seed variety. Finding the best seed, not only for this variety, but also for plant varieties with similar problems, is a challenge for agriculture (Wakchaure et al., 2023). The reason lies in the fact that with such varieties, without breeding and finding a new variety that would be more resistant to climate changes, even if all conditions are met, the quality of production cannot be increased. Namely, seeds of lower quality in production will lead to a lower quantity, even if all the conditions for cultivation are provided. Dry beans are native to the Americas, while there is genetic diversity around the world (Koklu et al., 2020).

You can see how the production of beans in the world is moving in Figure 1. The biggest producer is Asia with 43.9%, than America with 31.9%, Africa 21.6%, Europe 2.3% and Oceania 0.3%.

Figure 1. Production/Yield quantities of Beans, dry in World + (Total) 1994 – 2022



Source: downloaded from https://www.fao.org/faostat/en/#data/QCL/visualize

Analysis and classification (Halder et al., 2023) of dry bean genotypes, which are very common in Turkey and around the world, constitute one of the main processes in crop production. Characteristics of dry beans determined by physical dimensions, such as appearance, size, color, interior and variety increase market value. Identification of bean varieties helps farmers to use seeds for planting and marketing. Manual classification and sorting of bean varieties is a long, inefficient and painstaking job. The problem of detecting the variety of beans based on an image, using artificial intelligence, is one of the current ones that scientists of various profiles deal with. For these reasons, it is good to apply artificial intelligence techniques and use its power, in order not only to facilitate this process and similar ones, but also to speed up development (Shukla et al., 2023). In the last few years, the application of artificial intelligence to various problems in agriculture is current, from image processing problems, up to prediction and optimization.

The idea of this paper is to show through a concrete example the application of mathematical models that can improve and speed up the solution of this problem. The work is divided into several sections. The next section will deal with the description of the distances used in the clustering procedure presented in this paper, as well as their applications not only in genetics and agricultural sciences, but also in other applied sciences where problems of a similar nature have arisen. In the first subsection, the definition of the fuzzy metric and the proven claim that the function that is used as a criterion function and whose arguments of the fuzzy metric are also fuzzy metrics are proven. We will also show how we used the variable environment method applied to this problem. In the Data Description section, the data that were used, what they represent and how the final conclusions were obtained based on them will be shown. The conclusions will be discussed in more detail in that section.

Materials and methods

Materials

In order to perform the analysis and classification of bean genotypes, we will consider its characteristics determined by physical dimensions (appearance, size, color, interior and variety), which we can examine using appropriate sample images. That's why the so-called are used in image analysis. shape descriptors (area and perimeter of the shape) of which we use some of them, such as the 16 features listed below.

In this research, Koklu et al., 2020 set up a camera and recorded the types of beans. The camera was connected to a computer and the image was converted into a signal using MATLAB R2016a. In total, 13,611 dry bean samples were obtained from 236 images. A detailed description of each attribute (features vectors) with which the grains are described can be seen at https://archive.ics.uci.edu/dataset/602/dry+bean+dataset.

The types of beans considered are Seker, Barbunya, Bombay, Cali, Dermosan, Horoz and Sira.

Fuzzy metric

The concept of distance in different sciences and applications can be interpreted in different ways. Mathematically, the distance is a non-negative symmetric function that assigns a non-negative number to each pair of elements (their order is not important) of a set (objects), and assigns 0 to two of the same elements. It can also be interpreted as the similarity or difference of two objects, and often and belonging to a group. We will use this feature of the distance in this paper for the clustering procedure.

The distance defined over a nonempty set U encompasses a wide family of mappings d from U^2 to $[0,+\infty)$. Depending on the research needs, d is required to have some important properties. Thus, the distance $d: U^2 \to [0,\infty)$ that satisfies the conditions:

1) $d(\eta,\varsigma) = 0 \Leftrightarrow \eta = \varsigma$;

2) $d(\eta,\varsigma) = d(\varsigma,\eta)$;

3) $d(\eta,\psi) \le d(\eta,\varsigma) + d(\varsigma,\psi)$.

is called *metric* and (U,d) is called *metric space* (m.s). If $U = R$ (real set) the usual Euclidean distance $d(u,v) = |u - v|$ is a metric.

It can be noticed that the distance between two object does not have to be a non-negative real number, but it can be taken to be, for example, a fuzzy set. Such generalization of distances is very useful in applications.

In order to define the notions the fuzzy *S*-metric and the fuzzy T–metric, the triangular norm (shorter t−norm) and triangular conorm (shorter t−conorm) are used (more details, for example, in book Klir et al., 1995).

The function T: $[0,1]^2 \to [0,1]$ (S: $[0,1]^2 \to [0,1]$) is *t−norm* (*t−conorm*) if it satisfies: associativity, commutativity, monotonicity and 1 (0) is neutral element.

The most widely used t-norms are: $T_M(a,b) = \min\{a,b\}$ and $T_P(a,b) = ab$ and their dual conorms: $S_M(a,b) = \max\{a,b\}$ and $S_P(a,b) = a + b - ab$.

Let $g:[0,1] \to [0,1]$ be an increasing continuous mapping such that $g(0) = 0$ and $g(1) = 1$. We will say $g \in G$. It is easy to show that with

$$T_g(u_1,u_2) = u_1 \otimes u_2 = g^{-1}(g(u_1)g(u_2))$$

a t-norm is defines the so called *pseudo multiplication* generated by g.

We will give the definition of fuzzy metrics and some examples that we use in this paper (more details, for example, in the papers Gregori et al., 2011, Ralević et al., 2019, Ralević et al., 2022.

Definition 1. Let U be a non-empty set, and T (S) a continuous t-norm (t-conorm). A fuzzy set $\mathbf{t}$ ($\mathbf{s}$) defined on $U \times U \times (0,+\infty)$ is called a *fuzzy T−metric* (shortly f.T.m.) (*fuzzy S−metric* (shortly f.S.m.)) if the following conditions for all $\eta,\varsigma,\psi \in U, \alpha,\beta > 0$, hold:

1) $\mathbf{t}(\eta,\varsigma,\alpha) \in (0,1]$ $\quad \left(\mathbf{s}(\eta,\varsigma,\alpha) \in [0,1)\right)$;

2) $\mathbf{t}(\eta,\varsigma,\alpha) = 1 \Leftrightarrow \eta = \varsigma$ $\quad \left(\mathbf{s}(\eta,\varsigma,\alpha) = 0 \Leftrightarrow \eta = \varsigma\right)$;

3) $\mathbf{t}(\eta,\varsigma,\alpha) = \mathbf{t}(\varsigma,\eta,\alpha)$ $\left(\mathbf{s}(\eta,\varsigma,\alpha) = \mathbf{s}(\varsigma,\eta,\alpha)\right)$;

4) $T\left(\mathbf{t}(\eta,\varsigma,\alpha),\mathbf{t}(\varsigma,\psi,\beta)\right) \le \mathbf{t}(\eta,\psi,\alpha+\beta)$

$\left(S(\mathbf{s}(\eta,\varsigma,\alpha),\mathbf{s}(\varsigma,\psi,\beta)) \ge \mathbf{s}(\eta,\psi,\alpha+\beta)\right)$

5) $\mathbf{t}(\eta,\varsigma,\_\ ):(0,+\infty) \to [0,1]$ $\left(\mathbf{s}(\eta,\varsigma,\_\ ):(0,+\infty) \to [0,1]\right)$ is a continuous function.

Example 1. The mapping $\mathbf{t}_p : U \times U \times (0,+\infty) \to (0,1],\ U \subseteq R_0^+$, defined by

$$\mathbf{t}_p(\eta,\varsigma) = \frac{\left(\frac{\eta^p+\varsigma^p}{2}\right)^{1/p} + L}{\max\{\eta,\varsigma\}+L},$$

$L > 0, p \geq 1$ is a f.T.m. with respect to T_P (see Milosavljević et al., 2023) and its dual (with respect to standard fuzzy complement)

$$\mathbf{s}_p(\eta,\varsigma) = 1-\mathbf{t}_p(\eta,\varsigma) = \frac{\max\{\eta,\varsigma\} - \left(\frac{\eta^p+\varsigma^p}{2}\right)^{1/p}}{\max\{\eta,\varsigma\}+L}$$

is a f.S.m. with respect to S_P.

Specially, $\mathbf{t}_1(\eta,\varsigma) = \frac{\frac{\eta+\varsigma}{2}+L}{\max\{\eta,\varsigma\}+L}$, $\mathbf{s}_1(\eta,\varsigma) = 1-\mathbf{t}_1(\eta,\varsigma) = \frac{\frac{|\eta-\varsigma|}{2}}{\max\{\eta,\varsigma\}+L}$

Example 2. The mapping $\Delta : U \times U \times (0,+\infty) \to (0,1],\ U \subseteq R_0^+$, defined by

$$\Delta(\eta,\varsigma) = \frac{\min\{\eta,\varsigma\}+L}{\max\{\eta,\varsigma\}+L},$$

$L > 0,$ is a f.T.m. with respect to T_P and its dual (with respect to standard fuzzy complement)

$$\Sigma(\eta,\varsigma) = 1-\Delta(\eta,\varsigma) = \frac{|\eta-\varsigma|}{\max\{\eta,\varsigma\}+L}$$

is a f.S.m. with respect to S_P.

Example 3. If (U,d) is a m.s., then the mapping $\tau_{h(t)} : U^2 \times R^+ \to R$ defined by

$$\tau_{h(t)}(\eta,\varsigma,t) = \frac{h(t)}{h(t)+d(\eta,\varsigma)}$$

is a f.T.m. with respect to the T_P and its dual (with respect to the standard fuzzy complement) is a f.S.m. with respect to S_P.

$$\sigma_{h(t)}(\eta,\varsigma,t)=1-\tau_{h(t)}(\eta,\varsigma,t)=\frac{d(\eta,\varsigma)}{h(t)+d(\eta,\varsigma)}$$

The results for the special case, when function *h(t)=L*, *L = const* and $d(\eta,\varsigma)=|\eta-\varsigma|$, are shown in Table 2.

Theorem 1. Let $g\in G$ and $g_j\in G, j\in J=\{1,\dots,n\}$. If $\mathbf{d}_j : U_j\times U_j\to(0,1]$, $j\in J$ are f.T.m.s with respect to t-norms T_j, j□J generated by $g_j\in G$, respectively, then the mapping $\mathbf{d}: U^2\to[0,1]$, $U = U_1\times\dots\times U_n$ given by

$$\mathbf{d}(\eta,\varsigma,\alpha)=g^{-1}\Big(g_1\big(\mathbf{d}_1(\eta_1,\varsigma_1,\alpha)\big)\cdot g_2\big(\mathbf{d}_2(\eta_2,\varsigma_2,\alpha)\big)\cdot\ldots\cdot g_n\big(\mathbf{d}_n(\eta_n,\varsigma_n,\alpha)\big)\Big),$$

$\eta=(\eta_1,...,\eta_n)$, $\varsigma=(\varsigma_1,...,\varsigma_n)$, is a f.T.m. with respect to triangular norm *T* generated by $g\in G$.

Proof.

1. $\mathbf{d}_j(\eta_j,\varsigma_j,\alpha)\in(0,1], j\in J\Rightarrow\ g_j\big(\mathbf{d}_j(\eta_j,\varsigma_j,\alpha)\big)\in(0,1], j\in J\Rightarrow$

$$g_1\big(\mathbf{d}_1(\eta_1,\varsigma_1,\alpha)\big)\cdot g_2\big(\mathbf{d}_2(\eta_2,\varsigma_2,\alpha)\big)\cdot\ldots\cdot g_n\big(\mathbf{d}_n(\eta_n,\varsigma_n,\alpha)\big)\in(0,1]\quad\Rightarrow$$

$$\mathbf{d}(\eta,\varsigma,\alpha)\in(0,1]$$

2. $\eta=\varsigma\Leftrightarrow\ \eta_j=\varsigma_j, j\in J\Leftrightarrow$

$$\mathbf{d}_j(\eta_j,\ \varsigma_j)=1, j\in J\Rightarrow\mathbf{d}(\eta,\varsigma,\alpha)=g^{-1}(g_1(1)\cdot g_2(1)\cdot\ldots\cdot g_n(1))=g^{-1}(1\cdot\ldots\cdot1)=1.$$

How $g\in G$ an d $g^{-1}(a)=1\Leftrightarrow a=1$,then

$a_1,\dots,a_n\in[0,1], a_1\cdot\ldots\cdot a_n=1\Leftrightarrow a_1=1\wedge\ldots\wedge a_n=1$, and then

$$\mathbf{d}(\eta,\varsigma,\alpha)=g^{-1}\Big(g_1\big(\mathbf{d}_1(\eta_1,\varsigma_1,\alpha)\big)\cdot g_2\big(\mathbf{d}_2(\eta_2,\varsigma_2,\alpha)\big)\cdot\ldots\cdot g_n\big(\mathbf{d}_n(\eta_n,\varsigma_n,\alpha)\big)\Big)=1$$

$$\Leftrightarrow g_1\big(\mathbf{d}_1(\eta_1,\varsigma_1,\alpha)\big)=g_2\big(\mathbf{d}_2(\eta_2,\varsigma_2,\alpha)\big)=\ldots=g_n\big(\mathbf{d}_n(\eta_n,\varsigma_n,\alpha)\big)=1$$

$$\Leftrightarrow\mathbf{d}_1(\eta_1,\varsigma_1,\alpha)=\mathbf{d}_2(\eta_2,\varsigma_2,\alpha)=\ldots=\mathbf{d}_n(\eta_n,\varsigma_n,\alpha)=1$$

$$\Leftrightarrow\eta_j=\varsigma_j, j\in J\Leftrightarrow\eta=\varsigma.$$

3. $\mathbf{d}(\eta,\varsigma,\alpha)=g^{-1}\Big(g_1\big(\mathbf{d}_1(\eta_1,\varsigma_1,\alpha)\big)\cdot g_2\big(\mathbf{d}_2(\eta_2,\varsigma_2,\alpha)\big)\cdot\ldots\cdot g_n\big(\mathbf{d}_n(\eta_n,\varsigma_n,\alpha)\big)\Big)$

$$=g^{-1}\Big(g_1\big(\mathbf{d}_1(\varsigma_1,\eta_1,\alpha)\big)\cdot g_2\big(\mathbf{d}_2(\varsigma_2,\eta_2,\alpha)\big)\cdot\ldots\cdot g_n\big(\mathbf{d}_n(\varsigma_n,\eta_n,\alpha)\big)\Big)=\mathbf{d}(\varsigma,\eta,\alpha)$$

4. From the fourth axiom for $\mathbf{d}_j,\ j \in J$

$$\mathbf{d}(\eta,\varsigma,\alpha+\beta)=$$

$$g^{-1}\left(g_1\left(\mathbf{d}_1(\eta_1,\varsigma_1,\alpha+\beta)\right)\cdot g_2\left(\mathbf{d}_2(\eta_2,\varsigma_2,\alpha+\beta)\right)\cdot\ldots\cdot g_n\left(\mathbf{d}_n(\eta_n,\varsigma_n,\alpha+\beta)\right)\right)$$

$$\geq g^{-1}\left(g_1\left(T_1(\mathbf{d}_1(\eta_1,\psi_1,\alpha),\mathbf{d}_1(\psi_1,\varsigma_1,\beta))\right)\cdot\ldots\cdot g_n\left(T_n(\mathbf{d}_n(\eta_n,\psi_n,\alpha),\mathbf{d}_n(\psi_n,\varsigma_n,\beta))\right)\right)$$

$$=g^{-1}\left(g_1\left(\mathbf{d}_1(\eta_1,\psi_1,\alpha)\right)\cdot g_1\left(\mathbf{d}_1(\psi_1,\varsigma_1,\beta)\right)\cdot\ldots\cdot g_n\left(\mathbf{d}_n(\eta_n,\psi_n,\alpha)\right)\cdot g_n\left(\mathbf{d}_n(\psi_n,\varsigma_n,\beta)\right)\right)$$

$$=g^{-1}\left(g_1\left(\mathbf{d}_1(\eta_1,\psi_1,\alpha)\right)\cdot\ldots\cdot g_n\left(\mathbf{d}_n(\eta_n,\psi_n,\alpha)\right)\cdot g_1\left(\mathbf{d}_1(\psi_1,\varsigma_1,\beta)\right)\cdot\ldots\cdot g_n\left(\mathbf{d}_n(\psi_n,\varsigma_n,\beta)\right)\right)$$

$$=g^{-1}\left(g\circ g^{-1}\left(g_1\left(\mathbf{d}_1(\eta_1,\psi_1,\alpha)\right)\cdot\ldots\cdot g_n\left(\mathbf{d}_n(\eta_n,\psi_n,\alpha)\right)\right)\right.$$

$$\left.\cdot\, g\circ g^{-1}\left(g_1\left(\mathbf{d}_1(\psi_1,\varsigma_1,\beta)\right)\cdot\ldots\cdot g_n\left(\mathbf{d}_n(\psi_n,\varsigma_n,\beta)\right)\right)\right)$$

$$=g^{-1}\left(g\left(\mathbf{d}(\eta,\psi,\alpha)\right)\cdot g\left(\mathbf{d}(\psi,\varsigma,\beta)\right)\right)$$

$$=\mathbf{d}(\eta,\psi,\alpha)\otimes\mathbf{d}(\psi,\varsigma,\beta).$$

5) From continuity of metrics $\mathbf{d}_j$, j□J and mapp g, follows continuity of

$\mathbf{d}(\eta,\varsigma,\_\ ):(0,+\infty)\rightarrow[0,1]$. □

Note that we will use the distance determined in this way as a measure of the object's belonging to a cluster.

Variable neighborhood search

The variable neighborhood search (VNS) is a metaheuristic that was mathematically founded by Mladenović et al., 1997. It is based on a single solution and a search of its environments. The systematic use of multiple environments increases the efficiency of the search. VNS (Hansen et al., 2001.) relies on three simple facts:

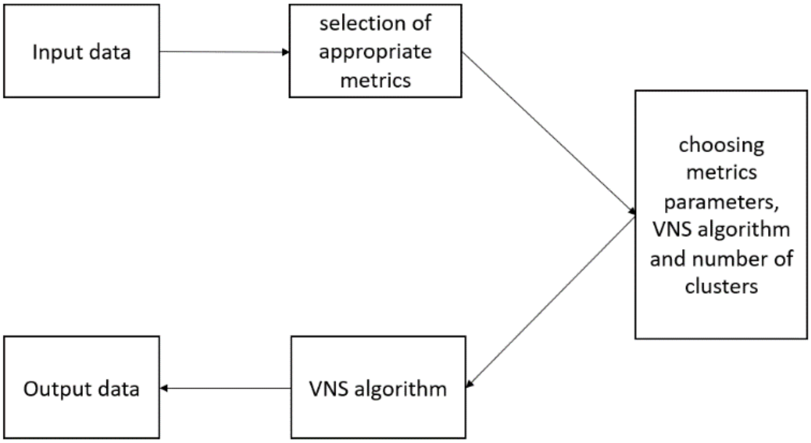
1) The local optimum in relation to one environment does not have to be the optimum in relation to another

2) The global optimum is local in relation to each environment

3) For most problems, the local optimums in relation to various environments are relatively close.

In this work, we applied this method in order to better cluster dry bean varieties. You can see the description of the algorithm in Figure 2.

Figure 2. Description of the step-by-step algorithm applied for clustering.

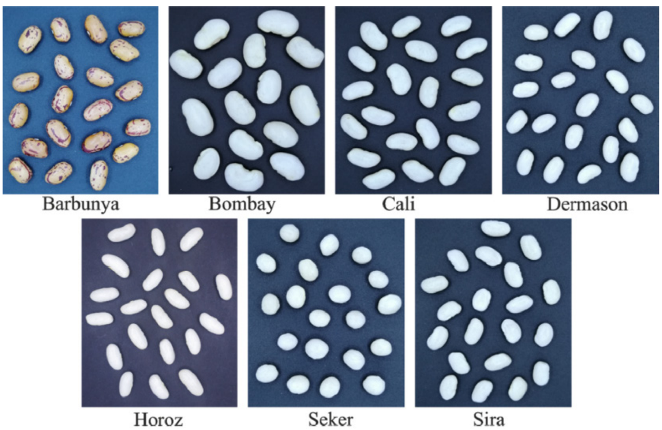


Source: Authors' figure

Results

We downloaded the data used in this work from the website https://archive.ics.uci.edu/dataset/602/dry+bean+dataset (Dry Bean Dataset. (2020). UCI Machine Learning Repository. https://doi. org/10.24432/C50S4B.). This database contains 7 different types of dry beans, where characteristics such as shape, type and structure are taken into account according to the market situation. For the classification model, 13,611 grains were used, 7 different registered dry grains (Figure 3.) were recorded with high-resolution cameras (more details in the paper Koklu et al., 2020. Images subjected to this kind of computer processing and segmentation, from which 16 characteristics were obtained for each grain, were used in the analysis of our proposed model.

Figure 3. Sample of taken dry bean images.



Source: Koklu et al., 2020

If we look at the descriptive statistics of the characteristic vectors of the beans (Table 1.) we can notice large changes in mean values, median, standard deviation and variance. These were some of the reasons why we do not use some of those values for centroids in our algorithm. Candidates for centroids in our algorithm are values from the database.

Table 1. Descriptive statistics feature vectors

Feature vectors	Mean	Median	Std. Deviation	Variance
1	53048.28	44652	29324.1	8.6E+08
2	855.2835	794.941	214.2897	45920.07
3	320.1419	296.8834	85.69419	7343.494
4	202.2707	192.4317	44.97009	2022.309
5	1.5832	1.5511	0.24668	0.061
6	0.7509	0.7644	0.092	0.008
7	53768.2	45178	29774.92	8.87E+08
8	253.0642	238.438	59.17712	3501.932
9	0.7497	0.7599	0.04909	0.002
10	0.9871	0.9883	0.00466	0
11	0.8733	0.8832	0.05952	0.004
12	0.7999	0.8013	0.06171	0.004
13	0.0066	0.0066	0.00113	0
14	0.0017	0.0017	0.0006	0
15	0.6436	0.642	0.099	0.01
16	0.9951	0.9964	0.00437	0

Source: Authors' calculations

The results we obtained compared with the results obtained in the work of Koklu et al., 2020 are shown in Table 2. The input data is the feature vector data of the publicly available database that we used. By metrics and their selection (Figure 2.), we mean the metrics presented in the methodology of this work. The output of the algorithm is the percentage of successful clustering of bean varieties. Since we obtained a unique best performance for changing the parameters of various phase metrics, we put it in Table 2. in bold. You can see how the performance values moved by varying certain parameters. We varied the parameter p, while the parameter K was 1.

Table 2. Results classification.

Name of methods	**Accuracy (%)**
MLP (Koklu et al., 2020,)	91.73
SVM (Koklu et al., 2020,)	93.13
DT (Koklu et al., 2020,)	87.92
kNN (Koklu et al., 2020,)	92.52
VNS metric Δ	96.01

Name of methods	Accuracy (%)
VNS metric $\mathbf{s}_1$	95.71
VNS metric $\mathbf{s}_2$	94.22
VNS metric $\mathbf{s}_3$	**96.12**
VNS metric t_1	95.71
VNS metric t_2	94.22
VNS metric t_3	**96.12**
VNS metric Σ	96.01
VNS euclidean distance	93.98
VNS special case metric τ	95.02
VNS special case metric σ	95.02

Source: Autor's and Koklu et al., 2020, calculations.

Discussions

In this paper, we have shown on one example, by changing only some values of the distance parameters, how the success of the clustering algorithm can be improved by using different distances. In comparison with the results obtained by Koklu et al., 2020 the success of some other clustering techniques and our VNS proposed model, we see that the success rates are also different. Our proposed model further performs better in all applications. Details of the results can be seen in Table 2. The application of each of the proposed metrics performed better than the performance of the model in Koklu et al., 2020. One can also see how the performance results change when the metrics and their parameters are changed. The best score is marked in bold in Table 2. Success in this case means how many bean varieties are distributed in those cluster groups to which it really belongs in reality. The percentage of that success is shown in the Table 2.

The proposed model has a wide application in agricultural problems. In addition to this problem, clustering methods can also solve the problems of detecting diseases that attack vegetables and fruits. Also, by analyzing the image in similar ways, we can detect the stages of fruit development. These are just some of the problems.

Conclusions

The proposed model showed better performance compared to others in Koklu et al., 2020. The entire work, which relies on the results presented in the work of other scientists, opens up possibilities for further research, and some of them are the possibilities of image processing in different ways, the possibility of improving and combining existing methods, as well as finding adequate distances that would lead to a more accurate model.

Khan et al. 2023 did similar research. They proposed other classifiers. They proposed an algorithm that reduces oversampling deviations and found better classifiers in that sense. Our results perform with higher accuracy. Dogan et al. 2023 in his study presents different proposals for finding classifiers. With him, the number of them has been reduced to 14. He combines various methods to achieve better precision. Taspinar et. al. 2022. in his work, he studies the classification of beans using deep learning techniques.

Diseases are an inevitable part of this topic. They limit the maximum yields. Little work is done in this field, especially to determine the degree of crop damage. The modeling that I deal with in this paper can be applied to these problems as well. Image processing is applicable to any type of images with minor modifications and refinements of the algorithm. Problems similar to this one were investigated by Hashemi-Beni et al. 2020. In their work, they propose a model that uses deep learning methods in precise agriculture. It especially deals with the problem of weeds and crop damage. He included a large number of paintings in his studio. Similar to these studies can be seen in Kumar et al. 2023. This paper deals with the sampling of vegetable leaves. Image processing uses different machine learning processes, using different classifiers.

Our future research will expand in these directions.

Acknowledgements

This research was partially supported by the Science Fund of the Republic of Serbia, #GRANT No 7632, Project ”Mathematical Methods in Image Processing under Uncertainty” - MaMIPU

Conflict of interests

The authors declare no conflict of interest.

References

1. de Almeida Costa, G. E., da Silva Queiroz-Monici, K., Reis, S. M. P. M., & de Oliveira, A. C. (2006). Chemical composition, dietary fibre and resistant starch contents of raw and cooked pea, common bean, chickpea and lentil legumes. *Food chemistry*, *94*(3), 327-330.
2. Dogan, M., Taspinar, Y. S., Cinar, I., Kursun, R., Ozkan, I. A., & Koklu, M. (2023). Dry bean cultivars classification using deep cnn features and salp swarm algorithm based extreme learning machine. *Computers and Electronics in Agriculture*, *204*, 107575.
3. Gregori, V., Morillas, S., & Sapena, A. (2011). Examples of fuzzy metrics and applications. *Fuzzy sets and systems*, (170), 95-111.
4. Halder, S., Bhattacharya, S., Roy, M. B., & Roy, P. K. (2023). Application of fuzzy C-means clustering and fuzzy EDAS to assess groundwater irrigation suitability and prioritization for agricultural development in a complex hydrogeological basin. *Environmental Science and Pollution Research*, *30*(20), 57529-57557.
5. Hansen, P., & Mladenović, N. (2001). Variable neighborhood search: Principles and applications. *European journal of operational research*, *130*(3), 449-467.

6. Hashemi-Beni, L., & Gebrehiwot, A. (2020). Deep learning for remote sensing image classification for agriculture applications. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 51-54. https://doi.org/10.1016/j.fss.2010.10.019
7. Karadžić Banjac, M., Kovačević, S., Tepić Horecki, A., Šumić, Z., Vakula, A., Podunavac Kuzmanović, S., & Jevrić, L. (2019). Toward consistent discrimination of common 742 BILJNI LEKAR / PLANT DOCTOR, 49, 6/2021 bean (Phaseolus vulgaris L.) based on grain coat color, phytochemical composition, and antioxidant activity. *J Food Process Preserv*. 2019; 43: e14246.
8. Khan, M. S., Nath, T. D., Hossain, M. M., Mukherjee, A., Hasnath, H. B., Meem, T. M., & Khan, U. (2023). Comparison of multiclass classification techniques using dry bean dataset. *International Journal of Cognitive Computing in Engineering*, *4*, 6-20.
9. Klir, G.J., & Yuan, B. (1995). *Fuzzy sets and fuzzy logic, theory and applications.* Prentice Hall, New Jersey.
10. Koklu, M., & Ozkan, I. A. (2020). Multiclass classification of dry beans using computer vision and machine learning techniques. *Computers and Electronics in Agriculture*, *174*, 105507.
11. Kumar, C., & Kumar, V. (2023, March). Vegetable plant leaf image classification using machine learning models. In *Proceedings of Third International Conference on Advances in Computer Engineering and Communication Systems: ICACECS 2022* (pp. 31-45). Singapore: Springer Nature Singapore.
12. Milosavljević, N. S., & Ralević, N. M. (2023). Fuzzy methaheuristic model for copy-move forgery detection on images. *Multimedia Tools and Applications*, DOI: 10.1007/s11042-023-17053-7
13. Mladenović, N., & Hansen, P. (1997). Variable neighborhood search. *Computers & operations research*, *24*(11), 1097-1100.
14. Ralević, N. M., Delić, M., & Nedović, Lj. (2022). Aggregation of fuzzy metrics and its application in image segmentation. *Iranian journal of fuzzy systems* 19(3), 19-37. DOI: https://doi.org/10.22111/ijfs.2022.6941
15. Ralević, N. M., Karaklić, D., & Pištinjat, N. (2019). Fuzzy metric and its applications in removing the image noise. *Soft Computing* 23(22), 12049–12061. DOI: 10.1007/s00500-019-03762-5, http://link.springer.de/link/service/journals/00500/index.htm
16. Ralević, N., & Paunović, M. (2021). Applications of the Fuzzy Metrics in Image Denoising nd Segmentation. *Tehnički vjesnik - Technical Gazette* 28(3), 819-826. DOI: https://doi.org/10.17559/TV-20200305075136
17. Šćiban, M. B., Vasić, M. A., Prodanović, J. M., Antov, M. G., & Klašnja, M. T. (2010). The investigation of coagulation activity of natural coagulants extracted from different strains of common bean. *APTEFF*, 41, 141- 147.

18. Shukla, M. K., & Sharma, P. (2023). Fuzzy k-means and principal component analysis for classifying soil properties for efficient farm management and maintaining soil health. *Sustainability*, *15*(17), 13144.
19. Taspinar, Y. S., Dogan, M., Cinar, I., Kursun, R., Ozkan, I. A., & Koklu, M. (2022). Computer vision classification of dry beans (Phaseolus vulgaris L.) based on deep transfer learning techniques. *European Food Research and Technology,* 248, 2707–2725. https://doi.org/10.1007/s00217-022-04080-1
20. Tepić, A., Vujičić, B., Vasić, M., & Lučić A. (2007). Amino acids and phytic acid in some Serbian varieties of dry beans (Phaseolus vulgaris); *Proc.of the 2nd Internat. Cong. on Food and Nutrition*, 24-26 October 2007, Istanbul, Turkey, 180-182.
21. Vasić, M., Malidža, G, & Rajković, M. (2019). Prinos pasulja u združenoj setvi sa kukuruzom tolerantnim na cikoksidim. Zbornik radova 1, *XXIV Savetovanje o biotehnologiji sa međunarodnim učešćem, Čačak*, 15-16 mart 2019., 141-148. [*in English*: Vasić, M., Malidža, G, & Rajković, M. (2019). Yield of beans in joint sowing with maize tolerant to cycoxidum. Proceedings 1, XXIV *Conference on Biotechnology with International Participation, Čačak*, March 15-16, 2019, 141-148.]
22. Wakchaure, M., Patle, B. K., & Mahindrakar, A. K. (2023). Application of AI techniques and robotics in agriculture: A review. *Artificial Intelligence in the Life Sciences*, 100057.