



RESEARCH ARTICLE

# An algorithm for creating semi-synthetic datasets for regional classification of cotton varieties

Nishanov Akhram Khasanovich<sup>1</sup>, Akhmedov Oybek Kamarbekovich<sup>1</sup> & Aktamov Shokhrukhbek Ulug'bek o'g'li<sup>2\*</sup>

<sup>1</sup>Department of System and Practical Programming of Tashkent University of Information Technologies named after Muhammad Al-Khwarizimi, Tashkent 100200,108, Uzbekistan

<sup>2</sup>Tashkent University of Information Technologies named after Muhammad Al-Khwarizimi, Tashkent 100200,108, Uzbekistan

\*Correspondence email - [aktamovshohruhbek@gmail.com](mailto:aktamovshohruhbek@gmail.com)

Received: 14 November 2025; Accepted: 13 April 2026; Available online: Version 1.0: 27 April 2026

**Cite this article:** Nishanov AK, Akhmedov OK, Aktamov SU. An algorithm for creating semi-synthetic datasets for regional classification of cotton varieties. *Plant Science Today* (Early Access). <https://doi.org/10.14719/pst.12737>

## Abstract

In this study, an algorithm for synthetic expansion of training datasets was developed and applied to the regional classification of cotton varieties recommended for cultivation in the Republic of Uzbekistan. The algorithm is based on heuristic logic, and the class objects were restructured based on similarity criteria. Initially, a space of textual, nominal, and quantitative features was formed using real data from the state register. Subsequently, the features were fully converted to nominal form, and the degrees of similarity between objects were determined through scaling. A proximity function and decision rules were developed, and the contribution of class objects to their respective classes was evaluated. Artificial objects were generated based on heuristic criteria, increasing the number of classes and their elements. This approach significantly improved the stability and accuracy of the classification model. Experimental results showed that the proposed algorithm achieved a precision of 95.8 %, which is substantially higher than that of the decision tree (87 %) and KNN (84.4 %) algorithms, demonstrating the effectiveness of the proposed method. The research results have created the opportunity to accurately classify cotton varieties by region.

**Keywords:** automatic classification; classification; cotton varieties; degree of similarity; nominal and quantitative; objects; regionalisation; synthetic training samples; text

## Introduction

The current global cotton industry demands the identification of cotton varieties adapted to specific climatic conditions to ensure high yields and quality. Cotton is one of the world's most economically significant agricultural commodities, supporting the livelihoods of hundreds of millions of people and anchoring global textile, trade, and manufacturing industries. Major cotton-producing countries like China, India, the USA, and Brazil have recommended varieties for each agro-ecological zone, which differ in their class, productivity, disease resistance, and fibre quality. In recent years, the use of artificial intelligence technologies for classifying cotton varieties has been intensifying. However, preparing high-precision classification models requires large volumes of training samples. High-precision artificial intelligence classification models require large training datasets to ensure sufficient variability and model reliability.

In many cases, due to insufficient real agrobiological data, the limited number of objects in each class reduces classification accuracy. From this perspective, generating synthetic replicas of cotton varieties based on real data is considered one of the pressing issues in international research. Agriculture, particularly cotton cultivation, is one of the main sectors of Uzbekistan's economy. Due to varying regional climatic and soil conditions, it is crucial to identify

and allocate suitable varieties for each province. In recent years, local breeders have developed varieties such as Zangi-Ota-2, Ishonch, Zafar and Niyat, which are currently being tested for their suitability in different regions. Zangi-Ota-2, a high-yielding cotton variety with good fibre quality and stable performance across different soil and climate conditions. It is valued for its adaptability and reliable productivity. Ishonch, an early-maturing, high-yield variety known for its tolerance to drought and soil salinity. It produces fiber that meets international quality standards. Zafar, an early-ripening cotton variety with consistently high fibre quality. It has shown strong results in regional field trials and is suitable for intensive cultivation. Niyat, a locally developed cotton variety characterised by good adaptation to local agro-ecological conditions and stable yields, making it suitable for practical farming systems.

Concurrently, the number of varieties included in the state register is increasing, though this process demands time and resources. If artificial objects were created based on the existing register, automated and intelligent systems for adapting varieties to specific regions could operate more quickly, reliably, and with higher precision. This would not only accelerate the selection process but also significantly contribute to the formation of the state register.

The rapid development of information technologies and the application of artificial intelligence algorithms in cotton farming are

creating significant advancements in the classification of cotton varieties. Although the existing real training datasets include 386 objects, the small number of objects in some classes reduces the reliability of classification. Therefore, by developing algorithms for creating synthetic datasets based on heuristic logic, the number of objects in each class is balanced, and the generalisation capability of machine learning models increases. Furthermore, this approach expands the possibilities for accelerating regional zoning and automating the analysis of varieties included in the registry.

Agriculture, systems for automatic classification of cotton varieties are being developed alongside digitalisation processes. However, the lack of a large, balanced training dataset consisting of clearly distinguishable objects remains a major challenge for these systems. In this context, synthetic training samples enable the achievement of reliable, reproducible, and highly accurate results in artificial intelligence and deep learning models. Simultaneously, the effectiveness of scientific and practical activities significantly increases through the selection of new varieties, their adaptation to specific regions, ensuring economic efficiency, and reducing trials in experimental fields. Thus, algorithms for creating synthetic samples are becoming an integral component of modern digital agriculture.

## Materials and Methods

Various algorithms for the territorial classification of agricultural crops were studied. Among them are hierarchical, density, network, and model methods, whose effectiveness was evaluated using accuracy, precision, and recall metrics. The importance of methods such as K-means and DBSCAN in the regionalisation process has been substantiated. Partitioning methods, also known as centroid-based methods. An example of an algorithm using the division method is the K-means algorithm. By finding the central points K and assigning a regionalisation corresponding to the center closest to each point, the divides the data points into K-separated regionalisations. Fränti P and Sieranoja S proposed algorithms and models that study the properties of the applied intelligence dataset across 6 clusters in K-means (1). Fränti and Sieranoja proposed K-means-based algorithms to analyse the structural properties of datasets through multi-cluster partitioning. Their work provides a methodological foundation for evaluating clustering behaviour, which is relevant to this study in designing and validating semi-synthetic datasets for regional classification of cotton varieties. Jianpeng Qi, Yanwei Yu, Lihong Wang, Jinglei Liu proposed effective methods of the K-means clustering algorithm (2). Density methods, an example of a density algorithm is DBSCAN (3), which is widely applied in clustering tasks. It groups nearby points based on distance (usually Euclidean) and the minimum number of points in a group. Ruitong Zhang, Hao Peng, Yingtong Dou, Jia Wu, Qingyun Sun, Jingyi Zhang, and Philip S Yu proposed a new automated deep learning-based system for searching DBSCAN parameters (4). Ala'raj M, Majdalawieh M and Abbod MF conducted research on improving binary classification using filtering based on K-NN proximity graphs (5), while Ignatyev NA carried out research on selecting structures for relationships between objects in metric classification algorithms (6).

As stated in the literature (7, 8) it is assumed that the set of training samples is expressed in the following form:

$$x_{p1}, x_{p2}, \dots, x_{pm_p} \in X_p, p = \overline{1, r}$$

Where,  $x_{pi}$  is a vector of N-dimensional text, nominal and

quantitative character spaces, each object  $x_{pi} = (x_{pi}^1, x_{pi}^2, \dots, x_{pi}^N), i = \overline{1, m_p}$ , is considered in N-dimensional character spaces,  $X_p, p = \overline{1, r}$  represents a set of classes consisting of  $m_p, x_{p1}, \dots, x_{pm_p}$  objects. To determine the degree of similarity between objects, a proximity function is introduced: each object belonging to a class is compared one by one with objects in its class, and the inter-object proximity function is below.

Suppose two objects  $x_{p1}, x_{p2}$  are given in the text, nominal, and quantitative character spaces of the class  $X_p$ . Here, the proximity function  $\rho_j(x_{p1}, x_{p2})$  between objects is entered in the text, nominal and quantitative character spaces as follows:

$$\rho_j(x_{p1}, x_{p2}) = \begin{cases} 1 & \text{if } (x_{p1}^j - x_{p2}^j) = 0, \quad j = \overline{1, N} \\ 0 & \text{then} \end{cases}$$

$$P_j(x_{p1}, x_{p2}) = \begin{cases} 1, & \text{if } x_{p1}^j - x_{p2}^j = 0, \quad j = 1, \dots, N \\ 0, & \text{otherwise} \end{cases}$$

The first condition indicates the degree of similarity between two objects, while the second condition indicates the magnitude of their difference from each other, meaning these components are not similar to one another.

In the space of textual, nominal, and quantitative features, the estimation of the value indicating the degree of similarity between an arbitrary p-class k-object and another t-object is calculated based on the following formula:

$$G_t(x_{pk}, x_{pt}) = \sum_{j=1}^N \rho_j(x_{pk}, x_{pt}), k = \overline{1, m_p}; t = \overline{1, m_p}; k \neq t$$

In the space of text, nominal, and quantitative features, the estimation of the value indicating the degree of similarity between an object k of any class p and all other objects is calculated using the following formula:

$$G_k(x_{pk}, x_{pt}) = \sum_{t=1}^{m_p} G_t(x_{pk}, x_{pt}), k = \overline{1, m_p}; t = \overline{1, m_p}; k \neq t$$

Consider a new unknown object  $w=(w^1, w^2, \dots, w^N)$  is given. To determine which of the given class objects this new object most closely resembles, the following formula is calculated:

$$G_w(w, x_{pk}) = \sum_{k=1}^{m_p} \sum_{j=1}^N \rho_j(w, x_{pk}), k = \overline{1, m_p};$$

If the inequality  $G_w(w, x_{pi}) > G_w(w, x_{pk})$  is satisfied, then the object  $w=(w^1, w^2, \dots, w^N)$  has a higher degree of belonging to the i-th object compared to other objects.

Classification and regionalisation of cotton varieties:

Based on the information provided above, the following issues have been addressed:

1. The task of numerically representing cotton characteristics by scaling them according to Table 1.
2. The problem of determining the degree of similarity between objects.

- The matter of classifying and regionalising cotton varieties in agriculture.
- Developing an algorithm for selecting a set of important features based on classification and regionalisation.

Based on the proposed theoretical framework and algorithm, the above-mentioned problems are addressed. The degree of similarity of cotton varieties according to the proximity function is carried out and expressed (4, 9).

The following classification results were obtained:

Objects  $x_{17}$  and  $x_{18}$  are considered to have the highest vote for  $x_1$  object. The similarity rating of these objects is 5. That is, 5 out of 12 possibilities.

Objects  $x_7$ ,  $x_8$ ,  $x_9$  and  $x_{19}$  are considered to have the highest vote for object  $x_2$ . The similarity rating between these objects is 5, meaning that 5 out of 12 evaluated features (attributes) are identical.

Objects  $x_{11}$  and  $x_{14}$  are considered to have the highest vote for object  $x_3$ . The similarity rating of these objects is equal to 9. That is, 9 out of 12 possibilities.

Using these classes, the third task focuses on the regionalisation of cotton varieties by evaluating their suitability and performance under different agro-ecological conditions. The degree of similarity of these objects is equal to 9.

Using these classes, the problem of regionalisation is investigated in the third problem.

As a result of preliminary data processing, the training sample has been divided into classes. For example, if we look at the first class, this class contains 10 objects, which consist of the following: The first class is composed of 10 objects, as follows:

$X_1 = \{x_1, x_2, x_5, x_6, x_7, x_9, x_{14}, x_{16}, x_{17}, x_{18}\} = \{\text{Bukhara-8, Chimboy - 5018, S- 6541, Khwarazm - 150, Do'stlik-2, Beshqahramon, S-6775, S -8284, Porloq, Porloq -2}\}$

As a result, the matrix representation of class  $X_1$ , along with its class objects and representative object, will be as follows:

$X_{11} = (322241122823)$ - this  $x_1$  is the representative object of the class.

A comprehensive analysis of the classification and regionalisation of agricultural industrial crops was conducted, focusing on cotton varieties. According to the analysis, cotton varieties belonging to the first class are recommended for planting primarily in the following regions: the Republic of Karakalpakstan, Kashkadarya region, Samarkand region, Surkhandarya region, and Khorezm region.

This research work has been implemented in practice at the Center for Innovative Developments and Consultations at the Andijan Institute of Agriculture and Agrotechnologies, the "Pakhtakul istiqboli" farm of the Council of Dehkan Farms and Household Plot Owners in Shakhrikhan District of Andijan region, as

**Table 1.** Initial information provided on cotton varieties

No.	Cotton name	Stalk height	Pattern	Stalk	Stalk color	Stem	Leaf	Leaf color	Cotton blossom	Boll size	Boll form	Boll color	Seed weight (1000)
1	$x_1$ Bukhara-8	90–20 sm	Columnar	Sparsely haired	Dark green, doesn't get stuck	2	Medium, 3–5 lobed	Green	Big	Big	Sparsely developed nostril	Dark green	138 g
2	$x_2$ Chimboy-5018	110 sm	Conical	Sparsely haired	Dark green, doesn't get stuck	1, 5	Medium, 5 lobed	Dark green	Medium	Medium	Small-nosed	Dark green	131 g
24	Future	120–130 sm	Conical	Loose-haired	Green, doesn't get stuck	1–2	Medium, 3–5 lobed	Green	Medium	Medium	Round-oblong	Green	122 g

Further, based on the data presented in Table 1 and the notations, the following Table 2 is formed. The row elements of the table represent objects (varieties), while the column elements represent characteristics (10–14).

**Table 2.** Text, nominal and quantitative features of cotton varieties

Cotton varieties	Features											
	$x_1^1$	$x_1^2$	$x_1^3$	$x_1^4$	$x_1^5$	$x_1^6$	$x_1^7$	$x_1^8$	$x_1^9$	$x_1^{10}$	$x_1^{11}$	$x_1^{12}$
$x_1$	3	1	1	1	1	1	1	1	1	1	1	4
$x_2$	2	2	1	1	2	1	2	2	2	2	1	3
$x_3$	3	3	2	2	3	1	1	2	2	3	2	1
$x_4$	3	2	3	2	4	2	1	2	2	4	2	2
-	-	-	-	-	-	-	-	-	-	-	-	-
$x_{23}$	5	2	3	2	7	1	1	2	2	5	2	1
$x_{24}$	5	2	3	2	7	1	1	2	2	5	2	2

well as at the Council of Dehkan Farms and Household Plot Owners of Andijan region. As a result of implementing the software "E-classification and zoning," which was developed based on methods and algorithms for classification, selection of informative feature sets, and assessing classification effectiveness during the preliminary processing of information about cotton types and varieties grown in the republic, 24 cotton varieties regionalised according to the state register were divided into 3 groups. The degree of similarity between objects within each group has been increased by 65 %. The set of informative features has been successfully reduced from 12 to 3, resulting in a 3.5 to 4-fold increase in efficiency levels. According to these results, it is possible to improve labor productivity by 17–22 %.

Algorithm for creating a synthetic training dataset aimed at regionalisation and classification of cotton varieties

This algorithm aims to create a new expanded dataset by evaluating the similarity of synthetic objects generated for a nominal feature space. The proposed algorithm expands the original real dataset by generating synthetic objects and evaluating the similarity between the real and synthetic objects in a nominal feature space. Before creating the dataset, it becomes necessary to increase the number of classes and objects in order to improve the results and indicators obtained from the research and enhance their percentage metrics.

For this purpose, in accordance with the state register, the cotton varieties to be sown in each region will be determined using the table of digitalisation of regions. Subsequently, varieties sown jointly in two regions, as well as those sown in three or more regions, are identified and recorded separately.

Classes are formed as follows:

As a result of the designations, 18 classes and class objects are formed in the table below.

The algorithm consists of the following steps:

Step 1: The objects of the general training sample are entered into the database. The initial database is formed in the context of all

objects  $x_i \in X, i = \overline{1, M}$ ; Let  $X = \{x_1, x_2, \dots, x_M\}$  represent the set of cotton varieties in the initial database, where M is the number of varieties. The initial database is formed using all objects  $x_i \in X, i = \overline{1, \dots, M}$ .

Step 2: The process of preliminary data processing is carried out. This stage involves data cleaning, which includes eliminating errors, restoring lost values, removing duplicates, and clarifying ambiguous information in the collected data. After these adjustments, data normalisation is performed.

Step 3: The general training dataset is preliminarily divided into

classes  $x_{p1}, x_{p2}, \dots, x_{pm_p} \in X_p, p = \overline{1, r}$ .

Step 4: For each formed class, the average similarity values of objects

$G_p(X_p), p = \overline{1, r}$  are determined.

Step 5: A new object  $\bar{x}$  is created for classes  $X_p$ , where  $p = \overline{1, r}$ , while preserving essential features that ensure an increase in the

Region	Cotton varieties
Republic of Karakalpakstan	Chimbay-5018; Dustlik-2; Bukhara-8; Sultan
Andijan	Istiqlol-14; Andijon-36; Andijon-37; Sultan
Bukhara	Bukhara-8
Jizzakh	C-6541; Pakhtakor-1; Ko'paysin; Porloq-2
Kashkadarya	Namangan-102; Beshqahramon; Bukhara-8; Sultan
Surkhandarya	C-6550; Beshqahramon
Syrdarya	Gulbakhor-2; Navbakhor-2; Ko'paysin
Tashkent	Kelajak; Andijon-36; Sultan
Fergana	C-6775; Porloq-4; Andijon-36; Sultan
Khorezm	Khorezm-150; Ibrat; Porloq-1; Porloq-2
Samarkand	Porloq-1; Namangan-34; Sultan
Navoi	Namangan-34; Sultan
Namangan	Porloq-2; Andijon-36; Sultan

average similarity level  $G_p(X_p)$  of objects within this class.

Step 6: If a class p contains only one object ( $m_p = 1$ ), then when creating a synthetic object, it is formed while preserving the features that distinguish the object of this class from objects of other classes.

Step 7: Steps 5 and 6 are repeated until K new synthetic objects are added to each class. As a result, classes consisting of hybrid objects are formed.

The problem mentioned in Section 3 is solved using the proposed algorithm.

## Results and Discussion

In the experimental tests, classes with fewer than 20 objects were selected from the general educational sample, on the basis of which synthetic objects were created.

The results obtained on the basis of the algorithm for creating a synthetic training sample based on the average similarity of objects in the class are reflected in Table 3. Based on the above procedure, the following Table 3 was completed.

The results of this experiment are aimed at assessing the effectiveness of the algorithm for creating a synthetic training sample, describing the number of objects belonging to different classes, their degree of similarity, and changes after the introduction of synthetic objects. Variant 1 experimental results demonstrate the algorithm's effectiveness in creating a synthetic training sample and highlight class composition, similarity among objects, and modifications resulting from the inclusion of synthetic objects. Variant 2 experimental results indicate that the algorithm effectively generates a synthetic training sample and modifies class composition and object similarity. The number of real objects in classes varies, and the degree of similarity between them is different. After the introduction of synthetic objects, these similarities increased accordingly. Initially, the average degree of similarity between classes was 52.77 %, but after implementing the synthetic learning sampling algorithm, the average degree of similarity between classes increased to 70.63 % (Fig. 1).

For a comparative assessment of the reliability of the developed algorithm, a classification of the created hybrid training sample using popular machine learning methods was carried out. The levels of recognition of each class by 3 different methods are presented in detail in Table 3.

Based on the results of the experiment, it was established that the reliability of the training sample obtained using the developed algorithm for creating a synthetic training sample is high. The proposed algorithm achieved high reliability in generating the synthetic training sample, with overall accuracy of 95.8 % and multiple classes reaching 100 % precision, surpassing decision tree and KNN methods.

## Conclusion

The identification of cotton varieties is one of the most pressing issues in agriculture today. This allows not only to increase the effectiveness of agrotechnical measures, but also to ensure maximum yields by selecting optimal varieties in the context of regions. The use of modern artificial intelligence technologies significantly improves the process of identifying and regionalising cotton varieties.

The use of modern artificial intelligence, as implemented in the proposed algorithm, significantly enhances the identification and regionalisation of cotton varieties, achieving 95.8 % overall accuracy and 100 % precision in several classes. In the study, an algorithm for creating a synthetic training sample was developed to fill the gap in data on low-sample varieties under conditions of insufficient actual data, especially in some regions.

The proposed algorithm successfully generated high-quality synthetic data for cotton varieties. Based on this data, artificial intelligence models were built that correctly identify varieties, and with their help, it became possible to identify varieties early and optimise them by region. Based on this data, artificial intelligence models were developed that achieved an overall accuracy of 95.8 % in identifying cotton varieties, enabling early identification and regional optimisation. The research results are important in the direction of increasing the accuracy of models through the use of synthetic data and supporting reliable decision-making for farmers and agronomists, especially in cases where the data collected in real field conditions are insufficient. The table above shows the accuracy (precision) indicators for the three compared approaches (decision tree, KNN, and the proposed new algorithm), which shows that the results of the proposed approach achieved the highest accuracy.

In conclusion, the developed new algorithm allows for the correct identification of cotton varieties, assessment of their significance, and improvement of samples by dividing them into clusters corresponding to the corresponding regions. This approach can be widely used in breeding processes, regionalisation and determining the optimal placement of varieties.

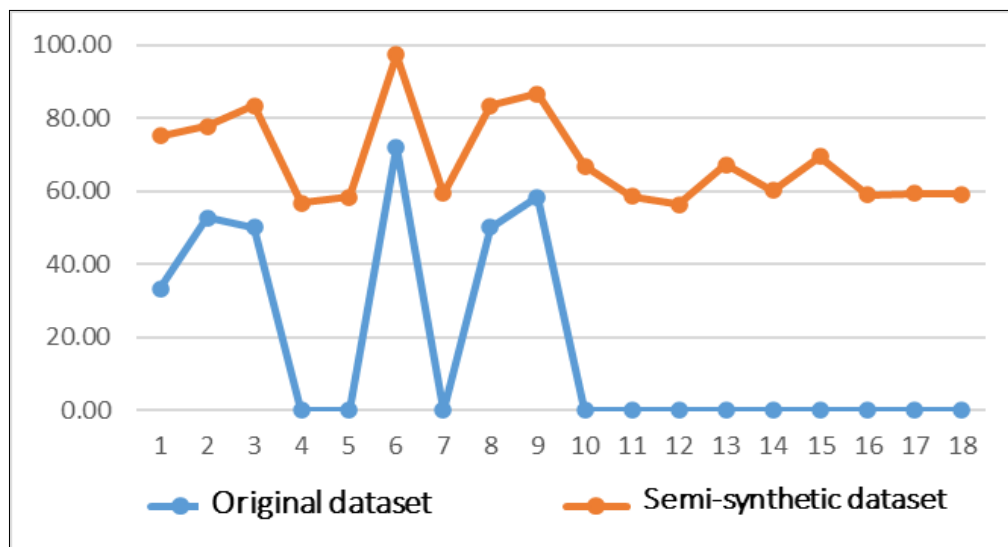


Fig. 1. Change in average similarity levels of class objects.

Table 3. Three levels of class recognition methods

No.	Decision tree (Precision)	KNN (Precision)	Proposed algorithm (Precision)
Class 1	62 %	100 %	83 %
Class 2	100 %	57 %	97 %
Class 3	88 %	100 %	88 %
Class 4	100 %	50 %	100 %
Class 5	83 %	100 %	100 %
Class 6	71 %	42 %	83 %
Class 7	100 %	100 %	100 %
Class 8	100 %	100 %	100 %
Class 9	86 %	100 %	83 %
Class 10	88 %	100 %	100 %
Class 11	100 %	80 %	100 %
Class 12	78 %	88 %	92 %
Class 13	100 %	100 %	100 %
Class 14	86 %	100 %	100 %
Class 15	100 %	100 %	100 %
Class 16	80 %	100 %	100 %
Class 17	80 %	75 %	100 %
Class 18	80 %	100 %	100 %
Accuracy	87 %	84.4 %	95.8 %

## Acknowledgements

The authors express their sincere gratitude to the management and academic staff of the Tashkent University of Information Technologies named after Muhammad al-Khwarizmi for their continuous support and for providing the necessary research environment and resources. Special thanks are extended to the Center for Innovative Developments and Consultations of the Andijan Institute of Agriculture and Agrotechnologies for their cooperation in implementing and testing the proposed algorithm in practical settings.

The authors also acknowledge the invaluable contributions of the “Pakhtakul istiqboli” farm of the Council of Dehkan Farms and Household Plot Owners in Shakhrikhan District, as well as the Council of Dehkan Farms and Household Plot Owners of Andijan Region, for granting access to real agricultural data and field information, which played a crucial role in validating the research outcomes.

The authors are grateful to all colleagues and experts who provided constructive feedback, methodological guidance, and technical assistance throughout the course of this study. Their insights significantly enhanced the quality and effectiveness of the developed algorithm.

## Authors' contributions

NAK conceptualised the research idea, supervised the overall study, and provided critical revisions to the manuscript. AOK contributed to the development of the algorithm, performed data preprocessing and carried out the experimental analysis. ASU collected and organised the dataset, conducted the literature review and assisted in interpreting the results. All authors participated in drafting the manuscript. All authors read and approved the final manuscript.

## Compliance with ethical standards

**Conflict of interest:** The authors declare that they have no conflict of interest.

**Ethical issues:** None

## References

- Fränti P, Sieranoja S. K-means properties on six clustering benchmark datasets. *Appl Intel*. 2018;48(12):4743–759. <https://doi.org/10.1007/s10489-018-1238-7>
- Qi J, Yu Y, Wang L, Liu J. K\*-Means: An effective and efficient K-means clustering algorithm. 2016 IEEE international conferences on big data and cloud computing (BDCloud), social computing and networking (SocialCom), sustainable computing and communications (SustainCom). 2016. p. 242–49. <https://doi.org/10.1109/BDCloud-SocialCom-SustainCom.2016.47>
- Geeks for Geeks. DBSCAN clustering in ML: Density-based clustering. 2025. <https://www.geeksforgeeks.org/dbscan-clustering-in-ml-density-based-clustering/>
- Zhang R, Peng H, Dou Y, Wu J, Sun Q, Zhang J, et al. Automating DBSCAN via deep reinforcement learning. Proceedings of the 31<sup>st</sup> ACM international conference on information and knowledge management (CIKM 2022). Association for computing machinery (ACM). 2022. <https://doi.org/10.1145/3511808.3557220>
- Ala'raj M, Majdalawieh M, Abbod MF. Improving binary classification using filtering based on k-NN proximity graphs. *J Big Data*. 2020;7(1):1–18. <https://doi.org/10.1186/s40537-020-00332-9>
- Ignatyev NA. Structure choice for relations between objects in metric classification algorithms. *Pattern recognition and image analysis. Adv Math Theory Appl*. 2018;28(4):695–702. <https://doi.org/10.1134/S1054661818040097>
- Nishanov A, Ruzibaev O, Tran N. Modification of decision rules “Ball Apolonia” for the problem of classification. 2016 international conference on information science and communications technologies (ICISCT 2016). 2016. <https://doi.org/10.1109/ICISCT.2016.7777382>
- Nishanov A, Akbarova M, Tursunov A, Ollamberganov F, Rashidova D. Clustering algorithm based on object similarity. *J Math Mech Comput Sci*. 2024;123(3):108–20. <https://doi.org/10.26577/JMMCS2024-v123-i3-4>
- Nishanov A, Tursunov A, Ollamberganov F, Rashidova D. Algorithm for clustering different types of drugs affecting blood pressure. *J Math Mech Comput Sci*. 2025;125(1).
- Abdukurimov DT, Lukov MQ. Cotton breeding and seed production. Textbook. Tashkent. 2015. p. 331.
- Avliyoqulov AE, Akhmedov J, Nuriddinov A. Agrotechnical measures for cultivating cotton varieties. Tashkent. 2016. p. 4–56.
- State register of agricultural crops of the Republic of Uzbekistan. Varieties and hybrids recommended for cultivation. 2007–2017.
- State register of agricultural crops of the Republic of Uzbekistan. Varieties and hybrids recommended for cultivation. 2008–2018.
- State register of agricultural crops of the Republic of Uzbekistan. Varieties and hybrids recommended for cultivation. 2022.

## Additional information

**Peer review:** Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

**Reprints & permissions information** is available at [https://horizonepublishing.com/journals/index.php/PST/open\\_access\\_policy](https://horizonepublishing.com/journals/index.php/PST/open_access_policy)

**Publisher's Note:** Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Indexing:** Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc See [https://horizonepublishing.com/journals/index.php/PST/indexing\\_abstracting](https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting)

**Copyright:** © The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited (<https://creativecommons.org/licenses/by/4.0/>)

**Publisher information:** Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.