ACTA IMEKO
ISSN: 2221-870X
September 2025, Volume 14, Number 3, 1 - 8

Algorithms for clustering, classification and selection of informative features for early detection of brain cancer

Akhram Nishanov[1], Fayzulla Ollamberganov[1], Malika Khasanova[2], Farkhod Mengturayev[3], Uktamjon Allayarov[4]

[1] *Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent 100084, Amir Temur Avenue 108, Uzbekistan*
[2] *Tashkent Medical Academy, Tashkent 100109, 2 Farabi Street, Uzbekistan*
[3] *Denov Institute of Entrepreneurship and Pedagogy, Denov 190507, 112 Garden Street, Uzbekistan*
[4] *Termez Branch of Tashkent Medical Academy, Surkhandaryo 132000, 64 Islam Karimov Street, Uzbekistan*

ABSTRACT
Early detection of brain cancer is crucial for improving treatment outcomes and patient survival rates. This study proposes a set of algorithms for clustering, classification, and selection of informative features that can aid in the early diagnosis of brain tumours. In collaboration with medical experts, a dataset comprising 218 patient records and 82 clinical and symptomatic features was constructed. Through clustering analysis, the dataset was grouped into four diagnostic classes: (1) Anaplastic astrocytoma in the right frontal region, (2) Adenoma in the chiasmal-sellar region, (3) Glioblastoma in the right frontal lobe, and (4) Meningioma in the right frontal lobe. A feature selection algorithm was then applied to identify the most diagnostically relevant attributes. From the initial 82 features, 19 were determined to be strongly indicative of the disease classes. Further refinement using the proposed algorithm resulted in a subset of six highly informative features, which successfully differentiated the classes with a minimum object similarity threshold of 65 %. The approach demonstrates the potential of data-driven techniques in enhancing the accuracy and efficiency of brain cancer diagnostics, offering a scalable method that could be integrated into clinical decision-support systems.

Section: RESEARCH PAPER

Keywords: brain cancer; early detection of diseases; selection of informative features; clustering; classification

Citation: A. Nishanov, F. Ollamberganov, M. Khasanova, F. Mengturayev, U. Allayarov , Algorithms for clustering, classification and selection of informative features for early detection of brain cancer, Acta IMEKO, vol. 14 (2025) no. 3, pp. 1-8. DOI: 10.21014/actaimeko.v14i3.2053

Section Editor: Francesco Lamonaca, University of Calabria, Italy

Received January 10, 2025; **In final form** July 29, 2025; **Published** September 2025

Corresponding author: Akhram Nishanov, e-mail: nishanov_akram@mail.ru

1. INTRODUCTION

Brain cancer remains one of the most aggressive and lethal forms of cancer, posing a significant threat to global public health. According to the International Agency for Research on Cancer (IARC), approximately 330,000 new cases of brain and Central Nervous System (CNS) cancers were diagnosed worldwide in 2022, accounting for about 1.7 % of all cancer cases. These cancers are associated with high mortality, resulting in nearly 250,000 deaths annually. Malignant brain tumours, in particular, show some of the lower survival outcomes, with a five-year survival rate of only 36 %. Among them, glioblastoma is especially severe, with a median survival time of just 15 months, even when standard treatment protocols are applied.

Given the high mortality and limited effectiveness of current treatments, early detection plays a critical role in improving patient outcomes. Research shows that diagnosing brain cancer at an early stage significantly enhances the likelihood of successful treatment, thereby improving both survival rates and quality of life. However, traditional diagnostic techniques, such as Magnetic Resonance Imaging (MRI) and biopsy, are often time-consuming, costly, and reliant on specialised expertise for accurate interpretation. These challenges underscore the need for advanced computational tools, particularly those based on Machine Learning (ML) and Artificial Intelligence (AI), to streamline and improve the accuracy of early diagnosis. This study explored algorithmic approaches to clustering, classification, and the selection of informative features for the early detection of brain cancer. By leveraging data-driven methodologies, the research aims to identify key predictive biomarkers, enhance diagnostic precision, and develop effective

classification models that can be integrated into clinical workflows. The proposed approach addresses the limitations of traditional diagnostic methods and offers a foundation for future AI-assisted diagnostic solutions in oncology.

Developing an effective algorithm for selecting a subset of informative features for early brain cancer detection involves multiple stages [1]–[5]. Existing literature highlights various methods and algorithms used to extract and select significant features, particularly those derived from MRI imaging, for the early classification of brain cancer. These studies explore a range of technical challenges, methodologies, and comparative evaluations that contribute to the accurate identification of relevant diagnostic characteristics. The following sections outline the general process and methodology employed in this research.

1.1. Data collection

A comprehensive dataset containing both brain cancer cases and healthy control samples is essential for the development of effective classification algorithms [6]–[10]. Numerous studies have emphasised the significance of data collection in brain cancer research, proposing various datasets that include clinical and imaging information from both affected and unaffected individuals. These studies provide valuable insights into dataset composition, including the number of samples, the types of features recorded, and the methodologies used for data acquisition, pre-processing, and interpretation. Each sample in such datasets is typically followed by a set of features that describe relevant characteristics of brain function, structure, or pathology, forming the foundation for reliable algorithmic analysis.

1.2. Pre-processing of data

The collected data must undergo purification and pre-processing to ensure its quality and suitability for analysis [11]–[15]. The referenced literature explores various techniques for cleaning and preparing brain cancer datasets, emphasising the importance of data integrity in achieving accurate analytical results. Key pre-processing steps include quality assurance, correction of inconsistencies, segmentation error detection, and normalisation of feature values. These processes address common issues, such as duplicate records, missing values, and data variability. The goal is to enhance the reliability and consistency of the dataset, thereby improving the effectiveness of subsequent classification and diagnostic modelling efforts.

1.3. Feature extraction (selection of a set of informative features)

Relevant features must be extracted from the pre-processed data to enable accurate classification and prediction tasks [16]–[20]. The referenced literature discusses a range of techniques for feature extraction in brain cancer analysis, including statistical analysis, texture-based descriptors, machine learning methods, and deep learning architectures. These approaches aim to identify the most informative characteristics that differentiate between disease types or stages. Feature extraction may involve dimensionality reduction techniques, such as Principal Component Analysis (PCA), as well as feature selection methods based on correlation analysis, cross-validation, or statistical significance testing. Given its critical role in improving model performance and interpretability, the development of robust algorithms for feature selection remains a highly relevant and active area of research in medical data analysis.

2. GENERAL CONCEPTS, DESIGNATIONS AND DEFINITIONS

Brain cancer, also known as a brain tumour, refers to the abnormal and uncontrolled growth of cells within the brain. These cells multiply rapidly, forming masses that interfere with normal brain function. Brain tumours can be classified into two broad categories: primary tumours, which originate in the brain, and metastatic tumours, which spread to the brain from other parts of the body.

The symptoms of brain cancer vary depending on the tumour's location, size, and type. Common symptoms include:

1) Headache: persistent or severe headaches that worsen over time;
2) Seizures: sudden, uncontrolled electrical activity in the brain;
3) Cognitive and memory issues: difficulty concentrating, memory loss, confusion, or impaired problem solving;
4) Personality or mood changes: irritability, depression, or noticeable changes in behaviours;
5) Vision and hearing problems: blurred or double vision, partial or complete hearing or vision loss;
6) Problems with arm and leg movements: weakness in limbs, difficulty with movement, balance, or coordination;
7) Nausea and vomiting: often accompanied by other symptoms such as headaches;
8) Fatigue: persistent tiredness or lack of energy;
9) Speech difficulties: trouble speaking, finding words, or understanding language;
10) Appetite or weight changes: unexplained weight loss or gain, or reduced appetite.

It is worth noting that many of these symptoms may be caused by other, less serious medical conditions. However, persistent or worsening symptoms warrant prompt medical evaluation and diagnosis.

In this study, an initial review of 541 medical records of patients diagnosed with brain cancer was conducted. Following a thorough assessment in collaboration with specialists from the Republican Scientific and Practical Medical Center of Oncology and Radiology of the Republic of Uzbekistan, 238 records were deemed suitable for inclusion in the study. Further collaboration with medical professionals from the Surkhandarya branch of the same institution and the Tashkent Medical Academy led to the identification of 19 key symptoms relevant to the diagnosis and classification of brain cancer (as listed in Table 1).

Table 1. Acceptable values of sympathocomplexes in brain cancer.

Features	Symptoms of brain cancer (name of features)	Acceptable values of features
x^1	Headache	1 - no headache 2 - headache 3 - strong headache - constant headaches
x^2	Dizziness	0 - no dizziness 1 - vertigo
x^3	Nausea (vomiting)	1 - nausea, no vomiting 2 - there is nausea 3 - there is vomiting 4 - there is vomiting and nausea
x^4	Convulsions	1 - no convulsions 2 - the convulsion occurred once 3 - the convulsions occurred once or twice a week 4 - the convulsions occurred once or twice a day

x^5	Standing in the Romberg pose (standing in the pose)	1 - can stand in the Romberg pose 2 - swings in the Romberg pose 3 - can't stand in the Romberg pose 4 - deviates to the right in the Romberg pose 5 - deviates to the left in the Romberg pose
x^6	General weakness (helplessness)	0 - no 1 - yes
x^7	Rapid fatigue	0 - does not get tired 1 -gets tired
x^8	Lack of strength (weakness) in the arm and leg	1 - hands and feet are healthy 2 - insufficient strength in the left arm and left leg 3 - insufficient strength in the right arm and right leg 4 - insufficient strength in the arms and legs 5 - insufficient strength of the left hand 6 - insufficient strength of the right hand 7 - insufficient strength in the left leg 8 - insufficient strength in the right leg
x^9	Memory weakening	0 - memory is not weakened 1 - memory is weakened can't remember
x^{10}	Loss of consciousness	0 - conscious 1 - unconscious
x^{11}	Hearing level	1 - hears in both ears 2 - the ears are weak to hear 3 - hearing loss in the right ear 4 - hearing loss in the left ear 5 - not audible in both ears 6 - deaf from birth
x^{12}	Speech disorder	1 - speech is not broken 2 - speech is distorted 3 - speech disorder (1 year before the onset of the disease or congenital)
x^{13}	Numbness in the arms and legs	1 - no numbness in the arms and legs 2 - there is a numbness on the leg 3 - there is a numbness on the right leg 4 - the left leg has numbness 5 - numb hands 6 - the right hand becomes numb 7 - the left arm becomes numb 8 - the right arm and right leg become numb 9 - the left arm and left leg become numb 10 - both arms and legs become numb 11 - both legs go numb 12 - both hands go numb
x^{14}	Restriction of movements in the legs and arms	1 - there are no restrictions on movements in the arms and legs 2 - limited movement of arms and legs 3 - limited movement of the left hand 4 - limited movement of the right hand 5 - limited movement of the left leg 6 - limited movement of the right leg 7 - slightly limited movement of the right arm and right leg 8 - slightly limited movement of the left arm and left leg 9 - limited movement of the left arm and left leg 10 - limited movement of the right arm and right leg
x^{15}	Central facial nerve paralysis	1 - the front area is unchanged 2 - central facial nerve paralysis is diagnosed on the left side 3 - central facial nerve paralysis is diagnosed on the right side 4 - central facial nerve paralysis is detected on both sides
x^{16}	Somnolence	1 - normal sleep 2 - somnolence 3 - from insomnia
x^{17}	Change of consciousness	1 - into consciousness 2 - consciousness is dim 3 - not conscious
x^{18}	Ingestion and swallowing activity	1 - the activity of swallowing is preserved 2 - difficulty swallowing 3 - can't swallow
x^{19}	Reduced vision	1 - visual acuity in both eyes is good 2 - decreased visual acuity in both eyes 3 - decreased visual acuity in the left eye 4 - decreased visual acuity in the right eye 5 - no vision in the left eye 6 - no vision in the right eye 7 - no vision in both eyes 8 - darkening of the front of the eye 9 - two eyes see things in a double image 10 - vision in the left eye congenital blindness 11 - vision in the right eye congenital blindness

The first column of Table 1 presents the list of symptoms used in the study. Each symptom can be interpreted in the context of a feature vector representation of a medical record. For example, x^{18} denotes the 18th feature (or symptom) in the vector $x = (x^1, x^2, \dots x^{18}, x^{19})$, which corresponds to the symptom "Ingestion and swallowing activity." The second column provides a detailed description of each symptom, while the third column includes the nominal values assigned to symptom severity or type, along with their textual interpretations.

Table 2 outlines the classification of brain cancer types. It includes the disease class labels, corresponding tumour types, and their associated primary and auxiliary symptoms.

The first column of Table 2 presents the classification of brain cancer into diagnostic classes. For example, the first entry refers to the class of diseases identified as "Anaplastic astrocytoma of the right frontal region of the brain", denoted as X_1. Each class X_p consists of a set of patient records (objects) $x_{pi} \in X_p$, where $p \in \{1, 2, 3, 4\}$ indicates the class index and $i \in \{1, \dots, m_p\}$ represents the number of objects (or patients) within that class. The second column of Table 2 provides the textual descriptions of the disease types corresponding to each class. For instance, class X_2 represents patients diagnosed with Adenoma of the chiasmal-sellar region of the brain. The third column of Table 2 lists the primary symptoms and diagnostic features that define

Table 2. Classification of symptom complexes for brain cancer diseases.

Class	Type of disease	Primary symptoms	Auxiliary symptoms
X_1	Anaplastic astrocytoma of the right frontal region of the brain	$x^1, x^2, x^4, x^5, x^6, x^7, x^8, x^9, x^{19}$	$x^{10}, x^{11}, x^{12}, x^{13}, x^{14}, x^{15}, x^{16}, x^{17}, x^{18}$
X_2	Adenoma of the chiasmal-sellar region of the brain	$x^1, x^2, x^4, x^5, x^6, x^7, x^{10}, x^{11} x^{12}, x^{13}, x^{19}$	$x^8, x^9, x^{14}, x^{15}, x^{16}, x^{17}, x^{18}$
X_3	Glioblastoma of the right area of the forehead of the brain	$x^1, x^2, x^4, x^5, x^6, x^7, x^{10}, x^{13}$	$x^8, x^9, x^{11}, x^{12}, x^{14}, x^{15}, x^{16}, x^{17}, x^{18}, x^{19}$
X_4	Meningioma of the right area of the forehead of the brain	$x^1, x^2, x^3, x^4, x^5, x^6, x^8, x^{14}, x^{15}, x^{16}, x^{17}, x^{18}, x^{19}$	$x^7, x^9, x^{10}, x^{11}, x^{12}, x^{13}$

each disease class. The fourth column of Table 2 includes auxiliary symptoms that may support or supplement the main diagnostic indicators.

To aid interpretation, a brief explanation of key medical terms related to brain cancer is provided below:

Malignant: Refers to cancerous tumours capable of invading nearby tissues and spreading to other parts of the body.

Benign: Non-cancerous tumours that do not invade surrounding tissues or metastasise.

Metastasis: The process through which cancer cells spread from the original tumour site to other areas via the bloodstream or lymphatic system.

Glioma: The most common type of brain tumour, originating from glial cells that support and protect neurons.

Astrocytoma: A subtype of glioma arising from astrocytes—star-shaped glial cells in the brain.

Meningioma: A typically benign tumour that develops from the meninges, the protective membranes surrounding the brain and spinal cord.

Biopsy: A medical procedure involving the extraction and microscopic examination of tissue to determine if a tumour is benign or malignant.

Radiation therapy: A treatment method that uses high-energy radiation to destroy or shrink cancer cells.

Chemotherapy: The use of drugs to eliminate or control the growth of cancer cells throughout the body.

Surgery: The physical removal of a tumour through operative techniques.

Treatment decisions for brain cancer depend on various factors, including the tumour's type, size, and location, as well as the overall health of the patient. A multidisciplinary team, typically composed of neurosurgeons, radiation oncologists, and medical oncologists, collaborates to provide the most effective care.

3. PROBLEM STATEMENT

Let N denote the number of features (symbols) that characterise each object $x_i \in X$, where $i = \overline{1,M}$, and X is the set of all objects in the dataset. Each object x_i is represented in the N-dimensional symbolic feature space as:

$$x_i = (x_i^1, x_i^2, \dots, x_i^N). \tag{1}$$

The goal is to construct training subsets (classes) from the general dataset X. These classes are defined as $x_{p1}, x_{p2}, \dots, x_{pm_p} \in X_p, p = \overline{1,r}$, and each class contains m_p objects. Each object x_{pi} in class X_p is represented in the N-dimensional symbolic space as:

$$x_{pi} = \left(x_{pi}^1, x_{pi}^2, \dots x_{pi}^N\right), i = \overline{1, m_p}\,. \tag{2}$$

The complete set of class samples is then expressed as the union:

$$\bigcup_{p=1}^{r} X_p = \left\{x_{p1}, x_{p2}, \dots, x_{pm_p}\right\}. \tag{3}$$

This formulation provides the foundation for defining similarity-based classification and clustering tasks within the N-dimensional symbolic feature space [21]–[26].

Task 1. Based on the given X of the general selection of the educational sample X_p class, $p = \overline{1,r}$ be formulated in such a way that the degrees of similarity of the resulting class objects are not less than a predetermined number of δ.

$\lambda = (\lambda^1, \lambda^2, \dots, \lambda^N)$, $\mu = (\mu^1, \mu^2, \dots, \mu^N)$ are vectors; let their components take the values of zero or one. Here, the components of the $\lambda = (\lambda^1, \lambda^2, \dots, \lambda^N)$ indicate which symbol is or is not involved in the calculations. If $\lambda^j = 1$, then the J-component is involved in computational work, otherwise, if $\lambda^j = 0$, then the J-component is not involved in computational work. There is, we must find such a vector $\lambda_{\text{optimal}} = (\lambda^1, \lambda^2, \dots, \lambda^N)$ that the objects of the class have a degree of similarity of at least 55 %.

Here, $\lambda_{\text{optimal}} = (\lambda^1, \lambda^2, \dots, \lambda^N)$ equal together components of the vector are understood as informative symbols, and values equal to zero as non-informative symbols.

Similarly, we can define this vector $\mu = (\mu^1, \mu^2, \dots, \mu^N)$. The components of this vector are counted as follows. Suppose we are given two $x_i, x_k \in X$ objects, then the components of the vector $\mu(x_i, x_k)$ are calculated as:

$$\mu^j(x_i, x_k) = \begin{cases} 1 \text{ if } \left|x_i^j - x_k^j\right| = 0, j = \overline{1,N} \\ \\ 0 \text{ else.} \end{cases} \tag{4}$$

Here, $\mu^j(x_i, x_k) = 1$ if the corresponding components of the two objects $x_i^j = x_k^j$ are mutually equal.

There, vector $\mu(x_i, x_k)$ will be $\mu^j = 1$ if the corresponding components of the two x_i, x_k objects are $x_i^j = x_k^j, i \neq k.$ are mutually equal.

Thus, it is possible to uniquely define the vector $\mu(x_i, x_k)$ for two arbitrary objects of the training sample. Let us suppose that through the value $\varkappa = \sum_{j=1}^{N} \mu^j(x_i, x_k)$ denote the similarity coefficient objects x_i, x_k. This value indicates the number of x_i, x_k object components that are identical.

Let us determine the degrees of similarity of these two objects through $\nu(x_i, x_k)$ and calculate it as a percentage as follows:

$$\nu(x_i, x_k) = \frac{\varkappa \cdot 100\ \%}{N} \tag{5}$$

$$X_p = \left\{\forall x_{pi}, x_{pk} : \nu\left(x_{pi}, x_{pk}\right) \geq \delta\,, i \neq k, p = \overline{1,r}\right\} \tag{6}$$

more than a number, that is

$$\begin{cases} \nu\left(\lambda, x_{pi}, x_{pk}\right) \geq \delta \\ \lambda \in \Lambda^{\ell} = \left\{ \lambda : \sum_{j=1}^{N} \lambda^j = \ell\,, \lambda^j \in \{0,1\}, j = \overline{1,N} \right\} \end{cases} \tag{7}$$

requires solving the problem.

This condition requires that the similarity within each class must meet or exceed a predefined threshold $\delta \geq 65\ \%$, regardless of the number of object pairs in the class. Accordingly, the classes are constructed so that the degree of similarity between any two objects within the same class is not less than the specified threshold δ.

Task 2. Suppose that two $X_p$and X_q criteria are given, providing the choice of components of the λ vector, which gives the difference between the objects of class $\boldsymbol{J}(\boldsymbol{\lambda}, X_p, X_q)$. Let us, from a given N-dimensional feature space, pass into such an $\ell \ll N$-dimensional space that, in the resulting feature space, objects of two classes clearly differ from each other.

To solve the second problem, we introduce the following notation:

1) $I(\lambda, X_p)$- through this functional with respect to the vector λ, the average degree of all interobject similarities of the class;
2) And through the functional $I(\boldsymbol{\lambda}, X_q)$- relative to the vector $\boldsymbol{\lambda}$, the average value of the degrees of mutual similarity of all objects of class X_q;
3) And using the criterion $I(\boldsymbol{\lambda}, X_p, X_q)$, we determine the average similarity of objects of classes X_p and X_q.

Consider the following optimisation problem:

$$\begin{cases} J(\lambda, X_p, X_q) = \dfrac{I(\lambda, X_p) + I(\lambda, X_q)}{I(\lambda, X_p, X_q)} \to \max \\ \lambda \in \Lambda^{\ell} = \left\{ \lambda : \sum_{j=1}^{N} \lambda^j = \ell \,, \lambda^j \in \{0,1\}, j = \overline{1,N} \right\}. \end{cases} \tag{8}$$

The content of this optimisation problem includes the following.

The sum in the figure of the functional, which gives in relation to the vector λ the tendency of the sums of the degrees of similarity of objects within each class to the maximum, and in relation to this vector λ, which gives in the denominator the tendency of the degrees of similarity between objects of two classes to the minimum, is equal to 1 component of the vector λ, are understood as a set of symptoms indicating the differences between objects of two classes from each other, i.e. a set of informative features. Therefore, the solution of this optimisation problem provides a transition from an N-dimensional character space to such an $\ell \ll N$-dimensional character space.

Task 3. Suppose an unknown object $w = (w^1, w^2, \dots, w^N)$ is given in an N-dimensional feature space. The task is to determine whether this object belongs to class X_p or X_q. To make this determination, the following similarity function is used:

$$I(\lambda, w, X_p)\,. \tag{9}$$

This functional represents the average similarity between the unknown object w, projected onto the feature subset defined by the vector λ, and all objects in class X_p. Similarly, the same functional $I(\lambda, w, X_q)$ is computed for class X_q.

The classification rule is defined as follows:

- If $I(\lambda, w, X_p) > I(\lambda, w, X_q)$, then the object w is considered to belong to class X_p;
- Otherwise, if $I(\lambda, w, X_p) < I(\lambda, w, X_q)$, then the object w is classified as a member of class X_q.

4. STAGES OF DIAGNOSIS OF BRAIN CANCER

The diagnosis of brain cancer typically involves multiple stages, which may vary depending on the diagnostic approach, available technologies, and the individual condition of the patient. A general overview of the key stages involved in the diagnostic process follows below:

Symptom identification: The initial stage involves recognizing symptoms that may suggest the presence of a brain tumour. These symptoms can vary widely depending on the tumour's size, location, and growth rate. Common signs include persistent headaches, seizures, visual or auditory disturbances, cognitive impairment, mood or personality changes, balance and coordination difficulties, and episodes of nausea or vomiting.

Medical history and physical examination: If brain cancer is suspected, the clinician begins by taking a comprehensive medical history to evaluate the nature, duration, and progression of the symptoms, as well as any relevant risk factors. A physical examination is then conducted, which typically includes a neurological assessment of reflexes, muscle strength, coordination, and mental status.

Neuroimaging (neurovisualisation): One of the most critical diagnostic steps is neuroimaging, which provides detailed visualisation of the brain. Magnetic Resonance Imaging (MRI) is the most commonly used modality, offering high-resolution images that help detect the presence, size, and precise location of tumours. In certain cases, contrast-enhanced MRI is performed to better delineate tumour boundaries. Computed Tomography (CT) scans may also be employed, especially in emergency settings or when MRI is contraindicated.

Biopsy: To confirm the diagnosis and determine the tumour type, a biopsy is often required. This involves surgically extracting a small tissue sample from the tumour, typically guided by MRI or CT imaging. The sample is then examined by a pathologist under a microscope, often supplemented by immunohistochemical staining and genetic profiling to identify specific tumour markers.

Molecular testing: Following biopsy, molecular testing may be conducted to identify genetic mutations or molecular alterations in the tumour cells. These tests provide critical information for treatment planning and prognosis, particularly in the context of personalized or targeted therapies.

Lumbar puncture (spinal tap): In some cases, a lumbar puncture is performed to collect Cerebrospinal Fluid (CSF) for analysis. This procedure can help detect cancer cells that have spread to the CSF, or identify other abnormalities that inform the diagnosis and staging.

Additional diagnostic tests: Depending on the patient's condition, further diagnostic tests may be necessary to assess tumour progression and its effect on brain function. These may include functional MRI (fMRI), Positron Emission Tomography (PET), Electroencephalogram (EEG), or cerebral angiography. Each test provides unique insights into the metabolic, electrical, or vascular characteristics of the brain.

It is important to note that the sequence and selection of these diagnostic procedures are tailored to each patient's clinical presentation and guided by the judgment of a multidisciplinary medical team.

Based on the results of the first stage, Table 3 has been compiled to present the solution to the initial classification problem. The first column lists the diagnostic classes, while the second column specifies the corresponding disease types associated with each class. The remaining columns enumerate the symptom values for each object within the class. Although the table provides detailed numerical data on symptom distribution across different classes, the density of information makes it challenging to visually detect patterns. To address this, a heat map representation of Table 3 is provided in Figure 1, allowing for easier interpretation of symptom clustering and class differentiation.

Table 4 presents the solution to the second problem, which involves the identification of a set of informative features under the conditions $\ell = 6$ and $\delta = 65\ \%$. The resulting informative feature set includes the following symbolic features: $x^2, x^6, x^{11}, x^{12}, x^{18}, x^{19}$. In Table 4, the first column identifies

Table 3. Results of the first stage.

№	Type of disease	x^1	x^2	x^3	x^4	x^5	x^6	x^7	x^8	x^9	x^{10}	x^{11}	x^{12}	x^{13}	x^{14}	x^{15}	x^{16}	x^{17}	x^{18}	x^{19}
1	Anaplastic astrocytoma of the right area of the forehead of the brain	2	1	1	1	2	1	1	1	0	0	1	1	1	10	1	3	2	1	2
		3	1	3	3	3	1	1	2	1	1	4	2	9	8	2	2	3	2	2
		3	1	2	4	5	1	1	3	0	1	1	1	8	2	1	2	3	2	1
		2	0	1	2	2	1	1	4	1	0	2	1	1	1	2	1	1	1	1
		4	1	2	3	4	1	1	3	0	0	1	1	10	3	1	1	2	1	1
		3	1	4	1	2	1	1	8	0	0	3	1	9	1	3	2	1	2	1
		2	1	1	4	2	1	1	5	1	1	5	2	7	6	1	1	2	2	2
		2	1	2	2	2	1	1	7	0	0	1	1	1	7	2	1	2	2	2
		2	1	4	2	5	1	1	2	0	0	4	1	5	9	1	1	2	2	1
		4	1	3	3	4	1	1	6	1	0	3	1	4	5	3	1	1	1	1
2	Adenoma of the chiasmal-cellular region of the brain	4	1	2	2	2	1	1	1	0	1	2	2	10	1	1	1	1	1	2
		4	1	1	3	3	1	1	2	0	1	1	1	9	1	1	2	1	1	8
		4	1	1	3	2	1	1	2	0	1	1	1	8	1	1	1	2	1	2
		1	1	1	2	5	1	0	2	0	1	1	1	9	8	1	3	1	1	3
		4	1	2	2	2	1	1	3	1	0	1	1	8	1	1	3	2	1	2
		2	1	2	3	2	1	1	8	1	1	2	1	9	1	1	2	3	1	2
		2	1	2	2	2	1	1	1	0	0	1	1	8	1	1	2	1	1	2
		2	1	4	4	2	1	1	2	0	1	1	1	8	9	1	2	2	1	4
		2	1	2	3	2	1	1	2	0	0	1	1	11	6	2	2	2	1	2
		2	1	2	4	2	1	1	4	0	1	1	1	9	8	1	1	1	1	2
3	Glioblastoma of the right area of the forehead of the brain	4	1	4	1	2	1	1	2	0	0	1	1	8	2	1	2	1	2	1
		2	1	2	1	2	1	1	2	0	0	1	1	9	8	1	2	1	1	2
4	Meningioma of the right area of the forehead of the brain	4	1	2	2	2	1	0	2	0	0	1	1	9	8	1	2	1	1	1
		4	1	2	1	2	1	1	2	0	0	1	1	9	8	1	2	3	1	1
		4	1	2	1	2	1	1	2	0	0	5	1	10	8	2	2	3	1	1
		4	1	1	1	2	1	1	2	0	0	1	1	8	8	3	3	3	1	2
		4	1	2	3	4	1	0	2	1	1	1	1	11	3	3	2	3	1	2

the diagnostic classes, the second column lists the corresponding disease types, and the third column outlines the most informative feature combinations that characterise each class. Additionally, representative diagnostic objects that contributed to the formation of these classes are specified, highlighting the features most relevant to class distinction.

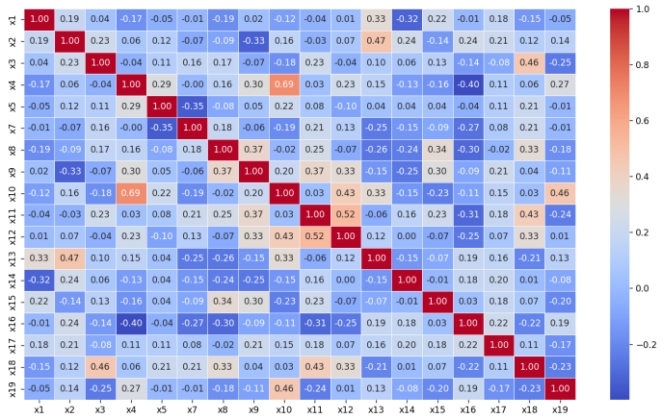

Figure 1. Heatmap of Table 3.

It was determined that the first class contained 10 representative objects, the second class included 5, and the third and fourth classes each comprised 2 objects.

5. CONCLUSION

This study concludes that classification models based on informative feature selection offer a promising approach for the early diagnosis of brain cancer. Compared to traditional diagnostic methods, these models demonstrate advantages in terms of efficiency, cost-effectiveness, and scalability. The proposed method is currently being utilised at the Karakalpak branch of the Republican Specialised Scientific and Practical Medical Center of Oncology and Radiology for the diagnosis of new patients. Field specialists have reported diagnostic accuracy rates ranging from 86 % to 91 %.

The integration of machine learning algorithms with medical diagnostic tools represents a significant advancement toward precision medicine and AI-driven healthcare. Moving forward, collaboration among clinicians, data scientists, and regulatory bodies will be essential to foster an interdisciplinary approach that supports the refinement and successful implementation of such systems in clinical settings.

Table 4. Solution to the second problem.

Class	Type of disease	Main features
X_1	Anaplastic astrocytoma of the right frontal region of the brain	1,1,1,1,1,1; 1,1,1,1,1,2; 1,1,3,1,1,2; 1,1,1,1,1,4; 1,1,1,1,1,3; 1,1,1,1,1,9; 0,1,1,1,1,2; 1,0,1,1,1,2; 1,1,1,1,1,7; 1,1,1,1,1,10.
X_2	Adenoma of the chiasmal-cellular region of the brain	1,1,1,2,1,2; 0,1,1,2,1,2; 1,1,1,2,2,2; 1,1,2,2,1,2; 1,1,1,2,1,1.
X_3	Glioblastoma of the right area of the forehead of the brain	1,1,2,1,1,1; 1,1,2,2,1,1.
X_4	Meningioma of the right area of the forehead of the brain	1,1,2,1,1,2; 1,1,2,1,2,2.

By addressing the identified limitations and enhancing the existing framework, this study lays the groundwork for the development of next-generation diagnostic tools with the potential to transform the early detection and treatment of brain cancer.

REFERENCES

[1] A. Chaddad, R. Colen, Statistical feature selection for enhanced detection of brain tumor, Applications of Digital Image Processing XXXVII, vol. 9217, 2014, pp. 260-267.
DOI: 10.1117/12.2062143

[2] H. Khan, W. Jue, M. Mushtaq, M. Mushtaq, Brain tumor classification in MRI image using convolutional neural network, Mathematical Biosciences and Engineering, vol. 17, no. 5, 2020, pp. 6203-6216.
DOI: 10.3934/mbe.2020328

[3] V. Gladis, P. Rathi, S. Palani, Brain tumor MRI image classification with feature selection and extraction using linear discriminant analysis, Int. Journal of Information Sciences and Techniques, vol. 2, no. 4, 2012. pp. 131-146.
DOI: 10.5121/IJIST.2012.2413

[4] R. Kaifi, A Review of Recent Advances in Brain Tumor Diagnosis Based on AI-Based Classification, Diagnostics, vol. 33, no. 18, 2023, pp.1-32.
DOI: 10.3390/diagnostics13183007

[5] M. Sasikala, N. Kumaravel, Comparison of Feature Selection Techniques for Detection of Malignant Tumor in Brain Images, Annual IEEE India Conf. - Indicon, Chennai, India, 11-13 December 2005, pp. 212-215.
DOI: 10.1109/INDCON.2005.1590157

[6] B. H. Menze, A. Jakab, S. Bauer, (+ another 65 authors), The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), IEEE Transactions on Medical Imaging, vol. 34, no. 10, 2015, pp. 1993-2024.
DOI: 10.1109/TMI.2014.2377694

[7] P. Korfiatis, T. L. Kline, L. Coufalova, D. H. Lachance, I. F. Parney, R. E. Carter, J. C. Buckner, B. J. Erickson, MRI Texture Features as Biomarkers to Predict MGMT Methylation Status in Glioblastomas, Medical Physics, vol. 43, no. 6, part 1, 2016, pp. 2835-2844.
DOI: 10.1118/1.4948668

[8] S. Bauer, T. Fejes, J. Slotboom, R. Wiest, L.-P. Nolte, M. Reyes, Segmentation of Brain Tumor Images Based on Integrated Hierarchical Classification and Regularization, in MICCAI BraTS Workshop. Nice: Miccai Society, vol. 11, 2012. Online [Accessed 11 September 2025]
http://www2.imm.dtu.dk/projects/BRATS2012/BauerBRATS2012.pdf

[9] D. Zikic, B. Glocker, E. Konukoglu, A. Criminisi, C. Demiralp, (+ another 5 authors), Decision Forests for Tissue-Specific Segmentation of High-Grade Gliomas in Multi-Channel MR, Proc. of the Int. Conf. on Medical Image Computing and Computer-Assisted Intervention, 2012.
DOI: 10.1007/978-3-642-33454-2_46

[10] E. Panagiotaki, S. Walker-Samuel, B. Siow, S. P. Johnson, V. Rajkumar, R. B. Pedley, M. F. Lythgoe, D. C. Alexander, Noninvasive Quantification of Solid Tumor Microstructure using VERDICT MRI, Cancer Research, vol. 74, no. 7, 2014.
DOI: 10.1158/0008-5472.can-13-2511

[11] M. J. Clark, B. Vendt, K. Smith, (+ another 9 authors), The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository, Journal of Digital Imaging, vol. 26, no. 6, 2013, pp. 1045-1057.
DOI: 10.1007/s10278-013-9622-7

[12] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, (+ another 4 authors), Advancing the Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Scientific Data, vol. 4, 2017, 170117, pp. 1-13.
DOI: 10.1038/sdata.2017.117

[13] N. Tustison, B. Avants, P. Cook, Y. Zheng, A. Egan et al., N4ITK: Improved N3 Bias Correction, IEEE Transactions on Medical Imaging, vol. 34, no. 11, 2015, pp. 1310-1320.
DOI: 10.1109/TMI.2010.2046908

[14] S. Gangadhar, C. Naveena, S. Poornachandra, V. Aradhya, Preprocessing of MR Images for Effective Quantitative Image Analysis, Fifth International Conference on Image Information Processing (ICIIP), 2019, pp. 63-67.
DOI: 10.1109/ICIIP47207.2019.8985817

[15] A. Makropoulos, I. Gousias, C. Ledig, P. Aljabar, A. Serag et al., Automatic whole brain MRI segmentation of the developing neonatal brain, IEEE Transactions on Medical Imaging, vol. 33, no. 9, 2014, pp. 1818-1831.
DOI: 10.1109/tmi.2014.2322280

[16] J. Huang, N. Huang, L. Zhang, H. Xu, A method for feature selection based on the correlation analysis, Int. Conf. on Measurement, Information and Control, Harbin, China, 18-20 May 2012, pp. 529-532.
DOI: 10.1109/MIC.2012.6273357

[17] Sh. Mao, Ch. Zhang, N. Gao, Y. Wang, Y. Yang et al., A study of feature extraction for Alzheimer's disease based on resting-state fMRI, 39th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Korea (South), 11-15 July 2017, pp. 517-520.
DOI: 10.1109/EMBC.2017.8036875

[18] Z. Wang, X. Xiao, Z. Zang, K. He, F. Hu, A Radiomics Model for Predicting Early Recurrence in Grade II Gliomas Based on Preoperative Multiparametric Magnetic Resonance Imaging, Frontiers in Oncology, vol. 11, 2021, pp. 1-10.
DOI: 10.3389/fonc.2021.684996

[19] W. Shen, M. Zhou, F. Yang, C. Yang, J. Tian et al., Multi-Scale Convolutional Neural Networks for Lung Nodule Classification, Information Processing in Medical Imaging, 2015, pp. 588-599.
DOI: 10.1007/978-3-319-19992-4_46

[20] J. van Griethuysen, A. Fedorov, C. Parmar, (+ another 7 authors), Computational Radiomics System to Decode the Radiographic Phenotype, Cancer Research, vol. 77, no. 21, 2017, e104-e107.
DOI: 10.1158/0008-5472.can-17-0339

[21] A. H. Nishanov, B. B. Akbaraliev, S. K. Tajibaev, About One Feature Selection Algorithm in Pattern Recognition, Advances in Intelligent Systems and Computing, 2021, 1323 AISC, pp. 103-112.
DOI: 10.1007/978-3-030-68004-6_13

[22] A. K. Nishanov, O. T. Allamov, O. B. Ruzibaev, A. S. Abdullaev,

S. T. Allamova, An approach to finding the most optimal route in a dynamic graph, Int. Conf. on Information Science and Communications Technologies: Applications, Trends and Opportunities, ICISCT 2021, Tashkent, Uzbekistan, 3-5 November 2021.
DOI: 10.1109/ICISCT52966.2021.9670385

[23] A. Nishanov, B. Akbaraliev, R. Beglerbekov, O. Akhmedov, S. Tajibaev, R. Kholiknazarov, Analytical method for selection an informative set of features with limited resources in the pattern recognition problem, E3S Web of Conferences, 2021, 284, 04018.
DOI: 10.1051/e3sconf/202128404018

[24] A. H. Nishanov, G. P. Djuraev, M. A. Khasanova, S. X. Saparov, F. M. Zaripov, Algorithm of diagnostics of medical datas based on symptom complexes, Proc. of the 2nd Int. Conf. on Computer Applications for Management and Sustainable Development of Production and Industry (CMSD-II-2022), vol. 12564, 125640W (2023).
DOI: 10.1117/12.2669449

[25] A. Nishanov, M. Akbarova, A. Tursunov, F. Ollamberganov, D. Rashidova, Clustering algorithm based on object similarity, Journal of Mathematics, Mechanics and Computer Science, 123(3), pp. 108-120.
DOI: 10.26577/JMMCS2024-v123-i3-4

[26] A. K. Nishanov, G. P. Juraev, M. A. Khasanova, F. M. Zaripov, S. X. Saparov, Algorithm for the Classification of Coronary Heart Disease Based on the Use of Symptom Complexes in the Cardiovascular Environment, Communications in Computer and Information Science, 2023, 1733 CCIS, pp. 147-167.
DOI: 10.1007/978-3-031-23744-7_12