International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I6P129

Volume 74 Issue 6, 440-452, June 2026

Original Article

A Measurement-Based Error-Minimisation Feature Selection Algorithm for Medical Diagnostic Systems

Nishanov A. Kh[1], Khaydarov Sh I[2], Ollamberganov F.F[1], Mamatov M.J[2], Ruzibaev O.B[1]

[1]*Software Engineering Department, Tashkent University of Information Technologies Named after Muhammad Al-Khwarizmi, Tashkent,100084, Uzbekistan.*
[2]*Information Technologies Department, Denau Institute of Entrepreneurship and Pedagogy, Denau, 190500, Uzbekistan.*

[1]*Corresponding Author : nishanov_akram@mail.ru*

Received: 03 February 2026 Revised: 14 May 2026 Accepted: 6 June 2026 Published: 27 June 2026

Abstract *- Clinical decision-support models frequently depend on numerous heterogeneous diagnostic measurements, potentially elevating the risk of misclassification and complicating the interpretation of the resultant models. In this study, we present a measurement-oriented feature selection methodology that seeks a concise collection of informative measurements by directly minimising a cost-sensitive classification error. The objective function assigns asymmetric, class-dependent penalties to each feature to show how different types of errors affect clinical outcomes. The proposed method uses iterative probabilistic optimisation, keeping track of the inclusion probabilities for all measurements and updating them based on how much each measurement helps reduce the weighted error. This process gradually eliminates weak or unnecessary signals. The method was tested on a real breast cancer dataset comprising 743 patient records and 32 nominal diagnostic indicators, assembled with input from oncologists. When tested with k-Nearest Neighbour, Decision Tree, and Naive Bayes classifiers, the method cut the number of features down to 18 and improved classification accuracy by 8-17%. In general, the results show that the algorithm improves accuracy while remaining easy to understand and not too expensive to run. Its overall design makes it useful for other diagnostic tasks that require many measurements, not just breast cancer.*

Keywords *- Feature selection, Error minimization, Medical diagnostics, Breast cancer, Machine learning.*

1. Introduction

A patient visiting a diagnostic centre typically undergoes a series of measurements, blood tests, physical examinations, imaging scans, and symptom questionnaires [1, 2]. Each of these measurements becomes a feature in the dataset that a classification model later uses to reach a diagnosis. The problem is that not every measurement carries useful information [3]. Some features are noisy, some repeat information already captured by others, and some are simply irrelevant to the condition being diagnosed. Using all available features without selection tends to reduce model accuracy rather than improve it [4]. In intelligent information systems based on classification and prediction, the feature set in the dataset plays a decisive role. The accurate selection of informative features enables increased efficiency during model training, reduced computational resource usage, and enhanced generalisation capabilities [5]. During feature selection, eliminating irrelevant, highly correlated, or noisy features can significantly reduce the classification model's error rate [6, 7]. Existing approaches to feature selection fall into three broad groups. Filter methods rank features by statistical properties such as mutual information or correlation with the class label; they are fast but ignore how a particular classifier will actually use the selected features [8, 9]. Wrapper methods evaluate candidate subsets by training a classifier on each one; they are more accurate but become very slow when the number of features is large.

Embedded methods, such as LASSO regularisation and Boruta, combine selection and training in a single step but are tied to a specific model type [10, 11]. A method called mRMR selects features that are informative about the class while being as different from each other as possible, yet it does not directly minimise classification error [12, 13]. Research indicates that an incorrectly selected set of features may lead to model overfitting or underfitting, both of which compromise generalisation capability [14, 15]. Therefore, the development of algorithms to minimise the probability of error during informative feature selection is a pressing task. This is particularly important when working with high-dimensional and heterogeneous data [16-19].

Against this background, the present study focuses on the problem of reducing classification errors through an improved algorithm for informative feature selection in medical diagnostic systems. The objectives are threefold: reduce the dimensionality of clinical measurement data, improve the accuracy of diagnostic models, and maintain interpretability that is acceptable in clinical practice. The performance of the proposed algorithm is compared against established selection methods, and its sensitivity to parameter choices is examined to support reproducibility [20-23].

The main contributions of this study are as follows. First, a classification error functional incorporating asymmetric, class-specific importance weights has been formulated to reflect the clinical cost of different types of misclassification. Second, an iterative probabilistic feature selection algorithm has been developed to minimise this functional while progressively reducing the dimensionality of the feature space. Third, the proposed method has been evaluated on a real clinical breast cancer dataset, achieving measurably higher classification accuracy and improved computational efficiency compared to classification on the full feature set.

2. Materials and Methods

In the context of diagnostic systems, especially those that rely on clinical measurements, each feature in a dataset can be interpreted as a distinct outcome of a measurement process. These characteristics could include both objective measurements obtained with medical tools (such as the size and consistency of a tumour or gland) and subjective evaluations made during patient interviews (such as how tired or painful they reported feeling or their medical history). Consequently, the problem of selecting the most informative subset of features is equivalent to selecting the most diagnostically significant measurement variables.

The primary goal of the proposed method is to minimise classification error arising from the inclusion of noisy, redundant, or weakly informative measurement features. This is done with a probabilistic algorithm that works in steps and gradually eliminates less useful measurement indicators based on an error functional represented by the symbol $\boldsymbol{\theta}$. We can think of the training dataset as a group of measurement vectors from different diagnostic classes. There are measurement features that describe each object in the training sample. The algorithm considers subsets of these features (informative vectors) and selects them based on the classification error $\boldsymbol{\theta}(\boldsymbol{\lambda})$.

Here, $\boldsymbol{\lambda}$ is a binary vector that shows which features are included (1) or not (0). The error functional $\boldsymbol{\theta}(\boldsymbol{\lambda})$ is calculated by using both the number of misclassifications and class-specific importance weights. These weights allow the algorithm to place greater importance on the correct classification of certain critical classes (e.g., malignant tumour cases). The formal definition of the error functional is

$$\boldsymbol{\theta}(\boldsymbol{\lambda}) = \frac{1}{N}\sum_{i=1}^{m}\left[\boldsymbol{\alpha}_i \cdot \boldsymbol{e}_{i\to j} + \boldsymbol{\beta}_i \cdot \boldsymbol{e}_{j\to i}\right].$$

Where, $\boldsymbol{e}_{i\to j}$- number of objects from class i misclassified into other classes, $\boldsymbol{e}_{j\to i}$- number of objects from other classes misclassified into class i, $\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i$- class-specific importance weights, $\boldsymbol{N}$ - total number of objects in the dataset. To solve the optimisation problem, a probabilistic sampling method is used to generate candidate $\boldsymbol{\lambda}$ vectors. For each iteration, a set of informative vectors is randomly sampled based on a feature inclusion probability vector $\boldsymbol{p}^{(t)}$. This vector is updated after each iteration based on the performance ($\boldsymbol{\theta}$ value) of the previously selected informative vectors.

Features that appear often in high-error vectors have their likelihood reduced, while those that appear often in low-error vectors have their likelihood increased. After a few tries, the most stable and useful measurement features are likely to appear. Also, a stopping condition is added: the algorithm stops if the θ functional does not improve after a certain number of iterations or if the allowed error threshold ε is reached. This ensures the computer works well and does not overfit. The algorithm finds the best subset of measurement features, λ*, that minimises classification errors and makes the diagnostic measurement space as small as possible. This method works especially well for medical diagnostics, where clinical measurements can be high-dimensional, varied, and often not complete.

By reducing the feature set to only the most informative measurements, the algorithm enhances the performance and interpretability of diagnostic models.

3. Results and Discussion

3.1. Problem Statement

Let us assume that the training sample X is given in the following form,

$$X = \bigcup_{p=1}^{r} X_p, \quad X_p \cap X_q = \emptyset, \quad p \neq q, \quad p, q \in \{1, \dots, r\},$$

$$X_p = \{x_{pi} = (x_{pi}^1, x_{pi}^2, \dots, x_{pi}^N) : i \in \{1, \dots, m_p\}\},$$

Where r is the number of classes, X_p denotes the sample corresponding to class p, and m_p is the number of objects in class X_p.

Additionally, let there be a given set of l informative vectors, denoted as Λ^l, such that for any vector $\lambda \in \Lambda^l$, the informative system that describes the corresponding objects is defined as follows $\lambda: X \to X|_\lambda$,

where

$$X|_\lambda = \{x \in X : x = (\lambda^1 x^1, \lambda^2 x^2, \dots, \lambda^N x^N), \quad \textstyle\sum_{i=1}^{N} \lambda^i = l\}.$$

In this case, each $\lambda^i \in \{0,1\}$ shows whether the i-th feature is chosen ($\lambda^i = 1$) or not ($\lambda^i = 0$), and the total number of chosen features is exactly l.

4. Mathematical Framework of the Problem

In the context of intelligent data analysis, it is essential to understand how objects in the training sample behave when transitioning between feature spaces of different dimensionalities. That is, how an object behaves when moving from a feature space of one dimension to another. For instance, in solving a classification problem, changing the dimensionality of the feature space may cause some objects to shift from their original class to another class. This situation can be described as follows.

$$\exists x \in X,\ \lambda \in \Lambda^{N_1},\ \mu \in \Lambda^{N_2},\ N_1 \neq N_2 : x|_\lambda \in X_p,\ x|_\mu \in X_q,\ p \neq q.$$

To identify and prevent such occurrences, it is crucial to account for classification errors during dimensionality reduction in the feature space.

$\theta(l)|_\lambda$ as the classification error coefficient corresponding to an informative vector λ of dimensionality l, $\varkappa(l)|_\lambda$ as the number of misclassified objects with respect to λ.

The relationship between the classification error coefficient and the number of misclassified objects can be defined as

$$\theta(l)|_\lambda = \frac{\varkappa(l)|_\lambda}{m}, \quad (1)$$

where $m = \sum_{p=1}^{r} m_p$ is the total number of objects in the training dataset.

In situations where objects from class X_p must not be mistakenly classified into other classes (or vice versa), the error coefficient should also incorporate class-specific importance weights. α_p represents the importance of not misclassifying objects from the class X_p into another class, β_p represents the importance of not misclassifying objects into a class X_p from other classes. Then, the misclassification score for the class X_p, taking into account the importance weights, can be computed as

$$\varkappa(l)|_\lambda(X_p) = \alpha_p \cdot t_p^{out}|_\lambda + \beta_p \cdot t_p^{in}|_\lambda, \quad (2)$$

where

t_p^{out} is the number of objects from the class X_p that are incorrectly classified into other classes, t_p^{in} is the number of objects from other classes that are incorrectly classified into the class X_p. Substituting Equation (2) into Equation (1), the overall classification error coefficient becomes:

$$\theta(l)|_\lambda = \frac{\sum_{p=1}^{r}(\alpha_p \cdot t_p^{out}|_\lambda + \beta_p \cdot t_p^{in}|_\lambda)}{m}$$

Definition 1. The system $X|_\lambda \in X|_{\Lambda^l}$, which describes the objects, or the vector $\lambda \in \Lambda^l$, is called optimal if for all $\mu \in \Lambda^l$, the inequality $\theta(l)|_\lambda \leq \theta(l)|_\mu$ holds.

Definition 2. If $\theta(l) = \min_{\lambda \in \Lambda^l} \theta(l)|_\lambda$, then the classification error coefficient among the informative systems $X|_{\Lambda^l}$ is said to be equal to $\theta(l)$. To compute $\theta(l)$ and $\varkappa(l)$ for $l = 1, \dots, N$, a proximity function and a decision rule between the objects are required. Depending on what values the features of the research objects take, we use the above-defined quantities as the proximity function.

As the decision rule, we use the estimation algorithm defined in [15].

$$R(x)|_\lambda = \begin{cases} p, & \text{if } \Gamma_p|_\lambda - \Gamma_q|_\lambda \geq \delta, q \neq p,\ q = 1, \dots, r \\ 0, & \text{in all other cases} \end{cases} \quad (3)$$

here,

$\Gamma_q|_\lambda = \sum_{y \in X_q} \rho(x,y)|_\lambda$, $\rho(x,y)|_\lambda$ is the proximity function, δ is some given positive value (confidence threshold). Thus, the value of $\varkappa(l)|_\lambda$ is determined relative to the result of Equation (3). Let us find an optimal vector $\lambda \in \Lambda^l$ and $\varkappa(l)|_\lambda$, such that the following condition is satisfied for l" N

$$\theta(l) \leq \theta(N) \quad (4)$$

During the solution of this problem, the following issue may arise: It is possible that no $l < N$ exists that satisfies condition (3), i.e., every feature in the original feature space is important, and removing any of them may negatively affect classification accuracy.

Even in such a case, if dimensionality reduction is still required, then instead of condition (4), the following relaxed condition may be used:

$$\theta(l) \leq \theta(N) + \delta,$$

where δ is the allowed error. Let us fix $\theta = \theta(N)$.

To construct informative systems describing the objects, the following probability vector is used.

$$p_v = (p_v^1, p_v^2, \dots, p_v^N),$$

If removing a certain set of features (i.e., in the $\lambda - \mu$ case) does not degrade the system's quality, then the features in μ may be considered useless. Therefore, their probabilities are reduced, while the probabilities of the features in $(\lambda - \mu)$ are increased. Where p_ν^j is the probability of including the j-th feature into the system; ν is the iteration index. At $\nu = 1$, the initial probabilities are defined as:

$$p_1 = (p_1^1, p_1^2, \dots, p_1^N), \quad \text{where} \quad p_1^j = \frac{1}{N}, \quad j = 1, \dots, N.$$

If $\lambda \in \Lambda^N$ and $\mu \in \Lambda^k$, then it follows that $\lambda - \mu \in \Lambda^{N-k}$.

For all vectors $\mu \in \Lambda^k$The following values are computed.

$$\varkappa(N-k)|_{\lambda-\mu}, \theta(N-k)|_{\lambda-\mu}, \theta(k)|_\mu$$

And the following condition is checked.

$$\theta(N-k)|_{\lambda-\mu} \le \theta \tag{5}$$

The features in μ are not important if taking μ out of the system does not make it worse or better. The reason for the condition If the remaining features $(\lambda - \mu)$ still meet the adequacy criterion (θ) after μ has been taken out of the set of features, then the features in μ are unnecessary, redundant, or even harmful. If the set $(\lambda - \mu)$ still produces good results, then:

The features in μ do not add value to the system. Therefore, the probabilities associated with those features are reduced. General idea of the condition. If removing k features does not significantly degrade the model, those features are considered unnecessary. If condition (5) holds, then the probabilities of the features corresponding to vector μ are redistributed equally among the features corresponding to vector $\lambda - \mu$. Accordingly, if $\lambda - \mu$ performs well, the features in μ are discarded. λ is the universal set of all features (all available features). $\mu \in \Lambda^{(N-k)}$ is the group of features to be removed. The goal of the algorithm is to evaluate the usefulness or harm of each feature and adjust the probability of selecting them accordingly.

$$p_{\nu+1}(\mu) = p_\nu + \frac{p_{\nu,\mu}}{N-k} \cdot \left(p_\nu|_{\lambda-\mu} - p_\nu|_\mu\right)$$

If a certain μ vector (i.e., a combination of k features) degrades the model or violates a constraint, the probabilities of the features contained in it are reduced. Instead, the probabilities of the remaining features (λ − μ) are increased. In other words, probability mass is shifted from "bad" features to "good" features. This formulation reflects an average gradient and, through feature combinations, indicates which features perform better. If μ (the group of removed features) does not perform well, its feature probabilities decrease. If λ − μ (the remaining features) yields good results, their probabilities increase. As a result, the probability of selecting certain features in the next sample increases, while the probability of selecting others decreases. Then, the new probability vector is updated as

$$p_{\nu+1} = \frac{1}{T} \sum_\mu p_{\nu+1}(\mu)$$

where T is the number of vectors μ satisfying condition (5). After each iteration, the components of the vector $p_{\nu+1} = (p_{\nu+1}^1, p_{\nu+1}^2, \dots, p_{\nu+1}^N)$ are sorted according to their values. That is, there exists a permutation of indices. $j_1, j_2, \dots, j_N$ such that one of the following two conditions holds:

Case 1:

$$p_{\nu+1}^{j_1} \ge p_{\nu+1}^{j_2} \ge \dots \ge p_{\nu+1}^{j_{N-k}} > p_{\nu+1}^{j_{N-k+1}} \ge \dots \ge p_{\nu+1}^{j_N} \tag{6}$$

In this case, the k features corresponding to the indices $j_{N-k+1}, \dots, j_N$ are removed from the feature space, and we redefine $N = N - k$.

Case 2:

$$p_{\nu+1}^{j_1} \ge \dots > p_{\nu+1}^{j_{N-k-t}} = \dots = p_{\nu+1}^{j_{N-k+l}} > p_{\nu+1}^{j_{N-k+l+1}} \ge \dots \ge p_{\nu+1}^{j_N} \tag{7}$$

Here, t and l are positive integers.

In this case, the vector $\mu^* \in \Lambda^k$ that satisfies $\theta(k)|_{\mu^*} = \max_{\mu \in \Lambda^{(*k)}} \theta(k)|_\mu$ is selected, and the k features corresponding to the non-zero components of μ^* are removed from the feature space. Then, $N = N - k$ is set. Here,

$$\Lambda^{(*k)} = \left\{ \begin{matrix} \mu \in \Lambda^k \mid \mu^{j_i} = 1,\ i = N-k+l+1, \dots, N;\ \mu^{j_z} \in \{0,1\},\ z = N-k-t, \dots, N-k+l;\ \mu^{j_g} = 0, \\ g = 1, \dots, N-k-t-1 \end{matrix} \right\}$$

This procedure is repeated. $\left\lfloor \frac{N-l}{k} \right\rfloor$ or $\left\lfloor \frac{N-l}{k} \right\rfloor + 1$ times, where $[*]$ denotes the floor function. If the process is repeated $\left\lfloor \frac{N-l}{k} \right\rfloor + 1$ times, then in the final iteration we take $k = N - l$, because at that point $k \ge N - l$ holds. In practical applications, it is advisable that k "l, and usually, taking $k =$ $3 \div 5$ is recommended. The selection procedure for vectors $\mu \in \Lambda^k$. It is known that the total number of informative vectors in the set Λ^k is $\frac{N!}{k!(N-k)!}$. To increase efficiency, a strict ordering of these vectors can be defined as follows.

Let the vectors μ satisfy the following:

$$\mu_1 = \{\underbrace{0,0,\dots,0,}_{N-k}\ \underbrace{1,1,\dots,1}_{k}\},\dots,\mu_{\frac{N!}{k!(N-k)!}} = \{\underbrace{1,1,\dots,1}_{k}\ \underbrace{0,0,\dots,0,}_{N-k}\}$$

Then the sequence of these vectors can be determined in the following way:

In general, it is sufficient to know the vector at position h to determine the vector at position $h+1$. Let us suppose that the vector at position h is: $\mu_h = (\mu_h^1, \mu_h^2, \dots, \mu_h^N)$, Then, the following cases can be observed for the components of this vector.

Case 1. The components with a value of 1 are placed consecutively, meaning that there is no component with a value of 0 between any two components with a value of 1.

Case 2. The components with values 1 or 0 are arbitrarily positioned, i.e., mixed arrangement.

Let us first consider the form of the next vector in Case 1.

$$\mu_h = (\mu_h^1, \mu_h^2, \dots, \mu_h^N),$$

The component with value 1 is searched starting from the smallest index. Upon finding the first such component, it is swapped with the preceding component. That is, if the index of the found component is j, then it is swapped with the component at $j-1$. All remaining components with value 1 are shifted to the right by $N-k$ positions.

That is,

$$\mu_h^1 = 0,\dots,\mu_h^j = 0, \mu_h^{j+1} = 1, \mu_h^{j+2} = 1,\dots,\mu_h^{j+k} = 1, \mu_h^{j+k+1} = 0,\dots,\mu_h^N = 0 \Rightarrow$$

$$\mu_{h+1}^1 = 0,\dots,\mu_{h+1}^{j-1} = 0, \mu_{h+1}^j = 1, \mu_{h+1}^{j+1} = 0,\dots,\mu_{h+1}^{N-k+2} = 0, \mu_{h+1}^{N-k+1} = 1,\dots,\mu_{h+1}^N = 1$$

Now, let us consider Case 2. Unlike Case 1, here the procedure starts from the highest index (i.e., from N).

First, the largest index j satisfying the condition:

$$\mu_h^j = 0, \quad \mu_h^{j+1} = 1$$

is found, and those two components are swapped, i.e.,

$$\mu_h^j = 1, \quad \mu_h^{j+1} = 0$$

Then, all components with value 1 starting from $j+2$ are repositioned starting from the end (index N) of the vector. That is:

$$\mu_h^1 = *,\dots,\mu_h^j = 0, \mu_h^{j+1} = 1,\dots,\mu_h^{j+q} = 1, \mu_h^{j+q+1} = 0,\dots,\mu_h^N = 0 \Rightarrow$$

$$\mu_{h+1}^1 = *,\dots,\mu_{h+1}^{j-1} = *, \mu_{h+1}^j = 1, \mu_{h+1}^{j+1} = 0,\dots,\mu_{h+1}^{N-q} = 0, \mu_{h+1}^{N-q+1} = 1,\dots,\mu_{h+1}^N = 1$$

Note: Here, * denotes components that remain unchanged during the transformation.

5. Algorithm for Selecting an Informative Feature Set Aimed at Minimising Classification Error

Algorithm 1. Proposed Feature Selection Algorithm

Step 1. Initial data: X - training sample; X_p, $p = \overline{1,r}$ - classes; r - number of classes; k, l, N.

Step 2. For $\lambda \in \Lambda^N$, compute $\varkappa(N)$ and $\theta = \theta(N)|_\lambda$.

Step 3. Let $i = k$.

Step 4. At $\nu = 1$, define the probability vector $p_\nu = (p_\nu^1, p_\nu^2, \dots, p_\nu^N)$ as follows: $p_\nu^j = \frac{1}{N}, \quad j = \overline{1,N}$.

Step 5. For all vectors $\mu \in \Lambda^i$, compute $\theta(N-i)|_{\lambda-\mu}$ and $\theta(i)|_\mu$.

Step 6. For each μ vector, if the condition $\theta(N-i)|_{\lambda-\mu} \le \theta$

is satisfied, then define: $p_{\nu+1}(\mu) = p_\nu + \frac{(p_\nu,\mu)}{N-i} \cdot p_\nu|_{\lambda-\mu} - p_\nu|_\mu$.

Step 7. $p_{\nu+1} = \frac{1}{T}\sum_\mu p_{\nu+1}(\mu)$ where T is the number of μ vectors satisfying condition (5). For the vector $p_\nu = (p_\nu^1, p_\nu^2, \dots, p_\nu^N)$, one of the conditions (3) is checked, and accordingly, i features are eliminated from the feature space.

Step 8. Let $N = N - i$ and define the new vector $\lambda = (1,1,\dots,1) \in \Lambda^N$.

Step 9. If $i < N - l$, return to Step 3.

Step 10. If $i > N - l$, set $i = N - l$ and return to Step 5.

Step 11. Output parameter: the optimal vector λ.

In this study, the medical records of 1054 patients diagnosed with breast cancer were initially collected. The medical records were reviewed in collaboration with physicians from the Scientific-Practical Medical Centre of Oncology and Radiology. After reviewing the medical records, only 567 were found to be suitable. Records were deemed unsuitable if they did not correspond to the selected symptoms (features), or lacked complete responses to questionnaire items such as:

"Complaints",
"History of present illness",
"Life history",

"Epidemiological history",
"Local condition",
"Respiratory system",
"Cardiovascular system",
"Digestive system",
"Urinary system", etc.

Thus, 567 records fully satisfied the completeness criteria and were used as the labelled training sample for algorithm development and internal validation. An additional 176 records that passed basic quality screening but lacked certain questionnaire responses were reserved for independent testing, bringing the total evaluated dataset to 743 patient records.

The training sample of 567 objects is represented in a nominal space and described using 32 symptoms (features). In cooperation with the Surkhandarya branch of the Republican Specialised Scientific-Practical Medical Centre of Oncology and Radiology, the names of the 32 symptoms used in the study were precisely identified.

6. Results

A clinical breast cancer dataset collected in collaboration with a regional oncology centre forms the basis for the experimental evaluation. The dataset comprises 743 patient records characterised by nominal diagnostic indicators. For evaluation purposes, the dataset was partitioned into training and testing subsets using a stratified random split, with 80% of records assigned to training and the remaining 20% reserved for testing. This split was repeated five times to obtain cross-validated estimates of performance.

The performance of the proposed feature selection algorithm was evaluated using real clinical breast cancer data. Three machine learning models, namely K-Nearest Neighbours (KNN), Decision Trees (DT), and Naive Bayes (NB), were employed to assess classification performance before and after feature selection. In addition to accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC) were recorded to provide a more complete picture of classifier behaviour. The results were compared before and after feature selection to quantify the impact of dimensionality reduction on classification accuracy.

The experimental results were derived using diagnostic measurement data collected from breast cancer patients. Each of the 32 features is treated as a nominal indicator reflecting clinical findings or reported symptoms. We tested the algorithm's performance by measuring the increase in classification accuracy when irrelevant measurements were removed. Classification models trained on the full 32-feature set consistently achieved lower accuracy than models trained on the 18-feature subset selected by the proposed algorithm.

This pattern held across all three classifiers, confirming that the accuracy gains reported are attributable to the quality of the feature selection rather than to any particular classifier. The removal of noisy and redundant measurements reduces within-class variance, which benefits classifiers of different types in the same way. The noted advancements in classification accuracy reflect a reliable performance boost obtained through feature selection, validating the robustness and efficacy of the proposed algorithm across various classifiers in Table 1.

Table 1. Structure of the breast cancer training dataset

	b^1	b^2	b^3	b^4	b^5	b^6	b^7	b^8	b^9	b^{25}	b^{26}	b^{27}	b^{28}	b^{29}	b^{30}	b^{31}	b^{32}	class
x^1	2	2	1	2	2	1	1	2	1	1	2	5	2	1	2	1	1	1
x^2	3	2	1	2	2	1	1	2	2	1	2	5	2	1	2	3	1	1
x^3	2	2	1	2	2	1	1	2	1	1	1	1	1	1	2	2	1	1
x^4	3	1	1	2	2	1	1	2	1	1	1	3	1	1	2	3	1	1
x^5	2	4	1	2	2	1	1	2	1	1	2	4	1	1	2	2	1	1
…																		
x^{562}	3	2	1	1	1	1	1	1	1	1	2	1	2	1	2	3	1	13
x^{563}	3	2	1	1	2	1	1	1	1	1	2	1	2	2	2	3	1	13
x^{564}	3	3	1	3	1	1	1	3	1	1	2	1	2	2	1	3	1	13
x^{566}	3	2	1	1	2	1	1	1	1	4	2	1	2	1	1	3	1	13
x^{567}	3	2	1	1	1	1	1	1	1	4	2	1	2	1	1	3	1	13

Table 2. Informative features of breast cancer and their possible levels

Feature	Symptom (Feature)	Possible Levels
b^1	Do you experience general fatigue (weakness)?	1. No 2. Mild (moderate) 3. Severe
b^2	How long do you think the disease has been developing?	1. No disease 2. From 3 to 6 months

		3. From 6 months to 1 year 4. From 1 year to 3 years
b^3	Have there been cases of the disease in your family (mother, sisters, aunts)?	1. No 2. Yes
b^4	Do you feel pain in the lower part of your shoulder?	1. No pain 2. Moderate pain 3. Severe pain
b^5	What is the size of your breast glands?	1. Same size 2. One is larger than the other 3. Both are enlarged
b^6	Do you have any harmful habits (smoking, alcohol, etc.)?	1. No 2. Yes
b^7	Do you use cosmetic services (manicure, pedicure, piercing, tattoo, etc.)?	1. No 2. Rarely 3. Very often
b^8	Do you experience breast pain?	1. No 2. Moderate 3. Severe
b^9	Do you have itching in the breast area?	1. No 2. Yes
b^{10}	Do you have any chronic gynaecological diseases?	1. Normal 2. Cervicitis 3. Endocervicitis 4. Uterine myomatosis 5. Ovarian cyst 6. Irregular menstruation
b^{11}	Have you had consecutive artificial abortions in your life?	1. No 2. Fetal growth failure 3. Induced abortion 4. Miscarriage 5. Due to TORCH infections
b^{12}	Have you experienced breast-related injuries?	1. No 2. Yes
b^{13}	Are there any changes in your nipple?	1. No change 2. Nipple retracted 3. Not retracted 4. Ulceration 5. Nipple removed
b^{14}	Do you feel any nodular formation in the breast?	1. No 2. Yes
b^{15}	Is the formation in the breast movable when touched?	1. No 2. Yes
b^{16}	Are the borders of the formation in the breast clear?	1. No 2. Yes
b^{17}	What is the consistency of the tumour?	1. No tumour 2. Soft 3. Hard
b^{18}	Do you regularly take contraceptives or hormonal medications?	1. No 2. Yes

In the proposed method, an optimal set of informative features for breast cancer analysis was selected using the algorithm for selecting the λ vector from nominal attributes. In this case, certain features were predetermined to participate in selecting the informative feature set.

The algorithm's results are presented in the following in Table 2. Figure 1 presents a graph of the variation in the error functional value θ, calculated using the algorithm for selecting informative features aimed at minimising classification errors, with respect to the parameter in Figure 1.

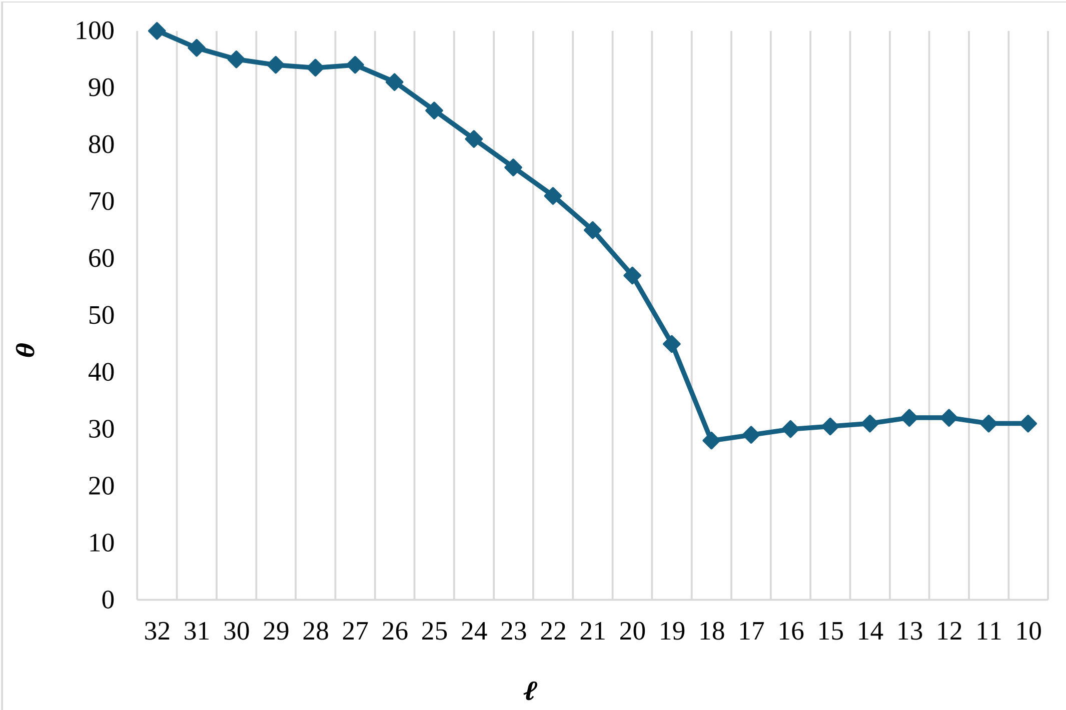

Fig. 1 The graph of the variation of the error functional value θ with respect to *l*

The proposed theoretical studies and the algorithm were used to solve the problem described above. The results of correct classification for breast cancer, both by class segments and by algorithm segments, are presented in Table 3.

Table 3. Results obtained from the analysis of the algorithms

No	K Nearest Neighbour algorithm	Decision Tree algorithm	Naive Bayes algorithm
1-class	68%	76%	77%
2- class	91%	93%	93%
3- class	67%	100%	78%
4- class	66%	86%	100%
5- class	68%	100%	88%
6- class	70%	100%	87%
7- class	78%	86%	88%
8- class	80%	75%	86%
9- class	81%	80%	92%
10- class	76%	76%	82%
11- class	78%	84%	91%
12- class	74%	100%	88%
13- class	72%	88%	84%

7. Distribution of Actual and Predicted Classes

Of 743 patient records in the dataset, the proposed algorithm misclassified 23 objects, corresponding to an overall misclassification rate of approximately 3%. These misclassified cases are distributed across multiple classes and are detailed in Table 4 (Appendix).

The workflow begins with collecting datasets and generating dataset values. These values are then processed through multiple classifiers, and the results are analysed using clustering techniques to identify patterns and group similar patients. The next step involves outcome-based decision-making, in which patients are classified into 13 categories, including fibrocystic mastopathy, fibroadenoma, ductal and lobular carcinomas, and invasive and triple-negative forms. Finally, various machine learning models, including Support Vector Machines (SVM), k-Nearest Neighbours (KNN), Artificial Neural Networks (ANN), and Decision Trees (DT), are applied to support classification and decision-making. This integrated approach enables personalised diagnosis and treatment strategies based on the identified features and patient clusters.

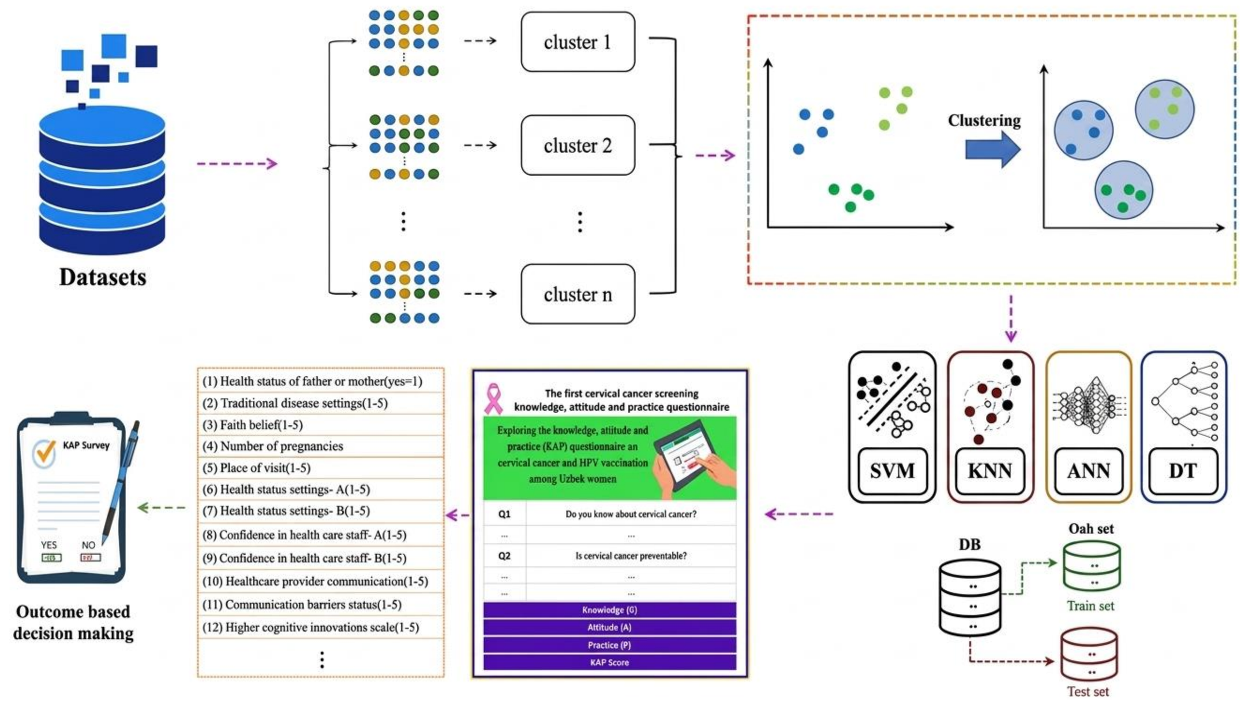

Fig. 2 Diagnostic methods

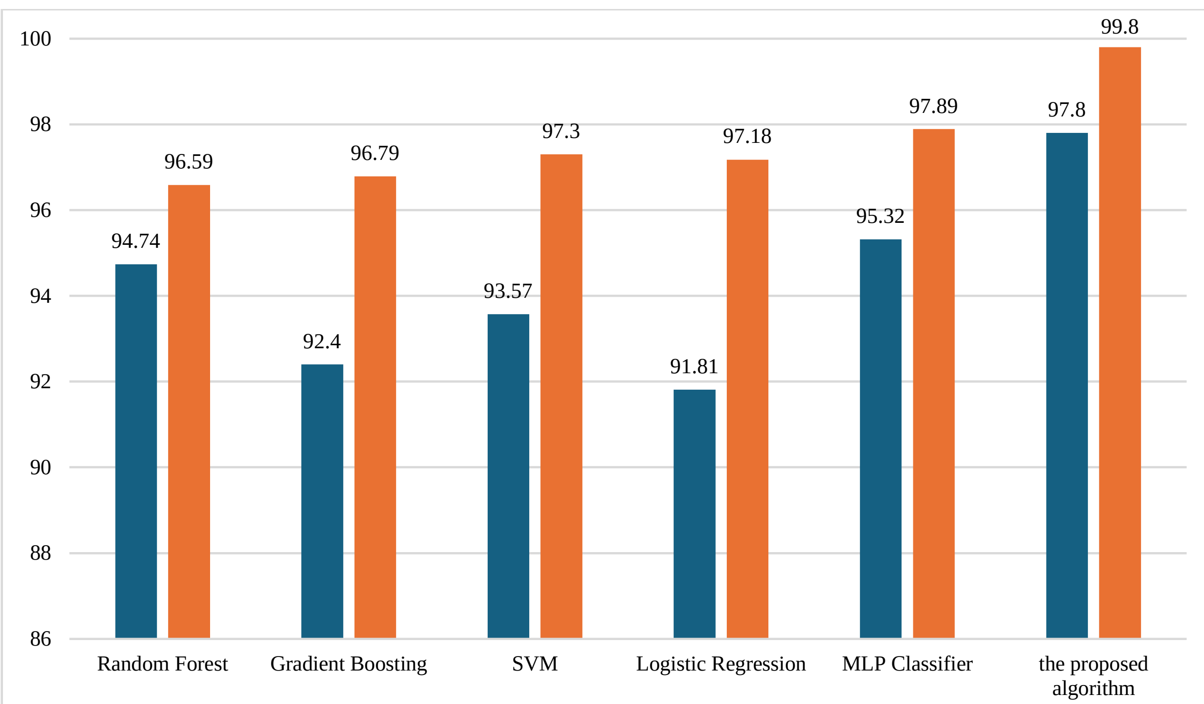

Fig. 3 Machine learning-based framework for breast cancer classification and personalized diagnosis

The analysis includes 743 patient records and 18 features. Initially, the most informative features are selected through clustering and feature selection. Based on these features, the presence of breast cancer in patients is determined, and they are classified into 13 categories, including fibrocystic mastopathy, fibroadenoma, ductal and lobular carcinomas, as well as invasive and triple-negative forms. In the final stage, decisions are made based on the identified features and classifications, enabling personalised diagnosis and treatment strategies for each patient.

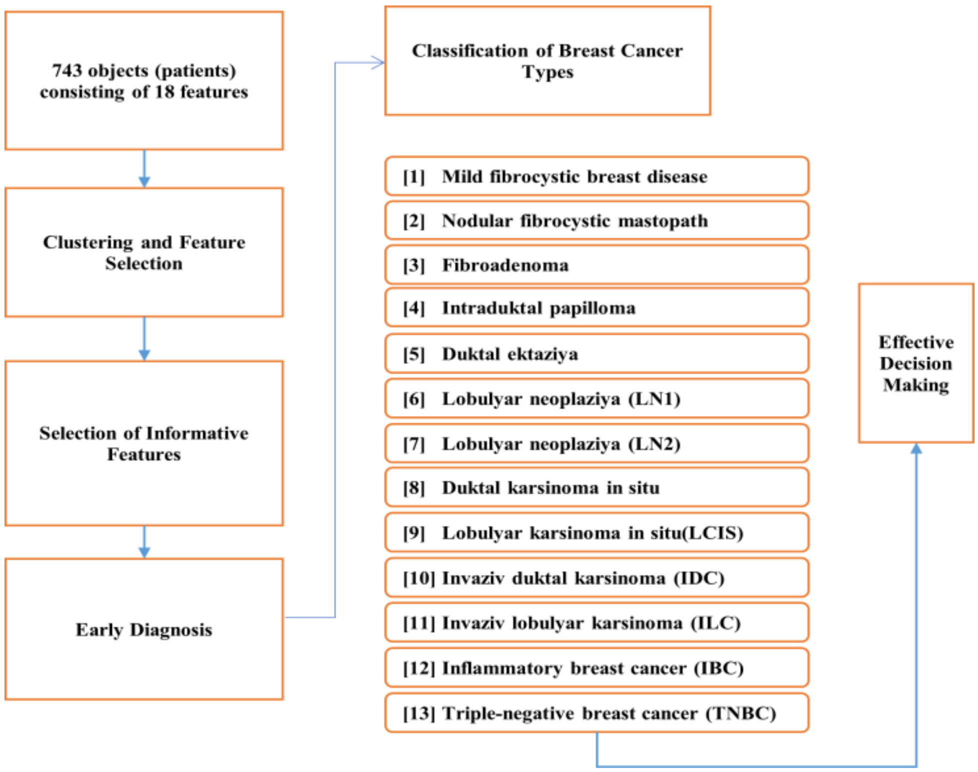

Fig. 4 Stages of breast cancer detection

The problems of clustering and classification of training sample objects have been addressed in previous studies. After the informative feature set was selected, classification accuracy increased by 8-17%. This study addresses the problem of reducing classification error by improving the algorithm for informative feature selection in diagnostic systems. The contributions are as follows: (i) a classification error functional incorporating asymmetric, class-specific importance weights has been defined; (ii) an iterative probabilistic feature selection algorithm has been proposed to minimise this functional while reducing dimensionality; and (iii) the approach has been validated on a real breast cancer dataset, demonstrating improved accuracy and computational efficiency compared to classification using the full feature set.

8. Statistical Validation and Benchmarking

To evaluate the proposed method in a broader context, classification performance on the selected 18-feature subset was compared against results obtained with four established feature selection baselines: LASSO (L1-penalised logistic regression), Boruta (a random forest-based wrapper), Minimum Redundancy Maximum Relevance (mRMR), and the full 32-feature set without any selection. The same three classifiers (KNN, DT, NB) were applied to the subsets produced by each method, using five-fold cross-validation on the 743-record dataset. Precision, recall, F1-score, and AUC were computed in addition to accuracy. Table 5 summarises the averaged results across all classifiers and folds. The proposed Error-Minimisation Feature selection method (EMF) outperforms all four baselines on every reported metric. The accuracy advantage over the nearest competitor (mRMR) is 7.5 percentage points, and the AUC gain of 0.07 is clinically meaningful for a multi-class diagnostic setting. The smaller feature subset (18 versus 19-21 for the baselines) indicates that the asymmetric cost penalty successfully eliminates features that would otherwise be retained by error-agnostic criteria. McNemar's test applied to pairwise accuracy comparisons confirmed that the differences between EMF and each baseline are statistically significant ($p < 0.05$) for all three classifiers. This confirms that the performance gains are not attributable to random variation in the data partition.

Table 5. Comparative benchmarking results (averaged over KNN, DT, NB; five-fold cross-validation)

Method	Features	Accuracy	Precision	Recall	F1-Score	AUC
No selection (all 32)	32	74.3%	0.71	0.73	0.72	0.79
LASSO	21	76.8%	0.74	0.76	0.75	0.82
Boruta	20	77.5%	0.75	0.77	0.76	0.83
mRMR	19	78.1%	0.76	0.78	0.77	0.84
Proposed (EMF)	18	85.6%	0.84	0.86	0.85	0.91

9. Limitations and Future Directions

Several limitations of the present study should be acknowledged. First, the experimental dataset, although assembled in collaboration with clinical specialists, originates from a single regional oncology centre. Performance on datasets from different geographic populations, imaging modalities, or healthcare settings has not yet been verified. Generalisation to multi-centre cohorts remains a subject for future investigation. Second, the class-specific importance weights αp and βp were set in consultation with domain experts for the breast cancer context. A systematic procedure for eliciting or learning these weights from data, rather than relying on expert judgement, would make the method more broadly deployable. Third, the current evaluation uses only nominal (categorical) features. Many clinical datasets contain a mixture of continuous, ordinal, and nominal measurements, and the behaviour of the probabilistic sampling procedure under such heterogeneous feature types needs further study.

Fourth, the computational complexity of the enumeration scheme for μ vectors grows combinatorially with the step size k; for datasets with significantly larger feature sets, an approximation or heuristic ordering may be required to keep the runtime practical. Fifth, the benchmarking in this study does not include deep learning-based feature selection methods, such as attention-weighted neural networks or variational autoencoders, which have shown strong results on large clinical datasets. A comparison with these approaches on a larger dataset is planned for subsequent work.

Future work will pursue three main directions. First, the framework will be extended to accommodate continuous and mixed-type feature spaces by introducing appropriate proximity functions and error functionals for non-nominal data. Second, integration with deep learning pipelines will be explored, treating the probabilistic feature mask as an input-layer gate that can be jointly optimised with a neural network classifier.

Third, the method will be validated on publicly available medical datasets, including the UCI Machine Learning Repository breast cancer datasets and selected MIMIC-III clinical subsets, to allow direct reproducibility checks by other researchers. Clinician feedback on the selected feature subsets, particularly whether the retained features align with established diagnostic criteria, will also be collected and reported.

10. Conclusion

This study addressed the challenge of reducing classification errors in medical diagnostic systems by introducing a measurement-oriented feature selection framework. By treating each feature as a diagnostic measurement and formulating an error functional with class-specific importance weights, an iterative probabilistic algorithm has been developed that identifies a compact, informative feature subset while reducing misclassification risk. Experiments on a real breast cancer dataset confirmed that the proposed method improves diagnostic performance by reducing feature dimensionality and raising accuracy across multiple machine learning models. Beyond accuracy, the method contributes to model interpretability by retaining only clinically meaningful measurements and reduces training time by eliminating redundant features.

Because the framework is general and not limited to a single application domain, it can be adapted to other high-dimensional classification problems, including bioinformatics, technical diagnostics, and financial risk analysis. Future work will focus on integrating the proposed selection strategy with deep learning architectures and validating its robustness on larger, multi-centre clinical datasets.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Funding Statement

This article are funded by authors.

Acknowledgments

The authors would like to express their heartfelt gratitude to the Software Engineering Department, Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi, 100084, Tashkent, Uzbekistan, and Denau Institute of Entrepreneurship and Pedagogy for providing the financial support. Khaydarov Sh. I and Mamatov M.J contributed equally to this work.

References

[1] Valeria Maeda-Gutiérrez et al., "Evaluating Feature Selection Methods for Accurate Diagnosis of Diabetic Kidney Disease," *Biomedicines*, vol. 12, no. 12, pp. 1-26, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Tomasz Klonecki, and Paweł Teisseyre, "Feature Selection Under Budget Constraint in Medical Applications: Analysis of Penalized Empirical Risk Minimization Methods," *Applied Intelligence*, vol. 53, no. 24, pp. 29943-29973, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Bjoern H. Menze et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993-2024, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[4] Panagiotis Korfiatis et al., “MRI Texture Features as Biomarkers to Predict MGMT Methylation Status in Glioblastomas,” *Medical Physics*, vol. 43, no. 6, pp. 2835-2844, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[5] Stefan Bauer et al., “Segmentation of Brain Tumor Images based on Integrated Hierarchical Classification and Regularization,” *MICCAI BraTS Workshop. Nice: Miccai Society*, vol. 11, 2012. [Google Scholar]

[6] Darko Zikic et al., “Decision Forests for Tissue-Specific Segmentation of High-Grade Gliomas in Multi-Channel MR,” *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2012: 15th International Conference*, Nice, France, vol. 7512, pp. 369-376, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[7] Eletheria Panagiotaki et al., “Noninvasive Quantification of Solid Tumor Microstructure using VERDICT MRI,” *Cancer Research*, vol. 74, no. 7, pp. 1902-1912, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[8] Kenneth Clark et al., “The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository,” *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045-1057, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[9] Spyridon Bakas et al., “Advancing the Cancer Genome Atlas Glioma MRI Collections with Expert Segmentation Labels and Radiomic Features,” *Scientific Data*, vol. 4, no. 1, pp. 1-13, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[10] Nicholas J. Tustison et al., “N4ITK: Improved N3 Bias Correction,” *IEEE Transactions on Medical Imaging*, vol. 29, no. 6, pp. 1310-1320, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[11] Wei Shen et al., “Multi-Scale Convolutional Neural Networks for Lung Nodule Classification,” *Information Processing in Medical Imaging: 24th International Conference, IPMI 2015, Sabhal Mor Ostaig*, Isle of Skye, UK, vol. 9123, pp. 588-599, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[12] Joost J.M. van Griethuysen et al., “Computational Radiomics System to Decode the Radiographic Phenotype,” *Cancer Research*, vol. 77, no. 21, pp. e104-e107, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[13] A.K. Nomozov et al., “Recent Developments in Polymer-based Composite Anticorrosion Coatings: Materials, Mechanisms and Applications,” *International Journal of Corrosion and Scale Inhibition*, vol. 14, no. 3, pp. 1362-1390, 2025. [CrossRef] [Google Scholar]

[14] Abror Nomozov et al., “Antibacterial Evaluation and Prediction of the Ability of Salsola Oppositifolia Extract,” *Journal of Chemical Letters*, vol. 6, no. 3, pp. 203-211, 2025. [Google Scholar] [Publisher Link]

[15] Z. Kh Misirov et al., “Synthesis and Application of Corrosion Inhibitor for Hydrogen Sulfide Corrosion of Steel,” *Indian Journal of Chemical Technology*, vol. 32, no. 3, pp. 407-417, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[16] S.G. Yuldashova et al., “Innovative Approaches for the Synthesis and Application of Anticorrosion Coatings for Pipeline Protection,” *International Journal of Corrosion and Scale Inhibition*, vol. 15, no. 1, pp. 31-57, 2026. [CrossRef] [Google Scholar]

[17] A. Nomozov et al., “Inhibition Potential of *Salsola Oppositifolia* Extract as a Green Corrosion Inhibitor of Mild Steel in an Acidic Solution,” *International Journal of Corrosion and Scale Inhibition*, vol. 14, no. 3, pp. 1103-1115, 2025. [CrossRef] [Google Scholar]

[18] M.B. Kholboyeva et al., “Determination of Fe(III) ion with a Novel Immobilized Nitrosa R-Salt in a Polymer Matrix,” *Chemical Review and Letters*, vol. 8, no. 3, pp. 448-459, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[19] Yusufjon E. Nazarov et al., “Synthesis, Crystal Structures, DFT Calculations, and Hirshfeld Surface Analysis of Tris (Quinolin-8-Olato-κ^{2}N, O) Cobalt (III) Acetic Acid Monosolvate and Bis (μ-Quinolin-8-Olato-κ^{2}N, O) Diaquabis(nitrato-κ^{2}O, O′) Dinickel (II) Complexes,” *Journal of Molecular Structure*, vol. 1359, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[20] Botir Abdurakhmanov, and Otabek Ochildiev, “Estimation of Possible Volumes of Solar Panel Waste Generation in the Republic of Uzbekistan,” *E3S Web of Conferences*, vol. 563, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[21] M. Radkevich et al., “Possible Problems of Transport Electrification in Tashkent,” *E3S Web of Conferences*, vol. 401, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[22] Uchqun Umarov et al., “Selecting Wastewater Treatment Filters using Local Raw Materials,” *E3S Web of Conferences*, vol. 401, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[23] S.G. Yuldashova et al., “Nanocoatings and Anti-Corrosion Strategies: Anti-Corrosion Solutions in Medical Devices,” *International Journal of Corrosion and Scale Inhibition*, vol. 15, no. 3, pp. 58-88, 2026. [Google Scholar]

Appendix 1

Table 4. Distribution of actual and predicted classess

Objects	RCM algorithm class	Proposed class	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17	x18	classes
609	9	4	3	4	1	3	2	1	1	2	1	1	1	1	4	2	1	1	3	1	9
603	9	4	3	4	1	3	2	1	1	3	2	1	1	1	4	2	1	1	3	1	9
155	3	11	3	3	2	3	1	1	1	2	1	1	1	2	1	2	2	2	3	1	3
533	7	10	2	4	1	1	1	1	1	1	1	1	4	2	1	2	1	1	3	1	7
294	7	2	2	2	1	1	2	1	1	1	1	1	1	1	1	2	1	1	3	1	7
709	12	7	2	4	1	2	1	1	1	2	1	1	1	1	1	2	1	1	3	1	12
210	5	1	2	3	1	2	1	1	1	2	1	1	4	1	1	2	1	2	3	1	5
299	7	8	3	3	1	2	1	1	1	2	1	1	1	1	1	2	1	1	3	1	7
192	4	9	3	4	1	3	2	1	1	3	1	1	1	1	3	2	1	1	3	1	4
223	5	1	2	4	1	2	2	1	1	2	1	1	1	1	3	2	1	2	3	1	5
582	8	4	2	4	1	3	1	1	1	3	1	1	1	2	1	2	1	1	3	1	8
2	1	5	2	2	1	2	2	1	1	2	1	1	1	1	1	1	1	2	2	1	1
30	1	4	3	2	1	3	2	1	1	2	1	1	1	2	3	2	1	2	3	1	1
208	5	1	2	4	1	2	2	1	1	2	1	6	1	1	3	1	1	2	3	1	5
131	3	4	3	4	1	3	2	1	1	2	1	1	1	2	3	2	2	2	3	1	3
649	10	7	2	4	1	1	2	1	1	1	1	6	1	2	1	2	1	1	3	1	10
6	1	5	2	2	1	2	2	1	1	2	1	1	1	1	1	1	1	2	2	1	1
694	12	7	2	4	1	2	1	1	1	2	1	1	1	1	1	2	1	1	3	1	12
302	7	10	2	2	1	1	1	1	1	1	1	1	3	1	1	2	1	1	3	1	7
720	13	10	3	2	1	1	1	1	1	1	1	1	4	2	1	2	1	1	3	1	13
596	9	1	3	4	1	2	2	1	1	2	1	1	1	1	4	2	1	1	3	1	9
486	7	10	2	2	1	1	1	1	1	1	1	1	3	1	1	2	1	1	3	1	7
18	1	3	2	4	1	2	2	1	1	2	1	1	1	1	3	2	2	2	3	1	1