SIR Publishers
CENTER FOR INTERNATIONAL SCIENTIFIC PUBLICATIONS
JOURNAL OF ENGINEERING SCIENCES
ISSN(Online): 2984-6714
Volume-9, Issue-01, Published |09-02-2026|

UDC: 004.852.5:616-07

THE USE OF ENSEMBLE METHODS FOR MEDICAL DIAGNOSIS PROBLEMS

Eshboyev Erkin Abdirashidovich
dots. Karshi SU city Karshi, Uzbekistan
Daminova Shaxzoda Egamberdiyevna
master. Karshi SU city Karshi, Uzbekistan
Shonazarov Uchqun Nurbek o'g'li
shonazarovuchqun@gmail.com
master. Karshi SU city Karshi, Uzbekistan

ABSTRACT

The article examines the problem of developing a medical diagnosis system based on ensemble methods and evaluating its effectiveness. The system automatically performs diagnosis using disease symptoms, laboratory test results, and other patient-related data, and this capability is confirmed by results obtained on a number of datasets. It is shown that the accuracy values achieved using the ensemble method are almost equivalent to those of other algorithms, and in certain cases demonstrate superior performance. Additionally, to verify the correct operation of the software package developed using the ensemble method, the Heart Disease dataset, the Diabetes dataset, and the Multiclass Diabetes dataset were used, and their accuracy levels and corresponding confusion matrices were calculated.

Keywords: bagging, boosting, stacking, XGBoost, RFC algorithm, Gradient Boosting.

INTRODUCTION

Developing mathematical models for diagnosis in medical research and applying appropriate algorithms holds great potential, offering the ability to significantly improve patient treatment processes and advance achievements in the field of medicine. Mathematical models and algorithms designed for automated medical diagnosis systems include a variety of methods such as machine learning algorithms, statistical models, decision trees, and ensemble methods. Their primary purpose is to comprehensively analyze patient data—including medical history, symptoms, age, and other relevant information—and provide accurate diagnosis. Integrating a mathematical approach into medical research offers new strategies for solving medical diagnostic problems. Sequential and logically structured approaches are applied in the diagnosis of various diseases, often employing diagnostic algorithms that use mathematical models adapted to medical reasoning [1]. Especially, the use of fuzzy set theory methods demonstrates the potential of this approach for creating information-mathematical models for diagnosis and prediction.

The article addresses the problem of developing a medical diagnosis system based on ensemble (artificial intelligence and machine learning) methods and assessing its effectiveness. The system is expected to automatically perform diagnosis using disease symptoms, laboratory test results, and other patient-related data.[2-4]

RESEARCH METHODS

Ensemble Method. The ensemble method is an approach based on combining multiple models to improve prediction accuracy in medical diagnosis problems. This method is used in medical diagnosis to enhance model stability and reliability, as errors in one model can be corrected by other models. In the ensemble method, models are trained separately. The predictions of each model are then combined to achieve more accurate and stable results. Ensemble methods are widely applied in areas requiring high accuracy, such as finance, bioinformatics, marketing, and medicine.

JOURNAL OF ENGINEERING SCIENCES
ISSN(Online): 2984-6714
Volume-9, Issue-01, Published |09-02-2026|

The ensemble method allows solving multiple problems simultaneously, including reducing errors, decreasing retraining, and increasing steadiness. The main principle of ensemble methods is to ensure diversity among models; the more diverse the base classifiers are, the more effectively they can correct each other's errors. There are several main approaches to building ensembles. The most popular ensemble methods include Bagging, Boosting and Stacking [7]:

Bagging – multiple models are trained on different small datasets obtained through random selection and resampling (bootstrap). The process of building an ensemble model using Bagging involves the following steps:

1. Several small datasets are generated from the main dataset using random selection and resampling;
2. A separate model is trained on each small dataset;
3. The predictions from all models are combined to determine the final result (for example, by voting in classification problems or by taking the average in regression problems).

Boosting – models are trained sequentially, with each subsequent model aimed at correcting the errors of the previous model. This algorithm includes the following steps:

1. The first model is trained on the collected medical data;
2. Errors made by the first model are identified, and weights are assigned to the data to focus more on these errors for the next model;
3. Each subsequent model is trained to correct the errors of the previous model;
4. The final prediction is determined based on the weighted sum of all model results.

Stacking – the results of multiple models are combined at the meta-model level. The predictions of the base models are used as input data. The steps of the stacking algorithm are:

1. Several base models are trained on the initial data, and their predictions are obtained;
2. A meta-model is trained based on the predictions of the base models;
3. The final prediction is generated using the meta-model, which can also take into account the relationships between the base model predictions.

Random Forest Algorithm. The Random Forest algorithm is an ensemble model based on the Bagging method, where decision trees are used as base models. In a random forest, each tree is trained on a randomly selected subset of the data and a randomly selected subset of features. The final prediction is determined by aggregating the results of all trees, either through voting or by averaging. The Random Forest algorithm is widely used in problems where interpretability and reliable results are important. This algorithm has shown effective results in bioinformatics, disease prediction, and many other fields.

XGBoost – this algorithm uses regularization to reduce overfitting and parallel computations to accelerate processing. Therefore, XGBoost is widely applied in problems requiring high accuracy and efficiency.

Selecting the Ensemble Method. The choice of an ensemble method depends on the characteristics of the problem and the requirements for the model:

- For problems requiring high steadiness and reliability, the Bagging method, particularly the Random Forest algorithm, is most suitable;
- For complex problems with a large number of features where error minimization is necessary, deep learning methods such as Gradient Boosting or XGBoost provide the best results;
- When achieving the highest accuracy and combining different types of models is required, the stacking method is preferable.

JOURNAL OF ENGINEERING SCIENCES
ISSN(Online): 2984-6714
Volume-9, Issue-01, Published |09-02-2026|

The research methodology consists of the following stages:

- **Data collection** – obtained from patient complaints and anonymized clinical observations. The collected data includes age, sex, body temperature, blood pressure, heart rate, laboratory test results, and disease symptoms;
- **Data preprocessing and preparation** – missing values are replaced with averages or removed entirely. Symptoms are normalized. Categorical variables are converted to numeric format using methods such as one-hot encoding. The dataset is mixed randomly and split into training and testing sets;
- **Model selection and training** – several intelligent, independent algorithms based on ensemble methods are tested in the system. For example, K-Nearest Neighbors, Naive Bayes classifier, Decision Trees, as well as ensemble methods like Random Forest and Gradient Boosting. Each model is trained using cross-validation and evaluated based on high accuracy;
- **Analysis and comparison of results** – model performance is assessed using criteria such as accuracy, reliability, recall, F1-score, and confusion matrix;
- **Visualization of results** – data distribution, model accuracy, and error levels are presented graphically.

Selection of Training Dataset for Medical Diagnosis. The study uses an open-source medical dataset obtained from Kaggle.com as the training set. The following criteria are considered when selecting the dataset:

- **Reliability of collected data** – patient medical data must be obtained from trustworthy sources;
- **Number of symptoms** – the dataset should contain at least 10 symptoms and 1 class label;
- **Balance** – the number of healthy and diseased patients should be approximately equal;
- **Dataset size** – the number of objects in the dataset should not be less than 500.

The data in the dataset should have the following structure:

№	Symptom Name	Description	Type
1	Age	Patient's age	Numerical
2	Sex	Patient's sex (male/female)	Nominal
3	Blood Pressure	Patient's blood pressure	Numerical
4	Cholesterol	Cholesterol level	Numerical
5	Pulse	Heart rate	Numerical
6	Temperature	Body temperature	Numerical
7	Symptom1, Symptom2, ...	Presence of disease symptoms	Nominal
8	Diagnosis	Type of disease (class variable)	Nominal

The correct interpretation of the research methodology and training datasets is considered one of the key factors of an intelligent diagnostic system. High-quality data preprocessing, proper algorithm selection, and result analysis enable the system to operate with high accuracy [8].

For medical diagnosis problems, the following training datasets were selected:

1. **Cardiovascular Disease Training Dataset – Heart Disease Dataset** (https://www.kaggle.com/datasets/ronanazarias/heart-desease-dataset?select=Heart-disease)

Below is a description of the selected training dataset containing cardiovascular disease features:

- **Dataset name** – heart.csv
- **Number of records** – 918
- **Number of classes** – 2
- **Number of features (symptoms)** – 11

№	Feature Name	Description	Type
1	Age	Patient's age (years)	Numerical
2	Sex	Patient's sex (M: Male, F: Female)	Nominal
3	ChestPainType	Type of chest pain (TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic)	Nominal
4	RestingBP	Resting blood pressure (mm Hg)	Numerical
5	Cholesterol	Cholesterol level (mg/dl)	Numerical
6	FastingBS	Blood glucose level (1: if FastingBS > 120 mg/dl, 0: otherwise)	Nominal
7	RestingECG	Resting ECG results (Normal: normal, ST: ST-T wave abnormality, LVH: probable or definite left ventricular hypertrophy by Estes criteria)	Nominal
8	MaxHR	Maximum heart rate achieved (numeric value, 60–202)	Numerical
9	ExerciseAngina	Exercise-induced angina (Y: Yes, N: No)	Nominal
10	Oldpeak	Depression level: 0 – normal; 1–2 – mild depression; >2 – severe depression (high risk of ischemia)	Numerical
11	ST_Slope	Slope of the peak exercise ST segment	Nominal

2. Diabetes Training Dataset – Diabetes dataset
(https://www.kaggle.com/datasets/akshaydattatraykhare/diabetes-dataset)

JOURNAL OF ENGINEERING SCIENCES
ISSN(Online): 2984-6714
Volume-9, Issue-01, Published |09-02-2026|

The description of the selected training dataset containing features for diabetes disease is as follows:

- Dataset name – diabetes.csv
- Number of records – 768
- Number of classes – 2
- Number of features – 8

№	Feature Name	Description (EN)	Type
1	Pregnancies	Number of times the woman has been pregnant	Numerical
2	Glucose	Blood glucose level (mg/dL)	Numerical
3	BloodPressure	Blood pressure (mm Hg)	Numerical
4	SkinThickness	Thickness of subcutaneous fat (mm)	Numerical
5	Insulin	Insulin level (mu U/mL)	Numerical
6	BMI	Body Mass Index	Numerical
7	DiabetesPedigreeFunction	Coefficient of hereditary risk of diabetes	Numerical
8	Age	Patient's age (years)	Numerical

3. Multiclass Diabetes Training Dataset – Multiclass Diabetes Dataset (https://data.mendeley.com/datasets/wj9rwkp9c2/1)

The description of the selected multiclass diabetes dataset is as follows:

- Dataset name – multiclass_diabetes.csv
- Number of records – 1000
- Number of classes – 3
- Number of features – 12

№	Feature Name	Description (EN)	Type
1	No_Patient	Unique identifier for each patient. Used for statistics; has no medical meaning (only serial number)	Numerical
2	Gender	Patient's sex (M: Male, F: Female)	Nominal
3	Age	Patient's age (years)	Numerical

JOURNAL OF ENGINEERING SCIENCES
ISSN(Online): 2984-6714
Volume-9, Issue-01, Published |09-02-2026|

4	Urea	Feature indicating kidney function	Numerical
5	Cr	Creatinine level	Numerical
6	HbA1c	Blood sugar level over the last 3 months	Numerical
7	Chol	Total cholesterol level in blood	Numerical
8	TG	Type of fat in the blood	Numerical
9	HDL	High-density lipoprotein (“good cholesterol”)	Numerical
10	LDL	Low-density lipoprotein (“bad cholesterol”)	Numerical
11	VLDL	Very low-density lipoproteins	Numerical
12	BMI	Body Mass Index	Numerical

RESULTS

Using the RFC algorithm, informative features are extracted from the training datasets mentioned above, and a software package for predicting cardiovascular diseases and diabetes is developed.

By using the created intelligent system, it is possible to diagnose a number of diseases in patients based on the provided training datasets and similar datasets.

To ensure the correct operation of the software package, results for each training dataset presented in Table 1 are obtained using the intelligent system and ensemble methods:

- Heart disease dataset – results obtained using the cardiovascular disease training dataset:
  - Number of samples: 918
  - Number of features: 12

RangeIndex: 918 entries, 0 to 917
Data columns (total 12 columns):

#	Column	Non-Null Count	Dtype
0	Age	918 non-null	int64
1	Sex	918 non-null	object
2	ChestPainType	918 non-null	object
3	RestingBP	918 non-null	int64
4	Cholesterol	918 non-null	int64
5	FastingBS	918 non-null	int64
6	RestingECG	918 non-null	object
7	MaxHR	918 non-null	int64
8	ExerciseAngina	918 non-null	object
9	Oldpeak	918 non-null	float64
10	ST_Slope	918 non-null	object
11	target	918 non-null	int64

dtypes: float64(1), int64(6), object(5)

JOURNAL OF ENGINEERING SCIENCES
ISSN(Online): 2984-6714
Volume-9, Issue-01, Published |09-02-2026|

Target column distribution:
target
1 508
0 410

Splitting data into training and test sets and normalization

✓ **Training set: 734 samples**

✓ **Test set: 184 samples**

Model evaluation

🔻 **Random Forest:**

Accuracy: 0.8859 (88.59%)
Precision: 0.8859
Recall: 0.8859
F1-Score: 0.8856

🔻 **Gradient Boosting:**

Accuracy: 0.8641 (86.41%)
Precision: 0.8660
Recall: 0.8641
F1-Score: 0.8644

🔻 **Extra Trees:**

Accuracy: 0.8913 (89.13%)
Precision: 0.8915
Recall: 0.8913
F1-Score: 0.8910

🔻 **Bagging:**

Accuracy: 0.8859 (88.59%)
Precision: 0.8859
Recall: 0.8859
F1-Score: 0.8856

BEST MODEL: Extra Trees
Accuracy: 89.13%

- **Diabetes Dataset – Results obtained using the diabetes training dataset:**
- Number of samples: 768
- Number of features: 9

RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):

#	Column	Non-Null Count	Dtype
0	Pregnancies	768 non-null	int64
1	Glucose	768 non-null	int64
2	BloodPressure	768 non-null	int64
3	SkinThickness	768 non-null	int64
4	Insulin	768 non-null	int64
5	BMI	768 non-null	float64
6	DiabetesPedigreeFunction	768 non-null	float64
7	Age	768 non-null	int64

JOURNAL OF ENGINEERING SCIENCES
ISSN(Online): 2984-6714
Volume-9, Issue-01, Published |09-02-2026|

8 target 768 non-null int64
dtypes: float64(2), int64(7)
Target column distribution:
Target
0 500
1 268
Splitting data into training and test sets and normalization:
✓ Training set: 614 samples
✓ Test set: 154 samples
Model Evaluation:
🔻 **Random Forest:**
Accuracy: 0.7532 (75.32%)
Precision: 0.7497
Recall: 0.7532
F1-Score: 0.7509
🔻 **Gradient Boosting:**
Accuracy: 0.7597 (75.97%)
Precision: 0.7556
Recall: 0.7597
F1-Score: 0.7568
🔻 **Extra Trees:**
Accuracy: 0.7273 (72.73%)
Precision: 0.7182
Recall: 0.7273
F1-Score: 0.7179
🔻 **Bagging:**
Accuracy: 0.7468 (74.68%)
Precision: 0.7423
Recall: 0.7468
F1-Score: 0.7437
BEST MODEL: Gradient Boosting
Accuracy: 75.97%

- **Multiclass Diabetes Dataset – Results obtained using the multiclass diabetes training dataset:**
  - Number of samples: 1000
  - Number of features: 13

RangeIndex: 1000 entries, 0 to 999
Data columns (total 13 columns):
Column Non-Null Count Dtype
0 No_Pation 1000 non-null int64
1 Gender 1000 non-null object
2 AGE 1000 non-null int64
3 Urea 1000 non-null float64
4 Cr 1000 non-null int64
5 HbA1c 1000 non-null float64

JOURNAL OF ENGINEERING SCIENCES
ISSN(Online): 2984-6714
Volume-9, Issue-01, Published |09-02-2026|

6 Chol 1000 non-null float64
7 TG 1000 non-null float64
8 HDL 1000 non-null float64
9 LDL 1000 non-null float64
10 VLDL 1000 non-null float64
11 BMI 1000 non-null float64
12 target 1000 non-null object
dtypes: float64(8), int64(3), object(2)
Target column distribution:
Target
Y 844
N 103
P 53
Splitting data into training and test sets and normalization:
✓ Training set: 800 samples
✓ Test set: 200 samples
Model Evaluation:
🔻 **Random Forest:**
Accuracy: 0.9700 (97.00%)
Precision: 0.9699
Recall: 0.9700
F1-Score: 0.9698
🔻 **Gradient Boosting:**
Accuracy: 0.9900 (99.00%)
Precision: 0.9909
Recall: 0.9900
F1-Score: 0.9902
🔻 **Extra Trees:**
Accuracy: 0.9600 (96.00%)
Precision: 0.9590
Recall: 0.9600
F1-Score: 0.9590
🔻 **Bagging:**
Accuracy: 0.9700 (97.00%)
Precision: 0.9713
Recall: 0.9700
F1-Score: 0.9699
BEST MODEL: Gradient Boosting
Accuracy: 99.00%

The comparative analysis table of the main classifiers in Weka using cross-validation and the results obtained via ensemble methods for the datasets used above is presented below:

JOURNAL OF ENGINEERING SCIENCES
ISSN(Online): 2984-6714
Volume-9, Issue-01, Published |09-02-2026|

Dataset	Decision Tree	Random Forest	NaïveBayes	Logistic Regression	Ensemble Method
Heart	0.8811	0.9045	0.8526	0.8700	0.8913
Diabetes	0.7057	0.7617	0.7565	0.7722	0.7600
Multiclass Diabetes	0.9900	0.9814	0.7720	0.9709	0.9900

The results indicate that the accuracy values obtained using the ensemble method are almost identical to the results of other algorithms and in some cases demonstrate superiority.

CONCLUSION

Ensemble methods are an effective approach for improving the accuracy and stability of machine learning models. Techniques such as bagging, boosting, and stacking offer different strategies for combining models and enable achieving high performance in both classification and regression tasks. This provides a unique approach to solving problems related to disease diagnosis and prediction. Applying ensemble methods requires careful attention and the correct choice of strategy; however, their practical implementation significantly enhances model efficiency and reliability across various domains.

REFERENCES

1. Djabbarov O.R., Eshboyev E.A. and Klicheva F.G. "Sun'iy immun tizimlari algoritmlarining kasalliklarni aniqlash va tasniflash masalalarida qo'llanilishi." MODERN PROBLEMS AND PROSPECTS OF APPLIED MATHEMATICS 1.01 (2024).
2. Andrew D. Chapman. 2023. *Artificial Intelligence and Machine Learning*. Nederland, CO: 439 pages.
3. Eshboyev E. et al. Using the RFC_PSO hybrid algorithm in sorting informative features //AIP Conference Proceedings. – AIP Publishing LLC, 2025. – T. 3356. – №. 1. – C. 030002.
4. Prasasd Babu M. S. and Katta S., "Artificial immune recognition systems in medical diagnosis," *2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China, 2015, pp. 1082-1087, doi: 10.1109/ICSESS.2015.7339240.
5. G. H. John, R. Kohavi, K. Pfleger, "Irrelevant feature and the subset selection problem," in Proc. of the Eleventh International Conference on Machine Learning, pp. 121-129, 1994.
6. Silva, G.C., Caminhas, W.M. and Palhares, R.M., 2017. Artificial immune systems applied to fault detection and isolation: A brief review of immune response-based approaches and a case study. Applied Soft Computing, 57, pp.118- 131.
7. Breiman L. Random Forests. Machine Learning, 2001, no. 45, pp. 5–32.
8. Ishan Gupta, Ruchir Shangleand others. Cardiovascular Disease Detection using Artificial Immune System and other Machine Learning Models. ICMAI 2021. Journal of Physics: Conference Series 1950 (2021) 012032 IOP Publishing doi:10.1088/1742-6596/1950/1/012032.