DOI: https://doi.org/10.15276/aait.09.2026.10
UDC 004.1

Development of a unified feature space for ensemble classification of polystructural heterogeneous small data

Olena O. Arsirii[1)]
ORCID: https://orcid.org/0000‒0001‒8130‒9613; e.arsiriy@gmail.com. Scopus Author ID: 54419480900
Oleksandr K. Andronati[1)]
ORCID: https://orcid.org/0009-0009-1794-5864; alex.andronati@gmail.com. Scopus Author ID: 58677655800
[1)] Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine

ABSTRACT

Relevance. The paper addresses the problem of classification of polystructural heterogeneous small data, which include structured tabular information and unstructured audio signals. **Objective**. The study proposes a unified methodological framework for constructing an ensemble classification system based on a common feature space that ensures compatibility between different data types and machine learning models. **Main Research Material.** A unified feature space model is developed, incorporating a linear representation for structured data and a three-level representation for audio data, including tabular spectral features, spectrogram-based image representation, and temporal sequences of features. To improve the quality of input data, a Backward Feature Elimination procedure is applied to adapt feature subsets to the specifics of individual classifiers and the removal of non-informative features. The classification framework is based on a stacking ensemble architecture that combines multiple base models, including classical machine learning algorithms and deep learning models. Three aggregation strategies are considered: hard voting, soft voting, and soft voting with Gompertz fuzzy ranking, which enables nonlinear adjustment of classifier probabilities and improves robustness under uncertainty. **Results**. Experimental evaluation was conducted on five datasets from different domains, including healthcare, finance, audio signal analysis, and deepfake detection. The results demonstrate that the proposed approach consistently improves classification performance compared to individual models. The application of feature selection and the integration of ensemble methods provides significant gains for polystructural data. **Conclusions.** The proposed model offers a flexible and scalable solution for handling heterogeneous small data and can be effectively applied across multiple domains, providing improved generalization, robustness to noise, and adaptability to different data representations.

Keywords: Polystructural data; heterogeneous data; small data; ensemble classifiers; Gompertz function; audio classification

***For citation*: Arsirii O. O., Andronati O. K. “Development of a unified feature space for ensemble classification of polystructural heterogeneous small data”. *Applied Aspects of Information Technology.* 2026; Vol.9 No.2: 147–157. DOI: https://doi.org/10.15276/aait.09.2026.10**

INTRODUCTION

The current stage of development in intelligent data analysis systems is characterized by the rapid rise of deep learning architectures, which have demonstrated impressive results in image recognition, natural language processing, and computer vision tasks. However, the functioning of such models critically depends on the availability of massive amounts of labeled data and colossal computational resources. At the same time, the real-world sectors of the economy and science often face a challenge known in global practice as “Small Data” – a situation where the volume of available data is limited and its nature is extremely diversive [1], [2].

Particularly challenging is heterogeneous polystructural small data, which combines structured tabular information (e.g., medical indicators or financial transactions) and unstructured time-series signals (audio recordings of emotional speech, music, or synthesized speech). Direct application of massive neural network models to such data often leads to overfitting and low robustness to noise [3].

Furthermore, the question of deploying models on local systems with limited resources remains relevant, where sequential ensemble training is more rational than maintaining massive neural network infrastructures.

The key scientific challenge lies in the absence of a unified methodological approach to feature space design that would allow for the unification of representations of such heterogeneous data. Traditional methods often treat different modalities separately, which negates the potential for combining them within ensemble classifiers. The need to create a model capable of transforming audio signals and tabular records into a standardized digital format adapted to the specifics of particular base algorithms (SVM, XGBoost, CNN, LSTM, etc.) determines the relevance of this research.

The aim of this article is to develop a model and methods for creating a unified feature space based on the use of a linear data model for structured data and a more complex three-level model for unstructured data, presented in this work using audio data. This approach allows for the optimization of input data for a wide range of machine learning and deep learning models, ensuring improved accuracy and reliability in the classification of small-scale, heterogeneous, polystructural data in multi-domain application tasks.

In previous works, the authors demonstrated the effectiveness of ensemble classifiers for processing credit and medical data, as well as audio data, such as human speech emotion recognition, music genre classification, and deepfake analysis [4].

LITERATURE REVIEW AND PROBLEM STATEMENT

The challenge of effectively processing multi-structured heterogeneous data in domains such as finance, medicine, audio data analysis, and others has attracted significant attention from researchers in the field of intelligent data analysis. The modern scientific literature proposes various approaches to analyzing structured and unstructured data using different machine learning and deep learning algorithms. However, despite the rapid development of deep neural networks, a significant number of practical problems are characterized by limited data samples, which complicates the use of complex parameterized models. In the scientific literature, this phenomenon is often described as the “Small Data problem,” which consists of an insufficient amount of training data for stable generalization of machine learning models. In literature sources [1], [5], it is noted that limited samples significantly complicate statistical analysis and increase the risk of model instability.

In this regard, feature space optimization is a crucial component of effective machine learning, as it allows for the removal of redundant or irrelevant features and improves the generalization ability of models. Various feature selection methods are widely used for this purpose. Among these approaches are forward feature elimination and backward feature elimination, which involve the gradual inclusion or exclusion of features based on an assessment of the model’s performance. An example of applying such approach to classification tasks is presented in [6], where it is shown that the use of sequential backward selection reduces data dimensionality and improves classification accuracy.

A separate area of research focuses on developing effective methods for representing audio signals for machine learning tasks. The most common approach is the use of mel-frequency cepstral coefficients (MFCCs), which allow for a compact description of a signal’s spectral characteristics and are widely used in tasks involving the recognition of audio emotions, music, and fake speech [7], [8], [9].

The next step in solving any classification problem is training and testing machine learning algorithms. Literature sources note that various machine learning methods are used to develop basic classifiers for further assembly, including K-nearest neighbors (KNN), support vector machines (SVM), random forest, extreme gradient boosting (XGB), logistic regression (LR), multilayer perceptions (MLP), convolutional neural networks (CNN), and recurrent neural networks (LSTM). Each of these methods has its advantages, but also has significant limitations.

According to literature sources [10], [11], [12], [13], [14], the limitations of basic models, such as sensitivity to noise, overfitting, and inability to effectively handle nonlinear dependencies or class imbalance, make them insufficiently versatile for working with polystructural heterogeneous small data. For example, regarding logistic regression, study [14] shows that logistic regression does not achieve sufficiently high classification quality on banking data, especially on data with nonlinear or complex dependencies.

To overcome these limitations, one of the best solutions is to use ensemble machine learning methods, which involve combining the results of several base models. Such approaches allow for improving the models’ robustness to noise and reducing the risk of over fitting by utilizing various generalization mechanisms. A detailed overview of modern ensemble algorithms, including bagging, boosting, and stacking, is provided in [15], where it is shown that combining models of different types allows for higher accuracy across a wide range of application tasks.

Despite a significant body of research in the fields of machine learning and audio signal processing, the scientific literature has not sufficiently explored the issue of constructing a unified feature space that would allow for the effective integration of structured tabular data and spectral characteristics of time-domain signals within a single classification system.

Thus, the task of developing a unified feature space model that adapts various data types to the specifics of machine learning algorithms with subsequent ensemble learning is a promising research direction.

THE AIM AND OBJECTIVES OF THE RESEARCH

The aim of the research is to develop a model and methods for creating a unified feature space that ensures effective integration of structured tabular data and unstructured audio signals with subsequent ensembling to improve classification accuracy and robustness when dealing with polystructural heterogeneous small data.

To achieve this aim, the following objectives were identified:

1) to develop a model of a unified feature space based on a linear representation for structured data and a three-level model for unstructured audio data (tabular, spectral, and temporal levels) for datasets in different domains such as healthcare, finance, information security and audio content analysis;
2) to adapt the input feature sets for specific machine learning and deep learning algorithms using the Backward Feature Elimination (BFE) procedure to reduce dimensionality and mitigate overfitting;
3) to design an ensemble classification framework based on the stacking architecture and different types of prediction aggregation to enhance the classification quality;
4) to evaluate the effectiveness of the proposed approach through experimental studies on diverse datasets, comparing the ensemble performance against individual base classifiers.

THE RESEARCH MATERIALS AND METHODS

A Unified Feature Space Model for Polystructural Heterogeneous Data

In this work unified feature space is defined as the space that is formed for the base classifiers, which is created using the BFE procedure described later in the paper.

In developing a unified feature space model for polystructural heterogeneous data, two types of data are considered: structured data, which consists of pre-formatted tabular data, and unstructured data. Audio data is considered as unstructured data in this work.

The study uses five datasets covering both structured and unstructured data.

Structured data is represented by two datasets.

1. Home Equity Line of Credit (HELOC) [16]: Contains 10,459 records with 23 attributes, including financial indicators (time since last delinquency, number of transactions). The target variable is credit risk (binary). The dataset was split into 7,844 records in the training set and 2,615 in the testing set.

2. Autistic Spectrum Disorder Screening Data for Adults (ASD) [17]: contains 800 records with 13 attributes, including binary behavioral characteristics (AQ-10) and demographic data (age, gender). The target variable is the presence/absence of autism. The dataset was split into 600 records in training set and 200 records in testing set.

Unstructured data is represented by three datasets.

1. Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) [18]: 1,440 audio recordings with 7 emotions (anger, happiness, etc.). The dataset was split into 1080 records in the training set, 360 in the testing set.

2. GTZAN [19]: 1000 audio files, 30 seconds each, 10 genres (100 files per genre). The dataset was split into 750 records in the training set and 250 in testing set.

3. Fake-or-Real (FOR) [20]: 17870 audio files for synthetic speech detection (real and fake audio). The dataset was split into 13,956 in the training set and 3,914 in the testing set.

According to [2] The concept of small data is fundamentally relative, as its definition depends on the complexity of the machine learning task and the specific model architecture; for instance, a dataset is considered “small” if it provides fewer data points than typically required to train a model optimally. For example in [2] it is also noted that for large language models (LLMs), the boundaries of “small data” shift significantly: even billions of text tokens can be classified as small data if they are insufficient for optimally training a model with a vast number of parameters.

For the complex architectures used in their research, such as CNNs and LSTMs, these sample sizes are insufficient for stable generalization, creating a high risk of overfitting that necessitates specialized techniques like Backward Feature Elimination and ensemble stacking. Thus, in this study, all the datasets are classified as small data.

For structured data, a linear data model is proposed:

$$X = \{a_1x_1, a_2x_2, a_3x_3, .. a_mx_m\}, x_i \in R^m, \qquad (1)$$

where m is the number of all features, $a_i = \begin{cases} 1, & when\, x_i \in R^n \\ 0, & otherwise \end{cases}$, R^n is optimal feature space, n is number of features in optimal feature space.

This type of data representation allows it to be used as input for algorithms such as Logistic Regression, Support Vector Machine, k-Nearest Neighbors, Random Forest, Extreme gradient boosting, Multilayer Perceptron.

For the analysis of audio data, a three-level model is proposed, which allows the same audio signal to be considered in multiple feature spaces. This approach ensures the extraction of diverse features and enhances the potential for the subsequent use of ensemble methods.

The three-level model includes the following levels of audio data representation:

1) tabular representation of spectral characteristics;

2) spectral representation of the signal as an image;

3) time representation of the audio signal as a numerical sequence.

Using multiple levels of representation allows us to extract different types of information from a single audio signal. For example, spectral characteristics reflect the frequency properties of the signal, while the time series allows analyzing the dynamics of the signal's change over time.

Let the audio signal be represented as a time sequence of amplitude values:

$$x(t) = \{x_1, x_2, \dots, x_T\}, \tag{2}$$

where T is the number of discrete time samples of the signal.

Within the proposed three-level model, the audio signal is transformed into three different representation spaces:

$$R(x) = \{R_{tab}(x), R_{img}(x), R_{seq}(x)\}, \tag{3}$$

where $R_{tab}(x)$ is a tabular representation of spectral characteristics; $R_{img}(x)$ is a representation of the signal as a spectral image; $R_{seq}(x)$ is a representation of the signal as a time series.

At the first level of the three-level model – the tabular representation of spectral characteristics – the audio signal is mapped to a feature vector of fixed length:

$$R_{tab}(x) = f(x) \in R^m, \tag{4}$$

where *f(x)* is a generalization of a function for obtaining spectral characteristics; *m* is the number of features obtained.

For a set of audio files, the following data matrix is formed:

$$X_{tab} = \{a_1 f_1(x), a_2 f_2(x), \dots, a_m f_m(x)\} \epsilon R^m, \tag{5}$$

where m is the number of all features, $f_m(x)$ is the m-th spectral feature; $a_i = \begin{cases} 1, & when\ x_i \in R^n \\ 0, & otherwise \end{cases}$, R^n is optimal feature space, n is number of features in optimal feature space.

The resulting tabular representation of the spectral characteristics of the signal X_{tab} is used as input data for machine learning algorithms that work with vector features.

Since an audio signal in the time domain does not allow for a direct analysis of the frequency-domain energy distribution, the Short-Time Fourier Transform (STFT) is used to obtain a frequency-time representation of the signal.

Let *x(t)* be a discrete audio signal. Then we denote the frequency-time representation of the signal (spectrum) as

$$X(k,t) = \left|STFT\left(x(t)\right)\right|. \tag{6}$$

The resulting frequency-time representation of the signal is used to compute spectral characteristics.

According to studies [21], [22], [23], the most promising spectral features for audio data classification are mel-frequency cepstral coefficients, spectral centroid, spectral decay, spectral flatness, zero-crossing frequency, spectral contrast, and others.

Mel-frequency cepstral coefficients (MFCCs), which are used for analyzing audio signals, allow for a compact description of the shape of a signal's spectral envelope, taking into account the peculiarities of human sound perception.

The process of obtaining MFCCs involves several stages [9]: the signal is converted into a frequency-time representation *X(k,t)* using the short-time Fourier transform. Next, the spectral energy of the signal is calculated. After that, Mel filtering is applied, which converts the frequency scale to the Mel scale. Since human perception of loudness is logarithmic, the mel spectrogram is converted to decibels using logarithmization. To obtain a compact representation of the spectral envelope, the discrete cosine transform is applied to the logarithm of the energy of the Mel filters.

In this study, K=20 MFCC coefficients are calculated for each audio file for all three unstructural datasets. Since a tabular representation of the signal is formed at the first level of the three-level model, averaging over all time frames is used for each coefficient.

The result is a fixed-length vector of MFCC features, which is used as part of a tabular representation of the audio signal's spectral characteristics.

Next, the remaining spectral characteristics are calculated and aggregated over time frames, after which they are added to the MFCC feature vector

The second level of the three-level model for representing audio data is based on the use of a spectral representation of the signal in the form of an image – a spectrogram.

Unlike the tabular representation, where spectral characteristics are aggregated by averaging over time, in this case the full frequency-time structure of the signal is preserved.

For this purpose, a decibel-normalized mel-spectrogram is interpreted as an image. Then it is scaled to a size of 128×128 pixels.

This representation allows algorithms that work with images, in particular convolutional neural networks, to be applied to audio data.

The third level of the proposed audio data representation model is based on the use of a time series of mel-frequency cepstral coefficients (MFCCs).

Thus, the audio signal is described by a sequence of MFCC vectors:

$$X = \{c(1), c(2), \ldots, c(T)\} . \quad (7)$$

In matrix form:

$$X = \begin{bmatrix} c_1(1) & c_2(1) & \cdots & c_K(1) \\ c_1(2) & c_2(2) & \cdots & c_K(2) \\ \vdots & \vdots & \ddots & \vdots \\ c_1(T) & c_2(T) & \cdots & c_K(T) \end{bmatrix}, X \in R^{K \times T}, \quad (8)$$

where T is the number of time frames; K is the number of MFCC coefficients.

Since the length of audio files in the dataset may vary, the number of time frames T may also differ. To ensure data compatibility with machine learning algorithms, the matrices are reduced to the same length.

As a result, the third level of the model represents the audio signal as a time series of spectral features, which allows for the use of models designed for sequential data, such as recurrent neural networks.

Deriving a feature space using the BFE algorithm

Using the full set of features does not always guarantee improved classification quality, as some features may be redundant, and individual variables may contain noise or be uninformative.

Forming different feature subspaces allows reducing the correlation of errors between base models and adapting the feature set to the specifics of a particular algorithm.

The algorithm operates according to an iterative scheme with a gradual reduction of the feature space.

1. Initially, the full feature set is used.
2. At each iteration, a candidate subset without one of the features is formed.
3. For each candidate set, the model is trained on a subset of the training dataset (validation set) and the weighted F1-score metric is calculated.
4. After evaluating all options, the feature is identified whose removal yields the greatest improvement in model quality.
5. If such an improvement exists, only one feature is removed from the current feature set – the one whose removal resulted in the maximum F1-score.
6. The procedure is repeated until removing any feature no longer improves the metric or the minimum number of features is reached.

Thus, at each iteration, the algorithm removes the least informative feature, allowing for the gradual formation of an optimal feature subspace for a specific model.

Table 1 lists the features selected by the BFE algorithm for each of the base classifiers for the structured HELOC dataset. In Table 1 a_i from formula (1) takes the values {1, 0}.

Table 1. **Features selected by BFE for HELOC**

Algorithm	**SVM**	**XGB**	**KNN**	**RF**	**LR**	**MLP**
MSinceOldestTradeOpen	0	1	1	1	1	1
MSinceMostRecentTradeOpen	1	1	1	1	0	+
AverageMInFile	1	1	1	1	1	1
NumSatisfactoryTrades	1	1	1	1	1	1
NumTrades60Ever2DerogPubRec	1	1	1	1	0	1
NumTrades90Ever2DerogPubRec	1	1	0	1	1	1
PercentTradesNeverDelq	1	1	1	1	1	1
MSinceMostRecentDelq	1	1	1	1	0	1
MaxDelq2PublicRecLast12M	1	0	1	1	1	1
MaxDelqEver	1	1	1	1	0	1
NumTotalTrades	0	1	1	1	0	1
NumTradesOpeninLast12M	1	1	1	1	1	1
PercentInstallTrades	1	1	0	1	1	1
MSinceMostRecentInqexcl7days	1	0	1	1	1	1
NumInqLast6M	1	1	1	1	1	1
NumInqLast6Mexcl7days	0	1	1	1	0	1
NetFractionRevolvingBurden	1	1	1	1	1	1
NetFractionInstallBurden	1	1	0	1	1	1
NumRevolvingTradesWBalance	1	1	0	1	1	1
NumInstallTradesWBalance	1	0	1	1	0	1
umBank2NatlTradesWHighUtilization	1	1	1	0	1	1
PercentTradesWBalance	1	0	1	1	0	1

***Source*: compiled by the authors**

Table 2 lists the features selected by the BFE algorithm for the first (tabular) level of representation of the RAVDESS unstructured dataset for each of the base classifiers. In Table 2 a_i from formula (5) takes the values {1,0}.

Additionally, mel-spectrograms were used as images for CNN, and MFCC were used as a time series input for LSTM.

A unified feature space was obtained in a similar manner for the GTZAN and HELOC datasets.

Unlike other datasets, the Backward Feature Elimination (BFE) algorithm was not applied to the ASD dataset because the feature set has a low dimensionality, and most variables directly correspond to the results of the AQ-10 survey.

Table 2. **Features selected by BFE for RAVDESS**

Algorithm	**SVM**	**XGB**	**KNN**	**RF**	**LR**	**MLP**
mfcc_0	1	1	1	1	1	1
mfcc_1	1	1	1	1	1	1
mfcc_2	1	1	1	1	1	1
mfcc_3	1	1	1	1	1	1
mfcc_4	1	1	1	1	1	1
mfcc_5	1	1	1	1	1	1
mfcc_6	1	1	0	1	1	1
mfcc_7	1	1	1	1	1	1
mfcc_8	1	1	1	1	1	1
mfcc_9	1	1	1	1	1	1
mfcc_10	1	1	1	1	1	1
mfcc_11	1	1	1	1	1	1
mfcc_12	1	1	1	1	1	1
mfcc_13	1	1	1	1	1	1
mfcc_14	1	1	0	1	1	1
mfcc_15	0	1	1	1	1	0
mfcc_16	1	1	0	1	1	1
mfcc_17	1	1	0	1	0	1
mfcc_18	1	1	1	1	1	1
mfcc_19	1	1	1	1	1	1
spectral contrast_0	1	1	1	1	1	1
spectral contrast_1	0	1	1	1	1	1
spectral contrast_2	1	1	1	1	1	1
spectral contrast_3	0	1	1	1	1	1
spectral contrast_4	1	1	1	1	1	1
spectral contrast_5	1	1	0	1	1	1
spectral contrast_6	1	1	1	1	1	0
spectral_centroid	1	1	0	1	0	0
spectral_flatness	1	1	1	1	1	1
spectral_rolloff	1	1	0	1	1	1
spectral_bandwidth	1	1	1	1	1	1
zero_crossing_rate	0	1	0	1	1	1
gender	1	0	0	1	1	1

***Source*: compiled by the authors**

After selecting the feature space, the base classifiers are trained for subsequent ensembling.

Building an Ensemble Classifier

To improve the classification accuracy of polystructural heterogeneous data, it is proposed to use an ensemble architecture based on the stacking method.

This approach involves combining the predictions of several base models by a meta-classifier. The advantage of the stacking architecture is its ability to effectively combine models of different types, allowing the strengths of each algorithm to be leveraged.

Within the proposed methodology, three approaches are used to aggregate the predictions of the base models: hard voting, soft voting, and soft voting using Gompertz fuzzy ranking.

In the hard voting method, the ensemble's final decision is determined by a majority vote of the individual classifiers.

Each model generates a predicted class, after which the class that received the most votes is selected:

$$\hat{y} = \mathrm{mode}(y_1, y_2, \ldots, y_n)\,, \qquad (9)$$

where $\hat{y}$ is the final predicted class determined by the ensemble; y_i is the predicted class by the i-th individual classifier, where i ranges from 1 to n; n is the total number of individual classifiers in the ensemble.

In the soft voting method, each classifier generates probabilities of an object's class membership, and the final decision is determined by averaging these probabilities:

$$\mathrm{p}_\mathrm{k} = (1/\mathrm{n})\ \Sigma\ \mathrm{p}_\mathrm{ik}\ , \qquad (10)$$
$$\hat{\mathrm{y}} = \mathrm{argmax}\ \mathrm{p}_\mathrm{k},$$

where $\hat{y}$ is the final predicted class determined by the ensemble; p_{ik} is the probability of class k predicted by the i-th classifier, where i ranges from 1 to n and k represents the class index; n is the total number of individual classifiers in the ensemble.

To improve the robustness against noise and uncertainty in the data, a modified version of soft voting is used, employing the Gompertz fuzzy function, which allows for the nonlinear transformation of classifier probabilities:

$$p'_{i,k} = a \cdot exp(-b \cdot exp(-c \cdot p_{i,k})),$$
$$p_k = (1/n)\ \Sigma\ p'_{i,k}, \qquad (11)$$
$$\hat{y} = \mathrm{argmax}_k\ p_k,$$

where $\hat{y}$ is the final predicted class determined by the ensemble; $p'_{i,k}$ is the adjusted probability of class k for the i-th classifier after applying the Gompertz function; p_k is the average probability of class k across all classifiers in the ensemble; n is the total number of individual classifiers in the ensemble; a, b, c are the parameters of the Gompertz function used for adjusting probabilities, where a controls the upper asymptote, and b and c control the shape of the curve. The values of a, b, c were obtained empirically.

Soft voting using fuzzy ranks (Gompertz) allows taking into account uncertainty and heterogeneity in the data, which improves the accuracy and stability of the ensemble classifier in the classification of data [4], [23], [24]. Statistically, the use of this approach improves the metrics in comparison with usual soft voting.

As part of the experimental study, ensembles with different combinations of base classifiers were constructed. For datasets representing structural data

– ASD and HELOC – six base models were used: SVM, XGB, KNN, Random Forest, Logistic Regression, and MLP. For the RAVDESS, GTZAN, and FOR audio datasets, an expanded set of eight models was applied, including CNN for analyzing the mel-spectrogram as an image and LSTM for analyzing the time series of mel-frequency cepstral coefficients as a sequence.

Taking into account various combinations of base classifiers and three aggregation methods, the following number of ensemble configurations were formed for ASD and HELOC:

$$C_n^k = 3\,(C_6^3 + C_6^4 + C_6^5 + C_6^6) = 126.$$

For RAVDESS, GTZAN and FOR datasets:

$$C_n^k = 3\,(C_8^3 + C_8^4 + C_8^5 + C_8^6 + C_8^7 + C_8^8) = 657,$$

where C_n^k is the number of combinations from n to k, multiplication with a factor of 3 is explained by the fact that we have three types of ensemble classifiers aggregation – hard voting, soft voting, soft voting with Gompertz aggregation.

Despite the large number of ensembles, the computational complexity remains moderate, since the main computations are associated with training the base models and feature selection using the BFE algorithm, which is performed only once for each model. Different ensembles are formed by combining pre-trained classifiers, which does not require significant additional computational resources.

The performance of the ensembles was evaluated on test datasets using key classification quality metrics: accuracy and F1-score. The final selection of the best ensemble configurations was based on the maximum values of accuracy and F1-score, with preference given to the model with the higher F1-score in cases of discrepancy.

RESULTS OF THE STUDY

To experimentally verify the effectiveness of the proposed model for constructing a unified feature space, a series of experiments was conducted on five datasets: ASD, HELOC, RAVDESS, GTZAN, FOR.

The first two datasets contain structured tabular features, while the last three represent audio signals. Thus, the experimental dataset encompasses small-scale, heterogeneous, polystructural data, allowing to assess the versatility of the proposed approach.

The experimental study was conducted in several stages.

1. Evaluation of the impact of the BFE procedure on individual classifiers

In the first stage, a comparison of the classification quality of individual machine learning models was conducted in two variants: using a standard feature set; using features after the BFE procedure (described by the authors in "*Deriving a feature space using the BFE algorithm"*).

This allowed evaluating the impact of the proposed method for forming a unified feature space before using ensemble models.

The results of comparing individual classifiers with a standard feature set and using the BFE procedure for structured data, using the HELOC dataset as an example, are presented in Table 3 (accuracy) and Table 4 (F1-score).

Table 3. **Accuracy difference for HELOC**

Algorithm	**Accuracy for HELOC**		
	Full features	***Features after BFE***	***Accuracy difference***
SVM	0.720	0.727	0.7%
MLP	0.725	0.715	-1%
XGB	0.718	0.712	-0.6%
RF	0.724	0.728	0.4%
KNN	0.685	0.691	0.5%
LR	0.720	0.728	0.8%

***Source*: compiled by the authors**

Table 4. **F1-score difference for HELOC**

Algorithm	**F1-score for HELOC**		
	Full features	***Features after BFE***	***Accuracy difference***
SVM	0.720	0.727	0.7%
MLP	0.723	0.716	-0.7%
XGB	0.717	0.712	-0.5%
RF	0.723	0.727	0.4%
KNN	0.685	0.691	0.6%
LR	0.720	0.728	0.8%

***Source*: compiled by the authors**

The results of comparing individual classifiers with a standard feature set and using the BFE procedure for unstructured data, using the RAVDESS dataset as an example, are presented in Table 5 (accuracy) and Table 6 (F1-score).

Table 5. **Accuracy difference for RAVDESS**

Algorithm	**Accuracy for RAVDESS**		
	Full features	***Features after BFE***	***Accuracy difference***
SVM	0.700	0.747	4.7 %
MLP	0.694	0.714	2 %
XGB	0.675	0.686	1.1 %
RF	0.586	0.625	3.9 %
KNN	0.617	0.669	5.2 %
LR	0.578	0.583	0.5 %

***Source*: compiled by the authors**

Table 6. **F1-score before and after BFE for RAVDESS**

Algorithm	**F1-score for RAVDESS**		
	Full features	***Features after BFE***	***Accuracy difference***
SVM	0.696	0.745	4.9 %
MLP	0.692	0.712	2 %
XGB	0.671	0.682	1.1 %
RF	0.571	0.615	4.4 %
KNN	0.606	0.660	5.4 %
LR	0.554	0.559	0.5 %

***Source*: compiled by the authors**

An analysis of the results showed that applying the BFE procedure leads to improved classification quality in most cases. This is because BFE adapts features to the specific characteristics of individual machine learning algorithms, thereby increasing the informativeness of the feature space. A similar improvement was observed for other datasets.

2. Comparison of individual models and ensemble classifiers

In the second stage, the performance of ensemble models built using a stacking architecture was analyzed. For each dataset, a set of ensembles with different combinations of base classifiers and three prediction aggregation methods was created.

For the HELOC dataset, which represents structural data, out of 126 designed ensemble classifiers; the best results were achieved by an ensemble consisting of KNN, SVM, Random Forest, XGB, and Logistic Regression with soft voting. The following metrics were obtained: accuracy = 0.739 (+1.4 % relative to the best classifier in the ensemble); F1 Score = 0.738 (+1.5 % relative to the best classifier in the ensemble).

Complete information comparing the best ensemble with individual classifiers before BFE is given in Table 7.

Table 7. **Best ensemble for HELOC**

Algorithm	**Metrics for HELOC**			
	Accuracy	***F1-score***	***Accuracy difference***	***F1-score difference***
Ensemble	***0.739***	***0.738***		
SVM	0.720	0.720	1.9 %	1.8 %
MLP	0.725	0.723	1.4 %	1.5 %
XGB	0.718	0.717	2.1 %	2.1 %
RF	0.724	0.723	1.5 %	1.5 %
KNN	0.685	0.685	5.4 %	5.2 %
LR	0.720	0.720	1.9 %	1.8 %

***Source*: compiled by the authors**

For the RAVDESS dataset which represents unstructural data, out of 657 designed ensemble classifiers, the best results were shown by ensemble consisting of KNN, SVM, XGB, CNN, LSTM with soft voting using fuzzy ranks (Gompertz) aggregation. The following metrics were obtained: accuracy = 0.822 (+12.2 % relative to the best classifier in the compound); F1 Score = 0.819 (+12.2 % relative to the best classifier in the compound).

Complete information comparing the best ensemble with individual classifiers before BFE is given in Table 8.

Table 8. **Best ensemble for RAVDESS**

Algorithm	**Metrics for emotion classification**			
	Accuracy	***F1-score***	***Accuracy difference***	***F1-score difference***
Ensemble	***0.822***	***0.819***		
SVM	0.700	0.696	12.2 %	12.3 %
MLP	0.694	0.692	12.8 %	12.7 %
XGB	0.675	0.671	14.7 %	14.7 %
RF	0.586	0.571	23.6 %	24.8 %
KNN	0.617	0.606	20.5 %	21.3 %
LR	0.578	0.554	24.4 %	26.4 %
CNN	0.586	0.570	23.6 %	24.9 %
LSTM	0.533	0.506	28.8 %	32.2 %

***Source*: compiled by the authors**

Similarly, the following metric values were obtained for the remaining datasets.

For the Autistic Spectrum Disorder Screening Data for Adults dataset, out of 126 designed ensemble classifiers, the best results were shown by ensemble consisting of MLP, SVM, XGB with soft voting with Gompertz fuzzy ranking. The following metrics were obtained: accuracy = 0.875 (+1.5 % relative to the best classifier in the compound); F1 Score = 0.873 (+1.3 % relative to the best classifier in the compound).

For the GTZAN dataset, out of 657 designed ensemble classifiers, the best results were shown by ensemble consisting of MLP, XGB, CNN, Random Forest, with soft voting using fuzzy ranks (Gompertz) aggregation. The following metrics were obtained: accuracy = 0.824 (+3.2 % relative to the best classifier in the compound); F1 Score = 0.822(+3.2 % relative to the best classifier in the compound).

For the Deepfake FOR dataset, out of 657 designed ensemble classifiers, the best results were shown by Ensemble consisting of KNN, RF, CNN with soft voting. The following metrics were obtained: accuracy = 0.935 (+3.9 % relative to the best classifier in the compound); F1 Score = 0.935 (+3.9 % relative to the best classifier in the compound).

In general, the obtained experimental results demonstrate a consistent and interpretable pattern across all considered datasets, confirming both the effectiveness and the robustness of the proposed methodology.

First, the comparison of individual classifiers with and without the application of the BFE procedure shows that feature space optimization plays a crucial role in improving classification quality. The removal of redundant and less informative features leads to an increase in Accuracy and F1-score.

Second, the transition from individual classifiers to ensemble models consistently results in further performance improvements. This confirms the hypothesis that combining heterogeneous models allows capturing complementary patterns in the data, which cannot be fully exploited by any single classifier. The stacking architecture, in particular, demonstrates its advantage by combining predictions of base models.

An important observation is that the relative gain from ensemble methods varies depending on the type of data. For structured tabular datasets, the improvement is moderate, which can be explained by the relatively lower complexity of feature interactions. In contrast, for unstructured audio data, the use of ensembles leads to significantly higher gains, indicating that complex signal characteristics benefit more from the three-level model integration.

Additionally, the experiments confirm that the proposed unified feature space model ensures compatibility between different types of classifiers, including both classical machine learning algorithms and deep learning models. This makes it possible to construct flexible ensemble configurations tailored to specific datasets without the need for separate preprocessing pipelines.

CONCLUSIONS

The proposed model and methods for constructing a unified feature space for ensemble classification of polystructural heterogeneous small data have demonstrated high effectiveness across multiple domains, including healthcare, finance, audio signal analysis, and information security.

The research objective has been achieved: a unified methodological framework has been developed that integrates polystructural data into a common feature space using a combination of linear model and a three-level model (tabular, spectral, and temporal levels). This approach enables the adaptation of feature representations to the specific requirements of various machine learning and deep learning algorithms.

A key component of the proposed methodology is the use of the Backward Feature Elimination (BFE) procedure, which allows constructing optimal feature subspaces for each individual classifier. Experimental results confirmed that BFE improves classification performance by reducing noise, eliminating redundant features, and increasing the informativeness of input data.

The ensemble classification framework based on the stacking architecture, combined with different aggregation strategies, further enhances model robustness and generalization ability.

Experimental validation on five datasets (ASD, HELOC, RAVDESS, GTZAN, FOR) confirmed the universality of the proposed approach. For structured tabular data, the application of BFE and ensemble methods resulted in moderate improvements (typically 1-2 % in Accuracy and F1-score), while for unstructured audio data, significantly higher gains were observed (up to 10-12 %), highlighting the effectiveness of the proposed feature space model in capturing complex signal characteristics.

Thus, the proposed approach provides a flexible and scalable solution for classification of polystructural heterogeneous small data, enabling the use of a single methodology for preprocessing, feature selection, model training, and evaluation across different data types and application domains.

Future research directions include extending the proposed framework to other types of unstructured data and ensemble techniques.

REFERENCES

1. McDermott, R. “On the scientific study of small samples: Challenges confronting quantitative and qualitative methodologies”. *The Leadership Quarterly*. 2023; 34 (3): 101675. DOI: https://doi.org/10.1016/j.leaqua.2023.101675.

2. Peng, J., Khuat, T., Musial, K. & Gabrys, B. “Machine learning methods for small data and upstream bioprocessing applications: A comprehensive review”. *arXiv*. 2025. DOI: https://doi.org/10.48550/arXiv.2506.12322.

3. Cawley, G. & Talbot, N. “On over-fitting in model selection and subsequent selection bias in performance evaluation”. *Journal of Machine Learning Research*. 2010; 11: 2079–2107.

4. Arsirii, O. O., Andronati, O. K. “Method for building ensemble classifiers of structured and unstructured data based on a unified approach”. *Herald of Advanced Information Technology.* 2025; 8 (3): 288–300, https://www.scopus.com/authid/detail.uri?authorId=58677655800. DOI: https://doi.org/10.15276/hait.08.2025.18.

5. Dou, B., Zhu, Z., Merkurjev, E., et al. “Machine Learning Methods for Small Data Challenges in Molecular Science”. *Chemical Reviews*. 2023; 123 (13): 8736–8780. DOI: https://doi.org/10.1021/acs.chemrev.3c00189.

6. Maldonado, S. & Weber, R. “A wrapper method for feature selection using Support Vector Machines”. *Information Sciences*. 2009; 179 (13): 2208–2217. DOI: https://doi.org/10.1016/j.ins.2009.02.014.

7. Khanjani, Z., Watson, G. & Janeja, V. P. “Audio deepfakes: A survey”. *Frontiers in Big Data*. 2023; 5: 1001063. DOI: https://doi.org/10.3389/fdata.2022.1001063.

8. Gourisaria, M. K., Agrawal, R., Sahni, M., et al. “Comparative analysis of audio classification with MFCC and STFT features using machine learning techniques”. *Discov Internet Things*. 2024; 4: 1. DOI: https://doi.org/10.1007/s43926-023-00049-y.

9. Xie, C., Cao, X. & He, L. “Algorithm of abnormal audio recognition based on improved MFCC”. *Procedia Engineering*. 2012; 29: 731–737. DOI: https://doi.org/10.1016/j.proeng.2012.01.032.

10. Uddin, S., Haque, I., Lu, H., et al. “Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction”. *Sci Rep*. 2022; 12: 6256. DOI: https://doi.org/10.1038/s41598-022-10358-x.

11. Stecking, R. & Schebesch, K. D. “Clustering large credit client data sets for classification with SVM”. 2009. – Available from: https://www.crc.business-school.ed.ac.uk/sites/crc/files/2023-10/Clustering-large-credit-client-data-sets-for-classification-with-SVM.pdf. – [Accessed: Jan 2025].

12. Fatima, S., Hussain, A., Amir, S. B., Ahmed, S. H. & Aslam, S. M. H. “XGBoost and random forest algorithms: An In-Depth analysis”. *Pakistan Journal of Scientific and Industrial Research*. 2023; 3 (1): 26–31. DOI: https://doi.org/ 10.57041/pjosr.v3i1.946.

13. Koshti, R. M., Molia, S. J. & Varia D. J. “Improving credit score classification using predictive analysis and machine learning techniques”. *International Journal on Science and Technology*. 2025; 16 (2): 1–12. DOI: https://doi.org/10.71097/IJSAT.v16.i2.4759.

14. Altunöz, U. “Prediction of banking credit risk using logistic regression and the artificial neural network models: A case study of english banks”. *Journal of Social Research and Behavioral Sciences*. 2024; 10 (21): 862–887. DOI: https://doi.org/10.52096/jsrbs.10.21.32.

15. Zhang Y, Liu J. & Shen W. “A review of ensemble learning algorithms used in remote sensing applications”. *Appl. Sci*. 2022; 12 (17): 8654, https://www.scopus.com/authid/detail.uri?authorId=55787621400. DOI: https://doi.org/10.3390/app12178654.

16. “HELOC”. – Available from: https://www.kaggle.com/datasets/averkiyoliabev/home-equity-line-of-creditheloc?select=heloc_dataset_v1+%281%29.csv – [Accessed: Feb 2025].

17. “Autistic spectrum disorder screening data for adult”. – Available from: https://www.kaggle.com/datasets/faizunnabi/autism-screening. – [Accessed: Feb 2025].

18. “RAVDESS”. – Available from: https://www.kaggle.com/datasets/uwrfkaggler/ravdess-emotional-speech-audio. – [Accessed: Feb 2025].

19. “GTZAN”. – Available from: https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification. – [Accessed: Feb 2025].

20. “The Fake-or-Real (FoR) Dataset (deepfake audio)”. – Available from: https://www.kaggle.com/datasets/mohammedabdeldayem/the-fake-or-real-dataset/data. – [Accessed: Feb 2025].

21. McFee, B., Raffel, C., Liang, D., Ellis D. P.W. , McVicar, M., Battenberg, E. & Nieto, O. “librosa: Audio and music signal analysis in Python”. In *Proceedings of the 14th Python in Science Conference,* 2015. p. 18–25. DOI: https://doi.org/10.25080/Majora-7b98e3ed-003.

22. Thukroo, I. A., Bashir, R. & Giri, K. J. “A comparison of cepstral and spectral featuresusing recurrent neural network for spoken language identification”. *Comput. Artif. Intell*. 2024; 2 (1). DOI: https://doi.org/10.59400/cai.v2i1.440.

23. Sharma, G., Umapathy, K. & Krishnan, S. “Trends in audio signal feature extraction methods”. *Applied Acoustics*. 2020; 158: 107020. DOI: https://doi.org/10.1016/j.apacoust.2019.107020.

24. Abhishek, A., Deb, S. D., Jha, R. K., Sinha, R. & Jha, K. “Ensemble learning using Gompertz function for leukemia classification”. *Biomedical Signal Processing and Control*. 2025; 100 (Part B): 106925, https://www.scopus.com/pages/publications/85204954946. DOI: https://doi.org/10.1016/j.bspc.2024.106925.

25. Baran, K. “Gompertz-fuzzy ensemble of lightweight convolutional neural networks for stress classification”. *Advances in Science and Technology Research Journal*. 2026; 20(4): 457–474, https://www.scopus.com/authid/detail.uri?authorId=57191337665.
DOI: https://doi.org/10.12913/22998624/214572.

Conflicts of Interest: The authors declare that they have no conflict of interest regarding this study, including financial, personal, authorship or other, which could influence the research and its results presented in this article

Received 22.12.2025
Received after revision 27.02.2026
Accepted 12.03.2026

DOI: https://doi.org/10.15276/aait.09.2026.10
УДК 004.1

Розробка уніфікованого простору ознак для ансамблевої класифікації поліструктурних гетерогенних даних малого обсягу

Арсірій Олена Олександрівна[1)]
ORCID: https://orcid.org/0000−0001−8130−9613; e.arsiriy@gmail.com Scopus Author ID: 54419480900
Андронаті Олександр Кирилович[1)]
ORCID: https://orcid.org/0009-0009-1794-5864; alex.andronati@gmail.com. Scopus Author ID: 58677655800
[1)] Національний університет “Одеська Політехніка”, пр. Шевченка, 1. Одеса, 65044, Україна

АНОТАЦІЯ

Актуальність. У статті розглядається проблема класифікації поліструктурних гетерогенних даних невеликого обсягу, що включають в себе структуровану табличну інформацію та неструктуровані аудіосигнали. **Мета**. У дослідженні розроблено уніфіковану методологічну основа для побудови ансамблевої системи класифікації на основі спільного простору ознак, що забезпечує сумісність між різними типами даних та моделями машинного навчання. **Основний матеріал дослідження**. Розроблено модель єдиного простору ознак, що включає лінійне представлення для структурованих даних та трирівневе представлення для аудіоданих, зокрема табличні спектральні ознаки, представлення зображень на основі спектрограм та часові послідовності ознак. Для підвищення якості вхідних даних застосовується процедура зворотного вилучення ознак з метою адаптації підмножин ознак до специфіки окремих класифікаторів та видалення неінформативних ознак. Структура класифікації базується на архітектурі ансамблю стекування, що поєднує декілька базових моделей, включаючи класичні алгоритми машинного навчання та моделі глибокого навчання. Розглядаються три стратегії агрегації: жорстке голосування, м’яке голосування та м’яке голосування з нечітким ранжуванням Гомперца, що дозволяє здійснювати нелінійне коригування ймовірностей класифікаторів та підвищує стійкість в умовах невизначеності. **Результати.** Експериментальна оцінка була проведена на п’яти наборах даних з різних галузей, включаючи медицину, фінанси, аналіз аудіосигналів та виявлення діпфейків. Результати демонструють, що запропонований підхід послідовно покращує ефективність класифікації порівняно з окремими моделями. Застосування відбору ознак та інтеграція методів ансамблю забезпечують значні переваги для поліструктурних даних. **Висновки.** Розроблена модель пропонує гнучке та масштабоване рішення для обробки гетерогенних даних малого обсягу і може ефективно застосовуватися в різних сферах, забезпечуючи покращене узагальнення, стійкість до шуму та адаптивність до різних представлень даних.

Ключові слова: поліструктурні дані; гетерогенні дані; дані малого обсягу; ансамблеві класифікатори; функція Гомперца; класифікація аудіоданих

ABOUT THE AUTHORS

Olena O. Arsirii - Doctor of Engineering Sciences, Professor, Head of Department of Information Systems. Odesa Polytechnic National University. 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/0000-0001-8130-9613; e.arsiriy@gmail.com Scopus Author ID 54419480900
Research field: Information technology; artificial intelligence; decision support systems; machine learning; neural networks

Арсірій Олена Олександрівна - доктор технічних наук, професор, завідувач кафедри Інформаційних систем. Національний університет “Одеська Політехніка”, пр. Шевченка, 1. Одеса, 65044, Україна

Oleksandr K. Andronati - graduate student, Department of Information Systems. Odesa Polytechnic National University. 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/0009-0009-1794-5864; alex.andronati@gmail.com. Scopus Author ID 58677655800
Research field: Data science; machine learning; ensemble classification

Андронаті Олександр Кирилович - аспірант кафедри Інформаційних систем. Національний університет “Одеська Політехніка”, пр. Шевченка, 1. Одеса, 65044, Україна