

How to RETIRE Tabular Data in Favor of Discrete Digital Signal Representation

Paweł Zyblewski^[0000–0002–4224–6709] (✉) and
Szymon Wojciechowski^[0000–0002–8437–5592]

Department of Systems and Computer Networks, Faculty of Information and
Communication Technology, Wrocław University of Science and Technology,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
`{pawel.zyblewski;szymon.wojciechowski}@pwr.edu.pl`

Abstract. The successes achieved by deep neural networks in computer vision tasks have led in recent years to the emergence of a new research area dubbed *Multi-Dimensional Encoding* (MDE). Methods belonging to this family aim to transform tabular data into a homogeneous form of discrete digital signals (images) to apply convolutional networks to initially unsuitable problems. Despite the successive emerging works, the pool of *multi-dimensional encoding* methods is still low, and the scope of research on existing modality encoding techniques is quite limited. To contribute to this area of research, we propose the *Radar-based Encoding from Tabular to Image REpresentation* (RETIRE), which allows tabular data to be represented as radar graphs, capturing the feature characteristics of each problem instance. RETIRE was compared with a pool of *state-of-the-art* MDE algorithms as well as with *XGBoost* in terms of classification accuracy and computational complexity. In addition, an analysis was carried out regarding transferability and explainability to provide more insight into both RETIRE and existing MDE techniques. The results obtained, supported by statistical analysis, confirm the superiority of RETIRE over other established MDE methods.

Keywords: Multi-Dimensional Encoding · Modality Encoding · Tabular data representation · Data visualization · Convolutional neural networks.

1 Introduction

Even though current technological developments are leading to a continuous increase in the amount of data in text or discrete digital signal form, tabular modality is still considered the most popular form of data [23], commonly found in tasks such as medical diagnosis [26], recommendation systems [30], cybersecurity [4] or psychology [27]. This information is essential because, despite the increasing proliferation of deep learning methods that excel at computer vision tasks or audio and text analysis, it is tabular data that provides enough of a challenge for it to be called the “last unconquered castle” for deep neural networks [12]. This is due to the heterogeneous nature of tabular data, which, unlike text,

discrete digital signals or acoustic signals – which are inherently homogeneous in nature – can contain different feature types. In brief, the literature currently identifies four significant challenges in analyzing tabular data: **(i)** Low-Quality Training Data, **(ii)** Missing or Complex Irregular Spatial Dependencies, **(iii)** Dependency on Preprocessing, and **(iv)** Importance of Single Features [3].

Due to these difficulties, using deep learning for tabular data analysis is currently an important and rapidly growing area of research. At present, a taxonomy of approaches to the application of deep learning for tabular data distinguishes three main paths: **(i)** Data Transformation Methods, **(ii)** Specialized Architectures, and **(iii)** Regularization Models [3]. While the vast majority of work in this area focuses on specialized network architectures or dedicated tabular data regularization mechanisms, this article focuses on data transformation methods, specifically *Multi-Dimensional Encoding* (MDE) techniques. While *Single-Dimensional Encoding* methods, such as ordinal or label encoding, are dedicated to transforming categorical variables into real numbers, MDE aims to turn entire feature vectors into image representation in order to facilitate employing convolutional neural networks for initially unsuitable problems. The upside of such approaches is the ability to leverage existing architectures without modification, as well as the potential benefit of *transfer learning* [32].

As already mentioned, *multi-dimensional encoding* methods, despite offering a highly intriguing way of dealing with tabular data and presenting potentially valuable results, are a definite minority in the pool of scientific articles addressing the application of deep learning in this area. As a result, the number of available MDE methods is relatively low, and their potential is still under-researched. In an effort to take the next step towards filling this gap in the literature and expanding the field of *multi-dimensional encoding*, this paper introduces the *Radar-based Encoding from Tabular to Image REpresentation* (RETIRE) method. RETIRE allows individual instances of tabular datasets to be depicted in the form of radar diagrams, capturing the characteristics of their feature values. The proposed method has been compared based on a robust experimental protocol with MDE algorithms known from the literature in order to determine which image representation offers the highest generalization ability while employing the *ResNet18* [10] as the architecture of choice. In addition, the *XGBoost* [5] algorithm was used as a strong baseline for tabular data classification tasks. In order to gain better insight into the interaction of selected MDE methods with convolutional networks, an attempt was made to analyze them in terms of transferability and explainability.

In brief, the main contributions of this article are as follows:

- The introduction of a novel *Radar-based Encoding from Tabular to Image REpresentation* (RETIRE) *multi-dimensional encoding* method.
- Extensive experimental evaluation of RETIRE on 22 benchmark datasets, including selected modality encoding methods and the *XGBoost* algorithm.
- Analysis of selected *state-of-the-art multi-dimensional encoding* methods and RETIRE in terms of transferability, explainability and computational complexity.

2 Related works

This chapter briefly introduces existing approaches to applying deep neural networks to the task of tabular data classification, with an emphasis on multi-dimensional encoding methods. In addition, to highlight the research area’s potential and the scientific community’s interest, the *Sentence Space* approach, which is similar to MDE but applied to textual modality, is described.

2.1 Deep Neural Networks for Tabular data

As already mentioned, the taxonomy for employing deep learning for tabular data distinguishes three groups: **(i)** Data Transformation Methods, **(ii)** Specialized Architectures, and **(iii)** Regularization Models [3]. While the *multi-dimensional encoding* methods are the main focus of this article, which is described in a separate section, the other approaches are briefly introduced below.

Specialized architectures, the largest group of approaches, focus on developing deep neural network structures dedicated to tabular data and considering its heterogeneous nature. The literature here distinguishes two subgroups: **(i)** Hybrid models and **(ii)** Transformer-based models. Solutions from the hybrid models group combine canonical machine learning algorithms with neural networks and include both fully and partly differentiable models. The Network-On-Network (NON) classification model by Luo et al. [18] consists of a fieldwise network containing unique deep networks for each problem feature, a cross-field network choosing optimal operations for a given dataset, and an operation fusion network, allowing for nonlinearities by connecting selected operations. In contrast, Ke et al. proposed the *DeepGBM* model, which combines neural networks with the preprocessing capability of gradient-boosted decision trees. *DeepGBM* consists of two networks, one dealing with dense numerical features and the other with sparse categorical features[13]. A subgroup of transformer-based approaches is inspired by transformers’ successes in text and image data tasks [14]. One of the most popular examples of a model in this category is *TabNet*, introduced by Arik and Pfister [2], consisting of multiple sequential hierarchical subnetworks, where each corresponds to a single decision step.

Regularization models are based on the assumption that in the case of heterogeneous tabular data, the flexibility of neural networks can be an obstacle, and strong regularization of parameters is needed. Kadra et al. [12] have shown that a multilayer perceptron using a set of 13 different regularization methods can outperform *state-of-the-art* tabular data classification methods at the cost of increased time.

Single-Dimensional Encoding, which belongs to the Data Transformation methods group, focuses on dealing with categorical features by encoding them into a form suitable for deep neural networks [9]. *Ordinal encoding*, *label encoding*, and *one-hot encoding* are among the most popular approaches. We can also distinguish *leave-one-out encoding*, which replaces each category with the target variable’s average value, and *hash-based encoding*, which transforms each category using a hash function to a fixed-size value.

2.2 Multi-dimensional Encoding

Numerous scientific articles attest to the effectiveness of convolutional networks in both unimodal and multimodal scenarios for natural language analysis [8] and image, video, and audio classification (e.g., in spectrogram form) [21]. Additionally, deep networks facilitate *transfer learning*, which enables models to apply previously learned information to the current challenge [32]. As a result of these achievements, a new field of study known as *Multi-Dimensional Encoding* [3] was created that focuses on converting tabular data into a more uniform form of digital discrete signals.

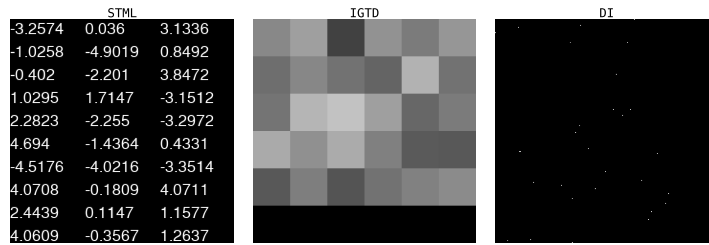


Fig. 1: An example of encoding a single sample of a synthetically generated tabular problem with 30 features, into a two-dimensional discrete digital signal using STML, IGTD and DI techniques.

This idea is best shown by *Super Tabular Machine Learning* (STML) approach, proposed by Sun et al. [25], which projects the feature values of a given problem instance onto the image. The *Image Generator for Tabular Data* (IGTD) [31] created by Zhu et al. offers an alternative approach. By minimizing the discrepancy between the ranking of distances among features and the ranking of distances between their corresponding pixels in the image, the algorithm looks for an optimal mapping. Damri et al. proposed *Feature Clustering-Visualization* (FC-Viz) [6] approach, where each instance of tabular data is converted into a 2D pixel-based representation, where pixels representing strongly correlated and interacting features are located in close proximity of each other. *SuperTML-Clustering*, as proposed by Zhang and Ding, embeds the indices of clustered continuous feature values onto an image [29]. *DeepInsight* (DI) by Sharma et al. [22] allows for the collective utilization of neighboring elements by arranging different elements or features farther away and similar elements or features closer together. The exemplary results of employing selected multi-dimensional encoding methods to tabular data is shown in Fig. 1. It is also worth noting that recently, there have been first works successfully employing MDE in data stream classification tasks while demonstrating relatively low computation time [33].

Although *multi-dimensional encoding* methods were developed to convert heterogeneous tabular data into homogeneous discrete digital signals, similar approaches have also been successfully applied to inherently homogeneous text

modality with remarkable results. Such techniques are usually based on *Sentence Space*, which is a reference to the method presented by Kim [15], in which text data is converted into an image with embeddings of individual words in each row. While these methods are not the subject of this article, and we will not describe them in depth, they undoubtedly provide evidence of the scientific community's interest in MDE methods and their derivatives.

3 Radar-based Encoding from Tabular to Image Representation (RETIRE)

The RETIRE algorithm is inspired by radar charts – a form of graphical representation of an instance's attributes. It can be often seen in cases where one need to compare available options with each other based on their properties. A radar chart is constructed by drawing a polygon over coordinates in a polar coordinate system. Such a figure has N vertices (same as object attributes), which coordinates (r - radius, ϕ - angle) can be determined as:

$$r_n = f_n, \quad \phi_n = (n - 1) \times \frac{2\pi}{N} \quad (1)$$

Where f_n is the value of an n -th attribute. Obviously, for the chart to be readable, values have to be relatable to each other. To achieve that, the scaling function S is often used so that $\{S(f_n) \mid l \leq S(f_n) \leq u\}$ where l and u is a set parameter (usually $l = 0, u = 1$).

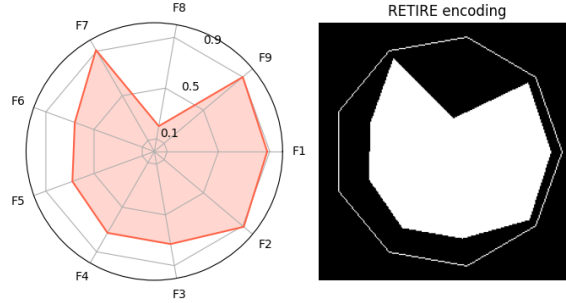
The radar charts often include the axis grid and labels. The figure can also introduce colors if used to compare multiple samples. The goal of the encoding algorithm, however, is to create a representation in the shortest possible time, so the RETIRE algorithm extracts only the binarized shape of the sample and a border indicating the upper limit of the value. An example of the original graph and the corresponding representation is presented in Fig. 2a.

Of course, there can be many modifications to the proposed method, which are the parameters of the algorithm. The very important one is assumed scaling. In this work, S will be function applied to each features as:

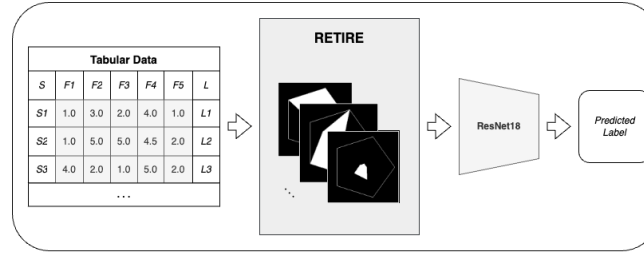
$$S(x_f) = \frac{(x_f - \min(X_f))}{(\max(X_f) - \min(X_f))} \times (u - l) + l \quad (2)$$

Where X is the full set of observations and x is a single sample. As can be seen, the scaling requires obtaining minimum and maximum values of X . Those have to be determined during the learning process – which indicates that RETIRE requires training before being applied to the dataset.

In addition, u and l are also important scaling parameters. Since the first part of the equation will be in the range of $< 0, 1 >$, setting $l > 0$ and $u < 1$ allows scaling to provide “space” for observations. It is important because when RETIRE is used for test data, values from the test set might be outside the observed min-max range. Finally, an important element is the retained figure



(a) Example of radar chart (left) and RETIRE encoding (right).



(b) Pipeline for classification system.

Fig. 2: Graphical abstract for the proposed method.

outline – it provides a reference point for the maximum observed value. The full pipeline of the proposed method is shown in Fig. 2b.

The image representations obtained utilizing encoding are then used by CNN – in this case, *ResNet18*. It is worth mentioning that although the image created by RETIRE is binarized, CNN uses all color channels as the input. Therefore, the resulting binary image is multiplied to obtain 3 identical color channels.

4 Experimental Evaluation

This section describes in detail the experimental evaluation process conducted to analyze the properties of the proposed RETIRE method. The experiments were designed to answer the following research questions:

- **RQ1** Does the discrete digital signal representation obtained using RETIRE benefits from ImageNet knowledge transfer in terms on *TransRate* metric?
- **RQ2** In the case of the *ResNet18* architecture, does the use of an image representation obtained using RETIRE offers a statistically significantly better balanced accuracy score compared to reference MDE methods?
- **RQ3** Can the RETIRE image representation be interpreted by a human?
- **RQ4** What is the computational complexity of the RETIRE encoding and how it compares with reference methods?

4.1 Set-up

Data The experiments were conducted using 22 benchmark datasets originating from the KEEL repository [7]. The datasets can be considered balanced - most have an *Imbalance Ratio* (IR) of less than 1.5, and only in one case the IR exceeds 2. The precise dataset characteristics are shown in Tab. 1.

Table 1: Datasets characteristics.

Dataset	#Instances	#Features	Dataset	#Instances	#Features
australian	690	14	mammographic	830	5
banknote	1372	4	monk-2	432	6
breastcan	683	9	monkone	556	6
breastcancoimbra	116	9	phoneme	5404	5
bupa	345	6	pima	768	8
cryotherapy	90	6	ring	7400	20
german	1000	24	sonar	208	60
haberman	306	3	spambase	4601	57
heart	270	13	titanic	2201	3
ionosphere	351	34	twonorm	7400	20
liver	345	6	wisconsin	699	9

Experimental protocol In order to guarantee a robust experimental protocol, the experiments were carried out using 5-times repeated 2-fold stratified cross-validation. The results were supported by the *Combined 5x2 CV F-test* [1] and the *Wilcoxon signed-rank test* [24] with $\alpha = 0.05$ (the higher the rank, the better). The *Combined 5x2 CV F-test* was used to examine the statistical correlations between methods within particular datasets, whilst the *Wilcoxon signed-rank test* allowed for a global comparison of the investigated approaches. Although the datasets used have a relatively low IR and can be considered balanced, the classification performance was evaluated based on the *balanced accuracy score* (BAC).

Reference methods In the course of the experiments, RETIRE was compared with three state-of-the-art MDE approaches described in Section 2: **(i)** *SuperTML*, **(ii)** *IGTD*, and **(iii)** *DeepInsight*. They were chosen because of the promising results obtained in previous studies and the access to a Python implementation offered by their authors (IGTD and DI). Despite the lack of an official implementation, we implemented STML due to its simplicity. IGTD and DI were used with default parameters. The size of the image representation for STML, DI, and RETIRE multi-dimensional encoding techniques was set to 224x224 pixels. The size of images created using IGTD, due to the characteristics of the method, depended on the number of features in each dataset. For each method the resulting representation is multiplied in order to achieve 3 identical color channels.

Regardless of the MDE approach used, the *ResNet18* architecture was chosen as a convolutional network due to its vast popularity and relatively small size. Training lasted for 20 epochs and the batch size was set to 8. An SGD optimizer with a learning rate of 0.001 and momentum of 0.9 was used. Cross-entropy loss was chosen as the loss function.

In addition, the *XGBoost* [5] algorithm with 100 estimators was used as the baseline. The other parameters remained default. This comparison was added to improve the interpretability of the results obtained through experiments, but it should be noted that *XGBoost* is a very strong benchmark and winning against it is not the primary goal of this work.

Reproducibility All experiments presented in this paper were carried out in *Python* using the *scikit-learn* [20], *PyTorch* [19], and *XGBoost* [5] libraries. The results obtained can be fully replicated using the code located on the *GitHub* repository¹. The computing platform used in all experiments was Mac Studio with Apple M1 Ultra with 20-core CPU, 64-core GPU, 32-core Neural Engine system, 128 GB RAM.

4.2 Experimental scenarios

Experiment 1 – Comparative experimental evaluation The purpose of Experiment 1 is to investigate which MDE method achieves statistically significantly the best balanced accuracy score for the 22 selected datasets and how it compares to the results obtained by the *XGBoost* algorithm. This is the main experiment of this paper, and it will allow us to investigate the quality of the image representations offered by each MDE method, thus answering **RQ2**.

Additionally, in order to verify the usefulness of *transfer learning* in conjunction with the studied representations obtained using MDE, all four encoding methods are evaluated in terms of *TransRate* metrics [11] and BAC. The *ResNet18* architecture offers the possibility of using pre-trained weights, but many scientific articles utilize this option without considering the possibility of negative transfer [28]. Since we are concerned with establishing the network’s inherent properties, *ResNet18* is not trained or finetuned in any way during this experiment. We only evaluate the initial compatibility of knowledge transfer with individual image representations. The results obtained will make it possible to answer **RQ1**.

Experiment 2 – Explainability As part of Experiment 2, the performance of the MDE algorithms offering representations potentially understandable to humans, namely STML and RETIRE, will be subjected to explainability methods derived from the SHAP (*SHapley Additive exPlanations*) library [17, 16]. The results of this analysis will be contrasted with the explained output of the *XGBoost* algorithm in order to investigate whether humans can interpret the images resulting from the MDE and whether the information relevant to the *ResNet18* convolutional network overlaps with that which most influences the decision of the algorithm based solely on tabular data. Based on the observations made, it will be possible to find an answer to **RQ3**.

Experiment 3 – Computational complexity analysis The purpose of the experiment is to study and analyze the computational complexity of selected methods. The main factor subject to this study is to demonstrate the relation of inference time to the dimensionality of the problem. For this experiment,

¹ <https://github.com/w4k2/mde-retire>

the synthetically generated datasets of a variable number of features will be used. Other parameters will be set as follows: MADELON generator, 100 samples, informative-only features, and random generation seed for each instance. Other parameters remain the default for each instance. The test will include measurement of encoding time and *ResNet18* model inference. To stabilize the results, the measurements will be repeated 100 times. The results obtained in this externality will make it possible to find the answer to **RQ4**.

4.3 Experiment 1 – Comparative experimental evaluation

Before the actual comparison experiment, a brief study was conducted to determine whether MDE representations could benefit from knowledge transfer derived from the *ImageNet* dataset. The results obtained in terms of BAC and *TransRate*, which measures the transferability as the mutual information between features of target examples extracted by a pre-trained model and their labels [11], are shown in Fig. 3.

It should be noted that although *TransRate* has a relatively linear relationship to generalization ability, in this case, we are dealing with a random classifier, and BAC values oscillating around 50% regardless of the use of *transfer learning* are expected. This is because, despite the use of *transfer learning*, the images in *ImageNet* deviate too much from the representations offered by MDE.

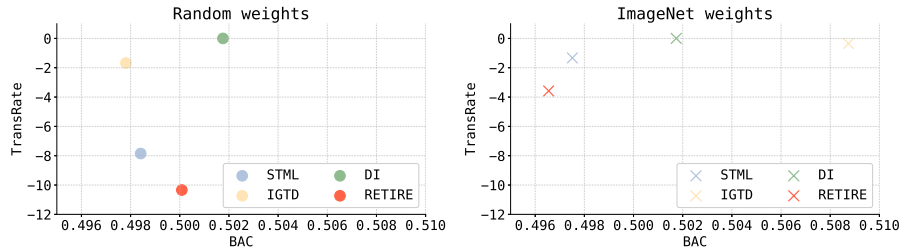


Fig. 3: Results of transferability estimation using the *TransRate* metric. Both BAC and *TransRate* values were averaged for all 22 datasets.

At the same time, interesting conclusions can be drawn from the obtained *TransRate* values, which clearly divide the four analyzed methods into two groups. IGTD and DI theoretically show better compatibility with *ResNet18*, but at the same time, the use of pretreated weights has virtually no effect on them. This is probably due to the abstract nature of the representations they offer, which cannot be compared to the images found in *ImageNet*.

On the other hand, STML and RETIRE, using a less abstract two-dimensional word embedding or geometric shape as the basis for encoding, despite theoretically lower compatibility with *ResNet18*, show significant improvement after using pretreated weights. RETIRE's lower *TransRate* than STML may be explained

by its use of encoding based on a single geometric shape, which inherently has a lower level of complexity than directly transcribing feature values onto the image. Without fine-tuning, such representation may carry slightly less information – especially for a large number of features.

Based on these results, we can answer **RQ1** and conclude that the RETIRE representation allows us to benefit from positive knowledge transfer from the *ImageNet* dataset. At the same time, due to the lack of negative transfer in the case of all tested MDE methods, the pre-trained *ResNet18* is used in further research.

The results of the experiment comparing RETIRE with the reference MDE methods and *XGBoost* can be seen in Fig. 4 and in Tab. 2. The first thing that catches the eye is the relatively low results achieved by IGTD. This is because the images produced by this method are relatively small, and their size depends directly on the number of features. Consequently, the *ResNet18*, adapted initially for 224x224 pixels images, is not the optimal solution. However, since the study aimed to evaluate the role of individual image representations in learning one selected architecture, we decided to utilize this method in its original form without any modifications.

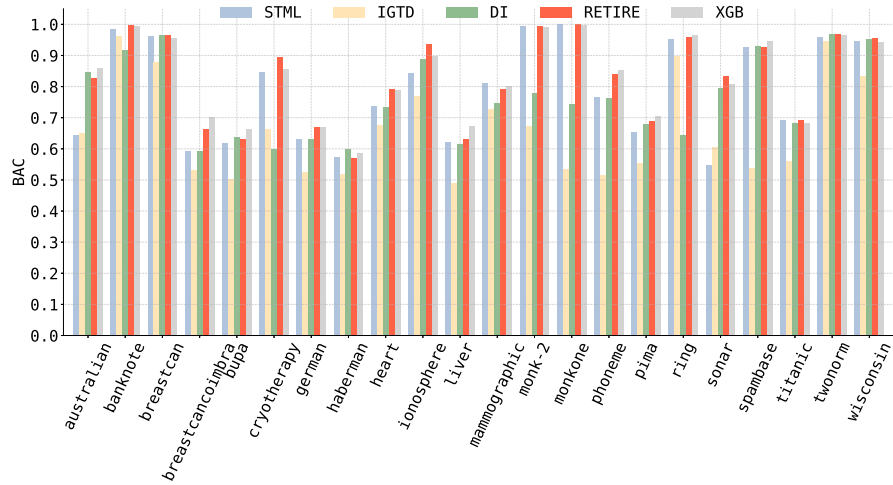


Fig. 4: The results of the experimental evaluation conducted in terms of BAC for each of the 22 datasets.

Among the other MDE methods, RETIRE shows the highest generalization ability by far. For 17 of the 22 datasets, it achieved the highest BAC value. For 6 datasets, it proved statistically significantly better than STML, and for 7 datasets - than DI. At the same time, none of the reference MDE methods ever obtained a statistically significantly better balanced accuracy score than RETIRE.

Table 2: Results of statistical analysis. The first row for each dataset shows the averaged BAC. The indices (by column) of the algorithms from which the given model is statistically significantly better based on the 5×2 *CV F-test* ($\alpha = 0.05$) are given below. The last two rows present the Wilcoxon signed-rank test. The highest BAC or rank values are **bold** or **highlighted in red**. **Bold** indicates the highest value among MDE methods, and **red** indicates the highest value overall (including *XGBoost*).

	STML ¹	IGTD ²	DI ³	RETIRE ⁴	XGB ⁵
australian	0.646	0.650	0.846	0.829	0.861
	—	—	1, 2	1, 2	1, 2, 4
banknote	0.986	0.962	0.918	0.999	0.994
	2	—	—	1, 2, 5	2
breastcan	0.962	0.880	0.965	0.966	0.957
	—	—	—	—	—
breastcancoimbra	0.594	0.531	0.593	0.664	0.703
	—	—	—	2	1, 2, 3
bupa	0.619	0.502	0.638	0.631	0.664
	2	—	2	2	1, 2
cryotherapy	0.846	0.664	0.599	0.895	0.857
	2, 3	—	—	2, 3	2, 3
german	0.631	0.525	0.630	0.671	0.671
	2, 3	—	2	2	2, 3
haberman	0.574	0.518	0.599	0.570	0.586
	—	—	—	—	2
heart	0.737	0.676	0.735	0.793	0.790
	—	—	—	1, 2	1, 2
ionosphere	0.843	0.770	0.889	0.937	0.897
	2	—	1, 2	1, 2, 3	1, 2
liver	0.621	0.490	0.614	0.631	0.672
	2	—	2	2	2
mammographic	0.811	0.727	0.747	0.793	0.802
	—	—	—	—	—
monk-2	0.993	0.672	0.778	0.995	0.990
	2, 3	—	—	2, 3	2, 3
monkone	1.000	0.536	0.743	1.000	0.999
	2, 3	—	—	2, 3	2, 3
phoneme	0.766	0.515	0.764	0.840	0.854
	2	—	2	1, 2, 3	1, 2, 3
pima	0.655	0.555	0.679	0.690	0.706
	—	—	2	2, 3	1, 2
ring	0.952	0.899	0.643	0.961	0.964
	2, 3	3	—	3	2, 3
sonar	0.546	0.605	0.796	0.832	0.807
	—	—	1, 2	1, 2	1, 2
spambase	0.927	0.538	0.931	0.927	0.945
	2	—	2	2	2, 3, 4
titanic	0.694	0.560	0.682	0.692	0.684
	2, 5	—	2	—	—
twonorm	0.958	0.947	0.968	0.968	0.966
	2	—	2	2	2
wisconsin	0.947	0.834	0.953	0.956	0.945
	2	—	2	2	2
Mean rank	2.932	1.227	2.682	4.114	4.045
	2	—	2	1, 2, 3	1, 2, 3

It is somewhat of a surprise that RETIRE for 11 datasets obtained a higher average BAC than *XGBoost*. The difference was statistically significantly better in only one case, but these minor differences translated into RETIRE obtaining a higher average rank value than *XGBoost*. At the same time, RETIRE is the only one of the MDE methods studied that, according to the *Wilcoxon signed-rank* test, globally scored statistically comparable to *XGBoost*. Also, like *XGBoost*, RETIRE is globally statistically significantly better than the other reference MDE methods, thus responding to **RQ2**.

4.4 Experiment 2 – Explainability

Fig. 5 presents the results of the explainability analysis for the RETIRE, STML, and *XGBoost* algorithms in the form of a SHAP image plot and Waterfall plot. Since the purpose of the analysis was to verify that the features relevant to *ResNet18* in the image representations overlapped with those that determined the *XGBoost* algorithm’s decision, the 4 samples analyzed were taken from the *monkone* dataset, for which all three approaches achieved near perfect BAC.

In the case of STML, we can see that regardless of the class, the first row of the image containing *Feature 0* and *Feature 1* values strongly influences the decision. In the case of *Sample 3* and *Sample 4*, *Feature 4*, located in the lower-left corner of the representation, also plays a significant role in making the correct decision. We can see similar correlations in the case of RETIRE, where the first two features (starting from the right and going clockwise) seem to be the most significant. In addition, in the case of *Class 0*, the area belonging to *Feature 4* also plays an important role in the decision-making process. The most significant observation here is that the components of the STML and RETIRE images that *ResNet18* found to be most relevant overlap significantly with the features that have the greatest impact on the decision of the *XGBoost* algorithm, based solely on tabular data. This confirms that despite minor differences due to the characteristics of the individual approaches and differing representations, convolutional networks employing MDE methods (including the proposed RETIRE) make their decisions based on the values of the problem’s features in a manner similar to algorithms operating on tabular data. This makes the proposed representation interpretable to humans, thus answering **RQ3**.

4.5 Experiment 3 – Computational complexity analysis

The computational complexity experiments were measured separately for data encoding time and model inference time. The results are shown in Figure 6. As can be seen, the lowest time values are achieved by the RETIRE and DI methods. Both graphs show a linear relationship between embedding preparation time and the number of features, although in the case of DI – in the first phases, it is flat. This can be justified by the fact that DI performs optimization (*Assymetric Greedy Search*), which can reach convergence in the first phases, but not for problems with higher dimensionality. As for RETIRE, this linear time characteristic was an assumed theoretical, computational complexity that responds to **RQ4**. It can be explained that drawing the figure is related to an additional condition check with each new feature. Lastly, the STML also has linear computational complexity, although the time values achieved by the algorithm increase much faster.

An important observation is also the inference time, which, for most cases, is constant and very low when compared to the encoding time. This is an expected behavior since the structure of the CNN remains constant. The only differences are due to the processing time of smaller resolution images, which is relatable for

IGTD, although these are not noticeable differences. In addition, IGTD is characterized by exponential computational complexity, which, with a large number of features, might not be feasible for computations.

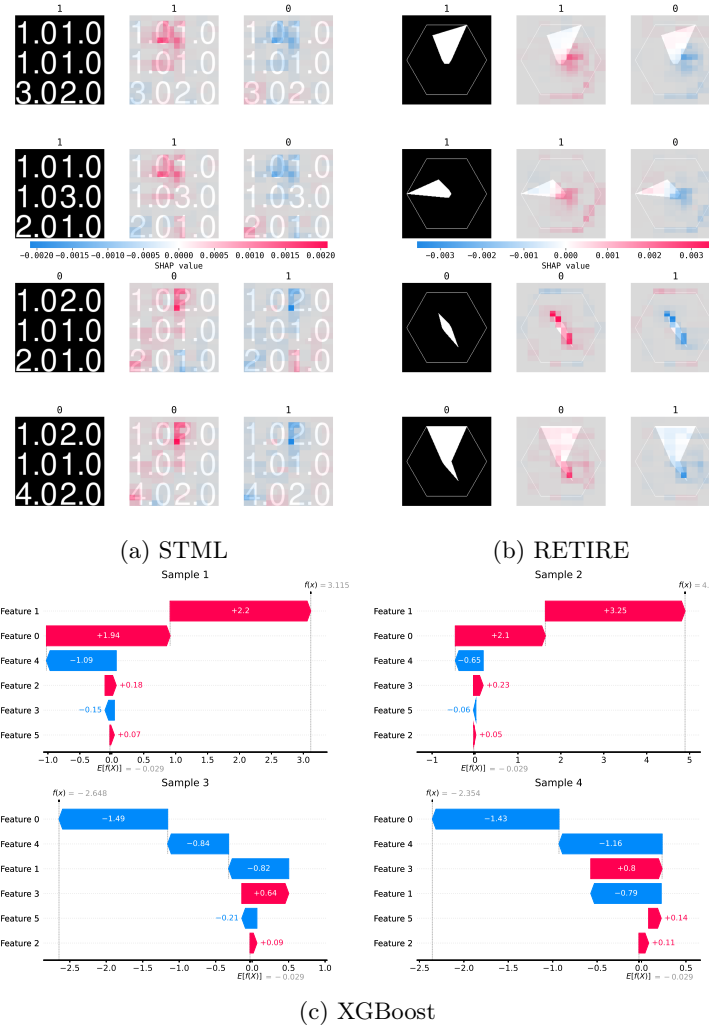


Fig. 5: Explainability of selected data transformation methods applied to the first data fold of the Monkone dataset. The sample numbers in subfigure (c) correspond to the subsequent rows in subfigures (a) and (b). In subfigures (a) and (b), the consecutive columns from the left represent: (i) the original image and its class, (ii) the classifier's decision and the explanation behind it, and (iii) the decision with less support value and its justification.

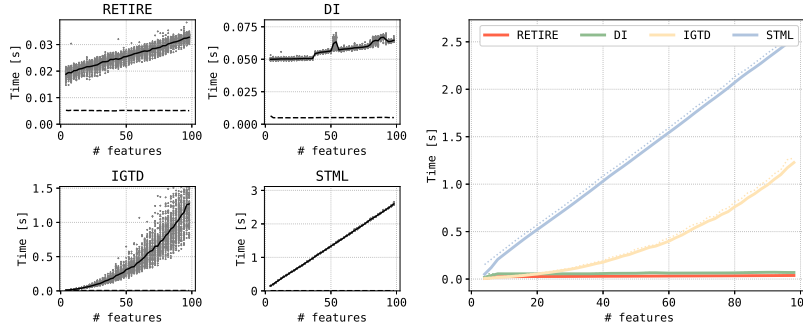


Fig. 6: Time complexity analysis. Results for individual methods on the left and the comparison (with applied Simple Moving Average) on the right.

5 Conclusions

The study aimed to expand the still under-exploited but extremely promising area of data transformation methods for tabular data analysis. This was done by proposing a novel *multi-dimensional encoding* method, namely *Radar-based Encoding from Tabular to Image REpresentation* (RETIRE), which allows obtaining an image representation depicting the characteristics of the features of each problem instance as a single geometric shape.

Extensive comparison experiments have been carried out, showing that the representation obtained using RETIRE in conjunction with the *ResNet18* architecture enables a balanced accuracy score statistically significantly better than *state-of-the-art* MDE methods with publicly available implementations. Most importantly, RETIRE is the only *multi-dimensional encoding* method tested that is not statistically significantly inferior to the *XGBoost* algorithm and even surpasses it in terms of average rank.

Additional strengths of the proposed method are its inherent explainability and relatively low linear computation time when compared to other MDE approaches. Analysis using SHAP confirmed that the features considered most important by *ResNet18* in the RETIRE representation overlap with those that have the greatest impact on the classification results of the *XGBoost* algorithm.

Future work may include a study of the transferability and explainability of MDE methods to provide more insight into the specifics of their work, as well as the use of color in methods so far focused on binary or grayscale images.

Acknowledgments. This work was supported by the statutory funds of the Department of Systems and Computer Networks, Faculty of Information and Communication Technology, Wrocław University of Science and Technology.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alpaydm, E.: Combined 5×2 cv f test for comparing supervised classification learning algorithms. *Neural computation* **11**(8), 1885–1892 (1999)
2. Arik, S.Ö., Pfister, T.: Tabnet: Attentive interpretable tabular learning. In: *Proceedings of the AAAI conference on artificial intelligence*. vol. 35, pp. 6679–6687 (2021)
3. Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., Kasneci, G.: Deep neural networks and tabular data: A survey. *IEEE Transactions on Neural Networks and Learning Systems* (2022)
4. Buczak, A.L., Guven, E.: A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications surveys & tutorials* **18**(2), 1153–1176 (2015)
5. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. pp. 785–794 (2016)
6. Damri, A., Last, M., Cohen, N.: Towards efficient image-based representation of tabular data. *Neural Computing and Applications* pp. 1–21 (2023)
7. Derrac, J., Garcia, S., Sanchez, L., Herrera, F.: Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *J. Mult. Valued Logic Soft Comput* **17**, 255–287 (2015)
8. Gimenez, M., Palanca, J., Botti, V.: Semantic-based padding in convolutional neural networks for improving the performance in natural language processing. a case of study in sentiment analysis. *Neurocomputing* **378**, 315–323 (2020)
9. Hancock, J.T., Khoshgoftaar, T.M.: Survey on categorical data for neural networks. *Journal of big data* **7**(1), 28 (2020)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778 (2016)
11. Huang, L.K., Huang, J., Rong, Y., Yang, Q., Wei, Y.: Frustratingly easy transferability estimation. In: *International conference on machine learning*. pp. 9201–9225. PMLR (2022)
12. Kadra, A., Lindauer, M., Hutter, F., Grabocka, J.: Well-tuned simple nets excel on tabular datasets. *Advances in neural information processing systems* **34**, 23928–23941 (2021)
13. Ke, G., Xu, Z., Zhang, J., Bian, J., Liu, T.Y.: Deepgbm: A deep learning framework distilled by gbdt for online prediction tasks. In: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. pp. 384–394 (2019)
14. Khan, S., Naseer, M., Hayat, M., Zamir, S.W., Khan, F.S., Shah, M.: Transformers in vision: A survey. *ACM computing surveys (CSUR)* **54**(10s), 1–41 (2022)
15. Kim, Y.: Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882* (2014)
16. Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I.: From local explanations to global understanding with explainable ai for trees. *Nature Machine Intelligence* **2**(1), 2522–5839 (2020)
17. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) *Advances in Neural Information Processing Systems* 30, pp. 4765–4774. Curran Associates, Inc. (2017)

18. Luo, Y., Zhou, H., Tu, W.W., Chen, Y., Dai, W., Yang, Q.: Network on network for tabular data classification in real-world applications. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 2317–2326 (2020)
19. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch (2017)
20. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011)
21. Satt, A., Rozenberg, S., Hoory, R., et al.: Efficient emotion recognition from speech using deep learning on spectrograms. In: Interspeech. pp. 1089–1093 (2017)
22. Sharma, A., Vans, E., Shigemizu, D., Borojevich, K.A., Tsunoda, T.: Deepinsight: A methodology to transform a non-image data to an image for convolution neural network architecture. *Scientific reports* **9**(1), 11399 (2019)
23. Shwartz-Ziv, R., Armon, A.: Tabular data: Deep learning is not all you need. *Information Fusion* **81**, 84–90 (2022)
24. Stapor, K., Ksieniewicz, P., García, S., Woźniak, M.: How to design the fair experimental classifier evaluation. *Applied Soft Computing* **104**, 107219 (2021)
25. Sun, B., Yang, L., Zhang, W., Lin, M., Dong, P., Young, C., Dong, J.: Supertml: Two-dimensional word embedding for the precognition on structured tabular data. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 0–0 (2019)
26. Ulmer, D., Meijerink, L., Cinà, G.: Trust issues: Uncertainty estimation does not enable reliable ood detection on medical tabular data. In: Machine Learning for Health. pp. 341–354. PMLR (2020)
27. Urban, C.J., Gates, K.M.: Deep learning: A primer for psychologists. *Psychological Methods* **26**(6), 743 (2021)
28. Wang, Z., Dai, Z., Póczos, B., Carbonell, J.: Characterizing and avoiding negative transfer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11293–11302 (2019)
29. Zhang, J., Ding, G.: Supertml-clustering: Two-dimensional word embedding for structured tabular data. In: International Conference on Image, Vision and Intelligent Systems. pp. 600–609. Springer (2023)
30. Zhang, Q., Cao, L., Shi, C., Niu, Z.: Neural time-aware sequential recommendation by jointly modeling preference dynamics and explicit feature couplings. *IEEE Transactions on Neural Networks and Learning Systems* **33**(10), 5125–5137 (2021)
31. Zhu, Y., Brettin, T., Xia, F., Partin, A., Shukla, M., Yoo, H., Evrard, Y.A., Doroshov, J.H., Stevens, R.L.: Converting tabular data into images for deep learning with convolutional neural networks. *Scientific reports* **11**(1), 11325 (2021)
32. Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., He, Q.: A comprehensive survey on transfer learning. *Proceedings of the IEEE* **109**(1), 43–76 (2020)
33. Zyblewski, P.: Employing two-dimensional word embedding for difficult tabular data stream classification. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 73–89. Springer (2024)