Journal of Biomedical Informatics 165 (2025) 104821

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Original Research

Transfer learning for a tabular-to-image approach: A case study for cardiovascular disease prediction

Francisco J. Lara-Abelenda [a,*], David Chushig-Muzo [a], Pablo Peiro-Corbacho [a], Vanesa Gómez-Martínez [a], Ana M. Wägner [b], Conceição Granja [c], Cristina Soguero-Ruiz [a]

[a] Department of Signal Theory and Communications, Telematics and Computing Systems, Rey Juan Carlos University, Madrid, Spain
[b] Instituto Universitario de Investigaciones Biomédicas y Sanitarias, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain
[c] Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

ARTICLE INFO

Keywords:
Tabular-to-image methods
Transfer learning
Low-dimensional data
Mixed-type data
Cardiovascular disease prediction

ABSTRACT

Objective: Machine learning (ML) models have been extensively used for tabular data classification but recent works have been developed to transform tabular data into images, aiming to leverage the predictive performance of convolutional neural networks (CNNs). However, most of these approaches fail to convert data with a low number of samples and mixed-type features. This study aims: to evaluate the performance of the tabular-to-image method named low mixed-image generator for tabular data (LM-IGTD); and to assess the effectiveness of transfer learning and fine-tuning for improving predictions on tabular data.

Methods: We employed two public tabular datasets with patients diagnosed with cardiovascular diseases (CVDs): Framingham and Steno. First, both datasets were transformed into images using LM-IGTD. Then, Framingham, which contains a larger set of samples than Steno, is used to train CNN-based models. Finally, we performed transfer learning and fine-tuning using the pre-trained CNN on the Steno dataset to predict CVD risk.

Results: The CNN-based model with transfer learning achieved the highest AUCORC in Steno (0.855), outperforming ML models such as decision trees, K-nearest neighbors, least absolute shrinkage and selection operator (LASSO) support vector machine and TabPFN. This approach improved accuracy by 2% over the best-performing traditional model, TabPFN.

Conclusion: To the best of our knowledge, this is the first study that evaluates the effectiveness of applying transfer learning and fine-tuning to tabular data using tabular-to-image approaches. Through the use of CNNs' predictive capabilities, our work also advances the diagnosis of CVD by providing a framework for early clinical intervention and decision-making support.

1. Introduction

Cardiovascular diseases (CVDs) are noncommunicable diseases (NCDs) that remain the leading cause of morbidity and mortality [1], imposing significant health and economic burden globally [2]. CVDs are among the most prevalent NCDs, affecting more than 500 million people worldwide [3]. Clinical studies indicate that CVD is highly prevalent in patients with type 1 diabetes (T1D), increasing the mortality rates [4–6]. Although T1D is frequently diagnosed in children and youth, many cases have been reported in adults [7]. Owing to multiple hospitalizations, adverse events, and frequent visits to primary and specialized care, patients diagnosed with NCDs significantly increase the cost and demand for healthcare services. Early identification of CVD cases, in cohorts diagnosed with other NCDs such as T1D, along with timely and effective interventions are crucial for reducing both health and economic burdens.

To address this global health challenge, several risk calculators have been developed to identify individuals at high risk of developing CVD, encouraging timely interventions and reducing the likelihood of acute events. Among the most popular calculators, we find the Framingham risk score [8], and the PROCAM calculator [9]. For individuals diagnosed with diabetes, more specialized CVD risk calculators have been created and internationally validated, including the Swedish T1D risk score [10], the Scottish T1D risk score [11] and the Danish Steno T1 Risk Engine (ST1RE) [12]. Despite their extensive usage, these CVD risk

* Corresponding author.
E-mail addresses: francisco.lara@urjc.es (F.J. Lara-Abelenda), david.chushig@urjc.es (D. Chushig-Muzo), pablo.peiro@urjc.es (P. Peiro-Corbacho), vanesa.gomez@urjc.es (V. Gómez-Martínez), ana.wagner@ulpgc.es (A.M. Wägner), conceicao.granja@ehealthresearch.no (C. Granja).

https://doi.org/10.1016/j.jbi.2025.104821
Received 28 November 2024; Received in revised form 20 February 2025; Accepted 27 March 2025
Available online 8 April 2025

calculators have been developed based on a set of rules defined by clinical experts and derived from specific cohorts, raising questions related to their scalability and applicability to different clinical scenarios.

Recent advancements in Machine Learning (ML) offer a promising opportunity to develop more generalizable and scalable models for CVD risk prediction [13–16]. Over the last decade, Deep Learning, a sub-field of ML, has made important breakthroughs across various domains, including speech recognition, natural language processing, and computer vision [17,18]. In particular, the Convolutional Neural Network (CNN)-based models have demonstrated exceptional results in multiple computer vision tasks [19–21]. Regarding CVD prediction, CNN-based models have achieved excellent results when applied to high-quality cardiac images obtained from MRI and CT scans [22]. However, the use of these imaging modalities is not preferred for predictions due to their prolonged acquisition times, limited availability, and associated radiation exposure. Additionally, the direct application of CNN-based models to tabular data, consisting of samples (rows) and features (columns), is still challenging due to the lack of inherent locality and spatial relationships that are crucial to effectively train CNNs [23,24]. CNNs' performance on tabular data is also limited by numerous challenges such as missing data, dependency on preprocessing, missing spatial dependencies, and the absence of prior knowledge about the data structure [18].

To mitigate these limitations, several approaches have been developed to transform tabular data into 2D images, allowing CNN-based models to be subsequently applied to them [25–28]. Among the most popular, we find DAFT [25], which integrates tabular data with 3D images within CNNs, DeepInsight [26], which converts non-image data into images for CNN architecture, REFINED [27], which represents features as images with neighborhood dependencies to be compatible with CNNs. Additionally, IGTD [28], addresses the conversion of tabular data into images but shows limitations with low-dimensional datasets. To overcome these challenges, DWTM [29] proposes a dynamic weighted tabular method for CNNs, while HACNet [30] proposes an interpretable end-to-end table-to-image converter for CNNs.

Although these approaches [25–30] are promising to address the challenges of applying CNN-based models on tabular data, they have been evaluated in scenarios with high-dimensional data. To our knowledge, no research has studied the transformation of tabular data with mixed-type features or datasets with few features (low-dimensionality) [31]. IGTD [28] is one of the most adaptive methods to transform tabular data into images, however, it does not perform well with low-dimensional and mixed-type datasets. To address these limitations, the authors developed the LM-IGTD [31], a modified version of IGTD that uses a novel technique for creating stochastic noisy features. Transforming tabular data into images facilitates the application of transfer learning, which is especially beneficial when data is limited or when training models from scratch is impractical. It leverages pre-trained models to enhance performance with smaller datasets, although most studies that have explored the use of transfer learning in tabular data have focused on embedding approaches [32]

In this study, the aim is two-fold. First, we aim to transform tabular data from patients diagnosed with CVDs into 2D images using the LM-IGTD method [31], addressing the inherent challenges associated with low-dimensional and mixed-type data (categorical and continuous features). Second, we seek to evaluate the effectiveness of transfer learning and fine-tuning for enhancing CVD prediction, using 2D images generated by LM-IGTD. To achieve this, we used two publicly available datasets, which are extensively used in the CVD risk assessment: the Framingham dataset [33] (hereafter referred to as Fram-data) and the Steno dataset [12] (shortened Steno-data). Fram-data was used as the large dataset to obtain the initial weights for the convolutional layers, which were then fine-tuned using the Steno-data. We compared the results of LM-IGTD and CNNs against traditional ML-based models extensively used for tabular data, including the Least Absolute Shrinkage and Selection Operator (LASSO), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Decision Tree (DT) [34,35]. Additionally, we evaluated TabPFN, a recently developed model that leverages foundational models for classification in small tabular datasets [36]. Lastly, we evaluated the importance of each feature for CVD prediction by using Gradient-weighted Class Activation Mapping (Grad-CAM) on the LM-IGTD generated images. To the best of our knowledge, this is one of the first studies that evaluate the applicability of transfer learning and fine-tuning on tabular data associated with CVDs using a tabular-to-image transformation approach over datasets with a low number of features, while also incorporating interpretability.

Table 1
Statement of significance.

Problem
Medical datasets often have a limited number of samples and features, which can lead to a reduced performance of machine learning (ML) models.
What is Already Known
ML models have been extensively used to predict medical hazards. Recent advancements focus on transforming tabular data into images to leverage convolutional neural networks' predictive capabilities. However, most methods struggle to handle datasets with few features and mixed data types.
What This Paper Adds
We present a transfer learning approach with fine-tuning using the tabular-to-image method LM-IGTD. This methodology enhances ML efficacy on small datasets with limited features, demonstrating its applicability in cardiovascular disease prediction.

This paper is organized as follows. A description of the public datasets used for CVD prediction and the methods employed, including the LM-IGTD method, are presented in Section 2. In Section 3, we detail the experimental setup and the results of the comparative analysis, where we assess the performance of CNN-based models, using LM-IGTD, against traditional ML-based models for CVD prediction. In Section 4, we discuss the implications of our findings, highlighting the advantages and limitations of using tabular-to-image transformation combined with transfer learning and fine-tuning techniques. Finally, Section 5 concludes the paper by summarizing our contributions and suggesting potential directions for future research, particularly in the application of tabular-to-image methods to other medical domains. A summary of the significance of this research and value added to the existing literature is presented in Table 1.

2. Materials and methods

This section provides a description of the public datasets used in this study, including the preprocessing steps applied to prepare the data for analysis. Following this, we present the notation used throughout the study and explain the tabular-to-image transformation method LM-IGTD. Finally, we introduce the foundational concepts of the CNNs, as well as the transfer learning techniques used. The source code implementing the methodology to ensure the reproducibility of the results presented in this paper is available at the following link: github.com/ai4healthurjc/tab2img.

2.1. Dataset description and preprocessing

For the development of this study, we used data from CVD patients from two public datasets: Fram-data and Steno-data. The Fram-data, collected from the *Framingham Heart Study* [1], contains a total of 4240 participants and 15 features relevant to CVD prediction. These features are: age (in years), sex (0: male, 1: female), level of education (1 to 4), smoking status (0: no, 1: yes), number of cigarettes smoked per day, total cholesterol level, systolic blood pressure (in mmHg), diastolic blood pressure (in mmHg), body mass index (in Kg/m2), heart rate, glucose level (mg/dL), use of blood pressure medications (0: no, 1: yes), presence of hypertension (0: no, 1: yes), history of stroke (0: no, 1: yes)

Table 2
A summary of statistics for continuous and categorical features associated with the Fram-data and Steno-data.

Feature	Fram-data			Steno-data		
	mean±std	minimum	maximum	mean±std	minimum	maximum
Age	49.58 ± 8.57	32.00	70	50.07 ± 13.67	10.00	95.00
Systolic BP	132.35 ± 22.03	83.50	295	124.9 ± 18.13	68.67	186.58
Glucose	81.96 ± 22.83	40.00	394	185.66 ± 41.10	53.87	321.64
	Categories	Count	Percentage	Categories	Count	Percentage
Sex	Male	1819	42.92%	Male	302	44.00%
	Female	2419	57.08%	Female	375	56.00%
Smoker	Smoker	2094	49.40%	Smoker	467	69.00%
	Non-smoker	2144	50.60%	Non-smoker	210	31.00%

or presence of diabetes (0: no, 1: yes). The dataset also includes a binary classification label, with '1' and '0' indicating whether the patient has a 10-year risk of developing CVD. Of the 4240 participants in the dataset, 3594 were labeled as '0' ('not-CVD'), while 644 were labeled as '1' ('CVD').

The Steno-data, collected from the *Steno Diabetes Center Copenhagen*, is composed of data from 1000 Danish adults diagnosed with T1D, including 10 clinical and lifestyle features [12]. These features include: age (in years), sex (male or female), diabetes duration (in years), systolic blood pressure (in mmHg), low-density lipoprotein (LDL) (in mmol/l), glycosylated hemoglobin (HbA1c) (in mmol/mol), albuminuria [categorized as normoalbuminuria (<30 mg/g), microalbuminuria (between 30 and 299 mg/g), and macroalbuminuria (≥ 300 mg/g)], estimated glomerular filtration rate (eGFR) (in ml/min/1.72 m^2), smoking status (0: non-smoker, 1:smoker), and regular physical activity (0:no, 1:yes). The majority of the features in Steno-data are continuous, with an exception for the binary features sex, smoking, physical activity; and albuminuria which presents three categories. All features are free of outliers or missing data. For both the Fram-data and Steno-data datasets, we applied and compared two normalization techniques: (1) the z-normalization that subtracts the mean and divides by the standard deviation; (2) the min–max normalization that rescales features to the $[0, 1]$ range. It is worth noting that patients with a history of CVD events were excluded from our analysis, resulting in a final dataset of 677 individuals.

In this study, Steno-data was used to evaluate the efficacy of transfer-learning and fine-tuning techniques in tabular-to-image transformation methods to predict the CVD risk. The selection of Steno-data is due to the low number of samples compared to Fram-data which can reduce the performance of training a neural network from scratch. It is worth noting that the ST1RE model provides a continuous CVD risk score in the [0...1] range, whereas Fram-data categorizes the CVD risk as binary. Consequently, a process of binarization of the label was required to align the risk labels for consistency in analysis. This binarization was conducted in two steps. First, we categorized the 10-year CVD risk scores provided by ST1RE into three levels: low, intermediate, and high. These levels were identified using the risk stratification guidelines from the National Institute for Health and Care Excellence [37], with the following cut-off thresholds: CVD risk <0.1 (low-risk patients); ≤ CVD risk < 0.2 (moderate-risk patients); and CVD risk≥0.2 (high-risk patients). Second, the patients categorized as low-risk (443 participants) were assigned a binary label of '0' (indicating 'not-CVD'), whereas both moderate-risk and high-risk patients (234 participants) were assigned a binary of '1' (indicating 'CVD').

To perform transfer learning from Steno-data using the weights of the convolutional layers derived from a model trained with Fram-data, in our approach, it was required that the two datasets shared common features. Hence, we identified the common features between both datasets, which included sex, age, systolic blood pressure, smoking status and glucose levels. Note that glucose was recorded differently in both datasets: the Fram-data presents glucose as blood glucose levels (mg/dL), while the Steno-data provides it as HbA1c levels. To standardize this feature across datasets, we converted the HbA1c from the Steno-data into equivalent glucose levels (mg/dL) through the formula [38]:

$$AGL = 28.7 * (HbA1c/10.929 + 2.15) - 46.7$$

Table 2 summarizes the statistics for Fram-data and Steno-data, including the mean, standard deviation (std), minimum and maximum, for continuous features, whereas for categorical features, the number of samples and percentage of each category are shown.

2.2. Notation

Let a dataset $\mathcal{D} = \{\mathbf{x}^{(i)}, \mathbf{y}^{(i)}\}_{i=1}^{n}$ consisting of n samples, where $\mathbf{x}^{(i)} \in \mathbb{R}^{D_f}$ is a vector representation composed of D_f features, and $\mathbf{y}^{(i)}$ is the corresponding label vector (with class 0 denoting 'not-CVD', and class 1 denoting 'CVD'). The LM-IGTD method is implemented in two stages: (1) increasing the dimensionality of the samples using a stochastic noise generation, transitioning to an augmented feature vector, *i.e.,* from $\{\mathbf{x}^{(i)}\}_{i=1}^{n}$ to $\{\mathbf{z}^{(i)}\}_{i=1}^{n}$, where $\mathbf{z}^{(i)} \in \mathbb{R}^{D'_f}$, and D' represents the total number of features after noise generation, with $D'_f > D_f$. The new feature dimension D'_f can be determined by two methods: Homogeneous Noise Generation (HoNG), where $D'_f = k \times D_f + D_f$ with k representing the number of noisy features generated per original feature; and Heterogeneous Noise Generation (HeNG), where $D'_f = f(k \times D_f) + D_f$, and f is a feature selection function applied to the generated noisy features; (2) transforming the augmented vectors $\{\mathbf{z}^{(i)}\}_{i=1}^{n}$ into 2D grayscale $\mathbf{Q}^{(i)} \in \mathbb{R}^{w \times h}$, where w and h represent the width and height of the image, respectively, and are derived from the augmented feature dataset as $w \times h = D'_f$.

The dataset $\mathcal{D}$ was divided into a training subset $\mathcal{D}_{train} = \{\mathbf{x}^{(j)}, \mathbf{y}^{(j)}\}_{j=1}^{n_{train}}$ and test subset $\mathcal{D}_{test} = \{\mathbf{x}^{(k)}, \mathbf{y}^{(k)}\}_{k=1}^{n_{test}}$, comprising 80% and 20% of the samples, respectively. The $\mathcal{D}_{train}$ was used for training the ML-based models, while $\mathcal{D}_{test}$ was reserved for assessing model performance. Lastly, prior to training and given the class imbalance problem present in the datasets, random under-sampling was used to ensure the same number of samples in both classes ('CVD' and 'not-CVD') only in $\mathcal{D}_{\text{train}}$. Regarding the Steno-data, $\mathcal{D}_{\text{train}}$ was composed of 187 samples labeled as 0 and 187 samples labeled as 1, while $\mathcal{D}_{\text{test}}$ was composed of 89 samples labeled as 0 and 47 samples labeled as 1. On the other hand, for the Fram-data, $\mathcal{D}_{\text{train}}$ was composed of 515 samples labeled as 0 and 515 samples labeled as 1, while $\mathcal{D}_{\text{test}}$ was composed of 719 samples labeled as 0 and 129 samples labeled as 1. After the pre-processing, different ML-based models were trained using the balanced $\mathcal{D}_{train}$ composed either of vectors $\{\mathbf{x}^{(i)}\}_{i=1}^{n_{train}}$ or 2D images $\{\mathbf{Q}^{(i)}\}_{i=1}^{n_{train}}$ to predict the estimated binary label $\hat{\mathbf{y}}^{(i)}$ (identifying CVD risk). After training, the performance of the models was evaluated using $\mathcal{D}_{test}$. Finally, to assess the generalization capability of the predictive models, we repeat the pre-processing and training pipeline four times using different seeds for train/test splitting and the random under-sampling technique. This process generates four distinct performance values for each ML-based model because the training and testing sets are different in each one of the iterations.

To quantitatively evaluate the predictive performance of the models, we used the figures of merit area under the receiver operating

characteristic curve (AUCROC), sensitivity, and specificity [35], which were calculated based on the model's ability to correctly predict the positive and negative classes, using the following categories: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

$$\text{AUCROC} = \int_0^1 \text{TPR}(\text{FPR})\, d(\text{FPR}) \tag{1}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3}$$

2.3. Tabular-to-image approaches for low-dimensional and mixed-type data

In the literature, several methods have been proposed for transforming tabular data into images [25–27,30], being IGTD one of the most adaptive for handling low-dimensional and mixed-type data. IGTD transforms tabular data into grayscale 2D images by assigning a pixel to each input feature, where the pixel intensity indicates the value of the feature [28]. IGTD determines the spatial arrangement of pixels by calculating the relationships among features and assigning pixels with similar relationships close to each other within the image. To achieve this, it computes two ranking matrices: (1) the *feature ranking matrix* that measures the similarity among features, and (2) the *pixel ranking matrix* that computes the similarity among samples. The algorithm optimizes the placement of features in the image by minimizing the differences between these two matrices. The IGTD's algorithm is a greedy iterative process that swaps pixel assignments to reduce the distance between similar features. At each iteration, the algorithm determines the feature that has not been selected for swapping and identifies another feature to replace it, aiming to minimize the dissimilarity between the *feature ranking matrix* and *the pixel ranking matrix*. The aim is to ensure that similar features are placed close to each other and dissimilar features are placed further apart in the resulting image.

LM-IGTD, an extension of IGTD [28] and proposed by the authors in [31], is a data-driven approach that allows us to transform tabular data into images when considering low-dimensional and mixed-type datasets. To address the limitations of existing tabular-to-image methods, LM-IGTD uses two main approaches: (1) the addition of noisy features using stochastic noise generation; and (2) the use of Gower distance [39] and a dynamic selection of correlations for continuous and categorical features. The inclusion of noisy features to input data seeks to extend the dimensionality of tabular data to create images with a reasonable number of pixels to leverage the potential of CNN-based models. The main hypothesis of the author's study [31] was that noisy features can help to create *new synthetic pixels* in the generated 2D image, thus helping capture intrinsic relationships among original features and improving the spatial relationship of the image. Noise generation is characterized by three parameters: the type of noise, the noise power and the number of noisy features to be created.

For continuous features, Gaussian noise was applied, while for categorical features swapping noise was used. To generate noisy features, we followed two approaches: HoNG and HeNG. HoNG generates k noisy features for each original feature by adding Gaussian noise to numerical features and swapping noise to categorical features. Note that various values of k were explored, and through experimentation, $k = 3$ was selected as the optimal value. HeNG follows a similar methodology but additionally selects a subset of D'_f generated in HoNG to improve the predictive performance [31].

After augmenting the dimensionality with noisy features, both the *feature ranking matrix* and the *pixel ranking matrix* were computed. To obtain the *feature ranking matrix*, the point-biserial [40], the phik [41] and the Pearson correlation [42] were considered, whereas the Gower distance was used for the *pixel ranking matrix*. Note that the arrangement of the pixels in the image is determined through an optimization process that leverages both the correlation and distance matrices to minimize dissimilarities, ensuring that similar features are placed close together in the resulting image. Notably, the same number of noisy features must be created in both the Fram-data and Steno-data datasets to perform transfer learning. To this end, Fram-data was used as the reference dataset, and the same noise power and number of features applied to the Fram-data were also used for Steno-data.

2.4. Convolutional neural networks and transfer learning

CNNs have gained great popularity in different domains due to their ability to effectively extract spatial information through convolution operations, leading to excellent results in image-related applications [20]. While several CNN-based models have been proposed in the literature, the fundamental structure of CNNs is typically composed of three types of layers: *convolutional*, *pooling*, and *fully-connected layers*. Despite their extensive usage, training CNNs from scratch presents significant challenges, particularly the need for large datasets to obtain high predictive results. To address this challenge, transfer learning has emerged as an effective approach. In this method, pre-trained CNN-based models are used, and their weights are fine-tuned to adapt to a new task, thus enhancing predictive results [43]. The transfer learning technique seeks to train CNN-based models using large datasets and subsequently leverage the pre-trained models by using learned weights, mainly those associated with the convolutional and pooling layers, to adapt to a new task.

Transfer learning begins with two datasets: a large dataset and a smaller dataset. A model that consists of multiple layers is initially trained using a large dataset. After this training, the layers of the model are 'frozen', meaning their parameters are fixed and will not be updated in subsequent training. New layers are then added to the model, resulting in a new architecture that combines the original layers with the newly added ones. The new model, with both original and added layers, is then trained on the smaller dataset. However, only the newly added layers are adjusted during this training, while the original layers remain unchanged. This method enables the model to retain the general knowledge from the large dataset while fine-tuning it for specific tasks using the smaller dataset.

The CNN architecture used in this project consists of two convolutional layers and two fully connected layers. Each convolutional layer is followed by a batch normalization layer, a rectified linear unit (ReLU) activation function, and a max-pooling layer. The first fully connected layer is followed by a ReLU activation function, a batch normalization layer, and a dropout layer. Finally, the second fully connected layer is followed by a sigmoid activation function to obtain a binary output. For training CNNs, we implemented an adaptive learning rate (which reduces the learning rate on the plateau) and early stopping, and the binary cross entropy was used as the cost function. In the fine-tuning approach, we trained a CNN with two convolutional layers and two fully connected dense layers using the Fram-data images. Then, we froze the two convolutional layers and trained the two dense layers with the Steno-data images.

2.5. Interpretability for models based on convolutional neural networks using class activation maps

Neural-based models have brought a revolution in both academia and industry due to their remarkable predictive performance. However, the adoption of these models in different domains remains limited due to the lack of interpretability [44]. To build models that have the ability to explain why they predict what they predict is crucial. Among the most extended interpretability methods for CNN-based models, we find layer-wise relevance propagation, and class activation maps (CAMs), to be one of the most used techniques owing to its adaptability and post-hoc approach. CAMs indicate which regions of an image are crucial for the network's predictions.

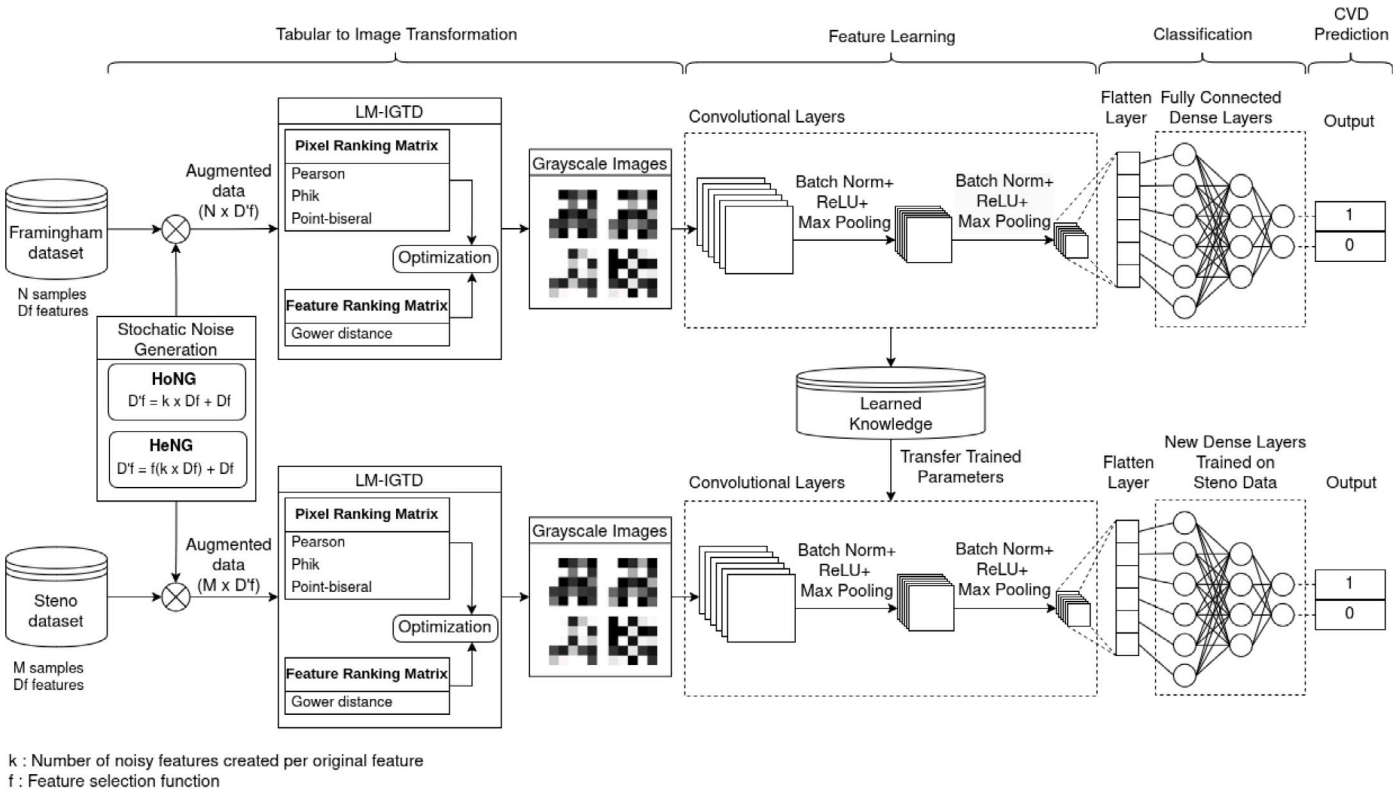

Fig. 1. Workflow of the data-driven approach for transfer learning on tabular data using information of patients with CVDs.

By extending CAMs, Grad-CAM was proposed aiming to be applied to any CNN-based architecture. It calculates a coarse-grained attribution map (with respect to a certain class) on the last convolutional layer and then multiplies it by the attribution map obtained from guided back-propagation [45]. It leverages gradient information and the property of CNNs where the spatial relationship in the data is maintained after it passes through the convolutional layers.

Grad-CAM generates saliency maps by computing the gradients of a target class score y^c with respect to the feature maps of convolutional layers. Given an input image I, a CNN-based model processes it through multiple convolutional layers, producing a set of feature maps A_k.

$$A_k = f_k(I) \quad \text{for } k = 1, 2, \ldots, K$$

where K is the number of feature maps in the selected convolutional layer. To determine the importance of each feature map A_k for a target class c, the gradients of the class score y^c are computed with respect to A_k. These gradients indicate how much the class score changes with respect to the feature map activation, providing insight into which regions contributed the most to the prediction. The gradients are global averages pooled over the spatial dimensions (height i and width j) of the feature maps to obtain the importance weights w_k^c:

$$w_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_k^{ij}}$$

where Z is the total number of pixels in the feature map. These weights quantify the importance of each feature map for predicting the class c. The Grad-CAM heatmap is computed as a weighted sum of the feature maps as follows:

$$L^c_{\text{Grad-CAM}} = \text{ReLU}\left(\sum_k w_k^c A_k \right)$$

$L^c_{\text{Grad-CAM}}$ is called class-discriminative localization map. Note that the ReLU function ensures that only positive contributions are visualized.

3. Experimental results

In this section, we present the experimental setup, some examples of images generated with LM-IGTD, and lastly the comparative analysis of predictive results obtained using: (1) tabular data and the ML-based models DT, KNN, LASSO, SVM and TabPFN; and (2) images generated with LM-IGTD and CNNs. Fig. 1 shows the workflow of the data-driven approach for transforming tabular data into images and the transfer learning technique on data belonging to patients with CVDs.

3.1. Experimental setup

The training of the CNN-based models and hyperparameter optimization were performed using an NVIDIA RTX 4000 ADA, while the tabular-to-image transformation was executed using 50 CPU cores from an AMD EPYC 7713P 64-core processor. CNN-based models were implemented with `Pytorch` (version: 2.1.2) [46]. To determine the best hyperparameter values in CNNs, we used the Python library `raytune` (version: 2.9.1) [47], which is used for distributed hyperparameter tuning. For each model, we conducted 200 different combinations. The following hyperparameter values were considered: the image size {25×25, 35 × 35, 45×45}; the kernel size {3×3, 4 × 4, 5×5}; the number of filters {8, 16, 32, 64} of the convolutional layers; the kernel size [2×2, 3×3] of the max pooling layer; the output size (dense_units) with values {32, 64, 128}, the dropout rate {0.1, 0.2, 0.4}, the optimizers Adam and RMSProp, and the learning rate values between [0.0001, 0.01].

We compared the performance of CNNs with traditional ML-based models trained with tabular data, including LASSO, KNN, SVM, and

DT [34,35]. To find the best hyperparameter values for these models, k-fold cross validation [34] (with k=5) over the balanced D_{train} was employed and using AUCROC as the figure of merit. The following hyperparameters were explored: C in the range $[1e^{-1.5}, 1e^{0.4}]$ for LASSO, K between [1...15] for KNN, and $\gamma \in \{1e^{-2}, 1e^{-3}, 1e^{-4}, 1e^{-5}\}$ and C in the range $[1e^{-1}...1e^{1}]$ for SVM. For DT, the maximum depth within the range [2...8] and the minimum samples per split are selected based on the training subset size. Regarding TabPFN, we evaluated different values for n_estimators, specifically $\{5, 10, 15\}$.

3.2. Tabular-to-image generation using IGTD with noisy features

The generation of noisy features resulted in medium size images that improved the pixel relationships in the images generated with the LM-IGTD algorithm. Fig. 2 presents the pairwise correlation between the features of D'_f (composed by the original features x and noisy features z) for both the Fram-data and Steno-data. The HoNG approach (Fig. 4 (a-b)) produced a matrix that quadrupled the number of features in the original dataset, increasing from five original features ($D_f = 5$) to 20 features ($D'_f = 20$) in the noisy dataset, due to the addition of three noisy features per original feature. The correlation matrix evidences this expansion as repeated patterns, indicating that while the noisy features are different, they retain a similarity to the original features. When applying the HeNG approach (Fig. 4(c–d)), we used only a subset of the noisy features generated with the HoNG to augment the dataset. The resulting noisy dataset comprised a total of 16 features ($D'_f = 16$), consisting of the five original features ($D_f = 5$) and nine noisy features selected.

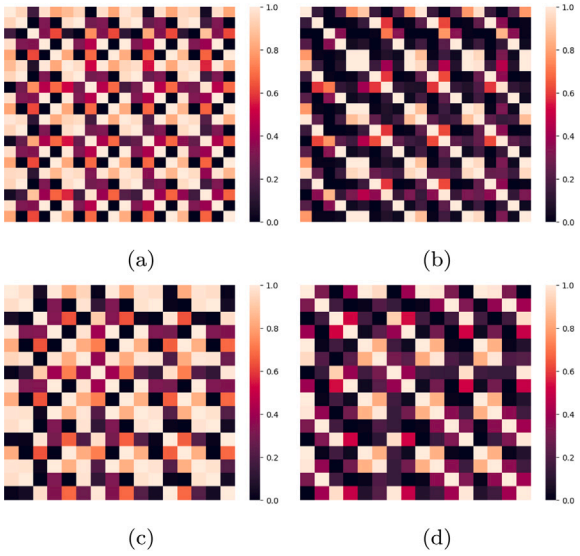

Fig. 2. Pairwise correlation heatmaps illustrating the relationship between original and noisy features. The noise was generated using: (a,b) HoNG; and (c,d) HeNG. These analyses were conducted on two datasets: (a,c) Fram-data; and (b,d) Steno-data. The color intensity reflects the strength of the correlation, with lighter colors indicating stronger correlations and darker colors indicating weaker correlations.

Fig. 3 illustrates the 2D grayscale images generated using the LM-IGTD algorithm applied to samples from Fram-data and Steno-data, using the HoNG and HeNG approaches to generate noisy features. The images generated using the HoNG have a dimension of 5 × 4 pixels as they are generated from a feature set z with $D'_f = 20$, whereas the HeNG images have a dimension of 4 × 4 pixels because they are generated from a feature set z with $D'_f = 16$. The LM-IGTD algorithm arranges the pixels within the generated images Q, resulting in a spatial configuration where similar features are placed close together, and dissimilar features are placed further apart [28].

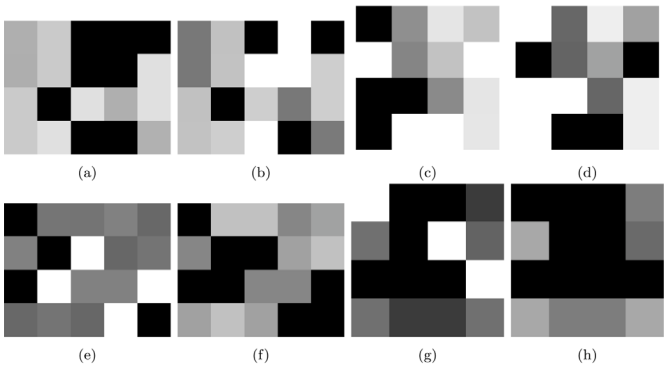

Fig. 3. 2D grayscale images generated with LM-IGTD applied to the: (a, b, c, d) Fram-data, and (e, f, g, h) Steno-data. The noisy features were generated using: (a,b,e,f) HoNG; and (c, d, g, h) HeNG (e–f). The grayscale value of each pixel indicates the value of the corresponding feature.

3.3. Cardiovascular disease classification results with tabular data and images generated with LM-IGTD

Table 3 presents the results of various ML-based models applied to the Fram-data. Among the tabular models, LASSO achieved the highest AUCROC, with a value of 0.644±0.012. When using images generated by LM-IGTD and analyzed with CNNs, the models achieved accuracies of 0.663 ± 0.017 and 0.655 ± 0.023 for HeNG and HoNG, respectively. HeNG demonstrated a performance improvement in AUCROC of 6.1% and 4..8% over DT and KNN, respectively. Meanwhile, the AUCROC value of the CNN which uses HeNG was more comparable to those obtained with LASSO, SVM, and TabPFN, with differences of 1.9%, 2.2%, and 2.9%. Regarding sensitivity and specificity, SVM achieved the highest sensitivity, with a value of 0.664 ± 0.028, while CNN-HeNG achieved the highest specificity, with a value of 0.688 ± 0.052.

Table 3
Classification results (measured by mean±std) using the Fram-data and the models DT, KNN, LASSO, SVM, TabPFN and CNNs. The best results for each metric are in bold. (N/A: not applicable).

Model Model	Type noise	AUCROC	Sensitivity	Specificity
DT	N/A	0.602 ± 0.017	0.612 ± 0.088	0.592 ± 0.098
KNN	N/A	0.625 ± 0.014	0.612 ± 0.0.027	0.637 ± 0.033
LASSO	N/A	0.644 ± 0.012	0.635 ± 0.012	0.653 ± 0.011
SVM	N/A	0.641 ± 0.008	**0.664 ± 0.028**	0.618 ± 0.027
TabPFN	N/A	0.634 ± 0.011	0.633 ± 0.011	0.634 ± 0.023
CNN	HeNG	**0.663 ± 0.017**	0.628 ± 0.056	**0.688 ± 0.052**
CNN	HoNG	0.655 ± 0.023	0.622 ± 0.067	0.669 ± 0.059

Table 4 presents the classification results using the Steno dataset. Among the tabular models, LASSO and TabPFN achieved the highest AUCROC values, with 0.835 ± 0.020 and 0.835 ± 0.015, respectively. By applying the LM-IGTD combined with CNNs methodology, an AUCROC of 0.842 ± 0.026 was achieved with HoNG, and an AUCROC of 0.844 ± 0.025 was achieved with HeNG, resulting in an improvement of 0.9% in AUCROC compared to the best tabular models. In addition, this methodology enhanced sensitivity (0.865 ± 0.055) compared to the baseline models LASSO (0.845 ± 0.046) and TabPFN (0.840 ± 0.048), while maintaining a comparable specificity (0.822) in the three models. Finally, using the proposed fine-tuning approach with HoNG resulted in an AUCROC of 0.853 ± 0.022, while fine-tuning with the HeNG approach achieved an AUCROC of 0.855 ± 0.018. These improvements represent a 2% (HeNG) and 1.8% (HoNG) increase, respectively, over the best tabular model. In addition, the fine-tuning process improved the balance between sensitivity (0.865 ± 0.065) and specificity (0.844 ± 0.037) for the HeNG compared to traditional models.

Table 4
Classification results (measured by mean±std) using the Steno-data and the models DT, KNN, LASSO, SVM and CNNs. The best results for each metric are in bold. (N/A: not applicable; N/U: not used).

Model	Type noise	Fine-tuning	AUCROC	Sensitivity	Specificity
DT	N/A	N/A	0.789 ± 0.018	0.792 ± 0.094	0.787 ± 0.083
KNN	N/A	N/A	0.805 ± 0.024	0.829 ± 0.049	0.782 ± 0.040
LASSO	N/A	N/A	0.835 ± 0.020	0.845 ± 0.046	0.824 ± 0.017
SVM	N/A	N/A	0.830 ± 0.021	0.841 ± 0.045	0.819 ± 0.023
TabPFN	N/A	N/A	0.835 ± 0.015	0.840 ± 0.048	0.829 ± 0.030
CNN	HeNG	N/U	0.844 ± 0.025	0.865 ± 0.055	0.822 ± 0.055
CNN	HoNG	N/U	0.842 ± 0.026	0.877 ± 0.057	0.798 ± 0.042
CNN	HeNG	✓	**0.855 ± 0.018**	0.865 ± 0.065	**0.844 ± 0.037**
CNN	HoNG	✓	0.853 ± 0.022	**0.881 ± 0.038**	0.823 ± 0.036

3.4. Identifying risk factors in cardiovascular diseases using class activation maps

To gain interpretability of the classification results using CNN-based models, we employed the post-hoc interpretability method named Grad-CAM. This method allows us to identify which parts of images are more relevant (represented by orange and red colors in heatmaps) for classification when CNN-based models are considered. Note that the LM-IGTD algorithm transforms individual features into pixels and rearranges them based on their correlations and similarities, and once the image is transformed, we can identify the features associated with each pixel. After model training, we apply Grad-CAM to determine the most saliency areas of the images involved in the classification, and then, which are the most relevant features.

Fig. 4 shows some examples of Grad-CAM heatmaps for the Fram and Steno datasets using the HeNG and HoNG. Figures on the left present the resulting images of LM-IGTD, and to identify the features in the interpretability analysis, we added colored circles (corresponding to the legend). The colors indicate the features used in the classification task, for example in Fig. 4(a): red (sex), blue (age), gray (sbp), black (glucose), and pink (smoker). Figures on the right show the heatmap generated by Grad-CAM, highlighting the importance of each feature in the model's predictions. The red and orange regions indicate areas where features had the highest relevance for classification, whereas blue regions represent areas of lower importance.

In Fig. 4, we observe that the pixels related to age, sex, and glucose were those with the highest impact on the model's decision according to the Grad-CAM images (see positions with warm colors in the Grad-CAM heatmaps). The feature age appears as a highly relevant feature for predictions using HoNG and HeNG on the Fram dataset, and HoNG on the Steno dataset. The sex feature is crucial for HoNG in the Fram dataset and HeNG in the Steno dataset. Lastly, the glucose variable is relevant across all the approaches. Fig. 4 presents only one sample per approach; however, in general, the same three features are the most relevant ones for CVD detection in the majority of the samples.

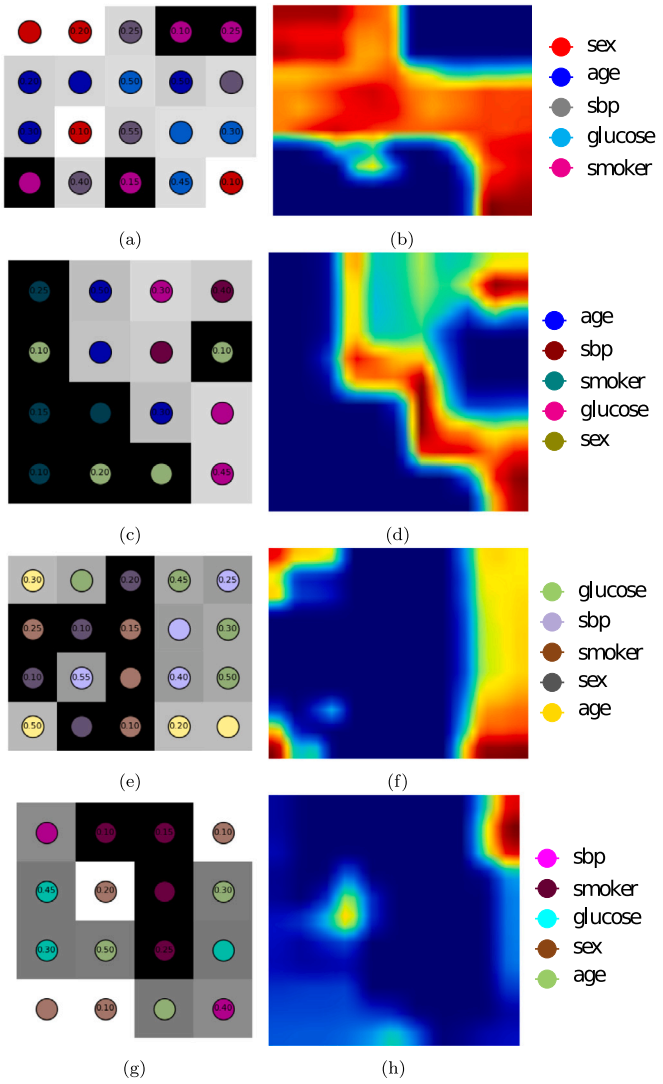

Fig. 4. Analysis of feature importance using Grad-CAM heatmaps. Interpretable images extracted from the CNN for patients who suffered a CVD from the Fram dataset (a, b, c, d) and the Steno dataset (e, f, g, h). The images were extracted from the HoNG noisy dataset (a, b, e, f) and the HeNG noisy dataset (c, d, g, h). Images (a, c, e, g) represent the features related to each pixel, while images (b, d, f, h) show the importance of each pixel to the final classification as determined by Grad-CAM.

4. Discussion

In this study, we evaluated the effectiveness of the tabular-to-image method named LM-IGTD for creating 2D images from tabular data belonging to patients with CVD, and the feasibility of using transfer learning in CNN-based models to improve CVD classification. The datasets Fram-data and Steno-data, which are extensively used in CVD risk identification, were considered. Initially, Fram-data and Steno-data were composed of 15 and 10 features, respectively, but some of these features were discarded to ensure the same set of features in both datasets, which is crucial for effective transfer learning and fine-tuning in our approach. To address the challenges caused by the number of features (low-dimensionality) and mixed-type data (categorical and continuous features) in tabular-to-image methods, LM-IGTD was used.

In LM-IGTD, first, we generated noisy features by aiming to augment the dimension of the datasets and increase the number of pixels in the resulting 2D images. This was achieved following two approaches: HoNG and HeNG. Regarding HoNG, we generated 3 noisy features per original feature, resulting in a noisy dataset with 20 features. Conversely, only 11 noisy features were added using HeNG, leading to a noisy dataset composed of 16 features. Subsequently, using the

augmented Fram-data and Steno-data, we employed LM-IGTD to transform the tabular data into 2D images. The HoNG approach generated images with dimensions of 5 × 4, while the HeNG approach generated images with dimensions of 4 × 4. Then, we trained a CNN with two convolutional layers and two fully connected dense layers using the Fram-data images. Finally, we froze the two convolutional layers and trained the two dense layers with images generated with LM-IGTD and belonging to Steno-data.

Regarding Fram-data, the transformation of the data into images and the use of CNN-based models did not improve the results obtained with tabular ML-based models (DT, KNN, LASSO, SVM, and TabPFN). The best performance of a tabular model was achieved by Lasso reaching an AUCROC of 0.644 ± 0.012, whereas the best LM-IGTD approach obtained an AUCORC of 0.663 ± 0.017 using HeNG. Therefore, the use of the LM-IGTD algorithm improved the AUCROC by 1.9% in the Fram-data

Regarding the Steno-data, the use of LM-IGTD improved CVD classification performance compared to ML-based models trained on tabular data. Using LM-IGTD with HeNG, we achieved an AUCROC of 0.844 ± 0.025, representing a 0.9% improvement over the best tabular model, TabPFN (0.835±0.015). Lastly, the best result was obtained by performing transfer learning and fine-tuning on the noisy dataset generated using HeNG. By training the CNN on the Fram dataset with HeNG and applying transfer learning and fine-tuning on the Steno dataset, an AUCROC of 0.855 ± 0.018 was achieved. This result represents a 1.1% improvement over the classical LM-IGTD approach and a 2% improvement over novel tabular models such as TabPFN.

Reviewing the results, LM-IGTD with HeNG proved to be effective in the generation of medium-sized images, which is crucial for training CNNs and obtaining predictive results compared to tabular ML models such as DT, KNN, LASSO, TabPFN, and SVM. The HeNG approach achieved better results compared to HoNG in the LM-IGTD methodology across both datasets. Thus, correctly selecting the number of noisy features created is crucial for enhancing model performance. There is a limit to the inclusion of noisy features to enlarge the dataset, as it can negatively affect the performance of the classification model. Consequently, adding too many noisy features does not result in an optimal model performance.

For transfer learning in our approach, it is essential that the *large* and *small* datasets present *common* features, such as age in both datasets. Furthermore, to achieve optimal performance, these common features must exhibit similar probability density functions (for continuous features) or probability mass functions (for categorical features). In this article, it is worth noting that despite using the same features in Steno-data and Fram-data, the characteristics of the participants in both datasets are different. Steno-data consists of data from adult patients diagnosed with T1D, which can lead to differences in the glucose level and age mean compared to patients from Fram-data (Table 2). Despite the differences in features between the datasets, the transfer learning approach allowed us to achieve the highest performance in the Steno-data.

Reviewing the interpretability images extracted from the Grad-CAM, the three most relevant variables for CVD prediction are age, sex, and glucose concentration. These results are closely aligned with the literature [16]. According to several studies, age is a crucial factor in the development of CVDs, as the risk of CVD increases exponentially with age [48,49]. Additionally, sex has been identified as one of the main risk factors in the development of CVDs in previous studies [48, 50], with men being at higher risk of suffering from CVD compared to women. Lastly, high blood glucose levels are closely related to several CVDs, such as coronary heart disease and heart failure, among others [51,52].

In future work, we plan to evaluate the performance of transfer learning on other datasets that have similar features in both the *large* and *small* datasets. Moreover, this proof-of-concept needs to be extended to other datasets with a diverse number of features and in different knowledge domains to ensure the effectiveness of this methodology. Additionally, to improve the predictive performance and interpretability, we can explore the integration of attention mechanisms, evaluating transformer-based models and vision transformers [53]. Attention layers have demonstrated the capability to capture feature interactions and can help to gain an understanding of the model's predictions.

5. Conclusion

In this study, we evaluated the use of a tabular-to-image approach LM-IGTD to improve CVD detection in imbalanced datasets with a low number of features, including Fram-data and Steno-data. The experimental results demonstrated the effectiveness of combining LM-IGTD and CNNs to create more generalizable models capable of improving CVD detection. LM-IGTD proved to be beneficial for improving predictive performance, achieving AUCROC values of 0.844 ± 0.020 and 0.842 ± 0.026 with CNNs and HeNG and HoNG for the Steno-data, respectively. Moreover, we proposed a pipeline that uses data from both datasets and involves transfer learning and fine-tuning. The proposed methodology requires training CNNs using the Fram-data after the tabular-to-image transformation with LM-IGTD. Then, we performed transfer-learning and fine-tuning using the pre-trained CNNs and data from Steno-data. As a result, we achieved an AUCROC of 0.855 ± 0.018, outperforming the results of traditional ML models by 2%. Additionally, we apply post-hoc interpretability to the generated images, which highlights the importance of age, sex, and glucose levels in predicting the potential onset of CVD in the future. This study represents one of the first attempts to apply and assess the effectiveness of transfer learning and fine-tuning using tabular-to-image approaches. Our work also contributes to enhancing CVD detection, which is crucial for effective decision-making and early intervention, thus aiming to avoid acute events related to CVDs.

CRediT authorship contribution statement

Francisco J. Lara-Abelenda: Writing – original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **David Chushig-Muzo:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Pablo Peiro-Corbacho:** Writing – review & editing, Software. **Vanesa Gómez-Martínez:** Writing – review & editing, Software. **Ana M. Wägner:** Writing – review & editing, Validation. **Conceição Granja:** Writing – review & editing, Validation, Project administration. **Cristina Soguero-Ruiz:** Writing – review & editing, Supervision, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the European Commission, through the H2020-EU.3.1.4.2, European Project WARIFA (Watching the risk factors: Artificial intelligence and the prevention of chronic conditions) under Grant Agreement 101017385; and by the Spanish Government by the Grant AAVis-BMR PID2019-107768RA-I00/AEI/10.13039/501100011033. The study sponsors have not been involved in any stage of the study.

References

[1] R.B. D'Agostino Sr, M.J. Pencina, J.M. Massaro, S. Coady, Cardiovascular disease risk assessment: insights from framingham, Glob. Hear. 8 (1) (2013) 11–23.

[2] S. Kaptoge, L. Pennells, D. De Bacquer, M.T. Cooney, M. Kavousi, et al., World health organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions, Lancet Glob. Heal. 7 (10) (2019) e1332–e1345.

[3] A. Giovanni, A. Enrico, B. Aime, B. Michael, B. Marianne, C. Jonathan, et al., Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study, J. Am. Coll. Cardiol. 76 (25) (2020) 2982–3021.

[4] B. Vergès, Cardiovascular disease in type 1 diabetes: A review of epidemiological data and underlying mechanisms, Diabetes & Metabolism 46 (6) (2020) 442–449.

[5] P. Balakumar, K. Maung-U, G. Jagadeesh, Prevalence and prevention of cardiovascular disease and diabetes mellitus, Pharmacol. Res. 113 (2016) 600–609.

[6] J.-S. Yun, S.-H. Ko, Current trends in epidemiology of cardiovascular disease and cardiovascular risk management in type 2 diabetes, Metabolism 123 (2021) 154838.

[7] International Diabetes Federation, IDF T1D index report 2022, 2022, (Accessed: 13 June 2024) URL https://diabetesatlas.org/idfawp/resource-files/2022/12/IDF-T1D-Index-Report.pdf.

[8] K.M. Anderson, P.M. Odell, P.W. Wilson, W.B. Kannel, Cardiovascular disease risk profiles, Am. Heart J. 121 (1) (1991) 293–298.

[9] G. Assmann, P. Cullen, H. Schulte, Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular munster (PROCAM) study, Circulation 105 (3) (2002) 310–315.

[10] B. Zethelius, B. Eliasson, K. Eeg-Olofsson, A.M. Svensson, S. Gudbjörnsdottir, J. Cederholm, et al., A new model for 5-year risk of cardiovascular disease in type 2 diabetes, from the Swedish national diabetes register (NDR), Diabetes Res. Clin. Pract. 93 (2) (2011) 276–284.

[11] S.J. McGurnaghan, P.M. McKeigue, S.H. Read, S. Franzen, A.-M. Svensson, M. Colombo, et al., Development and validation of a cardiovascular risk prediction model in type 1 diabetes, Diabetologia 64 (9) (2021) 2001–2011.

[12] D. Vistisen, G.S. Andersen, C.S. Hansen, A. Hulman, J.E. Henriksen, et al., Prediction of first cardiovascular disease event in type 1 diabetes mellitus: the steno type 1 risk engine, Circulation 133 (11) (2016) 1058–1066.

[13] M.M. Ali, B.K. Paul, K. Ahmed, F.M. Bui, J.M. Quinn, M.A. Moni, Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison, Comput. Biol. Med. 136 (2021) 104672.

[14] C. García-Vicente, D. Chushig-Muzo, I. Mora-Jiménez, H. Fabelo, I.T. Gram, M.L. Løchen, et al., Clinical synthetic data generation to predict and identify risk factors for cardiovascular diseases, in: VLDB Workshop on Data Management and Analytics for Medicine and Healthcare, Springer, 2022, pp. 75–91.

[15] A. Tiwari, A. Chugh, A. Sharma, Ensemble framework for cardiovascular disease prediction, Comput. Biol. Med. 146 (2022) 105624.

[16] D. Chushig-Muzo, H. Calero-Díaz, F.J. Lara-Abelenda, V. Gómez-Martínez, C. Granja, C. Soguero-Ruiz, Interpretable data-driven approach based on feature selection methods and GAN-based models for cardiovascular risk prediction in diabetic patients, IEEE Access 12 (2024) 84292–84305.

[17] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, in: J. Burstein, C. Doran, T. Solorio (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1, 2019, pp. 4171–4186.

[18] V. Borisov, T. Leemann, K. Seßler, J. Haug, M. Pawelczyk, G. Kasneci, Deep neural networks and tabular data: A survey, IEEE Trans. Neural Networks Learn. Syst. 35 (2021) 7499–7519.

[19] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM 60 (6) (2017) 84–90.

[20] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, et al., Recent advances in convolutional neural networks, Pattern Recognit. 77 (2018) 354–377.

[21] Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, A survey of convolutional neural networks: analysis, applications, and prospects, IEEE Trans. Neural Networks Learn. Syst. 33 (12) (2021) 6999–7019.

[22] M. Swathy, K. Saruladha, A comparative study of classification and prediction of cardio-vascular diseases (CVD) using machine learning and deep learning techniques, ICT Express 8 (1) (2022) 109–116.

[23] S.Ö. Arik, T. Pfister, Tabnet: Attentive interpretable tabular learning, in: Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, the Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021, 35, (8) 2021, pp. 6679–6687.

[24] R. Shwartz-Ziv, A. Armon, Tabular data: Deep learning is not all you need, Inf. Fusion 81 (2022) 84–90.

[25] T.N. Wolf, S. Pölsterl, C. Wachinger, Alzheimer's Disease Neuroimaging Initiative, et al., DAFT: a universal module to interweave tabular data and 3D images in CNNs, NeuroImage 260 (2022) 119505.

[26] A. Sharma, E. Vans, D. Shigemizu, K.A. Boroevich, T. Tsunoda, DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture, Sci. Rep. 9 (1) (2019) 11399.

[27] O. Bazgir, R. Zhang, S.R. Dhruba, R. Rahman, S. Ghosh, et al., Representation of features as images with neighborhood dependencies for compatibility with convolutional neural networks, Nat. Commun. 11 (1) (2020) 4391.

[28] Y. Zhu, T. Brettin, F. Xia, A. Partin, M. Shukla, et al., Converting tabular data into images for deep learning with convolutional neural networks, Sci. Rep. 11 (1) (2021) 11325.

[29] M.I. Iqbal, M.S.H. Mukta, A.R. Hasan, S. Islam, A dynamic weighted tabular method for convolutional neural networks, IEEE Access 10 (2022) 134183–134198.

[30] T. Matsuda, K. Uchida, S. Saito, S. Shirakawa, HACNet: End-to-end learning of interpretable table-to-image converter and convolutional neural network, Knowl.-Based Syst. 284 (2024) 111293.

[31] V. Gómez-Martínez, F.J. Lara-Abelenda, P. Peiro-Corbacho, D. Chushig-Muzo, C. Granja, et al., LM-IGTD: A 2D image generator for low-dimensional and mixed-type tabular data to leverage the potential of convolutional neural networks, 2024, arXiv:2406.14566.

[32] M. Bragilovski, Z. Kapri, L. Rokach, S. Levy-Tzedek, TLTD: Transfer learning for tabular data, Appl. Soft Comput. 147 (2023) 110748.

[33] C. Andersson, A.D. Johnson, E.J. Benjamin, D. Levy, R.S. Vasan, 70-Year legacy of the framingham heart study, Nat. Rev. Cardiol. 16 (11) (2019) 687–698.

[34] C.M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag, Berlin, Heidelberg, 2006.

[35] S. Shalev-Shwartz, S. Ben-David, Understanding machine learning: From theory to algorithms, Cambridge University Press, Cambridge, 2014.

[36] N. Hollmann, S. Müller, L. Purucker, A. Krishnakumar, M. Körfer, S.B. Hoo, et al., Accurate predictions on small data with a tabular foundation model, Nature 637 (8045) (2025) 319–326.

[37] M. Duerden, N. O'Flynn, N. Qureshi, Cardiovascular risk assessment and lipid modification: NICE guideline, Br. J. Gen. Pract. 65 (636) (2015) 378–380.

[38] D.M. Nathan, J. Kuenen, R. Borg, H. Zheng, D. Schoenfeld, et al., Translating the A1C assay into estimated average glucose values, Diabetes Care 31 (8) (2008) 1473–1478.

[39] G. Tuerhong, S.B. Kim, Gower distance-based multivariate control charts for a mixture of continuous and categorical variables, Expert Syst. Appl. 41 (4) (2014) 1701–1707.

[40] D.G. Bonett, Point-biserial correlation: Interval estimation, hypothesis testing, meta-analysis, and sample size determination, Br. J. Math. Stat. Psychol. 73 (2020) 113–144.

[41] M. Baak, R. Koopman, H. Snoek, S. Klous, A new correlation coefficient between categorical, ordinal and interval variables with pearson characteristics, Comput. Statist. Data Anal. 152 (2020) 107043.

[42] P. Di Lena, L. Margara, Optimal global alignment of signals by maximization of pearson correlation, Inform. Process. Lett. 110 (16) (2010) 679–686.

[43] H.-C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, et al., Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imaging 35 (5) (2016) 1285–1298.

[44] Y. Zhang, P. Tiňo, A. Leonardis, K. Tang, A survey on neural network interpretability, IEEE Trans. Emerg. Top. Comput. Intell. 5 (5) (2021) 726–742.

[45] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626.

[46] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, et al., PyTorch: An imperative style, high-performance deep learning library, in: H.M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E.B. Fox, R. Garnett (Eds.), Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada, 2019, pp. 8024–8035.

[47] R. Liaw, E. Liang, R. Nishihara, P. Moritz, J.E. Gonzalez, I. Stoica, Tune: A research platform for distributed model selection and training, 2018, arXiv preprint arXiv:1807.05118.

[48] P. Jousilahti, E. Vartiainen, J. Tuomilehto, P. Puska, Sex, age, cardiovascular risk factors, and coronary heart disease: a prospective follow-up study of 14 786 middle-aged men and women in Finland, Circulation 99 (9) (1999) 1165–1172.

[49] E.C. Leritz, R.E. McGlinchey, I. Kellison, J.L. Rudolph, W.P. Milberg, Cardiovascular disease risk factors and cognition in the elderly, Curr. Cardiovasc. Risk Rep. 5 (2011) 407–412.

[50] G. Mercuro, M. Deidda, A. Piras, C.C. Dessalvi, S. Maffei, G.M. Rosano, Gender determinants of cardiovascular risk factors and diseases, J. Cardiovasc. Med. 11 (3) (2010) 207–220.

[51] A.V. Poznyak, L. Litvinova, P. Poggio, V.N. Sukhorukov, A.N. Orekhov, Effect of glucose levels on cardiovascular risk, Cells 11 (19) (2022) 3034.

[52] H. Gerstein, Glucose: a continuous risk factor for cardiovascular disease, Diabetic Med. 14 (S3) (1997) S25–S31.

[53] Y. Wang, Y. Deng, Y. Zheng, P. Chattopadhyay, L. Wang, Vision transformers for image classification: A comparative survey, Technologies 13 (1) (2025) 32.