This article has been accepted for publication in IEEE Journal of Translational Engineering in Health and Medicine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JTEHM.2023.3307794

Self-supervised learning-based general laboratory progress pretrained model for cardiovascular event detection

Li-Chin Chen[1], Kuo-Hsuan Hung[1], *Student Member, IEEE,* Yi-Ju Tseng[2], *Member, IEEE,* Hsin-Yao Wang[3], Tse-Min Lu[4,5,6], Wei-Chieh Huang[4,6,7], Yu Tsao[1], *Senior Member, IEEE*

***Abstract*—Objective:** Leveraging patient data through machine learning techniques in disease care offers a multitude of substantial benefits. Nonetheless, the inherent nature of patient data poses several challenges. Prevalent cases amass substantial longitudinal data owing to their patient volume and consistent follow-ups, however, longitudinal laboratory data are renowned for their irregularity, temporality, absenteeism, and sparsity; In contrast, recruitment for rare or specific cases is often constrained due to their limited patient size and episodic observations. This study employed self-supervised learning (SSL) to pretrain a generalized laboratory progress (GLP) model that captures the overall progression of six common laboratory markers in prevalent cardiovascular cases, with the intention of transferring this knowledge to aid in the detection of specific cardiovascular event. **Methods and procedures:** GLP implemented a two-stage training approach, leveraging the information embedded within interpolated data and amplify the performance of SSL. After GLP pretraining, it is transferred for target vessel revascularization (TVR) detection. **Results:** The proposed two-stage training improved the performance of pure SSL, and the transferability of GLP exhibited distinctiveness. After GLP processing, the classification exhibited a notable enhancement, with averaged accuracy rising from 0.63 to 0.90. All evaluated metrics demonstrated substantial superiority ($p < 0.01$) compared to prior GLP processing. **Conclusion:** Our study effectively engages in translational engineering by transferring patient progression of cardiovascular laboratory parameters from one patient group to another, transcending the limitations of data availability. The transferability of disease progression optimized the strategies of examinations and treatments, and improves patient prognosis while using commonly available laboratory parameters. The potential for expanding this approach to encompass other diseases holds great promise.
Clinical impact: Our study effectively transposes patient progression from one cohort to another, surpassing the constraints of episodic observation. The transferability of disease progression contributed to cardiovascular event assessment.

***Keywords*—Cardiovascular diseases, cardiometabolic disease, disease progression, laboratory examinations, time-series data, pre-train model, representation learning, self-supervised learning, transfer learning.**

I. Introduction

REGULAR surveillance stands as an imperative facet within cardiovascular disorders management [1]. Laboratory analysis constitutes a vital component, involving multifarious chemical tests that scrutinize blood, urine, or body tissue specimens. These tests gauge the body's response to food intake, medication, and treatment, thus providing crucial insights into disease progression and signaling the need for medication or dietary modifications. For chronic diseases, laboratory results are more meaningful when observed longitudinally rather than episodically. A vast repository of longitudinal data has been accumulated for prevalent diseases such as hypertension (HTN) and diabetes mellitus (DM). Nonetheless, the inherent nature of these data, characterized by irregularity, absenteeism, and sparsity, presents challenges in leveraging their full potential for machine learning applications. Conversely, rare or specific cases are frequently associated with a limited patient population and episodic observations, also impeding the integration of machine learning technology in their progression assessment. This study employed self-supervised learning (SSL) to pretrain a **g**eneralized **l**aboratory **p**rogress (GLP) model. GLP captures the overall progression of common laboratory markers in prevalent cardiovascular cases, with the intention of transferring this knowledge to aid in the detection of target vessel revascularization (TVR) occurrences in patients undergoing percutaneous coronary intervention (PCI).

Challenging nature of laboratory data

Diverse methodological approaches can be adopted depending on the distinct characteristics of the population under study. Cross-sectional studies, which capture the patient status at a single time point, provide only a temporary snapshot and preliminary glimpses into future disease progression. In contrast, cohort studies, with their longitudinal observations, offer a more comprehensive understanding of disease development [2]. However, the collection of such data over an extended period can be intricate, time-consuming, and costly, frequently suffered from patient dropouts and incomplete data [3, 4].

Consequently, rare or specific diseases frequently resort to cross-sectional studies due to limited patient size, yielding episodic observations. In the context of prevalent cases, longitudinal data is more readily accessible owing to continuous follow-ups. Nonetheless, these observations heavily rely on

[1] Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan.
[2] Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu, Taiwan.
[3] Department of Laboratory Medicine, Chang Gung Memorial Hospital at Linkou, Taoyuan City, Taiwan.
[4] Division of Cardiology, Department of Internal Medicine, Taipei Veterans General Hospital, Taipei, Taiwan.
[5] Department of Health Care Center, Taipei Veterans General Hospital, Taipei, Taiwan.
[6] Department of Internal Medicine, School of Medicine, College of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan.
[7] Department of Biomedical Engineering, National Taiwan University, Taipei, Taiwan.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Translational Engineering in Health and Medicine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JTEHM.2023.3307794

patient adherence, insurance regulations, clinical guidelines, and the clinical judgment of physicians. Any disruption to these factors can lead to irregularity and sparsity, where observations may be skipped, or sampled irregularly over a prolonged period [5, 6]. Similarly to electronic health records (EHRs), laboratory test records encompass a wealth of abundant and longitudinal patient information, still, notable for their irregularity, temporality, and sparsity, often accompanied by noisy outliers and missing values [3, 4, 6].

Related works

Machine learning techniques have undergone extensive investigation, facilitating diverse applications that contribute to clinical care through the utilization of EHRs [7, 8, 9]. SSL has recently gained attention due to its ability to derive labels for training data directly from the data itself [10, 11, 12, 13, 14]. This offers a unique opportunity to leverage the vast amounts of data available without relying on quality annotations [15, 16, 13]. Generally, SSL can be categorized into generative and contrastive learning approaches [17, 18]. Generative models possess the capability to generate new samples from the underlying distribution or recover the original data distribution [13, 19]. For instance, Simone et al. [19] showcased the utilization of a generative adversarial network (GAN) for the synthesis of electrocardiography (ECG) data. Their study achieved remarkable results by generating a wide spectrum of ECG patterns that preserved synchronization and abnormalities. Yoon et al. [14] leveraged SSL strategies to impute corrupted values and train on unlabeled data in the domain of genomics and clinical data. Furthermore, Lee et al. [20] employed GPT-4 to summarize physician-patient conversations and generated clinical notes.

Meanwhile, contrastive learning aims to capture the relationship between input data and prediction targets, thereby generating a global contextual representation that is shared among samples [13, 16]. For instance, Kiyasseh et al. [16] devised a temporal and spatial discriminative approach for ECG analysis, extracting patient-specific representations through leveraging contrastive loss. Zhang et al. [21] employed contrastive learning to capture the distances between temporal and frequency components, and applied the pretrain representations to various time-series databases, including ECG, human activity recognition, and physical status monitoring. Wickstrøm et al. [22] proposed a contrastive framework based on mixing up augmentation for uni- and multivariate time-series data, which was then transferred to ECG classification. Furthermore, Ouyang et al. [23] trained an encoder using unlabeled retinal images through contrastive learning, and subsequently fine-tuned the encoder for the classification of reference and non-reference cases.

Many SSL applications focus on learning from a pretext task and transferring the learned representations to a different domain. This entails transfer learning, a concept that improves a classifier in one domain (pretext task) using more readily obtainable data and then applying this acquired knowledge to another domain (downstream task) [24]. For instance, Tang et al. [25] utilized a teacher-student self-training model to capture information from a large-scale unlabeled dataset of wearable and mobile sensing data, which was subsequently transferred to seven different datasets with varying sensor types, populations, and protocols. Similarly, Spathis et al. [26] used activity accelerometer sensor data as input to forecast heart rate and transferred the learned representations to capture physiologically meaningful and personalized information using linear classifiers.

Nontheless, the aforementioned works assume that signals and information were collected on a regular basis and do not address the issues of irregularity, absenteeism, and sparsity commonly encountered in EHRs. Tipirneni et al. [27] bypass the irregular, absent, and sparse nature of EHRs by training the model in a SSL manner and mask out unobserved forecasts in the loss function during training. Furthermore, SSL applications have been recognized as challenging in terms of finding effective pretrain tasks [17]. Most studies have relied on an empirical trial-and-error approach to identify the most suitable pretext tasks. The transferability of pretrain models has shown mixed results when applied to more specific domains such as medicine [28]. For instance, Liu et al. [29] were unable to successfully transfer ImageNet for detecting lymph node metastasis in pathology images, citing significant domain differences between natural scenes and pathology images as the reason for the transfer failure. Another example is the discussion surrounding language representation models for the biomedical domain [30]. Gu et al. [31] argued that models trained on domain-specific vocabulary outperform those trained on general corpora due to differences in word distributions between general and biomedical corpora .

Cardiovascular diseases and associated risk factors

Cardiovascular disease stands as one of the leading causes of global mortality. Established traditional risk factors incorporate HTN, DM, and smoking [32]. These risk factors can contribute to endothelial injury, plaque formation, and coronary thrombus formation [33], consequently driving the progression of cardiovascular disease. PCI has emerged as a widely employed treatment modality for cardiovascular disease [34].

Although the implementation of drug-eluting stents (DES) has significantly reduced the incidence of TVR in recent years [35, 36], the occurrence of TVR after DES implantation, with an incidence ranging from 3% to 20%, remains a prevalent clinical concern [37, 38, 39]. Hence, the prevention of TVR and the reduction of re-admission rates continue to be significant clinical challenges in the field of cardiovascular medicine following PCI. With respect to the timing of cardiac events, patients undergoing PCI face a risk of subsequent adverse events, including TVR [40, 41]. However, achieving a consensus on accurate preprocedural risk stratification and prognosis assessment to identify high-risk patients prior to PCI remains an ongoing pursuit.

TVR is associated with complex pathophysiological mechanisms involving lipid metabolic disorders [42] and inflammatory processes [43]. Several previous studies have explored potential predictive factors linked to a high incidence of TVR

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Translational Engineering in Health and Medicine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JTEHM.2023.3307794

based on patient and procedure-related variables [44, 45, 46]. However, dedicated applications for TVR prediction are yet to be developed. Most studies have focused on identifying comprehensive predictors for TVR or developing prediction models without specifically targeting individualized risks [45, 47, 48]. The collection of data for comprehensive predictors may introduce burden and complexities when integrating such applications into routine clinical practice and thus warrants careful consideration.

Laboratory markers of cardiovascular diseases

Earlier studies have indicated that preprocedural parameters are associated with cardiovascular disease. Total cholesterol levels (Chol) and low-density lipoprotein cholesterol (LDL-c) are strongly linked to cardiometabolic diseases and widely accepted in diagnostic practices. Conversely, the plasma level of high-density lipoprotein cholesterol (HDL-c) exhibits an inverse relationship with the risk of cardiovascular diseases [49]. Clinical studies have highlighted the connection between circulating white blood cells (WBCs) and cardiovascular outcomes, demonstrating that elevated WBC count increases the short- and long-term risk in patients with acute coronary syndromes (ACSs) [50]. Hage et al. [51] reported that baseline fasting blood glucose (glucose AC) predicts restenosis, suggesting that focusing on glucose reduction rather than solely normalizing glucose levels is more beneficial [51]. Additionally, serum uric acid (UA) has been identified as a prognostic cardiovascular biomarker, predicting total and cardiovascular mortality in the context of secondary prevention of coronary artery disease, as demonstrated by the Verona Heart Study [52]. Furthermore, the National Cholesterol Education Program III (NCEP III) recommends the use of Chol or LDL-c in conjunction with HDL-c (Chol/HDL-c, LDL-c/HDL-c) as markers for screening and treating patients with cardiovascular disease, along with the utilization of the 10-year risk Framingham scoring assessment [53].

II. Methodology

To address the challenges previously mentioned, we have devised the following propositions for our work: (1) In order to address the challenges stemming from irregularity, absenteeism, and sparsity within longitudinal observations, we have deployed interpolation and SSL techniques to infer absented data. (2) For patients with limited numbers and episodic observations, we developed a pretrain model specifically tailored to capture the temporal latent representation of prevalent cases and transfer disease progress knowledge to these smaller cohorts. (3) Our work is based on commonly available resources, avoiding the need to initiate new trials for extensive patient data collection. We focused on six laboratory parameters: the Chol/HDL-c ratio, LDL-c, the LDL-c/HDL-c ratio, glucose AC, WBC, and UA. (4) Our work leverages the intercorrelation among cardiometabolic diseases as an indication of pretext and downstream tasks.

Our objective is to construct a two-stage pretraining model that captures the laboratory progress of general cases and utilizes this information to predict cardiac events in another patient group. Prior research has indicated that incorporating a two-step training approach, involving pretraining the model on a domain-general dataset followed by training on domain-specific datasets, yields enhancements in transitioning representations to the downstream task [12, 30, 23]. Hence, we propose a two-stage training process: Stage 1 involves learning general laboratory progress information based on interpolated data, followed by Stage 2, where SSL is employed to refine the model's progression representation using non-interpolated data. Subsequently, GLP model is fine-tuned to classify the occurrence of TVR.

The following sections outline the detailed methodology of GLP. Initially, we introduced the interpolation and framing method for longitudinal data. Subsequently, we outlined the design of the GLP model, along with the training algorithm and the design of downstream classifier. We also elucidated the validation methods employed for both the pretext and downstream tasks. Lastly, we provided details regarding the patient recruitment process and the datasets utilized for both tasks.

Interpolation methods

Interpolation serves the purpose of inferring values that lie between two known observations. It aims to approximate the values of $\hat{f}(x)$ that fulfill the interpolation conditions $\hat{f}(x_j) = y_j$ for $j = 0, 1, \ldots, n$. This study encompassed the evaluation of three interpolation techniques: linear interpolation, piecewise cubic Hermite interpolating polynomial (PCHIP), and barycentric interpolation. The values in linear interpolation were derived by considering the gradients of the known observations, denoted as:

$$\hat{y}_j = y_i + (t_j - t_i)\frac{(y_k - y_i)}{(t_k - t_i)}, \tag{1}$$

where t signifies the time of estimation, $\hat{y}_j$ denotes the estimated value at t_j, and $i < j < k$. Meanwhile, PCHIP interpolation [54] defines $d_j = (y_k - y_j)/(t_k - t_j)$ as the slopes at x_j. If the signs of d_j and d_i differ or either of them equals zero, $\hat{y}_j$ is set to 0. Otherwise, it is determined using the weighted harmonic mean, expressed as:

$$\hat{y}_j = \frac{(w_1 + w_2)}{\frac{w_1}{d_i} + \frac{w_2}{d_j}}, \tag{2}$$

where $w_1 = 2(t_k - t_j) + (t_j - t_i)$ and $w_2 = (t_k - t_j) + 2(t_j - t_i)$. Finally, for barycentric interpolation [55, 56], a given set of nodes $x_0, x_1, \ldots, x_n$ and masses $w_0, w_1, \ldots, w_n$ are utilized to determine the functions $w_0(x), w_1(x), \ldots, w_n(x)$ satisfying:

$$x = \frac{\sum_{i=0}^{n} w_i(x) x_i}{\sum_{i=0}^{n} w_i(x)}. \tag{3}$$

Here, x represents the barycenter of the nodes, which can be employed for interpolation using:

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Translational Engineering in Health and Medicine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JTEHM.2023.3307794

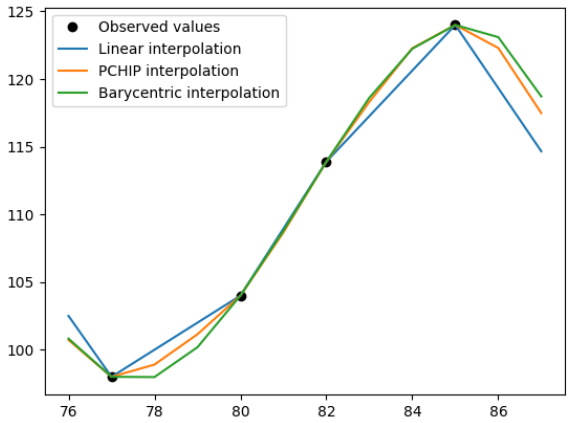

Fig. 1. Interpolation results based on different methods. This showcases a segmented period of glucose AC. The x-axis corresponds to the timeline, while the y-axis represents the laboratory values.

$$\hat{y} = \sum_{i=0}^{n} b_i(x) f_i, \tag{4}$$

where b_i corresponds to the linear function. Fig. 1 illustrates the segmented period of glucose AC, employing different interpolation methods.

Longitudinal data framing

By defining the laboratory observations for each patient during the study period as $y_{t_0}, y_{t_1}, \ldots, y_{t_n}$, where t_i represents the months in the timeline and $t_i \in \mathbb{R}$, we designate y_{t_0} and y_{t_n} as the actual observed values. Let y_{t_m} denote the second-to-last observed value of a patient. Interpolation takes place between y_{t_0} and y_{t_m} when y_{t_i} is missing. The observations were organized into a frame with a designated time interval r. Therefore, the longitudinal data is framed as $y_{t_i} : y_{t_{i+r}}$, with subsequent frames incrementing by one step $(i+1)$ while $i+r \leq m-1$. If $\|t_i, t_{m-1}\| \leq r$, the frame is omitted. These segmented frames, denoted as the interpolated data, serve as the input of Stage 1. Their corresponding prediction target is $y_{t_{i+r+1}}$, where $i+r+1 \leq m$.

From the previous stage, the frame $y_{t_{m-r}} : y_{t_m}$ and the last observed value y_{t_n} are isolated. They are the non-interpolated data in Stage 2, with $y_{t_{m-r}} : y_{t_m}$ serving as the input and y_n as the prediction target. Considering that coronary stent trials primarily focus on target vessel/lesion-related clinical outcomes within the shorter term, particularly in the first 12 months post-PCI [57], we set r as 12 months. The framing process is visually depicted in Figure 2.

GLP model design

GLP was designed to monitor the laboratory progress of patients and forecast their observations for the subsequent month. Fig. 3 presents the model architecture of GLP, which consists of two main components: the longitudinal iterative

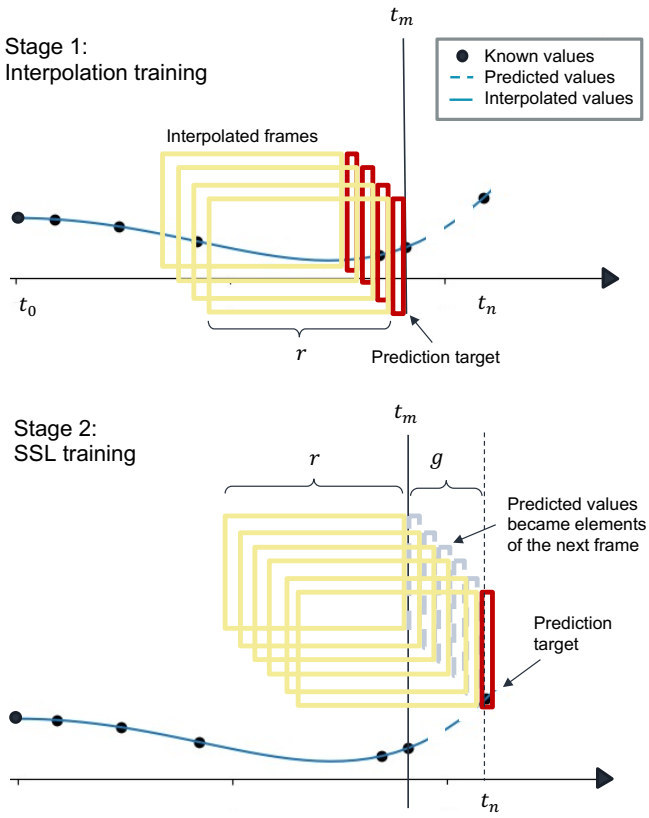

Fig. 2. Logitudinal data framing for two-stage training. SSL: self-supervised learning; r: time interval (frame size); g: time gap between the input frame and prediction target, where $g \leq \frac{r}{2}$.

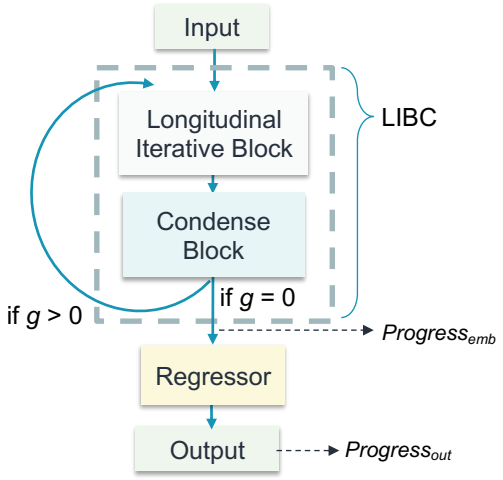

Fig. 3. GLP model design. g: time gap between the input frame and prediction target; $Progress_{emb}$: latent output of $LIBC$; $Progress_{out}$: latent output of $regressor$.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Translational Engineering in Health and Medicine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JTEHM.2023.3307794

block ($LIBC$) and a $regressor$. The $LIBC$ comprises a Bidirectional Long Short-Term Memory (BiLSTM) layer and a Fully Connected (FC) condensing layer. A Rectified Linear Unit (ReLU) activation function is applied after each layer to enhance the non-linearity of the model. The BiLSTM processes the input data in both the forward and backward directions [58], enabling the capture of contextual information from past data. The number of hidden nodes for the BiLSTM layer was set to 5. The FC condensing layer is utilized to compress the output of the BiLSTM layer back to the original input, facilitating an autoregressive flow.

On the other hand, the $regressor$ consists of two FC layers followed by a ReLU activation function in between. The number of hidden nodes for the $regressor$ is set as 5, 2, 2, and 1, respectively. The model used the Mean Squared Error (MSE) as loss function, which measures the average squared difference between the predicted values and the actual values.

Two-stage training algorithm

During Stage 1 training, the interpolated data were utilized in supervised learning. An additional parameter called the number of certainty mask ($certain$) was introduced in Stage 1. It denotes the required number of real observations within a frame, indicating the level of real observations the models require to generate reliable predictions. It also serves as a means to address uncertainty within the training data. Adhering to insurance regulations, patients were scheduled for cardiovascular disease examinations every three months, resulting in a maximum of four actual observations within a 12-month timeframe. Consequently, the range of 0 to 5 was explored for $certain$, and the optimal value was determined as a parameter setting for GLP.

In Stage 2 training, we employed an autoregressive based SSL, which takes inputs from a time series regressed on previous inputs from the same time series. The probability of each input is conditioned on the preceding input, and can be formulated as:

$$\max_{\theta} p_\theta(x) = \sum_{t=1}^{T} \log p_\theta(x_t \mid x_{1:t-1}), \tag{5}$$

where x_t represents the input at time t, p_θ denotes the probability, and $\max_\theta p_\theta$ signifies the maximized likelihood [13]. Starting from frame $y_{t_{m-r}} : y_{t_m}$, the model utilizes the parameters θ obtained from Stage 1 and predicts $y_{t_{m+1}}$. The subsequent input frame becomes $y_{t_{m-r+1}} : y_{t_{m+1}}$, with the prediction target located at $y_{t_{m+2}}$. This process continues until the prediction target reached y_{t_n}. The predicted data then becomes part of the training data for the next frame, allowing the model to learn from its own generation.

To summarize, in Stage 1, there is no time gap ($g = 0$) between the input frame and the prediction target. The input passes through the $LIBC$ once and then enters the $regressor$. On the other hand, in Stage 2, $g = n - m - 1$ and $g > 0$, and the process iterates until the prediction target is reached. The training algorithms for Stage 1 and 2 are explicitly outlined in Algorithm 1 and 2. Each laboratory parameter was trained individually and optimized to achieve the best performance. The six pretrained GLP models were subsequently concatenated in a multimodal fashion and utilized for domain transfer to perform TVR occurrence classification, as illustrated in Figure 4.

Algorithm 1 Stage 1 training

Input: $X_{interpolated}$, batch size N

for $epoch = 0$ to 50 **do**
  for $batch\{x_k\}_{k=1}^{N} \in X_{interpolated}$ **do**
    pass x_k through $LIBC$ and $regressor$
    Optimize using Adam optimizer
  end for
  Save the parameters θ of the model
end for

Output: $\hat{y}$

Algorithm 2 Stage 2 training

Input: $X_{non-interpolated}$, batch size N, time gap between the input frame and prediction target g, parameters θ from stage 1 training

for $epoch = 0$ to 50 **do**
  for $batch\{x_k\}_{k=1}^{N} \in X_{non-interpolated}$, $\{g_k\}_{k=1}^{N} \in g$ **do**
    if $max(g_k) = 0$ **then**
      pass x_k through $LIBC_\theta$
    end if
    if $max(g_k) > 0$ **then**
      for $t \in \{1, 2, \ldots, max(g_k)\}$ **do**
        pass x_k through $LIBC_\theta$
      end for
      for $g_i \in \{1, 2, \ldots, k\}$ **do**
        select estimated $\hat{y_k}$ at time stamp $g_i + 1$
      end for
    end if
    pass $\hat{y_k}$ through $regressor_\theta$
    Optimize using Adam optimizer
  end for
  Update the parameters θ of the model
end for

Output: $\hat{y}$

Input vector and normalization

The input of GLP consists of five features: age, gender, $certain$, discrete value encoding, and normalized laboratory values. Numeric values, such as age and laboratory values, are normalized using the natural logarithm of one plus the input ($log1p$). This transformation ensures that the values are projected into a vector space above zero, preventing potential errors that could arise from maldistribution between positive and negative values. Additionally, $log1p$ is accurate for small values of x, ensuring that $1 + x = 1$ with floating-point accuracy without significantly altering the original value [59]. The use of $log1p$ also prevents information leakage [60], unlike scaling which necessitates knowledge of the maximum and minimum values.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Translational Engineering in Health and Medicine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JTEHM.2023.3307794

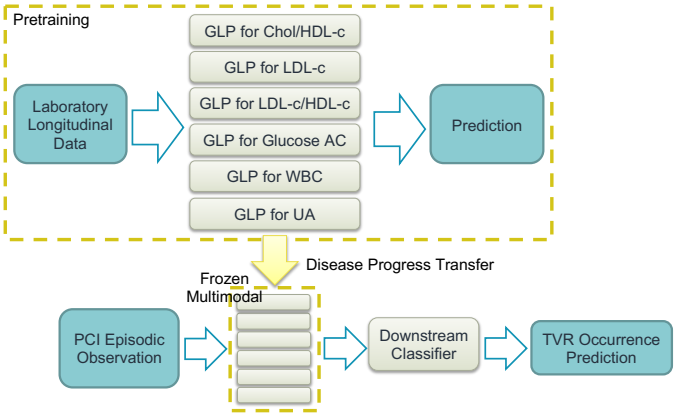

Fig. 4. Training pipeline. GLP: general laboratory progress pretrain model; Chol/HDL-c: ratio of total cholesterol and high-density lipoprotein cholesterol; LDL-c: low-density lipoprotein cholesterol; LDL-c/HDL-c: ratio of low-density and high-density lipoprotein cholesterol; Glucose AC: fasting blood sugar; WBC: white blood cells; UA: Uric Acid; PCI: percutaneous coronary intervention; TVR: target vessel revascularization.

One-hot encoding is employed for gender, certainty mask, and discrete value encoding. Gender is binary-encoded, with a value of 1 assigned to male and 0 assigned to female. The $certain$ is also binary-encoded. It indicated whether the value is a real observation or an estimation obtained through interpolation (true observation/estimated value). Discrete value encoding categorizes laboratory results into two groups (low and high) or three groups (low, normal, and high) based on different threshold values (summarized in Appendix B).

GLP validation

The validation of GLP was conducted by forecasting y_n without interpolation support within $\frac{r}{2}$. To further assess the performance of the proposed two-stage training process, we used SSL as baseline (Stage 2 training), and compared with the following approaches: (1) Supervised training based on interpolated data (Stage 1 training). (2) Hybrid training, which combines interpolated and non-interpolated data, and performs supervised and self-supervised training simultaneously. (3) Two-stage training, where GLP first reaches a local minimum loss using interpolated data and then undergoes domain-specific adjustments using SSL.

Regardless of the employed training approaches, the data were randomly divided into training and testing datasets in an 80:20 ratio. All training processes utilized the 5-fold cross-validation technique, and the reported results represent the mean value obtained from five repetitions of the training process [61]. Ablation studies were conducted to analyze the performance of different training processes combined with various interpolation methods. The outcome of the model was assessed using the R-squared (R^2) metric, which indicates the proportion of variance in the dependent variable that can be predicted by the independent variables in the model. R^2 values range from zero to one, with a value of one representing a perfect fit to the data and a value of zero indicating a poor fit. $R^2 < 0$ suggests that the model performs worse than a horizontal mean line passing through the mean value of the data. Statistical significance to determine differences in model performance was assessed using an independent T-test, with $p < 0.05$ indicating statistical significance.

Downstream classifier design and validation

The collection of the six laboratory values occurred at the time of performing PCI. Subsequently, g (month-based) was computed based on the temporal disparity between the PCI and TVR dates. Patient information, including gender, age, and the six laboratory parameters, was collected and normalized following the aforementioned procedures. Due to the imbalanced distribution of patients between those with TVR occurrence (42) and those without (441), TVR-negative cases were randomly downsampled. As a result, only 84 patients entered the training process. Patient data are processed by the frozen GLPs, which are concatenated in a multimodal fashion (as shown in Figure 4).

Owing to the limited patient volume, non-neural network algorithms were chosen for training the downstream classifier. The included methods are Light Gradient Boosting Machine (LightGBM), Support Vector Machine (SVM), Logistic Regression (LR), and K-Nearest Neighbors (KNN). These approaches were selected for their diverse mechanisms. Cohen's Kappa was calculated to assess the agreements between the classifiers, and the mean value of Cohen's Kappa represented the overall agreements among the pairwise classifiers. A value of Kappa ≤ 0 indicates no agreement, while an increase in Kappa signifies an increase in agreement, with a value of 1 indicating perfect agreement.

Additionally, we compared the latent progress representations produced by GLP by extracting the outputs of $LIBC$ ($Progress_{emb}$) and $regressor$ ($Progress_{out}$), depicted in Figure 3. The transferred results based on the original data (normalized but not processed with GLP), $Progress_{emb}$, and $Progress_{out}$ were compared. The reported performance represents the mean value obtained after executing the training process five times. The downstream classifier is a binary classifier that distinguishes TVR occurrence (positive/negative). The evaluation metrics for TVR incidence include the area under the receiver operating characteristic curve (AUROC), accuracy, sensitivity, specificity, precision, and F1 score. To determine the statistical significance of the differences in evaluation metrics between the original data and the extracted representations, an independent T-test was also conducted.

Furthermore, we plotted the distribution of the original data, $Progress_{out}$ at $g/2$, and $Progress_{out}$ at g to visually depict the contribution of GLP throughout the process. Here, $g/2$ signifies that the iteration before reaching the intended event.

Patient recruitment and datasets

Two datasets were obtained from two diverse medical institutes. The pretext dataset was obtained from the Chang Gung Research Database [62], a multi-institutional electronic medical records database comprising original medical records of seven medical institutes in Taiwan. We included patients who were diagnosed with HTN prior to DM. Patients diagnosed with hypertension before the age of 40, those with any

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Translational Engineering in Health and Medicine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JTEHM.2023.3307794

oncology visits, or individuals with observations spanning less than a year were excluded. The date of diagnosis was determined based on the International Classification of Diseases (ICD) encoding or the date of medication prescription. The ICD codes and medications used for the indications can be found in Appendix A. When a patient had been coded as having HTN or DM twice or more during a year, the onset date of the disease was defined as the first coded date. If the date of the first medication prescription preceded the ICD coded date, then the earlier date was designated as the onset date of the disease. The time interval between the HTN and DM diagnoses was set to be $\geq$ 3 months.

We gathered data on patients between the onset of HTN and DM; that is, the patient was diagnosed with HTN, but yet to be determined as DM. Demographic information and laboratory data of the enrolled patients were collected, including age, sex, Chol/HDL-c, LDL-c, LDL-c/HDL-c, glucose AC, WBC, and UA. Erroneous values such as "NA" or "." were excluded. A total of 9,720 patients were included, and laboratory data were collected between January 2001 and December 2019.

The downstream dataset was acquired from the Taipei Veterans General Hospital, a tertiary hospital situated in northern Taiwan. We recruited 891 patients with noninvasive evidence of myocardial ischemia who underwent PCI between January 2005 and January 2022. Patients with ACS, acute decompensated congestive heart failure, acute or chronic infections, autoimmune diseases, malignancies with a prognosis of less than one year, unstable hemodynamic status, or those unable to receive dual antiplatelet therapy were excluded. Angiographically successful coronary intervention was defined as residual stenosis of less than 30%, and coronary thrombolysis in myocardial infarction grade 3 flow was achieved at the conclusion of the procedure without any significant complications. All patients were followed up and to monitor the occurrence of TVR.

Patients necessitating TVR were labeled as positive, and the corresponding dates were recorded. Patients without TVR until the end of January 2022 were classified as negative, with the date of January 31, 2022 recorded as the endpoint. Within this dataset, the available information encompassed age, sex, PCI date, TVR date, and the aforementioned six laboratory values (collected during the PCI procedure). The time intervals between the PCI and TVR dates were calculated. Patients with incomplete data were excluded from the analysis to avoid deviations into this study, as the imputation of missing values could potentially have that effect. Consequently, a total of 483 patients were included in the subsequent analysis, comprising 42 TVR and 441 non-TVR cases.

The pretext dataset are longitudinal observations that consists of multiple events, whereas the downstream dataset are episodic records that consists of one observation event for each patient. All patient data were de-identified prior to analysis. This study was approved by the Institutional Review Board of the Chang Gung Medical Foundation (No. 202000376B0) and Taipei Veterans General Hospital (No. 2019-12-012CC).

TABLE I
DEMOGRAPHIC INFORMATION OF THE RECRUIT PATIENTS

Items		Values	
HTN to DM patient (Pretext dataset)	Gender (n, %)	Male Female	5032 (0.518) 4688 (0.482)
	HTN onset age (mean, SD)	56.965	9.653
	DM onset age (mean, SD)	61.884	9.878
	Duration between HTN and DM (year, mean, SD)	4.992	3.945
PCI patients (Downstream dataset)	Gender (n, %)	Male Female	402 (0.832) 81 (0.168)
	PCI treatement age (mean, SD)	66.828	12.471
	TVR occurrence (n, %)	Yes No	42 (0.087) 441 (0.913)
	Duration between PCI and TVR occurrence (year, mean, SD)	8.277	4.088
	Occurrence duration (year, mean, SD)	2.319	2.640
	Non-occurred duration (year, mean, SD)	8.844	3.734

n: number of samples; SD: Standard deviation; HTN: hypertension; DM: diabete mellitus; PCI: percutaneous coronary intervention; TVR: target vessel revascularization. Duration between HTN and DM signifies the cumulative observed timeframe for each individual. Occurrence duration delineates the temporal interval between PCI and TVR, and non-occurred duration characterizes the observed time span for patients who did not experienced TVR.

III. RESULTS

Table I illustrates the demographic information of the patients enlisted from the two datasets. It is noteworthy that patients in the pretext dataset (HTN to DM patients) exhibit a relatively younger age compared to those in the downstream dataset (PCI patients). It is observed that TVR typically occurs within a period of 2.32 $\pm$ 2.64 years. Fig. 5 depicted the overall performance of GLP, assessed by averaging the R^2 results across $certain$ ranging from 0 to 5. Fig. 5(a) provides an overview of the general disparities among different training approaches, incorporating all interpolation methods. It reveals that SSL and two-stage training exhibit similar performances, both achieving a mean R^2 value of 0.46 when rounded to the second decimal place. On the other hand, supervised training (Stage 1 training) and hybrid training achieved lower means with larger variations.

Fig. 5(b) to Fig. 5(d) delves into the exploration of two-stage training, supervised training, hybrid training, respectively. All figures inspect into the effect of different interpolation methods. Observing from the figures, the two-stage training approach (Fig. 5(b)) is the only approach that surpasses SSL (baseline). Using linear (R^2 = 0.49, p = 0.508) and PCHIP (R^2 = 0.48, p = 0.603) both yield better results than SSL, although the differences did not reach statistical significance. However, when compared to barycentric interpolation, both linear (p = 0.031) and PCHIP (p = 0.046) show significantly higher performance. Other training approaches, including supervised training depicted in Fig. 5(c) and hybrid training in Fig. 5(d), demonstrate weaker performance compared to the horizontal mean line ($R^2 < 0$), with larger variations.

Table II provides a summary of the optimal $certain$ settings for each GLP, which were found to be inconsistent across

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Translational Engineering in Health and Medicine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JTEHM.2023.3307794

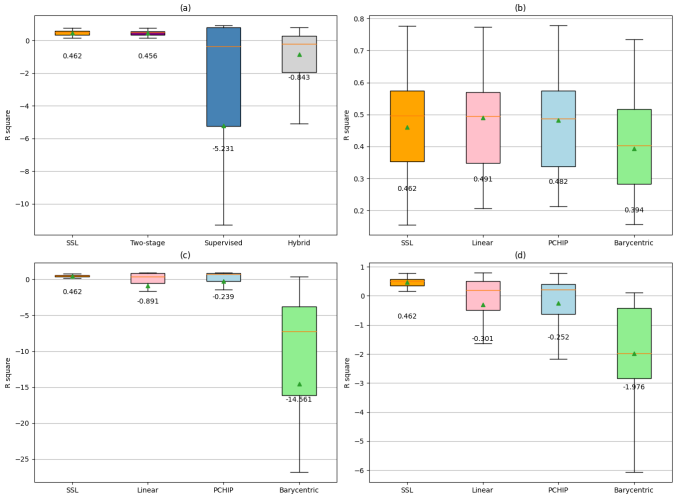

Fig. 5. R^2 values for different training approaches. SSL (baseline) is compared with (a) averaged R^2 across different training approaches; (b) two-stage training employing different interpolation methods; (c) supervised training employing different interpolation methods; and (d) hybrid training employing different interpolation methods. The mean value is represented by the green triangle and its corresponding numerical figure is provided in the text below. The presented outcomes are an average of five repetitions of training and prediction, determined by aggregating the R^2 outcomes across $certain$ values ranging from 0 to 5. SSL: self-supervised learning.

TABLE II
R^2 OF GLP BASED ON OPTIMIZED CERTAINTY MASK CONFIGURATIONS

	certain	SSL	Two-stage	*p-value*
Linear				
Chol/HDL-c	5 (2.4 months)	0.509	**0.602**	
LDL-c	3 (4 months)	0.373	**0.376**	
LDL-c/HDL-c	3 (4 months)	0.515	**0.522**	
Glucose AC	4 (3 months)	0.648	**0.666**	
WBC	4 (3 months)	**0.778**	0.774	
UA	4 (3 months)	0.353	**0.454**	
Avg.	3.8 (3.1 months)	0.529	**0.566**	0.773
PCHIP				
Chol/HDL-c	5 (2.4 months)	0.392	**0.586**	
LDL-c	3 (4 months)	**0.367**	0.341	
LDL-c/HDL-c	2 (6 months)	0.505	**0.521**	
Glucose AC	4 (3 months)	**0.644**	0.602	
WBC	5 (2.4 months)	0.751	**0.779**	
UA	5 (2.4 months)	0.316	**0.373**	
Avg.	4 (3 months)	0.496	**0.534**	0.702

Chol/HDL-c: ratio of total cholesterol and high-density lipoprotein cholesterol; LDL-c: low-density lipoprotein cholesterol; LDL-c/HDL-c: ratio of low-density and high-density lipoprotein cholesterol; Glucose AC: fasting blood sugar; WBC: white blood cells; UA: Uric Acid. The differitation did not reach the level of statistical significance ($p > 0.05$). The month indicated in the brackets signifies the necessary follow-up visits for patients based on the requisite count of actual observations within a 12-month period.

TABLE III
DOWNSTREAM TASK PREDICTIONS

	AUROC	Accuracy	Sensitivity	Specificity	Precision	F1
Original data (Mean Cohen's Kappa = 0.366)						
LGBM	**0.602**	0.671	**0.717**	0.632	0.587	0.645
SVM	0.313	**0.718**	0.603	0.836	**0.804**	**0.667**
LR	0.543	0.494	0.673	0.398	0.425	0.500
KNN	**0.602**	0.635	0.343	**0.939**	0.800	0.472
Avg.	0.515	0.630	0.584	0.701	0.654	0.571
$Progress_{emb}$ (Mean Cohen's Kappa = 0.307)						
LGBM	**0.706**	**0.671**	**0.526**	**0.925**	**0.933**	**0.656**
SVM	0.601	0.435	0.450	0.511	0.231	0.290
LR	0.536	0.494	0.362	0.657	0.510	0.357
KNN	0.394	0.518	0.142	0.830	0.267	0.183
Avg.	0.559	0.530*	0.370*	0.731	0.485	0.372*
$Progress_{out}$ (Mean Cohen's Kappa = 0.941)						
LGBM	**0.973**	**0.965**	**0.933**	0.978	0.978	**0.952**
SVM	0.907	0.882	0.767	0.969	0.933	0.822
LR	0.863	0.871	0.708	**1.000**	**1.000**	0.824
KNN	0.898	0.882	0.787	0.956	0.956	0.848
Avg.	**0.910****	**0.900****	**0.799****	**0.976****	**0.967****	**0.862****

LGBM: Light Gradient Boosting Machine; SVM: Support Vector Machine; LR: Logistic Regression; KNN: K-Nearest Neighbors (KNN); AUROC: area under the receiver operating characteristics curve; Avg.: average; *: Statistical significant differences between $Progress_{emb}$ and original data; **: Statistical significant differences between $Progress_{out}$ and original data, as well as between $Progress_{out}$ and $Progress_{emb}$.

different parameters. It is noteworthy that, in most parameters, the two-stage training process consistently outperformed SSL, although the degree of improvement remained marginal ($p > 0.05$). While specifying the $certain$ value resulted in enhancing GLP performance (evidenced by an increase from 0.49 to 0.57 for linear interpolation), the correlation between R^2 and $certain$ was determined to be weak based on Pearson correlation analysis (linear: -0.002, PCHIP: 0.026, and barycentric: -0.040). This result indicated that R^2 and $certain$ are not linear correlated. The detailed performances of $certain$ for each analysis can be found in Appendix C. Based the best-performing linear interpolation approach, we trained the GLP pretrain models by optimizing $certain$ value for each parameters individually.

Table III presents the result of the downstream task. After the processing by GLP, $Progress_{out}$ exhibited a significant enhancement in classification performance. On average, the measures attained an AUROC of 0.91, accuracy of 0.90, sensitivity of 0.80, specificity of 0.98, and F1 score of 0.86. $Progress_{out}$ displayed a substantial superiority over the performance of the original data ($p < 0.01$) and $Progress_{emb}$ ($p < 0.01$). The key distinction between $Progress_{emb}$ and $Progress_{out}$ lies in the fact that $Progress_{out}$ represents a condensed version that solely indicates the laboratory values for the subsequent month, whereas $Progress_{emb}$ still retains information on age, gender, $certain$, discrete value range encoding, and laboratory values. This condensed version has a distillation effect, where the $regressor$ was trained to provide a more valuable indication of future trends based on these information.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Translational Engineering in Health and Medicine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JTEHM.2023.3307794

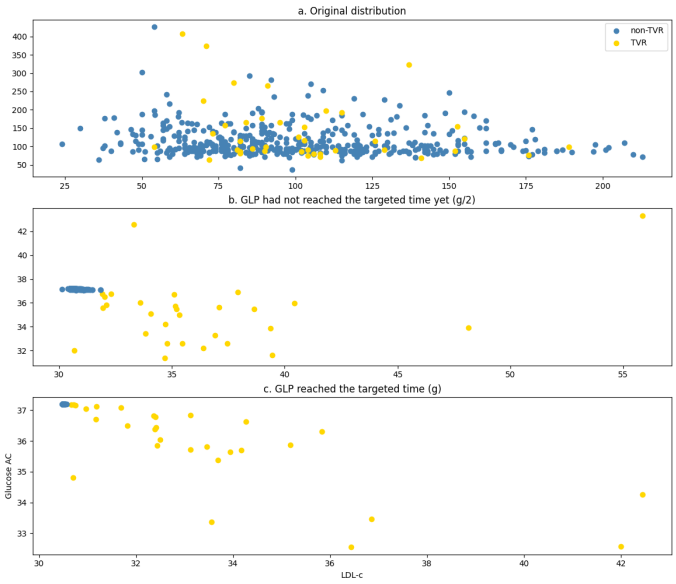

Fig. 6. Alterations in distribution prior to and subsequent to GLP processing. The values were inverse from normalization. g: time gap between the input frame and prediction target.

Notably, LGBM emerged as the best-performing algorithm for both $Progress_{emb}$ and $Progress_{out}$, while SVM demonstrated the highest performance when utilizing the original data. The Kappa score demonstrates that the agreement among classifiers increased from 0.37 to 0.94, indicating that GLP alters the distribution of the original data and simplifies the classification task, irrespective of the algorithm mechanism employed.

Fig.6 visually illustrates the changes in data distribution, using glucose AC and LDL-c as exemplars. The figure depicts that prior to GLP processing (Fig. 6a), distinguishing between TVR and non-TVR cases was challenging. However, after GLP processing, the non-TVR cases gradually converged towards a singular point, while the TVR cases displayed a more scattered distribution (as depicted in Fig. 6b to c).

IV. Discussion

We have successfully engage in translational engineering by transferring the progression of cardiovascular laboratory parameters from one patient group to another, without confining to episodic observations. To the best of our knowledge, this is the first study to apply the transfer of laboratory progression between patient groups, demonstrating that disease progression can be effectively transferred through deep neural network processing. The generality of our findings is assured. The datasets were sourced from two distinct medical institutions, signifying divergent patient populations, personnel, operational protocols, and measurement equipment. This finding opens up opportunities to leverage the trends observed in general cases for developing data-driven applications targeting patient groups with limited data availability.

Pretrained models capture the temporal dynamics and generate latent representations that enhance predictive capabilities for other tasks. With the widespread adoption of EHRs in modern hospitals, a large amount of data has been collected from prevalent cases, such as patients with HTN and DM. Conversely, specific cases, such as TVR occurance after PCI, remain limited in number. Given that HTN and DM are known risk factors for TVR [32], we successfully transferred the trend of laboratory progress observed in HTN patients (who had yet to be diagnosed as diabetic) to predict the progression of PCI patients.

Prior investigations [63] have highlighted that training deep neural network models directly through gradient descent can yield randomly initialized models that are less optimized. In contrast, commencing the training process with a pretrained model enables the preservation and utilization of previously acquired knowledge, thereby transforming our randomly initialized models into exceptional pretrained feature extractors [13]. The practice of acquiring knowledge from a more generalized domain and subsequently fine-tuning it to a specific domain during pretraining has previously exhibited efficacy [30]. It has been observed that representations acquired from supervised objectives tend to be more domain-specific and possess limited transferability to out-of-distribution domains [11, 13, 19]. Thus, modifying the model based on SSL confers enhanced flexibility and facilitates the extrapolation of knowledge across domains. This transformation enhanced our ability to infer patient progress. Despite the heterogeneous characteristics exhibited by the two patient cohorts, there existed intercorrelations among cardiometabolic diseases.

SSL proves viable in bridging the gap caused by irregularity, absenteeism, and sparsity in prolonged monitoring, enabling adequate prediction. Learning from interpolated data in advance exhibits the capacity to enhance prediction performance, however, the effects vary across interpolation methods. Linear and PCHIP methodologies are designed to acquire a continuous function, while barycentric interpolation is tailored to leverage the center of mass for interpolation. The distribution pattern of laboratory progress resembles more closely a continuous curve that extends over time, rather than being clustered in discrete bundles. Our findings reveal that linear and PCHIP interpolation yield more informative estimations, enhancing the performance of SSL. In the case of LDL-c and LDL-c/HDL-c examinations, interpolation expands the tolerance for periods of absence. Thus, our work implies that with sustainable estimation, patient risks can be monitored with less frequent returns and examinations (beyond the conventional 3-month interval), alleviating the burdens of travel and medical expenses for patients.

The results indicate that without the support of GLP, the original data lacked sufficient distinctiveness to enable substantial TVR predictions. Upon processing the data with GLP, the classifier successfully achieved a satisfactory separation. Non-TVR cases gradually converged towards a single point, while TVR cases exhibited a more scattered pattern. These findings align with clinical observations, which suggest that stable patients exhibit less variation, whereas those with scattered observations face increased risks. The distinctiveness is not constrained by algorithmic mechanisms. However, it can be observed that the data were transformed from a hyperplane cluster distribution (where SVM demonstrated highest perfor-

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Translational Engineering in Health and Medicine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JTEHM.2023.3307794

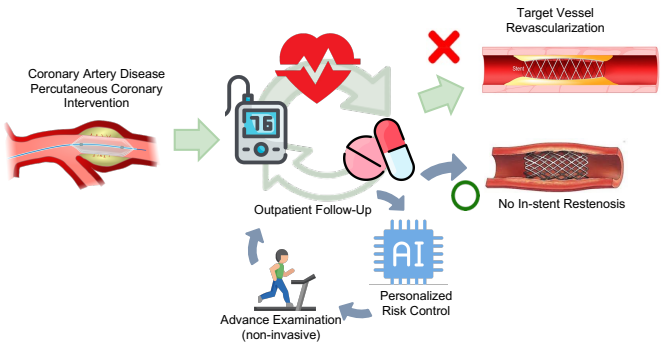

Fig. 7. The user scenario of GLP adoption. Advanced examinations indicates treadmill test, thallium scan, or coronary computerized tomography.

mance in the original distribution) to a more variable-specific representation (where LGBM outperformed other methods) and facilitating the creation of tree-based rules. Due to the limited size of the dataset, non-neural network algorithms were employed for TVR prediction, as deep neural networks did not perform well in this scenario.

The user scenario of GLP is depicted in Fig. 7. Following PCI treatment, patients are required to undergo follow-up visits at the outpatient setting on a 3-month basis. Clinicians assess patient risk to prevent TVR occurrence. GLP identify patients at a higher risk, prompting the initiation of more advanced examinations, such as treadmill tests, thallium scans, or coronary computerized tomography (CT). Conversely, patients with a relatively lower restenosis risk can also be separated, serving as a screening tool for more precise event detection and examination resource allocation. In general, GLP optimizes treatment strategies and improves patient prognosis while maintaining simplicity and user-friendliness without relying on accumulating comprehensive TVR predictors.

To broader the applications to other disease, an effective approach for identifying the appropriate pretext and downstream tasks becomes indispensable. Our findings suggested that through leveraging the intercorrelations among diseases, such as comorbidity or shared risk factors, represents a more proficient approach for selecting pretext and downstream tasks. The intercorrelations imply underlying similarities and suggest the possibility of translating disease progression between different patient groups. This approach offers a more precise alternative to empirical trial-and-error methods.

This study is subjected to certain limitations: While other approaches, such as GAN, have demonstrated effectiveness in recovering the distribution of absent observations, we opted for interpolation as it is considered a simpler and less computationally expensive method that still achieves excellent transfer performances. However, further exploration is required to determine the necessity or indispensable advantages of integrating a more advanced distribution generator; Our work is constrained by the unavailability of genuine patient observations at the target event, which prevented us from precisely reversing the exact output. Additionally, it is important to acknowledge that, at present, the output of GLP corresponds to a representation that cannot be straightforwardly reconstructed as real-world values through simple inverse-normalization. To achieve accurate inverse mapping, training an additional network or a decoder might be necessary to effectively map the latent output to real-world values; Furthermore, it is worth noting that this study exclusively focused on the analysis of numeric laboratory results and its applicability is limited to other numeric examinations. To make GLP a more comprehensive laboratory information extractor, it is necessary to incorporate different types of laboratory analyses, including categorical variables.

SSL also inherits certain limitations, such as the need for a larger amount of training data, more training iterations, and computational expenses compared to supervised training [15]. Additionally, deep neural networks employ a substantial number of trainable parameters and layers, making the model challenging to interpret [64, 19]. Thereby, SSL sacrificed interpretability for accuracy. This consideration should be taken seriously when utilizing the technology, as previous research has indicated that the level of trust in AI systems impacts the outcome of system utilization [65, 66]. Apart from interpretability, adopting AI system in to clinical workflow involved in multidisciplinary integration, encompassing areas such as the user interface design (research of Human-Computer Interactions) [67, 65], and factors that influence the acceptance of technology [66]. Ensuring system usability and reliability while resolving the disparities between proof-of-concepts and real-life environments is crucial for the successful adoption of AI systems into daily practice. This endeavor necessitates further collaboration to facilitate the reform of healthcare.

V. Conclusion

Our research successfully translate the progression trends of cardiovascular laboratory parameters between patient groups by capitalizing on the advantages of SSL and pretrained models. This discovery paves the road for wider data-driven applications in healthcare, and also functions as a screening tool for more precise event detection and judicious allocation of examination resources. Additionally, our study suggests that patients can benefits from diminished frequency of visits and onerous examinations by implementing sustainable estimation. To accomplish healthcare reform through the utilization of AI systems, the key to success and optimal system utilization lies in the multifaceted aspects involved multidisciplinary collaborations.

Acknowledgment

This work was supported by the National Science and Technology Council, Taiwan, NSTC 111-2628-E-A49-026-MY3, NSTC 112-2321-B-075-002- and Taipei Veterans General Hospital under Grant V1118-031.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Translational Engineering in Health and Medicine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JTEHM.2023.3307794

APPENDIX A
ONSET DEFINITION FOR PATIENTS WITH HYPERTENSION AND DIABETES MELLITUS.

HTN ICD code	
ICD-9-CM	codes starting with 401, 402, 403, 404, 405
ICD-10	codes starting with I10, I11, I12, I13, I15, I16
HTN Drug	
	Candesartan, Bisoprolol, Captopril, Enalapril, Amlodipine, Enalapril, Losartan Potassium, Carvedilol, Diltiazem HCL, Nifedipine, Felodipine, Lercanidipine, Nicardipine HCL, Nimodipine, Sotalol HCL, Verapamil HCL, Labetalol HCL, Fosinopril Sodium, Atenolol, Valsartan, Irbesartan, Metoprolol, Clonidine, Hydralazine HCL, Minoxidil, Benzyl Hydrochlorothiazide, Indapamide, Spironolactone, Pindolol, Nebivolol, Doxazosin SR, Eplerenone, Nicardipine, Ramipril, Imidapril HCL, Telmisartan, Olmesartan, Azilsartan Medoxomil.
DM ICD code	
ICD-9-CM	codes starting with 250
ICD-10	codes starting with E08, E09, E10, E11, E13
DM Drug	
	Metformin

HTN: hypertension; DM: diabetes mellitus; ICD: International Classification of Diseases (ICD); CM: clinical modification.

APPENDIX B
DISCRETE VALUE THRESHOLDS OF LABORATORY PARAMETERS

Item	Encoding	Value range
Chol/HDL-c	0	≤ 5
	1	> 5
LDL-c	0	≤ 160
	1	> 160
LDL-c/HDL-c	0	≤ 3.5
	1	> 3.5
Glucose AC	0	≤ 100
	1	$100 <$ value ≤ 125
	2	< 125
WBC	0	< 4
	1	$4 \leq$ value < 9
	2	≥ 9
UA	0	≤ 3.4
	1	$3.4 <$ value ≤ 7
	2	< 7

Chol/HDL-c: ratio of total cholesterol and high-density lipoprotein cholesterol; LDL-c: low-density lipoprotein cholesterol; LDL-c/HDL-c: ratio of low-density and high-density lipoprotein cholesterol; Glucose AC: fasting blood sugar; WBC: white blood cells; UA: Uric Acid.

APPENDIX C
R^2 VALUES OF EACH LABORATORY PARAMETERS BASED ON LINEAR AND PCHIP INTERPOLATION IN TWO-STAGE TRAINING

c	Linear			PCHIP			Barycentric		
	mean	v_{CI}	v_{2stage}	mean	v_{CI}	v_{2stage}	mean	v_{CI}	v_{2stage}
	Chol/HDL-c								
0	0.539	0.012	0.397	0.546	0.025	0.345	0.427	0.007	−0.238
1	0.565	0.024	0.368	0.552	0.058	0.343	0.486	0.018	−0.177
2	0.570	0.024	0.326	0.564	0.021	0.260	0.527	0.074	−0.686
3	0.586	0.043	0.348	0.575	0.054	0.309	0.490	0.025	−0.302
4	0.486	0.119	0.258	0.377	0.077	0.231	0.519	0.109	−7.851
5	**0.602**	**0.057**	**0.281**	**0.586**	**0.039**	**0.228**	0.295	0.094	−5.464
	LDL-c								
0	0.326	0.022	0.613	0.329	0.025	0.562	0.247	0.040	−0.130
1	0.348	0.034	0.598	0.338	0.018	0.568	0.257	0.040	0.182
2	0.326	0.221	0.618	0.344	0.019	0.472	0.299	0.018	−0.240
3	**0.376**	**0.011**	**0.568**	**0.341**	**0.033**	**0.573**	0.284	0.051	−0.034
4	0.207	0.349	0.272	0.213	0.119	0.382	−0.475	0.251	−6.924
5	0.234	0.095	0.402	0.311	0.070	0.526	0.157	0.164	−4.875
	LDL-c/HDL-c								
0	0.479	0.028	0.466	0.481	0.029	0.426	0.374	0.019	−0.070
1	0.505	0.074	0.446	0.503	0.033	0.407	0.417	0.014	0.050
2	0.509	0.043	0.399	**0.521**	**0.019**	**0.313**	0.444	0.020	−0.271
3	**0.522**	**0.053**	**0.414**	0.496	0.029	0.382	0.454	0.032	−0.026
4	0.488	0.413	−0.077	0.328	0.167	0.124	0.321	0.502	−8.539
5	0.297	0.122	0.433	0.513	0.022	0.272	0.290	0.106	−3.942
	Glucose AC								
0	0.480	0.040	0.426	0.480	0.053	0.392	0.360	0.014	−1.987
1	0.495	0.034	0.373	0.483	0.056	0.330	0.387	0.041	−1.799
2	0.546	0.096	0.262	0.508	0.090	0.268	0.451	0.085	−2.099
3	0.500	0.031	0.361	0.487	0.083	0.306	0.403	0.032	−2.496
4	**0.666**	**0.319**	**0.075**	**0.602**	**0.343**	**0.184**	0.516	0.121	−4.839
5	0.544	0.229	0.203	0.577	0.252	0.167	0.533	0.280	−3.504
	WBC								
0	0.664	0.024	0.245	0.657	0.022	0.226	0.564	0.027	−8.366
1	0.712	0.037	0.170	0.708	0.011	0.135	0.617	0.018	−3.608
2	0.744	0.022	0.106	0.758	0.022	0.060	0.701	0.013	−2.309
3	0.743	0.018	0.117	0.721	0.041	0.097	0.670	0.019	−3.949
4	**0.774**	**0.019**	**0.058**	0.761	0.012	0.051	0.716	0.026	−7.218
5	0.763	0.020	0.085	**0.779**	**0.016**	**0.010**	0.734	0.036	−6.471
	UA								
0	0.311	0.029	0.581	0.310	0.047	0.519	0.254	0.053	−1.894
1	0.333	0.038	0.539	0.325	0.047	0.472	0.256	0.079	−1.313
2	0.339	0.119	0.479	0.337	0.065	0.431	0.269	0.085	−1.872
3	0.355	0.040	0.521	0.328	0.033	0.415	0.257	0.058	−2.296
4	**0.454**	**0.179**	**0.290**	0.254	0.303	0.170	0.415	0.260	−3.265
5	0.272	0.231	0.235	**0.373**	**0.089**	**0.433**	0.283	0.149	−2.886
Avg.	**0.491**	0.091	0.340	0.482	**0.068**	**0.316**	0.394	0.083	−2.825

v_{CI}: boundary range of the 95% confidence interval; v_{2stage}: prediction variation between stage 1 and 2; c: $certain$. The bold figures are the chosen best performed settings.

REFERENCES

[1] N. Sattar, J. M. Gill, and W. Alazawi, "Improving prevention strategies for cardiometabolic disease," *Nature medicine*, vol. 26, no. 3, pp. 320–325, 2020.

[2] S. B. Hulley, S. R. Cummings, T. B. Newman, W. Browner, and D. Grady, "Designing cross-sectional and cohort studies," *Designing clinical research*, vol. 4, pp. 85–96, 2013.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Translational Engineering in Health and Medicine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JTEHM.2023.3307794

[3] J. Zhang, J. Gong, and L. Barnes, "HCNN: Heterogeneous convolutional neural networks for comorbid risk prediction with electronic health records," in *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pp. 214–221, 2017.

[4] Y. Cheng, F. Wang, P. Zhang, and J. Hu, "Risk prediction with electronic health records: A deep learning approach," in *Proceedings of the 2016 SIAM International Conference on Data Mining (SDM)*, pp. 432–440.

[5] S. Tang, P. Davarmanesh, Y. Song, D. Koutra, M. W. Sjoding, and J. Wiens, "Democratizing ehr analyses with fiddle: a flexible data-driven preprocessing pipeline for structured clinical data," *Journal of the American Medical Informatics Association*, vol. 27, no. 12, pp. 1921–1934, 2020.

[6] S. Wang, M. B. McDermott, G. Chauhan, M. Ghassemi, M. C. Hughes, and T. Naumann, "Mimic-extract: A data extraction, preprocessing, and representation pipeline for MIMIC-III," in *Proceedings of the ACM conference on health, inference, and learning*, pp. 222–235, 2020.

[7] Y. Si, J. Du, Z. Li, X. Jiang, T. Miller, F. Wang, W. Jim Zheng, and K. Roberts, "Deep representation learning of patient data from electronic health records (EHR): A systematic review," *Journal of Biomedical Informatics*, vol. 115, p. 103671, 2021.

[8] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Briefings in bioinformatics*, vol. 19, no. 6, pp. 1236–1246, 2018.

[9] A. Zaitcev, M. R. Eissa, Z. Hui, T. Good, J. Elliott, and M. Benaissa, "A deep neural network application for improved prediction of HbA_{1c} in type 1 diabetes," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 2932–2941, 2020.

[10] H. H. Mao, "A survey on self-supervised pre-training for sequential transfer learning in neural networks," 2020.

[11] M. Ragab, E. Eldele, Z. Chen, M. Wu, C.-K. Kwoh, and X. Li, "Self-supervised autoregressive domain adaptation for time series data," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–11, 2022.

[12] C. J. Reed, X. Yue, A. Nrusimha, S. Ebrahimi, V. Vijaykumar, R. Mao, B. Li, S. Zhang, D. Guillory, S. Metzger, K. Keutzer, and T. Darrell, "Self-supervised pretraining improves self-supervised pretraining," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 2584–2594, January 2022.

[13] X. Liu, F. Zhang, Z. Hou, L. Mian, Z. Wang, J. Zhang, and J. Tang, "Self-supervised learning: Generative or contrastive," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 1, pp. 857–876, 2023.

[14] J. Yoon, Y. Zhang, J. Jordon, and M. van der Schaar, "Vime: Extending the success of self- and semi-supervised learning to tabular domain," in *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, eds.), vol. 33, pp. 11033–11043, Curran Associates, Inc., 2020.

[15] D. Spathis, I. Perez-Pozuelo, L. Marques-Fernandez, and C. Mascolo, "Breaking away from labels: The promise of self-supervised machine learning in intelligent health," *Patterns*, vol. 3, no. 2, p. 100410, 2022.

[16] D. Kiyasseh, T. Zhu, and D. A. Clifton, "Clocs: Contrastive learning of cardiac signals across space, time, and patients," in *Proceedings of the 38th International Conference on Machine Learning* (M. Meila and T. Zhang, eds.), vol. 139 of *Proceedings of Machine Learning Research*, pp. 5606–5615, PMLR, 18–24 Jul 2021.

[17] R. Krishnan, P. Rajpurkar, and E. J. Topol, "Self-supervised learning in medicine and healthcare," *Nature Biomedical Engineering*, vol. 6, no. 12, pp. 1346–1352, 2022.

[18] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*, pp. 1597–1607, PMLR, 2020.

[19] L. Simone and D. Bacciu, "ECGAN: Self-supervised generative adversarial network for electrocardiography," in *Artificial Intelligence in Medicine*, pp. 276–280, Springer Nature Switzerland, 2023.

[20] P. Lee, S. Bubeck, and J. Petro, "Benefits, limits, and risks of gpt-4 as an ai chatbot for medicine," *New England Journal of Medicine*, vol. 388, no. 13, pp. 1233–1239, 2023. PMID: 36988602.

[21] X. Zhang, Z. Zhao, T. Tsiligkaridis, and M. Zitnik, "Self-supervised contrastive pre-training for time series via time-frequency consistency," in *Advances in Neural Information Processing Systems* (S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds.), vol. 35, pp. 3988–4003, Curran Associates, Inc., 2022.

[22] K. Wickstrøm, M. Kampffmeyer, K. Øyvind Mikalsen, and R. Jenssen, "Mixing up contrastive learning: Self-supervised representation learning for time series," *Pattern Recognition Letters*, vol. 155, pp. 54–61, 2022.

[23] J. Ouyang, D. Mao, Z. Guo, S. Liu, D. Xu, and W. Wang, "Contrastive self-supervised learning for diabetic retinopathy early detection," *Medical & Biological Engineering & Computing*, pp. 1–12, 2023.

[24] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big data*, vol. 3, no. 1, pp. 1–40, 2016.

[25] C. I. Tang, I. Perez-Pozuelo, D. Spathis, S. Brage, N. Wareham, and C. Mascolo, "SelfHAR: Improving human activity recognition through self-training with unlabeled data," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 5, no. 1, pp. 1–30, 2021.

[26] D. Spathis, I. Perez-Pozuelo, S. Brage, N. J. Wareham, and C. Mascolo, "Self-supervised transfer learning of physiological representations from free-living wearable data," in *Proceedings of the Conference on Health, Inference, and Learning*, CHIL '21, (New York, NY, USA), p. 69–78, Association for Computing Machinery, 2021.

[27] S. Tipirneni and C. K. Reddy, "Self-supervised transformer for sparse and irregularly sampled multivariate clinical time-series," *ACM Trans. Knowl. Discov. Data*, vol. 16, jul 2022.

[28] A. Chowdhury, J. Rosenthal, J. Waring, and R. Umeton, "Applying self-supervised learning to medicine: Review of the state of the art and medical implementations," *Informatics*, vol. 8, no. 3, 2021.

[29] Y. Liu, K. Gadepalli, M. Norouzi, G. E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P. Q. Nelson, G. S. Corrado, J. D. Hipp, L. Peng, and M. C. Stumpe, "Detecting cancer metastases on gigapixel pathology images," 2017.

[30] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang, "Biobert: a pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, 2020.

[31] Y. Gu, R. Tinn, H. Cheng, M. Lucas, N. Usuyama, X. Liu, T. Naumann, J. Gao, and H. Poon, "Domain-specific language model pretraining for biomedical natural language processing," *ACM Trans. Comput. Healthcare*, vol. 3, oct 2021.

[32] E. Braunwald, "Shattuck lecture—cardiovascular medicine at the turn of the millennium," *N Engl J Med*, vol. 337, pp. 1360–1369, 1997.

[33] V. Fuster, "Lewis a. conner memorial lecture. mechanisms leading to myocardial infarction: insights from studies of vascular biology.," *Circulation*, vol. 90, no. 4, pp. 2126–2146, 1994.

[34] G. W. Stone, "Percutaneous coronary intervention of chronic total occlusions: Conquering the final frontier," 2018.

[35] G. Giustino, A. Colombo, A. Camaj, K. Yasumura, R. Mehran, G. W. Stone, A. Kini, and S. K. Sharma, "Coronary in-stent restenosis: Jacc state-of-the-art review," *Journal of the American College of Cardiology*, vol. 80, no. 4, pp. 348–372, 2022.

[36] P. W. Serruys, M. J. Kutryk, and A. T. Ong, "Coronary-artery stents," *New England Journal of Medicine*, vol. 354, no. 5, pp. 483–495, 2006.

[37] G. D. Dangas, B. E. Claessen, A. Caixeta, E. A. Sanidas, G. S. Mintz, and R. Mehran, "In-stent restenosis in the drug-eluting stent era," *Journal of the American College of Cardiology*, vol. 56, no. 23, pp. 1897–1907, 2010.

[38] M. Nakano, F. Otsuka, K. Yahagi, K. Sakakura, R. Kutys, E. R. Ladich, A. V. Finn, F. D. Kolodgie, and R. Virmani, "Human autopsy study of drug-eluting stents restenosis: histomorphological predictors and neointimal characteristics," *European heart journal*, vol. 34, no. 42, pp. 3304–3313, 2013.

[39] J. M. Stolker, D. J. Cohen, K. F. Kennedy, M. J. Pencina, J. B. Lindsey, L. Mauri, D. E. Cutlip, and N. S. Kleiman, "Repeat revascularization after contemporary percutaneous coronary intervention: an evaluation of staged, target lesion, and other unplanned revascularization procedures during the first year," *Circulation: Cardiovascular Interventions*, vol. 5, no. 6, pp. 772–782, 2012.

[40] G. W. Stone, A. Maehara, A. J. Lansky, B. De Bruyne, E. Cristea, G. S. Mintz, R. Mehran, J. McPherson, N. Farhat, S. P. Marso, *et al.*, "A prospective natural-history study of coronary atherosclerosis," *New England Journal of Medicine*, vol. 364, no. 3, pp. 226–235, 2011.

[41] D. E. Cutlip, S. Windecker, R. Mehran, A. Boam, D. J. Cohen, G.-A. van Es, P. Gabriel Steg, M.-a. Morel, L. Mauri, P. Vranckx, *et al.*, "Clinical end points in coronary stent trials: a case for standardized definitions," *Circulation*, vol. 115, no. 17, pp. 2344–2351, 2007.

[42] X. Sun and T. Du, "Trends in cardiovascular risk factors among us men and women with and without diabetes, 1988–2014," *BMC Public Health*, vol. 17, no. 1, pp. 1–13, 2017.

[43] P. M. Ridker, B. M. Everett, T. Thuren, J. G. MacFadyen, W. H. Chang, C. Ballantyne, F. Fonseca, J. Nicolau, W. Koenig, S. D. Anker, *et al.*, "Antiinflammatory therapy with canakinumab for atherosclerotic disease," *New England journal of medicine*, vol. 377, no. 12, pp. 1119–1131, 2017.

[44] C. Kim, B.-K. Kim, S.-Y. Lee, D.-H. Shin, J.-S. Kim, Y.-G. Ko, D. Choi, Y. Jang, and M.-K. Hong, "Incidence, clinical presentation, and predic-

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Translational Engineering in Health and Medicine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JTEHM.2023.3307794

tors of early neoatherosclerosis after drug-eluting stent implantation," *American heart journal*, vol. 170, no. 3, pp. 591–597, 2015.

[45] A. Kastrati, A. Dibra, J. Mehilli, S. Mayer, S. Pinieck, J. Pache, J. Dirschinger, and A. Schömig, "Predictive factors of restenosis after coronary implantation of sirolimus-or paclitaxel-eluting stents," *Circulation*, vol. 113, no. 19, pp. 2293–2300, 2006.

[46] S. Cassese, R. A. Byrne, T. Tada, S. Pinieck, M. Joner, T. Ibrahim, L. A. King, M. Fusaro, K.-L. Laugwitz, and A. Kastrati, "Incidence and predictors of restenosis after coronary stenting in 10 004 patients with surveillance angiography," *Heart*, vol. 100, no. 2, pp. 153–159, 2014.

[47] R. W. Yeh, S.-L. T. Normand, R. E. Wolf, P. G. Jones, K. K. Ho, D. J. Cohen, D. E. Cutlip, L. Mauri, A. D. Kugelmass, A. P. Amin, *et al.*, "Predicting the restenosis benefit of drug-eluting versus bare metal stents in percutaneous coronary intervention," *Circulation*, vol. 124, no. 14, pp. 1557–1564, 2011.

[48] J. M. Stolker, D. J. Cohen, K. F. Kennedy, M. J. Pencina, S. V. Arnold, N. S. Kleiman, J. A. Spertus, E. Investigators, *et al.*, "Combining clinical and angiographic variables for estimating risk of target lesion revascularization after drug eluting stent placement," *Cardiovascular Revascularization Medicine*, vol. 18, no. 3, pp. 169–176, 2017.

[49] D. J. Gordon and B. M. Rifkind, "High-density lipoprotein—the clinical implications of recent studies," *New England Journal of Medicine*, vol. 321, no. 19, pp. 1311–1316, 1989.

[50] J. B. Zalokar, J. L. Richard, and J. R. Claude, "Leukocyte count, smoking, and myocardial infarction," *New England Journal of Medicine*, vol. 304, no. 8, pp. 465–468, 1981.

[51] C. Hage, A. Norhammar, L. Grip, K. Malmberg, N. Sarkar, B. Svane, and L. Rydén, "Glycaemic control and restenosis after percutaneous coronary interventions in patients with diabetes mellitus: a report from the insulin diabetes angioplasty study," *Diabetes and Vascular Disease Research*, vol. 6, no. 2, pp. 71–79, 2009.

[52] C. Mozzini, D. Girelli, A. Setti, J. Croce, F. Stefanoni, A. Castagna, F. Pizzolo, S. Friso, O. Olivieri, and N. Martinelli, "Serum uric acid levels, but not rs7442295 polymorphism of scl2a9 gene, predict mortality in clinically stable coronary artery disease," *Current Problems in Cardiology*, vol. 46, no. 5, p. 100798, 2021.

[53] E. Expert Panel on Detection *et al.*, "Executive summary of the third report of the national cholesterol education program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (adult treatment panel III)," *JAMA*, vol. 285, no. 19, pp. 2486–2497, 2001.

[54] "scipy.interpolate.PchipInterpolator — SciPy v1.9.1 manual." https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.PchipInterpolator.html. Accessed date: 2022-09-19.

[55] J.-P. Berrut and L. N. Trefethen, "Barycentric lagrange interpolation," *SIAM review*, vol. 46, no. 3, pp. 501–517, 2004.

[56] K. Hormann, "Barycentric interpolation," in *Approximation Theory XIV: San Antonio 2013*, pp. 197–218, Springer, 2014.

[57] D. E. Cutlip, A. G. Chhabra, D. S. Baim, M. S. Chauhan, S. Marulkar, J. Massaro, A. Bakhai, D. J. Cohen, R. E. Kuntz, and K. K. Ho, "Beyond restenosis: five-year clinical outcomes from second-generation coronary stent trials," *Circulation*, vol. 110, no. 10, pp. 1226–1230, 2004.

[58] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "The performance of lstm and bilstm in forecasting time series," in *2019 IEEE International Conference on Big Data (Big Data)*, pp. 3285–3292, 2019.

[59] "numpy.log1p - NumPy v1.23 manual." https://numpy.org/doc/stable/reference/generated/numpy.log1p.html. Accessed date: 2022-09-19.

[60] S. Kaufman, S. Rosset, C. Perlich, and O. Stitelman, "Leakage in data mining: Formulation, detection, and avoidance," *ACM Trans. Knowl. Discov. Data*, vol. 6, dec 2012.

[61] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, vol. 14, pp. 1137–1145, Montreal, Canada, 1995.

[62] S.-C. Shao, Y.-Y. Chan, Y.-H. Kao Yang, S.-J. Lin, M.-J. Hung, R.-N. Chien, C.-C. Lai, and E. C.-C. Lai, "The Chang Gung Research Database—a multi-institutional electronic medical records database for real-world epidemiological studies in Taiwan," *Pharmacoepidemiology and drug safety*, vol. 28, no. 5, pp. 593–600, 2019.

[63] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, and L. Bottou, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion.," *Journal of machine learning research*, vol. 11, no. 12, 2010.

[64] S. Nazir, D. M. Dickson, and M. U. Akram, "Survey of explainable artificial intelligence techniques for biomedical imaging with deep neural networks," *Computers in Biology and Medicine*, vol. 156, p. 106668, 2023.

[65] F. M. Calisto, J. a. Fernandes, M. Morais, C. Santiago, J. a. M. Abrantes, N. Nunes, and J. C. Nascimento, "Assertiveness-based agent communication for a personalized medicine on medical imaging diagnosis," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, (New York, NY, USA), Association for Computing Machinery, 2023.

[66] F. M. Calisto, N. Nunes, and J. C. Nascimento, "Modeling adoption of intelligent agents in medical imaging," *International Journal of Human-Computer Studies*, vol. 168, p. 102922, 2022.

[67] F. M. Calisto, A. Ferreira, J. C. Nascimento, and D. Gonçalves, "Towards touch-based medical image diagnosis annotation," in *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, (New York, NY, USA), p. 390–395, Association for Computing Machinery, 2017.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/