Reinforcement learning for active modality selection during diagnosis

Gabriel Bernardino, Anders Jonsson, Filip Loncaric, Pablo-Miki Marti Castellote, Marta Sitges, Patrick Clarysse, Nicolas Duchateau

Reinforcement learning for active modality selection during diagnosis

Gabriel Bernardino[1][0000−0001−8741−2566], Anders Jonsson[2], Filip Loncaric[3], Pablo-Miki Martí Castellote[2], Marta Sitges[4], Patrick Clarysse[1], and Nicolas Duchateau[1,5]

[1] Univ Lyon, Université Claude Bernard Lyon 1, INSA-Lyon, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69621, Lyon, France
[2] DTIC, Universitat Pompeu Fabra, Barcelona, Spain
[3] Department of Cardiovascular Diseases, Univ. Hospital Centre Zagreb, Croatia
[4] Cardiovascular Institute, Hospital Clínic Barcelona, University of Barcelona, Institut de Investigació Biomedica Augist Pi i Sunyer (IDIBAPS). CIBERCV, Instituto de Salud Carlos III, Spain
[5] Institut Universitaire de France (IUF)

Abstract Diagnosis through imaging generally requires the combination of several modalities. Algorithms for data fusion allow merging information from different sources, mostly combining all images in a single step. In contrast, much less attention has been given to the incremental addition of new data descriptors, and the consideration of their costs (which can cover economic costs but also patient comfort and safety).

In this work, we formalise clinical diagnosis of a patient as a sequential process of decisions, each of these decisions being whether to take an additional acquisition, or, if there is enough information, to end the examination and produce a diagnosis. We formulate the goodness of a diagnosis process as a combination of the classification accuracy minus the cost of the acquired modalities. To obtain a policy, we apply reinforcement learning, which recommends the next modality to incorporate based on data acquired at previous stages and aiming at maximising the accuracy/cost trade-off. This policy therefore performs medical diagnosis and patient-wise feature selection simultaneously.

We demonstrate the relevance of this strategy on two binary classification datasets: a subset of a public heart disease database, including 531 instances with 11 scalar features, and a private echocardiographic dataset including signals from 5 standard image sequences used to assess cardiac function (2 speckle tracking, 2 flow Doppler and tissue Doppler), from 188 patients suffering hypertension, and 60 controls.

For each individual, our algorithm allows acquiring only the modalities relevant for the diagnosis, avoiding low-information acquisitions, which both resulted in higher stability of the chosen modalities and better classification performance under a limited budget.

Keywords: Computer aided diagnosis · Reinforcement learning · Active feature selection · Acquisition costs · Cardiac imaging

1 Introduction

Medical diagnosis is not based on a single image, but usually considers several sources of information, each with different costs and accuracy at detecting different phenomena. Given the limited amount of resources in many real-life situations, it is crucial to select the most appropriate acquisitions for each patient [6]. In clinical practice, decisions are based on guidelines and consensus recommendations [4], which are in turn based on qualitative analysis of current evidence by experts. While machine learning has shown great success for quantitative analysis and diagnosis in medical images [1,13], quantifying the appropriateness of acquisitions has been neglected, as data are often considered an immutable input of the algorithms.

Cost-aware feature selection has received substantial attention from the machine learning community: the simplest methods are based on heuristics that promote sparsity at a population level, such as L_1 regularisation [9]; or decision trees and forests that include features' costs in the split criteria, thus doing patient-specific feature selection [11]. Recent approaches are based on Markov Decision Processes (MDPs), a formalisation of a time discrete process involving decisions with uncertain outcome, [5,12]. These methods build a common space that integrates all information, treating non acquired data as missing/censored. The current state is defined as a point/probability distribution in that space, which is updated after the acquisition of a new modality. Finally, Reinforcement learning (RL) is used to discover a policy, which is the optimal modality acquisitions at each point of this space.

A downside of the previous methods is that they heavily involve sampling, for both the data imputation and RL, which can become prohibitive when the data are high dimensional objects. In addition, the handling of "missing" data assumes that not-yet-acquired data can be estimated from the present data. In [10], *Wang et al.* proposed a method that considered all possible combinations of the N features, and a single-step policy, based on cost-sensitive-learning decision trees, had to be learnt for each of these 2^N combinations. However, this approach is only tractable for a small number of features, and therefore more modern literature has focused on partially-observed data approaches.

In this work, we propose a modality- and cost-aware RL method that sequentially proposes new acquisitions until it has enough confidence to produce a diagnosis. This work uses RL to extend to multiple modalities and more complex scenarios a recent two-stage strategy that recommends when a complex modality is needed instead of a simpler one [2]. Our method is similar to the value iteration algorithm, and we use kernel methods to estimate the state-action values. To avoid sampling, we use a strategy similar to *Wang et al.* [10]. We demonstrate the relevance of our method on two clinical datasets: the publicly available Heart Disease dataset [3], involving scalar measurements, and a private echocardiography dataset of patients with arterial hypertension with disease-related changes in cardiac function (i.e diastolic and systolic function), involving temporal signals along the whole cardiac cycle (flow and tissue motion data from speckle tracking and Doppler Imaging).

2 Methodology

2.1 Markov decision process and reinforcement learning

An MDP is a mathematical framework of sequential decisions with uncertain effects. Formally, an MDP consists of a tuple $(\mathcal{S}, \mathcal{S}_{end}, \mathcal{A}, \mathcal{R}, \mathcal{T})$, where $\mathcal{S}$ is the possible state spaces, $\mathcal{S}_{end} \subset \mathcal{S}$ is a set of ending states, $\mathcal{A}$ is the discrete action space, $\mathcal{T} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ is the transition function, which gives the probability distribution over $\mathcal{S}$ of the next state, knowing the current state and the action taken. $\mathcal{R} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ is the random reward function, that depends on the current and next states, as well as the action taken. An episode, starting in a given state s_1, consists of a sequence of states and actions:

$$((s_1, a_1, r_1), (s_2, a_2, r_2) \dots (s_n, a_n, r_n)), \tag{1}$$

where s_i is not a final state for $i \in [0, n-1]$ and s_n is a final state. The total reward of this episode is $\sum_i r_i$. The transitions and r_i follow the previously stated probability distributions $\mathcal{T}$ and $\mathcal{R}$. As we will show in the next section, our MDP is episodial with the number of steps being lower than the number of available modalities, so the use of a discount factor γ is not required.

A policy $\pi : \mathcal{S} \to \mathcal{A}$ is a function that chooses which action to perform at each state. We would like to find the optimal policy π^* that maximises the expected total reward over a random episode if we take actions following this policy. RL is a set of techniques to estimate such policy when the $\mathcal{R}$ and $\mathcal{T}$ distributions are unknown, but samples can be obtained. Given a policy, we define the *value* of a state s as the expected total reward over all episodes starting in s.

2.2 Problem definition

We formalise our cost-sensitive diagnostic problem under the notation of an MDP. An episode corresponds to an examination of a single patient, where each step (action) is the acquisition of a certain modality. The state will contain information on the already acquired data. The total reward will depend on whether a correct diagnosis was reached, and the costs associated with the used modalities.

Each modality of the set of modalities M is identified by an index i, and its measurements are elements of $\mathbb{R}^{n_i}$, allowing vector-valued measurements of different dimensionalities. Therefore, to represent a combination of measurements p, we use the Cartesian product ($\times$) of each space corresponding to a single measurement: $\times_{i \in p} \mathbb{R}^{n_i}$. To avoid data imputation, the full state space $\mathcal{S}$ is defined as the disjoint union of all possible combinations of acquired modalities $p \subset M$:

$$\mathcal{S} = \dot{\bigcup}_{p \subset M} \bigtimes_{i \in p} \mathbb{R}^{n_i} \tag{2}$$

where $\dot{\bigcup}$ denotes the disjoint union. $\mathcal{S}$ therefore consists of a connected domain for each element i of the powerset $p \subset M$. We will call each of these domains a

"superstate". By allowing multidimensional measurements, we can group modalities, so that the policy is forced to either ignore or acquire them together, thus reducing the number of superstates.

The set of actions contains one action a for each modality, associated with acquiring the a-th modality, and a special action to finish the episode and produce a diagnosis. Transitions associated which each "acquire" action are to move to a new state, which includes the previously acquired data, and the newly acquired measurement of the modality a. Therefore, the next state s_{n+1} will belong to the superstate identified by $p \cup \{a\}$, where p is the superstate of the current state s_n. Note that this means that even if we cannot know the exact state resulting of applying an action, since the measurement values are unknown, the transitions between "superstates" are completely deterministic, since they only depend on which modalities were acquired, but not on the measurement that was observed.

If the new state is not a final state, the reward of each action is set to 0. For a final state s_n, belonging to the superstate p, the reward is set as follows:

$$\mathcal{R}(s_f) = \mathbb{1}(y = y_p^{pred}) - \lambda \sum_{i \in p} c_i \tag{3}$$

where $\mathbb{1}$ is the indicator function, thus reflecting the accuracy (y_p^{pred} being the label prediction of the instance at state p, y being the true label of the instance); c_i is the cost of the i-th modality and λ is the coefficient weighting the relative contribution of the cost and accuracy. The reward of acquiring a modality that has already been observed is set to $-\infty$ to discourage the algorithm to take it, therefore each action is taken at most once during an episode. The class predictions y_p^{pred} are computed statically, using a classifier learnt at each superstate p from the available modalities. In our case, we chose a Support Vector Machine with Gaussian kernel.

2.3 Policy optimisation

We use a variation of value iteration to estimate π^*, where we estimate the state-action value $Q(s, a)$ for each state s, defined as the expected value if an action a is taken. Our method is a model-free strategy (meaning that it does not estimate the state transition function $\mathcal{T}$). We search for a solution to the recursive Bellman optimality equations:

$$Q(s, a) = E[\mathcal{R}(s, a, s') + \max_{a'} Q(s', a')], \tag{4}$$

where E refers to the expectation over the next states s', a' being the next action. For discrete spaces, this Q-function can be exactly stored in a table. However, in a continuous setting, function approximation is needed. Typical choices are neural networks, but given our limited amount of training data, and the success of kernel methods in similar applications, we used kernel ridge regression.

In a general case, Eq.4 cannot be solved for Q directly, since it requires knowledge of $\mathcal{R}$ and $\mathcal{T}$. Therefore, value iteration repeatedly performs the following

updates for all state-action pairs until convergence of Q:

$$Q^{n+1}(s,a) \leftarrow E[r_t + \max_{a'} Q^n(s',a')] \tag{5}$$

where s' is the next state.

In our case, the particular structure of the state transitions produced by the actions allows a direct solution of Equation 5, as shown in [10]. Since our state space is disconnected, we can train an independent Q-function for each of the $2^{N_{meas}}$ superstates, noted as $Q_p(s,a)$. And, since we know which will be the next superstate of each action, Eq. 5 becomes:

$$Q_p^{n+1}(s,a) \leftarrow E[r_t + \max_{a'} Q^n_{p\cup\{a_t\}}(s',a')]. \tag{6}$$

Using the fact that subsets form a directed acyclic graph, we can visit all superstates in postorder, and train the Q_p functions on the transitions derived from the collected data. As when visiting the superstate p, all $Q_{p\cup\{a_t\}}$ have already been trained, direct optimisation is possible.

2.4 Code availability

An implementation of the method, and the code used for the experiments, are publicly available at https://github.com/creatis-myriad/featureSelectionRL.

3 Datasets

3.1 Heart UCI dataset

We first tested our methods on the public Heart Disease Data Set [3] available at the University of California at Irving repository, whose objective is to diagnose heart disease for patients admitted to intensive care. These data correspond to a multicentric study, and we used all the available data except the imaging information since it was only available in a single center. Data therefore consisted of 11 features and associated costs per modality, which were split in 4 different feature groups (clinical history, laboratory, vital constants and exercise testing). We removed the individuals with missing data, which would induce additional challenges out of the scope of this paper, leaving a total of 207 controls and 324 cases.

3.2 Hypertension dataset

We also evaluated our methodology on temporal signals quantifying the cardiac function on an hypertense population. Details on the recruitment and cohort, as well as on the signal pre-processing (delineation and temporal alignment) can be found in [7]. The study protocol was approved by an internal ethical committee. We used the signals corresponding to flow (Mitral and Aortic Flow Doppler) and myocardial motion/deformation Septal Tissue Doppler, Global Longitudinal

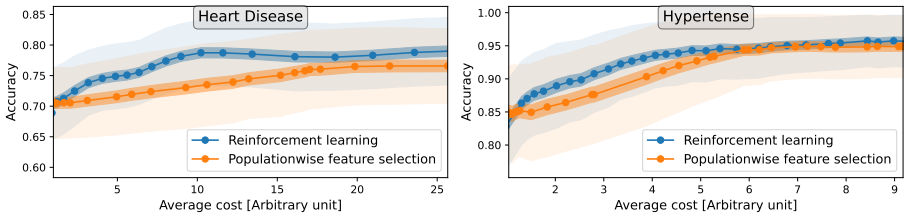

Figure 1: Average accuracy as a function of the average acquisition costs, where each point was generated by a specific value of the cost-coefficient λ. The solid line, light shadowed and darker shadowed depict the mean value, standard deviation, and 95% confidence interval, respectively.

Strain (GLS) and the local strain in the septal basal segment). These data were assigned arbitrary costs based on the frequency they are used in clinical practice (1,1, 2.5, 5, 10 respectively). Examples of the signals of three individuals can be found in Supplementary Material. Principal Component Analysis (PCA) was applied to the signals from each modality to reduce the dimensionality of these data before using them in the RL framework.

The interest of this dataset is twofold: the hierarchical representation provided by our method allows identifying different phenotypes, and simultaneously quantifying which modality is the most appropriate to detect each of these phenotypes. In addition to the cost-effective predictive power of our algorithm, we can interpret each decision of the policy by examining which biomarkers they capture, which can be seen as a data-based clinical guideline. This usage is very relevant when current human-generated clinical guidelines are long and complex.

4 Results

4.1 Prediction error against cost

For both datasets, we trained the policy for accuracy-cost coefficients λ (see definition in Eq.3) between 10^{-3} and 10^{-1}, and reported the mean accuracy and cost on a test set, which was a class-stratified split of the full dataset. We compared our algorithm with a classical feature selection using a validation set, in which we tested all possible combination and features and kept the one with a maximal validation reward. Results can be seen in Fig.1, where they were repeated over different train/test splits. We can observe that our method has a higher accuracy for a constrained budget, since RL allowed using expensive modalities only for a few individuals with a difficult diagnosis, while population-wise feature selection was forced to acquire the modality either for everybody or for nobody.

4.2 Stability under different training sets

We evaluated the consistency of the features selected by the algorithm, under different bootstrap samples of the training data. We separated a test subset

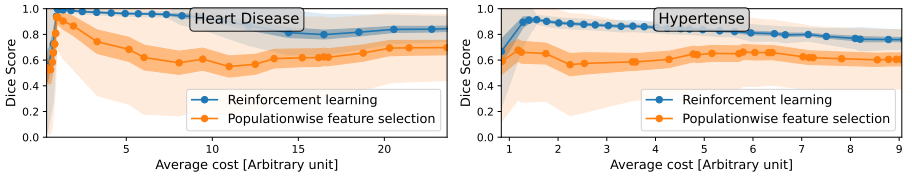

Figure 2: Stability of the policies to changes in the training set. Different policies resulting of bootstrap samples of a train dataset are evaluated on a fixed test-set, and the overlap of the sets of acquired modalities for each test patient on different bootstrap samples is quantified using the Dice score (y-axis). This experiment is repeated for several values of the λ weighting coefficient, resulting in different average costs (x-axis). The color code is equivalent to Fig.1.

from the dataset, and subsequently obtained several training samples of the remaining dataset, which were used to train the proposed RL model, and a classical feature selection using cross-validation. Then, we applied the trained models on the test set, and checked which were, for each subject, the selected features (ie. the recommended acquisitions) at their final state. This set was quantitatively compared to the features' set resulting from the other training bootstraps by computing the Dice coefficient. To improve statistical stability, the procedure was tested for different values of the accuracy-cost coefficient λ and test set splits. Figure 2 shows the results of this experiment, where our method is more robust than feature selection using validation: indeed, static feature selection forces that either all or none individuals acquire a modality, while our method allows a gradual process where only a few individuals acquire an expensive modality.

4.3 Policy interpretation on the Hypertense dataset

We further studied the learnt policy on the hypertense dataset. We trained the RL algorithm with $\lambda = 0.05$, chosen to guarantee a low number of acquired modalities. Afterwards, we evaluated the policy on all individuals from the training set, keeping track of the superstates they visited. The full decision graph can be seen in Fig.3. Figure 4 complements this by showing the representative signals associated to the individuals that were diagnosed at each superstate of the graph visited by more than 10 individuals, allowing us to examine and interpret the decisions of the algorithm.

The first acquisition recommended by RL was Mitral Doppler, which is consistent with clinical knowledge [8]. Hypertense individuals with a clear grade I diastolic dysfunction mitral inflow pattern (presenting E/A-wave fusion, A-peak larger than E-peak) or those clearly controls are differentiated based on the Mitral Doppler only (first row in Fig.4). The remaining patients had still an unclear diagnosis and were referred to either Aortic Doppler or GLS, depending on the findings in the Doppler.

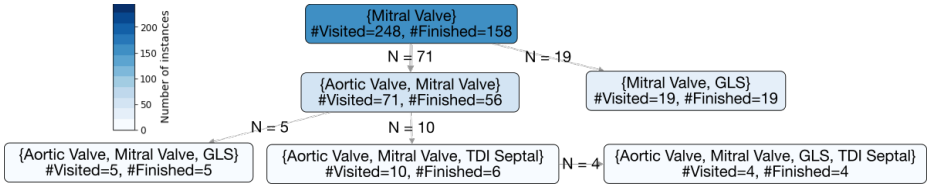

Figure 3: Graph showing the decision paths proposed by our algorithm on the full population. Each node represents a superstate (a set of modalities), and each edge represents an action (acquiring a new modality). The number of individuals that take the "finish" action is indicated at each node.

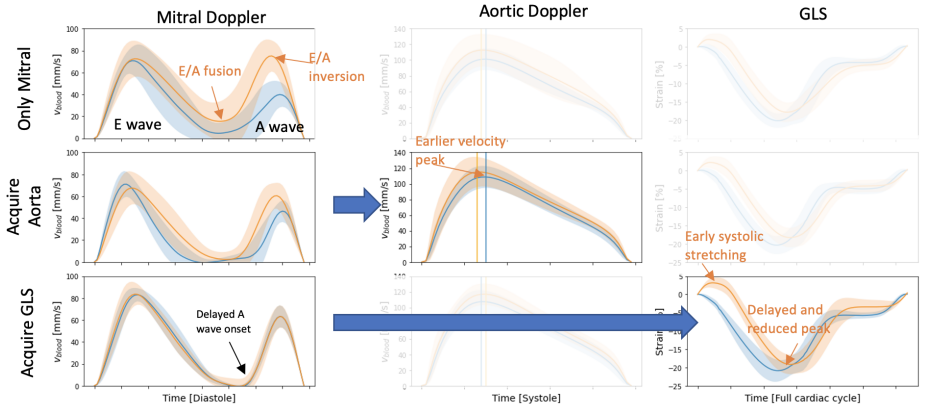

Figure 4: Class-wise mean and standard deviation (orange hypertension, blue control) for the Mitral, Aortic and GLS signals of the individuals whose diagnosis was done using Mitral only (1st row), Mitral and Aortic (2nd row) and Mitral and GLS (3rd row). Un-used modalities are displayed with low opacity.

The Mitral Doppler of patients for which RL suggested Aortic Doppler (2nd row) presented a slightly elevated A-wave, but still lower than the E-wave; lying on the boundary between controls and cases of the first subpopulation. Although the mitral inflow pattern is not suggestive of diastolic dysfunction, an earlier, higher velocity aortic flow peak reflected increased cardiac contractility in the setting of elevated afterload due to high blood pressure, identifying patients with altered systolic function in hypertension.

The last group (3rd row) consists of individuals for which GLS was recommended. Mitral Doppler showed a later onset of atrial contraction within the cardiac cycle. This is confirmed in the GLS curves, where hypertense subjects showed prolonged left ventricular stretching during the initial part of systole - identifying a disease-related pattern in the timing of cardiac events.

5 Conclusion

We presented an RL framework to obtain a cost-effective policy that sequentially proposes the best modality to acquire to produce a diagnosis. We thoroughly

evaluated it on a public dataset with scalar features and a private one with complex high-dimensional descriptors of hypertension. Compared to classical model selection using a validation set, our method improved the model performance at similar acquisition cost by making a more efficient use of expensive modalities. The proposed method also showed stability to changes in the training set.

In addition, we were able to interpret and explain the policy constructed in the diastolic dysfunction dataset, which captured physiological patterns consistent with current clinical knowledge. Our algorithm showed potential not only as a clinical decision support system for diagnosis, but also at a higher level to help clinicians to derive data-based guidelines.

On a broader perspective, the method is highly promising for assisting experts with the analysis of multiple descriptors in an efficient manner. It could lead to discovering cost-efficient policies in applications where high heterogeneity between individuals is expected, in particular for screening campaigns or in developing countries.

Acknowledgements. The authors acknowledge the partial support from the French ANR (LABEX PRIMES of Univ. Lyon [ANR-11-LABX-0063] and the JCJC project "MIC-MAC" [ANR-19-CE45-0005]) and the Spanish AEI [PID2019-108141GB-I00]. We thank Prof. B. Bijnens (IDIBAPS & ICREA, Barcelona, Spain) for fruitful discussions.

References

1. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Gonzalez Ballester, M.A., Sanroma, G., Napel, S., Petersen, S., Tziritas, G., Grinias, E., Khened, M., Kollerathu, V.A., Krishnamurthi, G., Rohe, M.M., Pennec, X., Sermesant, M., Isensee, F., Jager, P., Maier-Hein, K.H., Full, P.M., Wolf, I., Engelhardt, S., Baumgartner, C.F., Koch, L.M., Wolterink, J.M., Isgum, I., Jang, Y., Hong, Y., Patravali, J., Jain, S., Humbert, O., Jodoin, P.M.: Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? IEEE Transactions on Medical Imaging **37**(11), 2514–2525 (11 2018)
2. Bernardino, G., Clarysse, P., Sepúlveda-Martínez, A., Rodríguez-López, M., Prat-Gonzàlez, S., Sitges, M., Gratacós, E., Crispi, F., Duchateau, N.: Hierarchical Multi-modality Prediction Model to Assess Obesity-Related Remodelling. In: (STACOM - MICCAI workshop), LNCS. pp. 103–112 (2022)
3. Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J.J., Sandhu, S., Guppy, K.H., Lee, S., Froelicher, V.: International application of a new probability algorithm for the diagnosis of coronary artery disease. The American Journal of Cardiology **64**(5), 304–310 (8 1989)
4. Garbi, M., Edvardsen, T., Bax, J., Petersen, S.E., McDonagh, T., Filippatos, G., Lancellotti, P.: EACVI appropriateness criteria for the use of cardiovascular imaging in heart failure derived from European National Imaging Societies voting. European Heart Journal - Cardiovascular Imaging **17**(7), 711–721 (7 2016)
5. Gong, W., Tschiatschek, S., Nowozin, S., Turner, R.E., Hernández-Lobato, J.M., Zhang, C.: Icebreaker: Element-wise Efficient Information Acquisition with a Bayesian Deep Latent Gaussian Model. In: Wallach, H., Larochelle, H., Beygelzimer, A., Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 32. Curran Associates, Inc. (2019)
6. Hadian, M., Jabbari, A., Mazaheri, E., Norouzi, M.: What is the impact of clinical guidelines on imaging costs? Journal of Education and Health Promotion **10**, 10 (2021)
7. Loncaric, F., Marti Castellote, P.M., Sanchez-Martinez, S., Fabijanovic, D., Nunno, L., Mimbrero, M., Sanchis, L., Doltra, A., Montserrat, S., Cikes, M., Crispi, F., Piella, G., Sitges, M., Bijnens, B.: Automated Pattern Recognition in Whole-Cardiac Cycle Echocardiographic Data: Capturing Functional Phenotypes with Machine Learning. Journal of the American Society of Echocardiography **34**(11), 1170–1183 (11 2021)
8. Nagueh, S.F., Smiseth, O.A., Appleton, C.P., Byrd, B.F., Dokainish, H., Edvardsen, T., Flachskampf, F.A., Gillebert, T.C., Klein, A.L., Lancellotti, P., Marino, P., Oh, J.K., Popescu, B.A., Waggoner, A.D.: Recommendations for the Evaluation of Left Ventricular Diastolic Function by Echocardiography: An Update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. Journal of the American Society of Echocardiography : official publication of the American Society of Echocardiography **29**(4), 277–314 (4 2016)
9. Ng, A.Y.: Feature selection, L1 vs. L2 regularization, and rotational invariance. In: Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004 (2004)
10. Wang, J., Trapeznikov, K., Saligrama, V.: Efficient Learning by Directed Acyclic Graph For Resource Constrained Prediction. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 28. Curran Associates, Inc. (2015)

11. Xu, Z.E., Kusner, M.J., Weinberger, K.Q., Chen, M., Chapelle, O.: Classifier cascades and trees for minimizing feature evaluation cost. Journal of Machine Learning Research **15** (2014)
12. Yin, H., Li, Y., Pan, S.J., Zhang, C., Tschiatschek, S.: Reinforcement Learning with Efficient Active Feature Acquisition. arXiv (11 2020)
13. Zhou, T., Ruan, S., Canu, S.: A review: Deep learning for medical image segmentation using multi-modality fusion. Array **3-4**, 100004 (9 2019)