Computers in Biology and Medicine 158 (2023) 106841

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

ALEC: Active learning with ensemble of classifiers for clinical diagnosis of coronary artery disease

Fahime Khozeimeh [a], Roohallah Alizadehsani [a,*], Milad Shirani [b], Mehrzad Tartibi [c], Afshin Shoeibi [d,e], Hamid Alinejad-Rokny [f], Chandrashekhar Harlapur [g], Sayed Javed Sultanzadeh [h], Abbas Khosravi [a], Saeid Nahavandi [a,i], Ru-San Tan [j], U. Rajendra Acharya [k,l,m]

[a] Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Geelong, Australia
[b] Department of Mechanical Engineering, University of California, Berkeley, CA, 94720, USA
[c] Delbeat Inc., Berkeley, CA, USA
[d] Data Science and Computational Intelligence Institute, University of Granada, Spain
[e] Internship in BioMedical Machine Learning Laboratory, UNSW, Sydney, 2052, NSW, Australia
[f] BioMedical Machine Learning Lab (BML), The Graduate School of Biomedical Engineering, UNSW Sydney, Sydney, NSW, 2052, Australia
[g] Emergency Department, Epworth Hospital, Geelong, Australia
[h] Department of General Medicine, Gladstone Hospital, Gladstone, Australia
[i] Harvard Paulson School of Engineering and Applied Sciences, Harvard University, Allston, MA, 02134, USA
[j] Department of Cardiology, National Heart Centre Singapore, Singapore, Singapore
[k] School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, Australia
[l] Department of Biomedical Engineering, School of Science and Technology, Singapore University of Social Sciences, Singapore, Singapore
[m] Department of Bioinformatics and Medical Engineering, Asia University, Taichung, City, Taiwan

ARTICLE INFO

Keywords:
Active learning
Coronary artery disease
Classification
Data analysis
Ensemble method

ABSTRACT

Invasive angiography is the reference standard for coronary artery disease (CAD) diagnosis but is expensive and associated with certain risks. Machine learning (ML) using clinical and noninvasive imaging parameters can be used for CAD diagnosis to avoid the side effects and cost of angiography. However, ML methods require labeled samples for efficient training. The labeled data scarcity and high labeling costs can be mitigated by active learning. This is achieved through selective query of challenging samples for labeling. To the best of our knowledge, active learning has not been used for CAD diagnosis yet. An Active Learning with Ensemble of Classifiers (ALEC) method is proposed for CAD diagnosis, consisting of four classifiers. Three of these classifiers determine whether a patient's three main coronary arteries are stenotic or not. The fourth classifier predicts whether the patient has CAD or not. ALEC is first trained using labeled samples. For each unlabeled sample, if the outputs of the classifiers are consistent, the sample along with its predicted label is added to the pool of labeled samples. Inconsistent samples are manually labeled by medical experts before being added to the pool. The training is performed once more using the samples labeled so far. The interleaved phases of labeling and training are repeated until all samples are labeled. Compared with 19 other active learning algorithms, ALEC combined with a support vector machine classifier attained superior performance with 97.01% accuracy. Our method is justified mathematically as well. We also comprehensively analyze the CAD dataset used in this paper. As part of dataset analysis, features pairwise correlation is computed. The top 15 features contributing to CAD and stenosis of the three main coronary arteries are determined. The relationship between stenosis of the main arteries is presented using conditional probabilities. The effect of considering the number of stenotic arteries on sample discrimination is investigated. The discrimination power over dataset samples is visualized, assuming each of the three main coronary arteries as a sample label and considering the two remaining arteries as sample features.

* Corresponding author.
E-mail address: r.alizadehsani@deakin.edu.au (R. Alizadehsani).

https://doi.org/10.1016/j.compbiomed.2023.106841
Received 18 January 2023; Received in revised form 1 March 2023; Accepted 26 March 2023
Available online 31 March 2023

1. Introduction

There has been an exponential growth of data generated from digital devices and in diverse domains, including biomolecular research, commerce, social network, engineering sciences, and cybersecurity. By 2020, the generated data is estimated to be 35 trillion gigabytes [1]. It has become increasingly difficult for domain experts to extract useable information from large datasets and perform analysis for decision support systems. There is a need for systems that can automatically mine and discover novel knowledge from big data efficiently [2]. Traditionally, training data must be paired with class labels to solve learning problems. However, real-world data are often not labeled and require time and/or expertise for labeling to render them useable for learning [3]. In some applications, such as rating movies or flagging spam emails, labeled instances are accessible at low cost. Learning systems can use these labels to recommend movies or filter junk mail [4]. However, preparing labeled instances is usually costly and time-consuming. For example, consider classifying documents and media files, annotating knowledge domains, and labeling speech utterances. To circumvent labeling costs, new learning paradigms like active learning and semi-supervised learning have been proposed [5].

Active learning or query learning is a subfield of machine learning (ML) [6]. The difference between active and passive learning is depicted in Fig. 1. In contrast to passive learning, where the training data are provided in advance, and the learner makes no effort to obtain new data, the dataset is not provided in active learning, and the model is updated continually through querying and receiving new data [7]. As can be seen in Fig. 1, in active learning, the training process is begun with a limited set of labeled training samples. Throughout the training, the active learner may ask for the label of additional samples. The motivation is to reduce the number of labeled training samples needed for appropriate learning. This is achieved by asking for the label of the most informative training samples. There are three major approaches to active learning: pool-based, stream-based, and active learning with membership queries [8]. In the latter, the active learning system reduces labeling costs by querying unlabeled data that may need to be labeled by an oracle, such as a human annotator, without manually labeling the entire dataset. Active learning has been applied to diverse problems such as remote sensing, image classification [9], and multimedia annotation [7], where unlabeled samples are abundant, and their labeling cost is high.

Semi-supervised learning models reduce labeling costs by putting assumptions on unlabeled data based on attributes of labeled data [10] using techniques like graph-based methods, transductive support vector machines (SVMs), co-training, self-training, and expectation-maximization with generative mixture models. Discriminative semi-supervised models using graph-based methods and transductive SVM work best, provided that intraclass and interclass densities are high and low, respectively and classes do not overlap. Graph-based methods are preferred if two points or more with similar features are in the same class. Whenever features can be split into two sets, co-training is a good choice. In the case of well-clustered data, expectation-maximization with a generative mixture model is appropriate. Generative techniques are dependent on data distribution and can fail if inputs are not correlated with the classification task [11].

Several strategies can reduce annotation costs. Luo et al. [12] developed a hierarchical active learning framework where the entire population was divided into smaller subpopulations. The model was gradually refined through iterative learning from the subpopulations with their class labels. One of the common sample selection strategies in active learning is to select samples where the system is uncertain about their labels. Labeling these samples requires human feedback. Sharma et al. [13] showed that the cause of the uncertainty (either strong but conflicting evidence or insufficient evidence) could impact learning efficiency. They advocated an active learning approach that distinguished between the two causes of uncertainty.

Coronary artery disease (CAD) is a common cause of death worldwide. Invasive coronary angiography is the reference standard for diagnosing CAD but it is expensive and associated with risks. Many predictive models for CAD based on clinical and noninvasive imaging parameters have been developed from labeled datasets using data mining algorithms and ML but not (as far as we know) with active learning. For diagnosing CAD clinically, we propose an Active Learning with Ensemble of Classifiers (ALEC) method based on the interdependent outcomes of four classifiers that are predicated on the presence of CAD in the input sample as well as in each of its three branch coronary artery territories: left anterior descending (LAD), left circumflex (LCX), and right coronary artery (RCA). ALEC iteratively selects key unlabeled samples with uncertain (inconsistent) classification outcomes to be manually labeled by medical experts. The performance of the proposed approach is compared with existing active learning algorithms on the Z-Alizadeh Sani CAD dataset. Our original contributions are as follows.

- Active learning is used for CAD diagnosis by selective query of the label of samples for which the classifier outputs are inconsistent. Such samples are those that ALEC is uncertain about.
- The interdependent relationships between coronary artery stenosis per patient and vessel levels are represented as conditional probabilities.
- The soundness of our proposed method is theoretically justified through mathematical proof of posed lemmas.
- The correlation between pairs of features in the dataset is presented.
- The top 15 features contributing to CAD as well as stenosis of LAD, LCX, and RCA, are presented.
- The discrimination power over dataset samples is visualized, assuming each of the LAD, LCX, and RCA features as sample labels.
- The discrimination power over dataset samples is visualized assuming LAD/LCX/RCA as the label and {LCX, RCA}/{LAD, RCA}/{LAD, LCX} as sample features.

The rest of the paper is structured as follows. The literature review is presented in section 2. The dataset used in the experiments is introduced in section 3. The proposed method is explained in section 4. The experimental results and discussion are presented in section 5. Finally, section 6 is devoted to the conclusion and future works.

2. Active learning and automatic labeling applications

This section reviews active learning and automatic labeling

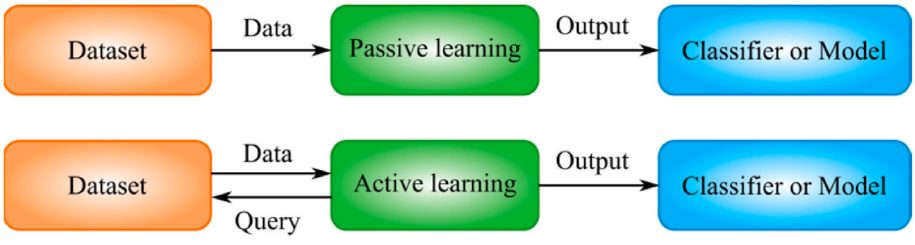


Fig. 1. General schema of active versus passive learning models.

applications in various fields, including cardiovascular diseases. At the end of the section, based on the reviewed related works, the research gap and contribution of our method are highlighted.

2.1. Science and medicine

Active learning of medical images [14–18], genes, mRNA, and protein classes [15] plays a crucial role in the design of decision support systems. Especially in medical applications, access to labeled training data is often limited. Data augmentation approaches have proved useful in increasing the training samples artificially. For example, generative adversarial networks (GAN) were exploited in combination with a novel mutation model for synthetic data generation [19]. These data were used for Ischemic Stroke Lesion Segmentation. The automatic generation of training data has been investigated as well [18]. To this end, a discriminative learning approach was proposed to automatically generate training data from fusion image datasets. The generated data were used to localize the transesophageal echography transducer in X-ray images with a 0.8 mm error. The fusion of image data from echo and X-ray fluoroscopy is useful in treatment of structural heart disease.

Given the promising results obtained by DL in medical diagnosis, it is critical to address the training data scarcity to achieve efficient training of DL models. By selecting the samples for labeling wisely, active learning can play a major role in achieving data efficiency. Examples of DL applications in the medical domain are breast cancer diagnosis [20], phasic dopamine release identification [21], and early diagnosis of Diabetes [22].

2.2. Multimedia, mobile, and security applications

Active learning has been applied to mobile app review labeling [23], video labeling [24], intrusion detection [25], television program review [26], cybersecurity [27], and security labeling [28]. By studying semi-structured interviews and app reviews, Sun et al. [23] identified eight psychological needs of mobile app users: self-actualization, cognitive, social, hedonic, health, security, low-cost, and utilitarian. They trained a classifier to filter app reviews that have words and phrases related to these eight psychological needs. A novel video labeling system VideoAL was proposed comprising XML rendering, classification, model learning, feature extraction, annotation, region, and shot segmentation [24]. VideoAL can generate MPEG-7 XML metadata from MPEG-1 sequence inputs.

2.3. Speech and text

The massive amount of speech and textual data is a rich source for designing, developing, and evaluating ML approaches. However, the labeling process of such massive data is tedious, which drives the need for automatic labeling and active learning. In the remainder of this section, some existing works on automatic labeling and active learning of speech and text data are reviewed.

2.3.1. Labeling of text

Automated topic labeling of speech and text has garnered growing research interest due to its application in search engines and document curation [29–33]. A probabilistic approach for automatic multinomial topic labeling of words in the text was proposed [31]. This method maximized the mutuality between labels, topic models, and Kullback-Leibler divergences between word distributions. Extending beyond word distributions, Allahyari et al. [32] designed a framework that incorporated multinomial distributions of ontological classifications to improve the automatic labeling of latent topics from text. With large document collections, a cluster approach to label closely related groups may prove more efficient than surveying individual documents. Motivated by the concept that words in text form a hierarchy of components with varying generality, Popescul et al. [34] developed two clustering techniques for the automatic labeling of document clusters. Damerau et al. [35] clustered unlabeled text data using nearest neighbor and/or running feature extraction methods and trained labeled data using supervised ML algorithms. Hacioglu et al. [36] presented an automatic time expression labeling model for English and Chinese texts. The model deployed deterministic, discriminative, token-by-token, and left-to-right classifiers for extraction of tags for each token, and yielded good performance for the one-versus-all multiclass SVM classification method.

2.3.2. Labeling of speech

Manual labeling of transcriptions of acoustic speech for training is time-consuming and costly, which is the motivation for active learning models. Li et al. [37] proposed grammar-based semi-supervised incremental learning for continuous speech recognition and labeling for large automatic vocabulary. The model initialization required sparse, manually labeled transcriptions. The model was then run to recognize greater numbers of unlabeled data. Clustering approaches can also be applied to acoustic models. Fu et al. [38] designed an unsupervised six-stage locally embedded clustering method for automatic labeling of high-dimensional acoustic speech data that attained efficient dimensionality reduction and data representation and improved clustering performance.

2.4. Active learning in cardiovascular disease diagnosis

Bizopoulos et al. [39] reviewed emerging research in deep learning cardiology applications based on signal and imaging data. Longstaff et al. [40] applied semi-supervised and active learning to the real-time classification of activity tracker signals on mobile devices, which typically used static classifiers. Among methods like democratic co-learning, en-co-learning, self-learning, and active learning, the former performed best. Active learning was also applied to task/patient-adaptive heartbeat classification of the cardiologist-benchmarked electrocardiogram (ECG) dataset with over 90% less patient-specific training data requirement compared to its rival methods [41]. Sun et al. [42] presented an automated -instance learning strategy called latent topic that applied SVM on the topic vectors of ECG. Without labeling heartbeats on the ECG data, the model successfully identified myocardial ischemia with better sensitivity and specificity than the state-of-the-art. Pasolli et al. [43] proposed hybridized SVM with query by committee, posterior probability, and margin sampling for ECG signal classification. These methods used active learning and were tested on simulated and MIT-BIH arrhythmia datasets with encouraging results.

Tissue segmentation is an important step in image processing that may be facilitated by active learning. An interactive active learning algorithm was able to generate accurate whole-heart segmentation models with reduced user inputs [44]. The method was seeded by a limited number of manually segmented pediatric short-axis cine cardiac magnetic resonance images. Chyzhyk et al. [45] studied stroke patients and reported good results for automatic tissue segmentation of multimodal anatomical, diffusion, and functional brain magnetic resonance images. They applied active learning selective sampling for training image data classifiers.

2.5. Research gap and contributions

To highlight the contribution of our proposed approach, the reviewed works are summarized in Table 1 and grouped according to their application domains. Inspecting Table 1 reveals that our proposed approach (ALEC) is the first to use disease-specific characteristics to its advantage. Careful analysis of the Z-Alizadeh Sani dataset revealed the correlation between stenosis of main coronary arteries and having CAD. To the best of our knowledge, such an approach has not been investigated before, which justifies the novelty of our approach. Moreover, we are the first to utilize active learning for implementing sample-efficient

Table 1
Summary of related works.

	Ref.	Year	Objective	Data	Approach
Multimedia	[20]	2003	Automatic video labeling	MPEG-1 video sequence	The labeling system is made of seven modules: shot segmentation, region segmentation, annotation, feature extraction, model learning (SVM), classification
	[38]	2007	Automatic labeling of high-dimensional data	Frey faces, AAI database	Unsupervised data clustering and automatic labeling
Cyber Security	[21]	2010	Classifying intrusion detection alerts in real-time	Alerts detected by Snort intrusion detection system	A two-phase approach: 1. Alert collection and clustering, 2. Learning rules for classifying alerts based on RIPPER [46] algorithm
	[23]	2013	Automatic text labeling to prepare a collection of annotated texts for cyber security applications	Text data obtained from multiple data sources	Domain-specific, structured data are exploited to train a Maximum Entropy Model [47,48]
	[24]	2015	Automatic determination of security labels for textual documents	Digital National Security Archive containing declassified US government documents publicly available	Textual content extraction using Abbyy [49] OCR service, documents represented by bag-of-words model [50], classification is done by SVM
Text processing	[22]	2013	Automatic filtering of Twitter messages containing the title of a TV program	Tweets containing names of specific TV programs	Building a list of unambiguous TV program names, collecting tweets containing selected program names, labeling gathered tweets, training a SVM for tweets automatic labeling
	[25]	2009	Topics automatic labeling according to an extracted hierarchy	Document and dictionary corpus	Topic hierarchy formed based on Google Drive service and OpenOffice English Thesaurus, Similarity between topics and the topic hierarchy evaluated using cosine/ overlap/ … similarity measures, topic extraction using Latent Dirichlet Allocation method [51]
	[26]	2012	Labeling hierarchical topics automatically	Documents gathered from Yahoo! Answers and Wikipedia	Exploiting parent-child and sibling relations between topics for assigning labels to topics, Ngram Testing [52] used for candidate label extraction, structural assisted label ranking: 1. Ranking by term weighting, 2. Ranking by Jensen-Shannon Divergence
	[32]	2015	Improved topic labeling using ontology concepts	British Academic Written English Corpus (BAWE), part of Reuters2 news articles	The proposed OntoLDA approach integrates ontological concepts with topic models in one framework, topic-concept relations exploited labeling accuracy improvement,
	[33]	2013	Automatic topic labeling to facilitate document collections analysis	BAWE, BBC corpus, StackExchange dataset	A topic graph is formed based on structured data provided by DBpedia,
	[34]	2000	Automatic labeling of clusters of documents by words	Computer science research papers obtained from the Cora search service	First method: X2 significance test employed to detect word usage in hierarchy categories. Second method: discriminative words with higher repetition frequency are chosen
	[35]	2004	Automatic text labeling for keyword search	Unlabeled text data	Members of a reference answer set are used as centroids for clustering unlabeled text samples, supervised ML method (e.g., a rule-based classifier) is trained on clustered (labeled) data
	[36]	2005	Automatic labeling of English and Chinese time expressions	ACE Temporal Expression Recognition and Normalization (TERN) corpus	Time expression labeling formulated as a tagging problem, 1-vs-all multi-class SVM is employed
Speech labeling	[37]	2012	Sample efficient automatic labeling of speech data	CCTV news multimedia label database	The small number of manually labeled data, Hidden Markov Model (HMM) used as a classifier, semi-supervised incremental learning
Medical	[14]	2014	Segmenting and recognizing cerebral vessels automatically	High-resolution 3D micro-CT images	The vascular structure is represented as a graph, each vessel segment is represented as an edge, and its local features (length, diameter, direction, etc.) are used for anatomical name assignment to it
	[15]	2001	Assigning protein, assignment of gene or mRNA class labels to biological terms in free text	Nine million words gathered from molecular biology journal articles, expressed in XML format	Naïve Bayes, decision tree, inductive rule learning [53]
	[16]	2007	Automated labeling of data for cancer diagnosis	Colonoscopy videos	Recording expert eye position during a colonoscopy video inspection
	[17]	2014	Automatic labeling of brain anatomies	Magnetic resonance images of the brain	Train multiple atlas forests [54], cluster atlas forests with similar performance, and train a new forest for each cluster
	[18]	2013	Domain adaptation: calculating the transform between *trans*-esophageal echography (TEE) and X-ray fluoroscopy imaging systems for structural heart disease treatment with minimal invasion	TEE and fluoroscopy images	Assigning weight to source domain sample, weights are based on target domain probability distribution to source domain probability distribution ratio
	[40]	2010	Automatic augmentation of user's activity classifier after deployment on mobile systems	GPS speed, acceleration	Evaluated multiple semi-supervised learning methods and active learning for classification of {standing still, walking, running} activities
	[41]	2010	Heartbeat classification in electrocardiogram (ECG) data	MIT-BIH Arrhythmia Database	Extracting features (Wavelet coefficients, normalized energy in different beat segments, etc.) from ECG signal, classification by SVM
	[43]	2010	Sample-efficient training of ECG classifier using active learning	MIT-BIH Arrhythmia Database	Active learning strategies used: 1. Margin sampling, 2. Posterior probability, 3. Query by committee, classification by SVM

(*continued on next page*)

Table 1 (*continued*)

Ref.	Year	Objective	Data	Approach
[44]	2015	Whole-heart segmentation for congenital heart disease utilizing active learning	Cardiac MRI	Regions of interest annotated manually by the user during active learning, Patch-based segmentation using KNN
[45]	2015	Lesion tissue segmentation in stroke patients' brain images using active learning	MRI	Classification performed by random forest (RF), samples actively selected for manual labeling according to RF prediction uncertainty about them
[55]	2021	Efficient vaccine allocation to COVID-19 patients	Gathered from Sherkat Naft hospital	The evaluated approach using multiple ML methods e.g., SVM, decision tree, etc.
[56]	2023	Multi-label classification	24 public datasets	Best members among an ensemble of classifiers are used to classify unseen samples
[57]	2022	Prioritize treatment of COVID-19 patients based on their health condition severity	Data gathered from Sherkat Naft hospital	The approach has two phases: 1. Classification by random forest, 2. Queuing optimization by Grasshopper optimization method
[58]	2023	Predicting drug-likeness	ZINC15 dataset	Active learning and ensemble learning are used to predict drug-likeness
ALEC	2023	Sample-efficient and accurate CAD diagnosis	Z-Alizadeh Sani dataset	Exploiting the correlation between stenosis of LAD, LCX, RCA, and having CAD, evaluating several classifiers, e.g., SVM, RF, KNN, etc., compared against existing active learning methods

ML for CAD diagnosis. Our approach superiority is validated by comparing its performance against multiple active learning methods for CAD diagnosis using the Z-Alizadeh Sani dataset.

3. Dataset

The Z-Alizadeh Sani dataset [59] used in the current study contains 303 subjects: 87 healthy subjects and 216 CAD patients. Each subject is represented by 54 features divided into four categories: clinical characteristics, symptoms and signs, ECG, echocardiography, and laboratory tests. The complete list of Z-Alizadeh dataset features grouped by categories is given in Table 2. This is the only public dataset that contains comprehensive and detailed information about individual artery stenosis involving the LAD, LCX, and RCA. Therefore, it has been widely used in various studies [60–64].

In the Z-Alizadeh Sani dataset, diagnosis of CAD was based on the detection of more than 50% coronary artery diameter stenosis on invasive angiography at per patient level as well as per vessel level in one or more of LAD, LCX, and RCA [65]. Among 216 CAD patients, 76, 118, and 113 had stenotic LAD, LCX, and RCA, respectively. Fig. 2 depicts the probability relationships between pairs of CAD, LAD, LCX, and RCA groups in the dataset, with the arrows indicating the direction of the probability statement. The probabilities in Fig. 2 are extracted from samples in the Z-Alizadeh Sani dataset. For example, the value (0.54) on LAD→LCX edge is the probability that LCX is stenotic given that LAD is already stenotic (P (stenotic LCX | stenotic LAD) = 0.54). To compute the aforementioned conditional probability, the number of patients with stenotic LAD and stenotic LCX was divided by the number of patients having stenotic LAD. The rest of the probabilities in Fig. 2 are computed in a similar manner.

Such depiction as in Fig. 2 is original and provides novel insights into the interdependent relationships among the CAD, LAD, LCX, and RCA groups. For example, if any of the LAD, LCX, or RCA groups is stenotic, CAD is already present (indicated as "1" on arrows from the LAD, LCX, and RCA groups pointing to the CAD group). Specific to this dataset, if there was LAD stenosis, the probability of LCX stenosis was 0.54; conversely, if there was LCX stenosis, the probability of LAD stenosis was 0.81, as shown in Fig. 2. These data interdependencies and probability statements are transferred downstream to classification labels, which provide a framework to identify inconsistencies in predicted outcomes.

3.1. Dataset analysis

This section carefully investigates the properties of the Z-Alizadeh Sani dataset. At first, the features of the dataset are analyzed (Section 3.1.1), and then the role of three main arteries in the discrimination of dataset samples is investigated (Section 3.1.2).

3.1.1. Analysis of dataset features

The correlation between pairs of dataset features is shown in Fig. 3, where only a subset of features with higher discrimination power for CAD diagnosis is included. These data have been collected from patients suspected of having CAD who have been referred to the hospital. For example, age positively correlates with features like hypertension (HTN), ST depression, and serum creatinine (CR). Not surprisingly, fasting blood sugar (FBS) and diabetes mellitus status (DM), as well as blood urea nitrogen (BUN) and CR, are positively correlated, whereas neutrophil (Neut) and lymphocyte counts (Lymph) are negatively correlated. The most important features contributing to the diagnosis of CAD and stenosis of LAD, LCX, and RCA are determined using information gain [66] and visualized in Fig. 4. As can be seen, features with higher levels of contribution to CAD diagnosis are shown with higher height in Fig. 4. Typical chest pain is the most dominant feature contributing to stenosis of LAD (Fig. 4a) and RCA (Fig. 4c), as well as having CAD (Fig. 4d). Regarding LCX stenosis, the age feature has the highest contribution.

Table 2
Features of Z-Alizadeh Sani dataset.

Feature type	Feature name	Range
Clinical characteristics	Age	30–86
	Weight	48–120
	Sex	Male, Female
	Body mass index (BMI), kg/m^2	18–41
	Diabetes Mellitus	Yes, No
	Hypertension)	Yes, No
	Current smoker	Yes, No
	Ex-smoker	Yes, No
	Family history	Yes, No
	Obesity	Yes if BMI >25, No otherwise
	Chronic renal failure	Yes, No
	Cerebrovascular accident	Yes, No
	Airway disease	Yes, No
	Thyroid Disease	Yes, No
	Congestive heart failure	Yes, No
	Dyslipidemia	Yes, No
Symptom and sign	Blood pressure, mmHg	90–190
	Pulse rate, beats per minute	50–110
	Edema	Yes, No
	Weak peripheral pulse	Yes, No
	Lung rales	Yes, No
	Systolic murmur	Yes, No
	Diastolic murmur	Yes, No
	Typical Chest Pain	Yes, No
	Dyspnea	Yes, No
	Function class	1, 2, 3, 4
	Atypical	Yes, No
	Nonanginal chest pain	Yes, No
	Exertional chest pain	Yes, No
	Low threshold angina	Yes, No
Electrocardiography	Rhythm	Sinus rhythm, Atrial fibrillation
	Q wave	Yes, No
	ST elevation	Yes, No
	ST depression	Yes, No
	T wave inversion	Yes, No
	Left ventricular hypertrophy	Yes, No
	Poor R wave progression	Yes, No
Laboratory and echo	Fasting blood sugar, mg/dl	62–400
	Creatine, mg/dl	0.5–2.2
	Triglyceride, mg/dl	37–1050
	Low-density lipoprotein, mg/dl	18–232
	High-density lipoprotein, mg/dl	15–111
	Blood urea nitrogen, mg/dl	6–52
	Erythrocyte sedimentation rate, mm/h	1–90
	Hemoglobin, g/dl	8.9–17.6
	Potassium, mEq/l	3.0–6.6
	Sodium, mEq/l	128–156
	White blood cell count, cells/ml	3700–18,000
	Lymphocyte count, %	7–60
	Neutrophil count, %	32–89
	Platelet count, 1000/ml	25–742
	Ejection fraction, %	15–60
	Region wall motion abnormality score, number[a]	0, 1, 2, 3, 4
	Valvular heart disease grade[b]	Normal, Mild, Moderate, and Severe

[a] The values are the average computed over wall segments.
[b] This grade is assigned to the valve with the highest stenosis level compared to the other valves.

3.1.2. Role of main arteries in sample discrimination

The dataset contains labels describing the status of individual arteries (LAD, LCX, and RCA) and overall CAD diagnosis, i.e., the presence of stenosis in at least one artery. In this section, the dataset is analyzed, and the results are visualized using dimensionality reduction methods such as principal component analysis (PCA) [67], partial least squares (PLS) [68], and t-SNE [69] methods. PCA and PLS perform unsupervised and supervised dimensionality reductions, respectively, allowing us to investigate the difference in discrimination power when dataset sample labels are considered or ignored. Fig. 5 shows the data partitioned according to the individual artery (LAD, LCX, RCA) and overall CAD status (normal/stenotic) using PCA, PLS, and t-SNE methods. All three dimensionality reduction methods are good at discriminating normal versus stenotic samples. Overall, PLS performance is slightly better than PCA and t-SNE. In addition, discrimination power is best for CAD, followed by LAD, RCA, and LCX. This stems from the fact that the CAD label contains all the information provided by the status of LAD, LCX, and RCA.

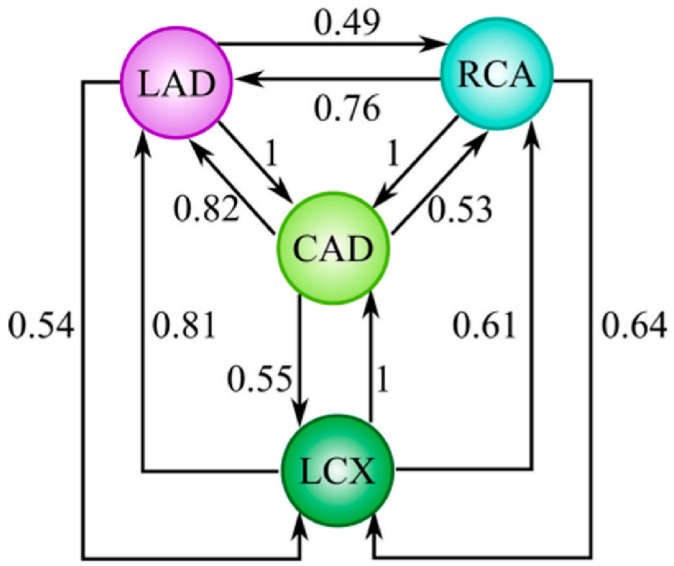


Fig. 2. Probability relationships between stenoses of the different coronary arteries.

The result of investigating correlations between the labeled artery and the other two arteries is shown in Fig. 6. Here, the PLS method is used since it gives better results compared with PCA and t-SNE when the arteries are analyzed independently. For example, in Fig. 6a, the status of the LAD artery is used to denote labels of the dataset samples, and the status of the LCX and RCA arteries is used as features. As the atherosclerotic process is driven by systemic factors, e.g., blood cholesterol level, it is reasonable to assume that knowledge of the status of the other two main arteries will provide some information about the status of the artery of interest.

We further investigate the discrimination power when the number of stenotic arteries is considered a label of dataset samples, again using the PLS method. The result is shown in Fig. 7. The four groups of data points correspond to four label values {Normal, One artery with a stenosis, Two arteries with stenosis, Three arteries with stenosis}. As seen in Fig. 7, samples with the labels "Normal" and "Three arteries with stenosis" are easily distinguishable, and between them are located samples with one or two stenotic arteries.

4. Active learning with ensemble of classifiers (ALEC)

ALEC is based on the interdependent outcomes of four classifiers that are predicated on the presence of CAD in the input sample data as well as in each of its three branch coronary artery territories. The high-level steps of ALEC are shown in Fig. 8. As the first step, the dataset is divided into equal-sized labeled and unlabeled data partitions, and the former is used to train the four classifiers ($C_{CAD}, C_{LAD}, C_{LCX},$ and C_{RCA}). Next, the predictions of the four classifiers (i.e., $\hat{f}_{CAD}, \hat{f}_{LAD}, \hat{f}_{LCX}, \hat{f}_{RCA}$) for unlabeled data are used to quantify the level of inconsistency: high, medium, and low. Inconsistency levels are shown in Table 3. For each unlabeled sample data, if the classifier outputs are consistent, the sample

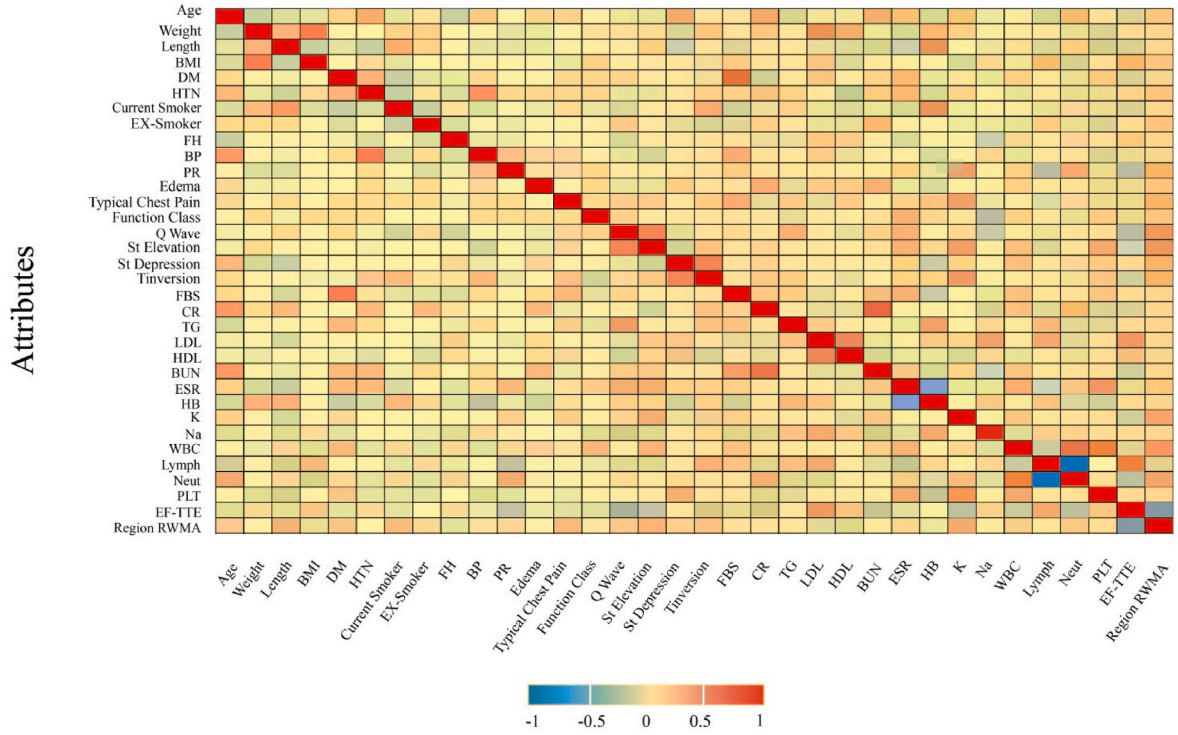


Fig. 3. Correlations between pairs of features in the Z-Alizadeh Sani dataset. In each, the strength of the correlation is indicated by the color scale.

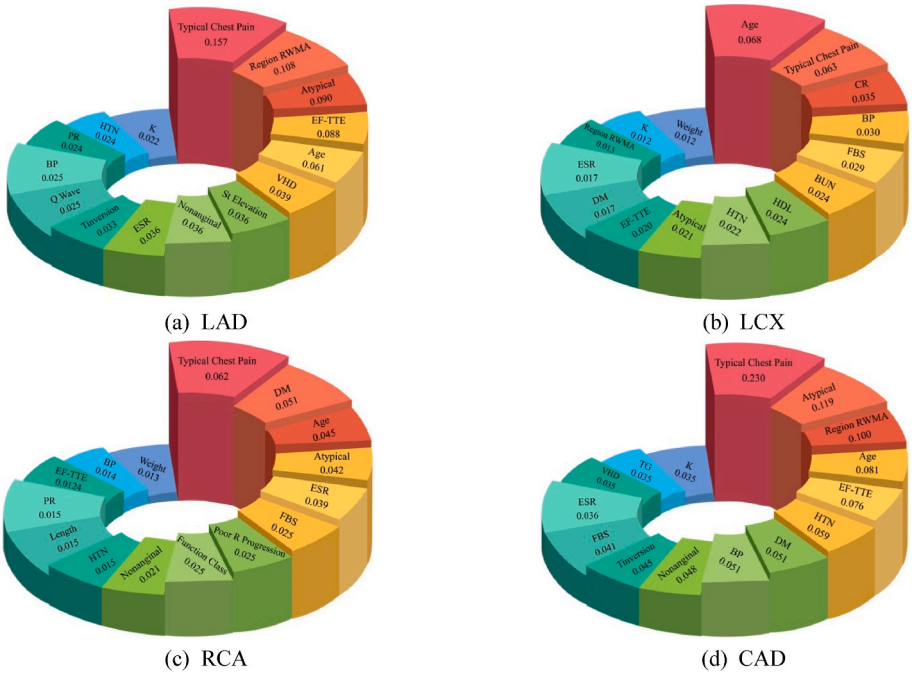


Fig. 4. The most important features contribute to the diagnosis of CAD as well as stenosis of LAD, LCX, and RCA.

is added to the pool of labeled samples for use during the next training phase. Ten percent of the inconsistent sample data are prioritized for manual labeling by medical experts according to the grade of inconsistency (from high to low) before being added to the training pool. The system is retrained on old and new labeled data, and the process is repeated until all data are labeled.

ALEC is a flexible framework that will work with various design choices. Any binary classifier can be used as one of the n classifiers. For example, neural networks can be used as classifiers and can be trained by standard optimizers such as Adam [70] using an appropriate loss function (e.g., binary cross entropy).

The pseudocode of ALEC is presented in Algorithm 1. The input is dataset D; and C_{CAD}, C_{LAD}, C_{LCX}, and C_{RCA} are the classifiers. In Line 1, dataset D is partitioned into labeled and unlabeled samples. In Line 2, the set of samples whose labels will be predicted by ALEC is initialized as an empty set. This set will be used to compute the labeling accuracy of ALEC during the training process (Lines 22–24). The main loop of ALEC is initialized at Line 3. This loop continues until no more unlabeled

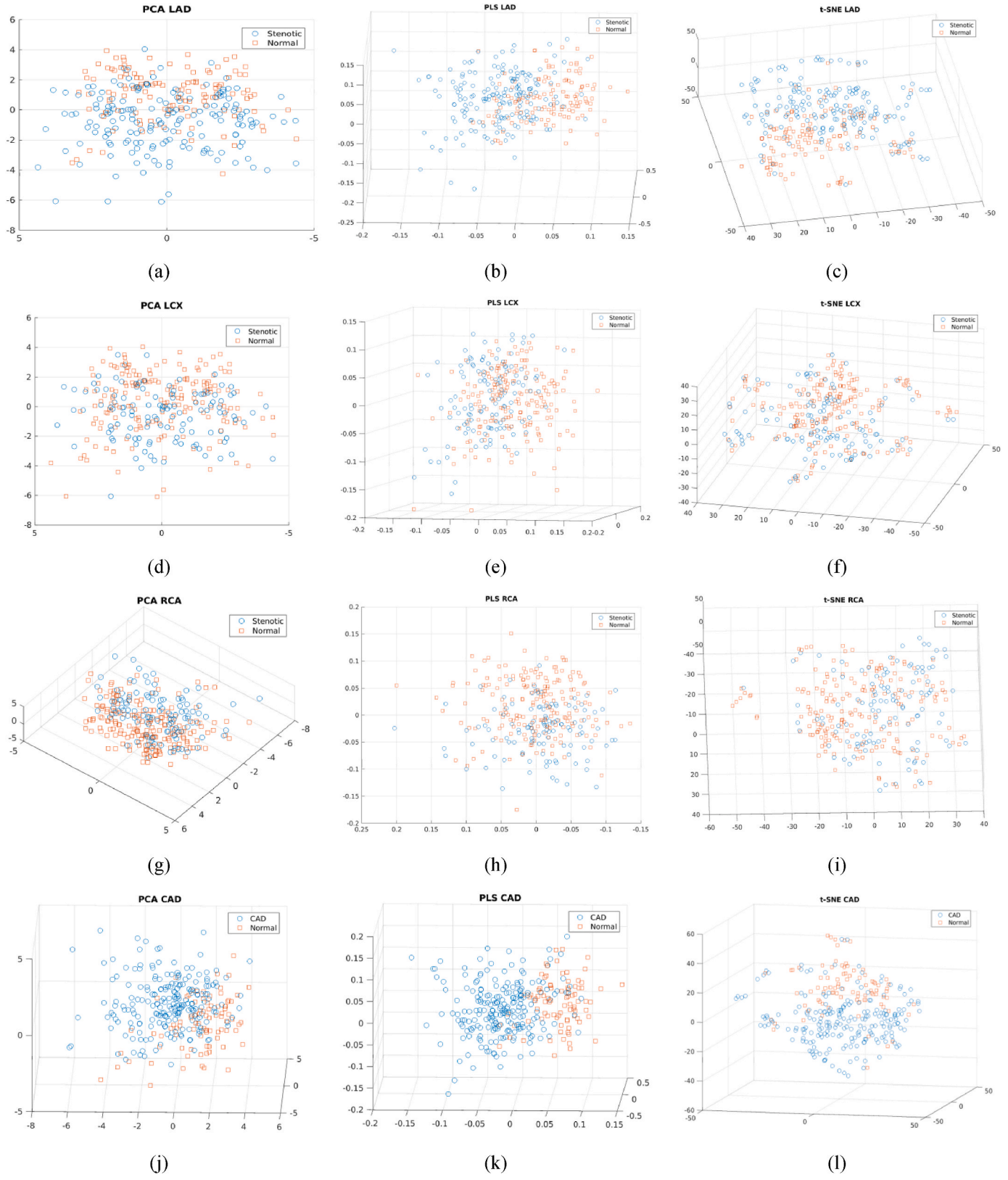


Fig. 5. Graphical representations of the discrimination power of LAD, LCX, RCA, and CAD labels produced using PCA, PLS, and tSNE dimensionality reduction methods.

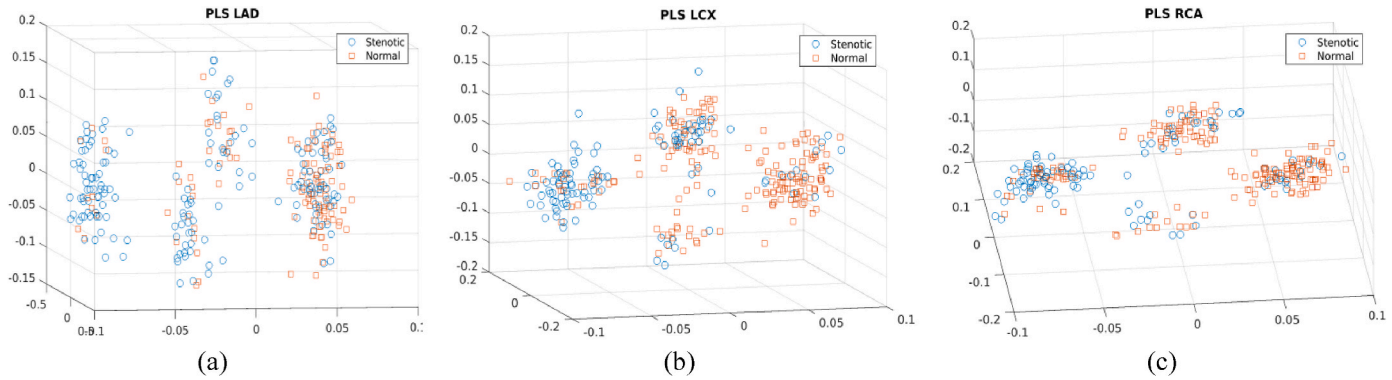


Fig. 6. Discrimination power of two out of three main arteries with the remaining artery as sample label using the PLS method. Discrimination power of LAD label when LCX and RCA are used as features (a); of LCX label when LAD and RCA are used as features (b); and of RCA when LAD and LCX are used as features (c). Part (a) shows the best classification quality, which suggests that {LCX, RCA} provide more information about LAD compared with what {LAD, RCA} and {LAD, LCX} provide about LCX and RCA, respectively.

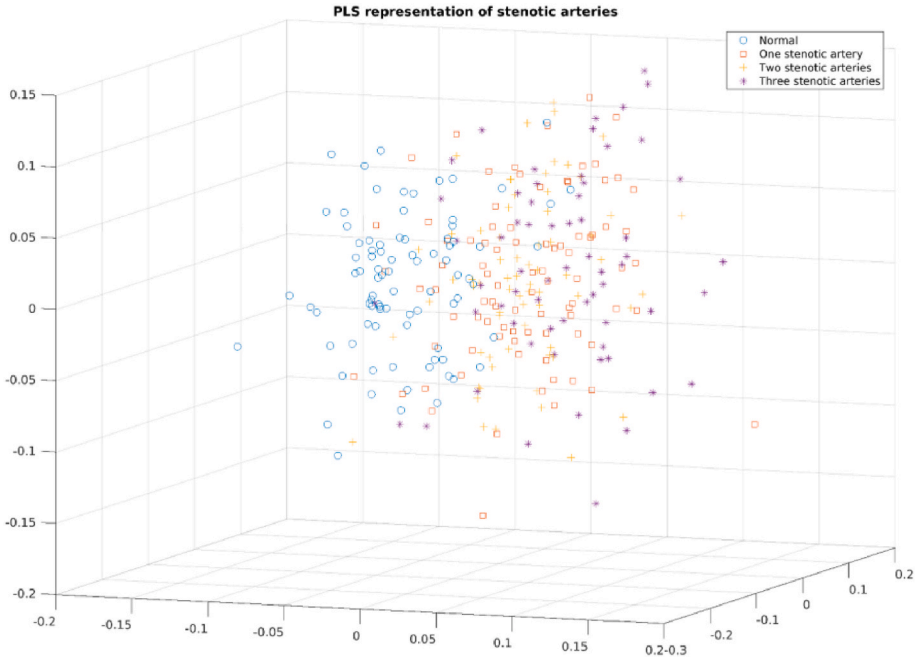


Fig. 7. Scatter plot of dataset samples when the number of stenotic arteries is considered as the label.

samples remain. In Lines 4–7, the four classifiers are trained on the current set of labeled samples. The set of labeled samples grows as more unlabeled samples are labeled either by ALEC or by medical experts. Next, the trained classifiers are used to classify unlabeled samples (Lines 9–13). At Line 14, every unlabeled sample undergoes a consistency check. The level of inconsistency is determined according to Table 3. Consistent samples will be added to the set of labeled samples (Line 15). The samples failing the consistency check will be added to the set of inconsistent samples (Line 18). In Line 19, the inconsistent samples are sorted based on their inconsistency level from high to low. At Line 20, starting from the most inconsistent samples, 10% of the sorted samples are manually labeled by medical experts and added to the set of labeled samples. At this point, one iteration of the "while" loop at Line 3 is completed, and a new iteration begins. After all unlabeled samples have been labeled, the "while" loop (Line 3) terminates, and the labeling performance of ALEC is evaluated at Lines 21–25. At Line 26, the trained classifiers and the computed accuracy are returned.

Algorithm 1. ALEC pseudocode

Algorithm 1: ALEC pseudocode

```
Input: CAD dataset D, Classifiers {C_CAD, C_LAD, C_LCX, C_RCA}
labeled_set, unlabeled_set ←Partition D to 50% labeled and 50% unlabeled sample sets
predicted_set ← {}   // Will contain samples labeled by ALEC
While unlabeled_set ≠ ∅
{
    C_CAD.Train(labeled_set)
    C_LAD.Train(labeled_set)
    C_LCX.Train(labeled_set)
    C_RCA.Train(labeled_set)
    For each sample in unlabeled_set
    {
        f̂_CAD ← C_CAD.Predict(sample)
        f̂_LAD ← C_LAD.Predict(sample)
        f̂_LCX ← C_LCX.Predict(sample)
        f̂_RCA ← C_RCA.Predict(sample)
        inconsistent_set ← {}
        If Is_consistent(f̂_CAD, f̂_LAD, f̂_LCX, f̂_RCA )  // According to Error! Reference source not found.
            labeled_set ← labeled_set ∪ sample
            predicted_set ← predicted_set ∪ sample
        Else
            inconsistent_set ←inconsistent_set ∪ sample
    }  // End for
    sorted_incosistent_samples ← Prioritize_samples(inconsistent_set)   // Prioritize based on inconsistency level (high,
                                                                        // medium, low)
    labeled_set ← labeled_set ∪ Expert_manual_labeling(10% of sorted_inconsistent_samples)
}   // End while
n_correct_prediciton  ← 0
For each sample in predicted_set
{
    If Predicted_label(sample) == true_label(sample)
        n_correct_prediction  ← n_correct_prediction + 1
}  // End for
Accuracy  ← n_correct_prediction / |predict_set|
Return labeled_set, Accuracy
```

5. Results

This section presents the details of the parameter settings used for the different classification methods. The performance of these methods with and without using ALEC is compared. Moreover, ALEC is compared with other active learning methods.

5.1. Classifier parameter settings

We experimented with several classifiers like SVM [71–73], random forest [74–76], k-nearest neighbors (KNN) [77–79], neural network [4, 80,81], C4.5 [82–84], and naïve Bayes [85–87]. The parameter details of these classifiers are presented in Table 4. In SVM, the parameter C of the radial basis function (RBF) kernel trades off the correct classification of training examples against the maximization of the decision function's margin. The parameter γ of the RBF kernel [88,89] can be considered the inverse of the influence radius of samples selected as support vectors. In C4.5, the parameters m, c, g, and s denote the tree construction stopping parameter, tree pruning confidence level parameter, splitting criterion, and subset split, respectively [90–92].

5.2. Comparison of classification methods

We compared the performance of various classifiers on the dataset with and without ALEC and reported the results in Table 5 and Table 6, respectively. With ALEC, the training process was started with an equal number of samples in the labeled and unlabeled sets, but the number of queried samples for manual labeling by experts varied among the individual classifiers. For a fair comparison of the classifiers with and without ALEC, the same number of training samples used by classifiers with ALEC (Equation (1)) was randomly selected from the dataset and provided to the corresponding classifiers for training without ALEC. The total number of training samples in equation (1) is computed based on the number of manually labeled samples with consistent labeling at the last training iteration of classifiers using ALEC.

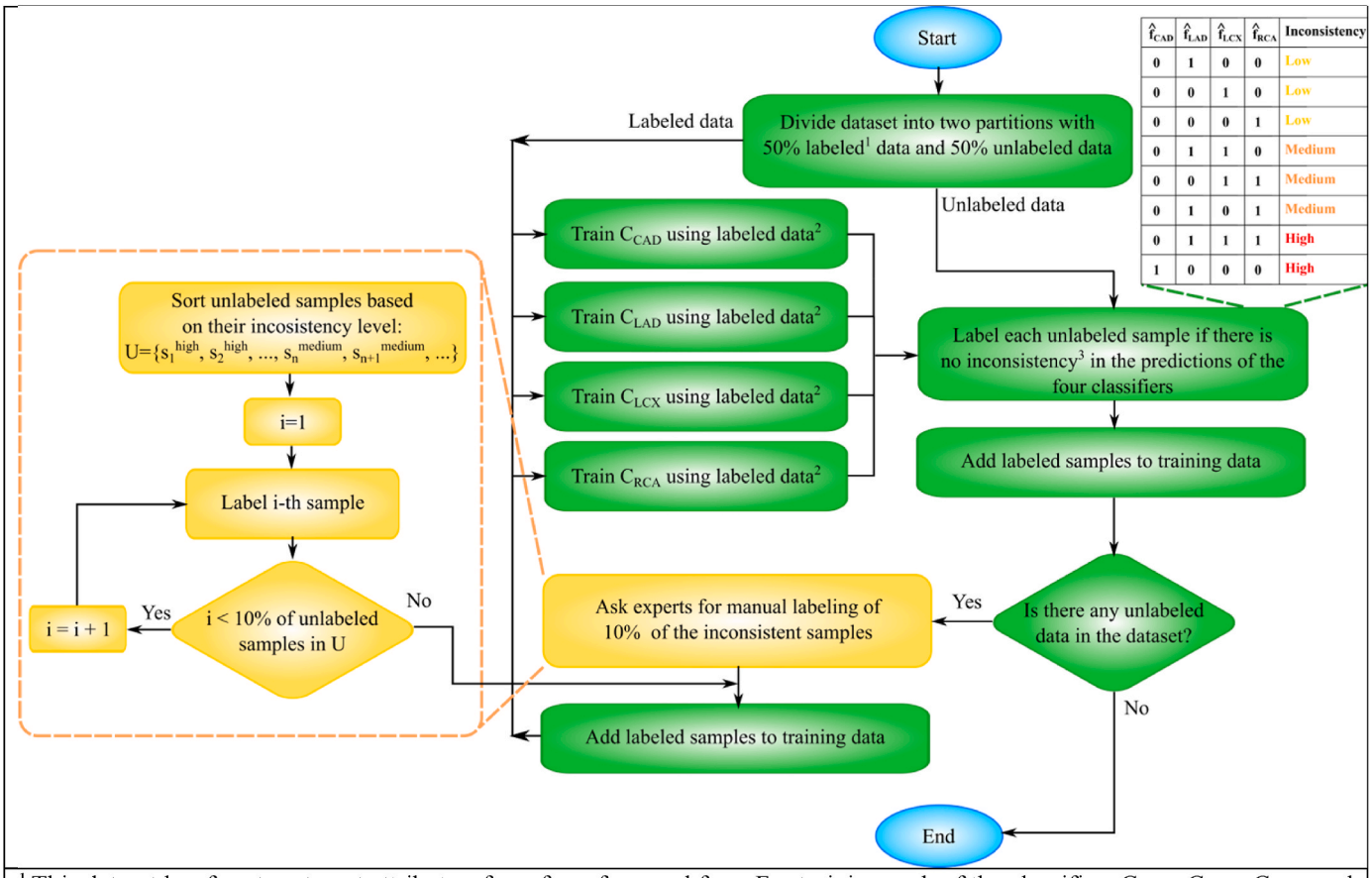
$$\#\ \text{training samples} = \left\lceil \frac{\#\ \textit{dataset samples}}{2} \right\rceil + \#\ \textit{manually labeled samples} + \#\ \textit{samples with consistent labeling} \tag{1}$$

Despite having access to the same number of training samples, classifiers without ALEC performed worse (Table 6) than the scenario in which ALEC was used (Table 5). This is mainly because in the absence of active learning, instead of asking the label of important samples deliberately, a random set of labeled training samples (Equation (1)) are given to the learner methods, reducing the chance of receiving more informative samples.

Among the evaluated classifiers, SVM attained the best performance with and without ALEC. Random forest gained the largest performance boost (absolute accuracy gain of 5.98%) from ALEC. As random forest relies on information theory to build an ensemble of decision trees, access to more informative samples via ALEC can plausibly enhance the quality of generated decision trees and improve classification performance. In contrast, C4.5 forms a single decision, cannot fully exploit the potential of actively selected samples, and consequently exhibited only a modest accuracy boost of 2.52% with ALEC. Considering that K-Nearest Neighbor relies on nearby samples for classification of test samples, it has gained considerable performance thank to more informative samples provided by ALEC. Using informative training samples also helps with reduction of training loss during neural network training as it is evident in Tables 5 and 6. Even Naïve Bayes which is a simple classifier has received 4.15% performance boost due to using good training samples suggested by ALEC.

5.3. Comparison with other active learning methods

We compared the performance of ALEC with the best classifier SVM with other active learning methods using the Active Learning in Python (ALiPy) package [93]. These 19 active learning algorithms are summarized below.



[1] This dataset has four true target attributes: f_{CAD}, f_{LAD}, f_{LCX}, and f_{RCA}. For training each of the classifiers C_{CAD}, C_{LAD}, C_{LCX}, and C_{RCA}, only the corresponding target attribute is kept, and others are ignored. For example, only target f_{CAD} is used for the training of C_{CAD}.

[2] Refers to all previously and recently labeled data.

[3] Inconsistent instances are depicted in Table 3.

Fig. 8. Flowchart of the proposed method.

Hierarchical Active Learning with Cost (HALC): This multi-label active learning strategy exploits label hierarchies for cost-effective queries by incorporating the potential contributions of ancestor and descendant labels to derive a criterion for expressing how much information each candidate query can offer [94]. As each instance may have multiple labels, the cost of annotation is higher. In addition, the labels are prepared into hierarchies from coarse to fine and have variable annotation costs as labels at different hierarchy levels contribute in different ways to model training.

Uncertainty sampling: This sequential algorithm selects the most uncertain instance-label pairs for the expert query [95], and can reduce by 500-fold the amount of training data that needs to be manually classified for the stipulated level of effectiveness. The method is widely used for text categorization but can also be extended to other classification tasks.

Active Self-Paced Learning (ASPL): Lin et al. [96] described a strategy for face identification that combined self-paced learning and active learning to automatically annotate new instances and incorporate

Table 3
Uncertain instances with inconsistency in predicted target attributes.

$\hat{f}_{CAD}$	$\hat{f}_{LAD}$	$\hat{f}_{LCX}$	$\hat{f}_{RCA}$	Level of inconsistency
0	1	0	0	Low
0	0	1	0	Low
0	0	0	1	Low
0	1	1	0	Medium
0	0	1	1	Medium
0	1	0	1	Medium
0	1	1	1	High
1	0	0	0	High

1 and 0 represent the presence and absence of stenosis in the predicted target attributes.

Table 4
Parameter settings for classifiers.

Algorithm name	Value of parameters
Support vector machine	C = 0; γ = 0.1; Kernel type = RBF
Random forest	Number of trees = 10; Minimal size for split = 4; Minimal leaf size = 2; Maximal depth = 20
K-nearest neighbor	K = 5
Neural network	Number of hidden layers = 1; Learning rate = 0.3; Number of epochs = 1000; Optimizer: Adam; Loss function: binary cross entropy
C4.5	m = 2; c = 25; g = off; s = off
Naïve Bayes	–

Table 5
Classification results with the proposed active learning algorithm.

Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)	Area under curve
Support vector machine	97.01	97.21	93.25	0.95
Random forest	96.50	97.22	88.62	0.91
K-nearest neighbor	86.25	87.64	79.25	0.80
Neural network	91.25	97.66	83.25	0.82
C4.5	93.35	95.27	86.98	0.89
Naïve Bayes	84.12	85.75	80.54	0.79

them into training under weak expert recertification. The method proved to be efficient as the performance was not compromised even though the number of annotated samples was drastically decreased. The method had high classification accuracy and was robust against noisy data.

Discriminative and Representative query for Batch Mode active learning (BMDR): Wang and Ye [97] devised an algorithm to select a batch of informative and representative instances by minimizing the empirical risk bound for active learning. A novel upper bound for the true risk in active learning was formulated, and a practical active learning batch mode method was derived by curtailing this upper bound. Their method was able to query the most informative samples while protecting the source distribution as much as possible.

Self-Paced Active Learning (SPAL): Starting the training with easier samples and gradually making the samples harder can improve performance. SPAL [98] is a technique that considers the easiness and potential value of an instance simultaneously. Using this method, the model was trained by querying the appropriate samples at the right time with minimum cost.

Learning Active Learning (LAL): a data-driven strategy was proposed based on training a regression model that predicted error reduction for a candidate in each learning iteration [99]. By formulating the query selection procedure as a regression problem, this method was not constrained to accessible active learning heuristics but instead learned techniques empirically from previous active learning outcomes. The strategy could be learned from either a subset of domain-specific data or simple synthetic two-dimensional datasets.

Expected error reduction: Roy and McCallum [100] proposed an active learning strategy that optimized expected future errors directly. This is contrary to other techniques that aim to lessen version space size. These techniques were popular because closed-form calculation of the expected future error is obstinate in many learning techniques. Their Monte Carlo approach made it feasible to approximate the expected error reduction because of the query labeling.

Cost-Effective Active Learning (CEAL): In traditional active learning, the ground truth of queried labels is returned by a single labeler. Huang et al. [101] performed active selection for both instances and labelers and evaluated instance-labeler pairs' cost-effectiveness for active selection to achieve high labeling accuracy by a selected labeler at a low cost. Such an approach is useful when there are multiple labelers with different qualities of labeling and costs.

Table 6
Classification results without the proposed active learning algorithm using the number of training samples in equation (1).

Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)	Area under curve
Support vector machine	94.25	97.93	91.25	0.91
Random forest	90.52	95.62	83.85	0.84
K-Nearest neighbor	80.65	86.25	77.86	0.75
Neural network	87.63	94.76	80.69	0.76
C4.5	90.83	92.49	85.62	0.85
Naïve Bayes	79.97	84.53	76.85	0.74

Active Feature Acquisition with Supervised Matrix Completion (AFASMC): an approach for training a classification model with a minimal acquisition cost is AFASMC [102] which combines supervised matrix completion and active feature query. The performance may degrade in applications where some features are missing from the dataset. As some features may be correlated with each other, it is possible to recover missing feature values based on these correlations. The quality of recovered feature values and the overall performance relies on choosing the most informative features. The new objective function could diminish reconstruction error and compensate for the loss of training data while completing the feature matrix.

Query-by-Committee (QBC): Mamitsuka [103] proposed a novel query-learning strategy that combined query by committee with bagging and boosting. QBC used many copies of a randomized algorithm as a committee. The randomized algorithms performed prediction on data samples. The sample with the highest uncertainty among members of the committee was chosen for the query. Such a sample provided the maximum information gain to select the next query point where the weighted majority voting by acquiring hypotheses had a minimal margin.

Active qUery on Relevance Ordering (AURO): as a simple selection strategy, AURO [104] was proposed to query triplets comprising one instance and two labels each, based on the relative order of relevance of the labels to the instance. This novel multi-label active learning approach, which demanded less proficiency of the annotator and yet obtained richer information with each query, significantly lessened the labeling task of annotators.

QUerying Informative and Representative Examples (QUIRE): this method adapted the active learning min-max view to measure sample representativeness based on prediction uncertainties in the labeled and unlabeled samples, respectively [105]. Using QUIRE, the ad hoc query of unlabeled instances was optimized according to their representativeness level.

Stability: Chakraborty et al. [106] proposed an ensemble/stability-based method to utilize conditional Gaussian distributions for uncertainty prediction of missing entries in the matrix to recover the complete matrix from partially observed entries (active matrix completion), thereby selecting query locations.

Interval Estimate Threshold (IEThresh): from multiple experts or oracles with different degrees of reliability, the ones with the highest labeling accuracies were selected [107]. To this end, labeling sources with measured upper-bound confidence intervals below a specified threshold were filtered out, which enriched the overall accuracy of the remaining ones. Therefore, the most informative samples for active learning were obtained with minimum labeling effort.

Active learning based on Uncertainty and Diversity for Incremental multi-label learning (AUDI): Instances may be simultaneously associated with multiple labels, increasing labeling costs. A novel multi-label strategy was proposed, which exploits measures of diversity and uncertainty in the label and instance spaces to select the optimal instance-label pairs for iterative query on whether each label was positive on each instance [108].

Adaptive: Multi-label classification involves dispersing each instance to multiple groups but can be expensive and time-consuming. Li and Guo [109] proposed an adaptive framework for active instance selection. To this end, two new multi-label active learning approaches i.e. label cardinality inconsistency approach and the max-margin prediction uncertainty approach were integrated.

Maximum loss reduction with Maximal Confidence (MMC): In a multi-label text classification paradigm, creating multiple labels for each document is impractical since multi-label text classifiers require large training data. A new multi-label active learning strategy was proposed which could reduce the amount of required labeled data without losing classification accuracy [110]. This method optimized the cutback in

expected loss, which was computed as the sum of all label losses. The size of the version space determined the model loss, and the reduction rate of the version space was optimized using SVM. The method could predict labels for each unlabeled data.

Repeated: repeated labeling of data items conferred advantages when it was costly to process unlabeled data, where it would be cheaper to label everything multiple times, or if the labels were noisy [111]. However, repeated labeling could not always enhance the model and label quality, and different notions of uncertainty should be considered when choosing data points for repeated labeling.

Random graph density: Ebert et al. [112] constructed an exploratory graph using dataset samples as graph nodes. Edges of the graph were weighted according to the Manhattan distance between pairs of nodes. The graph edges were the K-nearest neighbors of each sample in the graph. Highly connected samples were identified using the graph. These were the samples that belonged to the same class.

It is also useful to categorize the 19 active learning methods reviewed above according to their working mechanisms or predominant data issues. The categorization is shown in Table 7. As can be seen, most of the methods are based on the instance selection mechanism. The second most common category is multi-label data which contains four methods. The rest of the categories in Table 7 are less popular.

The 20 algorithms were run 100 times for a fair comparison, and the corresponding mean accuracies were compared. In each run, 50% of the samples were labeled and used for training, and the rest were unlabeled. The same set of training samples was given to all 20 active learning algorithms, and SVM with RBF kernel was used as the base classifier in all cases. SVM was acknowledged to be the best classifier in the field [122]. The comparison results are illustrated in Fig. 9, in which the x-axis represents the number of queries, i.e., the number of samples manually labeled. Our proposed method (ALEC) ranks first, followed by SPAL and the uncertainty method. The proposed method queries samples with higher inconsistency levels between predicted attributes (CAD, LAD, LCX, and RCA) (see Table 3). Such an approach not only eased the label identification of difficult unlabeled samples but also helped with learning the labels of simpler samples.

Curriculum learning has always been an effective approach to improving the learning process. Inspired by curriculum learning, SPAL starts learning with easier samples and tackles harder samples as the training unfolds. The performance of active learning methods heavily depends on the strategy based on which samples are chosen for manual labeling. The uncertainty method is based on such observation and chooses samples with the highest uncertainty for manual labeling. It is no surprise that this method has gained a third place among the evaluated methods.

Table 7
Categories of 19 active learning algorithms implemented on Active Learning in Python package.

Main Mechanism	Active Learning Algorithm
Instance selection	Active self-paced learning [96]
	Uncertainty [95]
	Query by committee [113]
	Expected error reduction [114]
	Random graph density [112]
	Discriminative and representative queries for batch mode active learning [97]
	Learning active learning [99]
	Self-paced active learning [98]
Multi-label data	Active learning based on uncertainty and diversity for incremental multi-label learning [115]
	Querying informative and Representative [105]
	Maximum loss reduction with maximal confidence [116]
	Adaptive [117]
Querying features	Active feature acquisition with supervised matrix completion [118]
	Stability [119]
Novel query type	Active query on relevance ordering [104]
Different labeling costs	Hierarchical active learning with cost [94]
	Cost-effective active learning [101]
Noisy oracles	Interval estimate threshold [120]
	Repeated [121]

HALC relies on label hierarchy. In CAD diagnosis, such label hierarchy exists, as shown by the probability relationships between stenosis of different coronary arteries (Fig. 2). By exploiting the labeling hierarchy between CAD and LAD, LCX, and RCA features, HALC has gained a fourth place among the evaluated methods. HALC is expected to obtain better performance in problems with multiple levels of hierarchy. In our case, the hierarchy is not very deep which is why HALC has exhibited lower performance compared to ALEC, SPAL, and uncertainty method.

The LAL method seeks to ask for labels of samples that yield maximum reduction in prediction error. ASPL chooses samples it is most uncertain about for manual labeling. These are exactly the samples that will give the classifier the largest performance boost. The expected error reduction method estimates the expected reduction in future errors upon asking for the label of a specific sample. The samples yielding the highest error reduction will be chosen for manual labeling. The sample selection strategies of LAL, ASPL, and Expected error reduction method are somewhat similar to our proposed method (ALEC) sample selection criterion for manual labeling. In our method, samples with the highest inconsistency level are chosen for manual labeling since they are the ones that can correct potential inconsistencies between our four classifiers. Therefore, LAL, ASPL, the expected error reduction method, and ALEC are practically the same, except that they come from different perspectives. The reason that ALEC outperforms LAL and ASPL is that ALEC incorporates medical domain knowledge regarding CAD diagnosis, whereas LAL and ASPL are general-purpose active learning methods.

CEAL method relies on actively selecting samples and labelers to enhance its performance. However, in our experiments, there is only one labeling source. Therefore, CEAL cannot reach its full potential. That is the reason for its poorer performance compared to ALEC and some other evaluated methods such as SPAL, HALC, and LAL.

BMDR balances the discriminative and representative information during empirical risk minimization, so BMDR is expected to outperform other active learning methods. However, BMDR is a batch-based method, so at each manual labeling phase, it queries the labels of multiple samples. Therefore, the batch mode approach leads to poor sample efficiency of the BMDR method. That is why it has been outperformed by some other methods.

AFASMC method can estimate missing values of features. However, the dataset used in our experiments has no feature with missing values, so AFASMC cannot take advantage of its ability to estimate missing values. That is why it has underperformed compared to some of the other methods. The stability method also focuses on the completion of missing values that are not present in the evaluated dataset. Like AFASMC, stability cannot exploit its ability for missing value completion.

QBC handles uncertainty by querying the labels of samples near the margin since these samples are the ones that the classifiers are most uncertain about. However, QBC is designed to be a general-purpose method, so it could not perform as well as our proposed method, which is specific to CAD diagnosis. Moreover, training multiple instances of the same classifier when the number of labeled samples is limited is not going to be that much useful. The reason is that the number of labeled samples is not enough to initialize the parameters of multiple classifiers to sensible values.

The AURO method focuses on multi-label problems. The samples of the dataset used in our experiments are not multi-label. Therefore, the AURO method could not use its ability to gain more information from multi-label samples.

QUIRE relaxes the label of the samples to continuous values. However, sample labels are discrete values. Such relaxation may hurt performance. IEthresh method has the ability to select among multiple

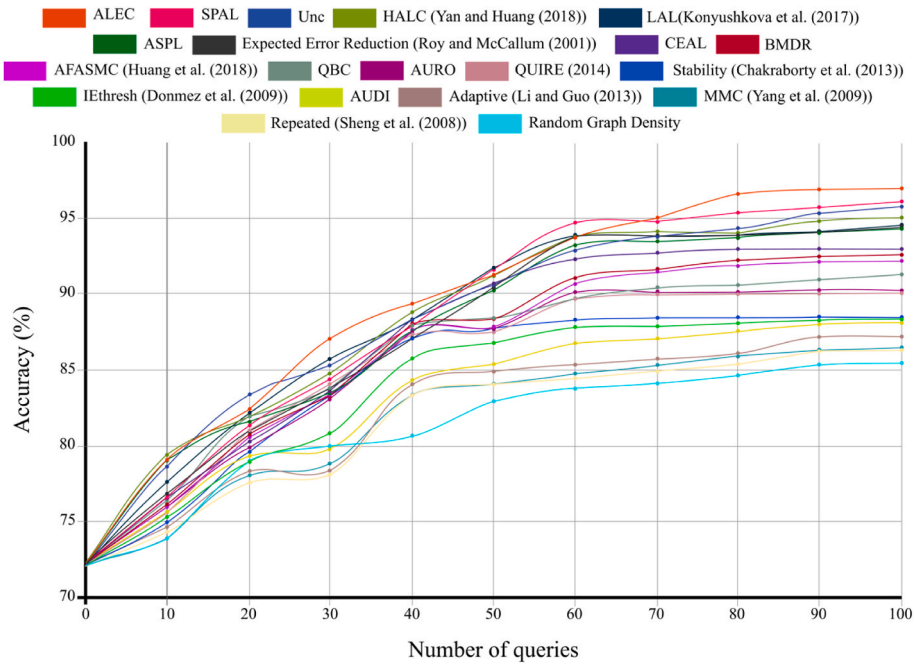


Fig. 9. Comparison of different active learning algorithms.

annotators. However, in our experiments, there is only one source for labeling queried samples. Therefore, IEthresh cannot benefit from selecting among multiple annotators.

AUDI, Adaptive, and MMC are all multi-label active learning methods. However, the evaluated dataset in our experiments is the single label, so the ability of these methods to handle multi-label data is not of much use here.

The effort-repeated method, which puts labels on each sample multiple times, is only beneficial when data samples are noisy. The evaluated dataset is not noisy, so the repeated method does not gain any advantage from labeling samples multiple times.

Random graph density has the worst performance among the evaluated methods. This is mainly due to heavy reliance on the K-nearest neighbor algorithm while forming the graph between dataset samples. The structure of the graph is very sensitive to the choice of hyperparameter K for the K-nearest neighbor method. It is not clear how the appropriate choice for K should be determined for different problems. An inappropriate choice of K can hurt performance badly.

5.4. Theoretical considerations

In this section, we lay out the theoretical basis of why ALEC has outperformed standard classifiers. $C_T(p)$ represents the consistency condition. Whenever the condition in equation (3) holds, the predictions of the four classifiers are consistent, and there is no need to query the sample label.

$$C_T(p) = (C_{LAD}(p) \vee C_{LCX}(p) \vee C_{RCA}(p)) \odot C_{CAD}(p) \tag{2}$$

The operator of ʘ shows the XNOR or logical complement of the exclusive OR operator. $C_{CAD}(P)$, $C_{LAD}(P)$, $C_{LCX}(P)$, and $C_{RCA}(P)$ are the outputs of classifiers used for diagnosing CAD and stenosis of LAD, LCX, and RCA. The error of $C_T(P)$ classifier is given by:

$$\begin{aligned} error(C_T) &= FPR(C_T) \times Pr(CAD=0) + FNR(C_T) \times Pr(CAD=1) \\ &= \Pr((X \cup Y \cup Z) \cap W | CAD=0) \times \Pr(CAD=0) \\ &+ \Pr\left(\overline{X} \cap \overline{Y} \cap \overline{Z} \cap \overline{W} | CAD=1\right) \times \Pr\left(CAD=1\right) \end{aligned} \tag{3}$$

where X, Y, Z, and W are stochastic events corresponding to C_{LAD} predicting LAD has stenosis, C_{LCX} predicting LCX has stenosis, C_{RCA} predicting RCA has stenosis, and C_{CAD} predicting patient has CAD. False positive rate (FPR) and false negative rate (FNR) are the probabilities of false alarm and miss rate, respectively [123]:

$$FPR = \frac{\sum False\ positive}{\sum Condition\ negative} \tag{4}$$

$$FNR = \frac{\sum False\ negative}{\sum Condition\ positive} \tag{5}$$

Next, we will prove that the error rate of our proposed method is lower than the standard classifier ($C_{CAD}(p)$) in Lemma 1. In Lemma 2, we also prove that our method error rate is lower than classification using the following rule:

$$C_T'(p) = C_{LAD}(p) \vee C_{LCX}(p) \vee C_{RCA}(p). \tag{6}$$

Accordingly, we will prove the two following lemmas.

Lemma 1. $error(C_T) \leq error(C_{CAD}(p))$.

It is obvious that,

$$P(A \cap B) \leq P(B) \tag{7}$$

Applying the inequality in Equation (7) to "$\Pr((X \cup Y \cup Z) \cap W | CAD = 0)$" in equation (3) yields:

$$\begin{aligned} error(C_T) &= \Pr((X \cup Y \cup Z) \cap W | CAD=0) \times \Pr(CAD=0) \\ &+ \Pr(\overline{X} \cap \overline{Y} \cap \overline{Z} \cap \overline{W} | CAD=1) \times \Pr(CAD=1) \leq \Pr(W | CAD=0) \\ &\times \Pr(CAD=0) + \Pr(\overline{X} \cap \overline{Y} \cap \overline{Z} \cap \overline{W} | CAD=1) \\ &\times \Pr(CAD=1) \leq \Pr(W | CAD=0) \times \Pr(CAD=0) \\ &+ \Pr(\overline{W} | CAD=1) \times \Pr(CAD=1) = \text{error}(C_{CAD}(p)) \end{aligned} \tag{8}$$

which proves the following inequality:

$$error(C_T) \leq error(\ C_{CAD}(p)) \tag{9}$$

Lemma 2. $error(C_T) \leq error(C_T')$.

We know that

$$error(C_T^{'}) = \Pr(X \cup Y \cup Z | CAD = 0) \times \Pr(CAD = 0) + \Pr\left(\overline{X} \cap \overline{Y} \cap \overline{Z} | CAD = 1\right) \times \Pr\left(CAD = 1\right) \tag{7}$$

So, according to Equation (3)

$$\begin{aligned} error(C_T) &= \Pr((X \cup Y \cup Z) \cap W | CAD = 0) \times \Pr(CAD = 0) \\ &+ \Pr(\overline{X} \cap \overline{Y} \cap \overline{Z} \cap \overline{W} | CAD = 1) \times \Pr(CAD = 1) \\ &\leq \Pr(X \cup Y \cup Z | CAD = 0) \times \Pr(CAD = 0) \\ &+ \Pr(\overline{X} \cap \overline{Y} \cap \overline{Z} \cap \overline{W} | CAD = 1) \times \Pr(CAD = 1) \\ &\leq \Pr(X \cup Y \cup Z | CAD = 0) \times \Pr(CAD = 0) \\ &+ \Pr(\overline{X} \cap \overline{Y} \cap \overline{Z} | CAD = 1) \times \Pr(CAD = 1) = error(C_T^{'}) \end{aligned} \tag{8}$$

Consequently, we show that $error(C_T) \leq error(C_T')$.

According to the results obtained in Lemmas 1 and 2, combining the four classifiers $C_{CAD}(P)$, $C_{LAD}(P)$, $C_{LCX}(P)$, and $C_{RCA}(P)$ leads to better classification performance for CAD diagnosis compared with using the single classifier $C_{CAD}(P)$ and the three classifiers $C_{LAD}(P)$, $C_{LCX}(P)$, and $C_{RCA}(P)$ in combination, respectively.

5.5. Discussion

We were motivated to develop a computer-aided diagnostic tool for the clinical prediction of CAD because of the significant impact of the disease and the high cost and associated risks of invasive angiography. However, high-quality real-world clinical data are scarce due to high labeling costs. Active learning approaches can circumvent these challenges by optimizing the selection of samples for manual labeling, thereby improving model training efficiency and reducing annotation costs. In this study, we have proposed a novel ALEC method for CAD diagnosis based on the outcomes of four classifiers predicated on the presence of CAD in the input sample and each of its three branch coronary artery territories: LAD, LCX, and RCA. ALEC is first trained using labeled samples. For each unlabeled sample, if the classifier outputs are consistent, the sample is added to the labeled samples pool for iterative training; if inconsistent, the sample is manually labeled by medical experts before being added to the pool. The process is repeated until all samples are labeled. Compared with 19 other established active learning algorithms, ALEC combined with the SVM classifier attained superior performance with 97.01% accuracy. The comprehensive nature of the Z-Alizadeh Sani dataset underpins the interdependent probabilistic relationships between per-patient CAD diagnosis and per vessel LAD, LCX, and RCA stenosis. The empirical arguments for the superiority of the proposed method are elegantly supplemented by theoretical justification. The solutions to the lemmas posed above established the theoretical foundation for the soundness of the proposed method. ALEC also owes its excellent performance to the input of domain-specific knowledge of medical experts. Using the domain knowledge, important samples are chosen for manual labeling by experts. While the reliance on problem-specific knowledge is a limitation with ALEC, the number of samples that are required to be labeled manually is significantly reduced, which should inspire interest and garner wider application.

To diagnose CAD efficiently, it is crucial to identify its key associated factors. This study tries to shed some light on the unexplored aspects of CAD disease by careful inspection of the Z-Alizadeh Sani CAD dataset. The correlation matrix (Fig. 3) and top features contributing to LAD, LCX, and RCA stenosis and having CAD (Fig. 4) provide important insights into the interplay between factors contributing to CAD. One of the important observations is that pairs of correlated features (according to Fig. 3) have approximately the same amount of contribution to the status of LAD, LCX, RCA, and CAD (Fig. 4). The data samples are also visualized based on the number of stenotic arteries (Fig. 7), demonstrating that the number of stenotic arteries can be used to organize dataset samples in a meaningful and consistent way. It is evident from Fig. 7 that normal samples (with no stenotic arteries) and CAD samples with three stenotic arteries form the two extremes of the dataset samples spectrum, while samples with one and two stenotic samples lie between the two extremes of the spectrum.

6. Conclusion and future work

In this research, we have proposed a new active learning method, ALEC, for CAD diagnosis that classifies stenosis of individual coronary arteries. To this end, demographic, clinical, ECG, and laboratory features were used to train the four classifiers to determine CAD and the stenosis status in LAD, LCX, and RCA vessel territories for each patient. , The inconsistency of the classifier outcomes was checked against the interdependent relationships of the four diagnostic categories and graded. For each test sample, if there was no inconsistency between the predictions of the four classifiers, the assigned label to the sample was confirmed. However, if there was inconsistency in the labels assigned to the sample, the sample was ranked according to the level of inconsistency. The higher the level of inconsistency in sample labels, the higher the priority the sample had to be queried for labeling by medical experts involved in the dataset acquisition. Combining this ensemble of four classifiers with SVM, ALEC attained superior performance compared to existing active learning modules implemented by ALiPy.

ALEC can be used as a diagnostic step before actually performing angiography on patients suspected of having CAD. This leads to reduced diagnosis costs and more importantly avoids the unnecessary risks and side effects associated with angiography. The only limitation of ALEC is that it is CAD-specific. The success of ALEC is due to the fact that domain knowledge about CAD disease is exploited to create an ensemble of classifiers to make robust predictions. Therefore, the proposed approach can only be applied to diseases with multiple related features (i.e., symptoms). Recall that in the case of CAD disease, the related features are stenosis of 3 main coronary arteries of the heart, i.e., LAD, LCX, and RCA. As future work, the possibility of applying ALEC to problems similar to CAD is worth investigating. Semi-supervised learning algorithms can also be incorporated to improve the efficiency of our algorithm on larger datasets with many samples.

Declaration of competing interest

None Declared.

References

[1] J. Qiu, Q. Wu, G. Ding, Y. Xu, S. Feng, A survey of machine learning for big data processing, EURASIP Journal on Advances in Signal Processing, 2016 (2016) 67–83.
[2] Y. Fu, X. Zhu, B. Li, A survey on instance selection for active learning, Knowl. Inf. Syst. 35 (2013) 249–283.
[3] F. Khozeimeh, D. Sharifrazi, N.H. Izadi, J.H. Joloudari, A. Shoeibi, R. Alizadehsani, M. Tartibi, S. Hussain, Z.A. Sani, M. Khodatars, RF-CNN-F: random forest with convolutional neural network features for coronary artery disease diagnosis based on cardiac magnetic resonance, Sci. Rep. 12 (2022) 1–12.
[4] K. Kakhi, R. Alizadehsani, H.D. Kabir, A. Khosravi, S. Nahavandi, U.R. Acharya, The internet of medical things and artificial intelligence: trends, challenges, and opportunities, Biocybern. Biomed. Eng. 42 (2022) 749–771.
[5] J.H. Joloudari, S. Mojrian, I. Nodehi, A. Mashmool, Z.K. Zadegan, S. K. Shirkharkolaie, R. Alizadehsani, T. Tamadon, S. Khosravi, M.A. Kohnehshari, Application of Artificial Intelligence Techniques for Automated Detection of Myocardial Infarction: a Review, 2022.
[6] B. Settles, Active Learning Literature Survey, University of Wisconsin-Madison Department of Computer Sciences, 2009.
[7] M. Wang, X.-S. Hua, Active learning in multimedia annotation and retrieval: a survey ACM, Trans. Intell. Syst. Technol 2 (2011) 1–21.
[8] A. Krishnakumar, Active Learning Literature Survey, Tech. Rep., University of California, Santa Cruz. 42, 2007. Tech. Rep.
[9] D. Tuia, M. Volpi, L. Copa, M. Kanevski, J. Munoz-Mari, A survey of active learning algorithms for supervised remote sensing image classification, IEEE Journal of Selected Topics in Signal Processing 5 (2011) 606–617.
[10] X. Zhu, Semi-supervised learning literature review, Technical Reporter (2005) 1530.
[11] N.N. Pise, P. Kulkarni, A Survey of Semi-supervised Learning Methods, International Conference on Computational Intelligence and Security, 2008, pp. 30–34.

[12] Z. Luo, M. Hauskrecht, Hierarchical active learning with group proportion feedback, IJCAI 4 (2018) 2532–2538.
[13] M. Sharma, M. Bilgic, Evidence-based uncertainty sampling for active learning, Data Min. Knowl. Discov. 31 (2017) 164–202.
[14] S. Ghanavati, J.P. Lerch, J.G. Sled, Automatic anatomical labeling of the complete cerebral vasculature in mouse models, Neuroimage 95 (2014) 117–128.
[15] V. Hatzivassiloglou, P.A. Duboué, A. Rzhetsky, Disambiguating proteins, genes, and RNA in text: a machine learning approach, Bioinformatics 17 (2001) S97–S106.
[16] F. Vilariño, G. Lacey, J. Zhou, H. Mulcahy, S. Patchett, Automatic Labeling of Colonoscopy Video for Cancer Detection, Pattern Recognition and Image Analysis, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 290–297.
[17] L. Zhang, Q. Wang, Y. Gao, G. Wu, D. Shen, Learning of Atlas Forest Hierarchy for Automatic Labeling of MR Brain Images, Machine Learning in Medical Imaging, Springer International Publishing, Cham, 2014, pp. 323–330.
[18] T. Heimann, P. Mountney, M. John, R. Ionasec, Learning without Labeling: Domain Adaptation for Ultrasound Transducer Localization, Medical Image Computing and Computer-Assisted Intervention, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 49–56.
[19] R. Ghnemat, A. Khalil, Q. Abu Al-Haija, Ischemic stroke lesion segmentation using mutation model and generative adversarial network, Electronics 12 (2023) 590.
[20] Q.A. Al-Haija, A. Adebanjo, Breast cancer diagnosis in histopathological images using ResNet-50 convolutional neural network, in: 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), IEEE, 2020, pp. 1–7.
[21] Q.A. Al-Haija, M. Smadi, O.M. Al-Bataineh, Identifying phasic dopamine releases using DarkNet-19 convolutional neural network, in: 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), IEEE, 2021, pp. 1–5.
[22] Q.A. Al-Haija, M. Smadi, O.M. Al-Bataineh, Early stage diabetes risk prediction via machine learning, in: Proceedings of the 13th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2021), Springer, 2022, pp. 451–461.
[23] Z. Sun, Z. Ji, P. Zhang, C. Chen, X. Qian, X. Du, Q. Wan, Automatic labeling of mobile apps by the type of psychological needs they satisfy, Telematics Inf. 34 (2017) 767–778.
[24] L. Ching-Yung, B.L. Tseng, M. Naphade, A. Natsev, J.R. Smith, VideoAL: a novel end-to-end MPEG-7 video automatic labeling system, International Conference on Image Processing (2003) 49–53.
[25] T. Subbulakshmi, G. Mathew, S.M. Shalinie, Real time classification and clustering of ids alerts using machine learning algorithms, International journal of Artificial & Application 1 (2010) 1–9.
[26] M. Erdmann, E. Ward, K. Ikeda, G. Hattori, C. Ono, Y. Takishima, Automatic Labeling of Training Data for Collecting Tweets for Ambiguous TV Program Titles, International Conference on Social Computing, 2013, pp. 796–802.
[27] R.A. Bridges, C.L. Jones, M.D. Iannacone, K.M. Testa, J.R. Goodall, Automatic labeling for entity extraction in cyber security, 3, 2013, pp. 1–11, arXiv preprint arXiv:1308.4941.
[28] H. Hammer, K.W. Kongsgård, A. Bai, A. Yazidi, N.A. Nordbotten, P.E. Engelstad, Automatic security classification by machine learning for cross-domain information exchange, in: IEEE Military Communications Conference, 2015, pp. 1590–1595.
[29] D. Magatti, S. Calegari, D. Ciucci, F. Stella, Automatic Labeling of Topics, Ninth International Conference on Intelligent Systems Design and Applications, 2009, pp. 1227–1232.
[30] X.-L. Mao, Z.-Y. Ming, Z.-J. Zha, T.-S. Chua, H. Yan, X. Li, Automatic labeling hierarchical topics, in: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, ACM, Maui, Hawaii, USA, 2012, pp. 2383–2386.
[31] Q. Mei, X. Shen, C. Zhai, Automatic labeling of multinomial topic models, in: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Jose, California, USA, 2007, pp. 490–499.
[32] M. Allahyari, K. Kochut, Automatic topic labeling using ontology-based topic models, in: International Conference on Machine Learning and Applications (ICMLA), 2015, pp. 259–264.
[33] I. Hulpus, C. Hayes, M. Karnstedt, D. Greene, Unsupervised graph-based topic labelling using dbpedia, in: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, ACM, Rome, Italy, 2013, pp. 465–474.
[34] D. Gildea, D. Jurafsky, Automatic labeling of semantic roles, Comput. Ling. 28 (2002) 245–288.
[35] F.J. Damerau, D.E. Johnson, M.C. Buskirk Jr., Automatic Labeling of Unlabeled Text Data, Google Patents, 2004, pp. 1–7.
[36] K. Hacioglu, Y. Chen, B. Douglas, Automatic Time Expression Labeling for English and Chinese Text, Computational Linguistics and Intelligent Text Processing, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 548–559.
[37] H. Li, T. Zhang, R. Qiu, L. Ma, Grammar-based semi-supervised incremental learning in automatic speech recognition and labeling, Energy Proc. 17 (2012) 1843–1849.
[38] Y. Fu, T.S. Huang, Unsupervised locally embedded clustering for automatic high-dimensional data labeling, IEEE International Conference on Acoustics, Speech and Signal Processing (2007) 1057–1060.
[39] P. Bizopoulos, D. Koutsouris, Deep learning in cardiology, IEEE Reviews in Biomedical Engineering 12 (2019) 168–193.
[40] B. Longstaff, S. Reddy, D. Estrin, Improving activity classification for health applications on mobile devices using active and semi-supervised learning, in: 4th International Conference on Pervasive Computing Technologies for Healthcare, 2010, pp. 1–7.
[41] J. Wiens, J.V. Guttag, Active learning applied to patient-adaptive heartbeat classification, Adv. Neural Inf. Process. Syst. (2010) 2442–2450.
[42] L. Sun, Y. Lu, K. Yang, S. Li, ECG analysis using multiple instance learning for myocardial infarction detection, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 59 (2012) 3348–3356.
[43] E. Pasolli, F. Melgani, Active learning methods for electrocardiographic signal classification, IEEE Trans. Inf. Technol. Biomed. 14 (2010) 1405–1416.
[44] D.F. Pace, A.V. Dalca, T. Geva, A.J. Powell, M.H. Moghari, P. Golland, Interactive whole-heart segmentation in congenital heart disease, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer International Publishing, Cham, 2015, pp. 80–88.
[45] D. Chyzhyk, R. Dacosta-Aguayo, M. Mataró, M. Graña, An active learning approach for stroke lesion segmentation on multimodal MRI data, Neurocomputing 150 (2015) 26–36.
[46] W.W. Cohen, Fast effective rule induction, in: Machine Learning Proceedings, Elsevier1995, 1995, pp. 115–123.
[47] A. McCallum, D. Freitag, F.C. Pereira, Maximum Entropy Markov Models for Information Extraction and Segmentation, Icml, 2000, pp. 591–598.
[48] A. Ratnaparkhi, A Maximum Entropy Model for Part-Of-Speech Tagging, Conference on Empirical Methods in Natural Language Processing, 1996.
[49] http://www.abbyy.com. (Accessed 2 February 2023).
[50] R. Baeza-Yates, B. Ribeiro-Neto, Modern Information Retrieval, ACM press, New York, 1999.
[51] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.
[52] J. Chen, J. Yan, B. Zhang, Q. Yang, Z. Chen, Diverse topic phrase extraction through latent semantic analysis, in: Sixth International Conference on Data Mining (ICDM'06), IEEE, 2006, pp. 834–838.
[53] W.W. Cohen, Learning Trees and Rules with Set-Valued Features, vol. 1, AAAI/IAAI, 1996, pp. 709–716. Citeseer.
[54] C. McIntosh, T.G. Purdie, Contextual atlas regression forests: multiple-atlas-based automated dose prediction in radiation therapy, IEEE Trans. Med. Imag. 35 (2015) 1000–1012.
[55] M. Alipour-Vaezi, R. Tavakkoli-Moghadaam, M. Samieinasab, Scheduling the COVID-19 vaccine distribution based on data-driven decision-making methods, Journal of Industrial Engineering and Management Studies 8 (2022) 196–206.
[56] X. Zhu, J. Li, J. Ren, J. Wang, G. Wang, Dynamic ensemble learning for multi-label classification, Inf. Sci. 623 (2023) 94–111.
[57] M. Alipour-Vaezi, A. Aghsami, F. Jolai, Prioritizing and queueing the emergency departments' patients using a novel data-driven decision-making methodology, a real case study, Expert Syst. Appl. 195 (2022), 116568.
[58] C. Cai, H. Lin, H. Wang, Y. Xu, Q. Ouyang, L. Lai, J. Pei, miDruglikeness: subdivisional drug-likeness prediction models using active ensemble learning strategies, Biomolecules 13 (2023) 29.
[59] R. Alizadehsani, J. Habibi, M.J. Hosseini, H. Mashayekhi, R. Boghrati, A. Ghandeharioun, B. Bahadorian, Z.A. Sani, A data mining approach for diagnosis of coronary artery disease, Comput. Methods Progr. Biomed. 111 (2013) 52–61.
[60] R. Alizadehsani, M.J. Hosseini, A. Khosravi, F. Khozeimeh, M. Roshanzamir, N. Sarrafzadegan, S. Nahavandi, Non-invasive detection of coronary artery disease in high-risk patients based on the stenosis prediction of separate coronary arteries, Comput. Methods Progr. Biomed. 162 (2018) 119–127.
[61] R. Alizadehsani, M. Roshanzamir, M. Abdar, A. Beykikhoshk, M.H. Zangooei, A. Khosravi, S. Nahavandi, R.S. Tan, U.R. Acharya, Model uncertainty quantification for diagnosis of each main coronary artery stenosis, Soft Comput 24 (2019) 10149–10160.
[62] R. Alizadehsani, M. Abdar, M. Roshanzamir, A. Khosravi, P.M. Kebria, F. Khozeimeh, S. Nahavandi, N. Sarrafzadegan, U.R. Acharya, Machine learning-based coronary artery disease diagnosis: a comprehensive review, Comput. Biol. Med. 111 (2019) 1–14.
[63] R. Alizadehsani, M.H. Zangooei, M.J. Hosseini, J. Habibi, A. Khosravi, M. Roshanzamir, F. Khozeimeh, N. Sarrafzadegan, S. Nahavandi, Coronary artery disease detection using computational intelligence methods, Knowl. Base Syst. 109 (2016) 187–197.
[64] R. Alizadehsani, M.J. Hosseini, Z.A. Sani, A. Ghandeharioun, R. Boghrati, Diagnosis of Coronary Artery Disease Using Cost-Sensitive Algorithms, 12th International Conference on Data Mining Workshops, 2012, pp. 9–16.
[65] Z. Arabasadi, R. Alizadehsani, M. Roshanzamir, H. Moosaei, A.A. Yarifard, Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm, Comput. Methods Progr. Biomed. 141 (2017) 19–26.
[66] L.E. Raileanu, K. Stoffel, Theoretical comparison between the gini index and information gain criteria, Ann. Math. Artif. Intell. 41 (2004) 77–93.
[67] M. Ringnér, What is principal component analysis? Nat. Biotechnol. 26 (2008) 303–304.
[68] O. Götz, K. Liehr-Gobbers, M. Krafft, Evaluation of Structural Equation Models Using the Partial Least Squares (PLS) Approach, Handbook of Partial Least Squares, Springer, 2010, pp. 691–711.
[69] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008).
[70] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, 3, 2014, pp. 1–15, arXiv preprint arXiv:1412.6980.
[71] E. Osuna, R. Freund, F. Girosi, An Improved Training Algorithm for Support Vector Machines, IEEE Signal Processing Society Workshop, 1997, pp. 276–285.

[72] A. Shoeibi, M. Khodatars, M. Jafari, N. Ghassemi, P. Moridian, R. Alizadehsani, S. H. Ling, A. Khosravi, H. Alinejad-Rokny, H.K. Lam, M. Fuller-Tyszkiewicz, U. R. Acharya, D. Anderson, Y. Zhang, J.M. Gorriz, Diagnosis of brain diseases in fusion of neuroimaging modalities using deep learning: a review, Inf. Fusion 93 (2023) 85–117.
[73] J.H. Joloudari, S. Mojrian, I. Nodehi, A. Mashmool, Z.K. Zadegan, S. K. Shirkharkolaie, R. Alizadehsani, T. Tamadon, S. Khosravi, M.A. Kohnehshari, Application of artificial intelligence techniques for automated detection of myocardial infarction: a review, Physiol. Meas. 43 (2022) 1–22.
[74] A. Liaw, M. Wiener, Classification and regression by randomForest, R. News 2 (2002) 18–22.
[75] R.Z. Nasab, M.R.E. Ghamsari, A. Argha, C. Macphillamy, A. Beheshti, R. Alizadehsani, N.H. Lovell, H. Alinejad-Rokny, Deep learning in spatially resolved transcriptomics: a comprehensive technical view, 2, 2022, pp. 1–39, arXiv preprint arXiv:2210.04453.
[76] P. Moridian, N. Ghassemi, M. Jafari, S. Salloum-Asfar, D. Sadeghi, M. Khodatars, A. Shoeibi, A. Khosravi, S.H. Ling, A. Subasi, Automatic autism spectrum disorder detection using artificial intelligence methods with MRI neuroimaging: a review, Frontiers in Molecular Neuroscience 15 (2022) 1–32.
[77] L.E. Peterson, K-nearest neighbor, Scholarpedia 4 (2009) 1883–1892.
[78] P. Moridian, A. Shoeibi, M. Khodatars, M. Jafari, R.B. Pachori, A. Khadem, R. Alizadehsani, S.H. Ling, Automatic diagnosis of sleep apnea from biomedical signals using artificial intelligence techniques: methods, challenges, and future works, Wiley Interdisciplinary Reviews: Data Min. Knowl. Discov. 12 (2022) e1478.
[79] M. Karami, R. Alizadehsani, K.H. Jahanian, A. Argha, I. Dehzangi, H. Alinejad-Rokny, Revolutionizing genomics with reinforcement learning techniques, arXiv (2023) 1–18.
[80] L.K. Hansen, P. Salamon, Neural Network Ensembles, IEEE Transactions on Pattern Analysis & Machine Intelligence, 1990, pp. 993–1001.
[81] S. Nahavandi, R. Alizadehsani, D. Nahavandi, S. Mohamed, N. Mohajer, M. Rokonuzzaman, I. Hossain, A comprehensive review on autonomous navigation, 3, 2022, pp. 1–48, arXiv preprint arXiv:2212.12808.
[82] J.R. Quinlan, Bagging, Boosting, and C4. 5, AAAI/IAAI, 1996, pp. 725–730.
[83] M. Jafari, A. Shoeibi, M. Khodatars, N. Ghassemi, P. Moridian, N. Delfan, R. Alizadehsani, A. Khosravi, S.H. Ling, Y.-D. Zhang, Automated diagnosis of cardiovascular diseases from cardiac magnetic resonance imaging using deep learning models: a review, 3, 2022, pp. 1–56, arXiv preprint arXiv:2210.14909.
[84] J.H. Joloudari, R. Alizadehsani, I. Nodehi, S. Mojrian, F. Fazl, S.K. Shirkharkolaie, H. Kabir, R.-S. Tan, U.R. Acharya, Resource allocation optimization using artificial intelligence methods in various computing paradigms: a Review, 4, 2022, pp. 1–23, arXiv preprint arXiv:2203.12315.
[85] I. Rish, An Empirical Study of the Naive Bayes Classifier, Workshop on Empirical Methods in Artificial Intelligence, 2001, pp. 41–46.
[86] S.S. Abedini, S. Akhavan, J. Heng, R. Alizadehsani, I. Dehzangi, D.C. Bauer, H. Rokny, A critical review of the impact of candidate copy number variants on autism spectrum disorders, 5, 2023, pp. 1–28, arXiv preprint arXiv:2302.03211.
[87] D. Sharifrazi, R. Alizadehsani, J.H. Joloudari, S. Shamshirband, S. Hussain, Z. A. Sani, F. Hasanzadeh, A. Shoaibi, A. Dehzangi, H. Alinejad-Rokny, CNN-KCL: Automatic Myocarditis Diagnosis Using Convolutional Neural Network Combined with K-Means Clustering, 2020.
[88] M. Roshanzamir, A. Shamsi, H. Asgharnezhad, R. Alizadehsani, S. Hussain, H. Moosaei, A. Mohammadi, U.R. Acharya, H. Alinejad, Quantifying Uncertainty in Automated Detection of Alzheimer's Patients Using Deep Neural Network, 2023.
[89] J.H. Joloudari, S. Hussain, M.A. Nematollahi, R. Bagheri, F. Fazl, R. Alizadehsani, R. Lashgari, BERT-deep CNN: state-of-the-art for sentiment analysis of COVID-19 tweets, 6, 2022, pp. 1–20, arXiv preprint arXiv:2211.09733.
[90] R. Kohavi, Wrappers for Performance Enhancement and Oblivious Decision Graphs, Carnegie-Mellon Univ Pittsburgh Pa Dept Of Computer Science, 1995, pp. 1–40.
[91] R. Alizadehsani, M. Roshanzamir, N.H. Izadi, R. Gravina, H.D. Kabir, D. Nahavandi, H. Alinejad-Rokny, A. Khosravi, U.R. Acharya, S. Nahavandi, Swarm intelligence in internet of medical things: a review, Sensors 23 (2023) 1466.
[92] M.H. Zangooei, J. Habibi, R. Alizadehsani, Disease Diagnosis with a hybrid method SVR using NSGA-II, Neurocomputing 136 (2014) 14–29.
[93] Y.-P. Tang, G.-X. Li, S.-J. Huang, ALiPy: active learning in Python, arXiv (2019) 1–5.
[94] Y.-F. Yan, S.-J. Huang, Cost-effective active learning for hierarchical multi-label classification, in: Proceedings of the 27th International Joint Conference on Artificial Intelligence, AAAI Press, Stockholm, Sweden, 2018, pp. 2962–2968.
[95] D.D. Lewis, W.A. Gale, A Sequential Algorithm for Training Text Classifiers, Springer London, London, 1994, pp. 3–12.
[96] L. Lin, K. Wang, D. Meng, W. Zuo, L. Zhang, Active self-paced learning for cost-effective and progressive face identification, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2018) 7–19.
[97] Z. Wang, J. Ye, Querying discriminative and representative samples for batch mode active learning, ACM Trans. Knowl. Discov. Data 9 (2015) 1–23.
[98] Y.-P. Tang, S.-J. Huang, Self-paced active learning: query the right thing at the right time, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2019, pp. 5117–5124.
[99] K. Konyushkova, R. Sznitman, P. Fua, Learning active learning from data, Adv. Neural Inf. Process. Syst. (2017) 4225–4235.
[100] N. Roy, A. McCallum, Toward optimal active learning through sampling estimation of error reduction, in: Int. Conf. On Machine Learning, Morgan Kaufmann, 2001.
[101] S.-J. Huang, J.-L. Chen, X. Mu, Z.-H. Zhou, Cost-effective active learning from diverse labelers, in: Proceedings of the 26th International Joint Conference on Artificial Intelligence, AAAI Press, Melbourne, Australia, 2017, pp. 1879–1885.
[102] S.-J. Huang, M. Xu, M.-K. Xie, M. Sugiyama, G. Niu, S. Chen, Active feature acquisition with supervised matrix completion, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, ACM, London, United Kingdom, 2018, pp. 1571–1579.
[103] N.A.H. Mamitsuka, Query Learning Strategies Using Boosting and Bagging, Machine Learning: Proceedings of the Fifteenth International Conference (ICML'98), Morgan Kaufmann Pub, 1998.
[104] S.-J. Huang, S. Chen, Z.-H. Zhou, Multi-label active learning: query type matters, in: Proceedings of the 24th International Conference on Artificial Intelligence, AAAI Press, Buenos Aires, Argentina, 2015, pp. 946–952.
[105] S.-J. Huang, R. Jin, Z.-H. Zhou, Active learning by querying informative and representative examples, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2014) 1936–1949.
[106] S. Chakraborty, J. Zhou, V. Balasubramanian, S. Panchanathan, I. Davidson, J. Ye, Active matrix completion, in: 2013 IEEE 13th International Conference on Data Mining, 2013, pp. 81–90.
[107] P. Donmez, J.G. Carbonell, J. Schneider, Efficiently learning the accuracy of labeling sources for selective sampling, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, Paris, France, 2009, pp. 259–268.
[108] S. Huang, Z. Zhou, Active query driven by uncertainty and diversity for incremental multi-label learning, in: 2013 IEEE 13th International Conference on Data Mining, 2013, pp. 1079–1084.
[109] X. Li, Y. Guo, Active Learning with Multi-Label SVM Classification, Twenty-Third International Joint Conference on Artificial Intelligence, 2013.
[110] B. Yang, J.-T. Sun, T. Wang, Z. Chen, Effective multi-label active learning for text classification, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, Paris, France, 2009, pp. 917–926.
[111] V.S. Sheng, F. Provost, P.G. Ipeirotis, Get another label? improving data quality and data mining using multiple, noisy labelers, in: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, Las Vegas, Nevada, USA, 2008, pp. 614–622.
[112] S. Ebert, M. Fritz, B. Schiele, RALF: A Reinforced Active Learning Formulation for Object Class Recognition, Conference on Computer Vision and Pattern Recognition, 2012, pp. 3626–3633.
[113] N.A.H. Mamitsuka, Query Learning Strategies Using Boosting and Bagging, ICML, Morgan Kaufmann Pub, 1998, pp. 1–9.
[114] N. Roy, A. McCallum, Toward optimal active learning through sampling estimation of error reduction, in: Int. Conf. On Machine Learning, Morgan Kaufmann, 2001, pp. 1–12.
[115] S. Huang, Z. Zhou, Active Query Driven by Uncertainty and Diversity for Incremental Multi-Label Learning, International Conference on Data Mining, 2013, pp. 1079–1084.
[116] B. Yang, J.-T. Sun, T. Wang, Z. Chen, Effective multi-label active learning for text classification, in: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, Paris, France, 2009, pp. 917–926.
[117] X. Li, Y. Guo, Active Learning with Multi-Label SVM Classification, Twenty-Third International Joint Conference on Artificial Intelligence, 2013, pp. 1479–1485.
[118] S.-J. Huang, M. Xu, M.-K. Xie, M. Sugiyama, G. Niu, S. Chen, Active Feature Acquisition with Supervised Matrix Completion, ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, ACM, London, United Kingdom, 2018, pp. 1571–1579.
[119] S. Chakraborty, J. Zhou, V. Balasubramanian, S. Panchanathan, I. Davidson, J. Ye, Active Matrix Completion, International Conference on Data Mining, 2013, pp. 81–90.
[120] P. Donmez, J.G. Carbonell, J. Schneider, Efficiently learning the accuracy of labeling sources for selective sampling, in: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, Paris, France, 2009, pp. 259–268.
[121] V.S. Sheng, F. Provost, P.G. Ipeirotis, Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, Las Vegas, Nevada, USA, 2008, pp. 614–622.
[122] R. Alizadehsani, M. Roshanzamir, M. Abdar, A. Beykikhoshk, A. Khosravi, M. Panahiazar, A. Koohestani, F. Khozeimeh, S. Nahavandi, N. Sarrafzadegan, A database for using machine learning and data mining techniques for coronary artery disease diagnosis, Sci. Data 6 (2019) 227.
[123] W. Zhu, N. Zeng, N. Wang, Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS Implementations, Nesug, 2010, pp. 1–9.