Facilitating innovation and knowledge transfer between homogeneous and heterogeneous datasets: Generic incremental transfer learning approach and multidisciplinary studies

Kwok Tai Chui[a,*], Varsha Arya[b,c,d,*], Shahab S. Band[e,*], Mobeen Alhalabi[f], Ryan Wen Liu[g], Hao Ran Chi[h]

[a] Department of Electronic Engineering and Computer Science, School of Science and Technology, Hong Kong Metropolitan University, Ho Man Tin, Kowloon, Hong Kong SAR, China
[b] Insights2Techinfo, India
[c] Asia University, Taiwan
[d] Lebanese American University, Beirut, 1102, Lebanon
[e] Department of Computer Science, National Yunlin University of Science and Technology, Douliou, Taiwan
[f] Department of Computer Science, King Abdulaziz University, Jeddah, Saudi Arabia
[g] Hubei Key Laboratory of Inland Shipping Technology, School of Navigation, Wuhan University of Technology, Wuhan 430063, China
[h] Instituto de Telecomunicações, 3810- 193, Aveiro, Portugal

ARTICLE INFO

ABSTRACT

Open datasets serve as facilitators for researchers to conduct research with ground truth data. Generally, datasets contain innovation and knowledge in the domains that could be transferred between homogeneous datasets and have become feasible using machine learning models with the advent of transfer learning algorithms. Research initiatives are drawn to the heterogeneous datasets if these could extract useful innovation and knowledge across datasets of different domains. A breakthrough can be achieved without the restriction requiring the similarities between datasets. A multiple incremental transfer learning is proposed to yield optimal results in the target model. A multiple rounds multiple incremental transfer learning with a negative transfer avoidance algorithm are proposed as a generic approach to transfer innovation and knowledge from the source domain to the target domain. Incremental learning has played an important role in lowering the risk of transferring unrelated information which reduces the performance of machine learning models. To evaluate the effectiveness of the proposed algorithm, multidisciplinary studies are carried out in 5 disciplines with 15 benchmark datasets. Each discipline comprises 3 datasets as studies with homogeneous datasets whereas heterogeneous datasets are formed between disciplines. The results reveal that the proposed algorithm enhances the average accuracy by 4.35% compared with existing works. Ablation studies are also conducted to analyse the contributions of the individual techniques of the proposed algorithm, namely, the multiple rounds strategy, incremental learning, and negative transfer avoidance algorithms. These techniques enhance the average accuracy of the machine learning model by 3.44%, 0.849%, and 4.26%, respectively.

Introduction

The current big data era has driven technological advancements in computational architecture, platforms, tools, and algorithms to bring ground truth data into valuable information (carrying innovation and knowledge). A recent research trend is to transfer innovation and knowledge via transfer learning algorithm. It has demonstrated its effectiveness in many smart city applications, such as bio-signal processing (Wan et al., 2021), cross-lingual transfer (Do et al., 2020), demand response for the electricity grid (Peirelinck et al., 2022), sentiment analysis (Chan et al., 2022), and intelligent transportation (Li et al., 2022). From the aspect of the machine learning model, the trained model in the source domain helps to fine-tune of the hyperparameters in the target domain. Typically, one may recommend employing transfer learning when the source and target domains are identical or similar domains. These datasets can be considered to be homogeneous datasets. However, in some disciplines, large-scale

* Corresponding authors.
E-mail addresses: jktchui@hkmu.edu.hk (K.T. Chui), varshaarya21@gmail.com (V. Arya), shamshirbands@yuntech.edu.tw (S.S. Band).

List 1 Summary of the acronyms and symbols

Acronym/Symbol	
C	classifier
D	discriminator
F	extractor for features
F_N	false negative rate
F_P	false positive rate
$L_{Adversarial}$	standard loss function of the generative adversarial network
$L_{Classification}$	loss function for the classification in source and target samples
L_{total}	loss function for the subtraction of the entropy of the label distribution from the entropy of the input
MITL-NTA	multiple rounds multiple incremental transfer learning and negative transfer avoidance
MRI	magnetic resonance imaging
N	number of equal-size subsets
Q	cluster for common feature
R	cluster for domain specific feature
S	association matrix
SVM	support vector machine
T	assignment matrix for sample cluster
T_N	true negative rate
T_P	true positive rate
X_{source}	feature matrix of the source domain
$\{\varphi_S^i\}_{i=1}^N$	feature map
$\{\omega_i\}_{i=1}^N$	set of weights
λ	hyperparameter controlling the loss functions
w_{sen}	weighting factor for the sensitivity
w_{spe}	weighting factor for the specificity

homogeneous datasets for robust deep learning may not be available due to various issues such as privacy, rare cases, high cost, and time-consuming. An emergent research initiative has thus been proposed in recent years to investigate the feasibility and benefits of transfer learning in heterogeneous datasets. The datasets are comprised of data from different domains. For instance, similarities exist in the problems of image recognition between different domains, such as the image noise. Nevertheless, it is a major challenge when a large level of dissimilarity is experienced between different domains. The machine learning model is prone to be corrupted by dissimilarities, which is a well-known issue of negative transfer (i.e., the performance of the target model decreases after transfer learning). In particular, the smart city applications cover a wide variety of disciplines in which one can find datasets with high similarities and datasets with very different data modalities. Regarding the methodology of the transfer learning algorithm, a generic approach is desired to address the nature of infinitely many combinations between the source and target domains. The algorithm does not need to converge to the global optimal solution at the very beginning, as long as the model performance improves when new data are available to update the trained target model.

The organization of this paper is structured as follows. The three subsections of Section 1 comprise (i) a literature review on incremental learning, transfer learning, and hybrid incremental learning and transfer learning algorithms; (ii) the limitations of the existing works; and (iii) the research contributions of the proposed algorithm. Section 2 presents a brief summary of the 15 benchmark datasets in 5 selected disciplines, along with the design and formulations of the proposed algorithm. The performance evaluation of the proposed

algorithm, and comparison between the proposed algorithm and existing works are conducted in Section 3. Ablation studies follow in Section 4. Finally, the conclusion summarizes the major results and contributions of our work and discusses several future research directions.

Literature review

Existing works focus on standalone incremental learning, standalone transfer learning, and hybrid incremental learning and transfer learning algorithms. Readers who are interested in the overview of the topics, are referred to the latest review articles on incremental learning (Belouadah et al., 2021; Delange et al., 2022) and transfer learning (Bashath et al., 2022; Pinto et al., 2022).

Standalone incremental learning

Incremental learning is generally featured with good adaptation to new training data. The trained model is updated where necessary to avoid duplicating features, thus enhancing the generalization. A fuzzy three-way algorithm was proposed to progressively update the machine learning model via incremental learning (Yuan et al., 2022). The difference between the new object and the centre of the class was used to determine if updating is necessary. The model was evaluated using 10 datasets with 70% of the data as the initial training phase where further update was proceeded with 3% of new training data in each iteration. The enhancement of the average accuracy via incremental learning was from 83.0% to 84.9%. The performance of the model was compared with k-nearest neighbour, fuzzy k-nearest neighbour, intuitionistic fuzzy k-nearest neighbour, and finite element neural network, with percentage improvement by 8.15−9.17%. Another study (Tabassum et al., 2021) presented a distributed network architecture comprised of generative, bridge, and classifier networks for the network attack detection. A positive constraint-based autoencoder was used for feature extraction to select relevant features. The general idea was to combine features with high similarity and minimize the residual of the objective function. Two datasets with 60% of the training data initialized the model and further fine-tuned the model with the rest of the data via the bridge network. The average accuracies before and after applying transfer learning were 81.8% and 95.2%, respectively, with a 16.4% improvement. In Kim et al. (2021), electrocardiogram-based authentication was achieved by a support vector machine model. Initially, a 10-min segment of the electrocardiogram signal was used to train the model which was updated by five 10-min segments on five consecutive days (one each per day). The feature space remained throughout the incremental learning process with the update on the decision boundaries of the SVM classifier. The performance evaluation reflected an accuracy improvement from 61.3% to 87.6%. An unsupervised Fisher's exact test-based supervised random forest algorithm was proposed for medical association rule discovery (Sánchez-de-Madariaga et al., 2022). The feature extraction was enhanced by combining supervised learning and unsupervised learning. Incremental learning was applied and evaluated with a grid search of [5, 200] on the seed size. An improvement of 0.61 to 0.80 was observed in the area under the ROC curve. A separable and compact feature learning algorithm was proposed in Li & Huang (2022) for vehicle detection. The novelty in this work was to freeze the feature selector during incremental learning using decoupled learning for compact and separable features. Under the optimal regularization parameter settings, the accuracy was improved from 78.1% to 80.2% in 10 iterations and to 80.7% in 20 iterations.

Standalone transfer learning

In Alzubaidi et al. (2021), a deep convolutional neural network was initially trained with 0.2 million unlabelled images of skin cancer. The model was transferred to a target domain with 33k labelled

images of skin cancer. Varying filter sizes of parallel convolutional layers were used for feature extraction of features with different granularity to enhance the overall feature representation. Transfer learning facilitated the accuracy improvement from 85.3% to 97.5%. A hybrid convolutional neural network and bidirectional long short-term memory were proposed to classify normal, viral pneumonia, and COVID-19 candidates in Aslan et al. (2021). Automatic feature extraction was achieved with a convolutional neural network. The model was fine-tuned using a pre-trained AlexNet via transfer learning. It achieved an accuracy of 98.7%. Another work (Han et al., 2021) implemented a multiclass domain discriminators to support adversarial training network-based feature domain adaptation for machinery fault detection. The optimization problems were to minimize the domain loss and label loss. The model was enhanced using a 2-round transfer learning, and the evaluation of four tasks showed an improvement in accuracy from 82.2% to 85.3% to 92.7%. A fine-grained recurrent neural network with transfer learning was proposed in Hua et al. (2022) for the estimation of the energy consumption of electric vehicles. The spatiotemporal features were iteratively transferred to the target model. Transfer learning reduced the mean square error, root-mean-square error, and mean average error from 4.89×10^7 to 4.23×10^7, from 6.99×10^3 to 6.50×10^3, and from 0.455×10^3 to 0.365×10^3, respectively. Transfer learning and adversarial networks were used to identify cross-lingual offensive speech (Shi et al., 2022). The unsupervised cross-lingual mapping facilitated the sharing of the language invariant features between languages. Four languages namely Greek, Turkish, Arabic, and Danish were considered for analysis. Ablation studies revealed the improvement of accuracy from 0.874 to 0.878, from 0.874 to 0.881, from 0.918 to 0.926, and from 0.919 to 0.932, in the four languages, respectively.

Hybrid incremental learning and transfer learning

A hybrid incremental learning and transfer learning algorithm, also named the incremental transfer learning algorithm, has received attention in recent years. It takes advantage of both incremental learning and transfer learning. Li et al. (2021) built a non-adaptive support vector machine (SVM) for electromyography-based recognition. It was initially built with 660 electromyography samples. It was then incrementally transferred to three target models using the variants of SVM, namely, TrAdaBoost-incremental SVM, TrAdaBoost SVM, and incremental SVM, using 2640 samples. Three approaches, namely, particle swarm optimization, sequential forward search, and no feature selection, were employed for feature extraction. With the aid of transfer learning, the incremental learning-based model enhanced the accuracy from 75% to 94%, 88%, and 95% for the three SVM algorithms. To recognize retinal pathologies, a pre-trained ImageNet was transferred to a Bayesian deep learning-based model (Hassan et al., 2021). The encoder of the ResNet utilized a fusion

mechanism for residual features so that useful features could be retained during the generation of latent vectors. Incremental learning was introduced to further update the model with 6 benchmark datasets. The accuracy was enhanced from 96% to 98%. In Martin et al. (2022), an evolution-aware model shifting algorithm was proposed to build a prediction model for the Linus kernel size. It was based on the pre-trained model using gradient boosting. The authors noted that the incremental learning process will lead to continual deletion of existing features and addition of new features in each round of iteration. The analysis showed that approximately 95% of the features remained in the feature space as top features in the evaluation of feature importance. Incremental learning was introduced to enhance the model, with accuracy improvement from 92.7% to 94.2%. In Francisco et al. (2022), researchers employed incremental transfer learning to improve the convolutional neural network for designing a rule checker. At most 10% of the weights could be updated to prevent the target model from learning too much irrelevant information from the source model, as negative transfer. The loss function dropped by approximately 0.15 in most cases. A dual-branch aggregation network with incremental transfer learning was proposed in Chen et al. (2022) for fault gearbox detection. A major portion of the network layers was frozen so that fine-tuning was performed in a small portion of the network layers to avoid negative transfer. Various scenarios showed that significant improvement was observed in the accuracy, from 77.1%, 81.4%, and 95.9% to 99.89%.

Comparisons between standalone incremental learning, standalone transfer learning, and hybrid incremental learning and transfer learning

Table 1 compares the three approaches namely standalone incremental learning, standalone transfer learning, and hybrid incremental learning and transfer learning, with a discussion of the technical limitations.

Limitations of existing works

The abovementioned existing works show the recent research trends in enhancing the performance of machine learning models using incremental learning and transfer learning. Although it may require additional computational power and training time for model updates, it is acceptable given that high-performance computing devices and platforms have become more affordable. There are some research limitations of the existing works that require investigation:

- Researchers have mainly focused on transfer learning with homogeneous datasets between the source and target domains. Limited studies (Day & Khoshgoftaar, 2017) have analysed the impact of transfer learning between heterogeneous datasets.

Table 1
Comparison between standalone incremental learning, standalone transfer learning, and hybrid incremental learning and transfer learning.

Technique	Purpose of formulation	Technical limitations
Standalone incremental learning	It aims at continually updating the machine learning model when more input data becomes available. One can also take advantage of this approach for a large-scale dataset where traditional approach requires huge computational resources.	The portion of new data to be selected to update the existing model requires customization and cannot remain fixed.
Standalone transfer learning	It aims at transferring knowledge from a pre-trained model of an application (source domain) to a target model of an application (target domain).	In principle, the source and target domains can be similar or different, to some extents. Selecting appropriate source domain is important to reduce the challenge of negative transfer from the source domain to the target domain.
Hybrid incremental learning and transfer learning	It aims to gradually transfer knowledge from the source domain to the target domain with the inclusion of new data to reduce the impact of negative transfer and provide flexibility to update the trained target model.	The complexity of the algorithm is higher compared with standalone incremental learning and standalone transfer learning algorithms. It also introduces more hyper-parameters by merging the algorithms.

- In transfer learning, migrating irrelevant information from the source domain to the target domain may deteriorate the performance of the target model. Thus, the design of transfer learning should be reformulated.
- The impact of multiple incremental transfer learning in both homogeneous and heterogeneous datasets on enhancing the performance of the target model has not yet been confirmed. Some of the existing works suggested that 2-round transfer learning may help increase the performance of the target model.

Research contributions of our work

A multiple rounds multiple incremental transfer learning and negative transfer avoidance (MITL-NTA) algorithm is proposed as a generic approach to transfer innovation and knowledge from the source domain to the target domain. The key research contributions are as follows:

- A negative transfer avoidance algorithm is proposed to lower the risk of transferring irrelevant information from the source domain to the target domain, particularly between heterogeneous datasets. The objective functions of the domain, instance, and feature, as a three-step approach, are optimally solved to reduce the impact of negative transfer. Therefore, positive transfer and thus the performance enhancement of the target model can be ensured.
- An in-depth analysis and discussion on the multiple incremental learning (multiple rounds strategy and incremental learning) are presented. As a consequence, adding more source heterogeneous datasets and model training via incremental learning can benefit the performance enhancement of the target model and enhance the ability of positive transfer. The complexity of the problem with selecting relevant samples is lowered when incremental learning is adopted, where a smaller dataset is introduced to update the model in each iteration.
- A detailed multidisciplinary analysis of the proposed algorithm was performed with 15 benchmark datasets across 5 disciplines (lung cancer, breast cancer, daily activities, fruits, and handwriting) to create scenarios for analysis via homogenous (between lung cancer and breast cancer) and heterogeneous datasets (between cancer datasets and the other three datasets). This reveals the ability and effectiveness of the proposed algorithm to handle very different source and target domains. At the same time, the restriction on the selection of the source dataset is relaxed. The implication is to perform transfer learning with any heterogeneous dataset when the performance of the target model is not satisfactory with traditional transfer learning with homogeneous datasets and small-scale datasets;
- The proposed algorithm enhances the accuracy by 4.35% compared with existing works. Ablation studies were also conducted to analyse the contributions of the individual techniques: multiple rounds strategy, incremental learning, and negative transfer avoidance algorithms. They show that the proposed algorithm enhances the average accuracy by 3.44%, 0.849%, and 4.26%, respectively.

Benchmark datasets and methodology

The 15 benchmark datasets in 5 disciplines are briefly summarized here. It is followed by the methodology of the proposed MITL-NTA algorithm, which comprises the incremental transfer learning algorithm, negative transfer avoidance algorithm, and strategy for multiple rounds of incremental transfer learning with negative transfer avoidance.

Benchmark datasets

Five disciplines were considered: (i) lung cancer with computerized tomography (CT) scans. Lung cancer is the leading cause of death amongst all types of cancers and causes 1.8 million deaths each year (Ferlay et al., 2020; Machacha et al., 2021); (ii) breast cancer with magnetic resonance imaging (MRI) scans, which is the most common type of cancer, and more than 2.26 million new cases are reported each year (Ginsburg et al., 2020; Masud et al., 2021); (iii) daily activities with body sensing data, in which recognising human activities provides various benefits in the understanding of health, physiological, and psychological status, as well as personal identity; (iv) fruits with images, in which the recognition of fruits is a well-known and one of the earliest problems in deep learning applications in image processing; and (v) handwriting with images, in which handwriting or hand text recognition focuses on the interpretation of handwritten inputs to computers.

In each of the five disciplines, three benchmark datasets were retrieved for performance evaluation and analysis of the proposed MITL-NTA algorithm. The summary of the details of a total of 15 benchmark datasets including the discipline, the released year, the number of Google Scholar citations (up to mid-June 2022), the size of the dataset, and the number of samples, is shown in Table 2.

Incremental transfer learning

Incremental transfer learning subdivides transfer learning into multiple subproblems so that the extent of potential negative transfer (arising from irrelevant training samples) can be reduced. The conceptual diagram of the incremental transfer learning algorithm is presented in Fig. 1. Solid lines relate to incremental learning whereas dotted lines relate to transfer learning. The illustration is based on two datasets (Dataset 1 from the source domain and Dataset 2 from the target domain), which are divided into N equal-sized subsets. The idea will be further extended in section 2.4. The value of N is closely related to the computational complexity in each iteration (from 1 to N), ability to learn new knowledge, and ability to mistake irrelevant information (negative transfer), which require further analysis which is presented in Section 3. Therefore, we have a source domain dataset, namely, Dataset 1 = {Dataset 1.1, Dataset 1.2, . . ., Dataset 1.N} and target domain dataset, namely, Dataset 2 = {Dataset 2.1, Dataset 2.2, . . ., Dataset 2.N}.

Here are the key steps to summarize the incremental transfer learning algorithm to build a target model (Model 2.N).

Step 1: Choose two datasets for which knowledge is transferred from one domain to another. Name Dataset 1 for the source domain and Dataset 2 for the target domain.

Step 2: Divide Dataset 1 and Dataset 2 into N equal-sized subsets.

Step 3: Train Model 1.1 using the first subset of Dataset 1, namely, Dataset 1.1.

Step 4: Train Model 2.1 using pre-trained Model 1.1 with knowledge transfer (including the negative transfer avoidance algorithm) and Dataset 2.1.

Step 5: Incrementally update Model 1.1 to Model 1.2 using Dataset 1.2 to obtain Model 1.2;.Step 6: Incrementally update Model 2.1 using Dataset 2.2 and pre-trained Model 1.2 with knowledge transfer (including negative transfer avoidance algorithm) to obtain Model 2.2. Step 7 Repeat Steps 5 and 6 until the last subsets of Dataset 1 Dataset 1.N) and Dataset 2 (Dataset 2.N) are used to obtain the final pre-trained model (Model 1.N) and final target model (Model 2.N).

Table 2
Summary of details of benchmark datasets.

| Dataset	Discipline	Released year	Number of google scholar citations	Size of the dataset	Number of samples
NSCLC-Radiomics-Genomics (Aerts et al., 2014)	Lung cancer	2014	3449	6.6 GB	13,482
SPIE-AAPM Lung CT Challenge (Armato III et al., 2015)	Lung cancer	2015	52	12.1 GB	22,489
LungCT-Diagnosis (Grove et al., 2015)	Lung cancer	2015	232	2.5 GB	4682
QIN-Breast (Li et al., 2015)	Breast cancer	2015	126	11.3 GB	100,835
QIN Breast DCE-MRI (Huang et al., 2014)	Breast cancer	2014	120	8.4 GB	76,328
Breast-MRI-NACT-Pilot (Newitt & Hylton, 2016)	Breast cancer	2016	16	19.5 GB	99,314
Daily and Sports Activities Dataset (Altun & Barshan, 2010)	Daily activities	2010	318	163 MB	9120
Activities of Daily Living Recognition Using Binary Sensors Dataset (Ordóñez & Sanchis, 2013)	Daily activities	2013	266	75.9 MB	2747
Smartphone-Based Recognition of Human Activities and Postural Transitions Dataset (Reyes-Ortiz et al., 2016)	Daily activities	2016	546	1.2 GB	10,929
Fruits 360 (Mureșan & Oltean, 2017)	Fruits	2017	270	1.38 GB	90,483
Fruit Image Data set (Marko, 2013)	Fruits	2013	10	428 MB	971
Fruits-262 (Minuț & Iftene, 2021)	Fruits	2021	3	6.59 GB	225,640
THE MNIST DATABASE of handwritten digits (LeCun et al., 1998)	Handwriting	1998	46,389	11.6 MB	70,000
GNHK: A Dataset for English Handwriting in the Wild (Lee et al., 2021)	Handwriting	2021	1	963.2 MB	687
The Chars74K dataset (De Campos et al., 2009)	Handwriting	2009	589	192.4 MB	74,107

It is worth noting that the design in Fig. 1 can be adjusted with nonequal-size subsets (requiring an update to Step 2); for instance, the future round of iteration may use a larger portion of the dataset so that the intake subset can retain a certain ratio (or percentage) compared with the total portion of the dataset being used in previous iterations. In addition, an initiative is proposed to the design that can be unrestrictedly applied to online incremental transfer learning when the dataset is ever-growing. The update will be executed when sufficient new data have gathered (fulfilling the predefined ratio). The iteration may be terminated when the optimal result of the target model has been achieved or negative transfer is observed (with the deterioration of the accuracy of the target model).

Negative transfer avoidance algorithm

Trivially, Dataset 1 and Dataset 2 are not identical datasets, or else they should be merged to train a single model. The degree of similarity between the homogeneous datasets is much larger than that between the heterogeneous datasets. Even though identical applications are considered, the datasets may not be homogeneous given the heterogeneity of sensors, devices, subjects (region-wise), etc. The issue of negative transfer from the source domain to the target domain becomes more severe in heterogeneous datasets. Negative transfer can be easily observed when the performance of the target model with transfer learning is less than that without transfer learning (Zhang et al., 2022). A negative transfer avoidance algorithm is thus introduced to leverage the ability of the incremental transfer learning algorithm to transfer relevant knowledge. The negative transfer avoidance algorithm is an essential part of the knowledge blocks in Fig. 1.

A number of feasible approaches exist to design a negative transfer avoidance algorithm. Three major ideas have been proposed in the literature: (i) redefining the objective functions such that the performance of the model with transfer learning will always be better than that without transfer learning (Li et al., 2019; Yoon & Li, 2018); (ii) employing transitive transfer learning, also known as distant
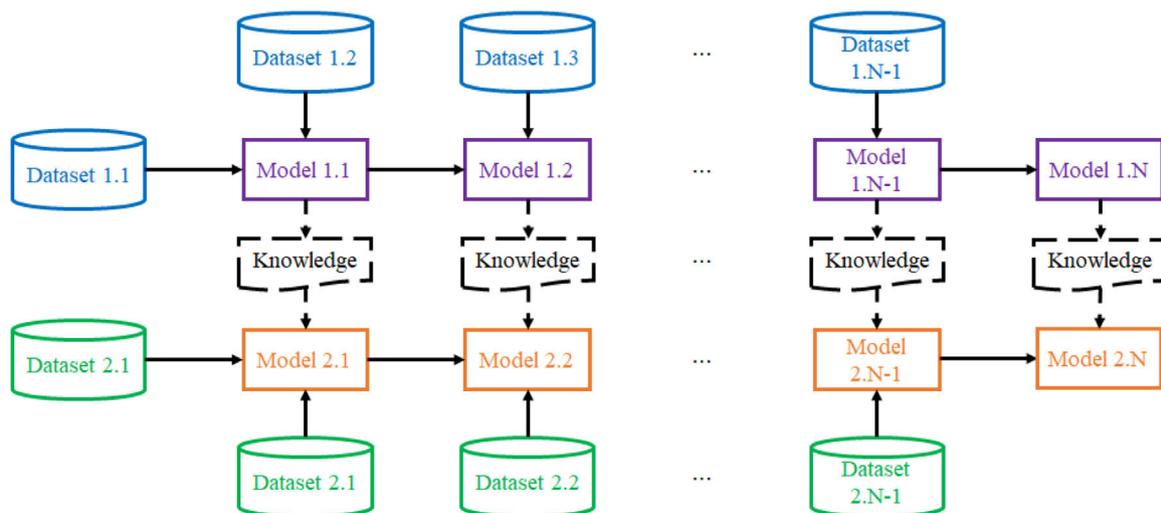
Fig. 1. Incremental transfer learning algorithm.

transfer learning (Niu et al., 2021; Ren et al., 2022); and (iii) enhancing the transferability of the source domain by considering data quality at various levels, such as the feature, instance, and domain levels (Peng et al., 2020; Zuo et al., 2021).

The first approach is not feasible with regard to the multiple rounds of MITL-NTA in later stages because it heavily relies on domain knowledge (too many domains to be considered) for objective functions and high computational power. Next, transitive transfer learning is less capable of negative transfer avoidance. Thus, this paper considers the third approach in the design of the MITL-NTA algorithm. In existing works, researchers have considered the enhancement of transferability at one of the levels (domain, instance, or feature) (Ahmed et al., 2021; Wang et al., 2019). To achieve robust negative transfer avoidance, a three-level transferability enhancement scheme based on multi-objective optimization using NSGA-III is proposed (Fig. 2). It will be incorporated into the knowledge blocks of the incremental transfer learning algorithm.

At the domain level, the minimization problem is formulated as:

$$\min_{\{\varphi_S^i\}_{i=1}^N, \{\omega_i\}_{i=1}^N} L_{total} \text{s.t.} \qquad \omega_i \geq 0 \ \forall\, i \in \{1, 2, \ldots, N\} \text{ and } \sum_{i=1}^N \omega_i = 1 \tag{1}$$

where $\{\varphi_S^i\}_{i=1}^N$ is the feature map, $\{\omega_i\}_{i=1}^N$ is the set of weights, and L_{total} is the loss function for the subtraction of the entropy of the label distribution from the entropy of the input. Thus, the total loss function can be expressed as:

$$L_{total} = H(Y) + H(X) \tag{2}$$

$$L_{total} = \sum_i P(Y_i)i(Y_i) + \sum_j P(X_j)i(X_j) \tag{3}$$

$$L_{total} = \sum_i P(Y_i)\log_2 P(Y_i) + \sum_j P(X_j)\log_2 P(Y_j) \tag{4}$$

with $P(Y_i)$ as the probability of the label Y_i, $i(Y_i)$ as the self-information of Y_i, $P(X_j)$ as the probability of input X_j, and $i(X_j)$ as the self-information of X_j.

At the instance level, a generative adversarial network is considered with the objective function as follows:

$$\underset{F,C}{\arg\min}\, \underset{F,D}{\arg\max}\, L_{Classification}(F, C) - \lambda L_{Adversarial}(F, D) \tag{5}$$

where F is the extractor for features, C is the classifier, D is the discriminator, $L_{Classification}$ is the loss function for the classification in source and target samples, $L_{Adversarial}$ is the standard loss function of the generative adversarial network, and λ is the hyperparameter

controlling the loss functions. $L_{Classification}$ and $L_{Adversarial}$ are defined as:

$$L_{classification}(F, C) = E_{x_s, y_s \sim S}[log(C(F(x_s))), y_s]$$
$$+ E_{x_l, y_l \sim T_L}[log(C(F(x_l))), y_l] \tag{6}$$

$$L_{Adversarial}(F, D) = E_{x_s \sim P_S(X)}[\log(1 - D(F(x_s)))]$$
$$+ E_{x_u \sim P_T(X)}[\log(D(F(x_u)))] \tag{7}$$

where x_s and y_s are the input and output of the source domain S, respectively; x_l and y_l are the input and output of the labelled target domain TL, and x_u is the sample from the unlabelled target domain.

At the feature level, a non-negative tri-factorization approach can be used to formulate the objective function:

$$\min_{Q,R,S,T \geq 0} \| X_{source} - [Q, R]ST^T \| \tag{8}$$

where X_{source} is the feature matrix of the source domain, Q is the cluster for common features, R is the cluster for domain specific features, S is the association matrix, and T is the assignment matrix for the sample cluster.

Strategy for multiple rounds MITL-NTA algorithm

The formulation of the MITL-NTA algorithm (Fig. 1) is between two datasets. In practice, the target model may not achieve satisfactory performance or a perfect result (for mission-critical applications) after one time transfer learning process. Hence, an extension is made toward a multiple rounds MITL-NTA algorithm with multiple datasets.

Fig. 3 shares the workflow of the prioritization algorithm. To prioritize the datasets to be transferred to the target model, domain similarity estimation is performed between the available datasets. The measures are based on a test performance-based approach: Similar datasets will be observed when the source model yields good performance using the data from the target domain. Thus, the performance of various potential source models with the target dataset is ranked in descending order. The source model with the highest similarity will serve as the first model to be transferred. Usually, homogeneous datasets will share higher similarities and thus contribute to transfer learning in earlier iterations compared with heterogeneous datasets. The higher the similarities between datasets, the lower the extent of negative transfer from the source domain to the target domain. In regard to the multiple rounds of the MITL-NTA algorithm (Fig. 4), for each iteration, incremental transfer learning between the source domain and target domain is performed (using
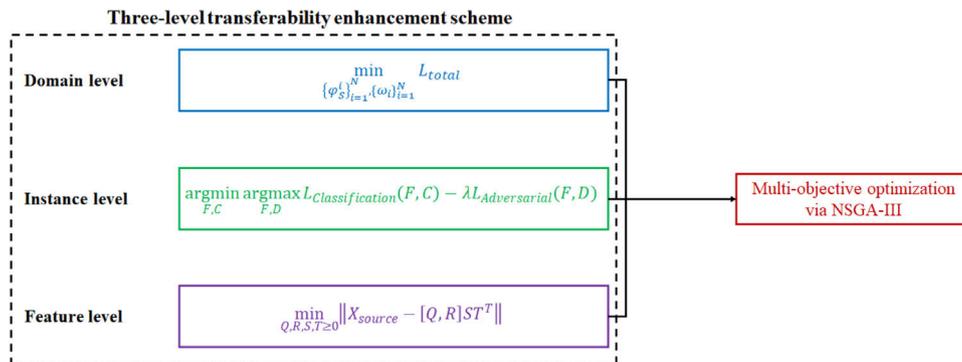

Fig. 2. Three-level transferability enhancement scheme based on multi-objective optimization.
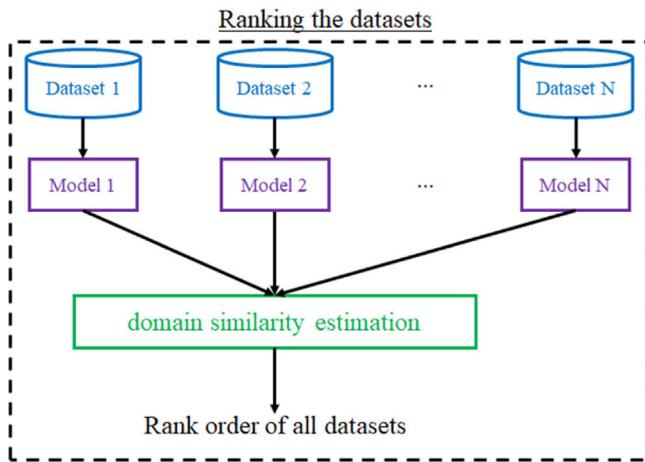
Fig. 3. Prioritization algorithm using domain similarity estimation.

the workflow in Fig. 1). The iterations continue until all available datasets have been used or the performance of the target model starts to drop (usually when there is no similarity between the source domain and the target domain). Taking 15 benchmark datasets as an example, the maximum number of iterations (i_{max}) turns out to be 14 to build a target model.

Performance evaluation and comparison

Since the focus of this research is not on feature extraction and classification algorithms, a convolutional neural network is employed for automatic feature extraction and classification. For example, the cancer-related target domain adopts the following CNN architecture: the 1st layer is a 2D convolutional layer with kernel size 64; the 2nd layer is a maximum pooling layer; the 3rd layer is a 2D convolutional layer with kernel size 128; the 4th layer is a maximum pooling layer; the 5th layer is a 2D convolutional layer with kernel size 256; the 6th

Fig. 4. Multiple rounds MITL-NTA algorithm.

layer is a maximum pooling layer; the 7th layer is a fully connected layer; and the 8th layer is the output layer. Notably, the convolutional layers have a filter size and stride of 3 × 3 and 1, respectively, whereas the maximum pooling layers have a filter size and stride of 2 × 2 and 2, respectively. With sufficient available labelled samples in the 15 benchmark datasets, as a trade-off between the computation time and the evaluation of model overfitting, 2-fold cross-validation is chosen in the performance evaluation and analysis instead of 5-fold (Chen et al., 2021; Chui, 2022) or 10-fold cross-validation (Almomani et al., 2022; Dwivedi et al., 2021). With regard to incremental learning, the training was divided into 10 iterations.

Performance evaluation of the multiple rounds MITL-NTA algorithm

The sensitivity, specificity, accuracy, and number of rounds of MITL-NTA were recorded for all 15 target models. The results are summarized in Table 3. The multiple rounds of the MITL-NTA algorithm achieve significant enhancement in improving sensitivity, the specificity, and accuracy of the target models. Details will be presented in the ablation studies in Section 4. The sensitivity, specificity, precision, F-Score, and accuracy are defined as follows (Hammad et al., 2021; Sedik et al., 2021):

$$Sensitivity = \frac{T_P}{T_P + F_N} \tag{9}$$

$$Specificity = \frac{T_N}{T_N + F_P} \tag{10}$$

$$Precision = \frac{T_P}{T_P + F_P} \tag{11}$$

$$F - Score = \frac{T_P}{T_P + 0.5(F_P + F_N)} \tag{12}$$

$$Accuracy = w_{sen}(Sensitivity) + w_{spe}(Specificity) \tag{13}$$

where T_P is the true positive rate, F_N is the false negative rate, T_N is the true negative rate, F_P is the false positive rate, w_{sen} is the weighting factor for the sensitivity, and w_{spe} is the weighting factor for the specificity, of the model.

Several key observations are:

- The data types of 12 datasets (except three datasets from daily activities) are images where similarities exist regardless of the domain because of the presence of image noise. Multiple rounds of the MITL-NTA algorithm were feasible across heterogeneous datasets in these datasets, which can be observed by the high value in the number of rounds.
- The models with 6 datasets related to cancers received higher values in the number of rounds because cancer images share similarities to a larger extent.
- Multiple rounds of the MITL-NTA algorithm were limited to the models with 3 daily activities. This suggests that if the data types between input samples are not similar, transfer learning becomes less feasible.
- The descending order of the average number of rounds in five disciplines is lung cancer, breast cancer, fruits, handwriting, and daily activities.

With regard to the cost and computational complexity of the proposed algorithm, the base of the classification model is a convolutional neural network that has cost and computational complexity comparable to existing works (Alencastre-Miranda et al., 2021; Roy et al., 2020; Wang et al., 2020). Since a multiple rounds strategy is adopted in the proposed algorithm, the overall cost and computational complexity are approximately equal to the sum of the
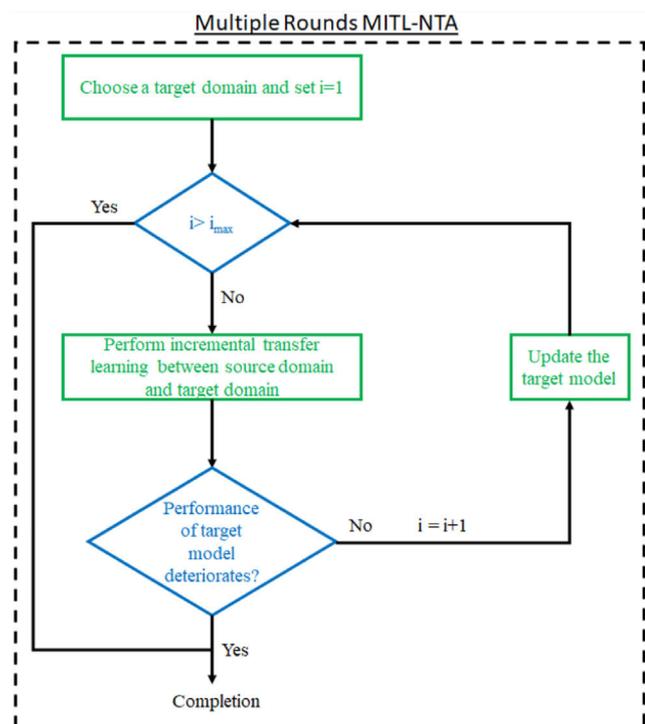
Table 3
Results of the 15 target models using the multiple rounds MITL-NTA algorithm.

Target Dataset/Model	Discipline	Sensitivity (%)	Specificity (%)	Precision (%)	F-Score (%)	Accuracy (%)	Number of rounds
NSCLC-Radiomics-Genomics (Aerts et al., 2014)	Lung cancer	98.3	97.9	98.1	98.2	98.2	8
SPIE-AAPM Lung CT Challenge (Armato III et al., 2015)	Lung cancer	98.7	98.1	98.3	98.5	98.4	9
LungCT-Diagnosis (Grove et al., 2015)	Lung cancer	98.0	97.4	97.6	97.8	97.7	7
QIN-Breast (Li et al., 2015)	Breast cancer	99.0	99.2	99.0	99.0	99.1	8
QIN Breast DCE-MRI (Huang et al., 2014)	Breast cancer	98.6	98.0	98.2	98.4	98.4	8
Breast-MRI-NACT-Pilot (Newitt & Hylton, 2016)	Breast cancer	98.4	97.7	97.9	98.2	98.0	7
Daily and Sports Activities Dataset (Altun & Barshan, 2010)	Daily activities	91.3	92.0	91.5	91.4	91.5	2
Activities of Daily Living Recognition Using Binary Sensors Dataset (Ordónez & Sanchis, 2013)	Daily activities	90.6	90.1	90.2	90.4	90.3	2
Smartphone-Based Recognition of Human Activities and Postural Transitions Dataset (Reyes-Ortiz et al., 2016)	Daily activities	92.4	91.6	91.8	92.1	92.1	2
Fruits 360 (Mureşan & Oltean, 2017)	Fruits	98.8	98.1	98.3	98.6	98.4	5
Fruit Image Dataset (Marko, 2013)	Fruits	95.3	94.6	94.8	95.1	94.9	4
Fruits-262 (Minuţ & Iftene, 2021)	Fruits	99.0	98.5	98.7	98.8	98.8	6
THE MNIST DATABASE of handwritten digits (LeCun et al., 1998)	Handwriting	94.3	93.4	93.6	94.0	93.8	4
GNHK: A Dataset for English Handwriting in the Wild (Lee et al., 2021)	Handwriting	92.5	91.8	92.1	92.3	92.2	3
The Chars74K dataset (De Campos et al., 2009)	Handwriting	96.8	96.0	96.3	96.5	96.5	4

individual multiple incremental transfer learning and negative transfer avoidance algorithms.

Performance comparison with existing works

The accuracies of the proposed work and existing works for the models with 15 benchmark datasets are summarized in Table 4. The key observations are highlighted as follows:

- The average accuracies in the five disciplines for our work, (Li et al., 2021), (Hassan et al., 2021), and (Martin et al., Early Access), are (98.1, 92, 93.5, 92.7)% for lung cancer, (98.5, 91.9, 94.2, 93.1)% for breast cancer, (91.3, 88.1, 89.6, 88.9)% for daily activities, (97.4, 93.6, 95.1, 94.2)% for fruits, and (94.2, 89.9, 91.1, 90.5)% for handwriting;
- Compared with Li et al. (2021), Hassan et al. (2021), and Martin et al. (Early Access), our work improves the average accuracy by (6.63, 4.92, 5.83)% for lung cancer, (7.18, 4.56, 5.80)% for breast cancer, (3.63, 1.90, 2.70)% for daily activities, (4.06, 2.42, 3.40)% for fruits, and (4.78, 3.40, 4.09)% for handwriting. The major reasons for the achievement by our work are the multiple rounds

strategy, incremental learning, and negative transfer avoidance algorithms. This will be analysed in Section 4.

Ablation studies

To evaluate the effectiveness of the components of the proposed multiple rounds MITL-NTA algorithm, ablation studies were conducted on multiple rounds strategy, incremental learning, and negative transfer avoidance algorithms.

Multiple rounds strategy

Fig. 5 compares the accuracy of the MITL-NTA algorithm with and without the multiple rounds strategy across all benchmark datasets. To provide clear visualization, the names of the datasets were renamed based on the concatenation of the discipline with _1, _2, and _3, according to the order of introduction in the previous sections, Tables 2, and 3. The range of improvement in the five disciplines is 3.72−5.02% for lung cancer, 3.59−5.24% for breast cancer, 1.23−1.78% for daily activities, 3.14−4.44% for fruits, and 2.44−3.88% for handwriting. The improvement of the average accuracy in the five

Table 4
Accuracies of the proposed work and existing works using different benchmark datasets.

| Target Dataset/Model	Discipline	Accuracy (%)			
		Proposed work	(Li et al., 2021)	(Hassan et al., 2021)	(Martin et al., Early Access)
NSCLC-Radiomics-Genomics (Aerts et al., 2014)	Lung cancer	98.2	92.4	93.7	93.0
SPIE-AAPM Lung CT Challenge (Armato III et al., 2015)	Lung cancer	98.4	92.1	93.9	92.6
LungCT-Diagnosis (Grove et al., 2015)	Lung cancer	97.7	91.5	92.9	92.4
QIN-Breast (Li et al., 2015)	Breast cancer	99.1	92.6	94.8	93.7
QIN Breast DCE-MRI (Huang et al., 2014)	Breast cancer	98.4	92.0	94.4	93.2
Breast-MRI-NACT-Pilot (Newitt & Hylton, 2016)	Breast cancer	98.0	91.2	93.3	92.5
Daily and Sports Activities Dataset (Altun & Barshan, 2010)	Daily activities	91.5	88.1	89.6	88.9
Activities of Daily Living Recognition Using Binary Sensors Dataset (Ordónez & Sanchis, 2013)	Daily activities	90.3	87.3	88.7	88.2
Smartphone-Based Recognition of Human Activities and Postural Transitions Dataset (Reyes-Ortiz et al., 2016)	Daily activities	92.1	88.8	90.4	89.5
Fruits 360 (Mureşan & Oltean, 2017)	Fruits	98.4	94.4	95.9	95.0
Fruit Image Dataset (Marko, 2013)	Fruits	94.9	91.6	92.8	92.1
Fruits-262 (Minuţ & Iftene, 2021)	Fruits	98.8	94.9	96.5	95.4
THE MNIST DATABASE of handwritten digits (LeCun et al., 1998)	Handwriting	93.8	89.7	90.7	90.1
GNHK: A Dataset for English Handwriting in the Wild (Lee et al., 2021)	Handwriting	92.2	88.0	89.3	88.8
The Chars74K dataset (De Campos et al., 2009)	Handwriting	96.5	92.0	93.3	92.6
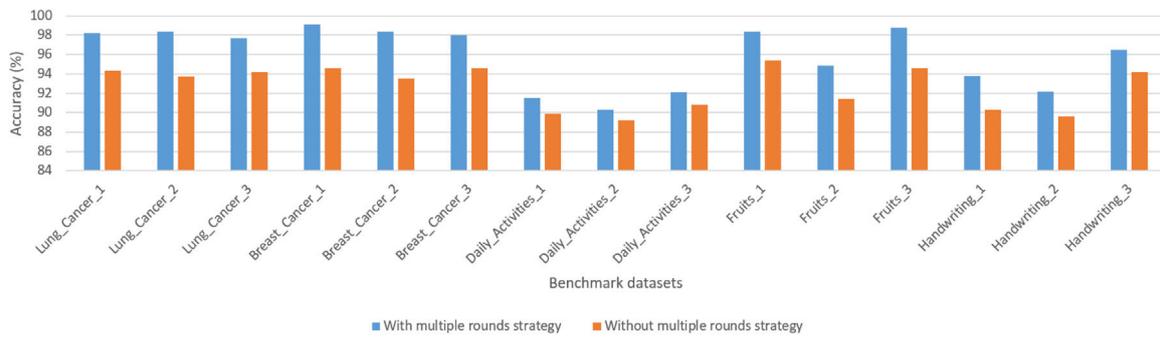
Fig. 5. Performance comparison between the MITL-NTA algorithm with and without the multiple rounds strategy.
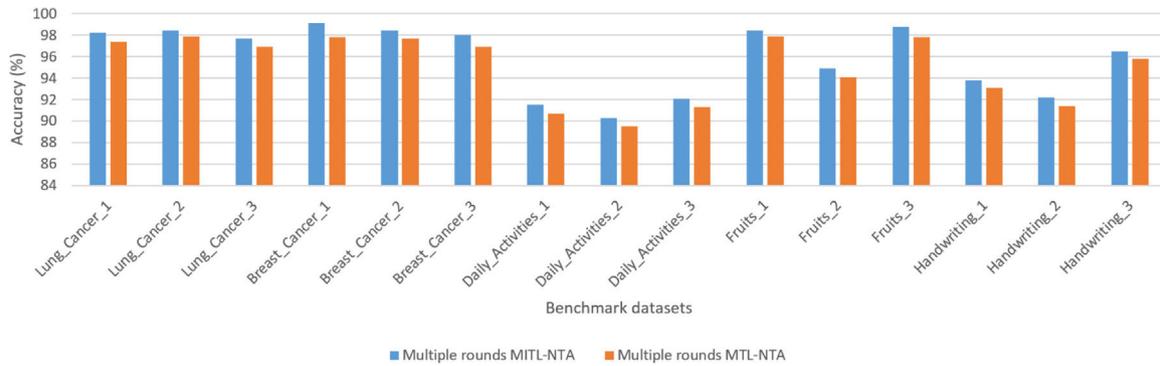

Fig. 6. Performance comparison between multiple rounds of the MITL-NTA algorithm and multiple rounds of the MTL-NTA algorithm.

disciplines by the proposed algorithm is 4.29% for lung cancer, 4.53% for breast cancer, 1.48% for daily activities, 3.80% for fruits, and 3.07% for handwriting. The deviations of enhancement in different disciplines are due to the nature of the similarities between the source and target domains. The more similar the domains are, the better the transferability to the target model is. The results suggest that the multiple rounds strategy can facilitate the transfer of innovation and knowledge to the target model based on multiple datasets.

Incremental learning

Fig. 6 compares the accuracy of the multiple rounds MITL-NTA algorithm with the multiple rounds MTL-NTA algorithm. The algorithm without incremental learning is equivalent to a one-step training of the target model without the division of datasets into subsets. The range of improvement in the five disciplines is 0.511−0.826% for lung cancer, 0.716−1.14% for breast cancer, 0.876−0.894% for daily activities, 0.509−1.02% for fruits, and 0.731−0.875% for handwriting. The improvement of the average accuracy in the five disciplines by

the proposed algorithm is 0.719% for lung cancer, 1.06% for breast cancer, 0.884% for daily activities, 0.794% for fruits, and 0.786% for handwriting. The results suggest that incrementally fine-tuning the model can enhance its performance due to the decrease in the level of negative transfer. It is worth noting that the negative transfer avoidance algorithm also contributes to the decrease in the impact of negative transfer.

Negative transfer avoidance

Fig. 7 compares the accuracy of the multiple rounds MITL-NTA algorithm with the multiple rounds MITL algorithm. The range of improvement in the five disciplines is 5.05−6.15% for lung cancer, 4.48−5.92% for breast cancer, 1.69−2.12% for daily activities, 4.02−4.99% for fruits, and 3.76−4.57% for handwriting. The improvement of the average accuracy in the five disciplines by the proposed algorithm is 5.45% for lung cancer, 5.16% for breast cancer, 1.93% for daily activities, 4.62% for fruits, and 4.13% for handwriting. These results confirm the effectiveness of the negative transfer avoidance
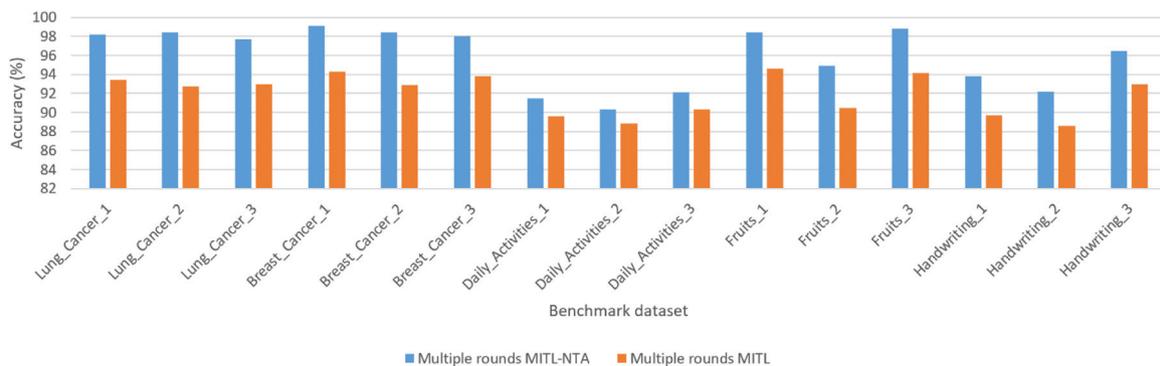

Fig. 7. Performance comparison between multiple rounds of the MITL-NTA algorithm and multiple rounds of the MITL algorithm.

algorithm which minimizes the transfer of irrelevant innovation and knowledge from the source domain to the target domain. Particularly, it is observed that the number of rounds is reduced when the negative transfer avoidance algorithm is omitted, which hits the stopping criterion of the iterations of the algorithm. Compared with incremental learning, the negative transfer avoidance algorithm plays a more significant role in the reduction of negative transfer (from 2 times up to 5 times improvement).

Conclusion and future research directions

In the current data explosion era, the availability of large-scale datasets contains valuable innovation and knowledge that could be utilized to build more accurate machine learning models for smart city applications. To share innovation and knowledge, transfer learning has served as a solid foundation to bridge both homogeneous and heterogeneous datasets. In this paper, the goal has been to leverage the performance of the target model by incorporating three novel ideas, namely, multiple rounds strategy, incremental transfer learning, and negative transfer avoidance. An in-depth performance evaluation and analysis was carried out on the proposed algorithm using 15 benchmark datasets in 5 disciplines including lung cancer, breast cancer, daily activities, fruits, and handwriting. Performance comparison between the proposed algorithm and existing works showed an average accuracy improvement of 4.35%. Ablation studies reveal that the proposed algorithm improves the average accuracy by 3.44%, 0.849%, and 4.26% compared with methodology without multiple rounds strategy, incremental learning, and negative transfer avoidance algorithms, respectively. The research implications are a breakthrough for the positive transfer of knowledge from the heterogeneous source domain to the target domain. Particularly, the research applications become very useful in regard to small-scale datasets and for those applications that may not have similar datasets. Transfer learning is usually accuracy-orientated to leverage the target model performance. Researchers may select a wide range of pre-trained models, such as AlexNet, GoogleNet, ResNet, VGG-16, and EfficientNet, from well-known datasets. Then, prioritization is carried out on the source domains for multiple rounds of incremental transfer learning. The target model not only receives knowledge from relevant source samples but also reduces the severity of overfitting of the model, thereby enhancing model performance. For image-related datasets, one may enhance the quality of the image before model implementation (Liu et al., 2022; Yu et al., Early Access). Several future research directions are recommended, including (i) reducing the number of rounds of transfer learning with the aid of the inclusion of a larger pool of source datasets; (ii) creating intermediate domains as bridges between the source and target domains to increase the similarities between domains; and (iii) generating additional training data from relevant samples via a data generation algorithm.

References

Aerts, H. J., Velazquez, E. R., Leijenaar, R. T., Parmar, C., Grossmann, P., Carvalho, S., et al. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications*, 5(1), 1–9. doi:10.1038/ncomms5006.

Ahmed, S. M., Raychaudhuri, D. S., Paul, S., Oymak, S., & Roy-Chowdhury, A. K. (2021). Unsupervised multi-source domain adaptation without access to source data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10103–10112). doi:10.48550/arXiv.2104.01845.

Alencastre-Miranda, M., Johnson, R. M., & Krebs, H. I. (2021). Convolutional neural networks and transfer learning for quality inspection of different sugarcane varieties. *IEEE Transactions on Industrial Informatics*, 17(2), 787–794. doi:10.1109/TII.2020.2992229.

Almomani, A., Alauthman, M., Shatnawi, M. T., Alweshah, M., Alrosan, A., Alomoush, W., et al. (2022). Phishing website detection with semantic features based on machine learning classifiers: A comparative study. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 18(1), 1–24. doi:10.4018/IJSWIS.297032.

Altun, K., & Barshan, B. (2010). Human activity recognition using inertial/magnetic sensor units. In Proceedings of the *international workshop on human behavior understanding* (pp. 38–51). Springer. doi:10.1007/978-3-642-14715-9_5.

Alzubaidi, L., Al-Amidie, M., Al-Asadi, A., Humaidi, A. J., Al-Shamma, O., Fadhel, M. A., et al. (2021). Novel transfer learning approach for medical imaging with limited labelled data. *Cancers*, 13, 1590. doi:10.3390/cancers13071590.

Armato, S. G., III, Hadjiiski, L., Tourassi, G. D., Drukker, K., Giger, M. L., Li, F., et al. (2015). LUNGx challenge for computerized lung nodule classification: Reflections and lessons learned. *Journal of Medical Imaging*, 2,(2) 020103. doi:10.1117/1.JMI.2.2.020103.

Aslan, M. F., Unlersen, M. F., Sabanci, K., & Durdu, A. (2021). CNN-based transfer learning—BiLSTM network: A novel approach for COVID-19 infection detection. *Applied Soft Computing*, 98, 106912. doi:10.1016/j.asoc.2020.106912.

Bashath, S., Perera, N., Tripathi, S., Manjang, K., Dehmer, M., & Streib, F. E. (2022). A data-centric review of deep transfer learning with applications to text data. *Information Sciences*, 585, 498–528. doi:10.1016/j.ins.2021.11.061.

Belouadah, E., Popescu, A., & Kanellos, I. (2021). A comprehensive study of class incremental learning algorithms for visual tasks. *Neural Networks, 135*, 38–54.

Chan, J. Y. L., Bea, K. T., Leow, S. M. H., Phoong, S. W., & Cheng, W. K. (2022). State of the art: A review of sentiment analysis based on sequential transfer learning. *Artificial Intelligence Review*, 1–32. doi:10.1016/j.neunet.2020.12.003.

Chen, B., Shen, C., Wang, D., Kong, L., Chen, L., & Zhu, Z. (2022). A lifelong learning method for gearbox diagnosis with incremental fault types. *IEEE Transactions on Instrumentation and Measurement*, 71, 3514010. doi:10.1109/TIM.2022.3177138.

Chen, S., Xu, H., Xu, D., Ji, W., Li, S., Yang, M., et al. (2021). Evaluating validation strategies on the performance of soil property prediction from regional to continental spectral data. *Geoderma*, 400, 115159. doi:10.1016/j.geoderma.2021.115159.

Chui, K. T. (2022). Driver stress recognition for smart transportation: Applying multiobjective genetic algorithm for improving fuzzy c-means clustering with reduced time and model complexity. *Sustainable Computing: Informatics and Systems*, 35, 100668. doi:10.1016/j.suscom.2022.100668.

Day, O., & Khoshgoftaar, T. M. (2017). A survey on heterogeneous transfer learning. *Journal of Big Data*, 4(1), 1–42. doi:10.1186/s40537-017-0089-0.

De Campos, T. E., Babu, B. R., & Varma, M. (2009). Character recognition in natural images. *VISAPP, 7*(2), 273–280.

Delange, M., Aljundi, R., Masana, M., Parisot, S., Jia, X., Leonardis, A., et al. (2022). A continual learning survey: Defying forgetting in classification tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7), 3366–3385. doi:10.1109/TPAMI.2021.3057446.

Do, P., Phan, T., Le, H., & Gupta, B. B. (2020). Building a knowledge graph by using cross-lingual transfer method and distributed MinIE algorithm on apache spark. *Neural Computing and Applications*, 1–17. doi:10.1007/s00521-020-05495-1.

Dwivedi, R. K., Kumar, R., & Buyya, R. (2021). Gaussian distribution-based machine learning scheme for anomaly detection in healthcare sensor cloud. *International Journal of Cloud Applications and Computing (IJCAC)*, 11(1), 52–72. doi:10.4018/IJCAC.2021010103.

Ferlay, J., Ervik, M., Lam, F., Colombet, M., Mery, L., Piñeros, M., et al. (2020). *Global cancer observatory: Cancer today*. Lyon: International Agency for Research on Cancer.

Francisco, L., Davis, W. R., & Franzon, P. (2022). A Deep transfer learning design rule checker with synthetic training. *IEEE Design & Test*. doi:10.1109/MDAT.2022.3162786.

Ginsburg, O., Yip, C. H., Brooks, A., Cabanes, A., Caleffi, M., Dunstan Yataco, J. A., et al. (2020). Breast cancer early detection: A phased approach to implementation. *Cancer*, 126, 2379–2393. doi:10.1002/cncr.32887.

Grove, O., Berglund, A. E., Schabath, M. B., Aerts, H. J., Dekker, A., Wang, H., et al. (2015). Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. *PloS One, 10*,(3) e0118261. doi:10.1371/journal.pone.0118261.

Hammad, M., Alkinani, M. H., Gupta, B. B., El-Latif, A., & Ahmed, A. (2021). Myocardial infarction detection based on deep neural network on imbalanced data. *Multimedia Systems*, 1–13. doi:10.1007/s00530-020-00728-8.

Han, T., Liu, C., Wu, R., & Jiang, D. (2021). Deep transfer learning with limited data for machinery fault diagnosis. *Applied Soft Computing*, 103, 107150. doi:10.1016/j.asoc.2021.107150.

Hassan, T., Hassan, B., Akram, M. U., Hashmi, S., Taguri, A. H., & Werghi, N. (2021). Incremental cross-domain adaptation for robust retinopathy screening via Bayesian deep learning. *IEEE Transactions on Instrumentation and Measurement*, 70, 1–14. doi:10.1109/TIM.2021.3122172.

Hua, Y., Sevegnani, M., Yi, D., Birnie, A., & Mcaslan, S. (2022). Fine-grained RNN with transfer learning for energy consumption estimation on EVs. *IEEE Transactions on Industrial Informatics*, 18, 8182–8190. doi:10.1109/TII.2022.3143155.

Huang, W., Li, X., Chen, Y., Li, X., Chang, M. C., Oborski, M. J., et al. (2014). Variations of dynamic contrast-enhanced magnetic resonance imaging in evaluation of breast cancer therapy response: A multicenter data analysis challenge. *Translational Oncology*, 7(1), 153–166. doi:10.1593/tlo.13838.

Kim, J., Yang, G., Kim, J., Lee, S., Kim, K. K., & Park, C. (2021). Efficiently updating ECG-based biometric authentication based on incremental learning. *Sensors*, 21(5), 1568. doi:10.3390/s21051568.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. doi:10.1109/5.726791.

Lee, A. W., Chung, J., & Lee, M. (2021). GNHK: A dataset for english handwriting in the wild. In *Proceedings of the* international conference on document analysis and recognition (pp. 399–412). Springer. doi:10.1007/978-3-030-86337-1_27.

Li, X., Abramson, R. G., Arlinghaus, L. R., Kang, H., Chakravarthy, A. B., Abramson, V. G., et al. (2015). Multiparametric magnetic resonance imaging for predicting pathological response after the first cycle of neoadjuvant chemotherapy in breast cancer. *Investigative Radiology*, *50*(4), 195–204. doi:10.1097/RLI.0000000000000100.

Li, Y. F., Guo, L. Z., & Zhou, Z. H. (2019). Towards safe weakly supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(1), 334–346. doi:10.1109/TPAMI.2019.2922396.

Li, D. W., & Huang, H. (2022). Few-shot class-incremental learning via compact and separable features for fine-grained vehicle recognition. *IEEE Transactions on Intelligent Transportation Systems*, *23*(11), 21418–21429. doi:10.1109/TITS.2022.3174662.

Li, X., Wei, J., & Jiao, H. (2022). Real-time tracking algorithm for aerial vehicles using improved convolutional neural network and transfer learning. *IEEE Transactions on Intelligent Transportation Systems*, *23*(3), 2296–2305. doi:10.1109/TITS.2021.3072872.

Li, Q., Zhang, A., Li, Z., & Wu, Y. (2021). Improvement of EMG pattern recognition model performance in repeated uses by combining feature selection and incremental transfer learning. *Frontiers in Neurorobotics*, *15*, 87. doi:10.3389/fnbot.2021.69917.

Liu, R. W., Guo, Y., Lu, Y., Chui, K. T., & Gupta, B. B. (2022). Deep network-enabled haze visibility enhancement for visual iot-driven intelligent transportation systems. *IEEE Transactions on Industrial Informatics*. doi:10.1109/TII.2022.3170594.

Niu, S., Liu, M., Liu, Y., Wang, J., & Song, H. (2021). Distant domain transfer learning for medical imaging. *IEEE Journal of Biomedical and Health Informatics*, *25*(10), 3784–3793. doi:10.1109/JBHI.2021.3051470.

Machacha, L., & Bhattacharya, P. (2021). Cancer classification from DNA microarray using genetic algorithms and case-based reasoning. *International Journal of Software Science and Computational Intelligence (IJSSCI)*, *13*(1), 17–37. doi:10.4018/IJSSCI.2021010102.

Marko, Š. (2013). Automatic fruit recognition using computer vision. *Mentor: Matej Kristan), Fakulteta za racunalništvo in informatiko, Univerza v Ljubljani*.

Martin, H., Acher, M., Lesoil, L., Jezequel, J. M., Khelladi, D. E., & Pereira, J. A. (2022). Transfer learning across variants and versions: The case of linux kernel size. *IEEE Transactions on Software Engineering*, *48*(11), 4274–4290. doi:10.1109/TSE.2021.3116768.

Masud, M., Hossain, M. S., Alhumyani, H., Alshamrani, S. S., Cheikhrouhou, O., Ibrahim, S., et al. (2021). Pre-trained convolutional neural networks for breast cancer detection using ultrasound images. *ACM Transactions on Internet Technology (TOIT)*, *21*(4), 1–17. doi:10.1145/3418355.

Minuţ, M. D., & Iftene, A. (2021). Creating a dataset and models based on convolutional neural networks to improve fruit classification. In *Proceedings of the 23rd international symposium on symbolic and numeric algorithms for scientific computing (SYNASC)* (pp. 155–162). IEEE. doi:10.1109/SYNASC54541.2021.00035.

Mureşan, H., & Oltean, M. (2017). Fruit recognition from images using deep learning. *arXiv preprint arXiv:1712.00580*. doi:10.48550/arXiv.1712.00580.

Newitt, D., & Hylton, N. (2016). Single site breast DCE-MRI data and segmentations from patients undergoing neoadjuvant chemotherapy. *The Cancer Imaging Archive, 2*.

Ordónez, F. J., De Toledo, P., & Sanchis, A. (2013). Activity recognition using hybrid generative/discriminative models on home environments using binary sensors. *Sensors*, *13*(5), 5460–5477. doi:10.3390/s130505460.

Peirelinck, T., Kazmi, H., Mbuwir, B. V., Hermans, C., Spiessens, F., Suykens, J., et al. (2022). Transfer learning in demand response: A review of algorithms for data-efficient modelling and control. *Energy and AI*, *7*, 100126. doi:10.1016/j.egyai.2021.100126.

Peng, Z., Zhang, W., Han, N., Fang, X., Kang, P., & Teng, L. (2020). Active transfer learning. *IEEE Transactions on Circuits and Systems for Video Technology*, *30*(4), 1022–1036. doi:10.1109/TCSVT.2019.2900467.

Pinto, G., Wang, Z., Roy, A., Hong, T., & Capozzoli, A. (2022). Transfer learning for smart buildings: A critical review of algorithms, applications, and future perspectives. *Advances in Applied Energy*, *5*, 100084. doi:10.1016/j.adapen.2022.100084.

Ren, Y., Liu, J., Zhang, H., & Wang, W. (2022). TDDA-Net: A transitive distant domain adaptation network for industrial sample enhancement. *Information Sciences*, *66*, 927–944. doi:10.1016/j.ins.2022.05.109.

Reyes-Ortiz, J. L., Oneto, L., Samà, A., Parra, X., & Anguita, D. (2016). Transition-aware human activity recognition using smartphones. *Neurocomputing*, *171*, 754–767. doi:10.1016/j.neucom.2015.07.085.

Roy, D., Panda, P., & Roy, K. (2020). Tree-CNN: A hierarchical deep convolutional neural network for incremental learning. *Neural Networks*, *121*, 148–160. doi:10.1016/j.neunet.2019.09.010.

Sánchez-de-Madariaga, R., Martinez-Romo, J., Escribano, J. M. C., & Araujo, L. (2022). Semi-supervised incremental learning with few examples for discovering medical association rules. *BMC Medical Informatics and Decision Making*, *22*(1), 1–11. doi:10.1186/s12911-022-01755-3.

Sedik, A., Hammad, M., El-Samie, A., Fathi, E., Gupta, B. B., El-Latif, A., et al. (2021). Efficient deep learning approach for augmented detection of Coronavirus disease. *Neural Computing and Applications*, 1–18. doi:10.1007/s00521-020-05410-8.

Shi, X., Liu, X., Xu, C., Huang, Y., Chen, F., & Zhu, S. (2022). Cross-lingual offensive speech identification with transfer learning for low-resource languages. *Computers and Electrical Engineering*, *101*, 108005. doi:10.1016/j.compeleceng.2022.108005.

Tabassum, A., Erbad, A., Mohamed, A., & Guizani, M. (2021). Privacy-preserving distributed IDS using incremental learning for IoT health systems. *IEEE Access Practical Innovations, Open Solutions*, *9*, 14271–14283. doi:10.1109/ACCESS.2021.3051530.

Wan, C., Yang, R., Huang, M., Zeng, N., & Liu, X. (2021). A review on transfer learning in EEG signal analysis. *Neurocomputing*, *421*, 1–14. doi:10.1016/j.neucom.2020.09.017.

Wang, Z., Dai, Z., Póczos, B., & Carbonell, J. (2019). Characterizing and avoiding negative transfer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11293–11302). doi:10.1109/CVPR.2019.01155.

Wang, J., Mo, Z., Zhang, H., & Miao, Q. (2020). Ensemble diagnosis method based on transfer learning and incremental learning towards mechanical big data. *Measurement*, *155*, 107517. doi:10.1016/j.measurement.2020.107517.

Yoon, H., & Li, J. (2018). A novel positive transfer learning approach for telemonitoring of Parkinson's disease. *IEEE Transactions on Automation Science and Engineering*, *16*(1), 180–191. doi:10.1109/TASE.2018.2874233.

Yuan, K., Xu, W., Li, W., & Ding, W. (2022). An incremental learning mechanism for object classification based on progressive fuzzy three-way concept. *Information Sciences*, *584*, 127–147. doi:10.1016/j.ins.2021.10.058.

Zhang, W., Deng, L., Zhang, L., & Wu, D. (2022). A survey on negative transfer. *IEEE/CAA Journal of Automatica Sinica*. doi:10.1109/JAS.2022.106004.

Zuo, Y., Yao, H., & Xu, C. (2021). Attention-based multi-source domain adaptation. *IEEE Transactions on Image Processing*, *30*, 3793–3803. doi:10.1109/TIP.2021.3065254.