This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Predicting Demands of COVID-19 Prevention and Control Materials via Co-Evolutionary Transfer Learning

Qin Song, Yu-Jun Zheng, *Senior Member, IEEE*, Jun Yang, Yu-Jiao Huang, Wei-Guo Sheng, *Member, IEEE*, and Sheng-Yong Chen, *Senior Member, IEEE*

***Abstract*—The novel coronavirus pneumonia (COVID-19) has created great demands for medical resources. Determining these demands timely and accurately is critically important for the prevention and control of the pandemic. However, even if the infection rate has been estimated, the demands of many medical materials are still difficult to estimate due to their complex relationships with the infection rate and insufficient historical data. To alleviate the difficulties, we propose a co-evolutionary transfer learning (CETL) method for predicting the demands of a set of medical materials, which is important in COVID-19 prevention and control. CETL reuses material demand knowledge not only from other epidemics, such as severe acute respiratory syndrome (SARS) and bird flu but also from natural and manmade disasters. The knowledge or data of these related tasks can also be relatively few and imbalanced. In CETL, each prediction task is implemented by a fuzzy deep contractive autoencoder (CAE), and all prediction networks are cooperatively evolved, simultaneously using intrapopulation evolution to learn task-specific knowledge in each domain and using interpopulation evolution to learn common knowledge shared across the domains. Experimental results show that CETL achieves high prediction accuracies compared to selected state-of-the-art transfer learning and multitask learning models on datasets during two stages of COVID-19 spreading in China.**

***Index Terms*—Co-evolutionary learning, deep learning, demand prediction, epidemic prevention and control, medical materials, transfer learning.**

Manuscript received October 6, 2021; revised January 26, 2022; accepted March 28, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61872123 and Grant 61873082, and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LR20F030002 and Grant LQY20F030001. This article was recommended by Associate Editor Z. Ju. *(Corresponding author: Yu-Jun Zheng.)*

Qin Song is with the School of Public Health, Hangzhou Normal University, Hangzhou 311121, China.

Yu-Jun Zheng and Wei-Guo Sheng are with the School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China (e-mail: yujun.zheng@computer.org).

Jun Yang is with the School of Public Health, Zhejiang University, Hangzhou 310009, China.

Yu-Jiao Huang is with the Zhijiang College, Zhejiang University of Technology, Shaoxing 312030, China.

Sheng-Yong Chen is with the School of Computer Science and Engineering, Tianjin University of Technology, Tianjin 300384, China.

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TCYB.2022.3164412.

Digital Object Identifier 10.1109/TCYB.2022.3164412

I. Introduction

THE ONGOING outbreak of the novel coronavirus pneumonia (COVID-19) has created great demands for medical resources. Predicting the demands of emergency medical resources in a timely and accurate manner is critical to the prevention and control of the epidemic. Some resource demands can be estimated based on the infection rate. For example, for most therapeutic drugs and diagnostic agents, their demands are considered as proportional to the infection rate. There have been many studies conducted on mathematical models and artificial intelligence methods for predicting the infection rate [1]–[5]. However, for many other epidemic prevention and control materials, such as face masks and disinfectants, their demands are not strictly proportional to the infection rate. In fact, their demands are also affected by many other factors, including population size and density, urban and rural environments, traffic flow, public panic, etc. The relationships between these influence factors and the demands of materials are complex, uncertain, and not yet entirely understood; therefore, it is difficult to develop a deterministic mathematical model for demand prediction. In recent years, machine-learning approaches, in particular deep neural networks (DNNs), have been increasingly used to model complex relationships between input and output variables [6], [7]. However, most machine-learning models require large amounts of historical data as training samples. COVID-19 is an emerging infectious disease, and so we do not have sufficient historical data of material demands for epidemic prevention and control.

The motivation of this article comes from our practice of planning epidemic prevention and control materials for local governments during the first outbreak of COVID-19 in Zhejiang Province, China. To overcome the data limitation, we intend to use transfer learning, an emerging paradigm of machine learning that reuses knowledge accumulated in a source domain (task) to better learn a different but related target domain [8], [9]. Nevertheless, the knowledge from other epidemics, such as severe acute respiratory syndrome (SARS), dengue fever, and bird flu, is still limited; therefore, we also intend to reuse knowledge from our past studies on material demand prediction and planning in various natural and manmade disasters [10]–[12]. For example, in a large-scale flood, we should plan sufficient materials, such as protective clothing and disinfectants to prevent potential epidemics, such as cholera and schistosomiasis, and we believe this knowledge can be shared for material preparation for epidemic prevention and control. Other potential knowledge includes the relationship between the demands and the affected area of the event, the relationship between the increase of demands and the spread of the event, the relationship among the demands of different materials, etc., which can be quite complex and difficult to learn by classical regression models or shallow neural networks. To effectively reuse knowledge from multiple domains, in this study, we propose a co-evolutionary transfer learning (CETL) method, which cooperatively evolves multiple DNNs, each for prediction in a domain, to simultaneously learn both domain-specific and domain-general knowledge. The main contributions of the article are as follows.

1) We propose a new fuzzy deep transfer learning model for learning a task with few data or knowledge by leveraging knowledge from a set of related tasks whose data can also be few and imbalanced;
2) We propose a new co-evolutionary learning algorithm that evolves multiple subpopulations of solutions, using intrapopulation evolution to learn task-specific components and interpopulation evolution to learn common components shared among the tasks;
3) We demonstrate the performance advantages of CETL over state-of-the-art transfer learning methods in demand prediction in eight cities during two stages of COVID-19 spreading in China.

It should be noted that, in general, there are two categories of public health strategies for containing COVID-19 [13]: 1) *suppression*, which aims to drastically reduce the transmission rate and halt endogenous transmission through strict nonpharmaceutical interventions and other management measures in the target population and 2) *mitigation*, which aims to slow but do not interrupt transmission completely, achieving herd immunity by allowing the virus to spread through the population while mitigating disease burden. In this study, we focus on the first strategy that has been taken by countries, such as China, Singapore, South Korea, and Italy. That is, we aim to predict material demands during the period from the outbreak of the epidemic to the achievement of suppression. This task has much more common features with most disasters.

In the remainder of this article, Section II describes related work, Section III describes the CETL method in details, Section IV reports the computational results, and Section V concludes.

II. Related Work

For an emerging epidemic, such as COVID-19, it is difficult to obtain sufficient data in a short period of time. Recently, there has been increasing interest in machine-learning tasks with limited data. These methods include transfer learning [8], few-shot learning [14], [15], metalearning [16], and multitask learning [17], although there are overlaps between these categories. In this section, we simply introduce the concept of each category, and then review its typical applications for epidemic prevention and control and other public health tasks.

A. Transfer Learning

Transfer learning aims to reuse knowledge accumulated in one or more source tasks to better learn a different but related target task with limited or no labeled training data. Deep learning models are particularly suited for transfer learning, as they use multiple hidden layers to capture intricate nonlinear representations of data, and the abstract features learned in one task may be useful for another [18]. As a shortage of reliable datasets of a running pandemic is a common phenomenon, transfer learning has been employed to reuse knowledge from other similar diseases to mitigate COVID-19 like pandemics in terms of stop spread, diagnosis of the disease, drug and vaccine discovery, treatment, patient care, and many more [19].

Minaee *et al.* [20] used transfer learning to identify COVID-19 from chest radiograms. Because of the limited number of patients, they reused ResNet/SqueezeNet/DenseNet convolutional neural networks (CNNs) pretrained on public non-COVID datasets by data augmentation and fine-tuning. Their test results on 3000 radiograms achieved a sensitivity rate of around 98% and a specificity rate of around 90%. The work of [21] added more DNNs, including MobileNet, VGG16, InceptionV3, and XceptionNet as the base models for the same task. Basu *et al.* [22] used a similar transfer learning method to classify four classes (normal, pneumonia, COVID-19, and other diseases) from chest X-Ray images, but their method replaces the last layer of the pretrained CNN with a new fully connected layer, which achieved an overall accuracy of around 90.13%. Ohata *et al.* [23] also transferred a CNN pretrained on ImageNet for COVID-19 detection based on chest X-ray images, but they combined the CNN with K-nearest neighbor, Bayes, and support vector machine (SVM) to achieve higher accuracy and $F1$-Score. To predict severity score for COVID-19 from chest X-ray images, Cohen *et al.* [24] started from a DNN pretrained on seven non-COVID-19 datasets and added a feature extraction layer to constructed features for COVID-19 images. The separate feature extractor restricts the model complexity and reduces the possibility of overfitting. Khalifa *et al.* [25] employed a different approach that first converts grayscale X-ray images into neutrosophic images to compensate the limited COVID-19 chest X-Ray dataset, and then uses these images to retrain existing DNNs to classify COVID-19 infection versus non-COVID-19 diseases. Khalifa *et al.* [26] further employed a generative adversarial network to generate more images to improve classification accuracy. In [27], Apostolopoulos and Mpesiana evaluated the performance of state-of-the-art CNN architectures adopting transfer learning for medical image classification related to COVID-19. The best accuracy, sensitivity, and specificity obtained are 96.78%, 98.66%, and 96.46%, respectively.

Using extensive machine-learning experiments on cancer omics data, Gao and Cui [28] found that current prevalent schemes of multiethnic machine learning are prone to generating significant performance disparities among ethnic groups. They analyzed the performance disparities caused by data inequality and data distribution, and revealed that transfer learning can improve model performance for data-disadvantaged ethnic groups so as to reduce health care disparities. Khamparia *et al.* [29] proposed an Internet of Health Things deep transfer learning framework, which extracts feature from cervical images on pretrained CNN models to learn for cervical cancer detection and classification in Pap smear images. The results showed that pretrained ResNet50 can achieve a classification rate of 97.89%. Song *et al.* [30] proposed a tridirectional transfer learning method for predicting the morbidity of gastric cancer based on an existing model for predicting the morbidity of another disease in another region by fusing two different directions of transfer learning, which achieves a significantly higher prediction accuracy compared with the state of the arts. However, the method requires that both two intermediate domains have sufficient data.

B. Few-Shot Learning

Originally, few-shot classification defined a scenario where only very few samples per class were accessible [31]. With the advent of deep learning, the scenario was broadened into that a classifier having large amounts of data for a number of base classes must be adapted to accommodate new classes bound by a scarce data regime [15]. In this sense, transfer learning leveraging knowledge or data from relevant tasks is an intuitive way to address few-shot tasks. Chen *et al.* [32] used contrastive learning to train an encoder that captures expressive feature representations on large and publicly available lung datasets, and adopted the prototypical network for the automated diagnosis of COVID-19 from chest CT images. The model has shown superior performance on annotated COVID-19 CT datasets compared with other competing methods. For the semantic segmentation of pneumonia-infected area segmentation in CT images, Voulodimos *et al.* [33] used few-shot learning based on U-Net architectures, the weights of which are dynamically tuned as new few samples are being fed into the network. Wei *et al.* [34] proposed a c-way k-shot learning method for leveraging small amount of samples to complete a brand new classification task. The method was combined with an attention similarity to perform cough classification in a COVID-19 detection system. Lwowski and Najafirad [35] used few shot learning to fine-tune a semisupervised model built on the unlabeled COVID-19 and previously labeled influenza dataset that can provide insights into COVID-19 that have not been investigated. The results showed the efficacy of the proposed model with an accuracy of 86% in the identification of Covid-19 related discussion using recently collected tweets.

C. Metalearning

Metalearning, as a concept of "learning to learn" inspired by human intelligence, is a task-level learning aims to accumulate experience from learning multiple tasks, while base-learning focuses on modeling the data distribution of a single task. It has been proposed as a framework to address the challenging few-shot learning setting [16]. Shorfuzzaman and Hossain [36] proposed a metalearning framework that uses a fine-tuned VGG16 CNN as base model to generate feature embeddings, which are then used by the Siamese network to learn a metric space for few-shot classification of COVID-19 cases with limited training chest X-Ray images. Naren *et al.* [37] used a model agnostic metalearning for COVID-19 detection based on chest X-ray images. They showed that increasing the number of images available to the sublearners can lead to diminishing returns in performance. To address the problem of differentiating COVID-19 from other pneumonia cases, Zheng *et al.* [38] proposed an unsupervised metalearning method, which first constructs a deep diagnosis model based on a relation network to capture and memorize the relation among different images, and then uses a self-knowledge distillation mechanism to distills knowledge to the model to enhance the performance. Singh *et al.* [39] proposed a MetaMed approach that relies on metalearning by formulating the medical image classification for low data regime as a few-shot learning problem. Test on three public medical datasets demonstrated that MetaMed exhibits an overall accuracy of over 70%. For predicting the population movement with the spread of COVID-19, Panagopoulos *et al.* [40] used a transfer graph neural network, where the nodes correspond to regions and the edge weights denote human mobility. The pandemic's asynchronous outbreaks across countries are capitalized using model-agnostic metalearning to transfer knowledge from one country to another. To analyze time influence in a causal analysis of the COVID-19 pandemic in Chile, Kristjanpoller *et al.* [41] used a set of machine-learning techniques, including metalearners and causal forest, and made final causal analysis with the best base learner. However, such an elitist strategy fails to utilize knowledge from different learners simultaneously.

D. Multitask Learning

Multitask learning aims to simultaneously learn a collection of similar regression or classification tasks by making the tasks learn from each other, often by using the same network for all tasks or optimizing all network parameters using the same samples [42]. Considering image segmentation, classification and reconstruction as three tasks, Amyar *et al.* [43] constructed an architecture composed of a common encoder for feature representation and two decoders and a multilayer perceptron for three tasks, respectively, to jointly identify COVID-19 patient and segment COVID-19 lesion from chest CT images. Goncharov *et al.* [44] considered identification of COVID-19 cases and severity quantification as two different triage problems, which are simultaneously learned using a CNN that applies classification layers to the most spatially detailed feature map. Hayhoe *et al.* [45] used multitask learning to tune a discrete-time epidemic model from various data sources and design optimal control strategies of human-mobility restrictions, where each region of interest was considered as a separate prediction task. Panda and Levitan [46] employed multitask learning to identify COVID-19 misinformation on social media in three languages, including English, Bulgarian, and Arabic, each of which is considered as a separate task. Rahimi and Gönen [47] proposed a multitask multiple kernel learning algorithm based on the Benders decomposition and treating the clustering problem as finding a given number of tree structures in a graph. They applied the algorithm to discriminate early- and late-stage cancers using genomic data, and the results show that the forest formulation becomes increasingly favorable with increasing number of tasks.

In the last few years, combining multitask learning with evolutionary computation to solve multiple optimization problems together has become an emerging research topic, where knowledge transfer is a key to share solutions across tasks. Liang *et al.* [48] presented an evolutionary multitasking based on subspace alignment and self-adaptive differential evolution, which uses a mapping matrix obtained by subspace learning to transform the search space and reduce negative knowledge transfer between tasks. The algorithm proposed by Chen *et al.* [49] treats the decision space of each task as a manifold, projects the joint manifold of all decision spaces to a latent space by solving a generalized eigenvalue decomposition problem, and represents the task relationships as the joint mapping matrix utilized for information transfer across decision spaces during the evolutionary process. As positive knowledge transfer facilitates superior performance characteristics, Lin *et al.* [50] proposed a multiobjective multitasking optimization algorithm that regards a transfer as positive if the transferred solution is nondominated in its target task and selects neighbors of this positive solution for the next generation. Chen *et al.* [51] proposed an evolutionary feature selection method for high-dimensional classification by using crossover to share information among tasks, where related tasks about the target concept are established by evaluating the importance of features.

E. Categories of Our Method

As aforementioned, the above categories have overlaps, and our method has more or less relations to each of these categories. As multitask learning, our method simultaneously learns a set of tasks that could benefit from each other; however, unlike most multitask learning problems, the amount of data we have for each task is imbalanced, and we do not use a big enough network for all tasks. Our method also addresses the problem of few examples as few-shot learning and metalearning, but our method does not use a single classifier as most few-shot learning tasks, nor does it explicitly differentiate base-learning and the specific target learning as most metalearning tasks. Because transfer learning leveraging knowledge or data from relevant tasks is more general and proposes a framework for more specific learning tasks, we define our method mainly in the category of transfer learning.

One distinct characteristic of our method is to propose a co-evolutionary algorithm to achieve simultaneous learning of multiple tasks, which borrows ideas from multitask optimization for transfer learning. Besides, as epidemic spreading and material consumption often involve great imprecision and uncertainty, we incorporate fuzzy information processing to the learning, borrowing ideas from recent studies that combines deep transfer learning with fuzzy systems [52]–[54].

III. Co-Evolutionary Transfer Learning

The proposed CETL method is for simultaneously learning m different but related tasks, formally denoted by $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m\}$, where each task $\mathcal{T}_i = \{X_i; Z_i; f_i : X_i \to Z_i\}$ is characterized by a feature space X_i, a label space Z_i, and a prediction function f_i. In this study, we consider nine tasks, including predicting prevention and control material demands for five types of epidemics (*COVID-19, SARS, dengue fever, H7N9 Influenza*, and *H1N1 Influenza*) and predicting relief material demands for four types of disasters/accidents (*earthquake, typhoon, flood*, and *dangerous chemical explosion*). The tasks have the same output labels, which are the demands of ten types of materials (*disposable medical masks, disposable medical gloves, goggles, protective clothing, oxygen bottles, surface disinfectant, disinfection machines, hand sanitizer, sprayers*, and *temperature guns*; the demand prediction tasks in disasters/accidents also involve other relief materials, such as emergency lights and tents, but they are not considered in COVID-19 material demand prediction in this study). As it is unnecessary to predict an accurate demand value in most cases, we describe each output demand as an interval (fuzzy) number $[z^L, z^U]$, where z^L and z^U are the lower and upper limits, respectively. The input features of the tasks are listed in Table I. Note that the number of input features is not fixed: a "repeatable feature" can be input multiple times and processed by recurrent neural networks; for example, for dengue fever, the area and mosquito density of each mosquito-rich region are input at a time. Also note that the relations between some features and a specific task can be unknown, but we use these features as inputs to the prediction task, leaving the deep learning model to implicitly infer the relations.

The nine tasks share the same output feature set Z and a subset of common input features X^c. Therefore, their prediction models can also share a part of latent representation Y, which can be useful for characterizing both input distributions $P(X_i)$ and edge distributions $P(Z|X_i)$ parameterized through $P(Z|Y)$ [55]. CETL cooperatively learns the shared representation of all tasks and simultaneously learns task-specific representations of different tasks, such that the knowledge of probability distribution learned from source tasks are automatically utilized in learning the target task.

In the following sections, we first describe the underlying model for prediction tasks and then describe the co-evolutionary learning algorithm in detail.

A. Model Structure and Pretraining

As the tasks share the same output space and a certain part of input space, we first construct a common network for all tasks; next, for each task, we add a task-specific part to the common network to obtain the corresponding prediction network. Fig. 1 illustrates the model structure.

The common network consists of a common underlying DNN and a common regression layer on top of the DNN. We propose a new fuzzy deep contractive autoencoders (CAEs) model for the common underlying DNN. A CAE [56] is a variant of autoencoder (AE) [57], which consists of an encoder and a decoder. The encoder transforms a D-dimensional input

TABLE I
INPUT FEATURES OF THE TEN CO-EVOLUTIONARY LEARNING TASKS

Domain (task)	**Specific input features**		**General input features**		
	standard	*repeatable*	*standard*		*repeatable*
COVID-19	Asymptomatic infection rate.		Infection rate, nosocomial infection rate, confirmed cases, confirmed severe cases, suspected cases, cured cases, transmission rate, vaccination coverage, vaccination success rate, floating population size, floating population density, garbage station density, public toilet density.	Total affected area, urban area, rural area, green area, mountainous area, water area, duration, peak duration, population size, population density, proportion of elderly, proportion of women, proportion of children, proportion of infants, proportion of disabled, average air temperature, max air temperature, min air temperature, average air humidity, max air humidity, min air humidity, peak rainfall, average rainfall, peak wind force, average wind force, average wind speed, peak snowfall, average snow fall.	Highly affected regions: - area - population density
SARS					
Dengue fever	Mosquito-rich area.	Mosquito-rich regions: - area - density			
H7N9	Wild-bird-rich area, domestic-bird breeding area.	Wild-bird-rich regions: - area - density domestic-bird farms: - area - density			
H1N1	Infant vaccination coverage.				
Earthquake	Magnitude, intensity, focal depth, distance from epicenter.	Aftershocks: - intensity - duration - affected area	Earthwork of collapsed buildings, death toll, number of injured, number of seriously injured, road network density, road network damage rate, power network density, power network damage rate.		
Typhoon	Distance from typhoon center, central wind force, central wind speed.				
Flood	Peak discharge, peak level, peak speed, average discharge, average level, average speed.				
Chemical explosion	Explosion equivalent, explosion heat, explosion radius, explosion coefficient, combustion heat, combustion radius, combustion coefficient, smoke concentration, smoke temperature, air velocity, spread rate.	Toxic gases: - toxicity grade - concentration - spread rate - affected area			

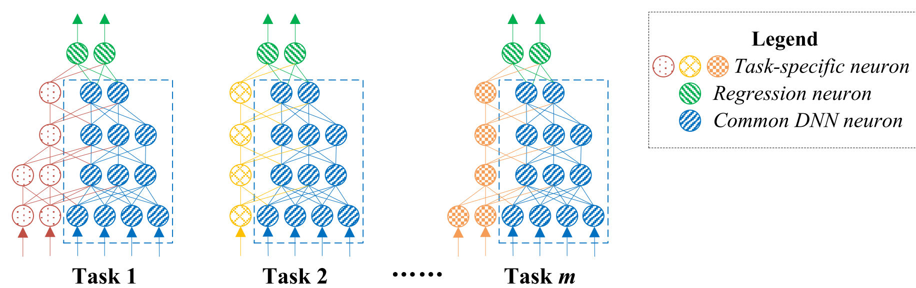

Fig. 1. Structure of co-evolutionary learning model for material demand prediction.

vector $\mathbf{x}$ to a hidden D'-dimensional representation $\mathbf{y}$ through an affine mapping

$$f(\mathbf{x}) = s(\mathbf{W}\mathbf{x} + \mathbf{b}) \tag{1}$$

where s is an activation function, $\mathbf{W}$ is a $D' \times D$ weight matrix, $\mathbf{b}$ is a D'-dimensional bias vector, and each input component is normalized into [0, 1].

The decoder maps the hidden representation $\mathbf{y}$ back to a reconstructed vector $\mathbf{x}'$ in the input space with appropriately sized parameters $\mathbf{W}^{\mathrm{T}}$ and $\mathbf{b}'$

$$f'(\mathbf{y}) = s\big(\mathbf{W}^{\mathrm{T}}\mathbf{y} + \mathbf{b}'\big). \tag{2}$$

The objective of AE learning is to minimize the reconstruction error $\mathcal{L}$ over the training set $\mathcal{D}$, while CAE improves the basic AE by incorporating the penalization of the sensitivity of the hidden representation $\mathbf{y} = f(\mathbf{x})$ to the input $\mathbf{x}$

$$\min J(\theta) = \sum_{\mathbf{x}\in\mathcal{D}} \Big(\mathcal{L}\big(\mathbf{x}, f'(f(\mathbf{x}))\big) + \lambda \|J_f(\mathbf{x})\|_F^2\Big) \tag{3}$$

where $\theta = \{\mathbf{W}, \mathbf{b}, \mathbf{b}'\}$ are model parameters to be learned, λ is a control parameter, and $\|J_f(\mathbf{x})\|_F^2$ is the Frobenius norm of the Jacobian of the nonlinear mapping for making the model less sensitive to input noise

$$\|J_f(\mathbf{x})\|_F^2 = \sum_{1\le i\le D} \sum_{1\le j\le D'} \left(\frac{\partial f_j(\mathbf{x})}{\partial x_i}\right)^2. \tag{4}$$

The common DNN uses stacked layers of CAE to learn higher order abstract and correlated representation from input features, where each layer captures the hidden representation of the layer below as input. In the early stage of an epidemic, the data from real life often contain much imprecise and uncertain information; to cope with such information, we express the model parameters as pythagorean-type fuzzy numbers (PFNs) [58]. Fuzzy number parameters enable the model to learn the fuzzy probability distribution over cross-layer units [59], and PFN based on Pythagorean fuzzy sets [60] that utilize both membership and nonmembership degrees satisfying the Pythagorean inequality enables each neuron to learn both how a feature *contributes* and *does not contribute* to the production of the output and allows for a larger body of membership grades than regular and intuitionistic fuzzy numbers [60], [61]. In this study, we use interval-valued PFN and employ $\exp(\cdot)$ as the activation function in (1), where the fuzzy exponential operation is defined as in [62] and [63]. To support CAE learning based on (3), we measure the reconstruction error in terms of the distance between $\mathbf{x}$ and the centroid of $\widetilde{\mathbf{x}}'$

$$\mathcal{L}(\mathbf{x}, \widetilde{\mathbf{x}}') = \sqrt{\sum_{d=1}^{D} \left(x_d - c(\widetilde{x}'_d)\right)^2} \tag{5}$$

where the centroid $c(A)$ of a fuzzy number A with membership function μ_A is calculated as follows [64], [65]:

$$c(A) = \frac{\int x\mu_A(x)dx}{\int \mu_A(x)dx}. \tag{6}$$

Using this centroid-based defuzzification method, the pretraining of fuzzy CAE is remained as a crisp optimization problem instead of a fuzzy optimization one. We employ the Hessian-free (HF) optimization algorithm [66] to pretrain the common DNN layer by layer in an unsupervised manner based on (3), using both labeled and unlabeled samples of all tasks. As X^c is shared among all tasks, the number of samples for training the common DNN is sufficient.

After unsupervised pretraining of the common DNN, we extend it to specific networks for different learning tasks by adding some neurons to each layer of the common DNN. For each task-specific DNN, we fix the common part and only pretrain the task-specific part layer by layer in an unsupervised manner. As the number of parameters of the task-specific part is only a small fraction of the entire network, we do not need a large number of samples for each task.

B. Co-Evolutionary Learning Algorithm

After unsupervised pretraining, we need to train the common regression layer and fine-tune all DNNs. We propose a co-evolutionary algorithm to simultaneously perform supervised training of the regressor and fine-tuning of all DNNs.

For supervised training of the regressor, because the regressor is on top of the fuzzy deep CAE, each actual output value of the regressor is a fuzzy number. As aforementioned, each expected output demand $\widehat{z}$ is also characterized by a fuzzy number $[\widehat{z}^L, \widehat{z}^U]$. For each actual output fuzzy number $z = [z^L, z^U]$, let $c(z)$ be the centroid of z, we measure the deviation of the z from the expected output value $\widehat{z}$ as

$$\mathrm{dis}(z, \widehat{z}) = \frac{1}{4}\left(\frac{c_l|z^L - \widehat{z}^L|}{\widehat{z}^L} + \frac{|z^U - \widehat{z}^U|}{\widehat{z}^U} + \frac{2|c(z) - c(\widehat{z})|}{c(\widehat{z})}\right) \tag{7}$$

where the coefficient c_l is 1 if $z^L \geq \widehat{z}^L$ and is 4 otherwise, such that the centroid distance plays a more important role than the lower limit distance and upper limit distance, and an estimated demand below the lower limit of the actual demand will be penalized more greatly.

The loss between an actual output fuzzy vector $\mathbf{z}$ and an expected output $\widehat{\mathbf{z}}$ is defined as the weighted deviations over ten types of materials ($D_o = 10$)

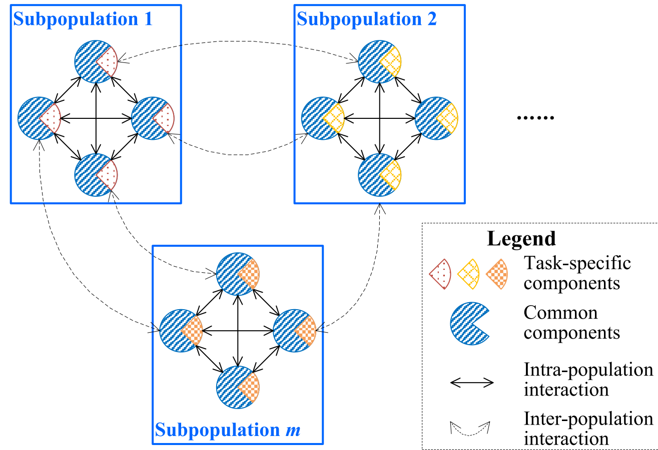

Fig. 2. Illustration of the co-evolutionary process for combined task-general and task-specific learning.

$$\min \mathcal{L}(\mathbf{z}, \widehat{\mathbf{z}}) = \sum_{d=1}^{D_o} w_d \mathrm{dis}(z_d, \widehat{z}_d) \tag{8}$$

where w_d is the importance weight of the dth type of materials. The weights are defined by the decision maker and normalized such that the sum of all weights equals to 1. The aim of supervised training of the regressor is to minimize the total losses on the labeled training set $\mathcal{D}^L$

$$\min J^S(\theta) = \sum_{(\mathbf{x}, \widehat{\mathbf{z}}) \in \mathcal{D}^L} \mathcal{L}\left(f'(f(\mathbf{x})), \widehat{z}\right). \tag{9}$$

For fine-tuning of each DNN for each task $\mathcal{T}_i$, we aim to minimize the reconstruction error over the training set $\mathcal{D}_i$ of all samples of task $\mathcal{T}_i$

$$J^{U(i)}(\theta) = \sum_{\mathbf{x} \in \mathcal{D}_i} \mathcal{L}\left(\mathbf{x}, f'(f(\mathbf{x}))\right). \tag{10}$$

To achieve the above two purposes, the proposed co-evolutionary algorithm uses m subpopulations, each of which uses an aggregate objective function as follows to optimize the pretrained prediction network of task $\mathcal{T}_i$ ($1 \leq i \leq m$, $m = 9$ in our study):

$$\min J^{(i)}(\theta) = J^{U(i)}(\theta) + \frac{|\mathcal{D}_i|}{|\mathcal{D}^L|} J^s(\theta) \tag{11}$$

where the ratio $|\mathcal{D}_i|/|\mathcal{D}^L|$ balances the two objectives according to the sizes of two training sets.

Each solution in a subpopulation has two parts: the first part consists of task-specific parameters of the DNN for the task and the second part consists of the parameters of the common regressor. The evolution of solutions consists of intrapopulation evolution and interpopulation evolution, as illustrated by Fig. 2.

1) Intrapopulation Evolution: Solutions in a subpopulation P_i evolve the network for the same task $\mathcal{T}_i$. At each iteration, each solution θ explores the solution space at each dimension j as follows:

$$\theta'_j = \theta_j + \lambda_\theta \cdot \mathrm{rand}(-1, 1) \tag{12}$$

where $\text{rand}(-1, 1)$ denotes a random number uniformly distributed in $(-1, 1)$, and λ_θ is for controlling the search range of θ. The initial value of λ_θ is set to 0.5; let $J^{(i)}_{\max}$ and $J^{(i)}_{\min}$ be the maximum and minimum reconstruction losses among all solutions in the subpopulation, respectively, and λ_θ is updated at each iteration as follows:

$$\lambda_\theta = \lambda_\theta \cdot \alpha^{-\left(J^{(i)}_{\max} - J^{(i)}(\theta) + \epsilon\right)/\left(J^{(i)}_{\max} - J^{(i)}_{\min} + \epsilon\right)} \tag{13}$$

where α is a control parameter set to 1.0026, and ϵ is a very small number to avoid division-by-zero. In this way, a fitter solution (with smaller $J^{(i)}(\theta)$ value) explores a smaller range to enhance local search, while a lower fitness solution explores a larger range to facilitate global search. Such an evolutionary mechanism is inspired by the motion of shallow water waves and proposed as the water wave optimization metaheuristic for solving optimization problems [67].

If a solution θ has not been improved after a predefined number of iterations, it is replaced by a new solution randomly generated between the old solution and the best known solution θ^* to avoid search stagnation

$$\theta_j = N\left(\frac{\theta^*_j + \theta_j}{2}, \frac{|\theta^*_j - \theta_j|}{2}\right) \tag{14}$$

where $N(\mu, \sigma)$ is a Gaussian random number with mean μ and standard deviation σ.

Whenever a new best known solution θ^* is found, the algorithm conducts a local search around θ^* by generating K neighbors, each of which is obtained by adding a small offset to a random dimension j a solitary wave $\mathbf{x}'$ as

$$\theta'_j = \theta_j + \beta \cdot N(0, 1) \tag{15}$$

where β is a control parameter that linearly decreases from 0.01 to 0.001 with iteration. The best neighbor, if better than θ^*, will replace θ^* in the subpopulation.

2) Interpopulation Evolution: Solutions in different subpopulations cooperatively evolve the common regressor shared among the tasks. Therefore, a solution with higher regression error can probably learn from solutions with lower regression errors in other solutions. Let $J^S_{\max}$ and $J^S_{\min}$ be the maximum and minimum regression errors among all solutions in all m subpopulation, respectively; at each iteration, we calculate a probability for each solution θ as follows:

$$p_\theta = \frac{J^S(\theta) - J^S_{\min} + \epsilon}{J^S_{\max} - J^S_{\min} + \epsilon} \tag{16}$$

such that a solution with higher regression error have higher p_θ value. At each iteration, each component of the common regression part of θ has a probability of p_θ being modified by learning from another solution θ' as follows:

$$\theta_j = \theta_j + \text{rand}(0, 1) \cdot \left(\theta'_j - \theta_j\right) \tag{17}$$

where the exemplar solution θ' is selected with a probability inversely proportional to $J^S(\theta')$, such that better solutions tend to export more features to other solutions [68].

Algorithm 1: Proposed Co-Evolutionary Algorithm for Mixed Feature and Regression Learning

for $i = 1$ *to* m **do**
  Randomly initialize a subpopulation P_i of solutions for task $\mathcal{T}_i$;
while *the stop criterion is not satisfied* **do**
  `// intra-population evolution`
  for $i = 1$ *to* m **do**
    Let θ^* be the best solution in subpopulation P_i;
    foreach θ *in subpopulation* P_i **do**
      Calculate λ_θ according to Eq. (13);
    foreach θ *in subpopulation* P_i **do**
      Produce an offspring θ' according to Eq. (12);
      if $J^{(i)}(\theta') < J^{(i)}(\theta)$ **then**
        $\theta \leftarrow \theta'$;
        if $f(\theta) < f(\theta^*)$ **then**
          $\theta^* \leftarrow \theta$;
          for $k = 1$ *to* K **do**
            Produce a neighbor θ' of θ^* according to Eq. (15);
            if $f(\theta') < f(\theta^*)$ **then** $\theta^* \leftarrow \theta'$;
      else
        if θ *remains for* $\hat{g}$ *consecutive iterations* **then**
          Reset θ according to Eq. (14);
  `// inter-population evolution`
  for $i = 1$ *to* m **do**
    foreach θ *in subpopulation* P_i **do**
      Calculate p_θ according to Eq. (16);
  for $i = 1$ *to* m **do**
    foreach θ *in subpopulation* P_i **do**
      foreach *dimension* j *in the second part of* θ **do**
        if $rand(0, 1) < p_\theta$ **then**
          Select another θ' with a probability inversely proportional to $J^S(\theta')$;
          Modify θ_j according to Eq. (17);

3) Algorithm Framework: The co-evolutionary algorithm alternatively performs an iteration of intrapopulation evolution and an iteration of interpopulation evolution until the stop condition is satisfied. Algorithm 1 presents the pseudocode of the algorithm. Let G be the maximum number of generations (stopping criterion), N_P be the (average) population size, and $O(\phi)$ be the time complexity of evaluating the objective function of model training; the average time complexity of the algorithm is $O(2GmN_P\phi)$ and the worst time complexity is $O(G(2m + K)N_P\phi)$ (in the case we obtain a new best solution at each generation, which is preferred because the algorithm converges rapidly and we can use a small number G of generations).

After evolution, we select the solution with best $J^S(\theta)$ among all subpopulations, and use its second part as the parameters of the common regressor. Then, for each task $\mathcal{T}_i$, we temporally combine the first part of each solution in the ith subpopulation with the selected second part, and select resulting solution with the best objective function value as the prediction network for the task.

TABLE II
NUMBERS OF LABELED/UNLABELED SAMPLES IN DATASETS OF THE EIGHT SOURCE TASKS

Task	Labeled	Unlabeled	Sum	Sources
Earthquake	13	21	34	Ministry of Emergency Management
Typhoon	37	41	78	
Flood	12	12	24	
Chemical-explosion	14	9	23	
SARS	10	7	17	Hubei, Zhejiang, Sichuang, Guangdong, GuangXi, Jiangxi provinces and Chongqing, Tianjin, Shanghai municipalities
Dengue fever	5	20	25	
H7N9	7	25	32	
H1N1	7	11	18	
Sum	105	146	251	

TABLE III
NUMBERS OF LABELED/UNLABELED SAMPLES IN DATASETS OF THE TARGET TASK

Stage	Week	Labeled	Unlabeled	Sum
1	#2	0	6	6
	#3	6	6	12
	#4	12	6	18
	#5	18	6	24
	#6	24	6	30
	#7	30	6	36
	#8	36	6	42
2	#1	42	2	44
	#2	44	2	46
	#3	46	2	48
	#4	48	2	50

IV. EXPERIMENTS

A. Experimental Settings

For source tasks, we collected datasets of material demands of four types of disasters and four types of other epidemics. The numeric values are normalized into [0,1] by dividing by the maximum value of each feature. There are around 10% feature missing values, for which we use average values of disasters of similar magnitudes. The labels (relief demands) are mainly from management departments. For missing labels, we estimate the demands based on the disaster magnitudes and number of victims; if the number of victims is still unknown or very inaccurate, the sample is considered as unlabeled. The data sources and number of samples of the datasets are summarized in Table II. As we can observe, the total number of samples was relatively small, the number of labeled samples was much less than that of unlabeled samples, and the number of samples of epidemics was much less than that of samples of disasters.

For the target task of material demand prediction for COVID-19, we applied our model in two stages.

1) The first stage was the peak of COVID-19 in China, 2020, where we collected datasets in six cities (Wuhan, Chengdu, Guangzhou, Chongqing, Hangzhou, and Wenzhou) during eight weeks from January 27 to March 22.
2) The second stage was the second wave of COVID-19 in China, 2021, where we collected datasets in two cities (Nanjing and Yangzhou) during four weeks from July 21 to August 1.

We obtained input data at the end of each week. However, determining the actual demands (labels) typically delays a week. Therefore, we trained the network since the second week of the first stage. Initially, we have only six unlabeled samples; on each next week, the number of labeled samples was increased by six; in the second stage, the number of labeled samples was increased by two on each week, as shown in Table III. At the beginning of the first stage, we have few experiences about the merging COVID-19, and the datasets have many missing values; with the progress of the epidemic, the portion of missing values decreases. In the second stage, the datasets are much completer and more accurate.

To validate the effectiveness of PFN parameters in our CETL model, we also implemented two variants: the first uses nonfuzzy parameters (denoted as CETL-NF) and the second uses regular fuzzy parameters (denoted as CETL-RF). The number of layers of the underlying DNN is five, and the numbers of neurons of the four hidden layers are 29, 17, 11, and 7, respectively, as a result of model structure tuning by evolutionary optimization [69].

For comparison, we tested a nontransfer-learning deep AE (DAE) model [70] and five popular transfer learning and/or multitask methods, including: 1) a shared-hidden-layer DNN (SHL-DNN) [71]; 2) a joint adaptation network (JAN) [72]; 3) a multitask Takagi-Sugeno-Kang fuzzy system (MTFS) [73]; 4) a distributed jointly sparse multitask (dFSMT) algorithm [42]; and 5) a fuzzy-residual-based transfer learning (ResTL) model [54], which are briefly described in Table IV. The nontransfer-learning methods only use the COVID-19 dataset. For each transfer learning method, we implement two version, the first only uses the datasets of epidemics (four source tasks), and the second uses the datasets of both epidemics and disasters (eight source tasks, denoted by a suffix "ex").

According to (7) and (8), we use $(1 - \mathrm{dis}_d)$ to evaluate the prediction accuracy of a material type and use $(\sum_{d=1}^{10} w_d dis_d)$ to evaluate the accuracy of a prediction. Moreover, we consider a prediction as "successful" if the predicted interval is in the range of the expected interval, and calculate the corresponding "success rate" over the test set. The experiments are conducted on a computer with i7-6500 CPU and 8-GB DDR3 memory.

To test the general applicability of the proposed CETL method, we also apply it to each of the eight tasks described in Table II, using other tasks as source tasks. Due to the page limit, the results are given in the supplementary material.

B. Experimental Results

For the first stage, Tables V–X present the prediction accuracies of the models in the six cities, respectively. In each column, the best accuracy among the 16 methods is shown in bold, a superscript $^+$ denotes that the result of CETLex is significantly better than that of the corresponding comparative method, a superscript $^-$ denotes vice versa, and otherwise there is no significant difference (with a confidence level of 95%, according to the rank-sum test). Fig. 3 illustrates the change of the average prediction accuracy of each model with week.

TABLE IV
COMPARATIVE DEEP AND/OR TRANSFER LEARNING MODELS

	Model	Ref.	Brief description
Non-transfer-learning	DAE	[70]	A basic deep AE model trained by HF algorithm [66].
Transfer-learning	SHL-DNN	[71]	A DAE trained sequentially by an unsupervised pre-training algorithm, a supervised back-propagation algorithm, and a supervised fine-tuning algorithm.
	JAN	[72]	A set of DNNs for learning transferable features by minimizing cross-domain discrepancy.
	MT-TSK-FS	[73]	A fuzzy system that is trained by balancing task-independent information and inter-task correlation information based on ϵ-insensitivity and L2-norm penalty.
	dFSMT	[42]	A multitask learning algorithm that exploits joint sparsity of the vectors of interest and uses adaptive cooperation strategy to consider intertask similarities.
	ResTL	[54]	A method that constructs the target model by adding fuzzy residual to a model-agnostic source model and reusing antecedent parameters of the source system to preserve the source distribution.

TABLE V
PREDICTION ACCURACIES (IN PERCENTS) OF THE COMPARATIVE METHODS IN WUHAN

Week	2	3	4	5	6	7	8
DAE		$^+$17.3	$^+$22.1	$^+$27.1	$^+$27.9	$^+$31.4	$^+$34.8
SHL-DNN	$^+$29.8	$^+$32.9	$^+$36.4	$^+$40.2	$^+$43.6	$^+$45.9	$^+$49.7
SHL-DNNex	$^+$36.5	$^+$40.5	$^+$44.6	$^+$48.5	$^+$50.5	$^+$58.1	$^+$62.0
JAN	$^+$26.8	$^+$28.9	$^+$33.7	$^+$36.2	$^+$38.7	$^+$40.3	$^+$41.5
JANex	$^+$33.2	$^+$38.6	$^+$41.5	$^+$43.2	$^+$46.6	$^+$50.1	$^+$53.8
MTFS	$^+$27.2	$^+$33.6	$^+$35.7	$^+$39.0	$^+$42.0	$^+$43.8	$^+$46.6
MTFSex	36.9	$^+$40.3	$^+$42.0	$^+$46.7	$^+$50.2	$^+$54.9	$^+$61.2
dFSMT	$^+$29.0	$^+$33.7	$^+$36.1	$^+$39.9	$^+$44.2	$^+$46.8	$^+$51.7
dFSMTex	$^+$35.8	$^+$40.2	$^+$43.4	$^+$47.5	$^+$53.9	$^+$59.0	$^+$63.7
ResTL	$^+$29.6	$^+$34.6	$^+$37.9	$^+$41.3	$^+$44.8	$^+$46.8	$^+$50.9
ResTLex	$^+$36.4	$^+$41.0	$^+$43.3	$^+$48.5	$^+$55.2	$^+$63.0	$^+$67.5
CETL-NF	$^+$30.0	$^+$33.9	$^+$36.0	$^+$40.9	$^+$42.3	$^+$47.5	$^+$51.6
CETL-NFex	$^+$34.5	$^+$38.9	$^+$43.5	$^+$46.0	$^+$57.3	$^+$66.4	$^+$69.8
CETL-RF	$^+$32.2	$^+$36.2	$^+$43.3	$^+$48.2	$^+$53.5	$^+$59.1	$^+$62.6
CETL-RFex	$^+$36.8	42.8	$^+$49.2	$^+$57.2	$^+$62.9	$^+$74.8	$^+$82.1
CETL	$^+$32.7	$^+$36.8	$^+$45.6	$^+$50.6	$^+$56.0	$^+$60.6	$^+$64.8
CETLex	**37.3**	**43.9**	**52.6**	**63.3**	**67.5**	**77.2**	**85.1**

TABLE VII
PREDICTION ACCURACIES (IN PERCENTS) OF THE COMPARATIVE METHODS IN GUANGZHOU

Week	2	3	4	5	6	7	8
DAE		$^+$16.2	$^+$18.5	$^+$23.2	$^+$25.5	$^+$29.1	$^+$33.2
SHL-DNN	$^+$26.4	$^+$29.7	$^+$35.6	$^+$40.4	$^+$42.3	$^+$45.4	$^+$50.5
SHL-DNNex	$^+$35.8	$^+$37.4	$^+$42.6	$^+$45.0	$^+$50.2	$^+$54.8	$^+$62.6
JAN	$^+$26.3	$^+$30.6	$^+$32.8	$^+$34.4	$^+$39.5	$^+$43.7	$^+$49.8
JANex	$^+$34.4	$^+$36.9	$^+$38.1	$^+$47.2	$^+$52.4	$^+$53.9	$^+$56.9
MTFS	$^+$30.0	$^+$32.4	$^+$36.0	$^+$39.6	$^+$44.8	$^+$47.0	$^+$50.2
MTFSex	$^+$36.4	$^+$41.6	$^+$44.7	$^+$46.9	$^+$50.8	$^+$57.3	$^+$60.9
dFSMT	$^+$30.8	$^+$32.9	$^+$36.7	$^+$39.2	$^+$44.2	$^+$48.9	$^+$53.3
dFSMTex	$^+$35.2	$^+$40.8	$^+$45.9	$^+$49.3	$^+$55.4	$^+$59.6	$^+$65.6
ResTL	$^+$30.6	$^+$33.8	$^+$36.9	$^+$41.8	$^+$45.4	$^+$49.2	$^+$57.1
ResTLex	$^+$36.7	42.0	$^+$48.2	$^+$52.5	$^+$57.2	$^+$66.8	$^+$73.2
CETL-NF	$^+$29.8	$^+$33.1	$^+$37.4	$^+$40.3	$^+$43.1	$^+$49.5	$^+$55.8
CETL-NFex	$^+$36.0	42.9	$^+$49.2	$^+$60.1	$^+$65.6	$^+$76.5	$^+$79.9
CETL-RF	$^+$33.3	$^+$36.4	$^+$39.5	$^+$44.7	$^+$49.2	$^+$52.0	$^+$56.9
CETL-RFex	$^+$37.5	**44.0**	$^+$49.2	$^+$59.8	$^+$67.9	78.2	83.3
CETL	$^+$33.6	$^+$37.1	$^+$40.3	$^+$46.0	$^+$49.5	$^+$53.5	$^+$58.4
CETLex	**38.9**	42.9	**51.1**	**63.4**	**70.0**	**79.5**	**84.5**

TABLE VI
PREDICTION ACCURACIES (IN PERCENTS) OF THE COMPARATIVE METHODS IN CHENGDU

Week	2	3	4	5	6	7	8
DAE		$^+$18.1	$^+$22.2	$^+$28.6	$^+$30.1	$^+$34.8	$^+$36.5
SHL-DNN	$^+$29.0	$^+$32.3	$^+$37.0	$^+$38.9	$^+$41.8	$^+$46.6	$^+$49.5
SHL-DNNex	$^+$37.3	$^-$**43.8**	**47.3**	$^+$51.4	$^+$55.0	$^+$58.3	$^+$63.6
JAN	$^+$28.1	$^+$31.6	$^+$33.7	$^+$37.8	$^+$39.4	$^+$41.8	$^+$43.1
JANex	$^+$34.1	$^+$36.7	$^+$40.6	$^+$44.6	$^+$48.2	$^+$51.3	$^+$55.0
MTFS	$^+$25.6	$^+$30.3	$^+$32.8	$^+$35.5	$^+$40.9	$^+$46.8	$^+$49.3
MTFSex	$^+$36.4	$^+$39.0	$^+$42.6	$^+$44.5	$^+$49.8	$^+$55.5	$^+$60.8
dFSMT	$^+$27.0	$^+$30.9	$^+$34.8	$^+$38.5	$^+$42.2	$^+$48.1	$^+$49.7
dFSMTex	$^+$35.3	$^+$37.9	$^+$43.0	$^+$44.9	$^+$50.1	$^+$53.9	$^+$62.0
ResTL	$^+$30.8	$^+$34.1	$^+$39.0	$^+$44.8	$^+$46.3	$^+$47.7	$^+$53.3
ResTLex	$^+$33.9	$^+$37.9	$^+$44.4	$^+$46.5	$^+$52.3	$^+$55.5	$^+$64.0
CETL-NF	$^+$31.2	$^+$33.8	$^+$39.9	$^+$44.2	$^+$45.8	$^+$47.3	$^+$51.9
CETL-NFex	$^+$36.4	$^+$38.9	$^+$44.9	$^+$53.0	$^+$55.5	$^+$59.5	$^+$68.8
CETL-RF	$^+$33.7	$^+$35.1	$^+$44.7	$^+$45.8	$^+$51.2	$^+$55.0	$^+$59.6
CETL-RFex	37.8	$^+$39.6	$^+$46.6	$^+$52.5	$^+$63.8	$^+$70.8	79.7
CETL	$^+$33.9	$^+$36.9	$^+$46.4	$^+$48.3	$^+$54.8	$^+$57.1	$^+$63.3
CETLex	**38.5**	42.1	46.8	**58.1**	**66.2**	**76.5**	**80.9**

TABLE VIII
PREDICTION ACCURACIES (IN PERCENTS) OF THE COMPARATIVE METHODS IN CHONGQING

Week	2	3	4	5	6	7	8
DAE		$^+$19.5	$^+$21.4	$^+$24.7	$^+$28.9	$^+$30.3	$^+$35.6
SHL-DNN	$^+$31.6	$^+$37.2	$^+$38.9	$^+$39.8	$^+$43.3	$^+$44.9	$^+$49.6
SHL-DNNex	36.5	$^+$38.8	$^+$45.2	$^+$48.3	$^+$49.9	$^+$52.2	$^+$56.4
JAN	$^+$26.0	$^+$33.3	$^+$35.1	$^+$37.8	$^+$36.9	$^+$43.0	$^+$46.5
JANex	$^+$32.9	$^+$36.7	$^+$42.5	$^+$42.8	$^+$45.2	$^+$45.9	$^+$50.6
MTFS	$^+$26.4	$^+$30.2	$^+$39.6	$^+$41.6	$^+$45.5	$^+$47.3	$^+$49.5
MTFSex	$^+$33.7	$^+$35.9	$^+$39.4	$^+$46.1	$^+$57.8	$^+$60.2	$^+$59.2
dFSMT	$^+$27.3	$^+$32.5	$^+$38.9	$^+$41.7	$^+$43.6	$^+$46.3	$^+$48.2
dFSMTex	$^+$32.2	$^+$37.4	$^+$41.1	$^+$45.0	$^+$50.0	$^+$56.3	$^+$58.2
ResTL	$^+$27.4	$^+$31.3	$^+$34.4	$^+$36.3	$^+$42.6	$^+$47.1	$^+$49.9
ResTLex	$^+$33.2	$^+$39.6	$^+$45.7	$^+$46.7	$^+$53.0	$^+$59.0	$^+$60.9
CETL-NF	$^+$28.9	$^+$30.6	$^+$33.7	$^+$35.0	$^+$39.2	$^+$44.7	$^+$46.6
CETL-NFex	$^+$35.5	$^+$39.3	$^+$43.0	$^+$45.9	$^+$53.6	$^+$61.0	$^+$62.5
CETL-RF	$^+$29.8	$^+$31.0	$^+$35.0	$^+$36.8	$^+$39.0	$^+$47.1	$^+$53.3
CETL-RFex	$^+$35.9	$^+$45.6	$^+$50.5	$^+$54.2	$^+$61.0	$^+$64.8	$^+$70.9
CETL	$^+$29.6	$^+$34.3	$^+$37.1	$^+$39.2	$^+$46.9	$^+$50.8	$^+$55.5
CETLex	**36.9**	**47.7**	**53.1**	**59.4**	**63.9**	**67.2**	**74.0**

As there are no labeled training samples during the first two weeks, we only test the nontransfer-learning DAE since the third week, where it always exhibits the lowest accuracy: its average accuracy is only 18.7% at the third week; although the accuracy increases with the weekly increment of the number of samples, it only reaches around 35% at the eighth week, which is still too low to support public health decisions. By utilizing samples of other epidemics/disasters, all other models achieve significantly higher accuracies than the nontransfer-learning model, indicating that the knowledge about material demand from the other tasks can be effectively used in learning material demand knowledge for the emerging COVID-19. In particular, for each transfer/multitask learning model, it achieves a significantly higher accuracy on the instance trained on both epidemic and disaster datasets than that trained on solely epidemic datasets, which demonstrates that the domains

TABLE IX
PREDICTION ACCURACIES (IN PERCENTS) OF THE COMPARATIVE METHODS IN HANGZHOU

Week	2	3	4	5	6	7	8
DAE		$^{+}$18.6	$^{+}$22.4	$^{+}$24.7	$^{+}$26.5	$^{+}$28.0	$^{+}$31.1
SHL-DNN	$^{+}$30.4	$^{+}$28.9	$^{+}$33.4	$^{+}$38.1	$^{+}$50.5	$^{+}$48.3	$^{+}$55.1
SHL-DNNex	$^{+}$32.1	$^{+}$37.7	$^{+}$46.7	$^{+}$52.0	$^{+}$54.9	$^{+}$58.3	$^{+}$65.0
JAN	$^{+}$24.8	$^{+}$29.2	$^{+}$30.8	$^{+}$33.0	$^{+}$39.1	$^{+}$42.7	$^{+}$48.3
JANex	$^{+}$33.5	$^{+}$36.4	$^{+}$40.1	$^{+}$46.3	$^{+}$49.3	$^{+}$52.8	$^{+}$56.0
MTFS	$^{+}$26.1	$^{+}$32.6	$^{+}$38.1	$^{+}$43.5	$^{+}$46.0	$^{+}$48.2	$^{+}$49.7
MTFSex	$^{+}$33.7	$^{+}$38.4	$^{+}$40.2	$^{+}$44.1	$^{+}$47.0	$^{+}$56.5	$^{+}$62.5
dFSMT	$^{+}$27.3	$^{+}$33.0	$^{+}$40.1	$^{+}$44.5	$^{+}$48.3	$^{+}$51.1	$^{+}$51.6
dFSMTex	$^{+}$34.9	$^{+}$35.6	$^{+}$42.5	$^{+}$47.7	$^{+}$52.0	$^{+}$54.3	$^{+}$63.8
ResTL	$^{+}$29.1	$^{+}$33.8	$^{+}$43.0	$^{+}$50.6	$^{+}$51.5	$^{+}$55.2	$^{+}$56.0
ResTLex	$^{+}$36.0	$^{+}$36.9	$^{+}$44.0	$^{+}$48.9	$^{+}$59.7	$^{+}$64.3	$^{+}$66.2
CETL-NF	$^{+}$27.8	$^{+}$34.3	$^{+}$42.0	$^{+}$47.4	$^{+}$51.1	$^{+}$53.2	$^{+}$54.3
CETL-NFex	$^{+}$35.1	$^{+}$43.2	$^{+}$49.0	$^{+}$54.5	$^{+}$58.3	$^{+}$65.1	$^{+}$70.2
CETL-RF	$^{+}$30.9	$^{+}$37.6	$^{+}$44.1	$^{+}$48.2	$^{+}$51.7	$^{+}$56.3	$^{+}$57.9
CETL-RFex	38.2	**47.0**	53.7	58.4	$^{+}$62.8	$^{+}$68.8	$^{+}$78.3
CETL	$^{+}$32.1	$^{+}$39.0	$^{+}$45.3	$^{+}$50.2	$^{+}$52.5	$^{+}$55.8	$^{+}$58.5
CETLex	**38.4**	46.2	**54.5**	**59.8**	**64.9**	**73.9**	**82.0**

TABLE X
PREDICTION ACCURACIES (IN PERCENTS) OF THE COMPARATIVE METHODS IN WENZHOU

Week	2	3	4	5	6	7	8
DAE		$^{+}$22.2	$^{+}$25.7	$^{+}$29.1	$^{+}$33.2	$^{+}$36.0	$^{+}$38.6
SHL-DNN	$^{+}$28.7	$^{+}$38.5	$^{+}$45.1	$^{+}$48.4	$^{+}$49.7	$^{+}$50.5	$^{+}$54.3
SHL-DNNex	$^{+}$33.1	$^{+}$40.0	$^{+}$46.9	$^{+}$53.5	$^{+}$55.9	$^{+}$57.8	$^{+}$60.0
JAN	$^{+}$27.2	$^{+}$34.0	$^{+}$36.9	$^{+}$39.5	$^{+}$45.0	$^{+}$47.2	$^{+}$50.1
JANex	$^{+}$29.3	$^{+}$42.1	$^{+}$46.5	$^{+}$47.8	$^{+}$52.3	$^{+}$55.0	$^{+}$56.8
MTFS	$^{+}$29.3	$^{+}$35.3	$^{+}$37.1	$^{+}$44.0	$^{+}$46.9	$^{+}$49.7	$^{+}$50.6
MTFSex	$^{+}$32.8	$^{+}$37.6	$^{+}$48.0	$^{+}$53.1	$^{+}$56.7	$^{+}$58.5	$^{+}$59.2
dFSMT	$^{+}$30.3	$^{+}$36.2	$^{+}$38.3	$^{+}$39.3	$^{+}$45.6	$^{+}$49.1	$^{+}$51.0
dFSMTex	$^{+}$33.1	$^{+}$39.2	50.0	$^{+}$55.1	$^{+}$53.2	$^{+}$63.2	$^{+}$61.9
ResTL	$^{+}$28.8	$^{+}$33.6	$^{+}$37.2	$^{+}$37.8	$^{+}$48.7	$^{+}$52.0	$^{+}$53.3
ResTLex	$^{+}$33.3	$^{+}$41.0	$^{+}$48.6	$^{+}$56.5	$^{+}$59.4	$^{+}$64.2	$^{+}$69.6
CETL-NF	$^{+}$27.3	$^{+}$32.6	$^{+}$34.0	$^{+}$40.3	$^{+}$45.3	$^{+}$48.2	$^{+}$52.5
CETL-NFex	$^{+}$33.8	$^{+}$36.2	$^{+}$43.9	$^{+}$48.8	$^{+}$57.6	$^{+}$65.9	$^{+}$71.0
CETL-RF	$^{+}$28.8	$^{+}$36.0	$^{+}$38.3	$^{+}$42.9	$^{+}$46.7	$^{+}$50.1	$^{+}$53.4
CETL-RFex	$^{+}$35.5	$^{+}$42.5	$^{+}$47.7	$^{+}$55.3	61.6	$^{+}$70.3	$^{+}$76.4
CETL	$^{+}$29.1	$^{+}$36.5	$^{+}$39.2	$^{+}$43.5	$^{+}$46.5	$^{+}$50.4	$^{+}$54.2
CETLex	**37.6**	**45.8**	**51.4**	**58.8**	**63.0**	**73.3**	**79.0**

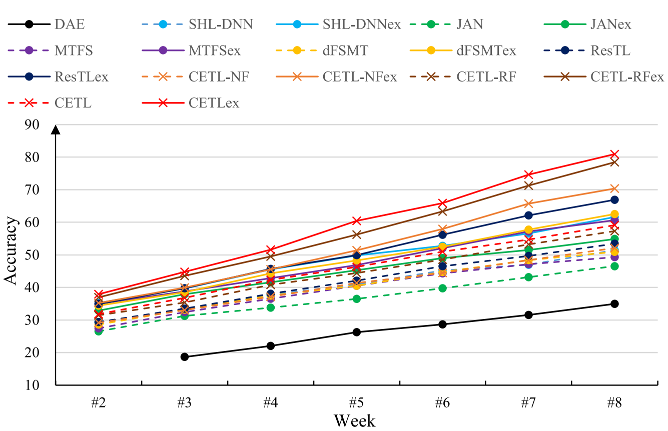

Fig. 3. Average prediction accuracies of the comparative methods on the test set of the first stage.

of natural and manmade disasters can share knowledge about material demands with those of epidemics.

The proposed CETLex obtains the highest prediction accuracies on all seven weeks in Wuhan, five of seven weeks in Chengdu, six of seven weeks in Guangzhou, all seven weeks in Chongqing, six of seven weeks in Hangzhou, and all seven weeks in Wenzhou. In more details, SHL-DNN obtains the highest prediction accuracies on weeks 3 and 4 in Chengdu, CETL-RFex obtains the highest prediction accuracy on week 3 in Guangzhou and week 3 in Hangzhou. Except that on week 3 in Chengdu the result of SHL-DNN is significantly better than that of CETLex, in other three cases, there are no statistically significant differences between the CETLex and the best comparative method. Among all 42 instances (seven weeks and six cities), CETLex performs significantly better than SHL-DNNex on 39 instances, better than JAN on 42 instances, better than MTFSex on 41 instances, better than dFSMex on 42 instances, and better than ResTLex on 41 instances. At the second week, the difference of the accuracies of CETLex and some popular models are not large: SHL-DNNex, MTFSex, dFSMex, and ResTLex are around 35%, and the accuracy of CETLex is near to 38%. However, the performance advantages of CETLex increase with the number of weeks; at the eighth week, the accuracies of those popular transfer/multitask learning models are around 60%–65%, while the accuracy of CETLex is near to 81%. The main reasons that CETLex improves its prediction accuracy more rapidly with time are as follows.

1) Compared to the learning mechanisms of the other popular transfer learning models that focus on the utilizing of knowledge of target tasks to improve the learning performance of the source tasks, the proposed coevolution strategy can make feature learning in different domains benefit from each other.
2) Compared to the learning mechanisms of the other popular multitask learning models that require (relatively) balanced data among the tasks, the proposed strategy can effectively handle imbalanced data (e.g., the labeled samples of epidemics are quite few) by using coevolutionary optimization to explore the solution space more thoroughly.

Therefore, in the proposed model, with the relatively small increase of the number of samples of the target task, the learning performances on other tasks simultaneously increase, which can feedback to further improve the model performance on the target COVID-19 task.

Compared to the CETL-NFex version using nonfuzzy parameters, the CETL-RFex version using regular fuzzy parameters achieves an accuracy improvement of around 4.7%; moreover, the CETLex version using PFN parameters achieves an accuracy improvement of around 4.7% over CETL-RFex. This demonstrates that, using fuzzy parameters in the coevolutionary learning model improves its capability of coping with imprecise and uncertain information contained in the datasets, and using PFN parameters with larger membership grades further improves the model learning abilities.

There are some small performance variances among the models in different cities. For example, the average accuracy of SHL-DNNex is better than that of MTFSex in Wuhan and Chengdu, but is worse than MTFSex in Chongqing. However, in general, the performance of the prediction models does not change greatly in different cities. In other words,

TABLE XI
PREDICTION SUCCESS RATES (%) THE COMPARATIVE METHODS FOR THE TEN TYPES OF MATERIALS

Material	d. m. masks	d. m. gloves	goggles	p. clothes	O. bottles	s. disinfectant	d. machines	h. sanitizer	sprayers	t. guns
DAE	36.11	38.89	30.56	33.33	30.56	44.44	33.33	41.67	38.89	33.33
SHL-DNN	50.00	52.38	45.24	47.62	50.00	47.62	47.62	54.76	52.38	47.62
SHL-DNNex	66.67	69.05	59.52	61.90	59.52	64.29	59.52	66.67	**71.43**	57.14
JAN	45.24	45.24	38.10	42.86	40.48	42.86	40.48	42.86	45.24	40.48
JANex	59.52	64.29	52.38	54.76	52.38	59.52	52.38	57.14	57.14	45.24
MTFS	52.38	57.14	42.86	45.24	52.38	50.00	50.00	50.00	52.38	57.14
MTFSex	69.05	69.05	64.29	69.05	61.90	71.43	64.29	61.90	64.29	61.90
dFSMT	54.76	57.14	45.24	45.24	47.62	52.38	52.38	52.38	47.62	50.00
dFSMTex	71.43	71.43	61.90	66.67	61.90	73.81	59.52	61.90	61.90	59.52
ResTL	57.14	61.90	50.00	47.62	54.76	52.38	50.00	57.14	54.76	54.76
ResTLex	73.81	76.19	61.90	73.81	66.67	69.05	64.29	66.67	64.29	57.14
CETL-NF	52.38	57.14	45.24	42.86	50.00	50.00	47.62	47.62	47.62	47.62
CETL-NFex	73.81	**73.81**	61.90	69.05	66.67	69.05	61.90	64.29	64.29	59.52
CETL-RF	59.52	64.29	54.76	54.76	57.14	59.52	52.38	57.14	57.14	54.76
CETL-RFex	80.95	80.95	69.05	80.95	71.43	78.57	73.81	71.43	69.05	61.90
CETL	59.52	64.29	54.76	57.14	59.52	64.29	52.38	59.52	57.14	54.76
CETLex	**83.33**	**85.71**	**73.81**	**80.95**	**76.19**	**83.33**	**76.19**	**73.81**	69.05	**64.29**

TABLE XII
PREDICTION ACCURACIES (%) OF THE COMPARATIVE METHODS IN NANJING

Week	1	2	3	4
DAE	$^+$28.8	$^+$30.3	$^+$32.9	$^+$36.8
SHL-DNN	$^+$42.5	$^+$45.8	$^+$49.7	$^+$54.1
SHL-DNNex	$^+$50.7	$^+$55.3	$^+$60.1	$^+$67.6
JAN	$^+$37.3	$^+$40.2	$^+$44.1	$^+$48.5
JANex	$^+$46.5	$^+$52.2	$^+$54.9	$^+$59.4
MTFS	$^+$39.6	$^+$40.9	$^+$44.8	$^+$47.7
MTFSex	$^+$45.0	$^+$50.5	$^+$53.0	$^+$56.8
dFSMT	$^+$49.0	$^+$49.9	$^+$53.4	$^+$60.3
dFSMTex	$^+$59.8	$^+$62.2	$^+$68.0	$^+$72.5
ResTL	$^+$42.2	$^+$50.1	$^+$51.5	$^+$55.7
ResTLex	$^+$57.1	$^+$66.3	$^+$68.9	$^+$70.2
CETL-NF	$^+$52.5	$^+$55.5	$^+$58.2	$^+$64.8
CETL-NFex	$^+$64.0	$^+$69.1	$^+$72.3	$^+$78.7
CETL-RF	$^+$55.5	$^+$59.2	$^+$63.1	$^+$65.0
CETL-RFex	$^+$66.9	$^+$71.1	$^+$74.5	$^+$80.3
CETL	$^+$56.4	$^+$58.9	$^+$66.2	$^+$67.7
CETLex	**73.2**	**77.5**	**78.9**	**82.4**

TABLE XIII
PREDICTION ACCURACIES (%) OF THE COMPARATIVE METHODS IN YANGZHOU

Week	1	2	3	4
DAE	$^+$29.5	$^+$34.6	$^+$35.6	$^+$38.7
SHL-DNN	$^+$40.8	$^+$44.6	$^+$50.1	$^+$55.3
SHL-DNNex	$^+$51.2	$^+$59.7	$^+$64.8	$^+$74.5
JAN	$^+$40.0	$^+$42.0	$^+$43.9	$^+$47.8
JANex	$^+$47.7	$^+$53.7	$^+$60.1	$^+$69.3
MTFS	$^+$38.5	$^+$42.2	$^+$47.2	$^+$49.9
MTFSex	$^+$50.9	$^+$54.7	$^+$57.4	$^+$64.6
dFSMT	$^+$49.2	$^+$54.0	$^+$57.1	$^+$58.5
dFSMTex	$^+$56.7	$^+$63.3	$^+$70.9	$^+$76.9
ResTL	$^+$44.8	$^+$46.6	$^+$52.6	$^+$59.1
ResTLex	$^+$59.3	$^+$65.8	$^+$71.7	$^+$73.7
CETL-NF	$^+$48.9	$^+$53.7	$^+$59.9	$^+$66.6
CETL-NFex	$^+$69.3	$^+$73.1	$^+$75.0	$^+$77.2
CETL-RF	$^+$52.4	$^+$58.1	$^+$64.1	$^+$66.3
CETL-RFex	$^+$70.5	77.9	79.2	$^+$80.9
CETL	$^+$55.9	$^+$59.7	$^+$64.5	$^+$68.0
CETLex	**75.7**	**79.3**	**80.8**	**83.1**

different input values, such as the number of cases and infection rate, do not seriously affect the model performance. Among all models, our CETL model using extensive datasets achieves the highest prediction accuracy in each city.

Table XI presents the prediction success rates the comparative methods for the ten types of materials in the first stage. The proposed CETLex model obtains the highest success rates for nine types of material, expect that for sprayers, SHL-DNNex obtains the highest success rate of 71.43% that is a bit higher that 69.05% of CETLex. Among the ten types of materials, the success rates of CETLex are over 80% for four types, between 70%–80% for four types, and between 60%–70% for two types. In particular, for disposable medical masks, disposable medical gloves, goggles, protective clothing, and surface disinfectant, which are the most important and widely used types of materials in epidemic prevention and control, the success rates of CETLex are 83.33%, 85.71%, 80.95%, and 83.33%, respectively.

For the second stage, Tables XII and XIII present the prediction accuracies of the models in Nanjing and Yangzhou, respectively, and Fig. 4 illustrates the change of the average prediction accuracy of each model with week. In these two cities, CETLex always obtains the highest prediction accuracy, which is significantly better than the other popular learning

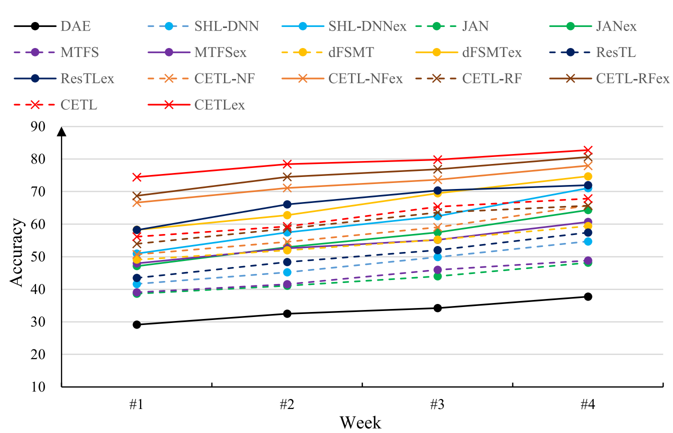

Fig. 4. Average prediction accuracies of the comparative methods on the test set of the second stage.

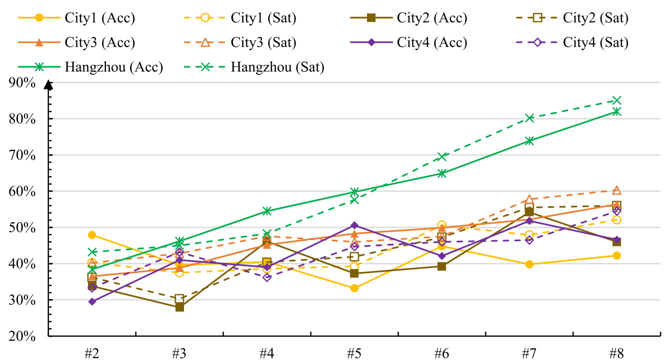

Fig. 5. Relationship between prediction accuracies (acc) and demand satisfaction rates (sat) of medical materials in five cities in Zhejiang Province, during the peak of COVID-19 in China, 2020.

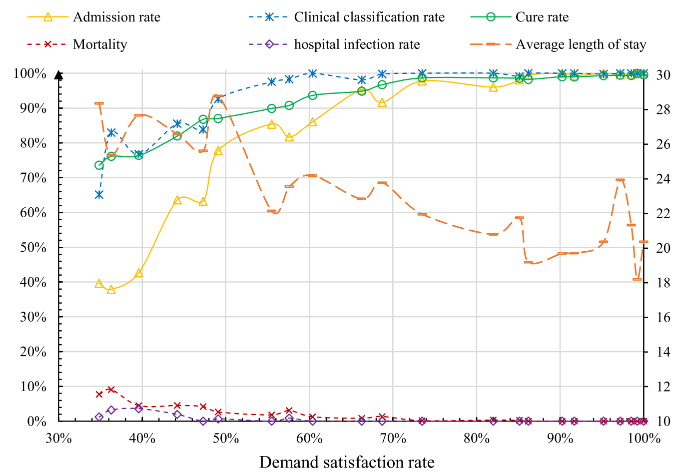

Fig. 6. Relationship between demand satisfaction rate and six medical quality control indices for COVID-19.

models selected from the literature. On the basis of the learning of the first stage, in the second stage, the proposed model exhibits a high accuracy of around 75% at the first week and achieves an average accuracy of around 80%.

C. Model Effectiveness

In general, underestimation of the demands would cause supply shortages and, hence, reduce the capacity of the healthcare system, while overestimation would cause waste of resources and economic losses and, when the total supply is insufficient, oversupply in a region often causes undersupply in other regions. However, it is difficult to quantitatively measure the effect of prediction accuracy on the effectiveness of epidemic prevention and control, because the effectiveness is affected by many other factors, such as the transmission rate, isolation strength, and existing medical capacity. Here, we evaluate the effectiveness of the proposed prediction model in COVID-19 in two parts. First, we evaluate the relationship between the prediction accuracy and demand satisfaction rate in hospitals. The data were from seven first-class hospitals, three in Hangzhou and four in other four cities in Zhejiang Province, during the first stage. The transmission rates and public health measures in these cities were similar. Medical material supply decisions in Hangzhou were based on the prediction results of our model, while those in the other cities were mainly based on manual experience or some basic regression models. Fig. 5 presents the prediction accuracies and demand satisfaction rates in the five cities, showing that the demand satisfaction rate is generally proportional to prediction accuracy; due to the higher prediction accuracy of CETL, the demand satisfaction rate in Hangzhou was significantly higher than those in the other four cities since the third week.

Second, we evaluate the relationship between the demand satisfaction rate and six key medical quality control indices for COVID-19: 1) hospital admission rate; 2) clinical classification rate; 3) cure rate; 4) mortality; 5) hospital infection rate; and 6) average length of stay (measured per week). Fig. 6 presents 22 data points collected from four hospitals (two in Wuhan and two in Hangzhou) during the first stage. The results shows that, in general, the hospital admission rate, clinical classification rate, and cure rate increase with the demand satisfaction rate, while mortality and hospital infection rate decrease with the demand satisfaction rate, demonstrating that there is a positive correlation between the demand satisfaction rate and the effectiveness of COVID-19 treatment in hospitals. Particularly, when the satisfaction rate exceeds 60%, the cure rate is over 30%, mortality is below 1.5%, and hospital infection rate is 0.

In summary, the evaluation results demonstrate that the prediction accuracy is positively related to the demand satisfaction rate, and the demand satisfaction rate is positively related to the effectiveness of COVID-19 treatment in hospitals in terms of the cure rate, mortality, and hospital infection rate.

V. Conclusion and Discussion

The article presented a CETL method for predicting medical material demands for COVID-19 prevention and control by utilizing knowledge from related tasks, including other epidemics and disasters. CETL uses multiple subpopulations to simultaneously evolve prediction networks for different prediction tasks and employs interpopulation evolution to cooperatively learn common knowledge shared across the tasks. Experimental results demonstrated that CETL achieves high prediction accuracies compared to the state of the arts. As CETL was mainly proposed for learning in a task with very few data by utilizing data from a set of related tasks that are also with few or relatively few data, there are many potential application areas, such as predicting the demands of police resources for anti-terrorism operations by co-evolutionary learning of tasks of other violent crimes, predicting the demands of unmanned aerial vehicle (UAV) resources in search and rescue by co-evolutionary learning of other UAV tasks [74], etc.

One shortage of the proposed method is that the co-evolutionary algorithm uses a large number of iterations and a relatively large population size and, therefore, the required computational resources are relatively large. In our experiments, a single run of training a transfer learning model from the literature on the entire test set typically requires 30 min to an hour, while that of training CETL requires around 4–6 h (although this computational time is affordable in the task and worth paying for the high prediction accuracies). Currently, we are integrating reinforcement learning into co-evolutionary optimization [75] to accelerate the algorithm.

REFERENCES

[1] C. Anastassopoulou, L. Russo, A. Tsakris, and C. Siettos, "Data-based analysis, modelling and forecasting of the COVID-19 outbreak," *PLoS One*, vol. 15, no. 3, 2020, Art. no. e0230405.

[2] J. Cao, X. Jiang, and B. Zhao, "Mathematical modeling and epidemic prediction of COVID-19 and its significance to epidemic prevention and control measures," *J. BioMed. Res. Innov.*, vol. 1, no. 1, pp. 1–19, 2020.

[3] V. K. R. Chimmula and L. Zhang, "Time series forecasting of COVID-19 transmission in Canada using LSTM networks," *Chaos Solitons Fractals*, vol. 135, Jun. 2020, Art. no. 109864.

[4] Z. Zhao, X. Li, F. Liu, G. Zhu, C. Ma, and L. Wang, "Prediction of the COVID-19 spread in African countries and implications for prevention and control: A case study in South Africa, Egypt, Algeria, Nigeria, Senegal and Kenya," *Sci. Total Environ.*, vol. 729, Aug. 2020, Art. no. 138959.

[5] N. Zheng *et al.*, "Predicting COVID-19 in China using hybrid AI model," *IEEE Trans. Cybern*, vol. 50, no. 7, pp. 2891–2904, Jul. 2020.

[6] J. V. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *J. Clin. Epidemiol.*, vol. 49, no. 11, pp. 1225–1231, 1996.

[7] M. Yang, J. Liu, L. Chen, Z. Zhao, X. Chen, and Y. Shen, "An advanced deep generative framework for temporal link prediction in dynamic networks," *IEEE Trans. Cybern.*, vol. 50, no. 12, pp. 4946–4957, Dec. 2020, doi: 10.1109/TCYB.2019.2920268.

[8] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[9] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Zhang, "Transfer learning using computational intelligence: A survey," *Knowl.-Based Syst.*, vol. 80, pp. 14–23, May 2015.

[10] Y.-J. Zheng, S.-Y. Chen, and H.-F. Ling, "Evolutionary optimization for disaster relief operations: A survey," *Appl. Soft Comput.*, vol. 27, pp. 553–566, Feb. 2015.

[11] Y.-J. Zheng, S.-Y. Chen, Y. Xue, and J.-Y. Xue, "A pythagorean-type fuzzy deep denoising autoencoder for industrial accident early warning," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 6, pp. 1561–1575, Dec. 2017.

[12] Q. Song, Y.-J. Zheng, Y. Huang, Z.-G. Xu, W.-G. Sheng, and J. Yang, "Emergency drug procurement planning based on big-data driven morbidity prediction," *IEEE Trans. Ind. Informat.*, vol. 15, no. 12, pp. 6379–6388, Dec. 2019.

[13] T. S. Brett and P. Rohani, "Transmission dynamics reveal the impracticality of COVID-19 herd immunity strategies," *Proc. Nat. Acad. Sci.*, vol. 117, no. 41, pp. 25897–25903, 2020.

[14] E. G. Miller, N. E. Matsakis, and P. A. Viola, "Learning from one example through shared densities on transforms," in *Proc. IEEE Conf. CVPR*, vol. 1, 2000, pp. 464–471.

[15] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Proc. 31st Int. Conf. NIPS*, 2017, pp. 4080–4090.

[16] Q. Sun, Y. Liu, T.-S. Chua, and B. Schiele, "Meta-transfer learning for few-shot learning," in *Proc. IEEE Conf. CVPR*, 2019, pp. 403–412.

[17] B. Bakker and T. Heskes, "Task clustering and gating for Bayesian multitask learning," *J. Mach. Learn. Res.*, vol. 4, pp. 83–99, Dec. 2003.

[18] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning," in *Proc. Int. Conf. Unsupervised Transf. Learn. Workshop*, vol. 27, 2012, pp. 17–37.

[19] A. Sufian, A. Ghosh, A. S. Sadiq, and F. Smarandache, "A survey on deep transfer learning to edge computing for mitigating the COVID-19 pandemic," *J. Syst. Arch.*, vol. 108, Sep. 2020, Art. no. 101830.

[20] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. Soufi, "Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101794.

[21] V. Arora, E. Y.-K. Ng, R. S. Leekha, M. Darshan, and A. Singh, "Transfer learning-based approach for detecting COVID-19 ailment in lung CT scan," *Comput. Bio. Med.*, vol. 135, Aug. 2021, Art. no. 104575.

[22] S. Basu, S. Mitra, and N. Saha, "Deep learning for screening COVID-19 using chest X-ray images," in *Proc. IEEE Symp. Series Comput. Intell.*, 2020, pp. 1–6.

[23] E. F. Ohata *et al.*, "Automatic detection of COVID-19 infection using chest X-ray images through transfer learning," *IEEE/CAA J. Automatica Sinica*, vol. 8, no. 1, pp. 239–248, Jan. 2021.

[24] J. Cohen *et al.*, "Predicting COVID-19 pneumonia severity on chest X-ray with deep learning," *Cureus*, vol. 12, no. 7, 2020, Art. no. e9448.

[25] N. E. M. Khalifa, F. Smarandache, and M. Loey, "A study of the neutrosophic set significance on deep transfer learning models: An experimental case on a limited COVID-19 chest X-ray dataset," *Symmetry*, to be published, doi: 10.1007/s12559-020-09802-9.

[26] N. E. M. Khalifa, M. H. N. Taha, A. E. Hassanien, and S. Elghamrawy, "Detection of coronavirus (COVID-19) associated pneumonia based on generative adversarial networks and a fine-tuned deep transfer learning model using chest X-ray dataset," 2020, *arXiv:2004.01184*.

[27] I. D. Apostolopoulos and T. A. Mpesiana, "COVID-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks," *Phys. Eng. Sci. Med.*, vol. 43, pp. 635–640, Apr. 2020.

[28] Y. Gao and Y. Cui, "Deep transfer learning for reducing health care disparities arising from biomedical data inequality," *Nat. Commun.*, vol. 11, p. 5131, Oct. 2020.

[29] A. Khampari, D. Gupta, V. H. C. de Albuquerque, A. K. Sangaiah, and R. H. Jhaveri, "Internet of Health Things-driven deep learning system for? Detection and classification of cervical cells using transfer learning," *J. Supercomput.*, vol. 76, pp. 8590–8608, Nov. 2020.

[30] Q. Song, Y.-J. Zheng, W.-G. Sheng, and J. Yang, "Tridirectional transfer learning for predicting gastric cancer morbidity," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 561–574, Feb. 2021.

[31] B. Lake, R. Salakhutdinov, J. Gross, and J. Tenenbaum, "Oneshot learning of simple visual concepts," in *Proc. 33rd Annu. Meeting Cogn. Sci. Soc.*, 2011, pp. 2568–2573.

[32] X. Chen, L. Yao, T. Zhou, J. Dong, and Y. Zhang, "Momentum contrastive learning for few-shot COVID-19 diagnosis from chest CT images," *Pattern Recognit.*, vol. 113, May 2021, Art. no. 107826.

[33] A. Voulodimos, E. Protopapadakis, I. Katsamenis, A. Doulamis, and N. Doulamis, "A few-shot U-net deep learning model for COVID-19 infected area segmentation in CT images," *Sensors*, vol. 21, no. 6, p. 2215, 2021.

[34] W. Wei, J. Wang, N. Cheng, Y. Chen, B. Zhou, and J. Xiao, "Epidemic Guard: A COVID-19 detection system for elderly people," in *Web and Big Data*, X. Wang, R. Zhang, Y.-K. Lee, L. Sun, and Y.-S. Moon, Eds. Cham, Switzerland: Springer, 2020, pp. 545–550.

[35] B. Lwowski and P. Najafirad, "COVID-19 surveillance through Twitter using self-supervised and few shot learning," in *Proc. EMNLP Workshop*, 2020, pp. 1–8.

[36] M. Shorfuzzaman and M. S. Hossain, "MetaCOVID: A siamese neural network framework with contrastive loss for n-shot diagnosis of COVID-19 patients," *Pattern Recognit.*, vol. 113, May 2021, Art. no. 107700.

[37] T. Naren, Y. Zhu, and M. D. Wang, "COVID-19 diagnosis using model agnostic meta-learning on limited chest X-Ray images," in *Proc. 12th ACM Conf. Bioinform. Comput. Biol. Health Inform.*, 2021, pp. 1–9.

[38] W. Zheng *et al.*, "Learning to learn by yourself: Unsupervised meta-learning with self-knowledge distillation for COVID-19 diagnosis from pneumonia cases," *Int. J. Intell. Syst.*, vol. 36, no. 8, pp. 4033–4064, 2021.

[39] R. Singh, V. Bharti, V. Purohit, A. Kumar, A. K. Singh, and S. K. Singh, "MetaMed: Few-shot medical image classification using gradient-based meta-learning," *Pattern Recognit.*, vol. 120, Dec. 2021, Art. no. 108111.

[40] G. Panagopoulos, G. Nikolentzos, and M. Vazirgiannis, "Transfer graph neural networks for pandemic forecasting," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, 2021, pp. 4838–4845.

[41] W. Kristjanpoller, K. Michell, and M. C. Minutolo, "A causal framework to determine the effectiveness of dynamic quarantine policy to mitigate COVID-19," *Appl. Soft Comput.*, vol. 104, Jun. 2021, Art. no. 107241.

[42] C. Li, S. Huang, Y. Liu, and Z. Zhang, "Distributed jointly sparse multitask learning over networks," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 151–164, Jan. 2018.

[43] A. Amyar, R. Modzelewski, H. Li, and S. Ruan, "Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation," *Comput. Biol. Med.*, vol. 126, Nov. 2020, Art. no. 104037.

[44] M. Goncharov *et al.*, "CT-based COVID-19 triage: Deep multitask learning improves joint identification and severity quantification," *Med. Image Analys.*, vol. 71, Jul. 2021, Art. no. 102054.

[45] M. Hayhoe, F. Barreras, and V. M. Preciado, "Multitask learning and nonlinear optimal control of the COVID-19 outbreak: A geometric programming approach," *Ann. Rev. Control*, vol. 52, pp. 495–507, May 2021, doi: 10.1016/j.arcontrol.2021.04.014.

[46] S. Panda and S. I. Levitan, "Detecting multilingual COVID-19 misinformation on social media via contextualized embeddings," in *Proc. 4th Workshop NLP Internet Freedom Censorship Disinform. Propaganda*, 2021, pp. 125–129.

[47] A. Rahimi and M. Gönen, "Efficient multitask multiple kernel learning with application to cancer research," *IEEE Trans. Cybern.*, early access, Mar. 11, 2021, doi: 10.1109/TCYB.2021.3052357.

[48] Z. Liang, H. Dong, C. Liu, W. Liang, and Z. Zhu, "Evolutionary multitasking for multiobjective optimization with subspace alignment and adaptive differential evolution," *IEEE Trans. Cybern.*, vol. 52, no. 4, pp. 2096–2109, APr. 2022, doi: 10.1109/TCYB.2020.2980888.

[49] Z. Chen, Y. Zhou, X. He, and J. Zhang, "Learning task relationships in evolutionary multitasking for multiobjective continuous optimization," *IEEE Trans. Cybern.*, early access, Nov. 18, 2020, doi: 10.1109/TCYB.2020.3029176.

[50] J. Lin, H.-L. Liu, K. C. Tan, and F. Gu, "An effective knowledge transfer approach for multiobjective multitasking optimization," *IEEE Trans. Cybern.*, vol. 51, no. 6, pp. 3238–3248, Jun. 2021.

[51] K. Chen, B. Xue, M. Zhang, and F. Zhou, "An evolutionary multitasking-based feature selection method for high-dimensional classification," *IEEE Trans. Cybern.*, early access, Dec. 31, 2020, doi: 10.1109/TCYB.2020.3042243.

[52] H. Zuo, G. Zhang, W. Pedrycz, V. Behbood, and J. Lu, "Fuzzy regression transfer learning in Takagi-Sugeno fuzzy models," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 6, pp. 1795–1807, Dec. 2017.

[53] Z. Deng, K. S. Choi, Y. Jiang, and S. Wang, "Generalized hidden-mapping ridge regression, knowledge-leveraged inductive transfer learning for neural networks, fuzzy systems and kernel methods," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2585–2599, Dec. 2014.

[54] G. Chen, Y. Li, and X. Liu, "Transfer learning under conditional shift based on fuzzy residual," *IEEE Trans. Cybern.*, vol. 52, no. 2, pp. 960–970, Feb. 2022, doi: 10.1109/TCYB.2020.2988277.

[55] G. Mesnil *et al.*, "Unsupervised and transfer learning challenge: A deep learning approach," in *Proc. ICML Workshop Unsupervised Transf. Learn.*, vol. 27, 2011, pp. 97–111.

[56] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio, "Contractive auto-encoders: Explicit invariance during feature extraction," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 833–840.

[57] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Advances in Neural Information Processing Systems*, vol. 19, J. P. B. Schölkopf and T. Hoffman, Eds. Cambridge, MA, USA: MIT Press, 2007, pp. 153–160.

[58] X. Zhang and Z. Xu, "Extension of TOPSIS to multiple criteria decision making with pythagorean fuzzy sets," *Int. J. Intell. Syst.*, vol. 29, no. 12, pp. 1061–1078, 2014.

[59] R. R. Yager, "Decision making with fuzzy probability assessments," *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 4, pp. 462–467, Aug. 1999.

[60] R. Yager, "Pythagorean fuzzy subsets," in *Proc. IFSA/NAFIPS*, 2013, pp. 57–61.

[61] Y.-J. Zheng, W.-G. Sheng, X.-M. Sun, and S.-Y. Chen, "Airline passenger profiling based on fuzzy deep machine learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 2911–2923, Dec. 2017.

[62] X. Gou, Z. Xu, and H. Liao, "Exponential operations of interval-valued intuitionistic fuzzy numbers," *Int. J. Mach. Learn. Cybern.*, vol. 7, no. 3, pp. 501–518, 2016.

[63] H. Garg, "New exponential operational laws and their aggregation operators for interval-valued Pythagorean fuzzy multicriteria decision-making," *Int. J. Intell. Syst.*, vol. 33, no. 3, pp. 653–683, 2018.

[64] Y.-M. Wang, J.-B. Yang, D.-L. Xu, and K.-S. Chin, "On the centroids of fuzzy numbers," *Fuzzy Sets Syst.*, vol. 157, no. 7, pp. 919–926, 2006.

[65] C. Chen, C.-Y. Zhang, L. Chen, and M. Gan, "Fuzzy restricted Boltzmann machine for the enhancement of deep learning," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 6, pp. 2163–2173, Dec. 2015.

[66] J. Martens, "Deep learning via hessian-free optimization," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 735–742.

[67] Y.-J. Zheng, "Water wave optimization: A new nature-inspired metaheuristic," *Comput. Oper. Res.*, vol. 55, no. 1, pp. 1–11, 2015.

[68] D. Simon, "Biogeography-based optimization," *IEEE Trans. Evol. Comput.*, vol. 12, no. 6, pp. 702–713, Dec. 2008.

[69] X.-H. Zhou, M.-X. Zhang, Z.-G. Xu, C.-Y. Cai, Y.-J. Huang, and Y.-J. Zheng, "Shallow and deep neural network training by water wave optimization," *Swarm Evol. Comput.*, vol. 50, pp. 1–13, Nov. 2019.

[70] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 110, pp. 3371–3408, 2010.

[71] Q. Hu, R. Zhang, and Y. Zhou, "Transfer learning for short-term wind speed prediction with deep neural networks," *Renew. Energy*, vol. 85, pp. 83–95, Jan. 2016.

[72] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Deep transfer learning with joint adaptation networks," in *Proc. ICML*, vol. 70, 2017, pp. 2208–2217.

[73] Y. Jiang, Z. Deng, F.-L. Chung, and S. Wang, "Multi-task TSK fuzzy system modeling using inter-task correlation information," *Inf. Sci.*, vol. 298, pp. 512–533, Mar. 2015.

[74] Y. Zheng, Y. Du, H. Ling, W. Sheng, and S. Chen, "Evolutionary collaborative human-UAV search for escaped criminals," *IEEE Trans. Evol. Comput.*, vol. 24, no. 2, pp. 217–231, Apr. 2020.

[75] K. Arulkumaran, A. Cully, and J. Togelius, "AlphaStar: An evolutionary computation perspective," in *Proc. Genet. Evol. Comput. Conf.*, New York, NY, USA, 2019, pp. 314–315.

Qin Song received the Ph.D. degree in environmental science and technology from the Zhejiang University of Technology, Hangzhou, China, in 2018.

She is currently a Research Assistant with Hangzhou Normal University, Hangzhou. Her current research interests include machine learning and health informatics.

Yu-Jun Zheng (Senior Member, IEEE) received the Ph.D. degree from the Institute of Software, Chinese Academy of Sciences, Beijing, China, in 2010.

He is currently a Professor with Hangzhou Normal University, Hangzhou, China. He has authored over 100 papers in IEEE TRANSACTIONS and other famous journals. His research interests include machine learning and evolutionary algorithms.

Prof. Zheng received the Runner-Up of IFORS Prize for OR in Development in 2014. In 2018, he was selected as a Finalist for the Daniel H. Wagner Prize for Excellence in Operations Research Practice.

Jun Yang received the B.S. degree in biochemistry and the M.S. degree in cell biology from Shandong University, Jinan, China, in 1991 and 1994, respectively, and the Ph.D. degree in biochemistry and molecular genetics from Georgia State University, Atlanta, GA, USA, in 2000.

He is currently a Professor with Zhejiang University, Hangzhou, China, and also with Hangzhou Normal University, Hangzhou. His main research interests are public health management and environmental health.

Yu-Jiao Huang received the B.S. degree in information and computer science, the M.S. degree in computational mathematics, and the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2008, 2010, and 2014, respectively.

She is currently a Lecturer with the Zhijiang College, Zhejiang University of Technology, Hangzhou, China. Her research interests are in areas of artificial neural networks, stability theory, and dynamical systems.

Wei-Guo Sheng (Member, IEEE) received the M.Sc. degree in information technology from the University of Nottingham, Nottingham, U.K., in 2002, and the Ph.D. degree in computer science from Brunel University London, Uxbridge, U.K., in 2005.

He was a Researcher with the University of Kent, Canterbury, U.K., and Royal Holloway, University of London, London, U.K. He is currently a Professor with Hangzhou Normal University, Hangzhou, China. His research interests include evolutionary computation, data mining, and machine learning.

Sheng-Yong Chen (Senior Member, IEEE) received the Ph.D. degree from the City University of Hong Kong, Hong Kong, in 2003.

He is a Professor and a Ph.D. Advisor with the School of Computer Science and Engineering, Tianjin University of Technology, Tianjin, China. He has authored over 100 scientific papers in reputable international journals and conferences. His research interests include computer vision and machine learning.

Dr. Chen achieved the National Outstanding Youth Fund of China in 2013. He is an IET Fellow.