A Study of the Effectiveness of Transfer Learning in Individualized Asthma Risk Prediction

Wan D. Bae
Seattle University
USA
baew@seattleu.edu

Shayma Alkobaisi*
College of IT, UAEU
United Arab Emirates
shayma.alkobaisi@uaeu.ac.ae

Matthew Horak
Lockheed Martin Space Systems
U.S.A.
matthew.horak@lmco.com

Sungroul Kim
Soonchunhyang University
South Korea
sungroul.kim@gmail.com

Choon-Sik Park
Soonchunhyang University Bucheon Hospital, South Korea
mdcspark@hanmail.net

Mark Chesney
Seattle University
USA
mchesney@seattleu.edu

ABSTRACT

Deep Learning classifiers require a vast amount of data to train models that generalize well and perform effectively on unseen data. However, small sizes of training data, especially in the medical domain, make this a challenging task. Transfer Learning (TL) can help overcome a scarcity of data by focusing on fine tuning a pre-trained model with a small amount of specialized training data. In the last few years, several studies have been performed on TL with medical images, and they point towards significant gains available with this method. However, to date no such studies have been performed in the area of individualized asthma prediction with limited training data for each patient. In this paper, we conduct a systematic study of transfer learning in this domain in the context of neural networks. Our TL approach trains the source model with population data of 25 asthma patients and then retrains the target model with a target patient's data. Our results show that transfer learning yields promising results in improving model performance on an individual basis. Further research directions that are worth investigating based on our results are pointed out as future work directions.

CCS CONCEPTS

• **Computing methodologies → Machine learning**;

KEYWORDS

Transfer learning, neural networks, personalized asthma risk prediction, exposome analytics, indoor air quality

ACM Reference Format:
Wan D. Bae, Shayma Alkobaisi, Matthew Horak, Sungroul Kim, Choon-Sik Park, and Mark Chesney . 2021. A Study of the Effectiveness of Transfer Learning in Individualized Asthma Risk Prediction. In *The 36th ACM/SIGAPP Symposium on Applied Computing (SAC '21), March 22–26, 2021, Virtual Event, Republic of Korea.* ACM, New York, NY, USA, Article 4, 4 pages. https://doi.org/10.1145/3412841.3442105

*Corresponding author.

SAC '21, March 22–26, 2021, Virtual Event, Republic of Korea

ACM ISBN 978-1-4503-8104-8/21/03.
https://doi.org/10.1145/3412841.3442105

1 INTRODUCTION

Asthma is one of the top chronic respiratory diseases affecting about 339 million people, and its prevalence and socioeconomic burden are rising [9, 10]. Possibly avoidable exacerbations account for 63% of total asthma cost due to hospital admissions and emergency room visits, and while asthma is not curable its symptoms can be controlled through effective management. With improvements in computing and sensor technologies, a new line of healthcare research and industry ventures has emerged that offers predictive health monitoring solutions [3–5, 11, 12, 14, 15]. Such solutions require automated and early recognition of symptoms or likelihood of exacerbated symptoms.

A recent review of machine learning approaches to prediction of chronic diseases focused on fairly small model types but still found that one of the main challenges to improving forecast quality is the limited availability of large high quality labeled data sets [1]. One way to overcome this challenge is with transfer learning which enables researchers to take advantage of high quality data sets in related fields to reduce the need for as much domain-specific data [13]. This strategy has shown great promise in the medical field in the context of image analysis of MRI or CT scan data and images from lab tests [6, 8]. However, to our knowledge no similar study has been performed in the context of asthma risk forecasting with non-image data. This paper reports results of a preliminary study of the effectiveness of transfer learning in this domain.

2 STUDY VARIABLES AND RISK METRICS

2.1 Individual-level Asthma Risk

One of the primary health indicators used in the management of asthma is the peak expository flow rate (PEFR) measurement. In this paper, we base our risk classifications on a simplified version of the individual-based asthma risk PEFR zoning method proposed in [2]. For the purposes of prediction of high risk days in this paper we use only two zones, a "safe zone" which we nominally take to be PEFR values in the upper 80% of the patient's historical PEFR values and a "red zone" nominally taken to be PEFR values in the lower 20% of the PEFR distribution. The classification task of the neural networks studied will be to predict when a patient next-day

PEFR value is expected to be in the red zone, which is assumed to be a high-risk condition.

2.2 Study Population and Variables

Study participants were recruited from the adult asthma patients who had joined the environmental health smart study with connectivity and remote sensing technologies (ESCORT) [16]. A total of 25 patients (15 women and 10 men) aged 32 to 78 years were consulted and monitored by doctors and medical practitioners at Soonchunhyang University Bucheon Hospital, South Korea. We took a comprehensive approach to tracking spatiotemporal exposure patterns and twice daily PEFR values for each participant between November 1, 2017 and May 31, 2019. While individual data set sizes varied, their sizes ranged from 100 to 200 days with a mean of 154 days.

The study's explanatory variables included both environmental as well as personal variables. Environmental variables included: levels of $PM_{2.5}$, CO_2 and humidity as well as temperature, all measured at 2 minute intervals by sensors within the home. Personal variables included: frequency of frying food (7-level scale), distance from home to nearest major road (5-level scale) and income (9-level scale).

3 TRANSFER LEARNING

Transfer learning strives to improve a specialized model's learning of a new task through the transfer of knowledge from a related task that has already been learned by a so-called *source model* [13]. The resulting specialized model is often referred to as the *target model*. For the purpose of our study, we analyze the task of forecasting whether or not a patient's next-day PEFR level will be above or below their critical PEFR value defined in Section 2. We treat this as a classification problem in which a sample's input data is the patient's current day environmental data and health data and the sample's class value is the patient's next day PEFR zone.

3.1 General Transfer Learning Paradigm

The basic process for applying transfer learning to a classification neural network involves first training a source model on a task for which a large dataset is available. This model is then optionally slightly modified, for example by replacing the last few layers with layers of a different size or type. The new model is then either fine tuned in full through further training on the target task or part of the model's parameters are frozen and the rest are fine tuned on the target task as in [8].

3.2 Present Transfer Learning Architecture

For image classification, datasets are often synthetically augmented by suitably transforming existing images. However, realistic transformations for exposome data have not been rigorously established, so we must retrain our target model with a very small dataset. Therefore, we focus only on the effect of freezing the weights of some initial layers of the source model and fine tuning only the later layers on the target data. All unfrozen layers for target models are initialized with the values from the source model.

For the present study, the source task and the target task are the same: classify a patient's next-day PEFR value as above or below their critical PEFR value. The source model is trained on data from the entire population minus the target patient data (24 patients' data) and the target model is trained on only the target patient's data, as illustrated in Figure 1.

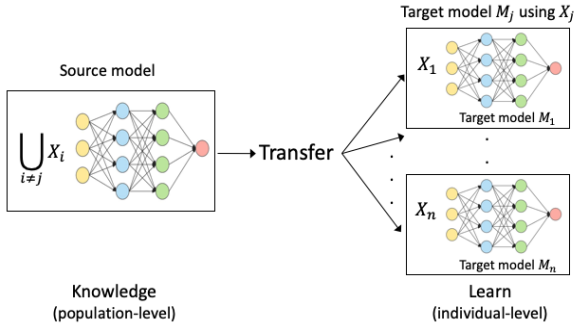

Figure 1: Transfer learning for personalized risk prediction

All of the networks studied are simple feed-forward neural networks with one, two or three hidden layers and an output layer returning a single value between 0 and 1 interpreted as the models confidence as to whether or not the patient's next-day PEFR value will fall below their critical PEFR. All model layers are fully connected with Rectified Linear Unit (ReLU) activation except for the output layer, which has sigmoid activation. All models were trained using Adaptive learning rate optimization (Adam) with binary cross entropy loss. The input layer of each model consists of 8 parameters which are yesterday's PEFR value, 4 indoor quality variables, and 3 personal environmental variables as described in Section 2.2.

Table 1: Neural Network and Transfer Learning design

model		architecture	unfrozen ratio	# unfrozen weights
NN		32	100.0%	321
TL	TL1-nf	32	100.0%	321
	TL1-1f	32*	10.3%	33
	TL2-1f	32*-10	50%	341
		32*-20	70.3%	681
		32*-32	79.1%	1089
	TL3-1f	32*-10-10	61.0%	451
		32*-20-20	79.3%	1101
	TL3-2f	32*-10*-10	16.4 %	121
		32*-10*-20	31.7 %	441

* represents a frozen layer.

The model architectures evaluated together with what layers were frozen for the target model and other model information are given in Table 1. In that table, the *architecture* column gives both the model architecture as well as the layers that were frozen for the target model. For example, "32 (frozen) -10 (frozen) - 10" indicates a source model with three hidden layers of size 32, 10 and 10 and a target model with the same layers that inherits the parameters of layers 1 and 2 from the trained source model but for which hidden layer 3 and the output layer are retrained on the target patient's

data. Table 1 also gives the total number of trainable (unfrozen) parameters in the target model as well as the fraction of the source model's parameters made up by parameters that are unfrozen in the target model.

4 EXPERIMENTAL RESULTS

4.1 Performance Metrics

Models in this paper were evaluated based on their ability to distinguish "red zone" examples from "safe zone" ones, as defined in Section 2.1. With our focus on high risk prediction, it is important to emphasize performance improvement on the target "red zone". We used four standard evaluation metrics: (1) balanced accuracy, (2) sensitivity, (3) precision average, and (4) F_1-score average, which are known to be good measures of a models ability to correctly predict the target class.

We trained the transfer learning models defined in Table 1 with 1000 epochs for the source model and 100 epochs for the target model. These values were taken from the optimal training durations found in the authors' previous work with similar networks. The source model was trained with 24 patients' datasets (full population excluding the target patient's dataset). The target model was fine tuned using the target patient's dataset. The models were developed in Python 3.7 and Keras framework. The hyperparameters in the results were selected through extended training and validation processes.

4.2 Evaluations of Transfer Learning Architectures

The architectures of the networks evaluated were selected to provide a range of percentages of unfrozen weights in the target model while keeping the total number of trainable model parameters within the constraints dictated by the size of the dataset. For each architecture, one target model was possible for each of the 25 patients and each one was evaluated through 3-fold cross validation giving 75 trained target models for each architecture. The performance statistics reported for these models are the averages of each metric over the 75 target models.

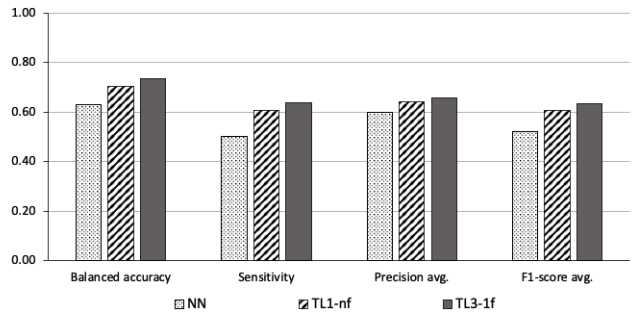

Figure 2: Performance results of NN and TL

Table 2 shows that all of the performance measures, optimal performance is achieved when approximately 80% of a model's parameters are fine tuned. Figure 2 shows the overall gains achieved in the two best transfer learning architectures over a 1-layer neural network model fully trained on target patients' data. Although these increases are more modest than achieved by transfer learning in other contexts like image classification, they are achieved without the benefit of data augmentation other than re-sampling to balance the class sizes. Optimal gain was achieved with relatively little of the source network frozen, which may indicate that while individuals differ from each other significantly enough for there to be benefit in training a new model for each individual, each individual is consistent enough for a large portion of a network to be reliably trained on relatively little individual data.

4.3 Population Size versus Relevancy Trade-off

When selecting the source model for transfer learning, one must often weigh the trade-off between using a high quality source model whose training data or target task is less related to the target model's task and using a lower quality source model whose task is closer to that of the target model, as studied in [8]. We performed a similar study by grouping patients into three sub-populations with similar medical and lifestyle characteristics such as income, cooking style, and home location. Each group had only 7 to 9 patients, with low patient-to-patient variation. For each of these sub-populations, we re-executed the evaluation regime performed on the full population and compared the average results of these 3 groupings to a baseline average obtained on 100 repetitions with randomly selected sub-populations of size 10. Table 3 presents the results of this study.

Opposite of the results in [8], almost all of the architectures show larger improvement over the base model (TL1-nf) than were obtained by the corresponding models in Table 2. Perhaps the source models themselves in the present study are lower quality due to the relatively small size of even the full population dataset, or the source task and target task may be more similar in this case.

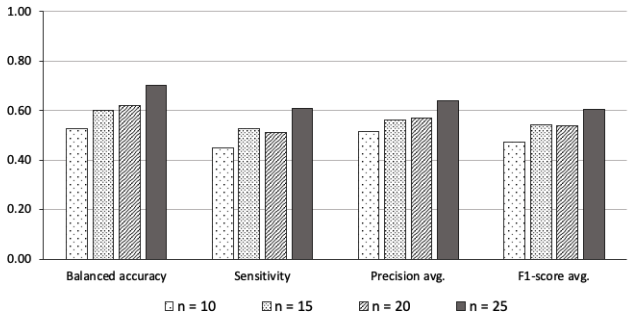

Figure 3: Performance results of transfer learning with varying population for base model

Our final study was to analyze the effect of the size of the source model's dataset on the quality of the target model when the relevancy level is unchanged. The average results achieved with each of these sizes was compared against the results achieved with the full population (size 25). Figure 3 presents the results of this study. As expected, in all performance metrics these results show increasing model quality as the population size increases.

Table 2: Performance comparisons of TL models (2 quantile zones: <= 20% and > 20%)

model	unfrozen ratio	balanced acc.	sensitivity	precision avg.	F_1 score avg.
TL1-nf	**100%**	**0.703**	**0.608**	**0.642**	**0.607**
TL1-1f	10%	0.705	0.660	0.644	0.567
TL2-1f	50%	0.701	0.660	0.627	0.574
	70%	0.699	0.636	0.638	0.594
	80%	**0.734**	**0.665**	**0.664**	**0.615**
TL3-1f	61%	0.687	0.586	0.644	0.595
	80%	**0.737**	**0.665**	**0.658**	**0.630**
TL3-2f	16%	0.671	0.665	0.608	0.536
	30%	0.711	0.657	0.643	0.600

Table 3: Performance comparisons of TL models by grouping patients (2 quantile zones: <= 20% and > 20%)

model	unfrozen ratio	group	balanced acc.	sensitivity	precision avg.	F_1 score avg.
TL1-nf	**100%**	Baseline average (random size 10 groups)	**0.523**	**0.450**	**0.516**	**0.471**
TL1-nf	**100%**	3 grouping average	**0.563**	**0.487**	**0.532**	**0.498**
TL1-1f	10%	3 grouping average	0.531	0.509	0.509	0.445
TL2-1f	50%	3 grouping average	0.549	0.511	0.520	0.468
	70%	3 grouping average	0.563	0.504	0.531	0.492
	80%	3 grouping average	**0.592**	**0.527**	**0.530**	**0.509**
TL3-1f	61%	3 grouping average	0.550	0.446	0.529	0.488
	80%	3 grouping average	**0.586**	**0.518**	**0.551**	**0.533**

5 CONCLUSIONS

Our results are consistent with the general trends in machine learning. First, transfer learning shows promise in partially alleviating the challenges of small datasets. Second, changes in amount of the source model used yielded observable trends in performance, useful for model optimization. These results suggest that future work investigating more complex transfer learning architectures may be worthwhile. In particular, output of a middle layer of the source model could be used to fuel a smaller target model such as a support vector machine or decision tree. One could also attempt some of the newer transfer learning methods available for other possibly less data-hungry models as described in [7].

ACKNOWLEDGMENTS

The authors would like to thank the medical staff in the Division of Allergy and Respiratory Medicine at the Soonchunhyang University Bucheon Hospital for the provision of the data. This material is based upon studies supported in part by the Korean Environmental Industry & Technology Institute, Ministry of Environment, South Korea, under Grant No. 2016001360002, and in part by the Seattle University under Grant No. CSE-01-2019.

REFERENCES

[1] 2020. Applications of machine learning predictive models in the chronic disease diagnosis. *Journal of Personalized Medicine* 10, 3 (2020).
[2] Shayma Alkobaisi, Wan D Bae, Matthew Horak, Sada Narayanappa, Jongwon Lee, Eman AbuKhousa, Choon-Sik Park, and Da Jung Bae. 2019. Predictive and exposome analytics: A case study of asthma exacerbation management. *Journal of Ambient Intelligence and Smart Environments* Preprint (2019), 1–26.
[3] Shubhi Asthana, Ray Strong, and Aly Megahed. 2016. HealthAdvisor: Recommendation System for Wearable Technologies enabling Proactive Health Monitoring. *arXiv preprint arXiv:1612.00800* (2016).
[4] Mario Hermann, Tobias Pentek, and Boris Otto. 2016. Design Principles for Industrie 4.0 Scenarios. *HICSS* (2016).
[5] Laleh Jalali, Minh-Son Dao, Ramesh Jain, and Koiji Zetts. 2015. Complex asthma risk factor recognition from heterogeneous data streams. In *IEEE International Conference on Multimedia and Expo Workshops*. 1–6.
[6] Xiangyun Liao, Yinling Qian, Yilong Chen, Xueying Xiong, Qiong Wang, and Pheng-Ann Heng. 2020. MMTLNet: Multi-Modality Transfer Learning Network with adversarial training for 3D whole heart segmentation. *Computerized Medical Imaging and Graphics* 85 (2020), 101785. https://doi.org/10.1016/j.compmedimag.2020.101785
[7] Mingsheng Long, Jianmin Wang, Guiguang Ding, Sinno Jialin Pan, and S Yu Philip. 2013. Adaptation regularization: A general framework for transfer learning. *IEEE Transactions on Knowledge and Data Engineering* 26, 5 (2013), 1076–1089.
[8] Afonso Menegola, Michel Fornaciali, Ramon Pires, Flávia Vasques Bittencourt, Sandra Avila, and Eduardo Valle. 2017. Knowledge transfer for melanoma screening with deep learning. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 297–300.
[9] Global Asthma Network. 2019. The Global Asthma Report 2018. http://www.globalasthmareport.org/Global_Asthma_Report_2018.pdf, Accessed: 09-01-2020.
[10] Carlos Nunes, Ana Margarida Pereira, and Mário Morais-Almeida. 2017. Asthma costs and social impact. *Asthma Research and Practice* 3, 1 (06 Jan 2017), 1. https://doi.org/10.1186/s40733-016-0029-3
[11] Yanfeng Shen and Victor Giurgiutiu. 2014. Predictive modeling of nonlinear wave propagation for structural health monitoring with piezoelectric wafer active sensors. *Journal of Intelligent Material Systems and Structures* 25, 4 (2014), 506–520.
[12] C Thuemmler and C Bai. 2017. Health 4.0: How virtualization and big data are revolutionizing healthcare.
[13] Lisa Torrey and Jude Shavlik. 2010. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI Global, 242–264.
[14] Paolo Vineis, Marc Chadeau-Hyam, Hans Gmuender, John Gulliver, Zdenko Herceg, Jos Kleinjans, Manolis Kogevinas, Soterios Kyrtopoulos, M Nieuwenhuijsen, David H Phillips, et al. 2017. The exposome in practice: design of the EXPOsOMICS project. *International journal of hygiene and environmental health* 220, 2 (2017), 142–151.
[15] Christopher Paul Wild. 2012. The exposome: from concept to utility. *International journal of epidemiology* 41, 1 (2012), 24–32.
[16] Jiyoung Woo, Guillaume Rudasingwa, and Sungroul Kim. 2020. Assessment of Daily Personal PM2. 5 Exposure Level According to Four Major Activities among Children. *Applied Sciences* 10, 1 (2020), 159.