Adversarial Training for Privacy-Preserving Deep Learning Model Distribution

Mohammed Alawad*, Shang Gao*, Xiao-Cheng Wu†, Eric B. Durbin‡§¶, Linda Coyle‖,
Lynne Penberthy**, Georgia Tourassi*
*Computational Sciences and Engineering Division, Health Data Sciences Institute,
Oak Ridge National Laboratory, Oak Ridge, TN, USA.
Email: {alawadmm,tourassig}@ornl.gov
†Louisiana Tumor Registry, Louisiana State University Health Sciences Center, New Orleans, LA, USA.
‡Kentucky Cancer Registry, University of Kentucky, Lexington, KY, USA.
§Division of Biomedical Informatics, College of Medicine, University of Kentucky, Lexington, KY, USA.
¶Cancer Research Informatics Shared Resource Facility, Markey Cancer Center, University of Kentucky, Lexington, KY, USA.
‖Information Management Services Inc, Calverton, MD, USA.
**Surveillance Research Program, Division of Cancer Control and Population Sciences,
National Cancer Institute, Bethesda, MD, USA.

***Abstract*—Collaboration among cancer registries is essential to develop accurate, robust, and generalizable deep learning models for automated information extraction from cancer pathology reports. Sharing data presents a serious privacy issue, especially in biomedical research and healthcare delivery domains. Distributing pretrained deep learning (DL) models has been proposed to avoid critical data sharing. However, there is growing recognition that collaboration among clinical institutes through DL model distribution exposes new security and privacy vulnerabilities. These vulnerabilities increase in natural language processing (NLP) applications, in which the dataset vocabulary with word vector representations needs to be associated with the other model parameters. In this paper, we propose a novel privacy-preserving DL model distribution across cancer registries for information extraction from cancer pathology reports with privacy and confidentiality considerations. The proposed approach exploits the adversarial training framework to distinguish private features from shared features among different datasets. It only shares registry-invariant model parameters, without sharing raw data nor registry-specific model parameters among cancer registries. Thus, it protects both the data and the trained model simultaneously. We compare our proposed approach to single-registry models, and a model trained on centrally hosted data from different cancer registries. The results show that the proposed approach significantly outperforms the single-registry models and achieves statistically indistinguishable micro and macro F1-score as compared to the centralized model.**

***Index Terms*—Privacy-preserving, convolutional neural network, natural language processing, information extraction.**

I. Introduction

Cancer pathology reports represent a critical component for cancer surveillance. They contain important information, such as cancer characteristics and medical history, that is necessary to support cancer incidence reporting. Manually extracting information from unstructured text in pathology reports is time-consuming and very costly. Moreover, training experienced staff will also take a lot of time. Therefore, automating the process is of interest to cancer registries. Traditional natural language processing (NLP) methods rely heavily on time-consuming and labor-intensive feature engineering. They cannot be easily generalized across data sources and information extraction tasks. Also, the variability and diversity in describing cancer characteristics lead to enormous syntactic and semantic variabilities which hinder the ability of conventional NLP techniques to effectively tackle the challenge. In recent years, different deep learning approaches have been developed to automatically extract information from cancer pathology reports and have shown unprecedented performance as compared to rule-based models and traditional machine learning techniques [1], [2]. To achieve this performance, deep learning models require a vastly larger labeled corpus for training, where data availability is a key consideration for effective deep learning solutions for real-world problems. Since the availability of labeled data samples at each cancer registry is limited, collaboration among them is essential to develop robust and generalized deep learning models [3]. The ideal scenario of collaboration is to share the raw data with a centralized host and train a deep learning model on all available data.

Such collaboration is hindered by security and privacy concerns that result from sharing pathology reports between cancer registries due to regulations [4]. Thus, privacy-preserving approaches have to be developed to offer secure

avenues for collaboration among data holders. One approach of collaboration without privacy violation is through text de-identification by detecting and scrubbing protected health information (PHI) from cancer pathology reports. However, manual de-identification approach is costly and time consuming, and automatic clinical text de-identification [5], [6] is highly challenging [7], [8]. Therefore, existing solutions typically cannot guarantee de-identification up to regulatory standards. Another approach of secure collaboration among cancer registries can be done through deep learning model distribution. This approach protects confidential features by distributing a trained model without sharing raw data. Model distribution methods, such as federated learning [9] and large batch synchronous stochastic gradient descent SGD [10], require sharing the model architecture, which can be attacked by recovering the raw data using adversarial networks [11]. Other techniques like SplitNN [4] protect the model parameters; however, they require a relatively larger overall communication bandwidth [12]. The challenge of model distribution increases when dealing with deep learning NLP models even if the model architecture is protected. Thus, simply distributing models across different institutes does not satisfy the privacy-preserving condition. Recently, data encryption techniques, such as differential privacy [13] and homomorphic encryption [14], have been used to protect deep learning model distribution. However, the additional computational cost required by these methods limits their application. Moreover, encrypted models can be attacked by untrusted platforms and unauthorized users.

In this paper, we present a novel privacy-preserving model distribution approach for cancer registries by exploiting the adversarial training framework. In the proposed approach, each registry dataset is assigned a private convolutional neural network (CNN) and a shared CNN. Our privacy-preserving model distribution protects the data such that no raw data needs to be shared across cancer registries. It also protects the dataset-dependent parameters – the private CNN – that captures the private features by only distributing the dataset-invariant parameters – the shared CNN – that capture the shared features among datasets. Our approach also attempts to maximize the benefits of collaboration across data holders by sharing useful features that are related to all sources. The proposed approach is used to develop a deep learning model for information extraction from cancer pathology reports collected from two different National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) cancer registries. We compare the distributed learning model performance with two training approaches: (1) the single-registry models, in which the model is trained and tested on one registry dataset and neither data nor model parameters shared across registries, and (2) the centralized model, in which a global NLP model trained on centrally hosted pathology reports collected from both registries. The centralized model works when there is no violation in sharing data across cancer registries.

II. Materials and Methods

A. Datasets and Pre-processing

The datasets used in this paper are 374,899 and 172,128 pathology reports obtained from two independent SEER program sources: the Louisiana Tumor Registry (LTR) and Kentucky Cancer Registry (KCR) respectively. The LTR corpus spans the period 2004-2018 while the KCR corpus spans the period 2009-2018. This research was conducted in accordance to the institutional review board protocol DOE000152 and under a data use agreement between UT-Battelle, LLC. and both registries. Each pathology report is identified by a combination of patient ID and tumor ID, which is called case ID. Documents generated within 10 days between the date of diagnosis and either path specimen collection date or the surgery date were identified as relevant to the specific case ID. The 10-day window was based on an analysis of the pathology report submissions with the vast majority of reports and addenda included within that time frame. The remaining pathology reports which were outside the 10-day window were excluded from the study. Ground truth labels associated with each unique case are obtained from the registry record associated with the pathology report. In this paper, we consider the International Classification of Diseases for Oncology, Third Edition (ICD-O-3), topography[1] (i.e., site/subsite) as the data element of interest as it is a fundamental information extraction task for cancer reporting. The total number of cancer subsite labels observed in LTR and KCR datasets is 313. These labels represent tumor topographies across 70 organs where cancer may appear. To simulate real world production environment, we used the pathology report date to split each registry dataset into train, validation and test sets. Specifically, reports collected in 2016 and later are used for testing, while the rest of the reports are used for train and validation with 80:20 ratio. Since multiple cancer pathology reports might have the same case ID, all reports associated with the same case ID are grouped together into only the train, validation or test set to avoid any positive bias in the reported results. The train, validation and test set sizes are summarized in Table I. After excluding metadata in cancer pathology reports, text is cleaned by removing any consecutive punctuation and lowercasing all alphabetical characters. To reduce the vocabulary space, all words with document frequency less than five are replaced with an "$unknown\_word$" token, all decimals are converted to a "$decimal$" word token, and all integers larger than 100 are converted to a "$large\_integer$" word token. Each cancer pathology report is clipped or padded to 1500 words, and each word in a report is represented by a length 300 word vector that is initialized randomly and learned through training. These hyperparameters were empirically set based on prior studies [2], [15].

B. Conventional Deep Learning Model

To process our cancer pathology reports, we use a CNN similar to the one described in [16]. This architecture is

[1] http://codes.iarc.fr/topography

TABLE I
TRAIN, VALIDATION, AND TEST SPLIT OF LTR, KCR, COMBINED DATASETS

Dataset	Train	Validation	Test
LTR	169,011	42,380	48,852
KCR	89,404	22,686	37,731
Combined	258,415	65,066	86,583

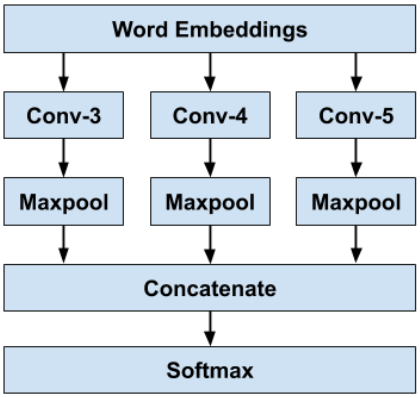

Fig. 1. Architecture diagram of the CNN model.

widely used across various clinical NLP tasks [17]–[19], and previous work has shown that it achieves strong performance in classifying cancer pathology reports while maintaining low computational cost and fast training time [15], [20].

The architecture of our CNN is shown in Figure 1. We use three parallel 1-D convolution layers which respectively examine three, four, and five consecutive words at a time – these layers act as feature extractors which identify important combinations of words for a given task. We concatenate the outputs of these convolutions and then use a temporal maxpool to filter the most salient word combinations generated by each convolution filter. The final selected features represent the most important word n-grams in the document for the task at hand; this is fed into a softmax layer for classification.

C. Privacy-preserving Deep Learning Model

Given a scenario in which there are multiple cancer registries that are unable to share data with each other, the simplest approach is for each registry to independently train its own CNN, as illustrated in Figure 2 (independent training). In this approach, each registry trains only on its own data and it is unable to benefit from data belonging to other registries.

In our privacy-preserving approach, we propose a privacy-preserving CNN which is shown in Figure 2 by exploiting the adversarial multitask learning architecture [21]. Similar to training an independent CNN for each registry, this architecture assigns separate convolution and maxpool layers for each registry – these registry-specific layers are designed to learn features useful for data from that specific registry. However, our privacy-preserving CNN also utilizes shared convolution

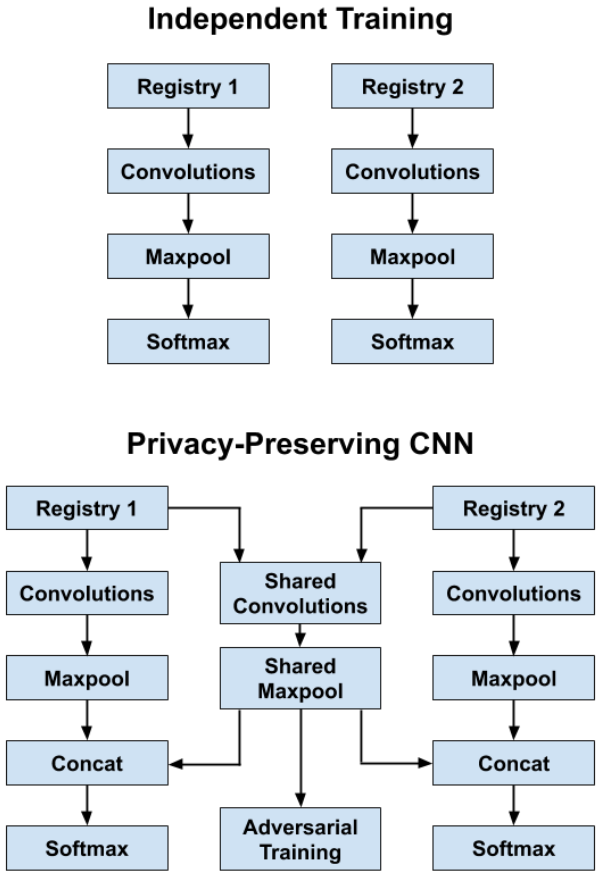

Fig. 2. Architecture diagrams of the independent training and the privacy-preserving CNN model.

and maxpool layers that are used for all registries – these shared layers are designed to learn registry-agnostic features that are useful for data from all registries. The outputs of these shared layers are concatenated with the output for the registry-specific layers before the final softmax classification.

In contrast to the independent training approach, our privacy-preserving CNN is able to train on data from all registries without requiring them to directly share their data. Also, the registry-specific model architecture and parameters are protected since they are not shared among registries. This can be achieved during training by keeping registry-specific data and layer parameters on secure servers, while the shared-layer parameters are stored on a central shared server. Once trained, the shared-layer parameters do not include any privacy information and can be shared with other registries.

D. Adversarial Training and Orthogonality Constraints

Our privacy-preserving CNN utilizes registry-specific layers for each registry and shared layers for all registries; our goal is to ensure that the registry-specific layers learn features private to that specific registry dataset, while the shared layers learn registry-agnostic features from all registry datasets. To achieve this, we utilize two additional losses in addition to the cross entropy loss associated with our classification task – (1) ad-

versarial training loss and (2) orthogonality loss. These losses are presented in [21] to separate the interference between task-specific and task-invariant feature spaces in multitask learning.

Adversarial training is used on the output from the shared layers to encourage the shared layers to learn registry-agnostic features. Given an input document, we train a separate softmax discriminator to attempt to identify the source registry using only the output generated from the shared layers. Simultaneously, we use adversarial loss to incentivize our privacy-preserving CNN to minimize the performance of the discriminator, thereby encouraging the shared layers to generate only features that do not contain any registry-specific information.

$$L_{adv} = \min_{\theta_s} \left(\lambda \max_{\theta_D} \sum_{r=1}^{R} \sum_{i=1}^{N_r} d_i^r \log[D(E(x^r))] \right) \quad (1)$$

Our adversarial loss is described in Equation 1, where d_i^r is the ground truth label indicating if an input document x^r belongs to registry r, $D(E(x^r))$ is the softmax probability assigned by the discriminator that the document belongs to registry r, and θ_s and θ_D are the parameters of the privacy-preserving CNN and discriminator.

Secondly, we use orthogonality constraints to further encourage registry-agnostic features to be learned in the shared layers rather than in the registry-specific layers. We achieve this by using an additional orthogonal loss term that penalizes the network if the outputs of the registry-specific convolution layers are similar to the outputs of the shared convolution layer.

$$L_{orth} = \sum_{r=1}^{R} \left\| \mathbf{S}^{r\top} \mathbf{H}^r \right\|_F^2 \quad (2)$$

Our orthogonal loss is described in Equation 2, where $\| \cdot \|_F^2$ is the squared Frobenius norm and $\mathbf{S}^r$ and $\mathbf{H}^r$ are the respectively the outputs generated by the shared convolutions and registry-specific convolutions for a given document.

$$L_{total} = L_{task} + \lambda L_{adv} + \gamma L_{orth} \quad (3)$$

Our combined loss function is described in Equation 3, where L_{task} is the cross entropy loss for the given classification task, L_{adv} and L_{orth} are the adversarial and orthogonal loss values, and λ and γ are tunable hyperparameters. We optimize our privacy-preserving CNN using a standard gradient descent optimizer, such as Adadelta, to minimize this total combined loss.

III. Experiments

A. Experimental Setup

In this study, we develop a privacy-preserving distributed CNN model as shown in Figure 2 for extracting cancer subsite from cancer pathology reports obtained from two SEER cancer registries. In this approach registry-specific parameters are updated based on their local data. Only the registry-invariant parameters are shared. The shared parameters are either updated by the LTR or KCR datasets. The parameter updates by one of the registry datasets are aggregated to generate a server model to be shared with the other registry dataset. Also, this approach protects the patient data, since raw data is not shared among cancer registries. However, we need to share the vocabulary list.

We benchmarked the proposed approach to two baseline models. The first model is single-registry model. In which a CNN model is trained and tested on each registry separately without sharing any information across them. This approach offers the highest data privacy and protection since nothing is shared across cancer registries. However, the limited dataset size available in each cancer registry may compromise overall model accuracy. The second model is the centralized model in which pathology reports are collected from LTR and KCR cancer registries and hosted in a centralized location. Then, a global CNN model is trained on the whole corpus and shared with cancer registries for testing. This approach does not offer any privacy with respect to raw data and model parameters sharing. It is expected though that this approach also offers the best classification accuracy as all the data is aggregated to train a global model.

B. Performance Evaluation

We evaluate the performance of all models using the standard NLP metrics of micro and macro F1-score. Micro F1-score is not sufficient to evaluate model performance when the dataset has extreme class imbalance because it assigns weight to each class proportional to the class prevalence in the dataset. Thus we also report macro F1-score, which averages by class without weighing by class prevalence, to evaluate the model performance on the less prevalent classes. For both the micro and macro F1-score, we calculate 95% confidence intervals by bootstrapping [22] from the test set to estimate the variability of each model performance metric. The confidence intervals are used to determine the statistical significance of the difference in performance between the various approaches.

IV. Results

The performance evaluation of the proposed privacy-preserving DL distribution and the baseline training approaches – single-registry model and centralized model – are summarized in Tables II and III. For both datasets, the single-registry model has the lowest performance as compared to other approaches. Specifically, the micro and macro F1-score of LTR and KCR datasets are (0.631,0.319) and (0.624,0.294), respectively. The inferior performance of the single registry model highlights the importance of collaboration among cancer registries by sharing raw data or trained model parameters. The results show that the centralized model, which is concurrently trained on LTR and KCR data, significantly outperforms the performance of the single-registry models for both the datasets. Specifically, it achieves micro and macro F1-score of (0.647,0.348) and (0.646,0.343) on LTR and KCR datasets, respectively. However, this approach does not provide any privacy protection to the raw data or the model architecture and parameters. The performance improvement is particularly notable for the macro F1-score, which suggests

TABLE II
MICRO AND MACRO F1-SCORE (WITH 95% CONFIDENCE INTERVALS) OF DIFFERENT MODEL TRAINING APPROACHES FOR LTR DATASET.

Model	Micro F1-score	Macro F1-score
Single-Registry	0.631 (0.627, 0.636)	0.319 (0.314, 0.336)
Centralized	0.647 (0.643, 0.651)	0.348 (0.343, 0.366)
Privacy-Preserving	0.647 (0.642, 0.651)	0.343 (0.338, 0.359)

TABLE III
MICRO AND MACRO F1-SCORE (WITH 95% CONFIDENCE INTERVALS) OF DIFFERENT MODEL TRAINING APPROACHES FOR KCR DATASET.

Model	Micro F1-score	Macro F1-score
Single-Registry	0.624 (0.619, 0.629)	0.294 (0.291, 0.312)
Centralized	0.646 (0.642, 0.651)	0.343 (0.336, 0.362)
Privacy-Preserving	0.642 (0.637, 0.647)	0.342 (0.337, 0.362)

that low prevalent classes benefit more from collaboration since the dataset size increases from single-registry data to multiple-registry data. The centralized model performance serves as the optimal one for the privacy-preserving model to reach while preserving patient privacy. The privacy-preserving distributed DL model training significantly outperforms the baseline single-registry model with micro and macro F1-score of (0.647,0.343) and (0.642,0.342) for LTR and KCR datasets respectively. Compared to the centralized model, the proposed approach is statistically indistinguishable for both registries. In addition, this approach offers high data privacy and protection since neither registry-specific data nor model parameters are shared across cancer registries. Although both LTR and KCR datasets benefit from the collaboration, the centralized and the privacy-preserving models improve the performance of KCR single-registry model more than the LTR single-registry model. This performance difference can be attributed to the larger size and more inclusive class representation of the LTR dataset compared to the KCR dataset.

To show the performance of data sharing and model distribution approaches on tackling the class imbalance problem, which is common in clinical datasets, we compute the F1-score per class label of all models. Since we have hundreds of subsite labels, we illustrate in Figures 3 and 4 the performance of different training approaches on the most and least represented classes with at least 100 samples. The figures clearly show that non-private data sharing, the centralized model, and privacy-preserving model distribution approaches improve the classification accuracy of low prevalent class labels. However, all models perform equally well on the more prevalent classes.

V. CONCLUSION

In this paper, we propose a privacy-preserving model distribution across cancer registries technique. It shares the registry-agnostic DL model parameters across cancer registries, and excludes any registry-specific parameters that may compromise privacy violations. The proposed method eliminates the need of a centralized host for datasets and the need of complex encryption techniques to privately distribute DL models across the collaborating institutes. The experiments demonstrate that our proposed DL training approach significantly outperforms the single-registry model and achieves a comparable performance to the centralized model. Future directions of this work include applying the proposed approach on more cancer registries and information extraction tasks as well as using the proposed approach to securely distribute deep learning model across registries through recently developed model distribution techniques, such as cyclical weight transfer [23].

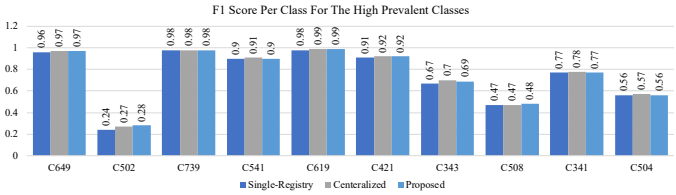

Fig. 3. F1-score of the most represented ten classes from KCR dataset.

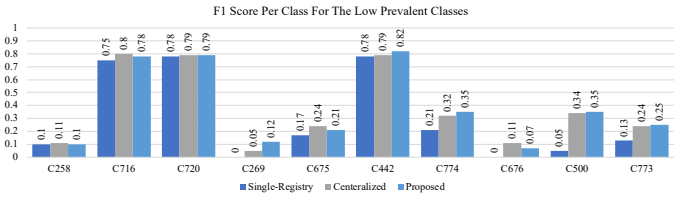

Fig. 4. F1-score of the least represented ten classes from KCR dataset with at least 100 samples.

ACKNOWLEDGMENT

This work has been supported in part by the Joint Design of Advanced Computing Solutions for Cancer (JDACS4C) program established by the U.S. Department of Energy (DOE) and the National Cancer Institute (NCI) of the National Institutes of Health. This work was performed under the auspices of the U.S. Department of Energy by Argonne National Laboratory under Contract DE-AC02-06-CH11357, Lawrence Livermore National Laboratory under Contract DEAC52-07NA27344, Los Alamos National Laboratory under Contract DE-AC5206NA25396, and Oak Ridge National Laboratory under Contract DE-AC05-00OR22725.

This work has also been supported by National Cancer Institute under Contract No. HHSN2612018000013I and NCI Cancer Center Support Grant (P30CA177558).

REFERENCES

[1] S. Gao, J. X. Qiu, M. Alawad, J. D. Hinkle, N. Schaefferkoetter, H.-J. Yoon, B. Christian, P. A. Fearn, L. Penberthy, X.-C. Wu, L. Coyle, G. Tourassi, and A. Ramanathan, "Classifying cancer pathology reports with hierarchical self-attention networks," *Artificial Intelligence in Medicine*, vol. 101, 2019.

[2] M. Alawad, S. Gao, J. X. Qiu, H. J. Yoon, J. BlairChristian, L. Penberthy, X.-C. Wu, L. Coyle, and G. Tourassi, "Automatic extraction of cancer registryreportable information from free-text pathologyreports using multi-task convolutional neuralnetworks," *J Am Med Inform Assoc*, 2019.

[3] M. Alawad, S. Gao, J. Qiu, N. Schaefferkoetter, J. D. Hinkle, H. Yoon, J. B. Christian, X. Wu, E. B. Durbin, J. C. Jeong, I. Hands, D. Rust, and G. Tourassi, "Deep transfer learning across cancer registries for information extraction from pathology reports," in *2019 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, pp. 1–4, May 2019.

[4] P. Vepakomma, O. Gupta, T. Swedish, and R. Raskar, "Split learning for health: Distributed deep learning without sharing raw patient data," *CoRR*, vol. abs/1812.00564, 2018.

[5] Y. Guo, R. Gaizauskas, I. Roberts, and G. Demetriou, "Identifying personal health information using support vector machines," in *i2b2 Workshop on Challenges in Natural Language Processing for Clinical Data*, 2006.

[6] F. Dernoncourt, J. Y. Lee, Ö. Uzuner, and P. Szolovits, "De-identification of patient notes with recurrent neural networks," *J Am Med Inform Assoc*, vol. 1;24(3), pp. 596–606, 2017.

[7] M. N. Sadat, M. M. A. Aziz, N. Mohammed, S. Pakhomov, H. Liu, and X. Jiang, "A privacy-preserving distributed filtering framework for nlp artifacts," *BMC Medical Informatics and Decision Making*, vol. 19, p. 183, Sep 2019.

[8] A. Stubbs, M. Filannino, and z. Uzuner, "De-identification of psychiatric intake records," *J. of Biomedical Informatics*, vol. 75, pp. S4–S18, Nov. 2017.

[9] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, 20-22 April 2017, Fort Lauderdale, FL, USA*, pp. 1273–1282, 2017.

[10] J. Chen, R. Monga, S. Bengio, and R. Jozefowicz, "Revisiting distributed synchronous sgd," in *International Conference on Learning Representations Workshop Track*, 2016.

[11] B. Hitaj, G. Ateniese, and F. Perez-Cruz, "Deep models under the gan: Information leakage from collaborative deep learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, CCS '17, (New York, NY, USA), pp. 603–618, ACM, 2017.

[12] P. Vepakomma, T. Swedish, R. Raskar, O. Gupta, and A. Dubey, "No peek: A survey of private distributed deep learning," *CoRR*, vol. abs/1812.03288, 2018.

[13] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, (New York, NY, USA), pp. 308–318, ACM, 2016.

[14] L. T. Phong, Y. Aono, T. Hayashi, L. Wang, and S. Moriai, "Privacy-preserving deep learning via additively homomorphic encryption," *Trans. Info. For. Sec.*, vol. 13, pp. 1333–1345, May 2018.

[15] J. X. Qiu, H. J. Yoon, P. A. Fearn, and G. D. Tourassi, "Deep learning for automated extraction of primary sites from cancer pathology reports," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, pp. 244–251, Jan 2018.

[16] Y. Kim, "Convolutional neural networks for sentence classification," *CoRR*, vol. abs/1408.5882, 2014.

[17] M. Hughes, I. Li, S. Kotoulas, and T. Suzumura, "Medical text classification using convolutional neural networks," *Studies in health technology and informatics*, vol. 235, pp. 246–250, 2017.

[18] B. He, Y. Guan, and R. Dai, "Classifying medical relations in clinical text via convolutional neural networks," *Artificial Intelligence in Medicine*, vol. 93, pp. 43–49, 2019.

[19] J. Liu, Z. Zhang, and N. Razavian, "Deep ehr: Chronic disease prediction using medical notes.," *arXiv preprint arXiv:1808.04928*, pp. 440–464, 2018.

[20] M. Alawad, H. Yoon, and G. D. Tourassi, "Coarse-to-fine multi-task training of convolutional neural networks for automated information extraction from cancer pathology reports," in *2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, pp. 218–221, March 2018.

[21] P. Liu, X. Qiu, and X. Huang, "Adversarial multi-task learning for text classification," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, (Vancouver, Canada), pp. 1–10, Association for Computational Linguistics, July 2017.

[22] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*. Chapman and Hall/CRC Monographs on Statistics and Applied Probability, Taylor and Francis, 1994.

[23] K. Chang, N. Balachandar, C. Lam, D. Yi, J. Brown, A. Beers, B. Rosen, D. L. Rubin, and J. Kalpathy-Cramer, "Distributed deep learning networks among institutions for medical imaging," *Journal of the American Medical Informatics Association*, vol. 25, pp. 945–954, 03 2018.