V. CONCLUSION

In order to progress machine learning research further in ultrasound, significant investment in data collection and annotation will be required, but this burden can be significantly reduced using targeted sampling methods. This paper proposed a 2 phase method to optimize the collection and annotation processes, demonstrated rigorously and repeatably per phase on public datasets, before being shown as effective together on a practical example. These methods provide ultrasound researchers with a method to not only identify the most effective sample size when collecting data but also a method to maximize annotation effectiveness potentially producing robust algorithms at reduced cost.

Using a simple statistical power curve to predict accuracy results will allow researchers to provide an estimate as to the usefulness of additional data in NN training. This allows the results of limited feasibility studies using relatively small datasets to inform the selection and design of subsequent larger studies just as with clinical trials. Active learning using uncertainty sampling to reduces annotation effort on a smaller subset, useful especially where a data collection is ongoing or a large unlabeled dataset is available. These methods can be used independently, but give best results when used together. This method is to be used to streamline future clinical trials by the authors, and applied to other ultrasound applications like NDT and industrial inspection to standardize methods.

While this study has focused on accuracy as the sole metric, additional validation metrics would allow for these methods to be significantly finetuned to produce the best network response at the lowest cost. Alexnet and per image classification provided an adequate baseline but further research into more complex architectures and ultrasound datasets with a more complex taxonomy would offer further insight into developing additional methods to reduce the cost of producing and annotating ultrasound data, especially where overlapping classifiers exist within the dataset.

REFERENCES

[1] B. Luijten *et al.*, "Adaptive ultrasound beamforming using deep learning," *IEEE Transactions on Medical Imaging,* vol. 39, no. 12, pp. 3967-3978, 2020.

[2] L. H. Lee, Y. Gao, and J. A. Noble, "Principled ultrasound data augmentation for classification of standard planes," in *International Conference on Information Processing in Medical Imaging*, 2021, pp. 729-741: Springer.

[3] P.-H. Yeung, M. Aliasi, A. T. Papageorghiou, M. Haak, W. Xie, and A. I. Namburete, "Learning to map 2D ultrasound images into 3D space with minimal human annotation," *Medical Image Analysis,* vol. 70, p. 101998, 2021.

[4] R. Huang, J. A. Noble, and A. I. Namburete, "Omni-supervised learning: scaling up to large unlabelled medical datasets," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018, pp. 572-580: Springer.

[5] L. Schmarje, *et al*., "A survey on semi-, self-and unsupervised learning for image classification," *IEEE Access,* vol. 9, pp. 82146-82168, 2021.

[6] K. Hemming, *et al*. "How to design efficient cluster randomised trials," *bmj,* vol. 358, 2017.

[7] D. J. Biau, S. Kernéis, and R. Porcher, "Statistics in brief: the importance of sample size in the planning and interpretation of medical research," *Clinical orthopaedics and related research,* vol. 466, no. 9, pp. 2282-2288, 2008.

[8] A. Rokem, Y. Wu, and A. Y. Lee, "Assessment of the need for separate test set and number of medical images necessary for deep learning: a sub-sampling study," *bioRxiv*, p. 196659, 2017.

[9] J. Cho, *et al*., "How much data is needed to train a medical image deep learning system to achieve necessary high accuracy?," *arXiv preprint arXiv:1511.06348,* 2015.

[10] H. G. Schnack and R. S. Kahn, "Detecting neuroimaging biomarkers for psychiatric disorders: sample size matters," *Frontiers in psychiatry,* vol. 7, p. 50, 2016.

[11] J. Escudero, J. P. Zajicek, and E. Ifeachor, "Machine Learning classification of MRI features of Alzheimer's disease and mild cognitive impairment subjects to reduce the sample size in clinical trials," in *2011 Annual international conference of the IEEE engineering in medicine and biology society*, 2011, pp. 7957-7960: IEEE.

[12] D. Calaprice-Whitty, K. Galil, W. Salloum, A. Zariv, and B. Jimenez, "Improving clinical trial participant prescreening with artificial intelligence (AI): a comparison of the results of AI-assisted vs standard methods in 3 oncology trials," *Therapeutic innovation & regulatory science,* vol. 54, no. 1, pp. 69-74, 2020.

[13] D. L. Labovitz, *et al*., "Using artificial intelligence to reduce the risk of nonadherence in patients on anticoagulation therapy," *Stroke,* vol. 48, no. 5, pp. 1416-1419, 2017.

[14] B. S. Glicksberg *et al*., "Automated disease cohort selection using word embeddings from electronic health records," in *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2018: Proceedings of the Pacific Symposium*, 2018, pp. 145-156: World Scientific.

[15] J. Romero, S. Chiang, and D. M. Goldenholz, "Can machine learning improve randomized clinical trial analysis?," *Seizure,* vol. 91, pp. 499-502, 2021.

[16] N. Zhou and P. Manser, "Does including machine learning predictions in ALS clinical trial analysis improve statistical power?," *Annals of Clinical and Translational Neurology,* vol. 7, no. 10, pp. 1756-1765, 2020.

[17] M. Hasenjäger and H. Ritter, "Active learning in neural networks," in *New learning paradigms in soft computing*: Springer, 2002, pp. 137-169.

[18] Y. Yang, Z. Ma, F. Nie, X. Chang, and A. G. Hauptmann, "Multi-class active learning by uncertainty sampling with diversity maximization," *International Journal of Computer Vision,* vol. 113, no. 2, pp. 113-127, 2015.

[19] D. D. Lewis and W. A. Gale, "A sequential algorithm for training text classifiers," in *SIGIR'94*, 1994, pp. 3-12: Springer.

[20] J. Yun, J. Oh, and I. Yun, "Gradually Applying Weakly Supervised and Active Learning for Mass Detection in Breast Ultrasound Images," *Applied Sciences,* vol. 10, no. 13, 2020.

[21] G. Liu *et al.*, "Breast Ultrasound Tumor Detection Based on Active Learning and Deep Learning," EasyChair 2021.

[22] L. Gao *et al.*, "Multi-modal active learning for automatic liver fibrosis diagnosis based on ultrasound shear wave elastography," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 2021, pp. 410-414: IEEE.

[23] L. Liu, *et al*., "Semi-supervised active learning for COVID-19 lung ultrasound multi-symptom classification," in *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)*, 2020, pp. 1268-1273: IEEE.

[24] W. Al-Dhabyani, *et al*., "Dataset of breast ultrasound images," *Data in brief,* vol. 28, p. 104863, 2020.

[25] J. Born *et al.*, "Accelerating detection of lung pathologies with explainable ultrasound image analysis," *Applied Sciences,* vol. 11, no. 2, p. 672, 2021.

[26] X. P. Burgos-Artizzu *et al.*, "Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes," *Scientific Reports,* vol. 10, no. 1, pp. 1-12, 2020.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems,* vol. 25, pp. 1097-1105, 2012.

[28] O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision,* vol. 115, no. 3, pp. 211-252, 2015.