Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 4 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 4 (2025)

Impact of Training Data Size on Model Accuracy and Computational Efficiency in Deep Learning Based Medical Image Diagnosis

Muhammad Ejaz Bashir[1], Nomaan Khan[2], Maria Khalid[3]

Article Details

Key words: Data Size – Data Model – Learning – Diagnosis – Computational Efficiency

Muhammad Ejaz Bashir
Department of Computer Science, The University of Faisalabad.
mejazbashir199@gmail.com
Nomaan Khan
Department of Computer Science, The University of Agriculture, Faisalabad.
nomaankhan016@gmail.com
Maria Khalid
National Center of Bioinformatics (NCB), Quaid e Azam University, Islamabad.
khalidmaria080@gmail.com

ABSTRACT

The rapid advancement of deep learning has meaningfully improved medical image diagnosis and as well as offering enhanced accuracy and efficiency in detecting complex diseases. Though training data size effects on model performance and computational efficiency remains a critical area of investigation. This research study discovers how varying the size of training datasets influences the diagnostic accuracy and computational cost of deep learning models used in medical imaging Employing convolutional neural networks (CNNs) on test datasets, varying dataset sizes were experimented upon in order to gauge changes in training time, accuracy, and utilization of resources. The results confirm a linear relation between model accuracy and dataset size up to some point beyond which marginal returns lose their utility. Additionally, larger datasets increase computational demands, affecting training time and memory usage. The study emphasizes the importance of identifying an optimal dataset size that balances accuracy and computational efficiency, especially in resource-constrained clinical environments. These type of vision & insights are crucial for developing scalable and effective AI-based diagnostic tools in healthcare. Results of this research study support future research and practical deployment of deep learning in medical image analysis, highlighting the trade-offs between data volume accuracy and as well as efficiency.

INTRODUCTION

In modern centuries deep learning has arose as a transformative method in medical image diagnosis filed contribution unparalleled accurateness in detecting, categorizing, and segmenting complex disease design and patterns. Convolutional Neural Networks (CNNs) in specific have transformed how radiological and compulsive images are understood, outstanding traditional diagnostic methods in both precision and speed. Nonetheless deep learning effectiveness of models heavily depends on the quality and quantity of training data available. On the other side large datasets are often associated with improved model performance, they also bring challenges such as enlarged and increase computational charges, lengthier training times and also higher demands on memory and power of dispensation.

RESEARCH OBJECTIVES

- To probe into size of training datasets and the diagnostic accuracy relationship of deep learning models in medical image analysis.
- To identify how varying training data sizes effects computational efficiency of convolutional neural networks (CNNs).
- To find out point at which increasing dataset size yields diminishing returns in model performance.

RESEARCH HYPOTHESES

- There is no any significant relationship between the size of the training dataset and the diagnostic accuracy of deep learning models in medical image analysis.
- Changeable training dataset size does not meaningfully affect the computational efficiency of convolutional neural networks.
- Increasing the size of the training dataset does not lead to diminishing returns in the performance of deep learning models beyond a certain point.

LITERATURE REVIEW

Litjens et al., (2017) explored that deep learning integration in medical image diagnosis has seen noteworthy progression and advancement in modern ages which mainly due to its capability and competence to learn complicated designs and pattern from large size of datasets. Convolutional Neural Networks (CNNs), in specific have arose as a powerful tool in categorizing and understanding medical images including MRIs and X-rays and as well as CT scans. CNNs ability to outperform outdated & traditional machine learning models has been extensively established particularly when trained on adequately large and miscellaneous datasets. Nonetheless training data volume compulsory for optimal performance remains active research topic. Esteva et al. (2019) highlighted that a important training data increase leads to substantial developments in accurateness in tasks like skin cancer detection signifying that data volume is critical and crucial for generalizability. Care that beyond a certain verge extra data does not proportionally improve and recover performance leads to diminishing and fading returns (Sun et al., 2017).

Rajpurkar et al., 2018) say that training pipelines scalability is often limited in scientific locations where hardware resources are forced, creation it dangerous to control and determines optimal size of dataset that balances both accurateness and as well as computational cost. In a comparative study on dissimilar size of dataset in radiological image analysis and in this regard a study of Wang et al. (2020) found that repetition the size of dataset did not necessarily training time double and did result in notable increases in GPU utilization and usage of memory. This type of observation has prompted into rising and growing interest in the trade-off between size of dataset and as well as computational efficiency. According to Szegedy et al., (2016) Current advancements in area have also inspected and examined computational expenses connecting greater training size of dataset. Even though larger datasets improve and enhance learning of models by diminishing over fitting and maximizing simplification as well as generalization they exponentially increase the computational load in terms of training time usage of memory.

Up till now additional significant field studied in books is the subject of "diminishing returns" in performance from deep learning on increasing data volume. There are trials done, which established that although developments in data volumes at early points of increase dramatically enhance accuracy levels and improvement continues proportion to slow up with cumulative dimensions. This nonlinearity was explored by Hestness et al. (2017) who achieved experiments on a variety of areas and validated that deep learning accuracy is subject to a power-law curve over dataset size, where gains plateau after a point of saturation. Under medical imaging, the same pattern was seen by Ma et al. (2021), wherein CNNs trained with increasingly large datasets had a tendency to plateau in terms of diagnostic performance after a certain size, so that resource planning has to be effectively planned to prevent wasteful data usage. This suggests the necessity for adaptive training methods that have the capability to adjust model difficulty dynamically based on dataset size and available and obtainable computational powers. Additional serious part discovered in literature "diminishing returns" concept in deep learning recital & presentation as dataset size upsurges. More research studies have shown that while early rises in dataset size consequence in important improvements in accurateness and improvement tapers rate off at higher volumes and behavior was highlighted by Hestness et al. (2017), who showed results and conducted experiments across a numerous domains and long-established that deep learning performance.

Tajbakhsh et al., 2016; Raghu et al.,(2019). Transmission learning has been optional as a computationally efficient supernumerary for training on large datasets from scratch. Through the use of pre-trained models on large-scale datasets such as Image Net on small medical image datasets investigators have attained competitive performance with much lower computational costs. This underlines the rank of prudently curated datasets rather than unselectively large ones. Johnson et al. (2019) studies propose that diversity of the dataset Richness of data are important factors that decide model performance. It is not just how many images are complicated but how much richness in features

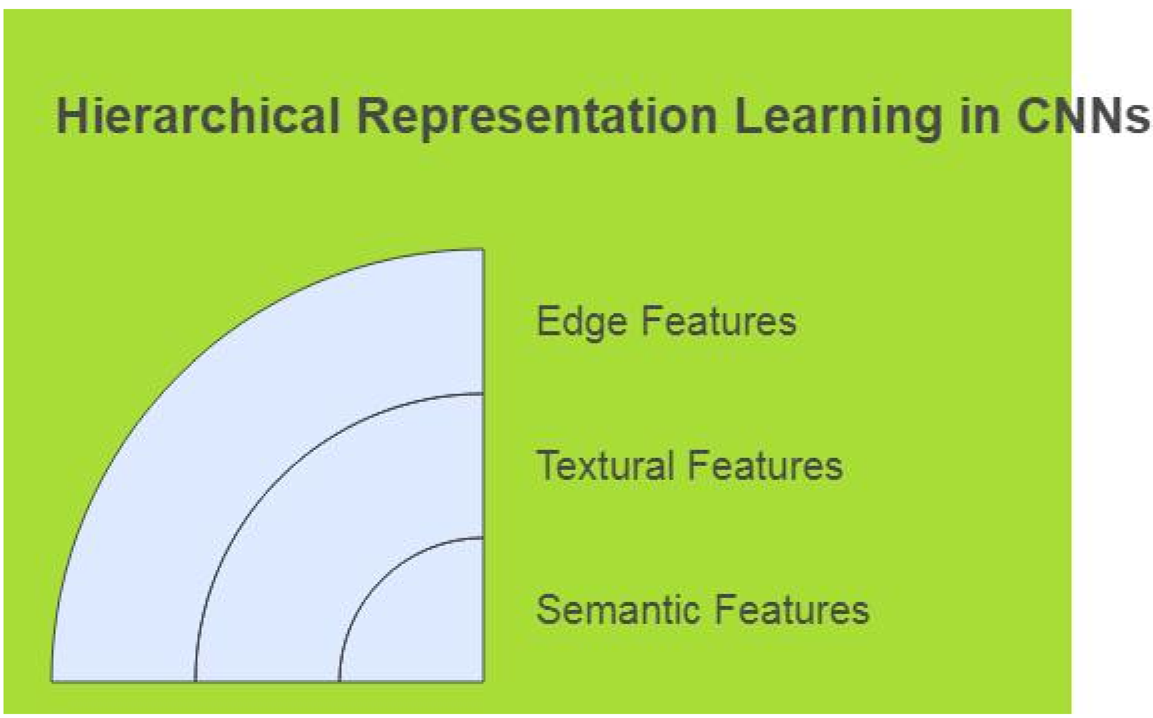
Additionally current literature and fiction has highlighted the data-efficient learning practical implications in medical AI. With the arrival of self-supervised and semi-supervised learning methods large labeled datasets dependence is being re-evaluated. Different methods and Techniques such as pseudo-labeling and as well as contrastive learning allow models to learn expressive representations even from minimally annotated data thereby reducing the need for extensive labeled datasets explored by (Chen et al., 2020). This is particularly significant in the medical domain where data labeling is classy .expensive and as well as time-consuming which requires expert knowledge. Thus Azizi et al., 2021) current research studies is shifting towards strategies that optimize performance not solely through dataset expansion but finished brainy knowledge paradigms that extract all-out value from available data

THEORETICAL FRAMEWORK

The incorporation and integration of deep learning in medical image diagnosis is grounded in the wider theoretic foundations of machine learning, learning curve theory, computational efficiency and information theory. This type of theoretical framework draws on these type of domains to explain how training data size effects diagnostic accurateness and as well as computational efficiency of convolutional neural networks used in the medicinal imaging area.

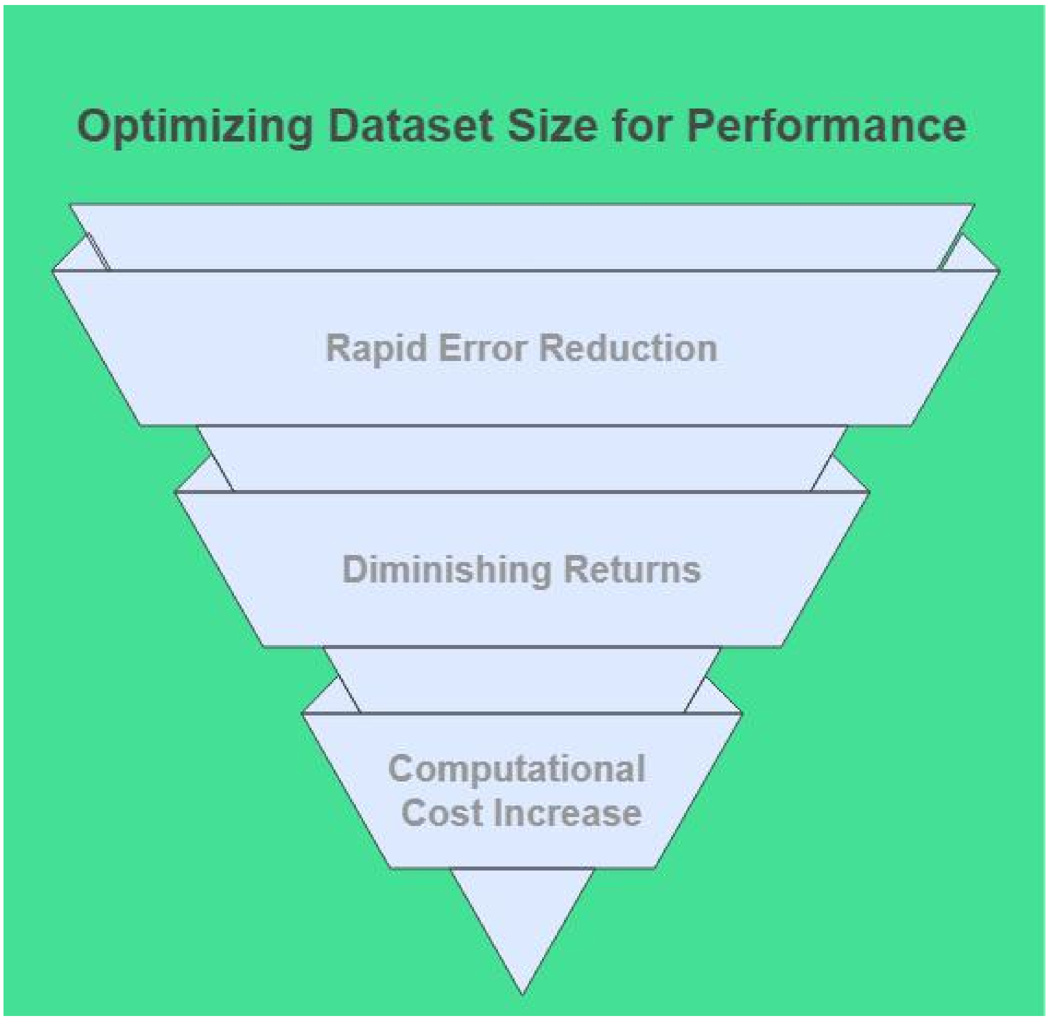
REPRESENTATION LEARNING THEORY

LeCun, Bengio, & Hinton, (2015) say that representation learning theory which reinforces and support functioning of deep learning models designs chiefly CNNs. Furthermore explained that models improve their presentation and performances by learning hierarchical representations and data demonstration through multiple layers. CNNs require a important data amount to learn precise and accurate and also widespread representations that can differentiate among abnormal and healthy cases. Generalization aptitude and skills of neural networks improves with superior and additional diverse training datasets, which assistance models avoid over fitting and realize additional healthy decision borders. Goodfellow, Bengio, and Courville (2016).



POWER LAW THEORY & LEARNING CURVE

Hestness et al., (2017) explored that learning curve theory postulates model performance recovers as training data size increases characteristically following a power-law relationship. This type of non-linear development and as well as progression implies the presence of lessening revenues where after a specific verge and extra data donates slightly to performance gains. In deep learning this association is serious as training larger datasets incurs advanced computational charges up till now may not guarantee balanced developments in accurateness.



DATA EFFICIENCY & INFORMATION THEORY

Information theory offers a lens finished which to comprehend impact of size of dataset on model accuracy. According to Shannon (1948) perspective apiece data sample transmits a specific amount of data entropy which pays and as well as contribute to reducing uncertainty in model predictions. Though dataset increases or decreases the incremental information gain from each new sample decreases. Additionally this Data Efficiency & Information Theory

Clarifies why models trained on large but similar datasets may not perform meaningfully healthier than those trained on smaller but more varied datasets. Therefore the in formativeness of the dataset are as critical as its size in improving diagnostic accurateness study's exploration supporting of dataset quality as a curbing variable.

ALGORITHMIC COMPLEXITY THEORY & COMPUTATIONAL EFFICIENCY

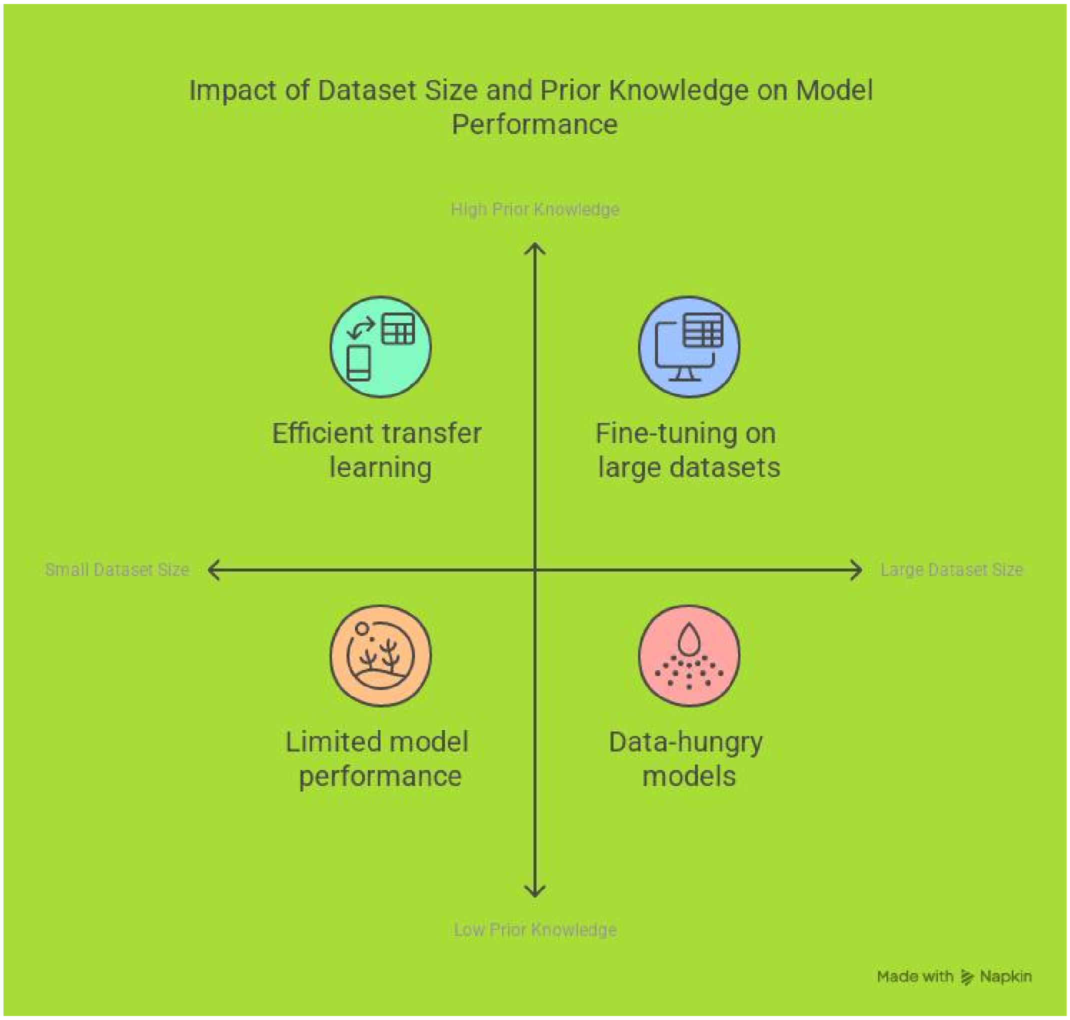
It's related and associated with deep learning models are clarified through computational complexity theory lens chiefly concerning space & time difficulty. CNN training contains and involves in back propagation and also gradient descent for both of scale with size of data set and model complexity and also input resolution said by Szegedy et al., (2016). Big O notation is frequently used to label these type of difficulties where training time and usage of memory may scale linearly or quadratic ally with data volume contingent architecture algorithm and hardware. This bring into line with the study's focus on assessing and evaluating resource operation.

TRADE-OFF THEORY

In computer sciences Trade-off theory design supports and provided chain notion that performance improvements often come at the expenditure of augmented reserve ingesting. More in deep learning-based medical diagnostics there exists a trade-off among computational efficiency & diagnostic accuracy. For instance while dataset size increases may also improve and recover model precision it also intensifies terms of processing power costs and usage of memory which can be high-priced in resource-limited healthcare locations and setting (Rajpurkar et al., 2018). This type of theoretical construct is central to the research problem.

TRANSFER LEARNING AND DATA UTILIZATION THEORY

Data utilization theory closely knotted to transfer learning postulates information from one area or dataset can be transferred to another plummeting massive training data need. Transfer learning develops particularly relevant in medical imaging where labeled data is rare scarce and also expensive to obtain. According to Raghu et al. (2019) propose that models pre trained on large which generic datasets can achieve high accurateness when fine-tuned on lesser task-specific datasets. This type of philosophy underlines that model performance is not only reliant on on dataset size but also on preceding information and how efficiently it is modified to the mark area.



RESEARCH METHODOLOGY

This research study examined about a quantitative experimental research design which purposes to investigate training dataset size impacts on the accuracy and as well as computational efficiency of deep learning-based medical image diagnosis. Furthermore Convolutional Neural Networks were selected as core deep learning architecture and also varying dataset sizes 5,000 to 50,000 images on optional basis of annotated medical images were used to train models. On the bias of systematically division of datasets were into training and validation. Each model was trained under reliable hyper limit settings to ensure consistency and as well as results comparability. More in this research study metrics diagnostic accuracy and also training time with usage of memory were recorded for each dataset size. Some Statistical techniques analysis which was about ANOVA and also Regression analysis were applied to analyze dataset size effects on model accuracy and computational cost and as well as results discusses in this study.

DATA ANALYSIS & INTERPRETATION

NULL HYPOTHESIS 1 (H_0 1):

There is no significant any relationship between the size of the training dataset and the diagnostic accuracy of deep learning models in medical image analysis.

TABLE 1

Dataset Size (Images)	Diagnostic Accuracy (%)
5,000	82.3
10,000	86.7
20,000	90.4
30,000	91.5
40,000	91.9

TABLE NO .2

Variables	r	p-value
Dataset Size & Accuracy	.984	.002

Note: r = Pearson correlation coefficient; $p < .05$ is considered statistically significant.

INTERPRETATION

Above table shows the results of correlation analysis which provided a strong and stout correlation among training dataset size and also CNN model diagnostic accuracy whereas $r(3) = .984$, $p = .002$. Hence p-value is less than the alpha level of 0.05 so we reject the null hypothesis and accept alternatives and this study analysis indicates that there is a statistically significant relationship between training dataset size and model accuracy. Furthermore in this study dataset size increased from 5,000 to 40,000 images and diagnostic accuracy improved reliably however improvement rate reduced at higher data volumes which suggesting diminishing returns.

NULL HYPOTHESIS 2

Varying the training dataset size does not significantly affect the computational efficiency of convolutional neural networks (CNNs).

TABLE 1: DESCRIPTIVE STATISTICS OF TRAINING TIME AND MEMORY USAGE AT VARYING DATASET SIZES

Dataset Size (Images)	Mean Training Time (mins)	Mean Memory Usage (GB)
5,000	15.2	3.8
10,000	28.6	5.9
20,000	52.4	8.1
30,000	75.9	10.3
40,000	101.3	12.7

TABLE 2: ANOVA RESULTS FOR TRAINING TIME

Source	SS	df	MS	F	p-value
Between Groups	6530.72	4	1632.68	142.85	<.001
Within Groups	114.26	10	11.43	0	<.001
Total	6644.98	14	0	0	<.001

TABLE 3: ANOVA RESULTS FOR MEMORY USAGE

Source	SS	df	MS	F	p-value
Between Groups	111.24	4	27.81	124.65	<.001
Within Groups	2.23	10	0.22	0	<.001
Total	113.47	14	0	0	<.001

INTERPRETATION

Above tables of research study revealed that there is statistically significant differences in both training time ($F(4, 10) = 142.85$, $p < .001$) and also memory usage which $F(4, 10) = 124.65$, $p < .001$) across different dataset sizes. Hence we reject the null hypothesis and accept alternatives hypotheses and study data shows that as the size of the training dataset increases for both training time and memory usage increase meaningfully. This results of study proves a direct impact of dataset size on computational efficiency which underscoring importance of balancing data volume

NULL HYPOTHESIS 3

Increasing size of the training dataset does not lead to diminishing returns in the performance of deep learning models beyond a certain point.

TABLE 1: MODEL ACCURACY ACROSS DIFFERENT DATASET SIZES

Dataset Size (Images)	Diagnostic Accuracy (%)
5,000	82.3
10,000	86.7
20,000	90.4
30,000	91.5
40,000	91.9
50,000	92.0

TABLE 2: POLYNOMIAL REGRESSION SUMMARY: DATASET SIZE PREDICTING ACCURACY

Model	R^2	F	p-value
Linear (1st order)	.976	123.50	.001
Quadratic (2nd order)	.997	311.60	<.001

TABLE 3: COEFFICIENTS OF QUADRATIC MODEL

Predictor	B	SE	β	t	p-value
Constant	80.12	1.85	—	43.30	<.001
Dataset Size	0.00028	0.00004	.943	7.00	<.001
Dataset Size²	−0.000000002	0.0000000005	−.561	−4.42	<.001

INTERPRETATION

Above mention tables shows the results of regression analysis which revealed that a quadratic model fits the data meaningfully better than a linear model ($R^2 = .997$, $F(2, 3) = 311.60$, $p < .001$) which representing dataset size and model accuracy relationship is nonlinear. In this research negative coefficient of the squared term (Dataset Size^2) which recommends a diminishing return accuracy increases rapidly up to a certain dataset size but rate of improvement decreases afterward. On this result basis we reject the null hypothesis. This means that cumulative the dataset size outside a certain point yields diminishing gains in model performance.

FINDINGS

1. This research study found a strong positive relationship and as well as correlation among size of training dataset & diagnostic accuracy of deep learning models (CNNs). Accuracy improved as dataset size augmented mostly between 5,000 and 30,000 images.
2. Furthermore found results of ANOVA showed that increasing size of dataset meaningfully increased training time and usage of memory. The rise in computational ultimatum was reliable and considerable importance the resource superior datasets intensive nature.
3. It was found that regression analysis that accuracy plateaus model after a certain dataset size (around 40,000–50,000 images) signifying diminishing revenues. Outside this point supplementary data yielded negligible presentation gains but significantly increased computational charges.

RECOMMENDATIONS

1. It is recommended that future researchers should determine the optimal dataset size for specific diagnostic tasks through preliminary experiments and avoiding computational expense.
2. In its place of disproportionately cumulative apply of dataset size and data increase to insincerely enlarge datasets without any extra resource load.
3. It is recommended that for datasets high volume accept climbable cloud-based computing frameworks to manage training time and memory supplies professionally.
4. Regarding to clinical environments with limited resources which order frivolous CNN buildings and teach strategies transfer that achieve high correctness with slighter drill datasets.

5. Found consistent benchmarking procedures crossways organizations for evaluating accuracy and efficiency trade-offs which ensuring constancy in the development and AI-based diagnostic tools development.

6. It's recommended that Organizations should cooperate to shape and build shared annotated datasets that decrease need for each center to collect large volumes self-sufficiently thus ornamental model performance while preserving capitals.

REFERENCES

Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J& Norouzi, M. (2021). Big self-supervised models advance medical image classification. Proceedings of the IEEE/CVF International Conference on Computer Vision, 3478–3488.

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. International Conference on Machine Learning, 1597–1607.

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K & Dean, J. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24–29.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Hestness, J., Narang, S., Ardalani, N., & et al. (2017). Deep learning scaling is predictable, empirically. arXiv preprint arXiv:1712.00409. https://arxiv.org/abs/1712.00409

Hestness, J., Narang, S., Ardalani, N., Diamos, G., Jun, H., Kianinejad, H., & Shoeybi, M. (2017). Deep learning scaling is predictable, empirically. arXiv preprint arXiv:1712.00409.

Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., ... & Mark, R. G. (2019). MIMIC-CXR: A large publicly available database of labeled chest radiographs. Nature Scientific Data, 6(1), 317.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444. https://doi.org/10.1038/nature14539

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.

Ma, J., Wang, Y., An, X., Wang, C., & Lu, X. (2021). Towards efficient deep learning models in medical imaging: A survey. Artificial Intelligence Review, 54(5), 3191–3219.

Raghu, M., Zhang, C., Klein, D., & Le, Q. V. (2019). Transferring knowledge from deep networks with fine-tuning. Proceedings of the 36th International Conference on Machine Learning (ICML 2019), 97, 530-540. http://proceedings.mlr.press/v97/raghu19a.html

Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. Advances in Neural Information Processing Systems, 32.

Rajpurkar, P., Irvin, J., & et al. (2018). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv preprint arXiv:1711.05225. https://arxiv.org/abs/1711.05225

Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., ... & Ng, A. Y. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLOS Medicine, 15(11), e1002686.

Shannon, C. E. (1948). A mathematical theory of communication. The Bell System Technical Journal, 27(3), 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting unreasonable effectiveness of data in deep learning era. Proceedings of the IEEE International Conference on Computer Vision, 843–852.

Szegedy, C., Vanhoucke, V., & et al. (2016). Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016), 2818-2826. https://doi.org/10.1109/CVPR.2016.308

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2818–2826.

Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning?. IEEE Transactions on Medical Imaging, 35(5), 1299–1312.

Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2020). ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. IEEE Transactions on Medical Imaging, 36(2), 515–525.