



# Accuracy of deep learning for automated detection of pneumonia using chest X-Ray images: A systematic review and meta-analysis



Yuanyuan Li <sup>a</sup>, Zhenyan Zhang <sup>a</sup>, Cong Dai <sup>a</sup>, Qiang Dong <sup>b,\*</sup>, Samireh Badrigilan <sup>c</sup>

<sup>a</sup> Department of Imaging, Yidu Central Hospital of Weifang, Weifang, 262500, China

<sup>b</sup> Department of Imaging, Qingzhou Hospital of Traditional Chinese Medicine, Qingzhou, 262500, China

<sup>c</sup> Department of Medical Physics, Kermanshah University of Medical Sciences, Kermanshah, Iran

## ARTICLE INFO

**Keywords:**  
Deep learning  
Artificial intelligence  
Meta-analysis  
CXR  
Pneumonia

## ABSTRACT

**Background:** Recently, deep learning (DL) algorithms have received widespread popularity in various medical diagnostics. This study aimed to evaluate the diagnostic performance of DL models in the detection and classifying of pneumonia using chest X-ray (CXR) images.

**Methods:** PubMed, Embase, Scopus, Web of Science, and Google Scholar were searched in order to retrieve all studies that implemented a DL algorithm for discriminating pneumonia patients from healthy controls using CXR images. We used bivariate linear mixed models to pool diagnostic estimates including sensitivity (SE), specificity (SP), positive likelihood ratio (PLR), negative likelihood ratio (NLR), and diagnostic odds ratio (DOR). Also, the area under receiver operating characteristics curves (AUC) of the included studies was used to estimate the diagnostic value.

**Results:** The pooled SE, SP, PLR, NLR, DOR and AUC for DL in discriminating pneumonia CXRs from controls were 0.98 (95% confidence interval (CI): 0.96–0.99), 0.94 (95% CI: 0.90–0.96), 15.35 (95% CI: 10.04–23.48), 0.02 (95% CI: 0.01–0.04), 718.13 (95% CI: 288.45–1787.93), and 0.99 (95% CI: 0.98–100), respectively. The pooled SE, SP, PLR, NLR, DOR and AUC for DL in discriminating bacterial from viral pneumonia using CXR radiographs were 0.89 (95% CI: 0.79–0.94), 0.89 (95% CI: 0.78–0.95), 8.34 (95% CI: 3.75–18.55), 0.13 (95% CI: 0.06–0.26), 66.14 (95% CI: 17.34–252.37), and 0.95 (0.93–0.97).

**Conclusion:** DL indicated high accuracy performance in classifying pneumonia from normal CXR radiographs and also in distinguishing bacterial from viral pneumonia. However, major methodological concerns should be addressed in future studies for translating to the clinic.

## 1. Introduction

With an annual fatality rate of nearly 4 million, pneumonia is one of the leading causes of death among both children and elderly people [1]. Based on the infectious pathogen, it can be viral, bacterial, and fungal which affects the small air sacs (Alveoli) in the lung [2]. It can be life-threatening for patients with background diseases such as asthma, patients with an impaired immune system, hospitalized infants, and elderly patients placed on ventilators, especially if not diagnosed early.

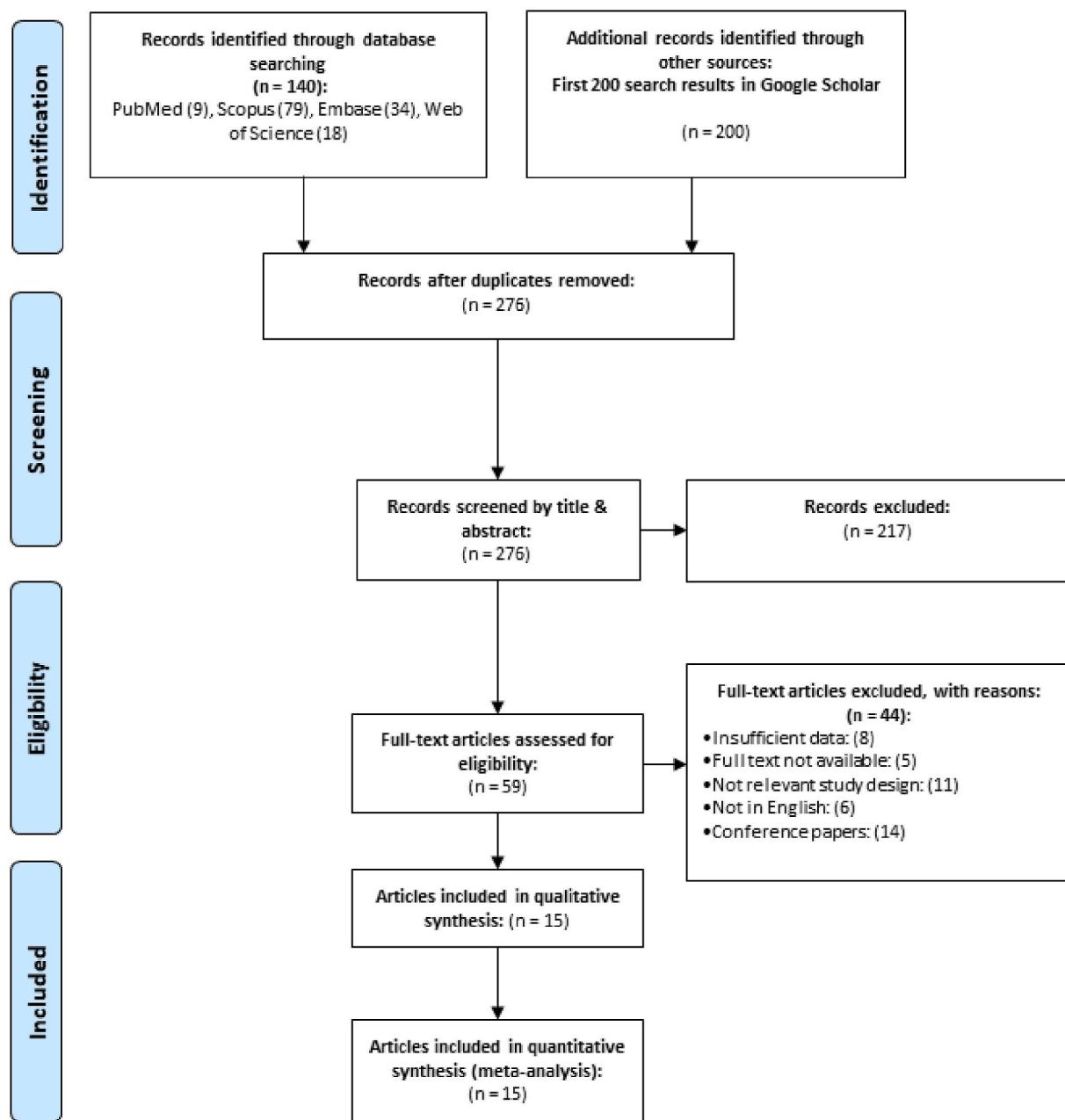
Various imaging modalities are applied to diagnose pneumonia in clinics, including magnetic resonance imaging (MRI), CXR, and computed tomography (CT). Due to the low cost and ease of access, CXR is the most common method used to detect pneumonia worldwide. Chest radiographs provide a significant amount of information about a

patient's condition, however, detecting pneumonia from CXRs is challenging even for highly experienced radiologists as these images have similar opacities for other various lung abnormalities such as lung cancer, and excess fluid. Therefore, conventional detection of pneumonia using CXR images is time-consuming and less accurate and can delay diagnosis and treatment process. As a result, the need for early detection of pneumonia, global use of CXR, and the complexity of interpreting these images make computer-aided detection (CAD) systems a promising approach for automated detection which can help physicians to overcome the above-mentioned problems and improve detection accuracy in a clinical setting [3].

Image preprocessing, extracting regions of interest (ROI) and their features, and feature-based classification of the disease are the main steps in CAD systems. Swift development of machine learning opens up

\* Corresponding author.

E-mail address: [dfty0522@163.com](mailto:dfty0522@163.com) (Q. Dong).



**Fig. 1.** PRISMA flowchart of the identification of eligible studies.

new opportunities for evolving artificial intelligence (AI)-based CADs in medical settings. Deep learning (DL) is a method that is globally used in medical image analysis due to its applicability in replacing the process of feature extraction and condition classification of CXR images compared with conventional CADs [4]. Furthermore, DL holds great promise in object detection [5–8] and semantic segmentation [9–11]. The main element of the high performance of DL is due to the proficiency of neural networks to learn high-level abstractions of introduced raw data by a learning procedure [12]. Also, advancements in DL, have improved the performance of health care professionals in several imaging modalities including pneumonia detection [13], arrhythmia detection [14], diabetic retinopathy diagnosis [15], and cancer [3]. As a result, DL with its capability of chest abnormality classification by automatically learning the image features is a hotspot research topic.

In this exploratory diagnostic systematic review and meta-analysis, we aimed at evaluating the diagnostic accuracy of DL algorithms in detecting and classifying pneumonia using CXR images.

## 2. Methods

### 2.1. Literature search strategy

A comprehensive literature search was carried out in PubMed, Embase, Scopus, and Web of Science databases up to April 26, 2020, to retrieve all relevant studies in literature which developed a DL model for the diagnosis and classification of pneumonia feature from CXR images. We further retrieved the first 200 search results in the Google Scholar to include any missed articles. The search strategy was generated by the following keywords with respect to each database search strategy: ("pneumonia") AND ("radiography" OR "chest film" OR "chest radiograph" OR "radiograph" OR "X-rays") AND ("deep learning"). There was no limitation with respect to the date or language of publications. An extra search was performed in the reference list of related literature to identify additional studies. This study was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-

**Table 1**  
Characteristics of the included studies.

StudyID Author, Year	Type of Pneumonia	Algorithm architecture	Dataset	No. of Images (Normal/ Pneumonia)	No. of Images for Training/Validation/ Testing	Pneumonia vs Normal	Bacterial vs Viral
						TP/FP/TN/FN	TP/FP/TN/FN
Acharya et al., 2020 [17]	Bacterial/ Viral	DSN	Kaggle [31]	5528 (NA/NA)	5328/NA/300	196/6/94/4	89/14/86/11
Altiparmakis et al., 2019 [18]	Bacterial/ Viral	ResNet-50	Kaggle [31]	5856 (NA/NA)	5216/16/624	316/12/108/4	–
Bhandary et al., 2020 [19]	Bacterial/ Viral	MAN-SVM	Kaggle [31]	2000 (NA/NA)	2000/NA/NA	868/117/871/144	–
Chouhan et al., 2020 [20]	Bacterial/ Viral	AlexNet, DenseNet121, InceptionV3, resNet18 and GoogLeNet neural networks	From [32]	5232 (1346/3883)	5232/NA/624	399/28/206/1	–
Liang et al., 2020 [23]	Bacterial/ Viral	CNN	From [32]	5232 (1346/3883)	5232/NA/624	377/46/13/188	–
Mittal et al., 2020 [2]	Bacterial/ Viral	CapsNet	From [32]	5857 (1583/4274)	4100/879/878	606/18/219/35	–
						625/22/212/16	–
						627/22/215/14	–
Rahman et al., 2020 [24]	Bacterial/ Viral	AlexNet, ResNet18, DenseNet201 and SqueezeNet	Kaggle [31]	5247 (1341/3906)	4828/NA/419	212/3/202/2	187/9/192/10
Rajaraman et al., 2018 [25]	Bacterial/ Viral	VGG16, GAP	From [32]	5232 (1346/3883)	5232/NA/624	375/15/225/9/	238/4/127/21
Sarkar et al., 2020 [26]	Bacterial/ Viral	Deep separable residual learning	From [32]	5856 (NA/NA)	4688/NA/1168	847/14/301/6	–
Sirazitdinov et al., 2019 [27]	NA	RetinaNet, Mask R-CNN	RSNA-PDC From Ref. [4]	26684 (7738/18946)	25684/NA/1000	280/89/558/73	–
Sousa et al., 2019 [28]	Bacterial/ Viral	CNN	From [32]	5232 (1346/3883)	5232/NA/624	389/28/206/1	221/45/103/21
Togacar et al., 2019 [29]	NA	AlexNet, VGG-16, GG-19	Custom	5849 (1583/4266)	4094/NA/1755	12754/471/5	–
Visapp et al., 2020 [22]	Bacterial/ Viral	SqueezeNet, Inception-v3	From [32]	5232 (1346/3883)	5232/NA/624	385/2/232/5	–
Wu et al., 2020 [30]	Bacterial/ Viral	CNN-RF	From [33]	5863 (1574/4265)	3911/NA/1928	513/57/1331/27	–
Gu et al., 2018 [21]	Bacterial/ Viral	FCN	From [32]	4513 (0/4513)	3211/802/500	–	153/43/89/215

Abbreviations: Not Available (NA), Deep Siamese Network (DSN); Residual Network (ResNet); Modified AlexNet- Support Vector Machine (MAN-SVM); Convolutional Neural Network (CNN); multi-layered capsules (CapsNet); Visual Geometry Group (VGG), Global Average Pooling (GAP) layer, Random Forest (RF); Fully Convolutional Networks (FCN), Radiological Society of North America-Pneumonia Detection Challenge (RSNA-PDCA).

Analyses (PRISMA) Statement [16].

## 2.2. Study selection and eligibility criteria

We included studies that met the following criteria: a) development of DL models for the detecting and classifying of pneumonia using CXR images (no restriction on the target population), b) sufficient data were reported or could be calculated from primary data (e.g., specificity (SP), sensitivity (SE), true positives (TP), false negatives (FN), true negatives (TN)) and false positives (FP)). The exclusion criteria were: a) review articles, conference abstracts, book chapters, meta-analysis, editorials, duplicate or non-English articles b) lacking sufficient data, and c) inability to obtain the full text.

Two reviewers independently excluded ineligible studies based on screening of the title and abstract of the retrieved studies, then the full text of remaining studies were retrieved for comprehensive assessment. Any discrepancies were resolved by consensus.

## 2.3. Data extraction

Two reviewers independently extracted the following information from each eligible study, the authors' name, year of publication, type of pneumonia, algorithms, data sets to acquire CXR images, number of images, included patients (e.g. pediatric or elderly), and also TP, FP, FN, and TN were extracted directly or calculated by given SE and SP.

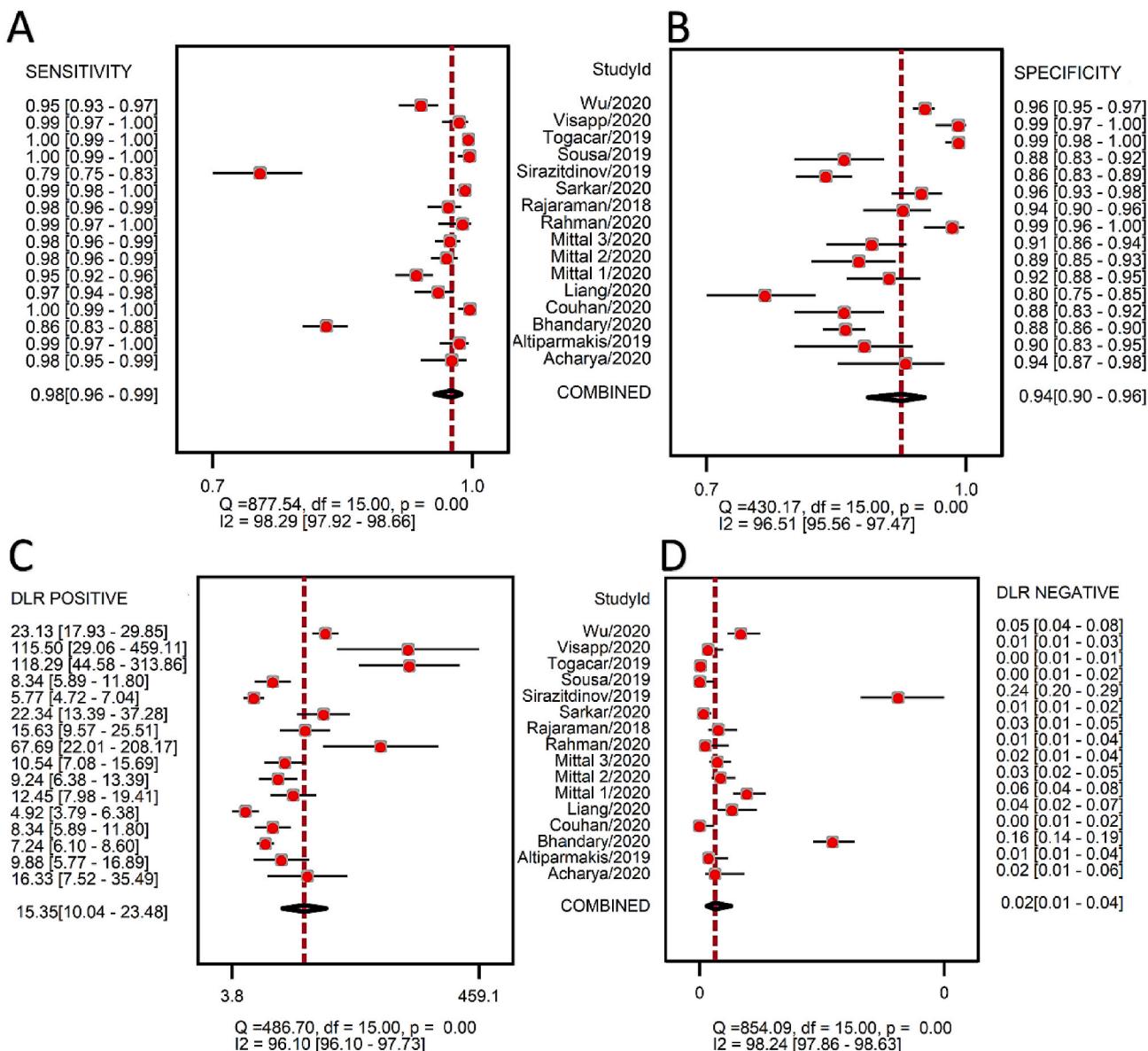
## 2.4. Statistical analysis

Stata 14.2 software (version 14.2; Stata Corporation, College Station, TX) was used to perform meta-analysis. A bivariate linear mixed model was used to pool the estimates. The following test accuracy measures for each study are calculated as follows; pooled SE and SP, positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR), and corresponding 95% confidence intervals (CIs). The analysis was based on the summary receiver operating characteristic (SROC) curves and the area under the curve (AUC) was calculated. Fagan monogram was applied to interpret clinical utility of DL algorithms for diagnosing pneumonia. Spearman correction coefficient test was used to evaluate the threshold effect. In case of non-threshold effect I<sup>2</sup> and χ<sup>2</sup> were used to evaluate the heterogeneity among studies. We used Deek's funnel plot asymmetry test to assess publication bias in the included literature.

## 3. Results

### 3.1. Collection and selection of literature

Fig. 1, demonstrated the process of retrieval of the relevant studies. We identified 340 relevant studies from several databases outlined in materials and methods. After removing 64 duplicated papers, 276 studies were screened by title and abstract, of which 217 studies were excluded due to irrelevant study design or insufficient data. The full text



**Fig. 2.** Forest plot for pooled estimates of DL diagnostic accuracy in discrimination of pneumonia from normal CXR images. A. Sensitivity B. Specificity C. Positive likelihood ratio D. Negative likelihood ratio. Abbreviations: diagnostic likelihood (DLR).

of the 59 remaining studies were retrieved for complete evaluation and out of those 59, 44 studies were excluded due to irrelevant design ( $n = 11$ ), insufficient data for calculating SE and SP ( $n = 8$ ), no access to full text (5), a publication in languages other than English (6), conference papers (14). Subsequently, the meta-analysis was conducted on 15 articles [2,17–30].

### 3.2. Baseline characteristics

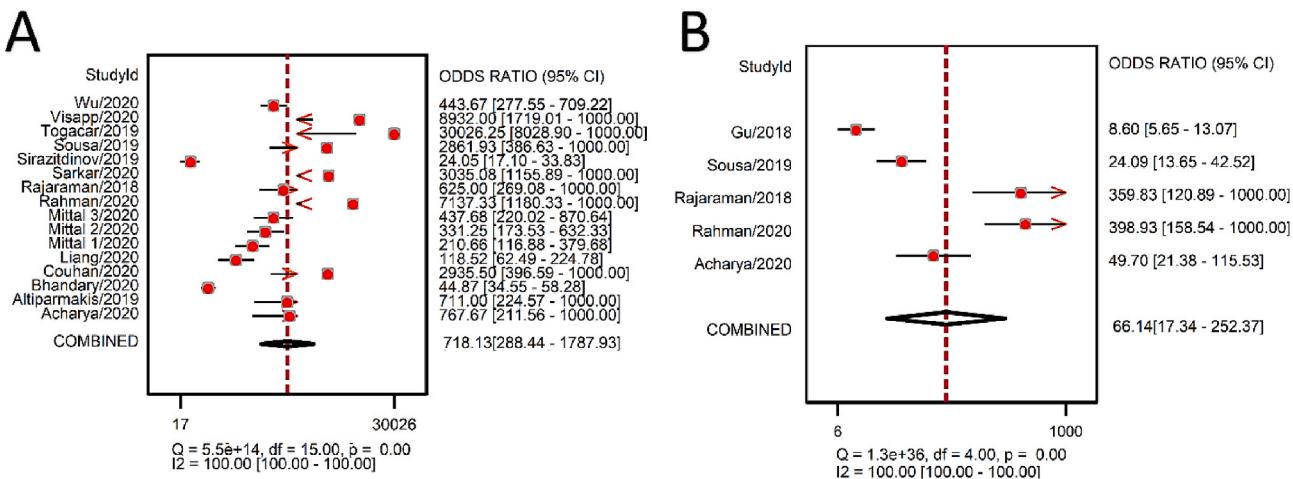
Table 1, presented the characteristics of the included papers. Fourteen studies reported the number of images in test set which consist of 11292 CXR images. Twelve studies separately reported the number of images for pneumonia and controls in test set. In total, 3357 images from patients with pneumonia and 7447 images from healthy controls were included in the test set (repeated images were not removed). Five studies provided data regarding the pathogenic type of pneumonia of which 781 images were from patients with bacterial pneumonia and 597 images were from patients with viral pneumonia.

### 3.3. Overall diagnostic accuracy

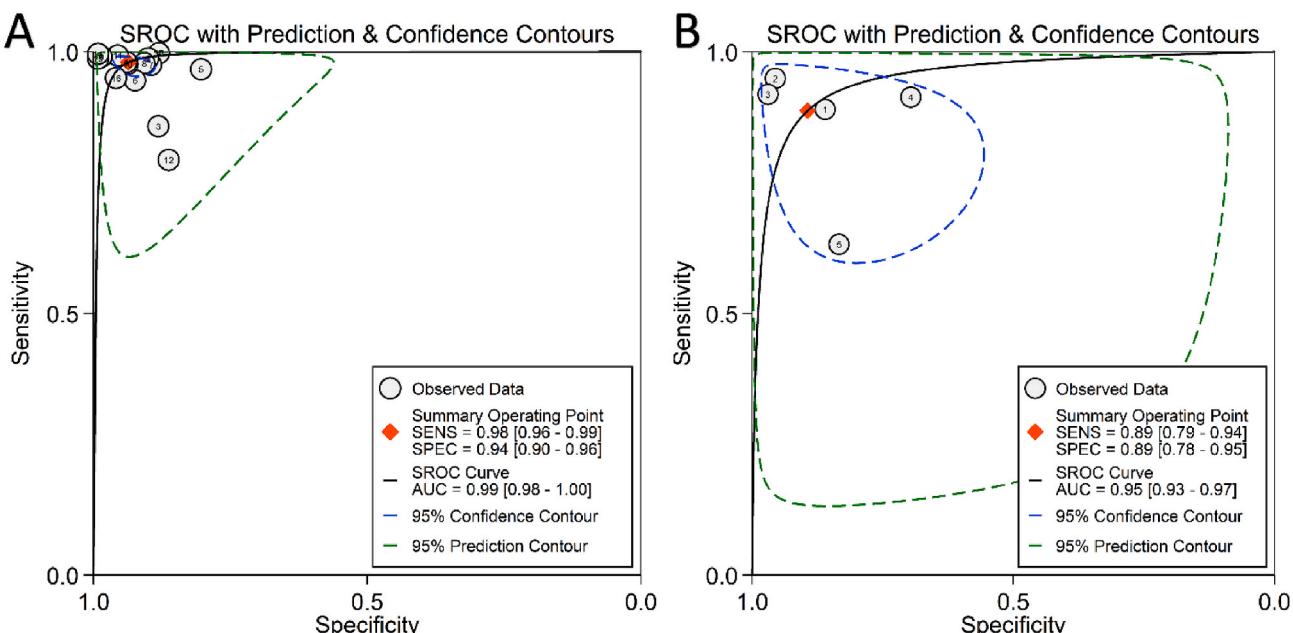
Pooled SE and SP for diagnostic performance of DL in detection of pneumonia using CXR images were 0.98 (95% CI: 0.96–0.99, I<sup>2</sup> statistic 98.29%), and 0.94 (95% CI: 0.90–0.96, I<sup>2</sup> statistic 96.51%) (Fig. 2A and B), respectively. PLR was 15.35 (95% CI: 10.04–23.48, I<sup>2</sup> statistic 96.10%) and NLR was 0.02 (95% CI: 0.01–0.04, I<sup>2</sup> statistic 98.24%) (Fig. 2C and D). The DOR calculated as 718.13 (95% CI: 288.45–1787.93, I<sup>2</sup> statistic 100%) (Fig. 3 A), and the AUC for SROC curves was 0.99 (95% CI: 0.98–1.00) (Fig. 4A).

### 3.4. Diagnostic accuracy of DL in distinguishing bacterial pneumonia from viral pneumonia

The pooled SE and SP for diagnostic performance of DL in distinguishing bacterial pneumonia from viral pneumonia were 0.89 (95% CI: 0.79–0.94, I<sup>2</sup> statistic 97.37%) and 0.89 (95% CI: 0.78–0.95, I<sup>2</sup> statistic 95.47%), respectively (Supplementary Material: Fig. S1A and B). The PLR, NLR and DOR were 8.34 (95% CI: 3.75–18.55, I<sup>2</sup> statistic



**Fig. 3.** Forest plot of diagnostic odds ratio for: A. Diagnostic accuracy of DL in discrimination of pneumonia from normal CXR images B. Diagnostic accuracy of DL in discrimination of bacterial pneumonia from viral pneumonia using CXR images.



**Fig. 4.** The SROC curves for: A. Diagnostic accuracy of DL in discrimination of pneumonia from normal CXR images B. Diagnostic accuracy of DL in discrimination of bacterial pneumonia from viral pneumonia using CXR images. Abbreviations: summary receiver operating characteristic (SROC), sensitivity (SENS), specificity (SPEC), area under the curve (AUC).

94.22%), 0.13 (95% CI: 0.06–0.26, I<sup>2</sup> statistic 97.88%), and 66.14 (95% CI: 17.34–252.37, I<sup>2</sup> statistic 100%), respectively (Supplementary Material: Fig. S1C and D). The AUC for SROC curves was 0.95 (0.93–0.97) (Fig. 4B).

### 3.5. Clinical utility

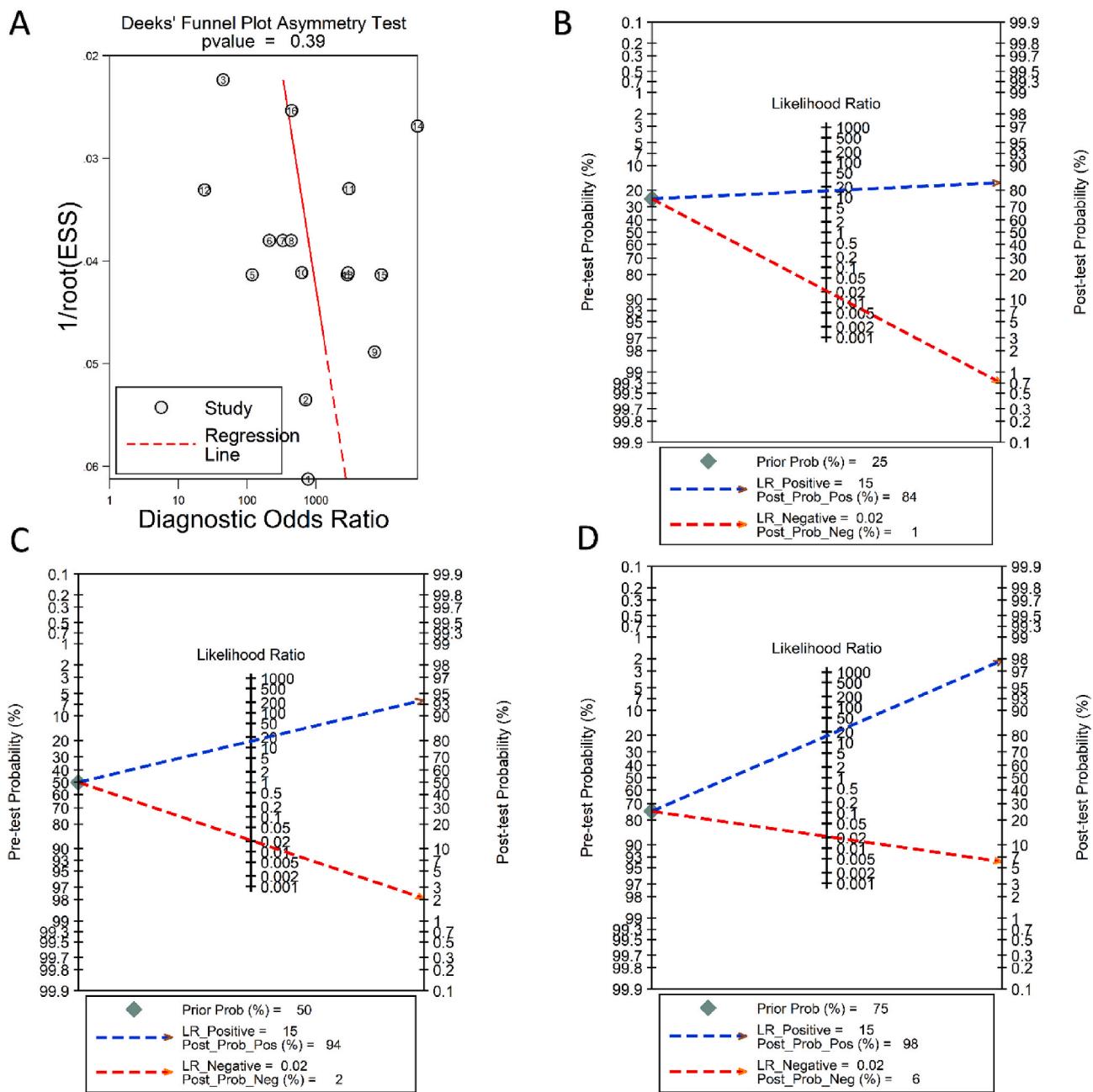
Fagan nomogram analysis showed that with a pretest probability of 25%, DL as an automated diagnostic method, can increase posttest probability of a positive result to 84% while the posttest probability of a negative test was only 1% (Fig. 5B). When setting the pretest probability to 50%, DL increased the probability of a correct detection to 94% while there is only 2% probability of ignoring a pneumonia patient with a negative result (Fig. 5C). Based on 75% pretest probability, DL increased the probability of a correct detection to 98%, while there is a posttest probability of 6% for ignoring a pneumonia patient with a negative result (Fig. 5D).

### 3.6. Publication bias

With a coefficient of 54.09 (95% CI: -76.87175–185.0545, p = 0.39), Deek's funnel plot asymmetry test revealed no publication bias in the included studies (Fig. 5A).

## 4. Discussion

CXR images are widely used for the detection of pneumonia as an easy, accessible and cost effective method. These radiographs provide massive amount of data on clinically valuable signals which are beyond the capability of traditional diagnostic methods to process. Currently, the detection of pneumonia using CXR radiographs necessitate highly experienced physicians since there is overlapping with several other lung abnormalities [19]. Therefore, manual detection of pneumonia can delay the diagnosis and initial treatment process. To the best of our knowledge, in the present study this is the first time estimates of



**Fig. 5.** A. Deeks' funnel plot asymmetry test B. Fagan monogram with 25% pretest probability C. Fagan monogram with 50% pretest probability D. Fagan monogram with 75% pretest probability. Abbreviations: probability (Prob), likelihood ratio (LR).

diagnostic performance of DL have been pooled for classifying pneumonia from normal CXR radiographs.

Pooled SE, SP, PLR, NLR, and DOR were used to evaluate the diagnostic accuracy of DL. With a 0.98 SE and 0.94 SP, DL is highly accurate in classifying pneumonia from normal CXR radiographs. Also, DL showed a high discriminatory performance for distinguishing bacterial pneumonia from viral type (SE: 0.89, SP: 0.89).

Since the likelihood ratios incorporated both SE and SP, they are more clinically applicable. PLR and NLR demonstrate the increase and decrease of the odds of a disease for positive and negative results, respectively [34,35]. Therefore, a greater value of PLR implies a greater probability of TP in the case of a positive test result. In contrast, a numerical value for NLR indicates the probability of the TN when a test result is negative [34–36]. In our study, PLR for DL was 15.35 which suggests that there is a 15–16 fold higher chance of patients with

pneumonia to have DL positive detection as opposed to normal individuals. In contrast, a NLR of 0.02 shows that if there is a negative result based on the DL, the probability of that the patient (based on CXR results) has pneumonia is 2%.

As a graphical tool, Fagan monogram is used to estimate the impact of a diagnostic test on the probability that a patient has a disease [35, 36]. This nomogram has three longitudinal axes. The pretest probabilities are represented on the left axis which is joined to the likelihood ratios on the central longitudinal axis, and the right axis represents posttest probabilities. With a likelihood ratio of 1, both the pre and posttest probabilities would be the same, suggesting that the diagnostic tool is not helpful. A PLR (for positive tests) greater than 1 increases the posttest probability and a NLR (for negative tests) of lesser than 1 decreases the posttest probability [35]. In this study, with pretest probabilities of 25, 50 and 75%, DL increased the probability of a correct

detection to 84, 94, and 98%, respectively. With the same points of pretest probability, the probability that a patient with pneumonia and a negative DL result would be overlooked were only 1, 2, and 6%, respectively. These results suggest the robust performance of DL for classifying pneumonia from normal CXR radiographs.

DOR, as a single indicator of test accuracy, is the ratio of the PLR relative to the NLR, and higher value of DOR is considered as equivalent to higher accuracy of the method [34]. Generally, a DOR value of  $>10$  is considered as a good discriminatory performance [37]. In the present study the DOR for DL in classifying pneumonia from normal CXR radiographs is calculated as 718.13 which highlights the remarkable performance of DL. For distinguishing bacterial from viral pneumonia the DOR was 66.14, which also suggesting a good performance of DL.

The AUC for SROC curves is another estimate for overall diagnostic performance [37] as the following evaluation criteria: low (AUC: 0.5–0.7), moderate (AUC: 0.7–0.9) or high (AUC: 0.9–1.0) accuracy. AUC of 0.99, and 0.95 represented a high overall diagnostic performance of DL for classifying pneumonia from normal CXR radiographs and for discriminating bacterial from viral pneumonia, respectively.

Regardless of the high discriminatory performance of DL, there are major limitations and methodological deficiencies in the enrolled studies including:

- The included articles assessed DL diagnostic accuracy in a manner that does not strictly reflect clinical practice, because was not compared with the performance of health-care professionals. Therefore, this is a major barrier to translate these results to the clinic.
- All studies were *in silico* retrospective with data from assembled datasets which poorly defined the criteria of absence or presence of the disease; this setting is far away from the clinical environment.
- A threshold at which accuracy estimates such as TP, FP, FN, and TN were reported should be specified in each study.
- Systematic quality assessment is not provided as there is not sufficient guidance to apply existing reporting guidelines for machine learning prediction models.
- It was not possible to do in depth analysis of observed significant heterogeneity among studies due to the poor reporting of different covariates in the included studies.

## 5. Conclusion

In this exploratory meta-analysis we showed that DL had high accuracy performance in the classifying of pneumonia from normal CXR radiographs and also in distinguishing bacterial from viral pneumonia however, there are many challenges that should be addressed before the introduction of DL into clinical setting. For the next step, investigators should incorporate such algorithms in a clinical environment. Clinical trials involving AI and DL interventions with rigorous standardized reporting guidelines are the ultimate way to reveal possible clinical implications of an algorithm in real life.

## Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Declaration of competing interest

The authors declare that there are no conflicts of interest.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.combiomed.2020.103898>.

## References

- [1] O. Ruuskanen, E. Lahti, L.C. Jennings, D.R. Murdoch, Viral pneumonia, *Lancet* 377 (2011) 1264–1275.
- [2] A. Mittal, D. Kumar, M. Mittal, T. Saba, I. Abunadi, A. Rehman, S. Roy, Detecting pneumonia using convolutions and dynamic capsule routing for chest X-ray images, *Sensors* 20 (2020) 1068.
- [3] A. Esteve, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, *Nature* 542 (2017) 115–118.
- [4] C. Qin, D. Yao, Y. Shi, Z. Song, Computer-aided detection in chest radiography based on artificial intelligence: a survey, *Biomed. Eng. Online* 17 (2018) 113.
- [5] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.* (2012) 1097–1105.
- [6] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014 arXiv preprint arXiv:1409.1556.
- [7] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
- [8] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
- [9] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
- [10] M. Mostajabi, P. Yadollahpour, G. Shakhnarovich, Feedforward semantic segmentation with zoom-out features, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3376–3385.
- [11] H. Noh, S. Hong, B. Han, Learning deconvolution network for semantic segmentation, *Proc. IEEE. Int. Conf. Comput. Vision.* (2015) 1520–1528.
- [12] M. Bakator, D. Radosav, Deep learning and medical diagnosis: a review of literature, *Multimodal technologies and interaction* 2 (2018) 47.
- [13] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, Chexnet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning, 2017 arXiv preprint arXiv:1711.05225.
- [14] P. Rajpurkar, A.Y. Hannun, M. Haghpanahi, C. Bourn, A.Y. Ng, Cardiologist-level Arrhythmia Detection with Convolutional Neural Networks, 2017 arXiv preprint arXiv:1707.01836.
- [15] V. Gulshan, L. Peng, M. Coram, M.C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, *Jama* 316 (2016) 2402–2410.
- [16] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, P. Group, Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, *PLoS Med.* 6 (2009), e1000097.
- [17] A.K. Acharya, R. Satapathy, A deep learning based approach towards the automatic diagnosis of pneumonia from chest radio-graphs, *Biomed. Pharmacol. J.* 13 (2020) 449–455.
- [18] N. Altiparmakis, Detecting and Understanding Pneumonia with Deep Learning.
- [19] A. Bhandary, G.A. Prabhu, V. Rajinikanth, K.P. Thanaraj, S.C. Satapathy, D. E. Robbins, C. Shasky, Y.-D. Zhang, J.M.R. Tavares, N.S.M. Raja, Deep-learning framework to detect lung abnormality—A study with chest X-Ray and lung CT scan images, *Pattern Recogn. Lett.* 129 (2020) 271–278.
- [20] V. Chouhan, S.K. Singh, A. Khamparia, D. Gupta, P. Tiwari, C. Moreira, R. Damaševičius, V.H.C. de Albuquerque, A novel transfer learning based approach for pneumonia detection in chest X-ray images, *Appl. Sci.* 10 (2020) 559.
- [21] X. Gu, L. Pan, H. Liang, R. Yang, Classification of bacterial and viral childhood pneumonia using deep learning in chest radiography, in: Proceedings of the 3rd International Conference on Multimedia and Image Processing, 2018, pp. 88–93.
- [22] K.T. Islam, S. Wijewickrema, A. Collins, S. O'Leary, A Deep Transfer Learning Framework for Pneumonia Detection from Chest X-Ray Images, 2020.
- [23] G. Liang, L. Zheng, A Transfer Learning Method with Deep Residual Network for Pediatric Pneumonia Diagnosis, *Computer Methods and Programs in Biomedicine*, 104964.
- [24] T. Rahman, M.E. Chowdhury, A. Khandakar, K.R. Islam, K.F. Islam, Z.B. Mahbub, M.A. Kadir, S. Kashem, Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray, *Appl. Sci.* 10 (2020) 3233.
- [25] S. Rajaraman, S. Candemir, I. Kim, G. Thoma, S. Antani, Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs, *Appl. Sci.* 8 (2018) 1715.
- [26] R. Sarkar, A. Hazra, K. Sadhu, P. Ghosh, A novel method for pneumonia diagnosis from chest X-ray images using deep residual learning with separable convolutional networks, in: *Computer Vision and Machine Intelligence in Medical Image Analysis*, Springer, 2020, pp. 1–12.
- [27] I. Sirazitdinov, M. Kholiavchenko, T. Mustafaev, Y. Yixuan, R. Kuleev, B. Ibragimov, Deep neural network ensemble for pneumonia localization from a large-scale chest X-ray database, *Comput. Electr. Eng.* 78 (2019) 388–399.
- [28] G.G.B. Sousa, V.R.M. Fernandes, A.C. de Paiva, Optimized deep learning architecture for the diagnosis of pneumonia through chest X-rays, in: *International Conference on Image Analysis and Recognition*, Springer, 2019, pp. 353–361.
- [29] M. Toğacar, B. Ergen, Z. Cömert, A Deep Feature Learning Model for Pneumonia Detection Applying a Combination of mRMR Feature Selection and Machine Learning Models, *IRBM*, 2019.
- [30] H. Wu, P. Xie, H. Zhang, D. Li, M. Cheng, Predict pneumonia with chest X-ray images based on convolutional deep neural learning networks, *J. Intell. Fuzzy Syst.* (2020) 1–15.

- [31] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R.M. Summers, Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2097–2106.
- [32] D.S. Kermany, M. Goldbaum, W. Cai, C.C. Valentim, H. Liang, S.L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, Identifying medical diagnoses and treatable diseases by image-based deep learning, *Cell* 172 (2018) 1122–1131, e1129.
- [33] D. Kermany, K. Zhang, M. Goldbaum, Labeled Optical Coherence Tomography (Oct) and Chest X-Ray Images for Classification, Mendeley data, 2018, p. 2.
- [34] R.M. Harbord, J.J. Deeks, M. Egger, P. Whiting, J.A. Sterne, A unification of models for meta-analysis of diagnostic accuracy studies, *Biostatistics* 8 (2007) 239–251.
- [35] A.K. Akobeng, Understanding diagnostic tests 2: likelihood ratios, pre-and post-test probabilities and their use in clinical practice, *Acta Paediatr.* 96 (2007) 487–491.
- [36] L. Irwig, P. Macaskill, P. Glasziou, M. Fahey, Meta-analytic methods for diagnostic test accuracy, *J. Clin. Epidemiol.* 48 (1995) 119–130.
- [37] C.M. Rutter, C.A. Gatsonis, A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations, *Stat. Med.* 20 (2001) 2865–2884.