SSA12

Science Session with Keynote: Informatics (Artificial Intelligence in Radiology: Cutting Edge Deep-Learning)

Sunday, Nov. 25 10:45AM - 12:15PM Room: S406B

AI IN

AMA PRA Category 1 Credits ™: 1.50
ARRT Category A+ Credit: 1.75

FDA Discussions may include off-label uses.

Participants

George L. Shih, MD, MS, New York, NY (*Moderator*) Consultant, Image Safely, Inc; Stockholder, Image Safely, Inc; Consultant, MD.ai, Inc; Stockholder, MD.ai, Inc;
An Tang, MD, Montreal, QC (*Moderator*) Research Consultant, Imagia Cybernetics Inc; Speaker, Siemens AG; Speaker, Eli Lilly and Company
Synho Do, PhD, Boston, MA (*Moderator*) Nothing to Disclose

Sub-Events

SSA12-01 Informatics Keynote Speaker: Cutting Edge AI in Radiology

Sunday, Nov. 25 10:45AM - 10:55AM Room: S406B

Participants
An Tang, MD, Montreal, QC (*Presenter*) Research Consultant, Imagia Cybernetics Inc; Speaker, Siemens AG; Speaker, Eli Lilly and Company

For information about this presentation, contact:

an.tang@umontreal.ca

Honored Educators

Presenters or authors on this event have been recognized as RSNA Honored Educators for participating in multiple qualifying educational activities. Honored Educators are invested in furthering the profession of radiology by delivering high-quality educational content in their field of study. Learn how you can become an honored educator by visiting the website at: https://www.rsna.org/Honored-Educator-Award/ An Tang, MD - 2018 Honored Educator

SSA12-02 Predicting Thyroid Nodule Malignancy with Efficient Convolutional Neural Networks

Sunday, Nov. 25 10:55AM - 11:05AM Room: S406B

Awards
Trainee Research Prize - Medical Student

Participants
Ian Pan, MA , Providence, RI (*Presenter*) Nothing to Disclose
Matthew T. Stib, MD, Providence, RI (*Abstract Co-Author*) Nothing to Disclose
William D. Middleton, MD, Saint Louis, MO (*Abstract Co-Author*) Nothing to Disclose
Derek Merck, PhD, Providence, RI (*Abstract Co-Author*) Nothing to Disclose
Michael D. Beland, MD, Providence, RI (*Abstract Co-Author*) Research Grant, Canon Medical Systems Corporation; Consultant, General Electric Company

For information about this presentation, contact:

ian_pan@brown.edu

PURPOSE

Various sonographic features of thyroid nodules have been described, and classification systems (i.e., TI-RADS) have been developed to aid radiologists in determining which suspicious nodules require fine needle aspiration (FNA). This study aims to improve our predictive ability by training convolutional neural networks (CNNs) to discriminate between pathology-confirmed benign and malignant thyroid nodules using ultrasound images.

METHOD AND MATERIALS

Our dataset consisted of 151 malignant and 500 benign thyroid nodules from 571 patients, where each nodule contributed 1 longitudinal and 1 transverse ultrasound view. Preprocessing included cropping the nodule of interest and resizing the image to 224 x 224 pixels. The data were divided into 10 training/validation/test folds following a stratified 80%/10%/10% split with no patient overlap. CNNs based on the MobileNet architecture were initialized with pretrained ImageNet weights. A fully-connected layer was first trained for 10 epochs, and the entire network was fine-tuned for 20 epochs. Data were sampled to achieve 50%/50% class balance for each epoch. Data augmentation probability, dropout probability, and learning rate were tuned via randomized search with 60 iterations. Weights with the highest area under the ROC curve (AUC) during validation were used for testing. A malignancy score is determined for each nodule by averaging the predictions for each view across 3 models.

RESULTS

Our model achieved a mean AUC of 0.863 (95% CI: 0.827, 0.898). The median malignancy scores for benign and malignant nodules were 0.162 and 0.618, respectively. With 5 strata, (0-0.2, 0.2-0.4, 0.4-0.6, 0.6-0.8, 0.8-1.0), 5.94%, 18.2%, 29.5%, 65.7%, and 81.4% of nodules in each respective stratum were malignant, compared to an overall malignancy rate of 23.2%. At a threshold of 0.10, the model reduces the number of negative FNAs by 36% while maintaining 95% sensitivity.

CONCLUSION

CNNs fine-tuned on limited data can accurately predict the malignancy potential of sonographically suspicious thyroid nodules. Larger datasets would likely further improve the performance of our classifier. External validation studies are necessary to verify the generalizability of this approach.

CLINICAL RELEVANCE/APPLICATION

CNN malignancy scores calculated from thyroid ultrasound images can be combined with a radiologist's interpretation for improved stratification of nodules to reduce the number of unnecessary FNAs.

SSA12-03 A Deep-learning Method for Fast Detection of Rib Fracture in CT Images: Effect of Computer-Aided Diagnosis to Radiologists

Sunday, Nov. 25 11:05AM - 11:15AM Room: S406B

Participants
Xiaodong LI, Linyi, China (*Presenter*) Nothing to Disclose

For information about this presentation, contact:

lxd2018rsna@163.com

CONCLUSION

2 reading modes of CAD(CR, SR) can significantly increase the sensitivities of RFD of radiologists. The reading time is shorter by CR than by SR.CR mode can be used as the first option to detect rib fracture by radiologists.

Background

To assess the effect of rib fracture computer-aided diagnosis(CAD) on diagnosis of radiologist.

Evaluation

85 trauma CTs(50 males) with follow-up review CTs were included in the retrospective study.All trauma CTs were subjected to CAD system to generate rib fracture bounding box.The procedure of the CAD system contains ribs segmentation, centerline extraction,rib fracture detection(RFD) based on deep learning algorithm(Faster RCNN),false positive removal and rib fracture localization. 2 senior(NO.1,2) and 2 junior radiologists(NO.3,4) independently evaluated the data using 3 reading modes(without CAD,CR,SR).The fracture line or bone callus growth is the criterion for determining the rib fracture.The follow-up review CTs verified the diagnosis of rib fracture and established the reference standard[1].All fractures detected by the 4 readers were compared to the reference standard. χ^2 test and rank-sum test were performed to test whether there was significant difference between sensitivities and reading times of 3 reading modes.Abbreviations:without CAD:Radiologists independently evaluated the data.CR:Radiologists apply CAD system as a concurrent reader(CR) to evaluate the data.SR:Radiologists evaluate the data as a first reader and then apply the CAD system as a second reader(SR) to review the results.

Discussion

The reference standard identified 281 rib fractures in 85 patients.The sensitivity of RFD with SR 97.2%(273/281;P<0.001) and CR 96.4%(271/281;P<0.001) were significantly higher than that of without CAD 89.7%(252/281).There was no significant sensitivity difference between CR and SR(P>0.3). Senior and junior radiologists used CAD as CR or SR and there was no significant sensitivity difference between 2 modes(P=0.067,P=0.067).Reading time was significantly shorter for CR(98s) compared to that of without CAD(148s;P<0.001) and SR(169s;P<0.001).Reading time of the 3 modes was less in the senior group than in the junior group with significant differences(P<0.001).

SSA12-04 Highly Sensitive Identification and Delineation of Hemorrhagic Stroke Lesion Using Cascaded Deep Learning Model

Sunday, Nov. 25 11:15AM - 11:25AM Room: S406B

Participants
Junghwan Cho, PhD, Lowell, MA (*Presenter*) Nothing to Disclose
Manohar Karki, Lowell, MA (*Abstract Co-Author*) Nothing to Disclose
Eunmi Lee, Lowell, MA (*Abstract Co-Author*) Nothing to Disclose
Jong Kun Kim, MD,PhD, Daegu, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose
Ki-Su Park, MD,PhD, Daegu, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose
Dong Eun Lee, MD, Daegu, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose
Jae Young Choe, MD, Daegu, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose
Sinyoul Park, MD,PhD, Daegu, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose
Suk Hee Lee, MD, Daegu, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose
Myungsoo Kim, MD, Daegu, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose
Jeoungwoo Son, MD, Gimcheon, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose
Jeong Ho Lee, MD,PhD, Daegu, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose

CONCLUSION

A cascaded deep learning model was developed to identify and delineate hemorrhagic stroke lesion, obtaining overall sensitivity accuracy for classification with 97.91% and segmentation with 83.43%, respectively.

Background

Highly accurate and timely detection of intracranial hemorrhagic stroke is a critical clinical issue for diagnosis decision and treatment in emergency room. Deep learning is a promising approach to solve delayed and missed diagnosis of stroke accident. Accordingly, we developed a cascaded deep learning model trained by a series of different CT window settings as a preprocessing step. It consists of two convolution neural networks (CNNs) for identifying bleeding or not and fully convolutional networks (FCNs) for delineating their lesions.

Evaluation

For this study, we acquired 135,000 CT images from 5,650 patients including 3,000 non-bleeding and 2,650 bleeding. In case of bleeding, five subtypes of intracranial hemorrhage (intraventricular, intraparenchymal, subarachnoid, epidural, and subdural hemorrhage) were well labeled by experts. At first, a cascaded deep learning model was trained to identity whether there is bleeding or not and 5-fold cross validation was conducted. We evaluated sensitivity accuracy by the cascaded model, enabling to review the negative case by the second CNN trained with more narrow window width (40/40 [level/width]) in case that CT image is recognized as negative by the first CNN trained with default brain window setting (50/100). It results in increasing around 1% sensitivity (97.91%) while preserving specificity (98.76%). To delineate lesion of bleeding, the FCNs was trained with 33,300 CT slices using DGX-1 system. We achieved overall precision accuracy ranging from 70% to 90% and recall accuracy ranging from 62% to 88% at different Dice coefficient threshold as true positive decision.

Discussion

In diagnostic accuracy, there is a tradeoff between sensitivity and specificity. But while preserving specificity, the cascaded deep learning model can increase sensitivity in diagnosis of hemorrhagic stroke. It has the capability to help doctors inform any suspected cases.

SSA12-05 Differentiation of Hepatic Masses in Abdominal CT Images Using Texture-Aware Convolutional Neural Networks with Texture Image Patches

Sunday, Nov. 25 11:25AM - 11:35AM Room: S406B

Participants
Hansang Lee, Daejeon, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose
Helen Hong, PhD, Seoul, Korea, Republic Of (*Presenter*) Nothing to Disclose
Heejin Bae, Seoul, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose
Sungwon Kim, MD, Seoul, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose
Joonseok Lim, MD, Seoul, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose
Junmo Kim, Seoul, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose

For information about this presentation, contact:

hlhong@swu.ac.kr

hlhong@swu.ac.kr

CONCLUSION

Our method can be applied to the differentiation of various subtypes of hepatic masses including cyst and hemangioma, and early diagnosis of hepatic cancer.

Background

Differentiation of hepatic masses into benign and malignant classes in CT images is an important task for early diagnosis and surgical decision of hepatic cancer. In the cases of small masses, acquisition of intensity and texture features is difficult, making the differentiation challenging. Thus, we propose a deep convolutional neural network (CNN) classification of hepatic masses using texture image patch (TIP) generation to enhance the classification efficiency in small masses.

Evaluation

Our method was evaluated on a dataset consisting of 349 abdominal CT scans including 576 benign and 210 malignant masses. Each mass was manually segmented by the radiologist. In TIP generation, the patches representing only the internal texture of the masses were created by filling the square patch with the segmented mass regions repeatedly. These TIPs have the effect of reflecting the texture information to CNN regardless of the original size of masses. Using these TIPs, the transfer learning (TL) was performed on the ImageNet pre-trained AlexNet to classify the patches into benign or malignant classes. To improve the classification efficiency, we re-trained the random forest (RF) classifier on the deep features extracted from the last feature layer of TL-AlexNet. In experiments, our framework was trained on 390 images(b282, m108), validated on 160 images(b113, m47), and tested on 236 images(b181, m55). The proposed method achieved the accuracy of 87.7% where the comparative methods achieved the accuracies of 83.5%, 80.1%, and 85.2%, without TIP, TL, and RF, respectively.

Discussion

Our TIPs improve the learning efficiency of CNN by augmenting the texture information of small masses and allowing the CNN to focus on the texture information. The TL also plays an important role in learning important imaging features for differentiating the hepatic masses. Instead of obtaining the CNN-classified outputs, re-training the RF classifier on the deep features improves the specificity of the proposed method by enhancing the malignancy detection.

SSA12-06 GrayNet: Medical Generic Image Representations for Deep Learning System of Urinary Stone Detection

Sunday, Nov. 25 11:35AM - 11:45AM Room: S406B

Participants
Hyunkwang Lee, Boston, MA (*Presenter*) Nothing to Disclose
Anushri Parakh, MBBS, MD, Boston, MA (*Abstract Co-Author*) Nothing to Disclose
Sehyo Yune, MD,MPH, Boston, MA (*Abstract Co-Author*) Nothing to Disclose
Jeong Hyun Lee, MD, Seoul, Korea, Republic Of (*Abstract Co-Author*) Nothing to Disclose
Myeongchan Kim, MD, Boston, MA (*Abstract Co-Author*) Nothing to Disclose

Isabelle Hartnett, Boston, MA (*Abstract Co-Author*) Nothing to Disclose
Dushyant V. Sahani, MD, Boston, MA (*Abstract Co-Author*) Research support, General Electric Company Medical Advisory Board, Allena Pharmaceuticals, Inc
Synho Do, PhD, Boston, MA (*Abstract Co-Author*) Nothing to Disclose

For information about this presentation, contact:

hyunkwanglee@seas.harvard.edu

PURPOSE

The performance of deep-learning based image analysis model developed from an institution is not guaranteed to be achieved when deployed in another if the institutions use different imaging systems with varying scan acquisition and reconstruction settings. We established a pretrained model enriched with medical generic image representations extracted from GrayNet, a dataset for human anatomy recognition with 23 labels and evaluated benefits of GrayNet pretrained models for detecting urinary stones.

METHOD AND MATERIALS

GrayNet contains 322 IV contrast-enhanced whole body CT scans with 120,182 axial slices obtained by CT scanners from two manufacturers (171 from GE and 151 from Siemens). The corresponding virtual unenhanced CT images were generated with a customized transform function. All slices were annotated as 23 radiologist-established anatomical labels. We randomly selected 40 cases for validation and the remainings were used for training of a deep convolutional neural network, Inception-v3. The best model, selected based on validation loss, was reserved as a pretrained model for urinary stone detection. Patients who underwent unenhanced CT scans from two manufacturers (GE and Siemens) for suspected urolithiasis were identified and categorized according to presence (n=128) or absence (n=161) of urinary stones, and then split into train, validation, and test subsets. Inception-v3 models initialized with random, ImageNet, and GrayNet pretrained weights were trained on training datasets from a single manufacturer and both. The optimal models were evaluated on test datasets. Area under the ROC curve (AUC) was measured for evaluation metric.

RESULTS

The performance of the GrayNet model trained on the GE dataset showed higher AUC (0.893) than the ImageNet model (0.833) when tested on the Siemens dataset. Similar trend was observed when models trained on the Siemens dataset and tested on the GE dataset (0.917 from GrayNet, 0.854 from ImageNet). When trained on the combined dataset, the GrayNet model obtained higher AUC than those of ImageNet and random models.

CONCLUSION

The GrayNet pretrained weights enabled better generalization performance, compared to the models initialized with ImageNet pretrained and random weights.

CLINICAL RELEVANCE/APPLICATION

The GrayNet pretrained weights enriched with generic medical image representations can be used as a baseline for deep learning systems for a successful deployment in varying settings.

Honored Educators

Presenters or authors on this event have been recognized as RSNA Honored Educators for participating in multiple qualifying educational activities. Honored Educators are invested in furthering the profession of radiology by delivering high-quality educational content in their field of study. Learn how you can become an honored educator by visiting the website at: https://www.rsna.org/Honored-Educator-Award/ Dushyant V. Sahani, MD - 2012 Honored EducatorDushyant V. Sahani, MD - 2015 Honored EducatorDushyant V. Sahani, MD - 2016 Honored EducatorDushyant V. Sahani, MD - 2017 Honored Educator

SSA12-07 Deep Learning of Clinically Relevant Chest X-Ray Findings on the Combination of Three Large Datasets

Sunday, Nov. 25 11:45AM - 11:55AM Room: S406B

Participants
Satyananda Kashyap, PhD, San Jose, CA (*Abstract Co-Author*) Nothing to Disclose
Mehdi Moradi, PhD, San Jose, CA (*Presenter*) Employee, IBM Corporation
Alexandros Karargyris, San Jose, CA (*Abstract Co-Author*) Employee, IBM Corporation
Joy T. Wu, MBChB,MPH, San Jose, CA (*Abstract Co-Author*) Employee, IBM Corporation
Murray A. Reicher, MD, Rancho Santa Fe, CA (*Abstract Co-Author*) Chief Medical Officer, Merge Healthcare Incorporated Board Member, Merge Healthcare Incorporated Co-CEO, Health Companion, Inc Former Chairman, DR Systems, Inc
Tanveer Syeda-Mahmood, PhD, San Jose, CA (*Abstract Co-Author*) Employee, IBM Corporation

For information about this presentation, contact:

mmoradi@us.ibm.com

mmoradi@us.ibm.com

CONCLUSION

We have presented a network trained on the largest collection of chest X-ray images with visually observable radiological findings that performs at similar accuracy to networks developed with 14 NIH labels.

Background

Despite deep learning networks now becoming the de facto method of image classification, their relevance to radiologists is limited by the semantics of label used for training such networks. Recent use of image labels such as pneumonia that are not diagnosable from imaging alone have raised concerns on the utility of the networks. We develop a new classifier for chest X-ray images by training it on labels that derived from visually observable radiological findings. We form a new combined data set of 335,688 images from three sources, namely, PLCO Chest X-rays [Gohagan 2000]), Cancer Screening Trial, National Institute of Health (ChestX-ray14 dataset [Wang 2017]) and the Indiana University dataset [Demner-Fushman 2016]. The 49 original labels assigned to these combined datasets were mapped to the corresponding visually observable findings and regrouped into 20 finding labels. For example, 'consolidation, pneumonia, infiltration, and infiltrate' were all mapped to 'alveolar opacity'. The consolidated dataset with the new labels was used to train a DenseNet121 network architecture [Huang 2016], with 512x512 size input images preprocessed with histogram equalization and intensity normalization.

Evaluation

The dataset was divided to 80% training and 20% validation. The areas under the ROC curve for the 20 findings are: Alveolar opacity: 0.81 , Hernia: 0.84, Pneumothorax: 0.86, Atelectasis: 0.87, Aortic atherosclerosis and/or Carotid artery calcification: 0.90, Bone Lesion: 0.77, Enlarged cardiac silhouette: 0.86, Enlarged Hilum: 0.75, Findings consistent with granulomatous disease: 0.76, Hyperaeration: 0.79, Increased reticular markings: 0.71, Mass and/or Nodule: 0.64, Pleural effusion: 0.92, Pleural mass and/or thickening: 0.71, Spinal degenerative changes: 0.89, Tortuous Aorta: 0.89, Vascular redistribution: 0.85, Catheter and/or Tube: 0.89, Missing plus NA: 0.82, Other: 0.69.

Discussion

Label validation particulalry in 'no finding' labels from NIH dataset is under way.

SSA12-08 Automatic Classification and Reporting of Multiple Common Thorax Diseases Using Chest Radiographs

Sunday, Nov. 25 11:55AM - 12:05PM Room: S406B

Participants
Xiaosong Wang, PhD, Bethesda, MD (*Presenter*) Nothing to Disclose
Yifan Peng, Bethesda, MD (*Abstract Co-Author*) Nothing to Disclose
Le Lu, PhD, Bethesda, MD (*Abstract Co-Author*) Nothing to Disclose
Zhiyong Lu, Bethesda, MD (*Abstract Co-Author*) Nothing to Disclose
Ronald M. Summers, MD,PhD, Bethesda, MD (*Abstract Co-Author*) Royalties, iCAD, Inc; Royalties, Koninklijke Philips NV; Royalties, ScanMed, LLC; Research support, Ping An Insurance Company of China, Ltd; Researcher, Carestream Health, Inc; Research support, NVIDIA Corporation; ; ; ;

For information about this presentation, contact:

xiaosong.wang@live.com

PURPOSE

Chest radiographs are one of the most common radiological exams in daily clinical routines. Reporting thorax diseases using chest radiographs is often an entry-level task for radiologist trainees, but it remains a challenging job for learning-oriented machine intelligence. It's due to the shortage of large-scale well-annotated medical image datasets and lack of techniques that can mimic the high-level reasoning of human radiologists. In this work, we show that clinical free-text radiological reports can be utilized as a priori knowledge for tackling these two difficult problems.

METHOD AND MATERIALS

We used a hospital-scale chest radiograph dataset, which consists of 112,120 frontal-view radiographs of 30,805 patients. 14 disease labels observed in images were mined using natural language processing techniques, i.e., atelectasis, cardiomegaly, effusion, infiltrate, mass, nodule, pneumonia, pneumothorax, consolidation, edema, emphysema, fibrosis, pleural thickening, and hernia. We propose a novel text-image embedding neural network (illustrated in the attached figure) for extracting the distinctive image and text representations. Multilevel attention models are integrated into an end-to-end trainable architecture for highlighting the meaningful text words and image regions. We first apply this combined convolutional and recurrent neural network (CNN-RNN) to classify the image by using both image features and text embeddings from associated reports. Furthermore, we transform the framework into a radiograph reporting system by taking only images as input and turning RNN into a generative model.

RESULTS

The proposed framework achieves high accuracy (0.96±0.03 in AUCs) in disease classification using both images and reports on an unseen and hand-labeled dataset (OpenI, 3,643 images). When using only the images as input, the system can also produce significantly improved results (0.80±0.07 in AUCs) compared to the state-of-the-art (0.74±0.08) with a p-value=0.0005. The figure shows sample classification results with generated reports (attended words in red).

CONCLUSION

We illustrate a framework for fully-automated classification and reporting of common thorax diseases in chest radiographs and demonstrate its superior performance compared to the state-of-the-art.

CLINICAL RELEVANCE/APPLICATION

The proposed multi-purpose CADx system can be applied for automatic classification and reporting of common thoracic diseases as a second opinion.

SSA12-09 Deep-learning for Distal Radius Fracture Detection in X-Ray Imaging

Sunday, Nov. 25 12:05PM - 12:15PM Room: S406B

Participants
Maximilian Russe, MD, Freiburg, Germany (*Presenter*) Nothing to Disclose
Mathias F. Langer, MD, PhD, Freiburg, Germany (*Abstract Co-Author*) Nothing to Disclose
Elmar C. Kotter, MD, MSc, Freiburg, Germany (*Abstract Co-Author*) Editorial Advisory Board, Thieme Medical Publishers, Inc

PURPOSE

Development of a robust algorithm for fracture detection of the distal radius in x-ray imaging for the use in an emergency department.

METHOD AND MATERIALS

Anterior-posterior x-ray images of the wrist from 2012-2017 were classified for fracture or absence of fracture by a consensus

Anterior-posterior x-ray images of the wrist from 2013-2017 were classified for fracture or absence of fracture by a consensus reading from a junior and a senior radiologist. Secondary reading was performed by a certified Msk radiologist. 1900 cases were exported for the deep learning study. Data leakage was excluded using only first-time images of the patients. Images where separated into 1351 images for training and 449 for validation. Besides the validation sample for the CNN learning, a separate analysis of the final model was performed using a separate test sample, containing 50 images with and without fractures. For the development of the deep learning model an established Convulsive neuronal network (CNN), GoogLeNet was used. Due network specifications an images distortion using a smashing transformation to 256*256 pixels was needed. The data was augmented using vertical flipping and up to +/- 10° rotation. No manual segmentation or image correction was made. Deep learning was performed by using Torch on Nvidia DIGITS with a standard workplace graphic unit (Nvidia Quadro P4000). Following parameters where used: 1000 training and validation epochs. AdaGrad was used as solver type and the initial learning rate was 0.01.

RESULTS

The training of the CNN took 4.34h of processing time. Final image processing of all 100 test images took 17 seconds. An overall accuracy of the validation sample was achieved with a final value of 94.2%, the overall accuracy of the separate test sample is 90%. The per-class accuracy in the validation sample was for fractures 87.5% and for no fractures 96.4% in the test sample 86% and 94%. These values are comparable and so overfitting of the CNN can be excluded.

CONCLUSION

The created algorithm shows a good detection rate for distal radius fractures. An exclusion of fracture was performed with even a higher accuracy. These results are promising for preliminary classification of x-rays within a clinical setting.

CLINICAL RELEVANCE/APPLICATION

X-ray reading is still a relevant task for fracture detection, fracture detection algorithms can be used to reduce the work load and already could be used for prioritizing work load. Instant preliminary fracture detection can be achieved with this deep learning model and easily implemented in clinical routine.