Clinical Radiology 77 (2022) e758–e764

Contents lists available at ScienceDirect

Clinical Radiology

journal homepage: www.clinicalradiologyonline.net

Automated detection of enteric tubes misplaced in the respiratory tract on chest radiographs using deep learning with two centre validation

D.H. Mallon [a,b,*], C.D. McNamara [a], G.S. Rahmani [c], D.P. O'Regan [a,b], D.G. Amiras [a]

[a] Imperial College Healthcare NHS Trust, London, UK
[b] MRC London Institute of Medical Sciences, Imperial College London, London, UK
[c] Galway University Hospitals, Galway, Ireland

ARTICLE INFORMATION

Article history:
Received 31 March 2022
Received in revised form
25 May 2022
Accepted 17 June 2022

AIM: To develop and test a model based on a convolutional neural network that can identify enteric tube position accurately on chest radiography.

MATERIALS AND METHODS: The chest radiographs of adult patients were classified by radiologists based on enteric tube position as either critically misplaced (within the respiratory tract) or not critically misplaced (misplaced within the oesophagus or safely positioned below the diaphragm). A deep-learning model based on the 121-layer DenseNet architecture was developed using a training and validation set of 4,693 chest radiographs. The model was evaluated on an external test data set from a separate institution that consisted of 1,514 consecutive radiographs with a real-world incidence of critically misplaced enteric tubes.

RESULTS: The receiver operator characteristic area under the curve was 0.90 and 0.92 for the internal validation and external test data sets, respectively. For the external data set with a prevalence of 4.4% of critically misplaced enteric tubes, the model achieved high accuracy (92%), sensitivity (80%), and specificity (92%) for identifying a critically misplaced enteric tube. The negative predictive value (99%) was higher than the positive predictive value (32%).

CONCLUSION: The present study describes the development and external testing of a model that accurately identifies an enteric tube misplaced within the respiratory tract. This model could help reduce the risk of the catastrophic consequences of feeding through a misplaced enteric tube.

Introduction

Enteric tubes, such as nasogastric and nasojejunal tubes, are commonly used to deliver nutrition, fluid, and medication. Approximately 10% of inpatients, often the most critically unwell, are reliant on an enteric tube for nutrition during their admission. More than 500,000 enteric tubes are inserted per year in the UK alone [1].

* Guarantor and correspondent: D. H. Mallon, Imperial College Healthcare NHS Trust, Charing Cross Hospital, Fulham Palace Road, London, W6 8RF, UK. Tel.: +44 20330 30718.
E-mail address: dmallon@ic.ac.uk (D.H. Mallon).

https://doi.org/10.1016/j.crad.2022.06.011

Following blind insertion, the rate of misplacement into the respiratory tract or the oesophagus has been estimated to be 3% and 20%, respectively [2–4]. Confirmation of correct enteric tube position within the stomach before use is essential. Introduction of fluids via an enteric tube into the airways can have dire consequences, ranging from aspiration pneumonia, respiratory failure, and death [5]. Therefore, feeding through an enteric tube within the respiratory tract, but not within the oesophagus, is classed as a "never event".

If safe enteric tube positioning cannot be confirmed by an acidic aspirate, a chest radiograph, the reference standard, is required [5,6]. The radiograph should be interpreted prior to use by a suitably trained clinician, typically a radiologist [7]. Despite guidelines that describe a robust method to protect against feeding via a misplaced tube and the release of multiple national patient safety alerts from NHS England [8], the rate of feeding through an enteric tube misplaced within the respiratory tract has not reduced. Each year in the UK, between 20 and 30 cases of feeding through a misplaced enteric tube are reported; associated with approximately six deaths [9,10]. In the most recent Never Event Report from NHS England, out of 29 cases of feeding through an enteric tube within the respiratory tract, 15 were due to radiograph misinterpretation and one was due to miscommunication [9]. Therefore, a rapid and automated method of identifying an enteric tube misplaced within the respiratory tract on a chest radiograph would be valuable in the effort to improve patient safety.

In this study, a model based on a convolutional neural network that can accurately identify enteric tube position on chest radiography was developed and tested using an external data set.

Materials and methods

The study was undertaken and reported in line with the RSNA CLAIM checklist [11]. This study was approved at the respective local ethics review committees for both Imperial College Healthcare NHS Trust and University Hospital Galway.

Data set collection and splitting

Data were collected retrospectively with model training and validation performed on posteroanterior (PA) and anteroposterior (AP) chest radiographs retrieved from Imperial College Healthcare NHS Trust. The model was tested on radiographs retrieved from University Hospital Galway. All chest radiographs were acquired as part of routine clinical care between 2011 and 2019 on patients above the age of 18 years. A flow chart of radiographs included in the study is shown in Fig 2.

As critically misplaced enteric tubes are a relatively rare event, to reduce class imbalance during training, a data set that was enriched for both correctly placed and misplaced enteric was identified by searching radiology reports for the terms "nasogastric", "nasojejunal", "NGT", NJ", "NJT", "Ryles tube", "feeding tube", or "enteric tube", and "malpositioned", "misplaced", "unsafe", "bronchus", "bronchial", or "trachea". An approximate ratio of 1:4 for critically misplaced (enteric tube within the airway) and non-critically misplaced (enteric tube not within the airway) enteric tubes was obtained. A list of all radiographs in chronological order was compiled. The radiographs that were likely to show a misplaced radiograph were identified (by searching for the above key phrases in the radiology report). Duplicates were removed so that a maximum of one radiograph per patient was included. Finally, in this list of radiographs (where there was only one radiograph per patient) the two subsequent and two prior radiographs were retained.

After annotation (described below), 50 radiographs with a critically misplaced and 50 radiographs without a critically misplaced radiograph were selected randomly to give a validation data set consisting of 100 radiographs. Radiographs were extracted and anonymised from PACS using *Radiomics Enabler* (version v1.5.5.1).

The external testing data set from Galway University Hospitals was extracted and anonymised within PACS (AGFA HealthCare Enterprise Imaging). The test data set comprised consecutive radiographs acquired to check the position of an enteric tube that were acquired consecutively between 1 June 2019 and 31 May 2020. Consecutive chest radiographs were obtained to ensure a realistic class balance. As for the training and validation test data set, a maximum of one radiograph per patient was included in the external test data set.

Annotation

A critically misplaced tube was defined as a tube projected over the trachea or bronchi (which may include cases where the tube was in fact within the proximal oesophagus but appears the same as a tracheal tube position on a single projection). Where this condition was not met, the enteric tube was considered to be in either a safe position (tube bisecting the carina with the radiopaque tip below the left hemidiaphragm) or in an unsafe, non-critical, position (most commonly within the oesophagus).

Enteric tube position for all chest radiographs was classified as either "safe", "unsafe non-critical", or "unsafe critical" by at least three radiologists (D.H.M., C.D.M. and G.S.R. with 6, 4, and 4 years of experience, respectively). The model was trained to differentiate between enteric tubes that were critically misplaced ("unsafe critical") and those that were not critically misplaced (i.e., "unsafe non-critical" or "safe"). Where there was a discrepancy, radiographs were reviewed with a fourth radiologist (D.G.A. with 13 years of experience) for a consensus to be reached. If a consensus could not be achieved, the radiograph was excluded. Radiographs were excluded if no nasogastric tube was included in the radiograph, in cases of poor technical quality (significant rotation, underexposure, or overexposure), excessive overlying apparatus (such as electrocardiogram [ECG] leads and cardiopulmonary resuscitation electrodes), failure to include the majority of the lungs, and excessively large borders created by shutters. Examples are provided in Electronic Supplementary Material Fig. S1. No radiographs were excluded from the test data set.

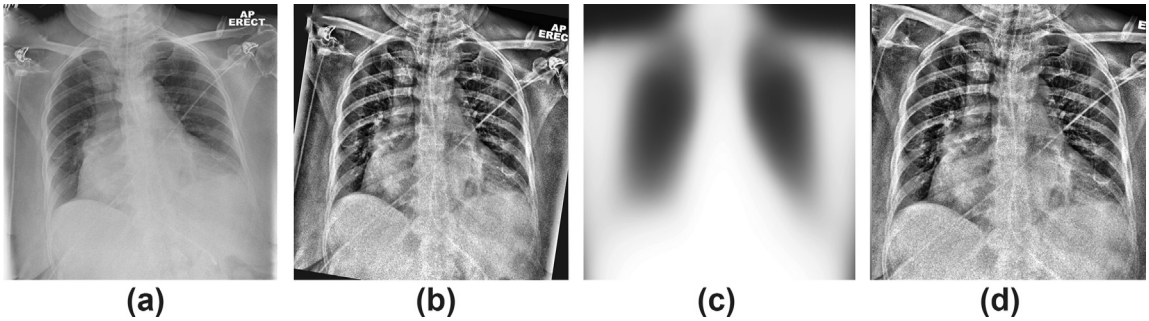

Figure 1 (a) Schematic of image pre-processing steps reduce technical variation between radiographs. (b) Contrast Limited Adaptive Histogram Equalisation was used to equalise regional contrast of the original chest radiograph. (c) Images were then rotated and resized to best align with a "template" radiograph. (d) Example of the final pre-processed image.

Image pre-processing

All code for image processing and model development is available at https://doi.org/10.5281/zenodo.6569554. All data processing and model generation was performed in *Python* (version 3.6). DICOMs were converted to *NumPy* (version 1.19.1) arrays at a resolution of 299 × 299 × 1. Image contrast was enhanced using the Contrast Limited Adaptive Histogram Equalization (CLAHE) method as implemented in *sci-kit learn* (version 0.23.2) with a clipping limit of 0.02 and a kernel size of 23. Radiographs were aligned to a "template". The template was calculated based on the average pixel values from 3,256 radiographs obtained from the National Institutes of Health Chest X-Ray Dataset [12]. Alignment was performed using affine transformation as implemented in the *findTransformECC* method in the *OpenCV* (version 4.4.0.42) library. Pre-processing steps are outlined in Fig 1.

Data augmentation

Images were augmented using geometric (translation <10%, cropping <10%, shearing <10°, rotation 10°), and photometric (brightness adjustments <10%), contrast rescaling (<10%), as implemented in the *exposure.rescale_intensity* method in scikit-learn, and the addition of random Gaussian noise <2.5%) transformations.

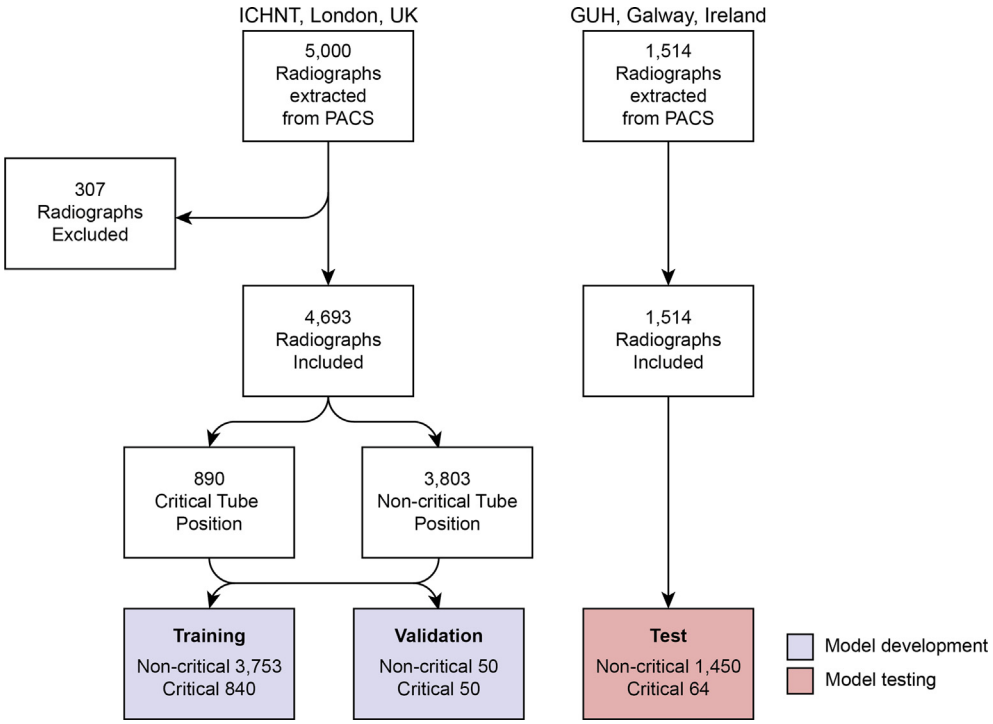

Figure 2 Flow chart of radiographs included in the training, validation, and test data sets. The model was developed using a data from Imperial College Healthcare NHS Trust. External testing was performed on an independent data from University Hospital Galway. "Critical tube position" refers to enteric tubes within the airways. "Non-critical tube position" refers to enteric tubes safely inserted within the stomach or unsafely positioned within the oesophagus.

Model development

The model was developed using the *TensorFlow* framework (version 1.15.2). Model input was a 299 × 299 × 1 image. Model output was a binary classifier representing either a "critical" enteric tube position or a "non-critical" enteric tube position.

Models based on the 121-layer DenseNet were trained using random weights initialisation and pre-trained weights (transfer learning from CheXNet [13] or ImageNet [14]). Where pre-trained weights were used, the top layer (containing 1,000 and 14 classification nodes for ImageNet and CheXNet, respectively) were replaced with two dense fully connected layers.

A random hyperparameter search was performed for the following model characteristics: input resolution, optimiser, fully connected layer nodes, dropout rate, learning rate, class weighting, batch size, data augmentation, random or pre-train weights initialisation and number of trainable layers (when using pre-trained weights). Full details on the hyperparameters considered are provided in Electronic Supplementary Material Table S1. Training continued for a maximum of 25 epochs or until validation loss did not decrease by more than 5% for more than 5 epochs.

Model selection was based on binary cross entropy loss in the validation data set. After the final model was selected, it was "locked down" (7 October 2020) such that no further modifications were made. The final model was assessed only once on the external validation set. The model was developed on a HP Z8 workstation with dual Intel Xeon 4112 CPUs, 128 GB of RAM, and an NVIDIA TITAN RTX.

Statistics

Patient age and sex between the training, validation and external test data sets were compared using the *t*-test and the Fisher's exact test, respectively, using the *statsmodel* package (version 0.11.1). A *p*-value of <0.05 was considered significant.

Accuracy, sensitivity, specificity, positive predictive value, and negative predictive values were calculated. SoftMax function as implemented in SciPy (version 1.5.2) was used to calculate the area under the receiver operator characteristic (ROC) curve (AUC). Predictions were based on a SoftMax cut-off of 0.5. Confidence intervals (CI) for the ROC curves were estimated using bootstrapping of 1,000 random samples.

Failure analysis was performed using gradient weighted class activation maps (CAM), which highlight regions within the input with the greatest contribution to the predicted class, were produced using TensorFlow as described previously [15].

Results

A total of 5,000 radiographs from unique patients were extracted from the PACS at Imperial College Healthcare NHS Trust. Three hundred and seven radiographs were excluded due to the absence of an enteric tube (n=211), due to technical factors (n=50, see Electronic Supplementary Material Fig. S1), or failure to reach consensus on enteric tube position (n=46). Out of a total of 4,693 radiographs, 890 and 3,803 radiographs were identified with either critical or non-critical enteric tube positions, respectively. For testing, 1,514 radiographs from unique patients were extracted from University Hospital Galway. No radiographs were excluded from the test data set. Details of the training (Imperial College Healthcare NHS Trust), validation (Imperial College Healthcare NHS Trust), and external test (University Hospital Galway) cohorts are given in Fig 1.

There was no statistically significant difference between the age of the patients in the training, validation, and external test sets. In all three sets, men were more common than women. The proportion of men in the training set (59%) was statistically significantly lower than both the validation set (69%, p=0.04) and the test set (p=0.002; Table 1).

After a random hyperparameter search, the lowest validation loss was observed after 16 epochs in a model using the 121-layer DenseNet model architecture within ImageNet pre-trained weights, training of weights within the last 14 layers, stochastic gradient descent optimiser with a learning rate of 0.001, batch size of 32, two fully connected layers with 32 and 16 nodes, dropout of 40% and L1 regularisation factor of 0.001. The training lasted for approximately 100 seconds per epoch. Inference took less than 1 second per radiograph.

The ROC AUC for the training, validation, and external test cohorts was 0.92, 0.90, and 0.92, respectively. There was significant overlap in the CIs for the ROC AUC for the training, validation, and test cohorts indicating model performance was similar (Table 2). ROC curves are shown in Fig 3. The model achieved a binary accuracy of 85% and 92% in the internal validation set and the external test set, respectively. For the external test set, the sensitivity, specificity, positive predictive value, and negative predictive value were 80%, 92%, 32%, and 99%. Diagnostic statistics are given in Table 2. A confusion matrix for the external test set is given in Fig 4.

Electronic Supplementary Material Fig. S2 shows failure analysis using CAMs for individual radiographs in correct and incorrect predictions. Electronic Supplementary Material Fig. S3 shows the summation of all CAMs in the validation and training data sets for each class. The non-critical class showed high activation values to be concentrated within the inferior mediastinum and the left upper quadrant of the abdomen. Conversely, the critical class showed activation mainly within both the right and left lower lung zones.

Table 1
Demographics of the patients included in training, validation, and test cohorts.

	Training (n=4,693)	Validation (n=100)	Test (n=1,514)
Age (IQR)	68 (53–78)	69 (56–75)	67 (60–77)
Sex (M:F)	2732:1911 (59% male)	69:31 (69% male)	969:545 (64% male)

Table 2
Diagnostic statistics of the model for the training (Imperial College Healthcare NHS Trust), validation (Imperial College Healthcare NHS Trust), and test cohort (University Hospital Galway).

	Training	Validation	Test
AUC	0.92 (0.91–0.93)	0.90 (0.87–0.93)	0.92 (0.90–0.94)
Sensitivity	0.85 (0.84–0.86)	0.90 (0.76–0.96)	0.80 (0.67–0.88)
Specificity	0.82 (0.81–0.83)	0.80 (0.69–0.91)	0.92 (0.91–0.94)
Positive predictive value	-	-	0.32 (0.27–0.36)
Negative predictive value	-	-	0.99 (0.98–0.99)
Accuracy	0.84 (0.83–0.85)	0.85 (0.76–0.91)	0.92 (0.90–0.93)

Ninety-five per cent confidence intervals are included in parentheses. Positive and negative predictive values were not obtained for the training and validation data sets as they were enriched with critically misplaced enteric tubes.
ROC AUC, area under the receiver operator characteristic curve.

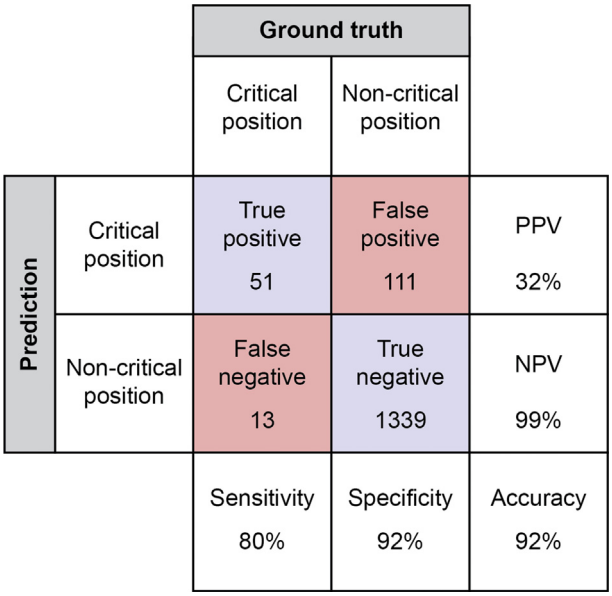

Figure 4 Confusion matrix for external test data set. A positive test refers to a critically misplaced enteric tube.

Discussion

Feeding through an enteric tube misplaced in the respiratory tract can have devastating consequences. The misinterpretation of enteric tube position on chest radiographs is the most common cause of feeding through an enteric tube misplaced within the respiratory tract [6]. Therefore, an automated method to interpret enteric tube position on chest radiographs, incorporated alongside other safety protocols, may reduce "never events" associated with chest radiograph misinterpretation. The present study describes, for the first time in the literature, the development and the external validation in a dataset of consecutive radiographs of an algorithm that can accurately identify critically misplaced enteric tubes.

Multiple studies have demonstrated the value of deep learning in the assessment of the position of in-dwelling tubes and catheters [16,17]. Singh *et al.* achieved an AUC of 0.87 in identifying critical enteric tube position (using a cohort consisting of 174 critically positioned enteric tubes) [18]. Crucially, however, their model performance was not assessed on an independent external data set, which is increasingly considered to be a critical part of algorithm assessment [19–23].

The present study describes the development and external testing of an algorithm that accurately identifies enteric tubes misplaced within the respiratory tract on chest radiography. The model achieved an AUC of 0.90 in the internal validation set and 0.92 in the external test set. In the external dataset, the algorithm was more specific (92%) than it was sensitive (80%) and had a greater NPV (92%) than PPV (32%).

Notably, while a drop-off in performance is expected in external data sets due to over-fitting, the binary accuracy and the ROC AUC were both higher in the external data set than the internal validation data set. This increase in model performance may reflect the much lower incidence of critically misplaced enteric tubes in the external test data set in the context of a model that can more confidently identify safely positioned enteric tubes (which would be

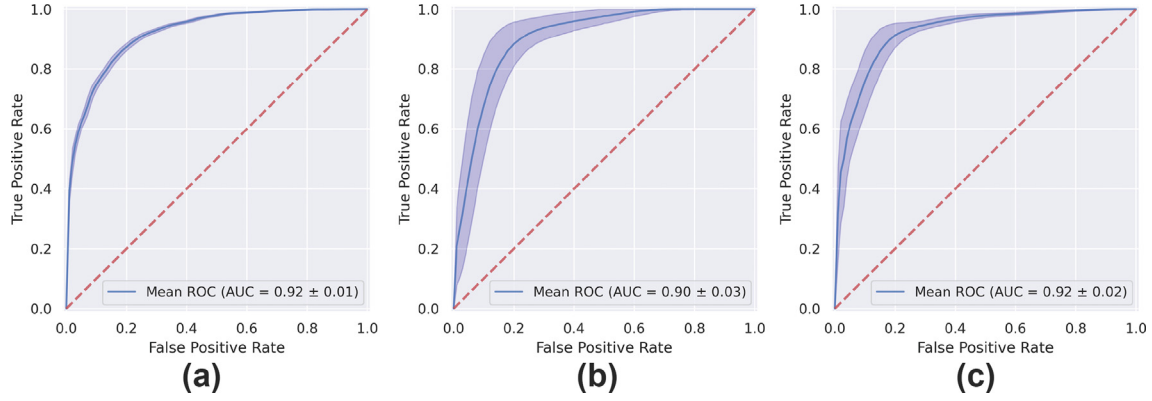

Figure 3 Receiver operator characteristic (ROC) curves with 95% confidence intervals for (a) training, (b) validation, and (c) test data sets showing a ROC area under the curve of 0.92, 0.90, and 0.92, respectively.

expected as safely positioned nasogastric tubes made up approximately 75% of the training data set). As there was no significant drop-off in model performance on the external test set, this suggests that the model generalises well to independent data. The model's generalisability is likely to have been aided by the multiple pre-processing and data augmentation steps, such as contrast equalisation and re-alignment, which reduces the effect of technical variation caused by differences in the radiograph acquisition using different hardware.

The 121-layer DenseNet architecture was used as there are multiple examples in the literature of its successful implementation in chest radiograph interpretation [13,18,24–26]. Perhaps unsurprisingly, training a naive model caused unavoidable model overfitting as the model, consisting of approximately 7 million parameters, is capable of effectively memorising the training data set. Therefore, transfer learning using pre-trained weights from ImageNet and CheXNet were used. Interestingly, pre-trained weights from ImageNet (trained to classify everyday objects such as food items, cats, and airplanes) outperformed pre-trained weights from CheXNet (used to classify disease on chest radiographs). A possible explanation for this is that sharp edges, likely to be a feature required for the identification of an enteric tube, are well represented within the ImageNet, but not the CheXNet, pre-trained weights. This observation highlights how the nature of both the input images and the task to be performed should be considered in the choice of pre-trained weights to fully exploit the benefits of transfer learning.

Failure analysis using CAMs highlighted shortcomings in the model. False positives were caused by sharp edges within the lower zones, for example, due to ECG leads and endobronchial barium (Electronic Supplementary Material Fig. S2). False negatives occurred when multiple tubes were included in the radiograph. For example, a radiograph of a patient with two enteric tubes, one safely positioned and one misplaced in the left main bronchus, was falsely classified as safe. This error may have arisen because the algorithm more confidently identifies, and therefore prioritises, safely positioned enteric tubes over critically misplaced enteric tubes, reflecting the class imbalance in the training set. Although the CAMs highlight regions surrounding the enteric tube in the vast majority of cases, there were multiple cases where regions unrelated to the enteric tube were highlighted. This may reflect either the inherent limitations of CAMs (which is a 2D representation of a multi-dimensional neural network) or that in certain cases, the model is activated by irrelevant features when making predictions. Deep-learning techniques that analyse the entire image may therefore classify the radiograph using irrelevant image features. This drawback could be avoided by segmentation of the enteric tube prior to classification; however, while this method has been attempted with other catheters [27], it adds further complexity to the algorithm and is therefore a source of further error. As such, segmentation of catheters or tubes has not been shown to be more accurate than analysis of the whole image.

The continuous variable output of a neural network can be considered a measure of the confidence in the prediction. Therefore, an optimal cut-off can be determined based on the consequences of either a false negative or a false positive result. Arguably, the potentially dire complications associated with a misplaced enteric tube would warrant the loss of specificity (and accuracy) in favour of improved sensitivity. The algorithm could operate cooperatively with clinicians; the prediction of a misplaced enteric tube with a high level of confidence could issue an entry into the patient's electronic health record, prohibit the prescription of enteral feeding, and automatically alert the clinical team. The prediction of a safely positioned enteric tube with a high level of confidence could permit an appropriately qualified clinician to "clear" the enteric tube for use without the input of a radiologist before initiating feeding. Predictions made with a lower level of confidence could simply flag the radiograph for expedited review by a radiologist.

There are several limitations of this study that should be considered. Firstly, while performance has been tested using an external data set consisting of consecutive radiographs with a real-world incidence of critically misplacement of the enteric tube, prospective evaluation is the optimum method of determining model performance. Secondly, while the model achieves a high level of accuracy, false negatives (13/64, 20%) are observed, and therefore, the model would not operate independently; radiologist review would still be necessary. Notably, it would be possible to reduce the number of false negatives by changing the classification cut-off to favour sensitivity over specificity. Thirdly, enteric tubes are more commonly misplaced in the right main bronchus than the left main bronchus, trachea, or oropharynx. Therefore, it is possible that enteric tubes placed in these locations will be less reliably identified than those in the right main bronchus. Fourthly, enteric tubes have either a radiopaque tip ($\sim$2 cm) or a full-length radiopaque guidewire. Identification of an enteric tube with only a radiopaque tip, safely positioned or otherwise, is a more challenging task. It is therefore likely that the classification accuracy of these enteric tubes is overestimated. Fifthly, as detailed above, in a method that does not segment that path of the enteric tube, the algorithm is likely to classify radiographs using information that is irrelevant to enteric tube placement, e.g., lower lobe collapse. Lastly, neonates and children were not included in this study due to insufficient data. One study has shown that the rate of enteric tube misplacement is higher in the neonatal and paediatric population than in adults [28]. Although it is possible that this model can identify misplaced enteric tubes in the paediatric population, this has not been formally assessed.

In conclusion, the present study describes the development and external testing of a model that can identify a misplaced enteric tube rapidly and accurately on a chest radiograph. Clinical implementation of this algorithm, in conjunction with all other safety measures for ensuring correct enteric tube positioning, could help reduce the risk of patients suffering from the catastrophic consequences of

feeding through an enteric tube misplaced within the respiratory tract.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

D.H.M. is supported by the Imperial Health Charity and National Institute for Health Research Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. D.P.O'R. was supported by the Medical Research Council (MC-A651-53301) and the National Institute for Health Research Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. Code is available at: https://www.doi.org/10.5281/zenodo.4298955.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crad.2022.06.011.

References

1. Coombes R. NHS safety agency issues guidance on nasogastric tubes. *BMJ* 2005;**330**:438. https://doi.org/10.1136/bmj.330.7489.438.
2. Hanna G, Investigator P. *Improving the safety of nasogastric feeding tube insertion position—a decision analysis approach. A report for the NHS patient safety Research portfolio*. London: Imperial College London; July 2010.
3. Marderstein EL, Simmons RL, Ochoa JB. Patient safety: effect of institutional protocols on adverse events related to feeding tube placement in the critically ill. *J Am Coll Surg* 2004 Jul;**199**(1):39–47.
4. Huffman S, Pieper P, Jarczyk KS, *et al*. Methods to confirm feeding tube placement: application of research in practice. *Pediatr Nurs* 2004 Jan-Feb;**30**(1):10–3. Erratum in: Pediatr Nurs. 2004 Nov-Dec;30(6):458.
5. Smith AL, Santa Ana CA, Fordtran JS, *et al*. Deaths associated with insertion of nasogastric tubes for enteral nutrition in the medical intensive care unit: clinical and autopsy findings. *Proc (Bayl Univ Med Cent)*. 2018 May 21;**31**(3):310–6.
6. Metheny NA, Krieger MM, Healey F, Meert KL. A review of guidelines to distinguish between gastric and pulmonary placement of nasogastric tubes. *Heart and Lung* 2019;**48**(3):226–35. https://doi.org/10.1016/j.hrtlng.2019.01.003.
7. Jones B. *A position paper on nasogastric tube safety. "Time to put patient safety first*. British Association for Parenteral and Enteral Nutrition; 2020 (September (accessed 22 Jan 2022), https://www.bapen.org.uk/pdfs/ngsig/a-position-paper-on-nasogastric-tube-safety.pdf.
8. Nhs England. Provisional publication of Never Events reported as occurring between 1 April 2020 and 28 February 2021. 8 April 2021 (accessed 22 Jan 2022), https://www.england.nhs.uk/wp-content/uploads/2021/04/Provisional-publication-NE-1-April-28-February-2021.pdf.
9. Nhs England. *Nasogastric tube misplacement: continuing risk of death and severe harm*. 2016 (accessed 22 Jan 2022), https://www.england.nhs.uk/publication/patient-safety-alert-nasogastric-tube-misplacement-continuing-risk-of-death-and-severe-harm/.
10. Nhs England. Never Events reported as occurring between 1 April 2019 and 31 March 2020–final update (accessed 22 Jan 2022), https://www.england.nhs.uk/wp-content/uploads/2021/01/Final-update-NE-1-April-2019-31-March-2020.pdf.
11. Mongan J, Moy L, Kahn CE. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell* 2020;**2**:e200029, https://doi.org/10.1148/ryai.2020200029.
12. Wang X, Peng Y, Lu L, *et al*. *ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the 30th IEEE conference on computer vision and pattern recognition, CVPR 2017*. Piscataway, NJ: IEEE; 2017. p. 3462–71. https://doi.org/10.1109/CVPR.2017.369.
13. Rajpurkar P, Irvin J, Zhu K, *et al*. CheXNet: radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv* 2017:3–9. https://doi.org/10.48550/arXiv.1711.05225.
14. Huang G, Liu Z, van der Maaten L, *et al*. Densely connected convolutional networks. In: *Proceedings of the 30th IEEE conference on computer vision and pattern recognition, CVPR 2017*. Piscataway, NJ: IEEE; 2017. p. 2261–9.
15. Zhou B, Khosla A, Lapedriza A, *et al*. Learning deep features for discriminative localization. In: *Proceedings of the IEEE computer society conference on computer vision and pattern recognition 2015*. Piscataway, NJ: IEEE; 2016. p. 2921–9.
16. Ramakrishna B, Brown M, Goldin J, *et al*. An improved automatic computer aided tube detection and labeling system on chest radiographs. In: van Ginneken B, Novak CL, editors. *Med Imaging 2012: Computer-Aided Diagn*, **8315**. SPIE; 2012. 83150R. https://doi.org/10.1117/12.911839.
17. Lakhani P. Deep convolutional neural networks for endotracheal tube position and X-ray image classification: challenges and opportunities. *J Digit Imaging* 2017;**30**(4):460–8. https://doi.org/10.1007/s10278-017-9980-7.
18. Singh V, Danda V, Gorniak R, *et al*. Assessment of critical feeding tube malpositions on radiographs using deep learning. *J Digit Imaging* 2019;**32**(4):651–5. https://doi.org/10.1007/s10278-019-00229-9.
19. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal–external, and external validation. *J Clin Epidemiol* 2016;**69**:245–7. https://doi.org/10.1016/j.jclinepi.2015.04.005.
20. West H, Leach JL, Jones BV, *et al*. Clinical validation of synthetic brain MRI in children: initial experience. *Neuroradiology* 2017 Jan;**59**(1):43–50. https://doi.org/10.1007/s00234-016-1765-z.
21. Mårtensson G, Ferreira D, Granberg T, *et al*. The reliability of a deep learning model in clinical out-of-distribution MRI data: a multicohort study. *Med Image Anal* 2020;**66**:101714, https://doi.org/10.1016/j.media.2020.101714.
22. Zech JR, Badgeley MA, Liu M, *et al*. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS Med* 2018 Nov 6;**15**(11):e1002683, https://doi.org/10.1371/journal.pmed.1002683.
23. Aggarwal R, Sounderajah V, Martin G, *et al*. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ Digit Med* 2021;**4**(1):65. https://doi.org/10.1038/s41746-021-00438-z.
24. Nash M, Kadavigere R, Andrade J, *et al*. Deep learning, computer-aided radiography reading for tuberculosis: a diagnostic accuracy study from a tertiary hospital in India. *Sci Rep* 2020;**10**(1):1–10. https://doi.org/10.1038/s41598-019-56589-3.
25. Kim DHH, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. *Clin Radiol* 2018;**73**(5):439–45. https://doi.org/10.1016/j.crad.2017.11.015.
26. Dunnmon JA, Yi D, Langlotz CP, *et al*. Assessment of convolutional neural networks for automated classification of chest radiographs. *Radiology* 2019;**290**(2):537–44. https://doi.org/10.1148/radiol.2018181422.
27. Yi X, Adams SJ, Henderson RDE, *et al*. Computer-aided assessment of catheters and tubes on radiographs: how good is artificial intelligence for assessment? *Radiol Artif Intell* 2020;**2**(1):e190082, https://doi.org/10.1148/ryai.2020190082.
28. Quandt D, Schraner T, Ulrich Bucher H, *et al*. Malposition of feeding tubes in neonates: is it an issue? *J Pediatr Gastroenterol Nutrt* 2009;**48**(5):608–11. https://doi.org/10.1097/MPG.0b013e31818c52a8.