SCIENTIFIC DATA

Check for updates

OPEN

DATA DESCRIPTOR

Creation and validation of a chest X-ray dataset with eye-tracking and report dictation for AI development

Alexandros Karargyris [1 ✉], Satyananda Kashyap[1,4 ✉], Ismini Lourentzou[1,2,4], Joy T. Wu[1,4], Arjun Sharma[1], Matthew Tong[1], Shafiq Abedin[1], David Beymer[1], Vandana Mukherjee[1], Elizabeth A. Krupinski [3] & Mehdi Moradi[1 ✉]

We developed a rich dataset of Chest X-Ray (CXR) images to assist investigators in artificial intelligence. The data were collected using an eye-tracking system while a radiologist reviewed and reported on 1,083 CXR images. The dataset contains the following aligned data: CXR image, transcribed radiology report text, radiologist's dictation audio and eye gaze coordinates data. We hope this dataset can contribute to various areas of research particularly towards explainable and multimodal deep learning/machine learning methods. Furthermore, investigators in disease classification and localization, automated radiology report generation, and human-machine interaction can benefit from these data. We report deep learning experiments that utilize the attention maps produced by the eye gaze dataset to show the potential utility of this dataset.

Background & Summary

In recent years, artificial intelligence (AI) has been extensively explored for enhancing the efficacy and efficiency of the radiology interpretation and reporting process. As the current prevalent paradigm of AI is deep learning, many of the works in AI for radiology use large data sets of labeled radiology images to train deep neural networks to classify images according to disease classes. Given the high labor cost of annotating images with the areas depicting the disease, large public training datasets often come with global labels describing the whole image[1,2] without localized annotation of the disease areas. The deep neural network model is trusted with discovering the relevant part of the image and learning the features characterizing the disease. This limits the performance of the resulting network. Furthermore, the black-box nature of deep neural networks and lack of local annotations means that the process of developing disease classifiers does not take advantage of expert's knowledge of disease appearance and location in medical images. The result is a multi-layer and nonlinear model with serious concerns with respect to the explainability of its output. Another well-studied concern is the generalization capability (i.e., when the model is deployed to infer the class labels for images from other sources or distributions) and how deep neural networks are affected by scanner differences and/or demographic changes[3].

In the past five decades eye-tracking has been extensively used in radiology for education, perception understanding, and fatigue measurement (example reviews[4–7]:). More recently, efforts[8–11] have used eye-tracking data to improve segmentation and disease classification in Computed Tomography (CT) by integrating them in deep learning techniques. With such evidence and with the lack of public datasets that capture eye gaze data in the chest X-Ray (CXR) space, we present a new dataset that can help improve the way machine learning models are developed for radiology applications and we demonstrate its use in some popular deep learning architectures.

This dataset consists of eye gaze information recorded from a single radiologist interpreting frontal chest radiographs. Dictation data (audio and timestamped text) of the radiology report reading is also provided. We also generated bounding boxes containing anatomical structures on every image and share them as part of this dataset. These bounding boxes can be used in conjunction with eye gaze information to produce more meaningful analyses.

We present evidence that this dataset can help with two important tasks for AI practitioners in radiology:

[1]IBM Research, Almaden Research Center, San Jose, CA, 95120, USA. [2]Department of Computer Science, Virginia Tech, Blacksburg, VA, 24061, USA. [3]Department of Radiology and Imaging Sciences, Emory University, Atlanta, GA, 30322, USA. [4]These authors contributed equally: Satyananda Kashyap, Ismini Lourentzou, Joy T. Wu. ✉e-mail: akarargyris@gmail.com; satyananda.kashyap@ibm.com; mmoradi@us.ibm.com
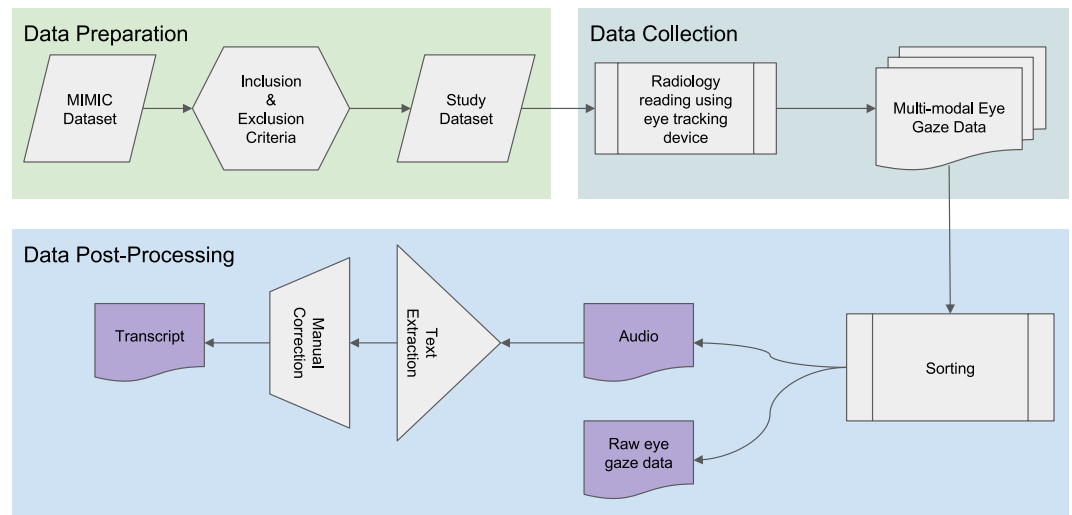
Fig. 1 Flowchart of Study.

- The coordinates marking the areas of the image that a radiologist looks at while reporting a finding provide an approximate region of interest/attention for that finding. Without altering a radiologist's routine, this approach presents an inexpensive and efficient method for generating a locally annotated collection of images for training machine learning algorithms (e.g., disease classifiers). Since we also share the ground truth bounding boxes, the validity of the eye-tracking in marking the location of the finding can be further studied using this dataset. We demonstrate the utilization of eye gaze in deep neural network training and show that performance improvements can be obtained.
- Tracking of the eyes can characterize how radiologists approach the task of reading radiographs. The study of the eye gaze of radiologists while reading normal and disease radiographs, presented as attention maps, reveals a cognitive workflow pattern that AI developers can use when building their models.

We invite researchers in the radiology community who wish to contribute to the further development of the dataset to contact us.

Methods

Figure 1 provides an overview of the study and data generation process. In this work we use the publicly available MIMIC-CXR Database[2,12] in conjunction with the publicly available Emergency Department (ED) subset of the MIMIC-IV Clinical Database[13]. The MIMIC-IV-ED subset contains clinical observations/data and outcomes related to some of the CXR exams in the MIMIC-CXR database. Inclusion and exclusion criteria were applied to the patient attributes and clinical outcomes (via the discharge diagnosis, a.k.a the ICD-9 code) recorded in the MIMIC-IV Clinical Database[13], resulting in a subset of 1,083 cases that equally cover 3 conditions: Normal, Pneumonia and Congestive Heart Failure (CHF). The corresponding CXR images of these cases were extracted from the MIMIC-CXR database[2]. A radiologist (American Board of Radiology certified with over 5 years of experience) performed routine radiology reading of the images using the Gazepoint GP3 Eye Tracker[14] (i.e., eye-tracking device), Gazepoint Analysis UX Edition software[15] (i.e., software for performing eye gaze experiments), a headset microphone, a desktop computer and a monitor (Dell S2719DGF) set at 1920 × 1080 resolution. Radiology reading took place in multiple sessions (i.e., 30 cases per session) over a period of 2 months (i.e., March – May 2020). The Gazepoint Analysis UX Edition[15] exported video files (.avi format) containing eye fixations and voice dictation of radiologist's reading along with spreadsheets (.csv format) containing eye tracker's recorded eye gaze data. The audio was extracted from the video files and saved in wav and mp3 format. Subsequently, these audio files were processed with speech-to-text software (i.e., https://cloud.google.com/speech-to-text) to extract text transcripts along with dictation word time-related information (.json format). Furthermore, these transcripts were manually corrected. The final dataset contains the raw eye gaze signal information (.csv), audio files (.wav,.mp3) and transcript files (.json).

Ethical statement. The source data from the MIMIC-CXR and MIMIC-IV databases have been previously de-identified, and the institutional review boards of the Massachusetts Institute of Technology (No. 0403000206) and Beth Israel Deaconess Medical Center (2001-P-001699/14) both approved the use of the databases for research. We have also complied with all relevant ethical regulations regarding the use of the data for our study.

Data preparation. *Inclusion and exclusion criteria.* Figure 2 describes the inclusion/exclusion criteria used to generate this dataset. These criteria were applied to the MIMIC-IV Clinical Database[13] to identify the CXR studies of interest. The studies were used to extract their corresponding CXR images from the MIMIC-CXR Database[2].
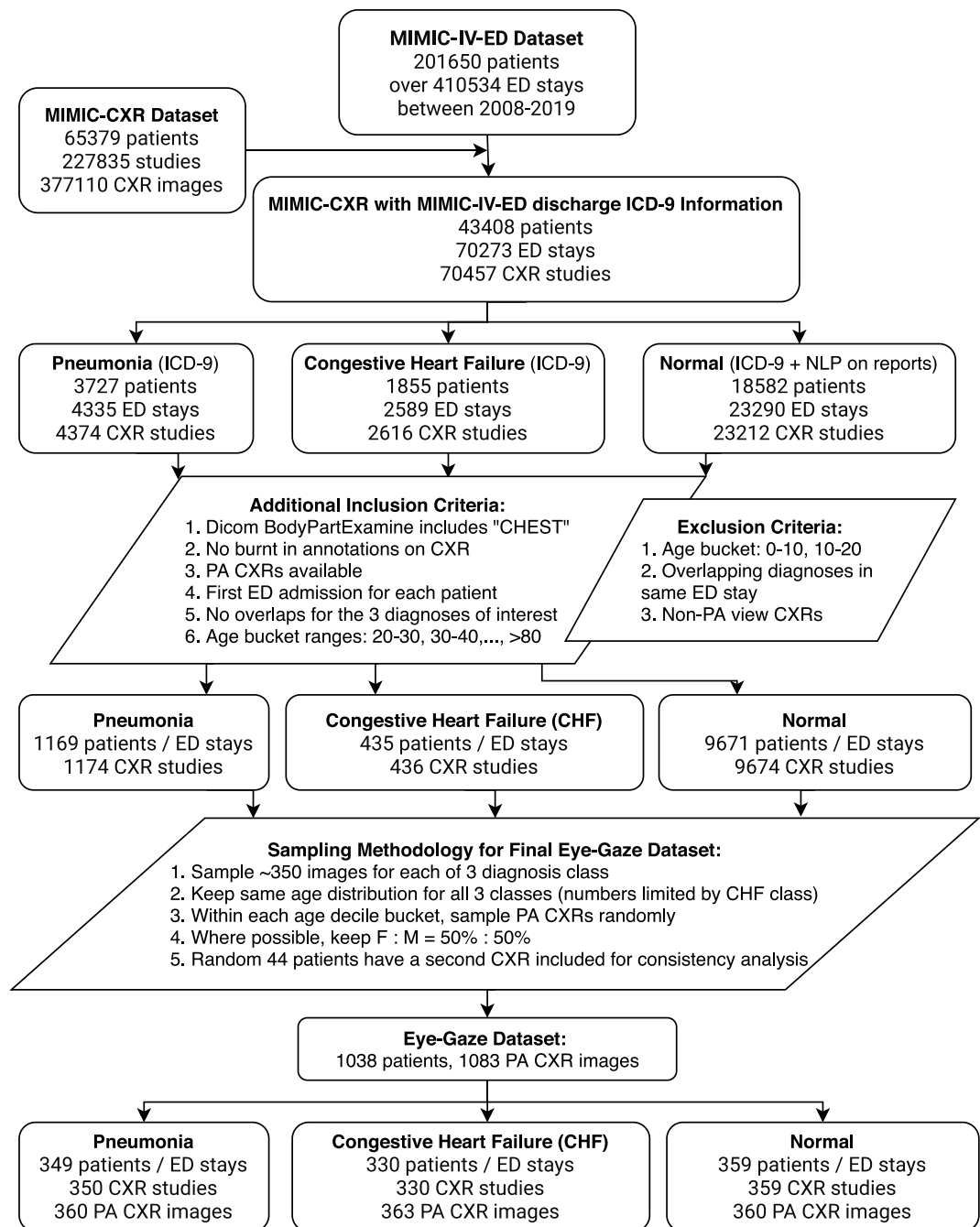
Fig. 2 Sampling flowchart for selecting images for this study from the MIMIC-IV (the ED subset) and the MIMIC-CXR datasets.0[2,13].

We selected two clinically prevalent and high impact diseases, pneumonia and congestive heart failure (CHF), in the Emergency Department (ED) setting. We also picked normal cases as a comparison class. Unlike related CXR labeling efforts[1], where the same labels are derived from radiology reports using natural language processing (NLP) alone, the ground truth for our pneumonia and CHF class labels were derived from unique discharge ICD-9 codes (verified by our clinicians) from the MIMIC-IV-ED tables[13].

This ensures the ground truth is based on a formal clinical diagnosis and is likely to be more reliable, given that ICD-9 discharge diagnoses are typically derived from a multi-disciplinary team of treating providers after having considered all clinically relevant information (e.g., bedside observations, labs) in addition to the CXR images. This is particularly important since CXR observations alone may not always be specific enough to reach a pneumonia or CHF clinical diagnosis. The normal class is determined by excluding any ICD-9 codes that may result in abnormalities visible on CXRs and also having no abnormal labels extracted from the relevant CXR reports using CXR report labeler[16]. The code to run the inclusion and exclusion criteria is available on our Github repository https://github.com/cxr-eye-gaze/eye-gaze-dataset.
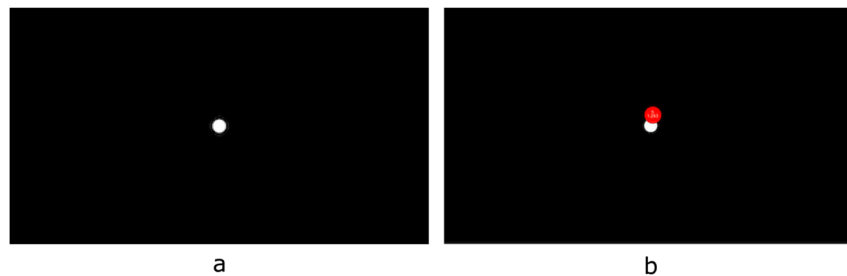
Fig. 3 Calibration images presented during data collection with the radiologist's fixation super-imposed. Left: Calibration image presented to radiologist during data collection, Right: Radiologist's fixation super-imposed in red.

In addition, our sampling criteria prioritized the strategy for getting a good number of examples of disease features across a range of ages and sex from the source ED population. The goal is to support building and evaluation of computer vision algorithms that do not overly rely on age and sex biases, which may depict prominent visual features[17], to predict disease classes.

Preparation of images.   Preparation of images The 1,083 CXR images (Inclusion/Exclusion section) were converted from DICOM (Digital Imaging and Communications in Medicine) format to .png format: normalized (0–255), resized and padded to 1920 × 1080 to fit the radiologist's computer's monitor resolution (i.e., kept same aspect ratio) and to enable loading into Gazepoint Analysis UX Edition[15].

A calibration image (i.e., resolution: 1920 × 1080 pixels) consisting of a white dot (30 pixels in radius) was generated (see Fig. 3 - left). The calibration image was presented to the radiologist randomly during data collection to measure eye gaze offset (see Fig. 3 - right).

The 1,083 images and calibration images were split into 38 numbered folders (i.e., '01', '02', '03', …'38') with no more than 30 images per folder. These folders were then uploaded to IBM's internal BOX™ and shared with the radiologist who finally downloaded and loaded them to Gazepoint Analysis UX Edition software[15] to perform the reading (i.e., data collection).

Data collection.   *Software and hardware setup.*   The Gazepoint GP3 Eye Tracker[14] is an optical eye tracker that uses infrared light to detect eye fixation. The Gazepoint Analysis UX Edition software[15] is a software suite that comes along with the Gazepoint GP3 Eye Tracker and allows performing eye gaze experiments on image series.

The Gazepoint GP3 Eye Tracker[14] was set up in the radiologist's routine working environment on a Windows desktop PC (connected at USB 3 port). The Gazepoint Analysis UX Edition software[15] was also installed on the same computer. Each session was a standalone experiment that contained up to 30 images for reading by the radiologist. The radiologist's eyes were 28 inches away from the monitor. The choice of this number of images was intentional to avoid fatigue and interruptions and to allow for timely offline review and quality assurance of each session recordings by the rest of the team. Gazepoint Analysis UX Edition software[15] allows for 9-point calibration which occurred in the beginning of each session. In addition, Gazepoint Analysis UX Edition[15] allows the user to move to the next image either by pressing the spacebar on the keyboard when done with a case or by waiting for a fixed time. In this way the radiologist was able to move to the next CXR image when he was done with a given image, making the experiment easier.

Radiology reading.   The radiologist read 1,083 CXR images reporting in unstructured prose, same as what he would perform in his routine working environment. The goal was to simulate a typical radiology read with minimal disruption from the eye gaze data collection process. The order of the CXR images was randomized to allow a blinded radiology read. Furthermore, we intentionally withheld the reason for exam information from our radiologist in order to collect an objective CXR exam interpretation based only on the available imaging features.

The original source MIMIC-CXR Database[2] has the original de-identified free text reports for the same images, which were collected in real clinical scenarios where the reading radiologists had access to some patient clinical information outside the CXR image. The radiologists may even have had discussions about the patients with the bedside treating physician. Interpreting CXRs with additional patient clinical information (e.g., age, sex, other signs or symptoms) has the benefit of allowing radiologists to provide a narrower list of disease differential diagnosis by reasoning with their extra medical knowledge. However, it may also have the unintended effect of narrowing the radiology finding descriptions or subconsciously biasing what the radiologists look for in the image. In contrast, our radiologist only had the clinical information that all the CXRs came from an ED clinical setting.

By collecting a more objective read, we ensured that the CXR images used in this dataset have associated reports from both kinds of reading scenarios (read with and without patient clinical information). The goal is to broaden the range of possible technical and clinical research questions that future researchers working with the dataset may ask and explore.

The only clinical information that our radiologist was given when reading the CXRs in the eye gaze dataset was that the CXR exams were taken in the Emergency Department (ED) setting. The ED setting is broad and can include several common acute and chronic conditions that may be apparent on CXR images. The radiologist was
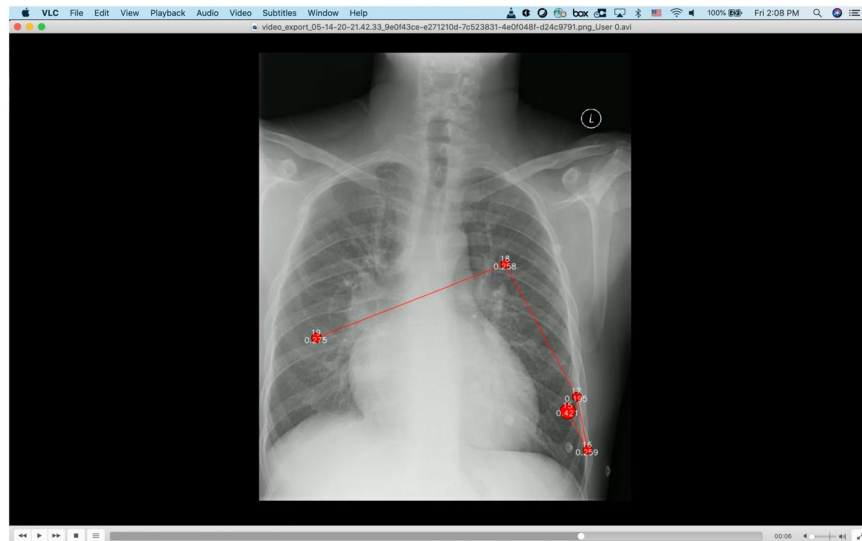
Fig. 4 Sample video exported from Gazepoint Analysis UX Edition[15] showing a CXR case image with overlayed fixations.

blinded to the final diagnosis for the ED admission, asked to read the CXR as he would typically in an ED case, and the CXR image order was randomized to remove potential clinical patterns.

The eye gaze radiology read setting is set to contrast the routine radiology report setting under which the original MIMIC reports were collected because our intention is to add new value to the dataset. In the routine workflow, information not visible on the CXR images is often available to the radiologists. As a result, the reports are influenced by complex multimodal inference. We made our best attempt to separate the multimodal information in the eye gaze dataset. A future direction of this work is to integrate all the multimodal clinical information available in this dataset in the experiments.

We observed that pneumonia and heart failure (disease terminologies) are described much less frequently by our radiologist in his eye gaze reports, which we attribute to the effect of creating the reports with just the image information and no prior patient clinical information. In contrast, the findings such as lung opacity, which do not require additional clinical information and can be deducted based on image appearance alone, are consistent between original reports and those produced during the eye gaze experiments.

Data post-processing.   At the end of each session the radiologist exported the following information from the Gazepoint Analysis UX Edition software[15]: (1) fixation spreadsheet (.csv) containing fixation information for each case in the session, (2) eye gaze spreadsheet (.csv) containing raw eye gaze information for each case in the session, and (3) videos files (.avi) containing audio (i.e., radiologist's dictation) along with his eye gaze fixation heatmaps per session case (see Fig. 4). These files were uploaded and shared over IBM's internal BOXTM subscribed service. A team member reviewed each video for any technical quality issues (e.g., corrupted file, video playback stopped abruptly, bad audio quality).

Once data collection (i.e., 38 sessions) finished, the following post-processing tasks were performed.

Spreadsheet merging.   From all sessions (i.e., folders), the fixations spreadsheets were concatenated into a single spreadsheet file: fixations.csv, and the raw eye gaze spreadsheets were concatenated into a single spreadsheet file: eye_gaze.csv. Mapping of eye gaze and fixation from the screen coordinate system to the original MIMIC image coordinate system was also performed at this stage.

Detailed descriptions of these tables are provided in the Data Records section.

Audio extraction and transcript generation.   For each session video file (i.e., containing radiologist's eye gaze fixations and dictation in .avi format, Fig. 4) the dictation audio was extracted and saved in `audio.wav` and `audio.mp3` files. We used Google Speech-To-Text service https://cloud.google.com/speech-to-text to transcribe the audio (i.e., wav file) into text. The transcribed text was saved in `transcript.json` containing timestamps and corresponding words based on the API example found in https://cloud.google.com/speech-to-text/docs/async-time-offsets. Furthermore, the transcripts were corrected manually by three (3) team members (all verified by the radiologist) using the original audio. An example of a transcript json is given in the Data Records section.

Segmentation maps and bounding boxes for anatomies.   Two supplemental datasets are also provided to enrich this dataset:

- **Segmentation maps:** Segmentation Maps Four (4) key anatomical structures per image were generated: i) `left_lung.png`, ii) `right_lung.png` and iii) `mediastinum.png` and iv) `aortic_knob.png`. These anatomical structures were automatically segmented by an internal segmentation model and then
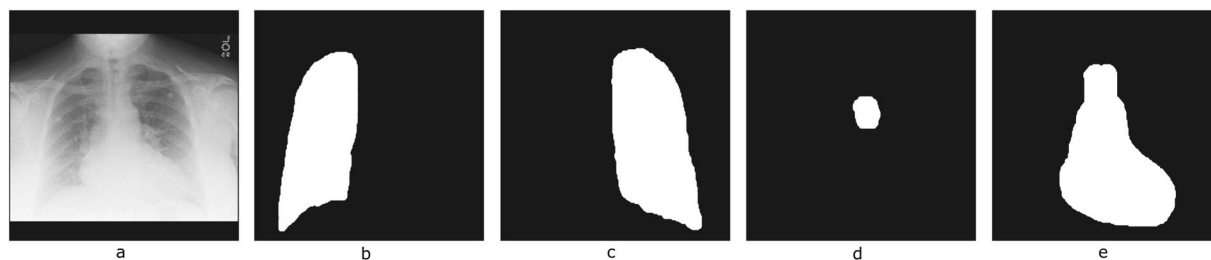
Fig. 5 Representative chest X-ray image and the corresponding segmentation maps. From Left to Right: (**a**) CXR image, (**b**) Right lung, (**c**) Left lung, (**d**) Aortic knob and (**e**) Mediastinum.
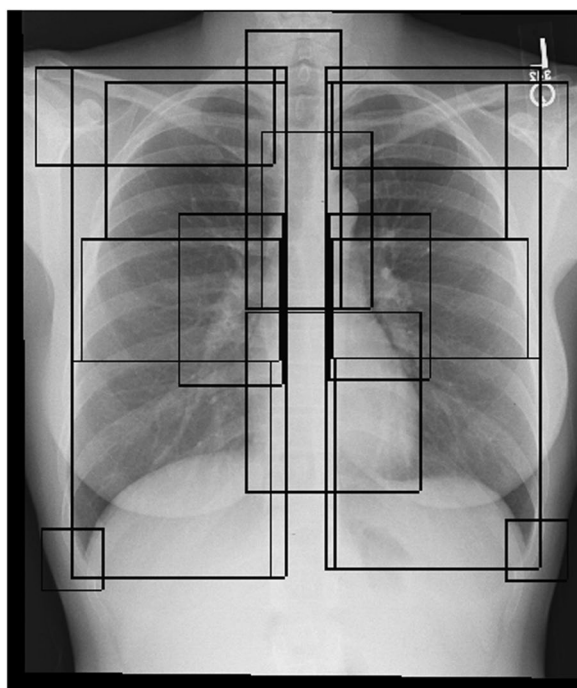

Fig. 6 Sample CXR case with 17 overlaying anatomical bounding boxes. The anatomies in the chest overlay one another on CXRs since the image is the 2D X-ray shadow capture of a 3D object.

    manually reviewed and corrected by the radiologist. Each image has pixel values 255 for anatomy and 0 for background. Figure 5 presents a sample case with its corresponding segmentation maps.

- **Bounding boxes**: An extension of a bounding box extraction pipeline[18] was used to extract 17 anatomical bounding boxes for each CXR image, which include: 'right lung', 'right upper lung zone', 'right mid lung zone', 'right lower lung zone', 'left lung', 'left upper lung zone', 'left mid lung zone', 'left lower lung zone', 'right hilar structures', 'left hilar structures', 'upper mediastinum', 'cardiac silhouette', 'trachea', 'right costophrenic angle', 'left costophrenic angle', 'right clavicle', 'left clavicle'. These zones cover the clinically most important anatomies on a Posterior Anterior (PA) CXR image. These automatically produced bounding boxes were manually corrected (when required). Each bounding box is described by the top left corner point (X_{X1}, Y_{Y1}) and bottom right corner point (X_{X2}, Y_{Y2}) on the original CXR image coordinate system. Figure 6 shows an example of anatomical bounding boxes. The information for bounding boxes of the 1,083 images are contained in `bounding_boxes.csv`.

    Researchers can utilize these two (2) supplemental datasets to improve segmentation and disease localization algorithms by combining them with the eye gaze data. In the Statistical analysis on fixations subsection we utilize the `bounding_boxes.csv` to perform statistical analysis between fixations and condition pairs.

Data Records

An overview of the released dataset with their relationships is provided in Fig. 7. Specifically, four (4) data documents and one (1) folder are provided:

1. `master_sheet.csv`: Spreadsheet containing MIMIC DICOM ids along with study clinical indication sentence, report derived finding labels, and ICD-9 derived outcome disease labels.
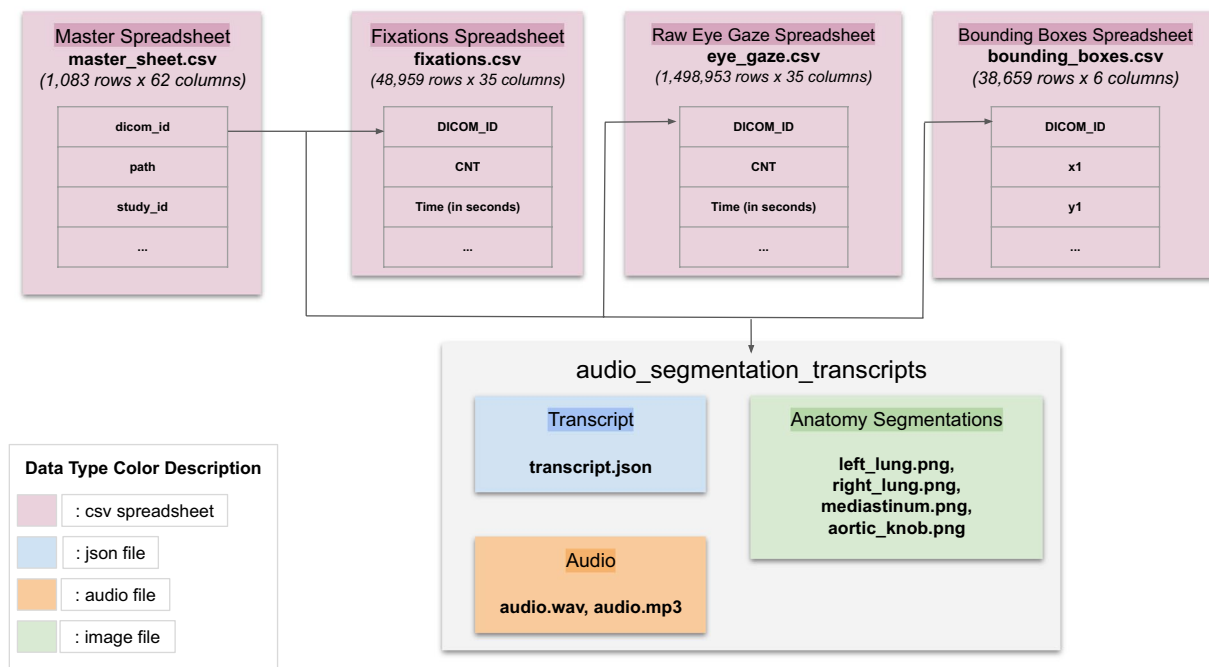
Fig. 7 Overview of Dataset.

2. `eye_gaze.csv`: Spreadsheet containing raw eye gaze data as exported by Gazepoint Analysis UX Edition software[15].

3. `fixations.csv`: Spreadsheet containing fixation data as exported by Gazepoint Analysis UX Edition software[15].

4. `bounding_boxes.csv`: Spreadsheet containing bounding box coordinates for key frontal CXR anatomical structures.

5. `audio_segmentation_transcripts`: Folder containing dictation audio files (i.e., mp3, wav), transcript file (i.e., json), anatomy segmentation mask files (i.e., png) for each dicom id.

The dataset is hosted at https://doi.org/10.13026/qfdz-zr67[19]. To utilize the dataset, the only requirement for the user is to obtain Physionet access to the MIMIC-CXR Database[2] in order to download the original MIMIC CXR images in DICOM format. The `dicom-id` tag found throughout all the dataset documents maps records to the MIMIC CXR images. A detailed description of each data document is provided in the following subsections.

Master spreadsheet. The master spreadsheet (`master_sheet.csv`) provides the following key information:

- The `dicom-id` column maps each row to the original MIMIC CXR image as well as the rest of the documents in this dataset.
- The `study-id` column maps the CXR image/dicom to the associated CXR report, which can be found from the source MIMIC-CXR dataset[2].
- For each CXR study (`study-id`), granular radiology 'finding' labels have been extracted from the associated original MIMIC reports by two different NLP pipelines – first is the CheXpert NLP pipeline[1], and second is an NLP pipeline developed internally[16].
- Additionally, for each CXR study (`study-id`), the reason for exam indication has been sectioned out from the original MIMIC CXR reports. The indication sentence(s) tend to contain patient clinical information that may not otherwise be visible from the CXR image alone.

Table 1 describes in detail each column found in the master spreadsheet.

Fixations and eye gaze spreadsheets. The eye gaze information is stored in two (2) files: a) `fixations.csv`, and b) `eye_gaze.csv`. Both files were exported by the Gazepoint Analysis UX Edition software[15]. Specifically, the `eye_gaze.csv` file contains one row for every data sample collected from the eye tracker, while `fixations.csv` file contains a single data entry per fixation. The Gazepoint Analysis UX Edition software[15] generates the `fixations.csv` file from the `eye_gaze.csv` file by averaging all data within a fixation to estimate the point of fixation based on the eye gaze samples, stopping when a saccade is detected. Table 2 describes in detail each column found in the fixations and eye gaze spreadsheets.

Bounding boxes spreadsheet. The bounding boxes spreadsheet contains the following information:

Column Name	Description
dicom-id	DICOM ID in the original MIMIC dataset[2]
path	Path of DICOM image in the original MIMIC dataset
study-id	Study ID in the original MIMIC dataset
patient-id	Patient ID in the original MIMIC dataset
stay-id	Stay ID in the original MIMIC dataset
gender	Gender of patient in the original MIMIC dataset
anchor-age	Age range in years of patient in the original MIMIC dataset
image-top-pad, image-bottom-pad, image-left-pad, image-right-pad	Padding (top, bottom, left, right respectively) in pixels applied after re-scaling MIMIC image to 1920 × 1080
normal-reports	No affirmed abnormal finding labels or descriptors documented in the original MIMIC-CXR reports, extracted using an internal CXR labeling pipeline[16].
Normal	No abnormal chest related final diagnosis from the Emergency Department (ED) discharge ICD-9 records AND have normal-reports as defined above.
CHF	A clinical diagnosis of heart failure (includes ICD-9 for congestive heart failure, chronic or acute on chronic heart failure) from the ED visit as determined from the associated ICD-9 discharge diagnostic code.
Pneumonia	A clinical diagnosis of any lung infection (pneumonia) including bacterial and viral, as determined from the ICD-9 discharge diagnosis code of the ED visit.
dx1, dx2, dx3, dx4, dx5, dx6, dx7, dx8, dx9	The descriptive ICD-9 diagnosis name associated with the Emergency Room admission for which the CXR study was ordered. ICD-9 final diagnoses are used to identify the 3 classes in the eye gaze analysis and experiments.
dx1-icd, dx2-icd, dx3-icd, dx4-icd, dx5-icd, dx6-icd, dx7-icd, dx8-icd, dx9-icd	ICD-9 code for corresponding dx
consolidation, enlarged-cardiac-silhouette,linear-patchy-atelectasis,lobar-segmental-collapse,not-otherwise-specified-opacity,pleural-parenchymal-opacity, pleural-effusion-or-thickening, pulmonary-edema-hazy-opacity, normal-anatomically, elevated-hemidiaphragm, hyperaeration, vascular-redistribution	Abnormal finding labels derived from the original MIMIC-CXR reports by an internal IBM CXR report labeler[16]. 0: Negative, 1: Positive
atelectasis-chx, cardiomegaly-chx consolidation-chx, edema-chx,enlarged-cardiomediastinum-chx fracture-chx,lung-lesion-chx,lung-opacity-chx,no-finding-chx,pleural-effusion-chx,pleural-other-chx,pneumonia-chx,pneumothorax-chx,support-devices-chx	ChexPert[1] report derived abnormal finding labels for MIMIC-CXR. 0: negative, 1: positive, −1: uncertain
cxr_exam_indication	The reason for exam sentences sectioned out from Indication section of the original MIMIC-CXR reports[1]. They briefly summarize patients' immediate clinical symptoms, prior medical conditions and or recent procedures that are relevant for interpreting the CXR study within the clinical context.

Table 1. Master Spreadsheet.

- `dicom_id`: DICOM ID as provided in MIMIC-CXR Database[2] for each image.
- `bbox_name`: These are the names for the 17 rectangular anatomical zones that bound the key anatomical organs on a frontal CXR image. Each lung (right and left) is bounded by its own bounding box, as well as subdivided into common radiological zones (upper, mid and lower lung zones) on each side. The upper mediastinum and the cardiac silhouette (heart) bounding boxes make up the mediastinum anatomy. The trachea is a bounding box that includes the visible tracheal air column on a frontal CXR, as well as the beginnings of the right and left main stem bronchi. The left and right hilar structures contain the left or right main stem bronchus as well as the lymph nodes and blood vessels that enter and leave the lungs in the hilar region. The left and right costophrenic angles are key regions to assess for abnormalities on a frontal CXR. The left and right clavicles can have potential fractures to rule out, but are also important landmarks to assess whether the patient (hence the anatomies on the CXRs) is rotated or not (which affects the appearance of potential abnormalities). Some of the bounding boxes (e.g clavicles) could be missing for an image if the target anatomical structure is cut off from the CXR image's field of view.
- `x1`: x coordinate for starting point of bounding box (upper left).
- `y1`: y coordinate for starting point of bounding box (upper left).
- `x2`: x coordinate for ending point of bounding box (lower right).
- `y2`: y coordinate for ending point of bounding box (lower right).

Please see Fig. 6 for an example of all the anatomical bounding boxes.

Audio, segmentation maps and transcripts. The `audio_segmentation_transcripts` folder contains subfolders for all the cases in the study with case `dicom_id` as name. Each subfolder contains: a) the dictation audio file (mp3, wav), b) the segmentation maps of anatomies (png), as described in Segmentation maps and bounding boxes for anatomies subsection above, and c) the dictation transcript (json). The dictation `transcript.json` contains the following tags:

Data Type/Column Name	Description
DICOM-ID	DICOM ID from original MIMIC dataset.
CNT	The counter data variable is incremented by 1 for each data record sent by the server. Useful to determine if any data packets are missed by the client.
TIME(in secs)	The time elapsed in seconds since the last system initialization or calibration. The time stamp is recorded at the end of the transmission of the image from camera to computer. Useful for synchronization and to determine if the server computer is processing the images at the full frame rate. For a 60 Hz camera, the TIME value should increment by 1/60 seconds.
TIMETICK(f = 10000000)	This is a signed 64-bit integer which indicates the number of CPU time ticks for high precision synchronization with other data collected on the same CPU.
FPOGX	The X- coordinates of the fixation POG, as a fraction of the screen size. (0,0) is top left, (0.5,0.5) is the screen center, and (1.0,1.0) is bottom right.
FPOGY	The Y-coordinates of the fixation POG, as a fraction of the screen size. (0,0) is top left, (0.5,0.5) is the screen center, and (1.0,1.0) is bottom right.
FPOGS	The starting time of the fixation POG in seconds since the system initialization or calibration.
FPOGD	The duration of the fixation POG in seconds
FPOGID	The fixation POG ID number
FPOGV	The valid flag with value of 1 (TRUE) if the fixation POG data is valid, and 0 (FALSE) if it is not. FPOGV valid is TRUE ONLY when either one, or both, of the eyes are detected AND a fixation is detected. FPOGV is FALSE all other times, for example when the participant blinks, when there is no face in the field of view, when the eyes move to the next fixation (i.e., a saccade)
BPOGX	The X-coordinates of the best eye POG, as a fraction of the screen size.
BPOGY	The Y-coordinates of the best eye POG, as a fraction of the screen size.
BPOGV	The valid flag with value of 1 if the data is valid, and 0 if it is not.
LPCX	The X-coordinates of the left eye pupil in the camera image, as a fraction of the camera image size.
LPCY	The Y-coordinates of the left eye pupil in the camera image, as a fraction of the camera image size.
LPD	The diameter of the left eye pupil in pixels
LPS	The scale factor of the left eye pupil (unitless). Value equals 1 at calibration depth, is less than 1 when user is closer to the eye tracker and greater than 1 when user is further away.
LPV	The valid flag with value of 1 if the data is valid, and 0 if it is not.
RPCX	The X-coordinates of the right eye pupil in the camera image, as a fraction of the camera image size.
RPCY	The Y-coordinates of the right eye pupil in the camera image, as a fraction of the camera image size.
RPD	The diameter of the right eye pupil in pixels
RPS	The scale factor of the right eye pupil (unitless). Value equals 1 at calibration depth, is less than 1 when user is closer to the eye tracker and greater than 1 when user is further away.
RPV	The valid flag with value of 1 if the data is valid, and 0 if it is not.
BKID	Each blink is assigned an ID value and incremented by one. The BKID value equals 0 for every record where no blink has been detected.
BKDUR	The duration of the preceding blink in seconds.
BKPMIN	The number of blinks in the previous 60 second period of time.
LPMM	The diameter of the left eye pupil in millimeters.
LPMMV	The valid flag with value of 1 if the data is valid, and 0 if it is not.
RPMM	The diameter of the right eye pupil in millimeters.
RPMMV	The valid flag with value of 1 if the data is valid, and 0 if it is not.
SACCADE-MAG	Magnitude of the saccade calculated as distance between each fixation (in pixels).
SACCADE-DIR	The direction or angle between each fixation (in degrees from horizontal).
X_ORIGINAL	The X coordinate of the fixation in original DICOM image.
Y_ORIGINAL	The Y coordinate of the fixation in original DICOM image.

Table 2. Fixations and Eye Gaze Spreadsheets.

- `full_text`: The full text for the transcript.
- `time_stamped_text`: The full text broken into timestamped phrases:
  - `phrase`: Phrase text in the transcript.
  - `begin_time`: The starting time (in seconds) of dictation for a particular phrase.
  - `end_time`: The end time (in seconds) of dictation of a particular phrase.

Figure 8 shows the structure of the `audio_segmentation_transcripts` folder, while Fig. 9 shows a transcript json example.

Technical Validation

We subjected two aspects of the released data to reliability and quality validation: eye gaze and transcripts.

The code for the validation tasks below can be found at https://github.com/cxr-eye-gaze/eye-gaze-dataset.

Validation of eye gaze data. As mentioned in the Preparation of images subsection, a calibration image was interjected randomly within the eye gaze sessions to measure the error of the eye gaze on the X- and Y- axis
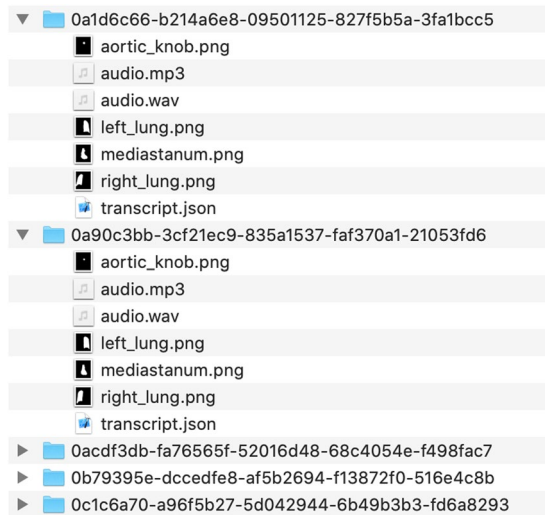
Fig. 8 `audio_segmentation_transcripts` folder structure.

(Fig. 3). A total of 59 calibration images were presented throughout the data collection. We calculated the error by using the fixation coordinates of the last entry of each calibration image (i.e., the final resting fixation by the radiologist on the calibration mark). The overall average percentage error on X, Y axes was calculated with (error_X, error_Y) = (0.0089, 0.0504), and std: (0.0065, 0.0347) respectively. In pixels, the same error was: (error_X, error_Y) = (17.0356, 54.3943), with std: (12.5529, 37.4257) respectively.

Validation of transcripts. As mentioned in the Preparation of images subsection, transcripts were generated using Google Speech-to-Text https://cloud.google.com/speech-to-text on the dictation audio with timestamps per dictated word. The software produced two (2) types of errors:

- Type A: Incorrect identification of a word at a particular time stamp (please see example in Fig. 10).
- Type B: Missed transcribed phrases of the dictation (please see example in Fig. 11).

The transcripts were manually corrected by three (3) experts and verified by the radiologist. Both types of errors were completely corrected. For Type B error, the missing text (i.e., more than one (1) word) was added with an estimation of the `begin_time` and `end_time` manually. To measure the potential error in the transcripts, the number of phrases with multiple words in a single time stamp was calculated (i.e., Type B error):

- Total number of phrases: 19,499
- Number of phrase with single words: 18,434
- Number of phrases with multiple words: 1,065

$$\text{Type B error} = 1 - \frac{1065}{19499} = 5.46\%$$

Statistical analysis on fixations. We performed t-test analysis to measure any significant differences between fixations for each condition within anatomical structures. More specifically, we performed the following steps:

1. We examined the average number of fixations made in each disease condition, and found that the expert made significantly more overall fixations in the two diseased conditions than in the normal condition ($p < 0.01$).
2. For each image we calculated the number of fixations that their coordinates (i.e., X_ORIGINAL, Y_ORIGINAL in `fixations.csv`) fall into each anatomical zone (bounding box) found in `bounding_boxes.csv`.
3. We performed t-test for each anatomical structure between condition pairs: i) Normal vs. Pneumonia, ii) Normal vs. CHF, iii) Pneumonia vs CHF.

Figure 12 shows the duration of fixations per image for each disease condition and anatomical area, while Table 3 shows p-values from each t-test. Fixations on Normal images are significantly different from Pneumonia and CHF. More fixations are made for images associated with either the Pneumonia or CHF final diagnoses. Moreover, fixations for the abnormal cases are mainly concentrated in anatomical regions (i.e., lungs and heart) that are relevant to the diagnosis, rather than distributed at random. Overall, the fixations on Pneumonia and CHF are comparatively similar, although still statistically different (e.g., Left Hilar Structure, Left Lung, Cardiac Silhouette, Upper Mediastinum). These statistical differences demonstrate that the radiologist's eye-tracking

```
1  {
2      "full_text": " prominent heart. cardiomegaly in fact. suspect small left effusion with atelectasis. right lung
   appears clear.",
3      "time_stamped_text": [
4          {
5              "begin_tim": 1.0,
6              "end_time": 1.7000000000000002,
7              "phrase": "prominent"
8          },
9          {
10             "begin_tim": 1.7000000000000002,
11             "end_time": 2.0,
12             "phrase": "heart."
13         },
14         {
15             "begin_tim": 3.0,
16             "end_time": 3.8,
17             "phrase": "cardiomegaly"
18         },
19         {
20             "begin_tim": 3.8,
21             "end_time": 4.2,
22             "phrase": "in fact."
23         },
24         {
25             "begin_tim": 6.3,
26             "end_time": 7.1,
27             "phrase": "suspect"
28         },
29         {
30             "begin_tim": 8.2,
31             "end_time": 8.9,
32             "phrase": "small"
33         },
34         {
35             "begin_tim": 8.9,
36             "end_time": 9.0,
37             "phrase": "left"
38         },
39         {
40             "begin_tim": 9.0,
41             "end_time": 9.4,
42             "phrase": "effusion"
43         },
44         {
45             "begin_tim": 10.7,
46             "end_time": 11.2,
47             "phrase": "with"
48         },
49         {
50             "begin_tim": 11.2,
51             "end_time": 11.6,
52             "phrase": "atelectasis."
53         },
54         {
55             "begin_tim": 17.9,
56             "end_time": 18.3,
57             "phrase": "right"
58         },
59         {
60             "begin_tim": 18.3,
61             "end_time": 18.5,
62             "phrase": "lung"
63         },
64         {
65             "begin_tim": 18.5,
66             "end_time": 18.6,
67             "phrase": "appears"
68         },
69         {
70             "begin_tim": 18.6,
71             "end_time": 19.0,
72             "phrase": "clear."
73         }
74     ]
75 }
```

Fig. 9 Transcript example.

information provides insight into the condition of the patient, and shows how a human expert pays attention to the relevant portions of the image when interpreting a CXR exam. The code to replicate the t-test analysis can be found on our Github repository (https://github.com/cxr-eye-gaze/eye-gaze-dataset).

Usage Notes

The dataset is hosted at https://doi.org/10.13026/qfdz-zr67[19] The user is also required to apply for access to MIMIC-CXR Database[2] to download the images used in this study. Our Github repository (https://github.com/cxr-eye-gaze/eye-gaze-dataset) provides a detailed description and source code (Python scripts) on how to use this dataset and reproduce the published validation results (e.g., post-processing, machine learning experiments, etc.). The data in the MIMIC dataset has been previously de-identified, and the institutional review boards of the Massachusetts Institute of Technology (No. 0403000206) and Beth Israel Deaconess Medical Center (2001-P-001699/14) both approved the use of the database for research.

```
1  {
2      "begin_time": 2.9,
3      "end_time": 3.3,
4      "phrase": "hard"
5  }
6
7  {
8      "begin_time": 2.9,
9      "end_time": 3.3,
10     "phrase":"heart"
11 }
12
```

Fig. 10 Top: Example of incorrect detection. Bottom: Manual correction.

```
1  {
2      "begin_time": 5.8,
3      "end_time": 5.9,
4      "phrase": "night."
5  }
6
7  {
8      "begin_time": 5.8,
9      "end_time": 8.5,
10     "phrase":"pacer with leads in the right atrium and ventricle."
11 }
12
```

Fig. 11 Top: Missed and incorrect transcript phrase. Bottom: Manually corrected phrase.
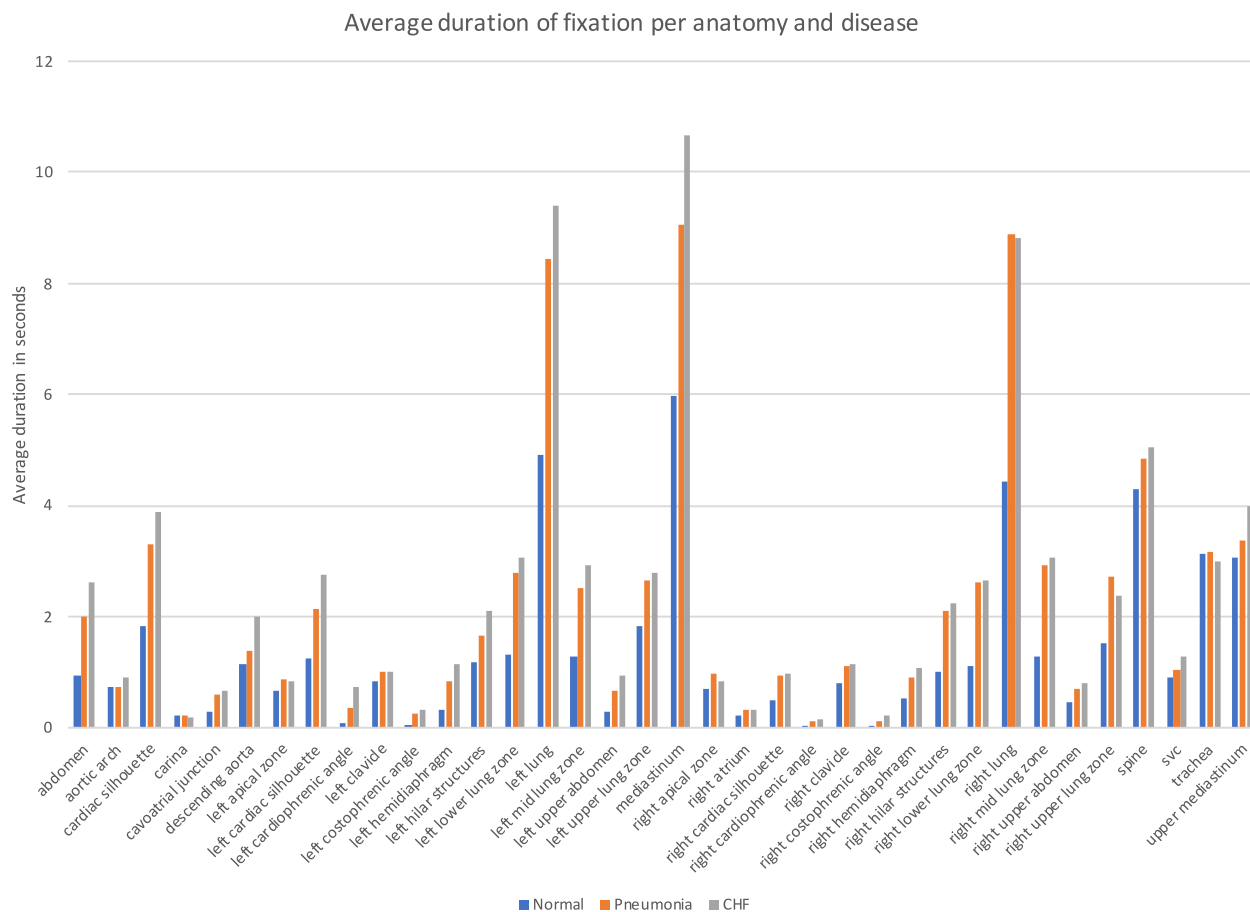
Fig. 12 Fixations vs. anatomical structures vs. conditions.

	Normal vs Pneumonia	Normal vs CHF	Pneumonia vs CHF
Cardiac Silhouette	**0.000**	**0.000**	**0.002**
Left Lung	**0.000**	**0.000**	**0.008**
Left Clavicle	**0.023**	0.110	0.406
Left Costophrenic Angle	**0.000**	**0.000**	0.171
Left Hemidiaphragm	**0.000**	**0.000**	0.195
Left Hilar Structures	**0.000**	**0.000**	**0.000**
Left Lower Lung Zone	**0.000**	**0.000**	0.169
Left Mid Lung Zone	**0.000**	**0.000**	**0.008**
Left Upper Lung Zone	**0.000**	**0.000**	0.676
Mediastinum	**0.000**	**0.000**	**0.000**
Right Lung	**0.000**	**0.000**	0.668
Right Clavicle	**0.000**	**0.000**	0.766
Right Costophrenic Angle	**0.000**	**0.000**	0.270
Right Hemidiaphragm	**0.000**	**0.000**	0.460
Right Hilar Structures	**0.000**	**0.000**	0.136
Right Lower Lung Zone	**0.000**	**0.000**	0.723
Right Mid Lung Zone	**0.000**	**0.000**	0.437
Right Upper Lung Zone	**0.000**	**0.000**	**0.044**
Spine	**0.008**	**0.001**	0.542
Trachea	0.397	0.367	0.097

Table 3. p-values (at 3 decimal places) for each pair of condition and anatomy. p-values in bold demonstrate statistical significant differences.

Use of the dataset in machine learning.   To demonstrate the effectiveness and richness of the information provided in this dataset, we performed two sets of machine learning multi-class classification experiments, with models designed to leverage the eye gaze data. These experiments are provided as dataset applications with simple and popular network architectures and they can function as a starting point for researchers.

Both experiments used the eye gaze heatmap data to predict the multi-class classification of the aforementioned classes (i.e., Normal, CHF, Pneumonia in Table 1) and compare model performances with and without the eye gaze information. Our evaluation metric was AUC (Area Under the ROC Curve). The first experiment incorporates information from the temporal eye gaze fixation heatmaps while the second experiment utilizes static eye gaze fixation heatmaps. In contrast to temporal fixation heatmaps, static fixation heatmaps are the aggregation of all the temporal fixations into a single image.

Temporal heatmaps experiment.   The first model consists of a neural architecture, where the image and the temporal fixation heatmap representations are concatenated before the final prediction layer. We denote an instance of this dataset as $X^{(i)}$, which includes the image $X^{(i)}_{CRX}$ and the sequence of m temporal fixation heatmaps $X^{(i)}_{eyegaze} = \left\{ X^{(i)}_j \right\}^m_{k=1}$, where $k \in \{1, \ldots, m\}$ is the temporal heatmap index. To acquire a fixed vector CRX representation $\mathbf{v}^{(i)}_{CRX}$, the image is passed through a convolutional layer with 64 filters of kernel size 7 and stride 2, followed by max-pooling, batch normalization and a dense layer of 64 units. The baseline model consists of the aforementioned image representation layer, combined with a final linear output layer that produces the classification prediction. Additionally, for the eye gaze, each heatmap is passed through a similar convolutional encoder and then the sequence of heatmaps is summarized with a 1-layer bidirectional LSTM with self-attention[20,21]. We denote the heatmap representation as $\mathbf{u}^{(i)}_{eyegaze}$. Here, the image and heatmaps representations are concatenated before passed through the final classification layer. Figure 13 shows the full architecture. We train with Adam[22], 0.001 initial learning rate and triangular schedule with fixed decay[23], 16 batch size and 0.5 dropout[24]. The experimental results in Fig. 14 show that incorporating eye gaze temporal information, without any preprocessing, filtering or feature engineering, results in 5% AUC improvement for this prediction task when compared to the baseline model with just CXR image data as input.

Static heatmaps experiment.   The previous section demonstrated performance improvements over baselines, originating from the use of temporal fixation heatmaps on a simple network architecture. In this experiment, we pose the classification problem in the U-Net architecture framework[25] with an additional multi-class classification block at the bottleneck layer (see Fig. 15). The encoding and bottleneck arm of the U-Net can be any standard pre-trained classifier without the fully connected layer. The two combined will act as a feature encoder for the classifier. The CNN decoder part of the network runs deconvolution layers to predict the static eye gaze fixation heatmaps. The advantage is that we can jointly train to output the eye gaze static fixation heatmap as well as predict the multi-class classification. Then, during testing on unseen CXR images, the network can predict the disease class and produce a probability heatmap of the most important locations pertaining to the condition.

We used a pre-trained EfficientNet-b0[26] as the encoder and bottleneck layers. The classification head was an adaptive average pooling followed by flatten, dropout[24] and linear output layers. The decoder CNN consisted of three convolutions followed by upsampling layers. The loss function was a weighted combination (γ) of the
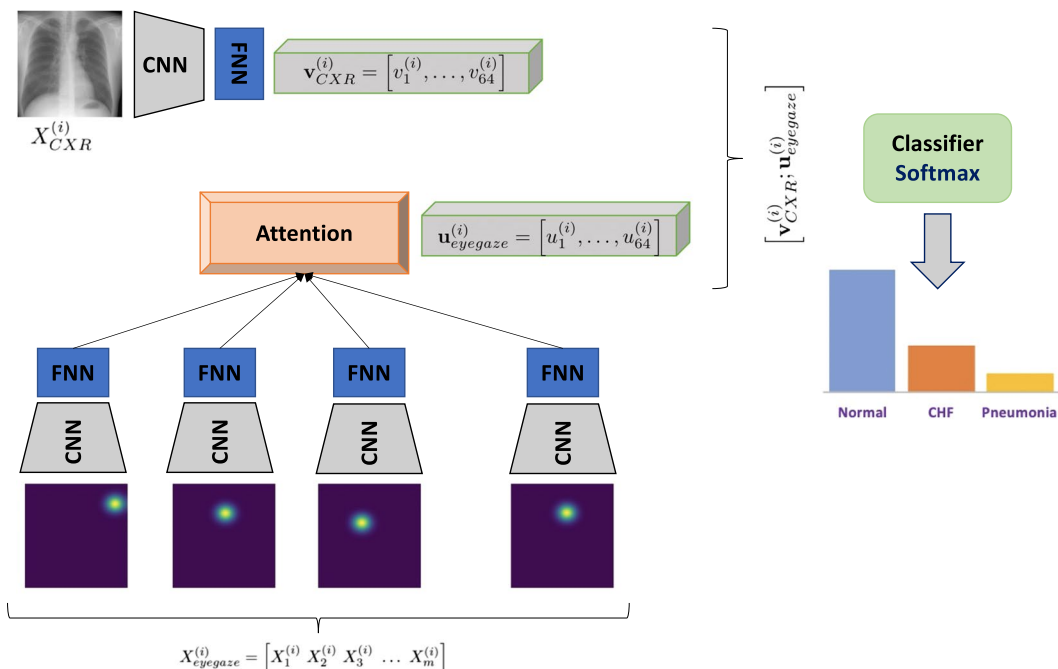
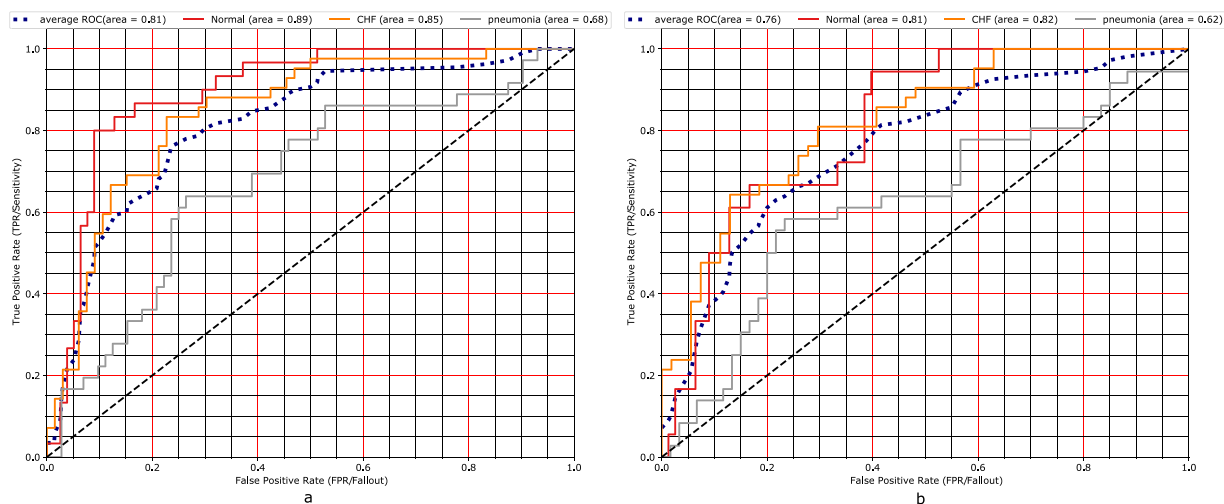**Fig. 13** Model architecture for leveraging temporal eye gaze information.

Fig. 14 Experimental results with and without temporal eye gaze information, i.e., (**a**). Temporal model versus (**b**). Baseline.

classification and the segmentation losses both of which used a binary cross-entropy loss function. The baseline network consisted of just the encoder and the bottleneck arm followed by the classification head.

The hyperparameter tuning for both the U-Net and the baseline classifier was performed using the Tune library[27] and the resulting best performing hyperparameter combination is shown in Table 4. Figure 16 shows the U-Net and baseline AUCs. Both had similar performance. However, for this experiment, we are interested in seeing how network interpretability improved with the use of static eye gaze heatmaps. Figure 17 shows a qualitative comparison of the GradCAM[28]. The GradCAM approach is one of the common methods to visualize activation maps of convolutional networks. While the GradCAM-based heatmaps don't clearly highlight the disease locations, we see clearly that the heatmap probability outputs of the U-Net highlight similar regions to what the static eye gaze heatmap shows.

With both experiments, we tried to demonstrate different use cases of the eye gaze data into machine learning. With the first experiment, we wanted to show how eye gaze data can be utilized in a human-machine setting where radiologist's eye gaze information is fed into the algorithm. The second experiment shows how eye gaze information can be used for explainability purposes through generating verified activation maps. We intentionally did not include other modalities (audio, text) because of the complexity of such experiments and the scope of
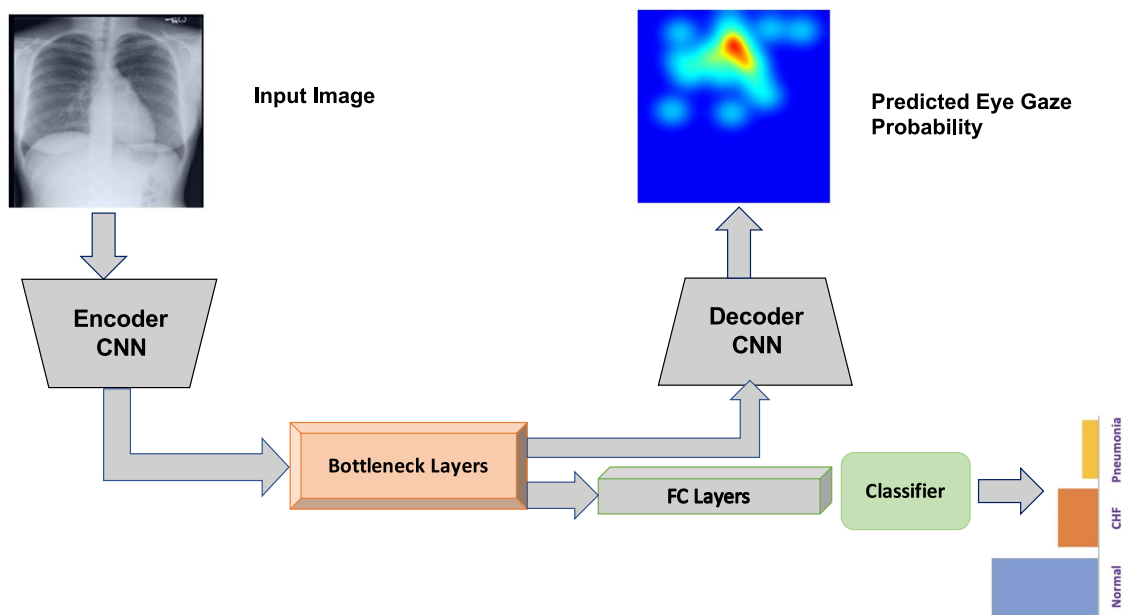
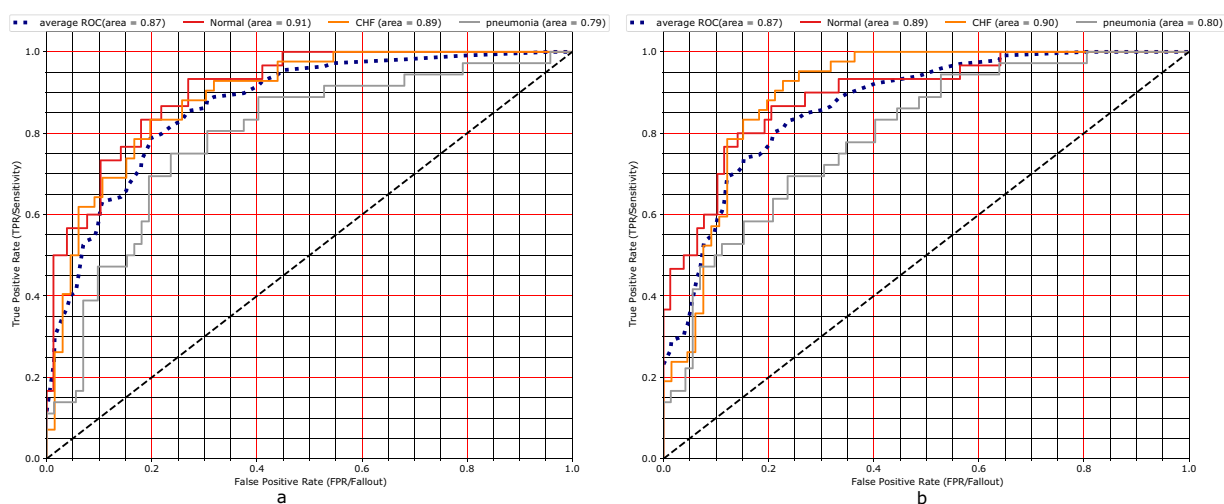**Fig. 15** Block diagram of U-Net utilizing the static heatmap combined with a classification head.

Fig. 16 AUC results with (**a**). U-Net and (**b**). Baseline classifier using static eye gaze information.

ExperimentName	Optimizer	Initial Learning rate	Scheduler[23] Step Size	Epochs	Dropout[24]	γ
UNet	Adam[22]	0.0091	2	35	0.5	0.417
Baseline Classifier	Adam[22]	0.0065	8	20	0.0	N/A

Table 4. Best performing hyper-parameters used for the static heatmap experiments found using the Tune[27] library.

this paper (i.e., dataset description). We hope that these experiments can serve as a starting point for researchers to explore novel ways to utilize this multimodal dataset.

Limitations of Study
Although this study provides a unique large research dataset, we acknowledge the following limitations:

1. The study was performed with a single radiologist. This can certainly bias the dataset (lacks inter-observer variability) and we aim to expand the data collection with multiple radiologists in the future. However, given the relatively large size and richness of data from various sources, i.e., multimodal, we believe that the

| CXR | GradCAM (Baseline) | GradCAM (UNet) | Static Heatmap | UNet Prob. Map |
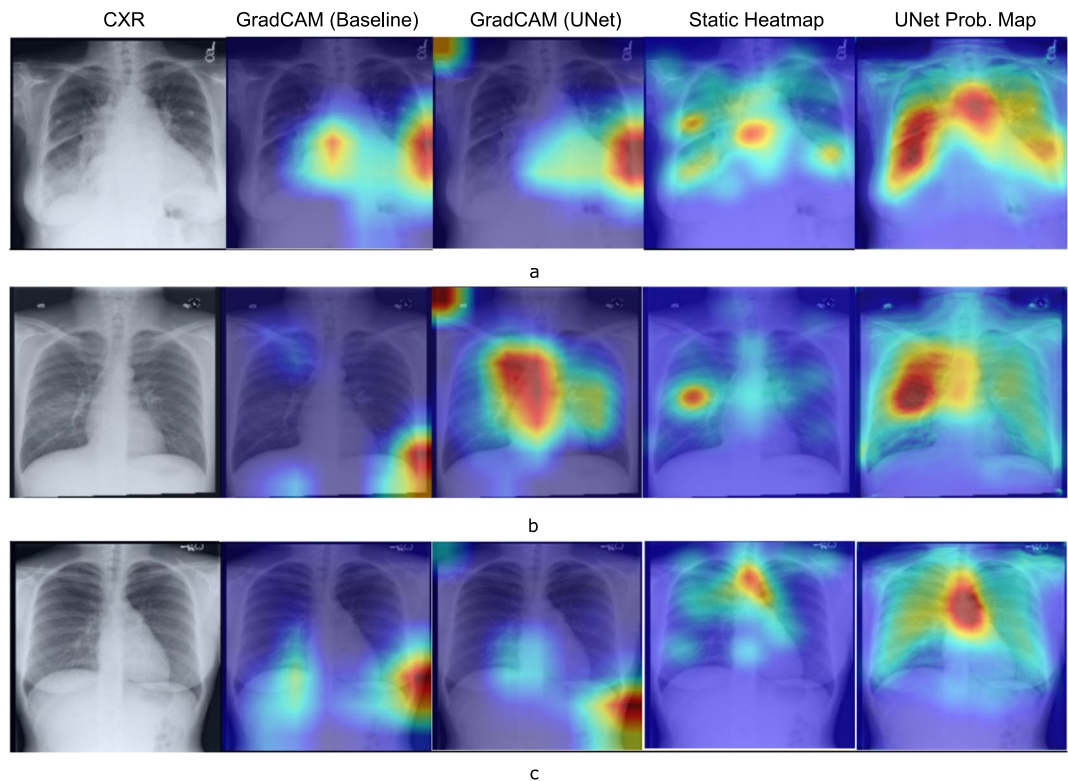
Fig. 17 Qualitative comparison of the interpretability of U-Net based probability maps in comparison with GradCAM for a few example use cases. (**a**) CHF. The physician's eye gaze tends to fall on the enlarged heart and hila, as well as generally on the lungs, (**b**) Pneumonia. The physician's eye gaze predictably focuses on the focal lung opacity and (**c**). Normal. Because the lungs are clear, the physician's eye gaze skips around the image without focus.

current dataset already holds great value to the research community. Given the availability of the images and our code, researchers (including our team) can expand the dataset with more reads per image in the future. In addition, we have shown with preliminary machine learning experiments that a model trained to optimize on a radiologist's eye-tracking pattern has improved diagnostic performance as compared to a baseline model trained with weak image-level labels.

2. The images used during the radiology reading were in 'png' format and not in DICOM. That's because the Gazepoint Analysis UX Edition[15] doesn't support DICOM format. This had the shortcoming that the radiologist could not utilize windowing techniques. However, the png images were prepared using the windowing information in the original DICOM images.

3. This dataset includes only Posterior Anterior (PA) CXR images as selected from the inclusion/exclusion criteria (Fig. 2). This view position criterion was clinically chosen because of its higher quality images compared to Anterior Posterior (AP) CXR images. Therefore, any analysis (e.g., machine learning models trained on only this dataset) may suffer from generalizability to AP CXR images.

Code availability

Our Github repository (https://github.com/cxr-eye-gaze/eye-gaze-dataset) contains code (Python 3) for:

1. Data Preparation

   - Inclusion and exclusion criteria on MIMIC dataset (see details in Inclusion and exclusion criteria section).
   - Case sampling and image preparation for eye gaze experiment (see details in Preparation of images section).

2. Data Post -Processing

   - Speech-to-text on dictation audio (see details in Audio extraction and transcript generation section).
   - Mapping of eye gaze coordinates to original image coordinates (see details in Fixations and eye gaze spreadsheets section).
   - Generate heatmap images (i.e temporal or static) and videos given eye gaze coordinates. The temporal and static heatmap images were used in our demonstrations of machine learning methods in Use of the Dataset in Machine Learning section.

3. Technical Validation
   - Validation of eye gaze fixation quality using calibration images (see details in Validation of eye gaze data).
   - Validation of quality in transcribed dictations (see details in Validation of transcripts section).
   - The t-test for eye gaze fixations for each anatomical structure and condition pairs (see details in Statistical analysis on fixations section)

4. Machine Learning Experiments, as described in Use of the Dataset in machine learning section.
Software requirements are listed in https://github.com/cxr-eye-gaze/eye-gaze-dataset.

References

1. Irvin, J. *et al*. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. *Proceedings of the AAAI Conference on Artificial Intelligence* **33**, 590–597 (2019).
2. Johnson, A. E. *et al*. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific Data* **6** (2019).
3. Bluemke, D. A. *et al*. Assessing Radiology Research on Artificial Intelligence: A Brief Guide for Authors, Reviewers, and Readers-From the Radiology Editorial Board. *Radiology* **294**, 487–489 (2020).
4. Waite, S. A. *et al*. Analysis of perceptual expertise in radiology–Current knowledge and a new perspective. *Frontiers in human neuroscience* **13**, 213 (2019).
5. Van der Gijp, A. *et al*. How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology. *Advances in Health Sciences Education* **22**, 765–787 (2017).
6. Krupinski, E. A. Current perspectives in medical image perception. *Attention, Perception, & Psychophysics* **72**, 1205–1217 (2010).
7. Tourassi, G., Voisin, S., Paquit, V. & Krupinski, E. Investigating the link between radiologists' gaze, diagnostic decision, and image content. *Journal of the American Medical Informatics Association* **20**, 1067–1075 (2013).
8. Khosravan, N. *et al*. A collaborative computer aided diagnosis (C-CAD) system with eye-tracking, sparse attentional model, and deep learning. *Medical image analysis* **51**, 101–115 (2019).
9. Stember, J. N. *et al*. Eye Tracking for Deep Learning Segmentation Using Convolutional Neural Networks. *Journal of digital imaging* **32**, 597–604 (2019).
10. Aresta, G. *et al*. Automatic lung nodule detection combined with gaze information improves radiologists' screening performance. *IEEE Journal of Biomedical and Health Informatics* (2020).
11. Mall, S., Brennan, P. C. & Mello-Thoms, C. Modeling visual search behavior of breast radiologists using a deep convolution neural network. *Journal of Medical Imaging* **5**, 035502 (2018).
12. Goldberger, A. L. *et al*. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *circulation* **101**, e215–e220 (2000).
13. Johnson, A. *et al*. MIMIC-IV. *PhysioNet* https://doi.org/10.13026/A3WN-HQ05 (2020).
14. Gazepoint. GP3 Eye Tracker.
15. Gazepoint. Gazepoint Analysis UX Edition.
16. Wu, J. T. *et al*. AI Accelerated Human-in-the-loop Structuring of Radiology Reports. In *AMIA* (2020).
17. Karargyris, A. *et al*. Age prediction using a large chest x-ray dataset. In Mori, K. & Hahn, H. K. (eds.) *Medical Imaging 2019: Computer-Aided Diagnosis*, vol. 10950, 468–476, https://doi.org/10.1117/12.2512922. International Society for Optics and Photonics (SPIE, 2019).
18. Wu, J. *et al*. Automatic bounding box annotation of chest x-ray data for localization of abnormalities. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 799–803 (2020).
19. Karargyris, A. *et al*. Eye gaze data for chest x-rays. *PhysioNet* https://doi.org/10.13026/QFDZ-ZR67 (2020).
20. Cheng, J., Dong, L. & Lapata, M. Long Short-Term Memory-Networks for Machine Reading. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 551–561 (2016).
21. Vaswani, A. *et al*. Attention is all you need. In *Advances in neural information processing systems*, 5998–6008 (2017).
22. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization.
23. Smith, L. N. Cyclical learning rates for training neural networks. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 464–472 (IEEE, 2017).
24. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* **15**, 1929–1958 (2014).
25. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241 (Springer, 2015).
26. Tan, M. & Le, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946* (2019).
27. Liaw, R. *et al*. Tune: A research platform for distributed model selection and training. *arXiv preprint arXiv:1807.05118* (2018).
28. Selvaraju, R. R. *et al*. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, 618–626 (2017).

Author contributions

clinical design of the study. M.M. contributed to the design of experiments, supervision of the work and writing of the manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.K., S.K. or M.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.