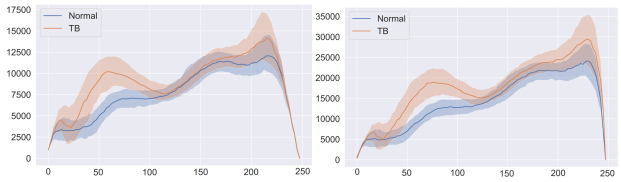
Supplementary Material

Further Figures for Topological Descriptors: In this part, we provide the Betti-0 and Betti-1 figures for TB-Shenzen dataset (Figure 6), and Viral and Bacterial Pneumonia images in Ped-Pneumonia dataset (Figure 7). In TB-Shenzhen, one can see the different patterns especially between 30-80 grayscale intervals.

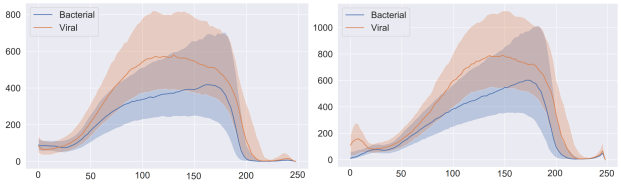


(a) TB-Shenzhen (Betti-0) (b) TB-Shenzhen (Betti-1)

Figure 6: In the figures above, we give the median curves and 40% confidence bands of our topological feature vectors (Betti functions) for each class in TB-Shenzhen datasets. x-axis represents grayscale values and y-axis represents count of components (Betti-0) or count of loops (Betti-1).

On the other hand, we did not see a distinction between bacterial and viral classes. This is not surprising since Bacterial vs Viral pneumonia cannot be determined based on imaging findings alone. In fact, viral pneumonia is often a misnomer. Viral infections generally do not cause lobar or consolidation process in the lungs. They can lead to a superimposed bacterial infection or coinfection resulting in a bacterial pneumonia [67]. Our approach verifies this as topological patterns are very similar in both classes.

CXR images are very useful to detect pneumonia, but it is not that helpful to distinguish the subtypes. One needs to use other diagnostic tools to distinguish viral or bacterial pneumonia[8].



(a) Bacterial vs. Viral (Betti-0) (b) Bacterial vs. Viral (Betti-1)

Figure 7: In the figures above, we give the median curves and 40% confidence bands of our topological feature vectors (Betti functions) for bacterial and viral pneumonia classes for Ped-Pneumonia dataset.

Confidence Bands: As our median curves and confidence bands figures (Figure 1 and Figure 5) indicates, our topological feature vectors successfully capture the topological patterns in the CXR images for normal and abnormal classes. Here, we explain the process of obtaining confidence bands for our Betti curves in these figures. This is a common method called *nonparametric confidence band for the median* [23]. We explain the process in an example. Assume there are 1000 Betti functions $\{\beta_1(t), \beta_2(t), \ldots, \beta_{1000}(t)\}$ where $t \in [a, b]$. For each t_0, rank the values of $\{\beta_1(t_0), \beta_2(t_0), \ldots, \beta_{1000}(t_0)\}$. In particular, for each $1 \leq i \leq 1000$, we define a ranking function $r_i : [a, b] \to \{1, 2, \ldots, 1000\}$ which assign t_0 to the index of i^{th} ranked Betti function, i.e, $\beta_{r_1(t_0)}(t_0) \leq \beta_{r_2(t_0)}(t_0) \leq \cdots \leq \beta_{r_{1000}(t_0)}(t_0)$. We first get the median curve as $\overline{\beta}(t) = \beta_{r_{500}(t)}(t)$. Then, we define 40% confidence bands around the median as the region between the functions $\widehat{\beta}_{300}(t)$ and $\widehat{\beta}_{700}(t)$ where $\widehat{\beta}_{300}(t) = \beta_{r_{300}(t)}(t)$ and $\widehat{\beta}_{700}(t) = \beta_{r_{700}(t)}(t)$. Here, the count 1000, and 40% are arbitrary, and the general definition can be easily adapted to general settings.