Inception-based Deep Learning Architecture for Tuberculosis Screening using Chest X-rays

Dipayan Das  
Dept. of Electronics and Communication Engg.  
National Institute of Technology  
Durgapur, India  
dipayan.das2010@gmail.com

K.C. Santosh  
IEEE Senior Member  
Dept. of Computer Science  
University of South Dakota  
Vermillion, SD 57069, USA  
santosh.kc@ieee.org

Umapada Pal  
Computer Vision & Pattern Recognition Unit  
Indian Statistical Institute Kolkata,  
India umapada@isical.ac.in
Motivation

- Primary need of screening TB (Tuberculosis) in resource constrained regions of the world.

- Chest X-rays are considered as promising indicators for the onset of TB, and there is an existent infrastructure of low-cost chest radiography around the world.

- Lack of skilled radiologists hinders the screening process.

- Computer vision based automated diagnosis systems offers a powerful alternative.

- Deep learning based Convolutional Neural Networks (CNNs) are adept at learning custom features from a given spatial data distribution.
We propose an **end-to-end CNN architecture** that automates the screening process of TB using chest X-rays.

The proposed model is derived from the **Inception Net V3** architecture, which makes use of **multiscale feature extracting modules** (**Inception modules**).

The model learns **TB related visual features** from the chest X-ray data and performs better than the state-of-the-art deep learning methodology on the used datasets.

We study the **computational efficiency** of the model from a **resource perspective** (availability of **GPU** and **model size**).

The effect of input chest X-ray **image size** on model performance is also studied.
Dataset Description

Table I: Dataset composition

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Tuberculosis +ve</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shenzhen, China</td>
<td>342</td>
<td>340</td>
</tr>
<tr>
<td>Montgomery County, USA</td>
<td>58</td>
<td>80</td>
</tr>
</tbody>
</table>

Fig. I: Randomly sampled chest X-rays from the used datasets. Anatomical and recording instrument variability between the datasets can be well observed.
Preprocessing

1. Raw Chest X-ray
2. Mask generated using pre-trained CNN
3. Applying mask & ROI cropping
4. CLAHE (Contrast Equalization)

Fig. II: Preprocessing pipeline

Shenzhen, China chest X-rays
Montgomery County, USA chest X-rays

Fig. III: Sample preprocessed images
Proposed Methodology

Fig. IV: The Inception Module

Fig. V: Truncated Inception Net V3 architecture (Ours).

Fig. VI: Original Inception Net V3 architecture.
## Results and Analysis

### Table II: Shenzhen, China

<table>
<thead>
<tr>
<th>Fold</th>
<th>ACC</th>
<th>AUC</th>
<th>SEN</th>
<th>SPEC</th>
<th>PREC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>93.93</td>
<td>0.96</td>
<td>0.91</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>2</td>
<td>93.93</td>
<td>0.94</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>3</td>
<td>93.93</td>
<td>0.99</td>
<td>0.94</td>
<td>0.93</td>
<td>0.94</td>
</tr>
<tr>
<td>4</td>
<td>88.05</td>
<td>0.94</td>
<td>0.85</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>5</td>
<td>95.45</td>
<td>0.95</td>
<td>0.94</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>6</td>
<td>81.81</td>
<td>0.94</td>
<td>0.75</td>
<td>0.87</td>
<td>0.86</td>
</tr>
<tr>
<td>7</td>
<td>89.55</td>
<td>0.95</td>
<td>0.88</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>8</td>
<td>90.90</td>
<td>0.94</td>
<td>0.87</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>9</td>
<td>96.96</td>
<td>0.97</td>
<td>0.97</td>
<td>0.96</td>
<td>0.97</td>
</tr>
<tr>
<td>10</td>
<td>92.42</td>
<td>0.98</td>
<td>0.87</td>
<td>0.96</td>
<td>0.96</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\mu & = 91.70 & \pm 4.16 \\
\sigma & = 0.96 & \pm 0.02
\end{align*}

### Table III: Montgomery County, USA

<table>
<thead>
<tr>
<th>Fold</th>
<th>ACC</th>
<th>AUC</th>
<th>SEN</th>
<th>SPEC</th>
<th>PREC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>83.89</td>
<td>0.90</td>
<td>0.75</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>2</td>
<td>89.97</td>
<td>0.92</td>
<td>0.80</td>
<td>0.96</td>
<td>0.94</td>
</tr>
<tr>
<td>3</td>
<td>79.16</td>
<td>0.89</td>
<td>0.75</td>
<td>0.89</td>
<td>0.91</td>
</tr>
<tr>
<td>4</td>
<td>96.89</td>
<td>0.95</td>
<td>0.84</td>
<td>0.98</td>
<td>0.96</td>
</tr>
<tr>
<td>5</td>
<td>86.85</td>
<td>0.90</td>
<td>0.79</td>
<td>0.95</td>
<td>0.94</td>
</tr>
<tr>
<td>6</td>
<td>88.73</td>
<td>0.91</td>
<td>0.80</td>
<td>0.96</td>
<td>0.94</td>
</tr>
<tr>
<td>7</td>
<td>82.37</td>
<td>0.89</td>
<td>0.75</td>
<td>0.93</td>
<td>0.92</td>
</tr>
<tr>
<td>8</td>
<td>90.97</td>
<td>0.92</td>
<td>0.82</td>
<td>0.96</td>
<td>0.95</td>
</tr>
<tr>
<td>9</td>
<td>91.11</td>
<td>0.94</td>
<td>0.84</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td>10</td>
<td>84.17</td>
<td>0.90</td>
<td>0.79</td>
<td>0.94</td>
<td>0.94</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\mu & = 87.47 & \pm 4.3 \\
\sigma & = 0.92 & \pm 0.04
\end{align*}

\[
\begin{align*}
\mu & = 0.76 & \pm 0.05 \\
\sigma & = 0.95 & \pm 0.03
\end{align*}
Results and Analysis

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>ACC</td>
<td>83.00</td>
<td>84.00</td>
<td>84.40</td>
<td>86.74</td>
<td>90.00</td>
<td>91.70</td>
</tr>
<tr>
<td></td>
<td>AUC</td>
<td>0.92</td>
<td>0.92</td>
<td>0.90</td>
<td>0.92</td>
<td>0.94</td>
<td>0.96</td>
</tr>
<tr>
<td>USA</td>
<td>ACC</td>
<td>67.00</td>
<td>76.00</td>
<td>79.00</td>
<td>77.14</td>
<td>--</td>
<td>87.47</td>
</tr>
<tr>
<td></td>
<td>AUC</td>
<td>0.88</td>
<td>0.83</td>
<td>0.81</td>
<td>0.90</td>
<td>--</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Table IV: Comparative Study

**Fig. VII:** Distribution of prediction time for 100 randomly sampled CXRs.

**Fig. VIII:** Speed-accuracy trade-off using images of larger sizes.
Results and Analysis

Fig. IX: Grad-CAM visualization of learnt feature maps.

Input chest X-ray images

Grad-CAM activation map overlayed on input images.
An end-to-end CNN model is proposed to screen TB using chest X-ray images.

The model learns TB specific visual/texture features using convolutional kernels of varying sizes within Inception modules. This enables a multiscale analysis of the chest X-rays.

Contrary to ensemble methods, the proposed model is computationally lighter while maintaining superior performance.

The model is also robust to dataset variability like anatomical and chest X-ray recording instrument parameters.

The proposed methodology provides a fast yet accurate procedure to undertake mass screening of TB in resource constrained regions of the world.


