EXTRACTING SPO2 BY PULSE OXIMETRY , X-RAY IMAGE PROCESSING, ALONG WITH COMBINING CLINICAL SYMPTOMS AND MACHINE LEARNING TO DETECT PNEUMONIA USING NON-INVASIVE METHOD

A.B.M. Zawadul Islam
Bangladesh University of Engineering and Technology
Email: islamzawadul@gmail.com

Md. Shakhrul Iman Siam
Bangladesh University of Engineering and Technology
Email: Shakhrulsiam268@gmail.com

Md. Hasibur Rahman
Bangladesh University of Engineering and Technology
Email: hasibrahman471@gmail.com

Rahad Arman Nabid
Bangladesh University of Engineering and Technology
Email: ran.nabid@gmail.com

Md. Kamrul Hasan
Bangladesh University of Engineering and Technology
Email: mkh.kamrulhasan@gmail.com

Nafhatul Zinan Shifa
Bangladesh University of Engineering and Technology
Email: shifazinann@gmail.com
]

Dr. Celia Shahnaz
Bangladesh University of Engineering and Technology
Email: celia@eee.buet.ac.bd

Arik Subhana
Bangladesh University of Engineering and Technology
Email: arik@eee.buet.ac.bd

***Abstract*—Pneumonia is one of the most common infectious diseases in the lungs especially attacked infant and elderly people which causing cough with phlegm or pus, fever, chills, difficulty breathing, fill lungs with pus, and sometimes leads to deaths in some case. In Bangladesh, pneumonia is responsible for around 28% of the deaths of children under five years of age. Around 50,000 children die of pneumonia every year because of not having proper medical doctors, especially in rural areas. But early detection can help to prevent this disease. Developing a device to detect pneumonia in a rapid and non-invasive way, which will be affordable for Bangladeshi people is our aim. Helping medical professionals to detect pneumonia early to facilitate the treatment of the patients is our motivation. That why an affordable, rapid detection model is designed. In our device, we used machine learning techniques using a non-invasive technique method by collecting x-ray images, spo2 level along with clinical symptoms that can help to detect pneumonia in absence of a doctor. Moreover, only an x-ray or only spo2 can not specify a pneumonia patient. But our model takes all the info and is gathered in machine learning model to detect the chance of having pneumonia with great accuracy with respect to other proposed models.**

***Index Terms*—CNN, Non-Invasive, SpO2**

I. INTRODUCTION

Pneumonia is an inflammatory condition of the lung primarily affecting the small air sacs known as alveoli [1][2]. Each year, pneumonia affects about 450 million people globally (7% of the population) and results in about 4 million deaths [3][4]. With the invention of modern antibiotics and vaccines, the survival rate has increased. Nevertheless, pneumonia remains a leading cause of death in developing countries, and also among the very old, the very young, and the chronically ill [3][5]. Detection of pneumonia in the early stage can eventually cause less harm to the patients especially infants and old persons who are mostly at risk. Chest X-rays are currently the best available method for diagnosing pneumonia [6], playing a crucial role in clinical care [7].

Detection of pneumonia from chest x-ray needs an experienced radiologist and it may be time-consuming sometimes which may lead the patient to a critical situation. Along with the chest x-ray, some vital symptoms of the patient are also taken into consideration for determining pneumonia. There is a known variability between radiologists in the interpretation of chest radiographers [8]. To improve the efficiency and accuracy of diagnostic services computer-aided diagnosis systems for pneumonia detection have been widely exploited in the last decade [9,10,11].

Convolution Neural Network, a deep learning algorithm, is normally used to detect pneumonia from a chest x-ray.

Fig. 1. Normal chest x-ray

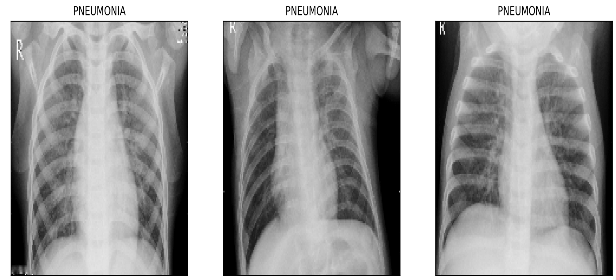

Fig. 2. Pneumonia chest x-ray

Recently, a transfer learning-based algorithm is also used in detecting pneumonia. In case of detecting pneumonia, radiologist only gives decision derived from the chest x-ray images which doesn't include any other symptoms or oxygen saturation measurements, both of them are also very important for detecting pneumonia efficiently.

Oxygen Saturation plays a crucial role in determining the severity of pneumonia patients. In this article, we tried to detect pneumonia using a low-cost oximeter to determine oxygen saturation, and patients' physical symptoms are taken into account to get better accuracy and efficiency in detecting pneumonia using a Convolution Neural Networking (CNN) and Random Forest with machine learning approach.

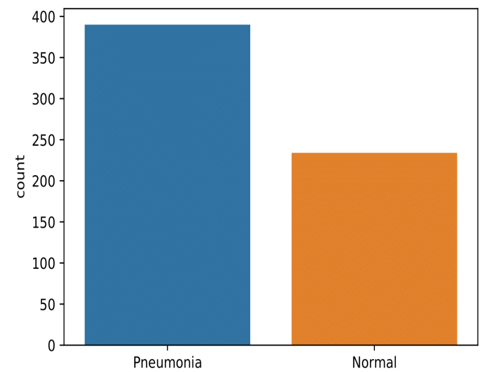

Fig. 3. Test Set

II. Dataset

The first dataset has 5,863 X-Ray images (JPEG) and two categories (Pneumonia/Normal). Fig.1 and Fig.4 show the distributions of the test and the train set. Chest X-ray images (anterior-posterior) were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients' routine clinical care. For the analysis of chest x-ray images, all chest radiographs were initially screened for quality control by removing all low quality or unreadable scans. The diagnoses for the images were then graded by two expert physicians before being cleared for training the AI system. To account for any grading errors, the evaluation set was also checked by a third expert. This dataset is used for our first model training where we train the dataset for detection of the x-ray images of pneumonia patients. Moreover, our second dataset having 624 x-ray images with patients oxygen saturation and some symptoms such as chest pain, cough, fatigue, nausea, fever, vomiting, diarrhea, shortness of breath from Bangladesh Specialized Hospital, Shamoli which we have used for first model testing and second model's testing and training. If any patient has any kinds of symptoms we take it as a binary one otherwise the symptom has zero, corresponding to the symptom.

III. Proposed methodology

According to WHO guidelines, three basic steps are followed to detect pneumonia. One, by the physical symptoms; two, physical examination and three by ordering diagnostics. In our project, all these steps are being followed. After finding out the symptoms, physical examination thus blood saturation level, and X-ray which is a well-known diagnosis tool are taken into account. As we know, the blood saturation level directly links with lung performance, and pneumonia is an infection that affects the lungs. Image processing of the x-ray images strengthens our prediction. A neural network is designed to train itself to correct its prediction mechanism. We have used two models one for detecting x-ray images whether it is pneumonia or normal. Then we add other parameters such as oxygen saturation, chest pain, cough, fatigue, nausea, fever, vomiting, diarrhea, shortness of breath for the same patients to train in random forest model Fig. 5. Patients oxygen saturation, symptoms along with x-ray image gives us a good overview of prediction whether it is pneumonia or not. Moreover, we prefer a low-cost pulse oximeter that can be used as a low-cost device for the helping hand of rural areas of third world countries.

A. CNN architecture

We used the Convolution Neural Network (CNN) architecture for detecting x-ray images. Our CNN architecture consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consist of a series of 2D convolution layers that convolve with multiplication or other dot product. Then we used batch normalization to normalize the layer. Following this, we use the max-pooling layer. The activation function is a ReLU layer and is subsequently

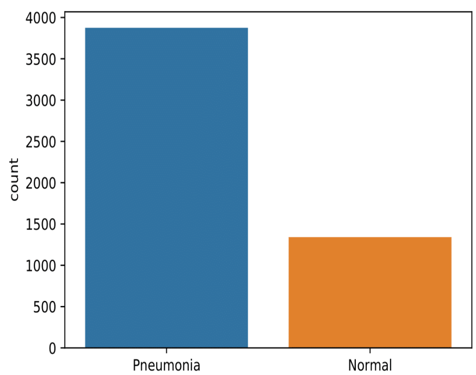

Fig. 4. Train Set

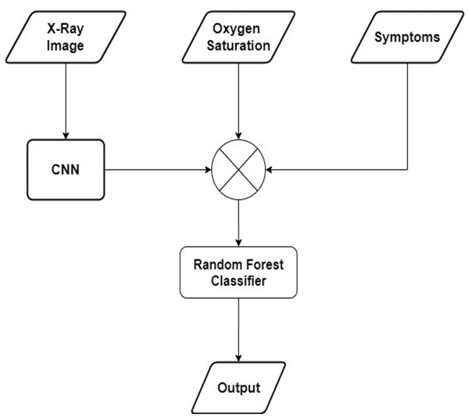

Fig. 5. Flow-Chart of proposed methodology

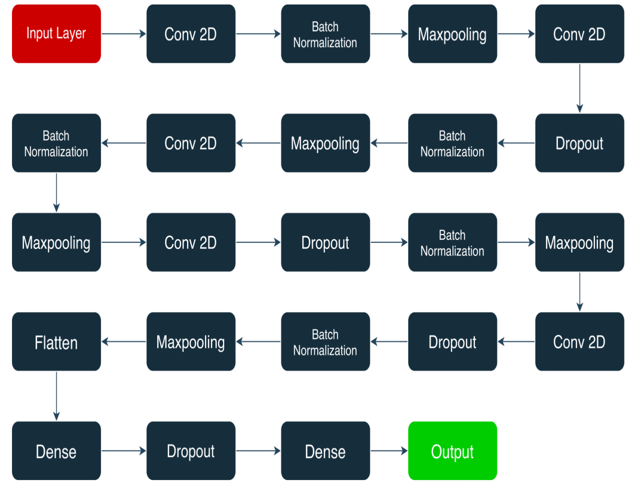

Fig. 6. Proposed CNN architecture

TABLE I
RESULTS COMPARISON

Item Name	Quantity	Price(BDT)
Raspberry Pi 4 Computer	1	4400
3.2inch 320x240 Touch LCD	1	1400
st1kl3b	1	50
RED LED	5	10
IR LED	2	10
Battery, Charger	1	1000
Resistance	several	100
Capacitance	several	20
CA3130 OPAMP	1	66
MAX30105(For checking our data)	1	1561

followed by additional convolutions such as pooling layers, fully connected layers, and normalization layers, referred to as hidden layers because their inputs and outputs are masked by the activation function and final convolution. Summary of our architecture is shown in Fig. 6

B. Low Cost Spo2 Unit

Developing a device to detect pneumonia in a rapid and non-invasive way, which will be affordable for Bangladeshi people is our aim. Helping medical professionals to detect pneumonia early to facilitate the treatment of the patients is our motivation. That why an affordable, rapid detection model is designed Fig. 7. This circuit is low cost having 8617 BDT that can be afforded by the village dispensary. Details price and component list are given in TABLE I.

C. Random Forests Model

Random Forests is an ensemble learning method for classification, regression, and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean/average prediction (regression) of the individual trees. We used the random forest model for our second model where we take features from x-ray predictions, some symptoms such as chest pain, cough, fatigue, nausea, fever, vomiting, diarrhea, shortness of breath as a binary number, and float value of oxygen saturation.

IV. EXPERIMENTATION AND RESULTS

A. Data Processing

Without prepossessing, our dataset may cause an over-fitting problem which may reduce model accuracy in validation and test set. Moreover, big-size images are not easy to train with moderate computational power. Before starting the Training process for our first CNN model all the images were resized to 150*150 pixels to reduce computational complexity. Some data augmentation methods like random rotation in a range of certain angles, random zoom, horizontal and vertical shift, horizontal flip were performed to introduce variability in the training dataset.

For our second model we take input binary one from the CNN model's output if it is pneumonia, otherwise zero from our second x-ray dataset. Moreover, we take symptoms as binary inputs and float values for oxygen saturation. Before training the second model, we normalize our dataset for giving each feature the same weight.

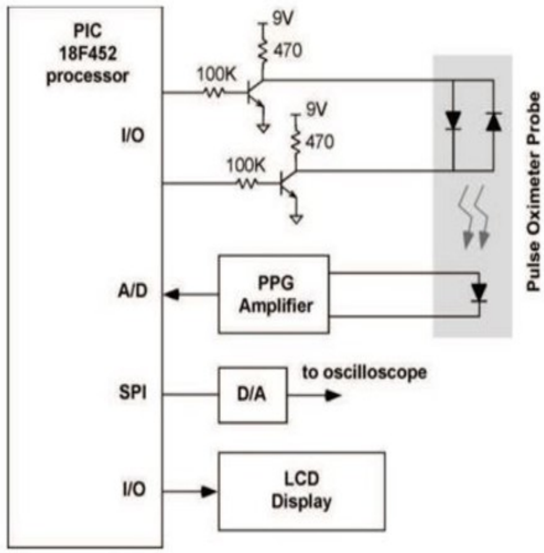

Fig. 7. Spo2 Detection Device Circuit

TABLE II
RESULTS COMPARISON

Algorithms	**Accuracy**	**Sensitivity**	**Specificity**	**AUC**
Reference Model	92.8	93.2	90.1	96.8
Proposed CNN Model(Stage-1)	93.109	97.436	85.897	97.44
Proposed Random Forest Model(Stage-2)	95.20	99.3	88.9	99.44

B. Result

The proposed CNN model was trained with 5232 x-ray images (3883 Pneumonia, 1349 Normal) and tested with 624 x-ray images (390 pneumonia, 234 normal). Accuracy as well as Sensitivity, Specificity, Area Under ROC was measured for the test dataset, which is shown in table-1. We achieved high sensitivity (97.4%) which indicates the possibility of false-negative cases is very low. Fig. 8 and Fig. 9 shows confusion matrix of predictions and Receiver Operating Characteristics (ROC) curve respectively. We have found our model performs better concerning reference paper in chest x-ray detection.

After getting the output from the test image dataset from the CNN model, we trained our second model with 624 data (390 pneumonia, 234 normal) having binary feature values for every symptom and float values from our low-cost spo2 device. In this Random forest model, we split the dataset 60% train Set (374 profiles) 40% test Set (250 profiles). We found high accuracy (95.20%), sensitivity (99.3%), specificity (88.9%), AUC (99.44%) which indicates that the model has boosted its performance from our first model and predicts more accurately Fig. **??**. We have shown the comparison of our two models and the reference model in TABLE II.

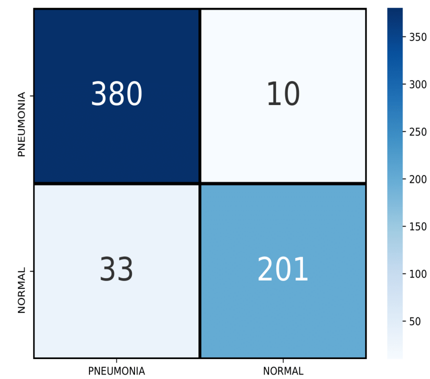

Fig. 8. Confmatrix for CNN Model

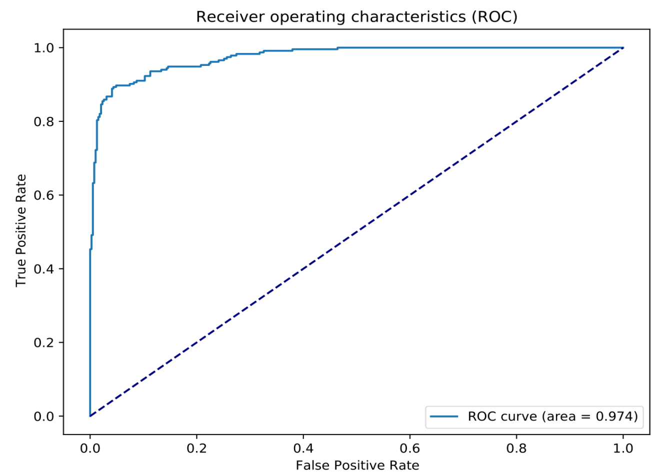

Fig. 9. ROC curve

V. CONCLUSION

Developing a device to detect pneumonia in a rapid and non-invasive way, which will be affordable for Bangladeshi people is our aim. Helping medical professionals to detect pneumonia early to facilitate the treatment of the patients

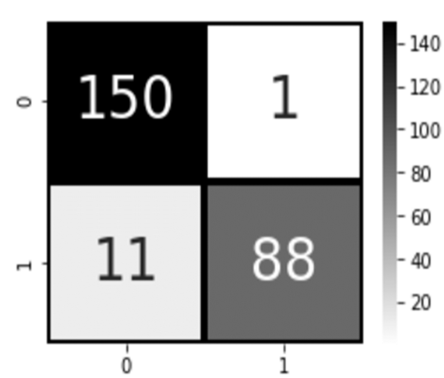

Fig. 10. Confmatrix for Random Forest Model

is our motivation. That why an affordable, rapid detection model is designed. This circuit is low cost having 8617 BDT that can be afforded by the village dispensary. Adding a machine learning model makes our cheap device more accurate and effective for delectation which can be used as an IOT device for mass use in an undeveloped remote area. Our two models where we used one for x-ray image detection and the decisions of CNN model used in the second Random Forest model along with some symptoms such as chest pain, cough, fatigue, nausea, fever, vomiting, diarrhea, shortness of breath and oxygen saturation makes our proposed model framework more robust and practically feasible for IoT device. Moreover, we believe that our model can perform well in other data set as well for detecting pneumonia with a non-invasive process. Moreover, a data set with a large amount of x-ray, oxygen saturation, and clinical symptoms and improve the efficiency of our model.

ACKNOWLEDGMENT

This work is made possible through continuous consultation and cross-checking dataset by Dr. Masud Parvez, Pathology, (Dhaka Medical College), M.B.B.S (Chittagong Medical College).

REFERENCES

[1] McLuckie A, ed. (2009). Respiratory disease and its management. New York: Springer. p. 51. ISBN 978-1-84882-094-4.

[2] 2.Leach RE (2009). Acute and Critical Care Medicine at a Glance (2nd ed.). Wiley-Blackwell. ISBN 978-1-4051-6139-8.

[3] Ruuskanen O, Lahti E, Jennings LC, Murdoch DR (April 2011). "Viral pneumonia". Lancet. 377 (9773): 1264–75. doi:10.1016/S0140-6736(10)61459-6. PMC 7138033. PMID 21435708.

[4] Lodha R, Kabra SK, Pandey RM (June 2013). "Antibiotics for community-acquired pneumonia in children". The Cochrane Database of Systematic Reviews. 6 (6): CD004874. doi:10.1002/14651858.CD004874.pub4. PMC 7017636. PMID 23733365.

[5] George RB (2005). Chest medicine : essentials of pulmonary and critical care medicine (5th ed.). Philadelphia: Lippincott Williams Wilkins. p. 353. ISBN 978-0-7817-5273-2.

[6] WHO. Standardization of interpretation of chest ra-diographs for the diagnosis of pneumonia in chil-dren. 2001.

[7] Franquet, T. Imaging of pneumonia: trends and algorithms. European Respiratory Journal, 18(1):196–208, 2001.

[8] Mark I. Neuman, Edward Y. Lee, Sarah Bixby, Stephanie Diperna, Jeffrey Hellinger, Richard Markowitz, Sabah Ser-vaes, Michael C. Monuteaux, and Samir S. Shah. Variability in the interpretation of chest radiographs for the diagnosis of pneumonia in children

[9] Norliza Mohd. Noor, Omar Mohd. Rijal, Ashari Yunus, and S.A.R. Abu-Bakar. A discrimination method for the detection of pneumonia using chest radiograph. Computerized Medical Imaging and Graphics, 34(2):160–166, mar 2010.

[10] Leandro Luıs Galdino Oliveira, Simonne Almeida e Silva,Luiza Helena Vilela Ribeiro, Renato Maurıcio de Oliveira,Clarimar José Coelho, and Ana Lúcia S. S. Andrade.Computer-aided diagnosis in chest radiography for detection of childhood pneumonia. International Journal of Medical Informatics, 77(8):555–564, aug 2008

[11] John R. Zech, Marcus A. Badgeley, Manway Liu, Anthony B. Costa, Joseph J. Titano, and Eric Karl Oermann. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. PLOS Medicine, 15(11):e1002683,nov 2018.

[12] Adamantia Liapikou, Catia Cillóniz Antoni Torres (2019) Emerging strategies for the noninvasive diagnosis of nosocomial pneumonia, Expert Review of Anti-infective Therapy, 17:7, 523-533, DOI: 10.1080/14787210.2019.1635010

[13] M. Tamayo, A. Westover and Y. Sun, "Microcontroller based pulse oximeter for undergraduate capstone design," Proceedings of the 2010 IEEE 36th Annual Northeast Bioengineering Conference (NEBEC), New York, NY, 2010, pp. 1-2, doi: 10.1109/NEBC.2010.5458198.

[14] M. Tamayo, A. Westover and Y. Sun, "Microcontroller based pulse oximeter for undergraduate capstone design," Proceedings of the 2010 IEEE 36th Annual Northeast Bioengineering Conference (NEBEC), New York, NY, 2010, pp. 1-2, doi: 10.1109/NEBC.2010.5458198.