Sensor signal processing

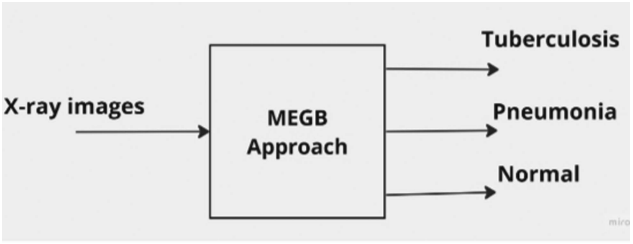
A Novel Approach for the Detection of Tuberculosis and Pneumonia Using Chest X-Ray Images for Smart Healthcare Applications

Subrat Kumar Kabi[1], Rajesh Kumar Tripathy[2*], Dipti Patra[1**], and Ganapati Panda[3***]

[1]*Department of Electrical Engineering, National Institute of Technology (NIT), Rourkela 769008, India*
[2]*Department of Electrical and Electronics Engineering, BITS-Pilani, Hyderabad Campus, Secunderabad 500078, India*
[3]*Department of Electronics and Communication Engineering, CV Raman Global University, Bhubaneswar 752054, India*
[*]*Member, IEEE*
[**]*Senior Member, IEEE*
[***]*Life Senior Member, IEEE*

Manuscript received 25 July 2023; revised 14 September 2023; accepted 21 October 2023. Date of publication 25 October 2023; date of current version 16 November 2023.

Abstract—The early discrimination of pneumonia and tuberculosis is difficult for radiologists due to similar pathological symptoms in the chest X-ray images. Therefore, there is a requirement for automated methods to detect and classify these two diseases accurately. This letter proposes the multiscale eigendomain gradient boosting (MEGB)-based approach to detect pneumonia and tuberculosis from chest X-ray images. The discrete wavelet transform is employed to evaluate subbands of chest X-ray images at different decomposition levels. The singular value decomposition (SVD) is utilized in each subband to evaluate singular values, left eigenmatrix, and right eigenmatrix, respectively. The maximum value of each column of both left and right eigenmatrices and singular values for each subband of the X-ray image are used as features. All subband eigendomain feature vectors are concatenated and given to the light gradient boosting model to detect pneumonia and tuberculosis diseases. The performance of the proposed MEGB approach-based detection is evaluated using chest X-ray images from a publicly available database. The suggested MEGB approach has achieved an accuracy value of 96.42%. The suggested approach performs better than the transfer learning and other reported methods to detect pneumonia and tuberculosis using chest X-ray images.



Index Terms—Sensor signal processing, accuracy, light gradient boosting model (LGBM), multiscale eigendomain features, pneumonia, tuberculosis.

I. INTRODUCTION

The Internet of Things (IoT) is vital in developing intelligent healthcare systems for patient monitoring and diagnosing various diseases using physiological signals and medical images [1]. Tuberculosis is a bacterial infection disorder that affects the person's lungs [2]. Similarly, pneumonia disease is a type of respiratory infection that occurs due to viruses or bacteria [3]. Chest X-ray imaging is widely used in the clinical standard to screen pneumonia, tuberculosis, and other thoracic diseases [3], [4]. Portable radiography imaging has been utilized to diagnose COVID and other diseases [5]. The pathological changes, such as the enhanced bronchovascular markings, consolidation in lungs, and alveolar infiltrates (affected areas of the lung look cloudy) in the chest X-ray images are used to detect pneumonia. Similarly, for tuberculosis, cavitary lesions (dark and fluid-filled regions within lung opacities), pleural effusion, and right-sided infiltration are observed in the chest X-ray images. Chest X-ray images are frequently generated in hospitals with large numbers of patients. It is a time-consuming task for radiologists to manually investigate the chest X-ray of each subject to diagnose tuberculosis and other thoracic diseases [4]. Therefore, automated approaches based on artificial intelligence (AI) algorithms are used to assist radiologists in predicting the type of disease from the chest X-ray images [4]. Developing novel AI-based methods for the automated detection of different thoracic diseases from chest X-ray images is important for IoT-based smart healthcare applications.

In recent years, various automated methods have been proposed to detect pneumonia and tuberculosis using chest X-ray images [3], [4], [6], [7]. The method reported in [8] has extracted the shape and texture-based features of chest X-ray images and used the support vector machine (SVM) classifier to detect tuberculosis disease. Similarly, Rahman et al. [9] used various transfer learning models to classify normal and tuberculosis using chest X-ray images. Pasa et al. [10] have considered a hybrid deep convolutional neural network (CNN) and SVM-based model to detect tuberculosis using chest X-ray images. Moreover, Zhang et al. [11] have used a ten-layer deep CNN model and different transfer learning models to detect pneumonia using chest X-ray images. They have obtained higher accuracy using their 10-layer based deep CNN model than the transfer learning models. Kundu et al. [12] have considered combining GoogLeNet, ResNet-18, and DenseNet-121-based transfer learning models to detect pneumonia using chest X-ray images. The abovementioned methods have considered only the normal versus tuberculosis and normal versus pneumonia classification schemes using chest X-ray images. Few methods have

Corresponding author: Subrat Kumar Kabi (e-mail: kabi.subrat@gmail.com).
Associate Editor: D. Senkal.
Digital Object Identifier 10.1109/LSENS.2023.3327580

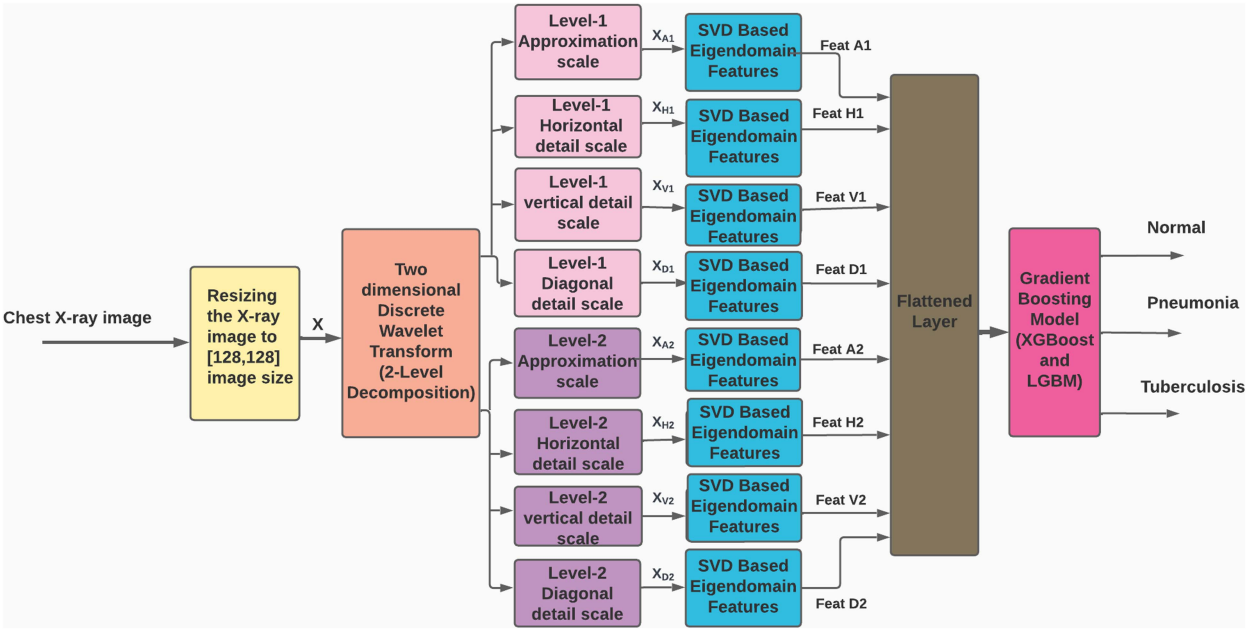


Fig. 1. Flowchart of the proposed MEGB approach to detect pneumonia and tuberculosis.

formulated the normal versus tuberculosis versus pneumonia classification strategy using chest X-ray images. Among them, Mogaveera et al. [13] have implemented a 12-layer deep CNN and obtained an accuracy value of 90% to detect tuberculosis and pneumonia diseases using chest X-ray images. Similarly, Abubakar et al. [14] have used the nine-layer deep CNN and different transfer learning models, such as VGG16 and InceptionNet-based models, to detect tuberculosis and pneumonia. They have obtained the maximum accuracy value of 92.97% using the 9-layer deep CNN model. The deep learning models have less classification performance for the detection of both tuberculosis and pneumonia diseases using chest X-ray images. Also, these models are not explainable as the layers in the CNN model produce random feature maps, which do not reveal the information of chest X-ray images for the automated detection of different diseases. Therefore, an explainable model with higher classification performance is required to detect pneumonia and tuberculosis pathologies using chest X-ray images automatically.

The multiscale analysis of chest X-ray images using discrete wavelet transform (DWT) helps to obtain the subbands at different frequency scales [15]. These subbands are interpretable feature maps that capture grossly segregated information from chest X-ray images at different scales. The singular value decomposition (SVD) is a matrix factorization approach to evaluate eigenmatrices and singular matrix [16]. The SVD has been used to extract features from the chest X-ray images [16]. The SVD has not been explored in the multiscale domain information or the subbands of chest X-ray images to detect pneumonia and tuberculosis diseases. Similarly, the light gradient boosting machine (LGBM) performs better for various biomedical applications than the other machine learning-based models [17], [18]. The novelty of this letter is to develop a new approach by considering the multiscale domain eigendecomposition and LGBM for the automated detection of pneumonia and tuberculosis using chest X-ray images. The rest of this letter is organized as follows. In Section II, the databases for chest X-ray images are described. The proposed MEGB approach is discussed in Section III. The results obtained in this work are presented and discussed in Section IV. Finally, Section V concludes this letter.

II. CHEST X-RAY IMAGE DATABASE

In this letter, we have used chest X-ray images from two publicly available databases to evaluate the performance of the proposed MEGB approach to detect pneumonia and tuberculosis diseases. The first database, which is available at DB1 contains 700 and 3500 chest X-ray images for tuberculosis and normal classes [9]. Similarly, the second database, which is available at DB2 consists of 4338 normal and 1525 pneumonia-based chest X-ray images [19], [20]. In this letter, we have considered 700 normal, 700 tuberculosis, and 700 pneumonia cases to evaluate the proposed MEGB approach. The sizes of chest X-ray images are different for both databases. In this work, we have resized each X-ray image into [128, 128] to reduce the computational time in the feature extraction stage.

III. PROPOSED METHOD

The flowchart depicting the architecture of the proposed MEGB approach to detect tuberculosis and pneumonia using chest X-ray images is shown in Fig. 1.

A. Multiscale Eigenanalysis of Chest X-Ray Images

In this work, the 2-D-DWT with the mother wavelet as Daubechies 4 (db4) is employed for multiscale analysis or decomposing X-ray images into subbands [15]. The two-level-based decomposition of the chest X-ray image is performed. Eight subband matrices are obtained after the two-level decomposition of the X-ray image using 2-D-DWT. The 2-D-DWT produces approximation **A1**, horizontal detail **H1**, diagonal detail **D1**, and vertical detail **V1** subbands in the first level decomposition of the chest X-ray image. Similarly, in the second level, the approximation subband of the first level is decomposed into new approximation (**A2**), new vertical detail (**V2**), new diagonal detail (**D2**), and new horizontal detail (**H2**) subbands. The eigenanalysis of each subband is performed using SVD. The SVD of the kth subband is given by $[\mathbf{U}_k, \mathbf{D}_k, \mathbf{V}_k] = \text{SVD}(S_k)$ [16], where S_k is the kth subband image matrix of the X-ray image. The S_k can

Table 1. Classification Results of All Classifiers Using Selected Multiscale Eigenspace Features With Hold-Out Validation

Classifiers	Acc (%)	Pre (%)	Rec (%)	F1-score	Kappa
KNN	89.76	90.00	90.00	0.900	0.846
DT	85.23	86.00	85.00	0.850	0.778
RF	90.47	91.00	91.00	0.910	0.857
XGBoost	95.95	96.00	96.00	0.960	0.939
LGBM	**96.42**	97.00	96.00	0.960	0.946
ERT	93.09	93.00	93.00	0.930	0.896

Table 2. Classification Results of the LGBM Model Using Multiscale Eigenspace Features From Different Subbands of Chest X-Ray Images With Hold-Out Validation

Features	Acc (%)	Pre (%)	Rec (%)	F1-score	Kappa
Feat A1	88.09	88.00	88.00	0.880	0.821
Feat H1	81.66	82.00	82.00	0.820	0.725
Feat V1	76.66	77.00	77.00	0.770	0.650
Feat D1	72.38	72.00	72.00	0.720	0.585
Feat A2	89.52	90.00	90.00	0.900	0.842
Feat H2	68.09	68.00	68.00	0.680	0.521
Feat V2	72.14	72.00	72.00	0.720	0.582
Feat D2	64.76	65.00	62.00	0.650	0.472
All features	96.19	96.00	96.00	0.960	0.942
Selected features	96.42	97.00	96.00	0.960	0.946

be written as $S_k \in \{\mathbf{A1}, \mathbf{H1}, \mathbf{D1}, \mathbf{V1}, \mathbf{A2}, \mathbf{H2}, \mathbf{D2}, \mathbf{V2}\}$. Similarly, the matrices $\mathbf{U}_k$, $\mathbf{D}_k$, and $\mathbf{V}_k$ are called the eigentriples (left eigenmatrix, diagonal matrix, and right eigenmatrix) for the kth subband of the chest X-ray image. The eigentriples are evaluated for all eight subband matrices of chest X-ray images.

In this letter, we have evaluated the maximum value of each column of the left eigenmatrix of the kth subband and used as features (Feat1). Similarly, the maximum value of each column of the right eigenmatrices of all subbands is evaluated and employed as features (Feat2). The singular values of all subbands are also used as features (Feat3). Therefore, a total of 1248 features (All features) are evaluated from each X-ray image by considering all subbands evaluated using multiscale eigendecomposition. Similarly, the individual subband eigendomain feature vectors are evaluated and these are denoted as FeatA1 (size as 201), FeatH1 (size as 201), FeatV1 (size as 201), FeatD1 (size as 201), FeatA2 (size as 111), FeatH2 (size as 111), FeatV2 (size as 111), and FeatD2 (size as 111), respectively. The analysis of variance (ANOVA) test-based feature selection [21] is employed to select relevant features out of all 1248 multiscale eigendomain features of chest X-ray images.

B. Detection of Pneumonia and Tuberculosis

In this letter, the LGBM classifier is used for the automated detection of pneumonia and tuberculosis diseases using multiscale eigendomain features of chest X-ray images. The LGBM is a leaf-wise split-based decision tree model in which the gradient-based side sampling and exclusive feature bundling-based methods are used to improve the computational efficiency and accuracy of the prediction [22]. The feature matrix computed using chest X-ray images' feature vectors is $\mathbf{F} \in R^{2100\times1248}$. Similarly, the class label vector is $\mathbf{y} \in R^{2100\times1}$. The elements of the class label vector belong to either class 1 (normal), class 2 (pneumonia), or tuberculosis (class 3). In this letter, we have used hold-out validation and tenfold cross-validation strategies [17] to select training and test chest X-ray image instances for the LGBM classifier. The learning rate and the maximum depth parameters in the LGBM classifier are 0.15 and 4, respectively. The optimal values of these parameters are chosen based on the grid-search method [18]. The other classifiers, such as the K-nearest neighbor (KNN), decision tree (DT), random forest (RF), cross gradient boosting (XGBoost), and extreme random trees (ERT) [23] are used, and the performances of these classifiers are compared with the LGBM model. We have used measures, such as accuracy (Acc), precision (Pre), recall (Rec), F1-score, and kappa [23], to evaluate the classification performance of LGBM and other classifiers.

IV. RESULTS AND DISCUSSION

This section shows the classification results obtained using the proposed MEGB approach to detect pneumonia and tuberculosis and their analyses. We have performed the ANOVA test to evaluate the statistical significance of the proposed multiscale eigendomain features of chest X-ray images. Notably, out of 1248 features (all scale eigendomain features), 841 features have p-values less than 0.001, and these 841 features are obtained as significant to detect pneumonia and tuberculosis-based thoracic diseases. The number of features selected in the A1, D1, V1, and H1 subbands of level 1 are 130, 152, 146, and 153, respectively. Similarly, 69, 46, 82, and 62 features are selected in the A2, D2, V2, and H2 subbands after the ANOVA test. In Table 1, we have given the classification results of all six models using selected 841 features of chest X-ray images to detect tuberculosis and pneumonia classes. It is observed that both XGBoost and LGBM models have demonstrated overall accuracy values of more than 95% in detecting tuberculosis and pneumonia classes. The LGBM classifier has obtained an accuracy value of 96.42%, which is the highest compared with the other five classifiers. The RF and ERT classifiers have produced overall accuracy values of 90.47% and 93.09%, respectively. Similarly, the highest Kappa score value of 0.946 has been obtained using the LGBM classifier. The number of true positives for normal, pneumonia, and tuberculosis classes evaluated using the LGBM classifier are 137, 130, and 138, respectively. The LGBM classifier has obtained more true positives than the XGBoost and other classifiers. The LGBM is faster and more memory efficient than that of the XGBoost model [22]. The accuracy values of LGBM classifiers are obtained as 94.52%, 96.12%, and 95.71%, respectively, using one, two, and three wavelet decomposition levels based features (804, 1248, and 1512 features for 1-level, 2-level, and 3-level DWT decomposition cases) extracted from chest X-ray images to detect pneumonia and tuberculosis pathologies. Hence, we have chosen the optimal decomposition level as two for the DWT-based analysis of X-ray images.

The classification results of the LGBM model using different subband features of chest X-ray images are given in Table 2. It is observed that second-level approximation band features (Feat **A2**) coupled with the LGBM model have higher accuracy as compared with other subband-based features of chest X-ray images to detect tuberculosis and pneumonia diseases. The LGBM model with all 1248 multiscale domain eigenspace features has obtained an accuracy value of 96.19% with the hold-out validation strategy. The accuracy value of the LGBM classifier is improved using ANOVA-based selected features compared with all multiscale domain eigenspace features of chest X-ray images. Moreover, we have evaluated the classification results of the LGBM model using all and selected multiscale eigenspace features with tenfold CV to detect pneumonia and tuberculosis using chest X-ray images and these results are shown in Table 3. It is noticed that the LGBM classifier has produced higher average accuracy using all features compared with the selected features of chest X-ray images with a tenfold CV strategy.

We have compared the classification performance of the proposed MEGB approach with different transfer learning-based methods (XceptionNet, VGG16, ResNet50, InceptionV3, and DenseNet121) [9] in Table 4 to detect pneumonia and tuberculosis diseases using

Table 3. Classification Results of LGBM Model Using All and Selected Multiscale Eigenspace Features From Chest X-Ray Images With 10-Fold CV

Features	Acc (%)	Pre (%)	Rec (%)	Kappa
All features	95.67 ± 1.86	95.45 ± 1.66	95.37 ± 1.72	0.93 ± 0.02
Selected features	95.18 ± 1.81	95.26 ± 1.76	95.18 ± 1.81	0.92 ± 0.02

Table 4. Classification Results of Different Transfer Learning Models to Detect Tuberculosis and Pneumonia Using Chest X-Ray Images With Hold-Out Validation

Classifiers	Acc (%)	Pre (%)	Rec (%)	F1-score	Kappa
XceptionNet [9]	91.46	92.06	91.27	0.912	0.871
VGG16 [9]	95.73	95.64	95.64	0.956	0.935
ResNet50 [9]	95.84	95.70	95.98	0.952	0.928
InceptionV3 [9]	85.78	87.20	85.45	0.851	0.786
DenseNet121 [9]	94.97	94.90	94.88	0.948	0.929
Proposed MEGB work	96.42	97.00	96.00	0.960	0.946

Table 5. Comparison With Existing Techniques to Detect Pneumonia and Tuberculosis

Methods used	Accuracy
12-layer based deep CNN model [12]	90%
9-layer based deep CNN model [13]	92.97%
Proposed MEGB approach	96.42%

hold-out validation with the same database chest X-ray images. It is observed that the proposed MEGB approach has obtained higher overall accuracy as compared with the transfer learning models. Furthermore, we have also compared the classification accuracy of the MEGB-based approach with existing deep CNN-based methods [13], [14] in Table 5 to detect pneumonia and tuberculosis diseases using chest X-ray images. It is found that the MEGB approach has outperformed the existing deep CNN-based methods to detect both tuberculosis and pneumonia diseases using chest X-ray images. The proposed MEGB approach decomposes the X-ray images at different scales and extracts discriminative features at each scale using eigendomain analysis. Hence, the LGBM has shown higher accuracy using these discriminative features of chest X-ray images. However, the CNN-based methods use convolution, pooling, and dense layers to classify X-ray images [13], [14]. The loss of relevant information may occur in the pooling layer stage. Due to this reason, the deep CNN model produces lower accuracy values than the MEGB approach. The proposed approach has not considered the segmentation of chest X-ray images for the automated detection of tuberculosis and pneumonia diseases. Unlike deep learning models, the proposed MEGB approach is interpretable, and the feature selection strategy has helped to improve the accuracy of the LGBM model. In this work, we have implemented the MEGB approach on the cloud-based framework for the automated classification of chest X-ray images. The suggested MEGB approach can be implemented on the edge device to design a smart healthcare system to detect tuberculosis and pneumonia diseases.

V. CONCLUSION

The MEGB-based detection of tuberculosis and pneumonia has been proposed in this letter using chest X-rays as input images. The subbands of chest X-ray images are obtained using 2-D-DWT. The eigendomain features have been determined from the subbands of chest X-ray images. The ANOVA-based feature selection technique has been employed to identify the relevant features of chest X-ray images. The LGBM classifier is used to detect pneumonia and tuberculosis diseases. The suggested MEGB approach with selected features has achieved higher accuracy than transfer learning and other existing methods.

REFERENCES

[1] M. H. Kashani, M. Madanipour, M. Nikravan, P. Asghari, and E. Mahdipour, "A systematic review of IoT in healthcare: Applications, techniques, and trends," *J. Netw. Comput. Appl.*, vol. 192, 2021, Art. no. 103164.

[2] World Health Org., WHO consolidated guidelines on tuberculosis. *Module 2: Screening-Systematic Screening for Tuberculosis Disease*. Geneva, Switzerland: World Health Org., 2021, pp. 5–15.

[3] A. Khan, M. U. Akram, and S. Nazir, "Automated grading of chest X-ray images for viral pneumonia with convolutional neural networks ensemble and region of interest localization," *PLoS One*, vol. 18, no. 1, 2023, Art. no. e0280352.

[4] J. Devasia, H. Goswami, S. Lakshminarayanan, M. Rajaram, and S. Adithan, "Deep learning classification of active tuberculosis lung zones wise manifestations using chest X-rays: A multi label approach," *Sci. Rep.*, vol. 13, no. 1, 2023,, Art. no. 887.

[5] A. Jacobi, M. Chung, A. Bernheim, and C. Eber, "Portable chest X-ray in coronavirus disease-19 (covid-19): A pictorial review," *Clin. Imag.*, vol. 64, pp. 35–42, 2020.

[6] W. Zhou et al., "Deep learning-based pulmonary tuberculosis automated detection on chest radiography: Large-scale independent testing," *Quantitative Imag. Med. Surg.*, vol. 12, no. 4, 2022, Art. no. 2344.

[7] K. Santosh, S. Allu, S. Rajaraman, and S. Antani, "Advances in deep learning for tuberculosis screening using chest X-rays: The last 5 years review," *J. Med. Syst.*, vol. 46, no. 11, 2022,, Art. no. 82.

[8] N. Singh and S. Hamde, "Tuberculosis detection using shape and texture features of chest X-rays," in *Proc. Innovations Electron. Commun. Eng.: Proc. 7th ICIECE*, 2019, pp. 43–50.

[9] T. Rahman et al., "Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization," *IEEE Access*, vol. 8, pp. 191586–191601, 2020.

[10] F. Pasa, V. Golkov, F. Pfeiffer, D. Cremers, and D. Pfeiffer, "Efficient deep network architectures for fast chest X-ray tuberculosis screening and visualization," *Sci. Rep.*, vol. 9, no. 1, 2019, Art. no. 6268.

[11] D. Zhang, F. Ren, Y. Li, L. Na, and Y. Ma, "Pneumonia detection from chest X-ray images based on convolutional neural network," *Electron.*, vol. 10, no. 13, 2021, Art. no. 1512.

[12] R. Kundu, R. Das, Z. W. Geem, G.-T. Han, and R. Sarkar, "Pneumonia detection in chest X-ray images using an ensemble of deep learning models," *PLoS One*, vol. 16, no. 9, 2021, Art. no. e0256630.

[13] R. Mogaveera, R. Maur, Z. Qureshi, and Y. Mane, "Multi-class chest X-ray classification of pneumonia, tuberculosis and normal X-ray images using convnets," in *Proc. ITM Web Conf.*, vol. 44, 2022, Art. no. 3007.

[14] M. Abubakar et al., "Classification of pneumonia and tuberculosis from chest X-rays," 2021, *arXiv:2103.14562*.

[15] R. Mostafiz et al., "Covid-19 detection in chest X-ray through random forest classifier using a hybridization of deep CNN and DWT optimized features," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3226–3235, 2022.

[16] S. Kabi, D. Patra, and G. Panda, "Detection of pneumonia from X-ray images using eigen decomposition and machine learning techniques," in *Proc. IEEE Int. Conf. Signal Inf. Process.*, 2022, pp. 1–5.

[17] R. K. Tripathy, S. Dash, A. Rath, G. Panda, and R. B. Pachori, "Automated detection of pulmonary diseases from lung sound signals using fixed-boundary-based empirical wavelet transform," *IEEE Sens. Lett.*, vol. 6, no. 5, pp. 1–4, May 2022.

[18] S. Dash, R. K. Tripathy, D. K. Dash, G. Panda, and R. B. Pachori, "Multiscale domain gradient boosting models for the automated recognition of imagined vowels using multichannel EEG signals," *IEEE Sens. Lett.*, vol. 6, no. 11, pp. 1–4, Nov. 2022.

[19] D. Kermany, K. Zhang, and M. Goldbaum, "Labeled optical coherence tomography (OCT) and chest X-ray images for classification (2018)," *Mendeley Data*, 2018. v2 10.17632/rscbjbr9sj. Accessed: Jan. 5, 2023. [Online]. Avaliable: https://nihcc.app.box.com/v/ChestXray-NIHCC

[20] D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[21] N. O. F. Elssied, O. Ibrahim, and A. H. Osman, "A novel feature selection based on one-way ANOVA F-test for e-mail spam classification," *Res. J. Appl. Sci., Eng. Technol.*, vol. 7, no. 3, pp. 625–638, 2014.

[22] G. Ke et al., "Light GBM: A highly efficient gradient boosting decision tree," *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 1–9, 2017.

[23] F. Pedregosa et al., "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 1–6, 2011.