https://doi.org/10.46335/IJIES.2023.8.5.11 e-ISSN: 2456-3463
Vol. 8, No. 5, 2023, PP. 48-52

International Journal of Innovations in Engineering and Science, www.ijies.net

Machine Learning for COVID-19 Image Analysis

Mahesh Shivling Sadavarte[1] , Dr. V.M.Deshmukh[2]

[1]*Assitant Professor in Electronics and Telecommunication*
Government College of Engineering,Jalgaon ,India,
[2]*Professor in Electronics and Telecommunication*
SSBT'S Colege of Engineering and Technology,Bambhori,India,

mahesh.sadavarte@gcoej.ac.in , vmdeshmukh947@gmail.com

***Received on**: 04 April, 2023* ***Revised on**: 08 May,2023* ***Published on**: 10 May,2023*

***Abstract—** COVID-19 is a worldwide epidemic, as announced by the World Health Organization (WHO) in March 2020. Image Processing methods can play vital roles in identifying COVID-19 patients by visually analyzing their chest x-ray images. As the cost and required time of conventional RT-PCR tests to detect COVID-19, researchers are trying to use medical images like X-Ray and Computed Tomography (CT) images to detect it with the help of Artificial Intelligence (AI) based systems. In this paper, we analyze different classification methodslike GLCM and Gabor features. Performance of SVM, KNN and Naïve Bayes emerging AI-based models that can detect COVID-19 from medical images using X-Ray or CT of lung images. We clant COVID images ssified by usng datasets, preprocessing techniques, segmentation, feature extraction, classification and experimental results which can be helpful for finding future research directions in the domain of automatic diagnosis of Covid-19 disease using Machine learning,*

Keywords: COVID-19, SVM, KNN, Machine Learning

I - INTRODUCTION

COVID 19 coronavirus named SRS CO V-2 which was first found China in December 2019. COVID 19 is a serious disease.[1] Ppeople suffer due to COVID. World Health Organization has declared COVID-19 pandemic by March 2020 [2]. Corona viruses widely spread in different countries.. Before SARS-CoV-2 two coronaviruses caused severe human respiratory disease namely SARS-CoV, which causes Severe Acute Respiratory Syndrome (SARS); and MERS-CoV, which causes Middle East Respiratory Syndrome (MERS) [4, 5]. But as compared with SARS and MERS, appearance of COVID-19 is in mild form but it allows infection spread from asymptomatic patients, which in turn has led to the current pandemic [6]. WHO has tackled the pandemic by emphasizing the need of massive testing and contact tracing. But all countries does not have the required laboratory infrastructure for addressing the task of massive testing and contact tracing. Also the test results of some of the laboratory test takes two or more days. This leads to spread of COVID-19 disease from patients who have mild or no symptoms. Computer assisted analysis of COVID-19 images has been of great use in case of COVID-19 image analysis using deep learning. Automatically analyzing Computed Tomography (CT) scans for diagnosis of the disease is important [7]. Fair results are obtained in detecting tuberculosis signs by radiographic analysis [8] and other multiple cardiothoracic abnormalities [9, 10]. RT-PCR testing requires lot of time. Hence from the beginning of COVID-19, and alternative approach of CT and X-ray imaging has been preferred by research community. During early days some pulmonary lesions are found in COVID-19 imagery of non-severe and even in recovered patients

Use of CT and X ray imaging has found useful in analysis of COVID-19. But when using it practically, it has some drawbacks like difficulty for use in massive COVID-19 screening. The cost of CT and X-ray equipment, availability, portability are the major concern. Analysis of

CT and X ray imaging is difficult for COVID-19 screening due to sterilization procedures of equipment.

I. COVID-19 Chest X-Ray Database

COVID 19 radiography database of Kaggle [11]is used during this work. The dataset consist of 21164 X ray images which are categorized into COVID-19, pneumonia, lung obesity and normal/healthy. These data included X-ray images with confirmed COVID-19, confirmed common pneumonia, lung obesity and normal\healthy individuals. This dataset is divided into 70% train dataset and 30% test dataset. Table 1 shows the final number of images distributed in the 70% train, 30% validation dataset used for the pre-trained model.

Table 1: Images for COVID-19 Dataset

Classification	Train	Validation	Total
COVID	2531	1084	3615
Normal	7135	3057	10192
Lung Obesity	4209	1803	6012
Viral Pneumonia	942	403	1345
Total	14817	6347	21164

Table 2 below shows some sample images from the dataset

Table 2: Images for COVID-19 Dataset

Classification			
COVID			
Normal			
Lung Obesity			
Viral Pneumonia			

II - Literature Review

Deep learning has shown a dramatic increase in the medical applications in general and specifically in medical image based diagnosis. Deep learning models performed prominently in computer vision problems related to medical image analysis. The ANNs outperformed other conventional models and methods of image analysis [7, 8]. Due to the very promising results provided by CNNs in medical image analysis and classification, they are considered as de facto standard in this domain [9, 10]. CNN has been used for a variety of classification tasks related to medical diagnosis such as lung disease [10], detection of malarial parasite in images of thin blood smear [11], breast cancer detection [12], wireless endoscopy images [13], interstitial lung disease [14], CAD-based diagnosis in chest radiography [15], diagnosis of skin cancer by classification [16], and automatic diagnosis of various chest diseases using chest X-ray image classification [17]. Since the emergence of COVID-19 in December 2019, numerous researchers are engaged with the experimentation and research activities related to diagnosis, treatment, and management of COVID-19. Researchers in [18] have reported the significance of the applicability of AI methods in image analysis for the detection and management of COVID-19 cases. COVID-19 detection can be done accurately using deep learning models' analysis of pulmonary CT [18]. Researchers in [19] have designed an open-source COVID-19 diagnosis system based on a deep CNN. In this study, tailored deep CNN design has been reported for the detection of COVID-19 patients using X-ray images. Another significant study has reported on the X-ray dataset comprising X-ray images belonging to common pneumonia patients, COVID-19 patients, and people with no disease [20]. The study uses the state-of-the-art CNN architectures for the automatic detection of patients with COVID-19. Transfer learning has achieved a promising accuracy of 97.82% in COVID-19 detection in this study. Another recent and relevant study has been conducted on validation and adaptability of Decompose-, Transfer-, and Compose-type deep CNN for COVID-19 detection using chest X-ray image classification [21]. The authors have reported the results of the study with an accuracy of 95.12%, sensitivity of 97.91%, and specificity of 91.87%. Having reviewed the relevant and recent research work on the design, development, and possible applicability of machine learning in COVID-19 detection using medical images, particularly X-ray images, due to the availability of a very less amount of X-ray images of COVID-19 patients and the poor quality of some images in the dataset, the accuracy of the models was affected.

III - Proposed Methodology

We worked on a machine learning approach for COVID 19 image classification. Different steps involved in the process are

1. Pre-processing: During this step noise free COVID 19 images prepared using median filtering. COVID 19 images are divided into three channels separately. Media filters are applied to each image

channel separately. Then the three image channels are combined again.

2. Features Extraction: During this stage, various color and texture based features are extracted from the noise free COVID 19 images. During features extraction we have used
    - Color Mean Pixel Value
    - Color moments
    - Edge Feature extraction using Pewit operator
    - Gabor Features Extraction
    - Histogram Features Extraction
    - GLCM Features
    - HAAR features
    - HOG Features Extraction
    - LBP feature extraction
3. Training: Various machine learning algorithms including LDA, KNN, Decision trees, Naïve Bayes, SVM are trained using 70% of the data.
4. Classification: COVID 19 images are classified into various types. The performance of the proposed architecture is compared with LDA, KNN, Decision trees, Naïve Bayes, SVM methods with 30% evaluation data.

IV-FEATURES EXTRACTION AND CLASSIFICATION METHODOLOGIES

a. Gabor Features Extraction

This filter here we used texture analysis, which means that it basically analyzes whether there are any specific frequency content in the image in specific directions in a localized region around the point or region of analysis. Frequency and orientation representations of Gabor filters done on the human visual system. Forr texture representation and discrimination. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave.

b. Statistical Features

stats = graycoprops(glcm, properties) determin the statistics specified in properties from the different gray-level co-occurence matrix glcm. glcm is an m-by-n-by-p array of valid gray-level co-occurrence matrices. If glcm is an array of GLCMs, stats is an array of statistics for each glcm. graycoprops normalizes the gray-level co-occurrence matrix (GLCM) so that the sum of its elements is equal to 1. Each element (r,c) in the normalized GLCM is the joint probability occurrence of pixel pairs with a defined spatial relationship having gray level values r and c in the image. graycoprops uses the normalized GLCM to calculate properties

Statistical features considered here are contrast, correlation, Energy and Homogeneity

c. Support Vector Machines

Spport-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms thatworks on analyze data used for classification and regression analysis. A set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is gives the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on the side of the gap on which they fall.

SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. When data are unlabelled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups.

d. KNN Classification

Here pattern recognitionit done withe k-nearest neighbors algorithm (k-NN) is a non-parametric method for classification and regression. The input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression.

. An object is classified by a plurality vote of its neighbors, with the object being give to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of it's single nearest neighbor.

The output is the property value for the object. This value is the average of the values of k nearest neighbors. k-NN is a type of instance-based learning, or slow learning, where the function is only approximated locally and all computation is deferred until classification. The k-NN algorithm is among the easiest algorithms- classification and regression, a useful technique can be used to assign weight to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones.

Naïve Bayes Classification

This classification isrithms based on **Bayes' Theorem**. It is t a family of algorithms where all uses common principle, that is. every pair of features being classified is independent of each other. Naïve Bayes is a algorithm that relies on strong assumptions of the independence of covariates in applying Bayes Theorem. The Naïve Bayes classifier assumes independence between predictor variables conditional on the response, and a Gaussian distribution of numeric predictors with mean and standard deviation computed from the training dataset.

Naïve Bayes models are commonly used as an alternative to decision trees for classification problems. When building a Naïve Bayes classifier, every row in the training dataset that contains at least one NA will be skipped completely. If the test dataset has missing values, then those predictors are omitted in the probability calculation during prediction.

V- RESULTS AND DISCUSSIONS

This classification of automatic COVID 19 image casing PYTHON. During preprocessing COVID 19 images are separated into three separate channels. Medial filter is applied to each image channel separately. And then these three separate noise free channel images are combined to form a noise free COVID 19 image. From COVID 19 images various color and texture features are extracted. The data is split into 70% data for training and 30% data for evaluation. Various classifiers are trained using 70%training data. Performance of the trained system is evaluated using 30% test data in terms of accuracy, precision, recall and f measure. SVM Classification using color mean pixel value features gives highest accuracy of 81.76% for classification of COVID 19 images.
Table 3 below depicts the accuracy of the proposed system for different classifiers under consideration.

Table 3: Accuracy for COVID 19 image Classification

	KNN	CART	NB	SVM	Average
Color Mean Pixel Value	0.77	0.693	0.571	**0.817**	0.713
Color Moments	0.51	0.591	0.444	0.485	0.508
GLCM	0.51	0.54	0.55	0.54	0.53
HOG	0.74	0.6	0.62	0.79	0.70
Gabor	0.42	0.65	0.43	0.66	0.54
Average	0.59	0.62	0.52	0.66	

As depicted in table 3, Color Mean Pixel Value features gives better accuracy as compared with other color and texture features under consideration. SVM classification using Color Mean Pixel Value features gives accuracy of 81.76% for 30% unseen test data of COVID 19 images.

Performance is also evaluated using precision, recall and F measure.

VI -CONCLUSION

The task of this work is to find a machine learning model that identifies whether a patient is suffering from COVID-19. To develop such a model, a literature study alongside an experiment will be set to identify a suitable algorithm. To assess the features that impact the prediction model.
In this have identified and localized COVID-19 abnormalities on chest radiographs. In particular, we have categorized the radiographs as negative for pneumonia or typical, indeterminate, or atypical for COVID-19. The model will work with imaging data and annotations from a group of radiologists.
This work will help radiologists diagnose the millions of COVID-19 patients more confidently and quickly. This will also enable doctors to see the extent of the disease and help them make decisions regarding treatment. Depending upon severity, affected patients may need hospitalization, admission into an intensive care unit, or supportive therapies like mechanical ventilation. As a result of better diagnosis, more patients will quickly receive the best care for their condition, which could mitigate the most severe effects of the virus.

REFERENCES

[1] *Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Kashem, S.B.A., Islam, M.T., Maadeed, S.A., Zughaier, S.M., Khan, M.S. and Chowdhury, M.E., 2020. Exploring the Effect of Image Enhancement Techniques on COVID-19 Detection using Chest X-ray Images. arXiv preprint arXiv:2012.02238.*

[2] *Q. Zhao, H. Wang and G. Wang, "LCOV-NET: A Lightweight Neural Network For COVID-19 Pneumonia Lesion Segmentation From 3D CT Images," 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France, 2021, pp. 42-45.*

[3] *J. Hua, M. Lin Huang, C. Zhao, S. Hua and C. Shih, "An Initial Visual Analysis of the Relationship between COVID-19 and Local Community Features," 2020 24th International Conference Information Visualisation (IV), Melbourne, Australia, 2020, pp. 718-722.*

[4] *D. D. Putra, M. Febriyanto, M. M. Nadra, W. Shalannanda, E. R. Firzal and A. Munir, "Design of Smart-Gate Based on Artificial Intelligence Possibly for COVID-19 Early Prevention at Public Area," 2020 14th International Conference on Telecommunication Systems, Services, and Applications (TSSA, Bandung, Indonesia, 2020, pp. 1-4.*

[5] *J. Luo, "Bitcoin price prediction in the time of COVID-19," 2020 Management Science Informatization and Economic Innovation Development Conference (MSIEID), Guangzhou, China, 2020, pp. 243-247.*

[6] J. Cheng et al., "Automated Diagnosis of COVID-19 using Deep Supervised Autoencoder with Multi-view Features from CT Images," in IEEE/ACM Transactions on Computational Biology and Bioinformatics.

[7] G. Tradigo, P. H. Guzzi, T. Kahveci and P. Veltri, "A method to assess COVID-19 infected numbers in Italy during peak pandemic period," 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Korea (South), 2020, pp. 3017-3020.

[8] S. Jadhav, G. Deng, M. Zawin and A. E. Kaufman, "COVID-view: Diagnosis of COVID-19 using Chest CT," in IEEE Transactions on Visualization and Computer Graphics.

[9] R. Ramachandran, S. Velan S and D. Imtiyaz Wadekar, "Statistical Comparison of COVID-19 Infections Based Upon the Food Habits/Diets in Countries Using RStudio," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2021, pp. 1095-1101.

[10] R. Chauhan, P. Goel, V. Kumar, N. Soni and N. singh, "Understanding Covid-19 using data visualization," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2021, pp. 555-559.

[11] M.E.H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M.A. Kadir, Z.B. Mahbub, K.R. Islam, M.S. Khan, A. Iqbal, N. Al-Emadi, M.B.I. Reaz, M. T. Islam, "Can AI help in screening Viral and COVID-19 pneumonia?" IEEE Access, Vol. 8, 2020, pp. 132665 - 132676.
https://www.kaggle.com/tawsifurrahman/covid19-radiography-database

[12] S. Tabik et al., "COVIDGR Dataset and COVID-SDNet Methodology for Predicting COVID-19 Based on Chest X-Ray Images," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 12, pp. 3595-3605, Dec. 2020.
doi: 10.1109/JBHI.2020.3037127

[13] H. Tahir, A. Iftikhar and M. Mumraiz, "Forecasting COVID-19 via Registration Slips of Patients using ResNet-101 and Performance Analysis and Comparison of Prediction for COVID-19 using Faster R-CNN, Mask R-CNN, and ResNet-50," 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 2021, pp. 1-6. doi: 10.1109/ICAECT49130.2021.939248

[14] J. Li, D. Zhang, Q. Liu, R. Bu and Q. Wei, "COVID-GATNet: A Deep Learning Framework for Screening of COVID-19 from Chest X-Ray Images," 2020 IEEE 6th International Conference on Computer and Communications (ICCC), Chengdu, China, 2020, pp. 1897-1902

[15] Y. -H. Wu et al., "JCS: An Explainable COVID-19 Diagnosis System by Joint Classification and Segmentation," in IEEE Transactions on Image Processing, vol. 30, pp. 3113-3126, 2021.

[16] A. Kundu, C. Mishra and S. Bilgaiyan, "COVID-SEGNET: Diagnosis of Covid-19 Cases on Radiological Images using Mask R-CNN," 2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII), Chennai, India, 2021, pp. 1-5.

[17] G. Guo, Z. Liu, S. Zhao, L. Guo and T. Liu, "Eliminating Indefiniteness of Clinical Spectrum for Better Screening COVID-19," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 5, pp. 1347-1357, May 2021.
doi: 10.1109/JBHI.2021.3060035

[18] R. Y. Wang, T. Q. Guo, L. G. Li, J. Y. Jiao and L. Y. Wang, "Predictions of COVID-19 Infection Severity Based on Co-associations between the SNPs of Co-morbid Diseases and COVID-19 through Machine Learning of Genetic Data," 2020 IEEE 8th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, 2020, pp. 92-96.
doi: 10.1109/ICCSNT50940.2020.9304990

[19] Arpit Jain; Abhinav Sharma; Jianwu Wang; Mangey Ram, "Use of AI, Robotics, and Modern Tools to Fight Covid-19," in Use of AI, Robotics, and Modern Tools to Fight Covid-19 , River Publishers, 2021, pp.ii-xxx.

[20] X. Li, C. Li and D. Zhu, "COVID-MobileXpert: On-Device COVID-19 Patient Triage and Follow-up using Chest X-rays," 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Korea (South), 2020,pp.1063-1067.doi:10.1109/BIBM49941.2020.9313217

[21] Kulkarni Chhaya; Dey Sandipan; Janeja Vandana, "7 Discovery of Robust Distributions of COVID-19 Spread," in Use of AI, Robotics, and Modern Tools to Fight Covid-19 , River Publishers, 2021, pp.89-110.