

AI based NLP Assistant for Automatic Detection of Coronavirus Disease (COVID-19) Using chest CT and X-ray Radiology Reports :A Technical Survey

Vaishali M. Kumbhakarna^{*}, Sonali B. Kulkarni^{**}, Apurva D.Dhawale^{***}

^{*} (Ph.D. student, Department of Computer Science, Dr.Babasaheb Ambedkar Marathwada University, Aurangabad 431004, India .)

^{**} (Assistant Professor, Department of Computer Science, Dr.Babasaheb Ambedkar Marathwada University, Aurangabad 431004, India .)

^{***} (Ph.D. student, Department of Computer Science, Dr.Babasaheb Ambedkar Marathwada University Aurangabad 431004, India.

Abstract: The rapid spread of novel COVID-19 disease is imminent. According to the World Health Organization (WHO), the coronavirus pandemic is putting even the best healthcare systems across the world under tremendous pressure. It is essential to identify infected individuals as early as possible for quarantine and treatment procedures. Artificial Intelligence technique has potential to control community spread of COVID-19 by providing a clinical diagnosis ahead of the pathogenic test. Through this study ,we can conclude that NLP technique provides machine learning algorithms for automatic identification of the imaging patterns of chest CT and X-ray .Thus CAD system provides the hallmarks of COVID-19 with low cost, less time ,less errors and minimum human interaction.

Keywords: CADe , CADx ,NLP, ML, AI, Clinical note,Radiology ,COVID-19.

Date of Submission: 15-01-2021

Date of acceptance: 31-01-2021

I. INTRODUCTION

The novel coronavirus 2019 (COVID-2019), which first appeared in Wuhan city of China in December 2019, caused by the SARS-CoV-2 virus .The outbreak of a person-to-person transmissible pneumonia caused by the Severe Acute Respiratory Syndrome coronavirus , also known as COVID-19 has caused a pandemic all over the world. There have been millions of confirmed cases of the COVID-19 throughout the world.Artificial Intelligence (AI) is a potential and powerful tool of computer science. It is the scramble for scientists and researchers to use AI as a weapon in the fight against corona.

There are Six areas where AI can contribute to the fight against COVID-19

Table 1.1 AI Based Application and Tools for COVID-19

Sr.No.	Application area	AI Tool
1)	To give early warnings and alerts about positive cases	BlueDot tool
2)	To tracking and prediction of positive cases	GLEAMviz model
3)	To provide data dashboards	Microsoft Bing's COVID-19 Tracker
4)	To diagnosis and prognosis the patient	CAD4COVID system
5)	To treat and cure of patient	DeepMind
6)	To control community spread	Aarogya Setu App

NLP is a subfield of artificial intelligence that enables computer programs to process and analyze unstructured data, such as free-text physician/clinical notes ,radiology reports, pathology reports etc.Clinical note and part of speech can be processed remotely,without coming in contact with suspected patient ,by using NLP techniques and thus social distancing can be maintained .As the doctors nurses ,hospital staff all are corona worriers and are working at high risk of getting infected while investigating the patient . So the early detection of this community spread type coronavirus will help in relieving the pressure of the healthcare systems.

Chest X-rays plays a crucial role in the accurate and automatic diagnosis of diseases . As COVID-19 is a type of influenza, it is possible to diagnose using this imaging technique. Machine Learning (ML) and Deep

learning, can develop intelligent systems to classify between COVID-19 positive or negative patients. Various studies are published on it, some are given below

II. LITERATURE REVIEW

In this paper, A Computer-Aided Detection System for Digital Chest Radiographs, author used 48 chest radiography images, those are high resolution DICOM X-ray images. Local binary patterns (LBP) histogram [1] was used to decide normality/pathology x-ray. For image processing, in the preprocessing stage, images were first resized by using supersampling interpolation. Then segmentation was done based on the template matching algorithm.

In the paper "A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)", Author had collected 1,065 CT images [2] of pathogen-confirmed COVID-19 cases. Author had modified the Inception transfer-learning model to establish the algorithm. The hallmarks of COVID-19 are bilateral distribution of patchy shadows and ground glass opacity in early stages. As the disease progresses, multiple ground glass and infiltrates in both lungs will appear. These features are quite similar to typical viral pneumonia with only slight differences, which are difficult to be distinguished by radiologists. Based on this, author believed that CNN might help to identify unique features that might be difficult for visual recognition. Hence, the purpose of this study was to evaluate the diagnostic performance of a deep learning algorithm [1] using CT images to screen for COVID-19.

In the paper "Accurate Prediction of COVID-19 using Chest X-Ray Images through Deep Feature Learning model with SMOTE and Machine Learning Classifiers", author used two publicly available datasets: Chest X-Ray Images (Pneumonia) [1] and COVID-19 public dataset from Italy [2]. For the training of the ResNet152 model architecture [3], a total of 5840 images are used. For the final classification of COVID-19 patients by the machine learning classifiers, we used the 2748 images

In order to effectively perform classification tasks, image preprocessing is performed. All input images are first converted to a standard size of 224x224 for a similar course of action in both the developed model

In the paper "Artificial Intelligence against COVID-19: An Early Review", author has specified six different application areas where artificial intelligence [4] can contribute to fight against pandemic. Author had concluded that AI plays a very important role for controlling the community spread of COVID-19 and helps in management of healthcare system.

In the paper "Automated detection of COVID-19 cases using deep neural networks with X-ray images", author had proposed CNN based model including 17 layers of convolutional neural network [5] to classify COVID-19 cases using x-ray images. Author also concluded that such type of AI based models will definitely fulfill the shortage of radiologists during the pandemic period as a burden is increased on healthcare profession

In the paper "Automatic Detection of Acute Bacterial Pneumonia from Chest X-ray Reports", author studied 292 chest x-ray reports. He had developed NLP system named as SymText [6] based on Bayesian network. From this study, author also concluded that NLP system gives better performance for extraction of pneumonia-related concepts from chest x-ray reports.

In this paper, author has used chest X-ray images from GitHub dataset to find COVID-19 positive patients automatically using image processing technique. In this study, author has developed CNN based ResNet50 [7] deep learning model. The model was developed using Python. Author concluded that the proposed model will help to detect COVID-19 patients at an earlier stage and thus prevent community spread.

In this study, author has taken survey on various open source datasets [8] available for Chest X-ray radiology images. In this survey, he came across various AI based techniques used in computer-aided detection and diagnosis system. Author has discussed many steps in image processing like preprocessing, ROI identification, feature extraction, classification, SVM, CNN etc. He also concluded that deep learning plays an important role in automatic diagnosis of disease from chest radiography.

In this paper, author has studied CT scan chest images. The dataset of 150 CT chest images [9] were grouped into 4 categories. For feature extraction, DWT i.e. discrete wavelet transform algorithm was used where as for SVM was used as classification technique. Author concluded that the proposed system may perhaps be used to diagnose the COVID-19 disease as an assistant system. He also concluded that to diagnose COVID-19 manual system takes 1 to 5 hr. whereas proposed system reduced the delineation time up to only 4 min. and thus prevent community spread.

In this paper, author taken literature review on the features found in CT images and information available related to novel corona virus. Author studied the various articles from PubMed, Elsevier, Google scholar and WHO. Author concluded various features of COVID-19 like ground-glass opacification (GGO) [10] appearance in lungs CT images, posterior distribution, presentation of consolidative opacities, septal thickening, bronchiectasis, pleural thickening, pleural effusion, pericardial effusion, lymphadenopathy, etc

In this paper author studied chest x-ray images[11] to automatic diagnose corona infected cases.author has proposed Convolutional Neural Network architectures along with transfer learning procedure.author concluded that by implementing this system, spreading period of the pandemic will be controlled and diagnosis of the covid -19 will be possible at a low-cost, rapid, and automatic with the proposed CNN architecture.

In this paper author used chest X ray images for COVID-19 screening.The proposed deep-learning model has three components named as backbone network,[12] classification head, and anomaly detection head.The system used 18-layer residual convolutional neural network .Stochastic gradient descent (SGD) algorithm was used for training dataset.author also concluded that as compare to CT-based screening , the proposed X-ray-based screening model gives better performance. Therefore, the proposed model was the effective computer-aided diagnosis (CAD) tool which could be widely accepted due to its faster imaging time and low-cost .

In this paper author proposed COVID-Net, a deep convolutional neural network model.It was a customized system for the detection of COVID-19 cases from chest X-ray images[13] . It was an open source model available to the general public. The architecture was based on projection-expansion-projection design pattern It was estimated using the Keras deep learning library with a TensorFlow backend. The proposed COVID-Net architecture was built using generative synthesis mechanism.

In this paper author proposed CVOIDX-Net framework which would assist radiologists to automatically detect covid -19 cases from 2D conventional CXR images.The model has 3 main components as, preprocessing using scaling ,trained using Stochastic Gradient Descent (SGD)[14][12] algorithm ,validation by applying evaluation metrics and deep learning classifiers implemented by using Python and the Keras package with TensorFlow2[13].the performance analysis of model was done based on confusion matrix.author also recommended that within all 7 deep learning classifiers; named as VGG19, DenseNet201, ResNetV2, InceptionV3, InceptionResNetV2, Xception, and MobileNetV2 ,the the VGG19 and DenseNet201 classifier if applied in CAD system ,it will give high performance score .

In this article, author had reviewed various applications of machine learning in the field of radiology.

recently CNN based CAD system can be found as common,widespread machine learning application[15] in radiology practice. Author also concluded that the AI techniques could definitely speed up the productivity of radiologists ,reduce the burden and hence enhance patient care and satisfaction.

In this paper, transverse-section lunges CT images were used for automatic detection of corona virus.in the proposed model ,images were first segmented by using a 3-D deep learning model ,then in the second stage, location-attention classification model was applied ,after that Noisy-or Bayesian function was used to calculate overall infection probability .The classical ResNet [16]was used for feature extraction . The author concluded that, appearances of COVID-19 CT imaging and other pneumonias'were unbiased .So one couldn't distinguish COVID-19 from other pneumonia just with human eyes. Therefore deep learning system-based screening models would be more reliable ,fast and accurate. It would definitely assist physicians to make a clinical decision automatically and quickly, which would benefit to manage the suspected corona case earlier and thus helps to control pandemic.

In this paper, author proposed DeCoVNet architecture which was based on PyTorch

Framework[17].It used lung CT images.It was a 3D deep convolutional neural Network developed to Detect COVID-19 cases automatically . The combination of epidemiologic features for example ,travel or contact history, clinical signs and symptoms,chest CT, laboratory findings and real-time RT-PCR nucleic acid testing was used for the final identification of COVID-19.If the CT reports found COVID-positive,then author mentioned label flag as one and zero for negative .author concluded that the deep learning algorithm would provide a fast and accurate method to identify COVID-19 patients, which would be beneficial to control the spread of corona virus.

In this paper author used X-ray images to detect corona infected patients.in the proposed ResNet50 model support vector machine classifier [18]was used.in this study author used the database of CXR[28] that is chest X-ray from GitHub, Kaggle and Open-I repository.after comparing total of 11 CNN model author concluded that the proposed model gives highest accuracy for detection of COVID -19 from chest X-ray images.

In this paper author used operation notes dataset for Extracting information .The proposed model used both rule-based and supervised machine learning NLP method[19]. rule-based method gives good performance on medical language processing .Conditional Random Fields (CRFs) also widely used to solve sequence tagging problem for medical natural language processing . author also concluded that, annotation from unstructured clinical notes would be a valuable information source for many clinical applications .

In this paper, author used computerised admission notes and discharge summaries for identification of assumptions.in the proposed system the method used for feature extraction from Chinese clinical text[21] was bag of characters, bag of words, character embedding, and word embedding. Conditional Random Fields was

used for extracting information from clinical notes. Author concluded that identification of speculation from Chinese clinical text was a spirited NLP task

In this paper author had developed a tool competent to the annotated corpus in Chinese clinical text. In this work sequence-labeling method was used to train statistical models and word segmentation, POS tagging, shallow parsing, and named entity recognition. Conditional random fields algorithm, was used to extract clinical information. The proposed model built for syntactic and semantic annotated corpus of Chinese clinical texts, [20][22] was trained on the bases of SVM algorithm. Author concluded that, it was the first widespread annotated corpus of Chinese clinical text, setting a solid establishment for future research in clinical text domain.

In this paper author had proposed the Medical information system Medis.NET [23] which was a data summarizing tool. It was developed using SQL. It used visit records, document, or prescription associated with a chronic diagnosis. The author designed extended data structure used as data summarization tool. The tool was working in both automatic and on-demand mode. The author concluded that the tool would perform better if the patient-diagnosis relation is not one-to-one but many-to-many. The tool was tested on a single type of the medical information extraction but could be customized.

In this paper, author had studied chest x-rays (CXR) [27] radiography for automated computer-aided interpretation of Tuberculosis (TB). [24] The developed system CAD4TB was based on machine learning method and k-NN was used to compute a cumulative abnormality score. Author concluded that the proposed system has potential to be used as a computerized tool for screening of symptomatic individuals of disease TB.

In this paper, author had proposed an automated surveillance system using NLP [26] for screening pneumonia in newborn babies [25]. In the proposed system had three phases such as Knowledge acquisition, Knowledge representation, Knowledge optimization and testing. The dataset used for this study was from two different neonatal intensive care units (NICU) radiology reports and monitoring system had components like MedLEE NLP system and rules based NLP.

In this paper author had proposed multiple CNN models [29] to classify Covid-19 positive patients based on chest X-ray scans. Kaggle dataset chest x-ray images were used. Author concluded that, chest x-ray has a significant lower cost process as compare to chest CT. Deep learning is the most successful technique of machine learning, for diagnosis of covid -19. Author also analysed the performance of model by comparing Inception V3, Xception, and ResNeXt models and examined the accuracy of system.

In this paper, author had taken a review on how NLP benefits to radiology domain [26]. For this review author studied relevant publications of NLP applications in radiology domain. During this survey author concluded that the Quality Assessment of Radiologic Practice [30] was the category which covers applications that fulfilling legal requirements. Automatic content analysis of radiology report for large databases can give awareness in the day-to-day routine and inner workings of the radiology department. NLP have been used to generate evocative statistics for large dataset of radiology reports like current pandemic situation.

In this paper author had discussed AI tool for analyzing chest x rays and MIMIC-CXR, [31] an online repository to study COVID-19 affected chest x-ray images. Author concluded that AI algorithms could give doctors an edge by enabling them to better distinguish COVID-19 from other diseases.

In this paper, the author had discussed the RADLogics, the AI-Powered solution to support chest CT imaging for COVID-19 patients [32]. It provides machine learning image analysis solutions to improve radiologists' productivity. Author also concluded that, RADLogics is one of the innovators in AI & machine learning image analysis and radical big data analytics to search and analyze CTs, MRIs, PET scans, and X-rays images which helped to ease diagnostics turnaround time into minutes by automating detection and report generation tasks. It had patented AI medical image analysis platform enabling fast development of AI algorithms, and provides unified assimilation into current radiology workflow.

In this paper author had developed U-Net, deep learning algorithm based on chest CT images for automated segmentation of multiple COVID-19 infection regions. It was a type of artificial neural network (ANN) [33] containing a set of convolutional layers and deconvolutional layers to perform the task of biomedical image segmentation.

In this paper, author had carried out Large scale observer tests to examine how radiologists can benefit from CAD systems. Author had studied chest x-ray systems for five different area of findings and evaluated the performance of radiologist work flow with and without using CAD system through ROC curves [34]. Author studied the chest x ray for following five different kind of findings

Table 2.1 NLP Method Based on X-ray Findings for COVID-19

Sr.No.	X-ray Findings	NLP Method
1.	Detection of pulmonary nodules	Rule-based analysis
2.	Temporal subtraction	Iterative image-warping technique
3.	Detection of interstitial lung disease	Fourier transformation
4.	Differential diagnosis of interstitial lung disease	Artificial Neural Network

5.	Distinction between benign and malignant pulmonary nodules	Linear discriminant analysis
----	--	------------------------------

author concluded that , CAD has the potential to improve the performance of radiologists in decision-making process for interpreting chest radiographs screening.

III. CONCLUSION

The corona virus shows partially similar behaviours with other viral pneumonia. Therefore, identification of the novel coronavirus has ambiguity ,which makes the situation difficult to be under control. Some clinical finding in lunges screening such as the bronchiectasis, lesion swelling symptoms, and different shadowiness in CT images and x-ray provide to diagnose COVID-19, easily with its own characteristics .Therefore, the clinical experts need lung CT images to diagnose the COVID-19 in early phase. In this study we concluded that machine learning and artificial intelligence are not expected to replace the radiologists but these techniques can potentially facilitate radiology workflow, increase radiologist productivity, improve detection and interpretation of findings, reduce the chance of error, and enhance patient care and help to control the pandemic situation .In this paper we had reviewed some NLP algorithms developed for automatic identification of covid -19 positive or negative cases based on chest CT or X-ray images. AI tools such as COVID-Net, CAD4COVID, “AI-assisted” radiology technologies , etc. definitely helps to Fast and accurate diagnosis of coronavirus .

REFERENCES

- [1]. Juan Manuel Carrillo-de-Gea,” A Computer-Aided Detection System for Digital Chest Radiographs”, *Journal of Healthcare Engineering*,2016.
- [2]. Shuai Wang,”A deep learning algorithm using CT images to screen for Corona Virus Disease(COVID-19)” ,*medRxiv April , 2020*.
- [3]. Rahul Kumar, Ridhi Arora,” Accurate Prediction of COVID-19 using Chest X-Ray Images through Deep Feature Learning model with SMOTE and Machine Learning Classifiers” *medRxiv, April 2020*.
- [4]. Wim Naude,” Artificial Intelligence against COVID-19:An Early Review”, *IZA – Institute of Labor Economics*,2020.
- [5]. Tulin Ozturk,” Automated detection of COVID-19 cases using deep neural networks with X-ray images”, *Computers in Biology and Medicine,Elsevier*,2020.
- [6]. MARCELO FISZMAN,” Automatic Detection of Acute Bacterial Pneumonia from Chest X-ray Reports” . *Journal of the American Medical Informatics Association*,2000.
- [7]. Ali Narin,” Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks” *arXiv*,2020.
- [8]. Chunli Qin ,”Computer- aided detection in chest radiography based on artificial intelligence :a survey”, *BioMed Eng OnLine* ,2018.
- [9]. Mucahid Barstugan ,”Coronavirus (COVID-19) Classification using CT Images by Machine Learning Methods”, *arXiv*,2020
- [10]. Sana Salehi ,”Coronavirus Disease 2019 (COVID-19): A Systematic Review of Imaging Findings in 919 Patients” , *AJR* 2020.
- [11]. Ioannis D. Apostolopoulos,” Covid-19: Automatic detection from X-Ray images utilizing Transfer Learning with Convolutional Neural Networks” , *Physical and Engineering Sciences in Medicine,Springer*, 2020.
- [12]. Jianpeng Zhang,” COVID-19 Screening on Chest X-ray Images Using Deep Learning based Anomaly Detection”, *ResearchGate* ,2020.
- [13]. Linda Wang,” COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images”, *ResearchGate* ,2020.
- [14]. Ezz El-Din Hemdan,” COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images”, *arXiv*,2020.
- [15]. Garry Choy,”Current Applications and Future Impact of Machine Learning in Radiology ”, *radiology.rsna.org* , 2018.
- [16]. Xiaowei Xu,” Deep Learning System to Screen Coronavirus Disease 2019 Pneumonia”,*Elsevier*,2020.
- [17]. Chuansheng Zheng, “Deep Learning-based Detection for COVID-19 from Chest CT using Weak Label ”*medRxiv*,2020.
- [18]. Prabira Kumar Sethy,” Detection of coronavirus Disease (COVID-19) based on Deep Features”,*ResearchGate.org*,2020.
- [19]. Hui Wang, “Extracting important information from Chinese Operation Notes with natural language processing methods”, *Journal of Biomedical Informatics,Elsevier*,2014.
- [20]. Vaishali M. Kumbhakarna, Sonali B. Kulkarni, Apurva D.Dhawale ,”Clinical text engineering using natural language processing tools in healthcare domain:A systematic review “,SSRN : ICICC 2020,3 rd International conference on innovative computing and communication *InfoSciRN:Health Informatics,Elsevier*,28 march 2020.
- [21]. Shaodian Zhang ,”Speculation detection for Chinese clinical notes: Impacts of word segmentation and embedding models”, *Journal of Biomedical Informatics,Elsevier*,2016.
- [22]. Bin He,” Building a comprehensive syntactic and semantic corpus of Chinese clinical texts” ,*Journal of Biomedical Informatics, Elsevier* ,2017.
- [23]. Dejan Aleksic,”Data summarization method for chronic disease tracking”, *Journal of Biomedical Informatics, Elsevier* ,2017
- [24]. Syed Mohammad Asad Zaidi,”Evaluation of the diagnostic accuracy of Computer-Aided Detection of tuberculosis on Chest radiography among private sector patients in Pakistan” *Scientific REPORTS* ,2018.
- [25]. Eneida A. Mendonc,” Extracting information on pneumonia in infants using natural language processing of radiology reports “,*Journal of Biomedical Informatics , Elsevier* ,2005.
- [26]. Vaishali M. Kumbhakarna, Sonali B. Kulkarni, Apurva D.Dhawale,” NLP Algorithms Endowed For Automatic Extraction Of Information From Unstructured Free-Text Reports Of Radiology Monarchy”, *IJITEE,ISSN: 2278-3075, Elsevier* , October 2020
- [27]. Ho Yuen Frank Wong,” Frequency and Distribution of Chest Radiographic Findings in COVID-19 Positive Patients” *radiology.rsna.org* , 2020.
- [28]. Ming-Yen,”Imaging Profile of the COVID-19 Infection: Radiologic Findings and Literature Review”, *radiology.rsna.org* , 2020.

- [29]. Rachna Jain," Deep learning based detection and analysis of COVID-19 on chest X-ray images", *Applied Intelligence, Springer,2020*.
- [30]. Ewoud Pons ,"Natural Language Processing in Radiology: A Systematic Review ", *radiology.rsna.org, 2016*
- [31]. Katherine Wright," An AI Assist for Spotting COVID-19 in X Rays", *Physics.asp.org,2020*.
- [32]. Travis Small," RADLogics Expands Deployment of its AI-Powered Solution to Support Chest Imaging for COVID-19 Patients", *PRWeb ebooks 2020*.
- [33]. Xiaocong Chen," Residual Attention U-Net for Automated Multi-Class Segmentation of COVID-19 Chest CT Images", *JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2015, arXiv,2020*.
- [34]. Hiroyuki Abe," Computer-aided Diagnosis in Chest Radiography: Results of Large- Scale Observer Tests at the 1996–2001 RSNA Scientific Assemblies", *IMAGING & THERAPEUTIC TECHNOLOGY,RSNA, 2003*.

Vaishali M. Kumbhakarna, et. al. "AI based NLP Assistant for Automatic Detection of Coronavirus Disease (COVID-19) Using chest CT and X-ray Radiology Reports :A Technical Survey." *International Refereed Journal of Engineering and Science (IRJES)*, vol. 10, no. 01, 2021, pp 11-16.