AI-enabled radiologist in the loop: novel AI-based framework to augment radiologist performance for COVID-19 chest CT medical image annotation and classification from pneumonia

Hemant Ghayvat, Muhammad Awais, A. K. Bashir, Sharnil Pandya, Mohd Zuhair, Mamoon Rashid, and Jamel Nebhen

Introduction

The novel coronavirus, which first was found in Wuhan, China, has exponentially spread to the entire world since December 2019 [1]. The World Health Organization (WHO) named the transmittable infection triggered by this virus as the coronavirus infection of 2019 (COVID-19) on Feb 11, 2020 [2]. As of May 31, 2020, approximately 6,126,802 cases of COVID-19 and 371,220 deaths related to COVID-19 have been confirmed worldwide [3, 4]. Thus, COVID-19 is the worst pandemic triggered by the coronavirus family burdening the complete health care system. In addition, the lockdown measures adopted by most countries to contain the spread are pushing economies to their limits.

The clinical characteristics of COVID-19 are nonspecific and may include fever, cough, and dyspnea in most cases. In complicated cases, respiratory failure with acute respiratory distress syndrome (ARDS) has been observed [5, 6]. Radiological imaging that includes chest radiographs and CT could be a possible alternative diagnostic instrument for COVID-19 as patients commonly exhibit two-sided patchy shadows or ground-glass opacity (GGO) within the lung [7], typically with a peripheral and lower-zone predominance. Based on these, recent studies have reported that chest imaging could be a feasible option to diagnose patients manifesting symptoms and diagnose asymptomatic subjects [8]. The bigger challenge is in delineating COVID-19 from pneumonia and other pneumonia type etiologies. These abnormalities' classification is instinctively extra composite, accounting for the nonspecific clinical ciphers of Covid-19 [9]. However, with an appropriate automated solution using AI, robust enough to be adapted to various parts of the world, it will drastically reduce the radiology burden and significantly support the diagnosis rapidly, consequently enabling instant isolation and curbing the disease spread. Moreover, the quick diagnosis will also aid in early treatment leading to better patient outcomes.

Our work proposes a RILML framework for automated segmentation where a radiologist is partially employed to annotate data for training sets, classification based on textural changes in CT, with clinical expandability via heat maps. Testing is performed on external test datasets to validate the robustness and generalizability of the system. We demonstrate that our technique is more robust to unseen datasets. Therefore, the proposed framework can be employed directly in clinical setups to diagnose Covid-19 and gain deeper clinical insights into the disease.

This investigation was directed in two phases.

- The first phase is the preliminary phase, which involved the SARS-CoV-2 infection (COVID-19 and Pneumonia) in datasets comprising CT scans and chest X-ray images. Section 3.1 offers the methodology of in-depth transfer learning model development and the definition of the experimental parameters. Section 3.2 presents the performance metrics, description, and results of the pre-trained models.
- The second phase of research involves developing ML-based humans in the loop system to annotate, segment, and classify SARS-CoV-2 infection from Covid-19, Pneumonia, by the CT

scan images. Section 4.1 presents the methodology, including designing a novel descriptor for the feature extraction from the patches, ML model design, learning, development, annotation-segmentation strategy, and evaluation of the proposed method radiologist-in-the-loop machine learning (RILML) System. The experiments and results Sect. 4.2 comprises computational cost, experimental approach, and results recorded. The observation and discussion from the tests are specified in Sect. 5. Finally, Sects. 6 and 7 cover the summary and future works, respectively.

The structure of the article is as follows. Section 2 presents the state-of-the-art techniques and necessity of research. The proposed scheme operates in two phases, namely phase I and phase II. Section 3 presents the phase 1 (transfer learning approach) and Sect. 4 presents the phase II (RILML model). Section 5 presents the observation and discussion of the proposed scheme.
Section 6 presents the conclusion and finally Sect. 7 presents the future work of the article.

State-of-the-art techniques and necessity of research

The false-positive of the current diagnostic gold standard at the primary staging of the COVID-19 is significantly performed the comparative analysis of non-contrast chest CT with RT-PCR and identified that viral nucleic acid is the current ground standard in recognition of COVID-19. Recent research claimed that the recognition of Covid-19 through non-contract chest CT was 98% compared to preliminary RT-PCR sensitivity of 71% [10, 11]. Studies have indicated that COVID-19 can be detected early by combining laboratory results with clinical image features [12]. Radiological images provide useful information for the diagnosis of COVID-19. Narin et al. used data mining methods for distinguishing typical Pneumonia and SARS in X-ray images [13]. CNN-based image processing identifies features in chest X-ray images and cystoscopic images to diagnose pediatric Pneumonia [14]. The X-ray-based imaging technique is more popular because it is economical as compared to a CT scan. In one of the researches, the investigator selected a deep learning-based transfer learning network to diagnose pneumonia by applying the ImageNet model and ensemble learning [15, 16]. Many studies have been conducted on AI-based COVID-19 recognition applied deep learning approaches on CT scan images to identify features associated with COVID-19 [17–19].

Diseases are diagnosed and treated based on medical image segmentation results. Suitable results have been obtained for many biomedical image segmentation tasks using AI techniques [20]. However, in contrast to images of natural scenes, labeled medical data for COVID-19 chest CT scans are expensive and rare. The annotation of COVID-19 images is a tedious task that can be effectively performed only by radiologists. Even radiologists do not have sufficient experience in annotating such images. Image annotation methods [21], active learning methods [22], and weakly supervised segmentation algorithms [23] have been proposed for alleviating the burden of manual annotation. As mentioned earlier, the approaches are applied to analyze natural scene images and not be straightforwardly applied in biomedical applications due to the limited availability of and large variations in biomedical training data [24–27]. Most of the approaches mentioned above do not explicitly exploit the interdependencies among annotations. These methods are not based on a standard lexicon, and they limit their labeling at an advanced level than diagnostics. This research was offering a semiautomatic RILML system. The projected system is founded on the Bayesian tree-structured model and SVM algorithm. It provides preliminary results for COVID-19 chest lesions in CT images. The RILML system allows radiologists to contribute the most crucial evidence in each iteration. Moreover, it uses a network model for updating the full annotation online. The proposed structure's efficiency was examined in this study.

To achieve this aim, we suggest opening the radiologist-in-the-loop machine learning (RILML) process by providing radiologists with labeled images (LIs) proximal to the present image (PI) for increasing the explicitness of the uncertainty and decision rationale. The aforementioned LIs act as supports for machine learning (ML) tools. In supervised methods, these LIs are utilized for training, cross-validation, and testing; however, the LIs are used for only validation and testing in unsupervised methods. If the PI is localized in a classification space and proximal LIs are chosen according to suitable metrics, the LIs can be presented to radiologists as COVID-19 chest CT scan images (their original form) with annotation; thus, the LI classification can be immediately interpreted. Furthermore, the LI density in the PI neighborhood and the classification confidence, COVID-19 chest CT scan image saliency maps, clinical information, and demographics can be obtained by radiologists. Thus, radiologists can determine the salient image regions enriched by LIs and the system output and use it for classification.

Phase 1 (Transfer learning approach)

This section presents the methodology of in-depth transfer learning model development and the definition of the experimental parameters in Sect. 3.1. The performance metrics, description, and results of the pre-trained models are presented in Sect. 3.2.

Methodology and experimental setup parameters

Pre-trained Transfer learning networks are the only solution when you have limited datasets, such as the case of the COVID-19 X-Ray and CT scan dataset [28]. Transfer learning networks are pre-trained from the large dataset and applied for the application with a comparatively smaller dataset. This helps deep researchers to overcome the limitation of the size of the dataset. Additionally, transfer learning networks considerably decrease the lengthy training period, as it is essential for deep learning models when it is designed from scratch [29]. For example, when we look into states of confirmed COVID-19 cases, there are millions of cases. However, the publicly available chest X-ray and CT scan image dataset is comparatively small, full of noise, unannotated, and scattered [28]. The block diagram of the first phase (Transfer Learning approach) research applying a pre-trained network is represented in Fig. 1.

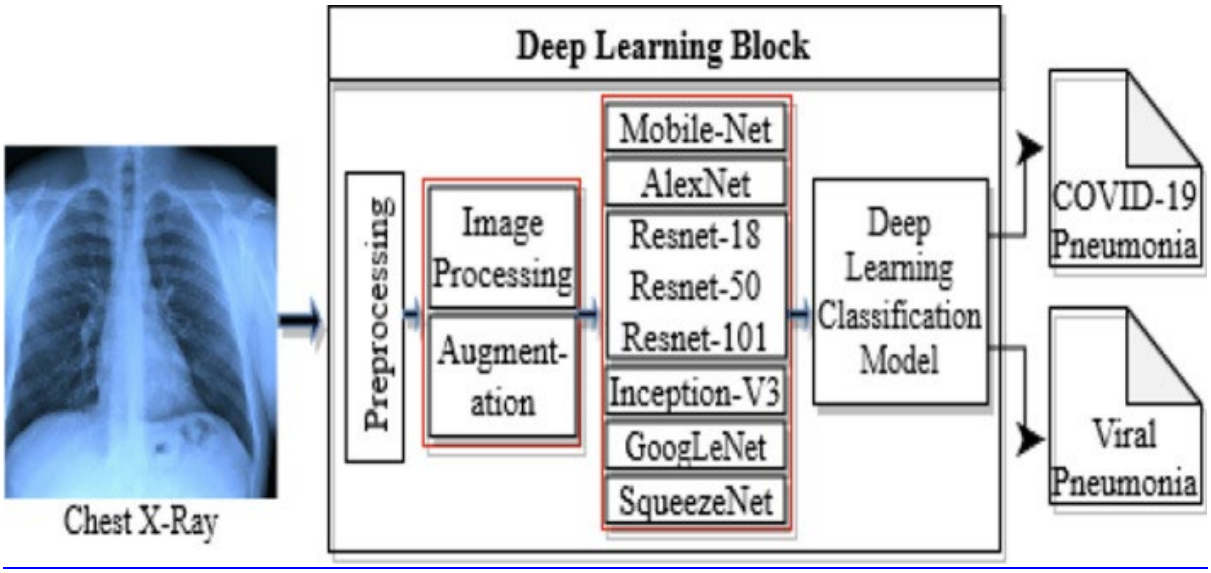

Fig. 1

Block diagram of the deep learning-based system

According to the transfer learning models, the preliminary step of the data pre-processing is to resize the X-ray and CT scan images as image input vary from model to model. Therefore, all the images were standardized according to the pre-trained network protocols and standards [30]. Unfortunately, this is a state where the number of annotations belonging to the COVID19 class is lower than those associated with the Pneumonia class. This problem is prominent in this situation as anomaly detection features belong to COVID19 to detect disease. In this condition, the pre-trained transfer learning model developed using deep convolutional neural networks could be biased and inaccurate. This occurs as deep learning algorithms are characteristically designed to advance accuracy by reducing the error. Thus, they do not give importance to the class distribution/proportion or balance of classes.

We adopted Synthetic Minority Over-sampling Technique (SMOTE) to balance Pneumonia to COVID19 [31]. SMOTE outperformed the conventional oversampling and under sampling methods in the current problem. In this learning, we have applied three augmentation approaches (Rotation, Scaling, and Translation) to produce a fivefold training set of COVID19 images [30]. The rotation process used for image augmentation was completed by spinning the images in the clockwise and counterclockwise direction with an angle of 15°, 30°, 45°, 60°, 75°, and 90° [32]. The scaling procedure is the amplification or lessening of the image's frame size. Image translation was done by transforming the image horizontally and vertically by 5% to 20%. Now, the system investigates the features from the images by spotting which region in the convolutional layers is activated and relating with the equivalent regions in the novel images. The activated region is normalized to 0, 1, as it takes a different range of values that needs to be normalized. The channels with the highest values were compared with the original image. The cross-validation approaches were performed, and the cross-validation approach with the best performance, namely fivefold cross-validation, was selected.

Experiment, results, and discussion

In the present research, we had proposed and implemented a classification of COVID-19 from other pneumonia using pre-trained transfer learning algorithms using both types of Radiographic images (CT scan, X-ray images). The pre-trained networks did not take the manual feature extraction, and it uses the end-to-end structure. Popular pre-trained models such as MobileNet, ResNet18, ResNet50, ResNet101, AlexNet, GoogleNet SqueezeNet, and InceptionV3 have been trained and tested on medical images. The research investigation found that MobileNet performed all other pre-trained networks. The ROC performance of selected transfer learning networks was significantly well during controlled testing from Fig. **2**. However, it dropped severely when the new dataset was applied for the testing, as presented by Fig. **3**.

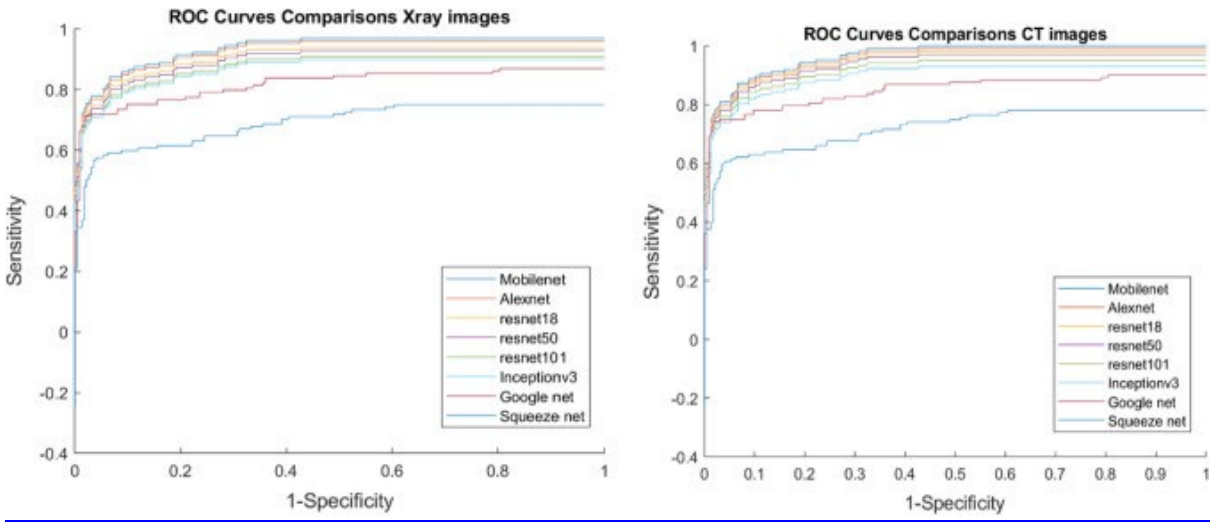

Fig. 2

Evaluation of the ROC curves for the pre-trained networks under a fivefold cross-validation test dataset for X-Ray and CT scan images

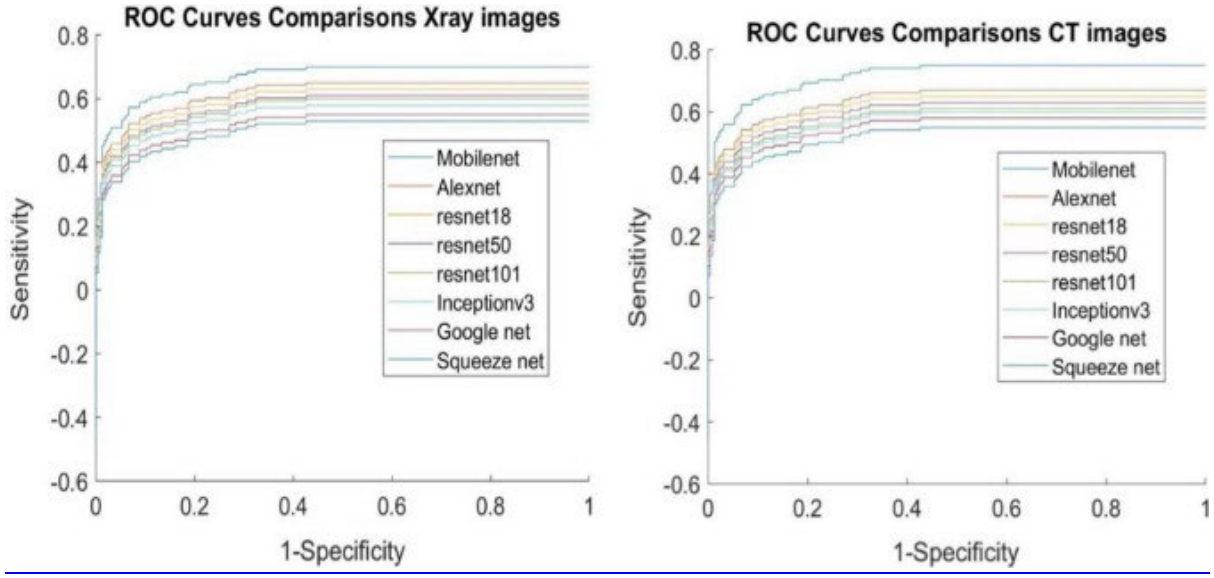

Fig. 3

Comparison of the ROC curves for the pre-trained networks under real-world dataset

Majority of the existing researches based on COVID-19 diagnosis using X-Ray and CT scan images documented the AUC from 0.90 to 1 [13, 30]. Our research using MobileNet secured AUC 1 under the testing dataset, though it was below 0.64 for the real-world new dataset. When these existing AI models are applied to real-world new datasets, they fail to repeat the performance. We tested these AI for COVID-19 models with our dataset, and their AUC was below 0.65 it displays the overfitting and high variance, which misguides the Radiologist instead of assisting in decision-making.

To prevent overfitting, the best arrangement is to utilize more training data, which is impossible in the present cases of Covid-19. As radiologists and doctors are busy in the treatment, they do not have enough to provide data for the research, so data access is limited. The pre-trained model prepared on more information will generally sum up better; otherwise, the best arrangement is to utilize regularization strategies [33]. In general, deep learning models will be acceptable at fitting to the training data, yet the genuine test is a generalization, not fitting. So to see the generalization, we brought a new real-world dataset for the testing. Then again, if the system has restricted memorization assets, it will not have the option to learn the mapping without difficulty. To limit its misfortune, it should learn packed portrayals that have increasingly predictive power. Simultaneously, on the off chance that you make your model excessively little, it will experience issues fitting to the training dataset. There is a trade-off between "too much capacity" and "not enough capacity" [34]. Tragically, there is no mystical equation to decide the model's correct size or design (regarding the number of layers or the right size for each layer). The only solution was to design a new integrated machine learning-based approach which can learn and generalize with inadequate training data.

Additionally, overfitting occurred in the existing pre-trained network implementation because a scheme learned the feature and noise in the training data to the degree that it destructively influenced the deep learning pre-trained model's performance on a new dataset (real-world dataset). This explained that the noise or arbitrary variations in the training data were elected and cultured as features by the model. The difficulty was that those features did not become applicable to the new dataset and deleteriously influenced the pre-trained network's aptitude to generalize. Therefore, the feature extraction methods need to apply, exploring the textures for the feature extraction and discards the error-noise.

Phase 2 (RILML model)

We created an ML-based RILML model to address the above poor performance relating to radiological information to detect COVID-19 from pneumonia patients. Sensitivity, specificity, accuracy, AUC, and Kappa metrics were used to investigate the classification performance. The optimum classification performance was obtained with fivefold cross-validation. A dataset of the offering chest CT scan images from 441 subjects for whom there was medical distress of COVID-19 was received. The skilled radiologists recognized from CT images that COVID-19(+) shows different characteristics from other viral pneumonia. Therefore, medical experts necessitate that COVİD-19 infection needs to be analyzed in the early phase. AI may deliver a technique to augment the timely uncovering of COVID-19 infection [35, 36]. Our objective was to project an AI scheme to classify COVID-19 infection built on preliminary chest CT scans that could swiftly categorize COVID-19 (+) patients in the preliminary stage. For uncovering the COVID-19, datasets were designed by taking patches from CT scan images of COVID-19 (+) and pneumonia. The feature mining process was applied to patches to advance the classification process. Projected research used the support vector machine (SVM) and Bayesian network classifiers to classify patients with COVID-19(+).

Methodology

The proposed RILML system involves image pre-processing, feature extraction through descriptors, a preliminary annotation according to SVM classification, and iterative radiologist-in-the-loop fine-tuning of annotations by applying a Bayesian model of interdependencies. The purpose behind the SVM selection is that it is impervious to over fitting, even in circumstances where the number of attributes is more than the number of observations or annotations. The SVM is an estimated execution of a bound on the generalization error that relies on the margin (fundamentally the

distance from the decision boundary to the closest pattern from for each class). However, it is autonomous of the dimensionality of the feature space. Furthermore, the Bayes classifier uses a classical linear hypothesis function. As an outcome, it undergoes high bias, or error follow-on from erroneousness in its hypothesis class, because its hypothesis cannot correctly characterize various complex circumstances. Alternatively, it displays low variance or failure to specify to concealed data based on its training set. Bayesian hypothesis class' uncomplicatedness stops it from overfitting to its training data. As an outcome of this attribute, the Bayes classifier has been presented remarkably well with minimal volumes of training data in the current Covid-19 CT scan images that most other classifiers would discover considerably inadequate. In the first stage, feature extraction methods mined the features, and the RILML system classified the features, as shown in Fig. 4. In the classification procedure, fivefold cross-validation approaches were applied. The mean classification outcomes subsequently cross-validations were attained. The proposed methodology is divided into two parts, which are described in the following sections.

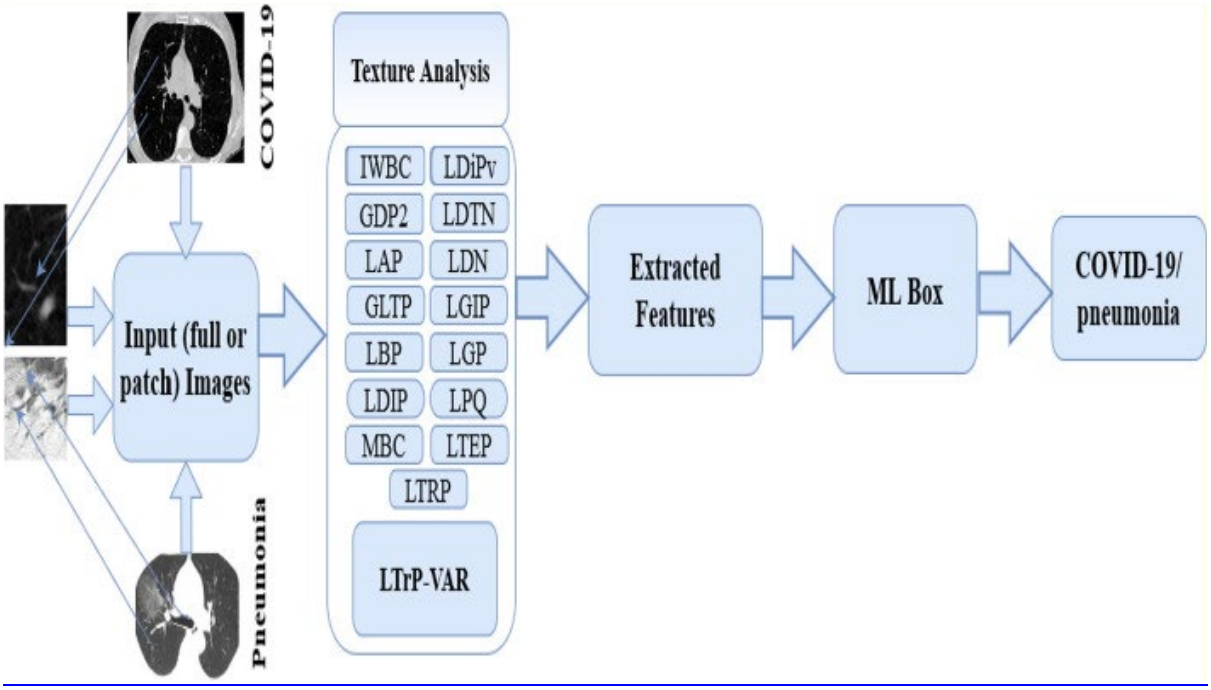

Fig. 4

Proposed textural analysis block diagram

Extraction of features from patches by the descriptor

Texture plays a vital role in biomedical image classification. An image can be extracted from an extensive medical image database according to its texture. To identify COVID-19 infection in CT data, we must first determine the primary feature types that define the image, such as its texture, color, gradient, and shape. Textural highlights are essential for extracting features from a medical image. They provide information regarding spatial tonal variations and object surfaces. Descriptors are used successfully to advance the accurateness of the diagnosis system by picking noticeable features.

In specific, we applied some options of descriptors, which includes gradient directional pattern (GDP.), gradient directional pattern (GDP2), Geometric Local Textural Patterns (GLTP), improved

Weber local descriptor (IWLD.), localized angular phase (LAP), a local binary pattern (LBP), Local directional pattern (LDIP), Local Directional Pattern Varianceldn (LDiPv), Inverse difference moment normalized (IDN), local directional number pattern (LDNP.), Local gradient increasing pattern (LGIP), local gradient patterns (LGP), local phase quantization (LPQ.), Local Ternary Pattern (LTeP), Local tetra pattern (LTrP), Monogenic Binary Coding (MBC), Local Frequency Descriptor (LFD), Local Mapped Pattern (LMP), those are extensively considered state of the art in texture descriptors [7, 30, 37, 38]. We performed new trials using the mentioned descriptors and offered features extracted from the patches of COVID-19(+) CT scan images.

The features mined from the descriptors are fed to the RILML model, LTrP, and LPQ. Performed significantly well as compare to other descriptors. To enhance the performance of RILML model, we modeled a novel descriptor called LTrP-VAR. The proposed descriptor variant is achieved by seeing diverse profiles for the locality calculation and different encodings to investigate the local grayscale variance. Extracted sets of features are then applied for training a machine learning-based RILML system. LTrP-VAR demonstrated salient micro-patterns to distinguish the infective region and train a RILML model to distinguish between the COVID-19(+) infection and the ones who are pneumonia. Most COVID-19 detection approaches depend on lesion detection. However, lesion segmentation may involve uncertainties and inaccuracies, which can cause classification errors. Therefore, in this study, specific lesion segmentation was not performed. Instead, the complete image information was used for decision-making. Also, with LTrP, which contributes a better spatial texture pattern, we figure rotation invariant contrast (VAR) modeling. LTrP and VAR measures complement each other, and we apply their joint functionality to advance the performance of the complete proposed algorithm to classify COVID-19(+) and pneumonia from chest CT scan images.

ML model for classification and segmentation

When expert radiologists encounter diagnostic uncertainties regarding COVID-19 chest CT scan images, they refer to textbook cases similar to the case regarding which uncertainty exists or digital atlases and obtain information regarding the classification confidence. Radiologists accept the classification obtained with an external source only when reasonable classification confidence and satisfactory similarity are achieved. However, if insufficient previous annotated images are available to understand the segmentation of the area infected by COVID-19, radiologists' task becomes challenging. Therefore, we propose an ML model-based outcome inspection strategy that imitates the behavior of radiologists in real scenarios. At present, when ML-based models are used to analyze complex cases, heat maps are matched with ground truth annotations so that radiologists can trust black box systems. In addition, the RILML method allows the processing of large-volume medical image data and the consolidation of multicenter data for big data analytics.

According to our hypothesis, when a PI without any annotation must be analyzed, decision-making support can be provided by similar LIs. Heat maps are useful for identifying the image part that guides the ML model to a verdict but does not indicate the verdict's cause. To achieve a secure link between the aforementioned image part and the classification result, radiologists should match this image part with the ground truth annotation. However, when investigating the PI, ground truth annotations are not obtained. Adding reference images or cases and fuzzy or probabilistic information for providing advanced decision-making support to the Radiologist is not tricky. These resolutions can be implemented, studied, and validated, as depicted in Fig. **5**. The usefulness of the projected scheme is described in the following text. First, it provides radiologists with relevant LIs. Second, in the proposed method, the proximal LIs provide the original images and ancillary information, including annotation agreements (radiologist confidence), annotation masks, heat maps localizing image regions for the classification, validation confidence, and the subject’s clinical

profile and demographics. Third, the proximity of the n-closest LIs indicates the library's density in which the subject lies, which suggests the training robustness and validation relevant to the PI. The proposed model considers the interdependencies among concepts.

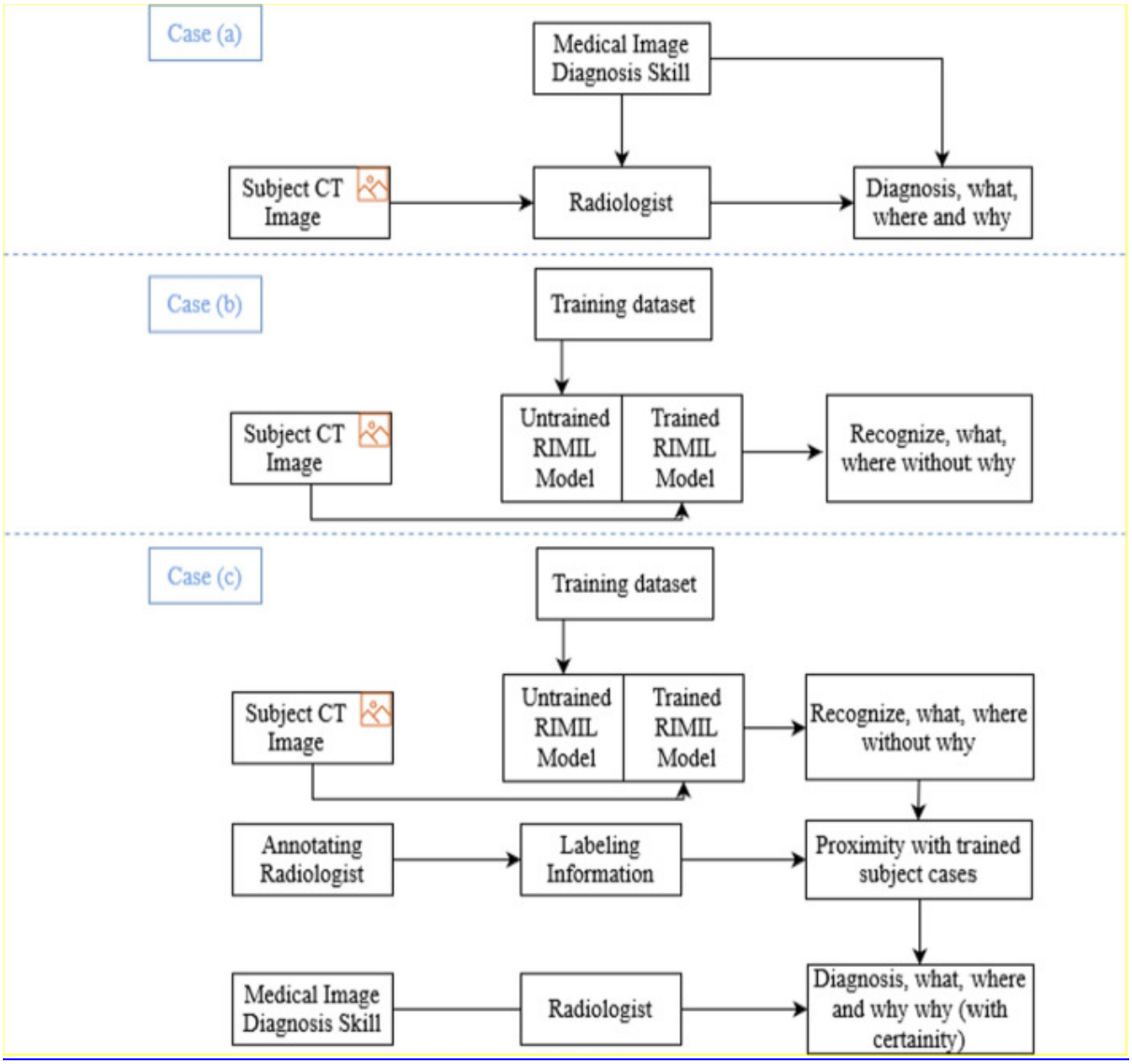

Fig. 5

Outlines the diagnostic procedure supported by machine learning tools to display procedure dissimilarities, **a**, Stand-alone Radiologist, without the assistance of RILML model, **b** Stand-alone RILML without Radiologist input and **c** Radiologist with RILML; classification outcomes (what), lesion localization-segmentation (where), and additional evidence on the diagnosis procedure (why) resultant from the annotated collection

Model formation

Three sets of random variables Xsvm, Xm, and Xh are modeled, represents SVM interpretations, modeled, and hidden variables. These are characterized in the setup as a bottom-up approach in the applied tree representation ranked representations. XSVM are children of Xm in individualized

communication, and Xh are presented by parents to several subsets of Xm to encrypt their dependencies [37].

The variance of low-level image features (256 features LTrP descriptor extracted) unfolding position, outline, and texture of chest and chest lesions are applied. The feature sets are briefing the chest's common graphical properties, all lesions, and replicating advanced levels of visual evidence associated with discrete lesions. The Bayesian network model, represented by Bnet, was built. The clustering mentioned above was applied to present hidden nodes. Bnet has a tiered arrangement through a set of unseen nodes, indexed from set $h = \{r + 1 = 31,\ldots, n\}$, signifying the lesion with conception classes, which are outline, texture (calcification as its sub-category), and locality (with closeness as its sub-category). Each modeled node is connected to an SVM-based observation node representing a multi-class RBF-SVM [37]. Classifier result that was prepared for modeled value. These classifiers order the lesion, utilizing low-level data highlights, into equivalent concept node positions. The projected scheme is tree-structured with Bnet carrying discrete variables [37]. Thus, scheme parameters Mm are conditional probabilities $p(Yi = Q|Xi;V)$ of a pixel Xi to the ith class, where V is the network parameter. Given each child-parent pair (Yi, Xi) in the tree, each of their probable state formations (Q, V).

Model learning

The machine learning model is adjusted with network parameters, and it iteratively records the scheme to take full advantage of the Bayesian Information Criterion (B.I.C.). Assumed a scheme Mp at the recent iteration with a set of hidden node h, assume S is the set of all sibling triplets in Mp with a mutual hidden parent. S's features are arranged in a descendant arrangement in a nested style primary according to the triplets' parent's entropies, E(Xh). A novel unseen node is introduced as the parent of the uppermost triplet in S and forming a contestant network model M. Forms all parents were exchanging transfers on the contestant network model, where a parent exchanging move is definite by means of varying the parent of a node from one unseen node to the next. The model learning executes the move that surges the BIC grade of the contestant network the most [37]. Then, the hidden nodes with one or no children are removed [31, 39]. The contestant network M is recognized as the new Mp if its BIC score is higher than the preceding model's; otherwise, S's next triplet is measured. The practice dismisses when none of the triplets in S succeeds for unseen node insertion.

The certainty score for a region Xh is defined as

;PXh=maxp(Ypi=i|Xh;V)

1

P_{Xh} is the maximum probability of the lesion area to all the sample pool PIs. In PIs, Xh's label is unidentified, and consequently, a higher P_{Xh} reflects greater certainty for the annotation. To correctly label hard samples, we simply define the selection criteria for hard samples in the segmentation task. The hard samples are small compared to easy background samples; easy samples are considered to be discovered, so add less to the development. Whereas the hard samples (e.g., mimics) are small, complex to be distinguished, and consequently offer more information to the model learning.

The entropy of a particular region of interest directly can be understood by local consistency Lc. Lc is defined as the average certainty score of the region forecasted as the lesion, but the probability is less than a threshold Ш.

$$L_c = \frac{\sum_{k=1}^{Ɲc} ÆP_{Xh}}{\sum_{k=1}^{Ɲc} P_{Xh}}$$

2

where Ɲc is the number of pixels recognized as the object COVID lesion. If *P*Xh < *Ɯ*, then *Æ* is 1. Otherwise, *Æ* is zero. If the local consistency is low, the object region's edge is not consistent and hence cannot be RILML annotated. The active learning approach is designed according to local consistency performance. If the confidence score *P*Xh of the pixel Xh is less than the threshold value Ɔ, the ML model's annotation is untrustworthy, and consequently, pixel Xh will not be machine labeled. Thus, the pixels with maximum confidence scores are machine labeled and contribute to the training model's fine-tuning procedure. As the performance of the ML is incrementally improved with iterations, we introduced a loss function. The loss function is defined as follows: Min (*X*, *V*, *Ɯ*, *ƴ*). The parameters *Ɯ* and *ƴ* for active learning are restructured during the training process. *Ɯ* is a threshold to confirm that the selected pixel is the object edge, and this edge is not reliable. *ƴ* characterizes the share of RILML annotations in the iteration. The approach is defined as where Ɯ0 and ƴ0 are the initial thresholds. Δ*Ɯ* and Δ*ƴ* control threshold variation with respect to the number of iterations.

Annotation strategy and classification

This paper presents a RILML scheme for annotating the *concept* states iteratively in a specific COVID-19 CT lesion image. The proposed scheme simultaneously allows users to provide modeled annotations in a definite order and update the probable states of annotations that have not yet been annotated according to the collected indication. An appropriate model is selected for user annotation in each step, which raises the confidence of unresolved assessments with a minimal extent of user energy until a discontinuing principle is fulfilled. An active learning approach and the Bayesian network model were combined to annotate CT scan images. First, the network model was trained using a small dataset. The trained model was then used for testing unlabeled data. Subsequently, a doctor judged the test performance's acceptability (fail or pass analysis of the entire dataset). If the test results did not meet the doctors' requirements, the uncertainty extracted with a well-trained ML network was used for determining which data had to be annotated. We retrained the model until a satisfactory performance was obtained by incorporating new annotation applicants into the novel training data. Finally, a stable model was achieved by annotating only useful data rather than all the unlabeled data. This procedure is better explained in the below algorithm 1.

Algorithm 1: Iterative Annotation

```
Train RILML applying a first train set, initialize the set
Verify the set PIs of unannotated cases with the correctly trained model by LIs;
Repeat
    Experts check;
    If the conduct is not suitable, Then
        Inquiry the furthermost ambiguous cases X of PIs, by measuring an average of the
        probability of the top two choices
        Pre-annotating applying correctly trained model LIs;
        Annotate the new enquired samples X by the Radiologist;
        Introduce the new annotated cases X within the train set and eliminate cases from the
        unannotated sample set PIs;
        Retrain the RILML applying the weights of the latest sequences commencement;
        Test the reorganized set of PIs unlabeled cases using the novel correctly trained
        model.
    End
Until the segmentation presentation is reasonable;
```

Evaluation of the RILML system

The process of medical image annotation was used for training, validating, and testing ML models. Thus, we obtained noteworthy ratings and clinical knowledge, which were ultimately encrypted within the trained model. However, as depicted in Fig. 6, the aforementioned information is not transferred to the clinician in charge of the PI, who must rely on their experience to justify the model prediction. Therefore, a communication barrier is created even when the entire procedure from development to application achieves classification rule consensus and provides public clinical knowledge. LIs from the library better convey the aforementioned information. The portion of the library relevant to the PI is triggered periodically according to a proximity concept. Thus, a radiologist can exploit the localization and classification capabilities of the model as well as examine reference cases to verify their decision (Fig. 6).

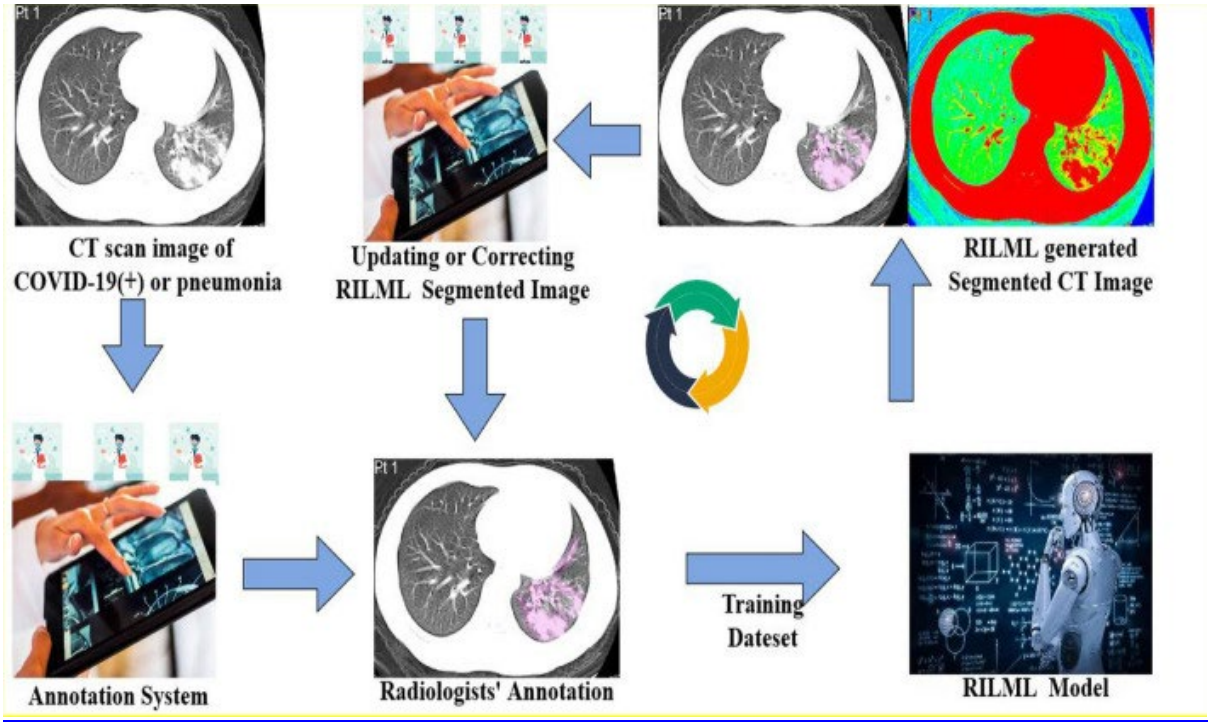

Fig. 6

Proposed RILML workflow

Moreover, the proposed scheme allows the detection of poorly addressed cases by the model; thus, users can provide feedback to designers. The obtained feedback can be verified and used to develop and test enhanced model versions prior to certifying and delivering them to the clinical community. Improvements in medicine are frequently achieved through the sharing of empirical observations among the clinical community.

Experimental result and approach

SARS-CoV-2 infection was detected and classified with the proposed model using CT scan images. The infection was classified into two categories: COVID-19 and Pneumonia. The proposed model's performance was assessed using fivefold cross-validation for the binary class. It is well recorded that the loss values increased significantly at the beginning of training and decreased considerably. This variation occurred due to the number of images in the COVID-19 class, which was considerably lower than in the pneumonia class. However, the magnitudes of the aforementioned rapid increases and decreases gradually declined in the latter part of the training when the proposed system repeatedly examined all the CT scan images.

A three-level hierarchical procedure was performed to segment the COVID-19 CT region in images. First, isotropic resampling was conducted on the extracted volume, which was subsequently processed using an edge-enhancing diffusion filter for noise suppression. Next, a modified MaxFlow or MinCut algorithm was used to segment the chest. In this algorithm, the shape representation based on the Poisson equation was used to generate chest boundary maps on 1D across-boundary CT profiles through autonomously trained KNN classifiers ($K = 20$). To avoid errors due to image processing, all the segmentations were manually verified and corrected if required. Each lesion's

low-level image features were computed. These features were used in separately trained radial basis function SVM classifiers to obtain markings for the observation nodes.

The RILML model SVM-based annotation method, which uses the SVM algorithm and linear collaborations of steerable Riesz wavelets, was compared with the proposed system to assess the projected plan's strength in contradiction of autonomous annotations (preliminary observations). We extracted 2D cross-sectional images from each lesion, selected image patches arbitrarily from peripheral and internal regions, and generated feature vectors to perform the comparison [37]. The study of specific SVM classifiers, trained on features, is useful to each concept value associated with texture and shape-related modeled value set to obtain a probability. Therefore, the SVM observations (X_{SVM}) setting depended on the maximum probability among the related value sets, and the proposed iterative online annotation was used [37].

The training data were divided into many groups. Initially, radiologists hand-labeled some CT images belonging to a group with a small amount of data. Descriptor methods were then used to extract features. Subsequently, the ML annotation model was trained as an initial model by using the aforementioned data group. This model was used to annotate infection areas in the images belonging to the following group. The radiologists performed manual checks to improve the annotation results obtained with the ML annotation model. The modified annotation outcomes were then used as new training data, and model retraining was performed with an augmented training dataset. The procedure as mentioned above resulted in a repeated increase in the training dataset's size and the final ML model generation. In the testing phase, infected regions were annotated on new CT images by using the trained annotation model. The proposed approach performs well after 4–5 repetitions.

Observation and discussion

Chest CT is a popular tool for diagnosing patients who are suspected of having a pulmonary infection. During the COVID-19 outbreak, chest CT has been extensively used in medical practice in certain countries, such as India, South Korea, and China, due to its availability and speed. According to the WHO, the most accurate COVID-19 diagnosis test involves detecting nucleic acid in secretion fluid collected from a throat swab by using the RT-PCR test. However, nucleic acid detection kits are scarce and provide outcomes in up to 2 days. Therefore, the chest CT has also been projected as an essential COVID-19 diagnostic tool. In addition, studies have used AI techniques to distinguish COVID-19 from other forms of pneumonia according to only chest CT images. However, the use of chest CT scans in diagnosis involves two potential limitations. First, health systems may be overburdened during epidemics. Consequently, radiologists may be unable to interpret CT scans on time. Second, the severity and morphology of pathological findings vary on CT scans. More specifically, mild cases may exhibit few or no anomalous findings on chest CT scans.

The limitations as mentioned earlier can be overcome using the proposed model. First, the proposed model can be used to assess CT scans immediately after their completion. Second, it supports radiologists in identifying COVID-19-positive patients who exhibit normal CT results in the primary phase of the sickness. Third, the proposed model reduces segmentation. In particular, the proposed model exhibited a significantly higher AUC than the pre-trained model that uses only CT images. Fourth, the proposed model can be run as an application on a simple workstation to assist radiologists. Finally, the proposed model must be integrated with communication systems, radiology picture archiving systems, clinical database systems, or other image storage databases, easily achieved in modern hospitals for hospital use.

Feature extraction and machine learning performance

The RILML system is based on texture feature mining and a machine learning classification approach. The features mined from the above the mentioned descriptor were applied to develop a machine learning algorithm and performance under 5-cross fold verification recorded. The modeled machine learning algorithm performed well for the local tetra pattern (LTrP) descriptor; the rest descriptors did not significantly perform. The testing under patches generated features (by LTrP), algorithm recorded sensitivity (Sen) 0.83 ± 0.05, (Spe) specificity 0.95 ± 0.04, kappa (*K*) 0.80 ± 0.03, and AUC 0.91. Whereas the results for testing under full image generated features, algorithm displayed Sen 0.87 ± 0.04, Spe 0.96 ± 0.02, *K* 0.83 ± 0.03, and AUC 0.97. When a similar approach applied and tested under real-world new data, the performance decreased, for patches generated features Sen 0.73 ± 0.06, Spe 0.95 ± 0.03, *K* 0.70 ± 0.05, and AUC 0.80 and for the full image generated features Sen 0.67 ± 0.04, Spe 0.76 ± 0.05, *K* 0.63 ± 0.05, and AUC 0.82 [3]. To enhance the classifier algorithm's performance, we derived a new descriptor variant LTrP-VAR, which improved performance. For the testing under patches generated features Sen 0.88 ± 0.03, Spe 0.94 ± 0.04, *K* 0.85 ± 0.05, and AUC 0.94 were recorded, and for the full image generated features Sen 89 ± 0.03, Spe 96 ± 0.02, *K* 0.85 ± 0.01, and AUC 0.98 were recorded. This performance dropped slightly for testing under real-world new dataset patches generated features Sen 0.80 ± 0.03, Spe 0.97 ± 0.01, *K* 0.79 ± 0.02, and AUC 0.88 were recorded, and the full image generated features Sen 0.71 ± 0.02, Spe 0.79 ± 0.05, *K* 0.70 ± 0.02, and AUC 0.88 was recorded [3]. Tables **Tables11** and **and22** present the comparison of test vs real-world database CT images using multiple textural descriptors.

Table 1

Patches annotated CT images texture analysis on test database versus real-world database

Patches	**Test database**				**Real-world database**			
Descriptors	**Kappa**	**Sensitivity**	**Specificity**	**AUC**	**Kappa**	**Sensitivity**	**Specificity**	**AUC**
GDP	0	0.5	1	NA	0	0.30	0.9	0.5
GDP2	0	0.5	1	NA	0	0.20	0.8	0.43
GLTP	0	0.5	1	NA	0	0.20	0.7	0.49
IWLD	0.05	0.51	1	0.76	0.03	0.49	0.7	0.59
LAP	0.33	0.6	1	0.80	0.26	0.5	0.9	0.66
LBP	0.77	0.85	0.9	0.85	0.67	0.75	0.8	0.80
LDIP	0	0.5	1	NA	0	0.2	0.65	0.55
LDIPV	0.63	0.73	1	0.86	0.53	0.64	0.92	0.79
IDN	0	0.5	1	NA	0	0.30	0.6	0.45

Patches	**Test database**				**Real-world database**			
Descriptors	**Kappa**	**Sensitivity**	**Specificity**	**AUC**	**Kappa**	**Sensitivity**	**Specificity**	**AUC**
LDNP	0	0.5	1	0,5	0	0.50	0.9	0.53
LGIP	0.36	0.66	0.71	0.68	0.26	0.59	0.601	0.63
LGP	0	0.5	0	NA	0	0.50	1	NA
LPQ	0.69	0.77	1	0.88	0.59	0.67	0.90	0.82
LTEP	0.58	1	0.71	0.85	0.48	0.90	0.61	0.78
LTrP	0.80	0.84	1	0.92	0.71	0.76	0.95	0.8
MBC	0	0.50	0	0.50	0	0.30	0.8	0.43
LFD	0.47	0.66	1	0.83	0.37	0.55	0.86	0.75
LMP	0.28	0.58	1	0.80	0.26	0.51	0.99	0.70
LTrP-VAR	0.852	0.88	1	0.94	0.79	0.80	0.97	0.88

Table 2

CT images texture analysis on test database versus real-world database

Full images	**Test database**				**Real-world database**			
Descriptors	**Kappa**	**Sensitivity**	**Specificity**	**AUC**	**Kappa**	**Sensitivity**	**Specificity**	**AUC**
GDP	0.58	0.72	0.92	0.81	0.48	0.62	0.82	0.73
GDP2	0.77	0.83	0.96	0.9	0.6778	0.73	0.86	0.80
GLTP	0.75	0.8	1	0.87	0.7	0.7	0.91	0.805
IWBC	0.97	0.97	1	0.98	0.8722	0.87	0.91	0.89
LAP	0.69	0.82	0.87	0.84	0.5944	0.72	0.77	0.74

Full images	**Test database**				**Real-world database**			
Descriptors	**Kappa**	**Sensitivity**	**Specificity**	**AUC**	**Kappa**	**Sensitivity**	**Specificity**	**AUC**
LBP	1	1	1	1	0.88	0.89	0.95	0.931
LDIP	0.69	0.76	1	0.88	0.69	0.76	1	0.84
LDIPV	0.83	0.89	0.94	0.91	0.733	0.79	0.84	0.81
IDN	0.72	0.82	0.9	0.86	0.62	0.72	0.8	0.76
LDTN	0.63	0.76	0.89	0.89	0.53	0.56	0.69	0.61
LGIP	0.38	0.63	0.85	0.72	0.3881	0.63	0.75	0.68
LGP	0.55	0.75	0.81	0.89	0.4556	0.65	0.71	0.67
LPQ	0.83	0.85	1	0.94	0.6333	0.65	0.8	0.71
LTEP	1	1	1	100	0.6	0.75	0.8	0.77
LTRP	0.83	0.87	0.96	0.97	0.6333	0.67	0.76	0.71
MBC	0.86	0.87	1	0.96	0.4611	0.67	0.81	0.73
LFD	0.4	0.65	0.83	0.74	0.45	0.66	0.84	0.74
LMP	0.57	0.7	0.60	0.73	0.5956	0.73	0.64	0.67
LTRP-VAR	0.85	0.89	0.96	0.98	0.7	0.71	0.79	0.64

When we analyzed, we found the transfer learning model had a better performance than the proposed RILML model under training and controlled training, whereas, during real-world new datasets used, pre-trained networks showed a significant decrease (MobileNet AUC controlled testing 0.97 and real-world testing 0.67). In contrast, the projected machine learning model performs decently during controlled testing AUC (0.94 (patches) and 0.98 (full image) to real-world new dataset AUC (0.88 (patches) and 0.88 (full image) with a slight decrease in performance. Tables **Tables33** and **and44** show the performance metric of the deep learning model on CT/X-rays images using with/without augmentation (Tables **(Tables55** and **and66**).

Table 3

Performance metrics for deep learning models [CT-Scan] (without augmentation)

Classification tasks	Deep learning models	Accuracy	Sensitivity	Specificity	Precision	F1 scores
	MobileNet	0.9756	0.9782	0.9729	0.9739	0.9760
	AlexNet	0.9413	0.9471	0.9356	0.9348	0.9413
	ResNet-18	0.9065	0.9083	0.9048	0.9043	0.9063
	ResNet-50	0.8587	0.8511	0.8667	0.8696	0.8602
	ResNet-101	0.8043	0.8182	0.7917	0.7826	0.8000
	Inception-V3	0.7755	0.7826	0.7692	0.7500	0.7660
Normal Pneumonia and COVID19 Pneumonia	GoogLeNet	0.7646	0.7671	0.7625	0.7304	0.7483
	SqueezeNet	0.7604	0.7626	0.7586	0.7261	0.7439

Table 4

Performance metrics for deep learning models [CT-Scan] (with augmentation)

Classification tasks	Deep learning models	Accuracy	Sensitivity	Specificity	Precision	F1 scores
	MobileNet	0.9770	0.9777	0.9762	0.9762	0.9770
	AlexNet	0.9762	0.9777	0.9747	0.9746	0.9762
	ResNet-18	0.9444	0.9375	0.9516	0.9524	0.9449
	ResNet-50	0.9246	0.9280	0.9213	0.9206	0.9243
	ResNet-101	0.8849	0.9008	0.8702	0.8651	0.8826
	Inception-V3	0.8532	0.8678	0.8397	0.8333	0.8502
Normal Pneumonia and COVID19 Pneumonia	GoogLeNet	0.8214	0.8347	0.8092	0.8016	0.8178
	SqueezeNet	0.7649	0.7680	0.7638	0.7619	0.7649

Table 5

Performance metrics for deep learning models [X-ray] (without augmentation)

Classification tasks	Deep learning models	Accuracy	Sensitivity	Specificity	Precision	F1 scores
	MobileNet	0.9883	0.9860	0.9906	0.9860	0.9883
	AlexNet	0.9852	0.9889	0.9814	0.9816	0.9852
	ResNet-18	0.9350	0.9355	0.9346	0.9346	0.9350
	ResNet-50	0.9085	0.9108	0.9061	0.9065	0.9087
	ResNet-101	0.8439	0.8459	0.8420	0.8411	0.8435
	Inception-V3	0.7921	0.7907	0.7934	0.7944	0.7925
Normal Pneumonia and COVID19 Pneumonia	GoogLeNet	0.7547	0.7619	0.7523	0.7477	0.7547
	SqueezeNet	0.5810	0.5128	0.6652	0.6542	0.5749

Table 6

Performance metrics for deep learning models [X-ray] (with augmentation)

Classification tasks	Deep learning models	Accuracy	Sensitivity	Specificity	Precision	F1 scores
	MobileNet	0.9944	0.9958	0.9930	0.9930	0.9944
	AlexNet	0.9884	0.9907	0.9861	0.9860	0.9883
	ResNet-18	0.9417	0.9439	0.9398	0.9395	0.9417
	ResNet-50	0.9186	0.9206	0.9167	0.9163	0.9184
	ResNet-101	0.8953	0.8972	0.8935	0.8930	0.8951
	Inception-V3	0.8256	0.8271	0.8241	0.8233	0.8252
Normal Pneumonia and COVID19 Pneumonia	GoogLeNet	0.7786	0.7804	0.7778	0.7767	0.7786
	SqueezeNet	0.7326	0.7336	0.7315	0.7302	0.7319

Human-in-the-loop strategy performance

The additional information provided by the proposed strategy may cause an increase in the reporting time. However, regular close inspection of similar cases should be avoided. Such inspection should be mainly conducted for critical cases, determining systematic classification flaws, and ML algorithm debugging (e.g., enriching a poorly represented class in the validation and training sets). Furthermore, additional information regarding the system decision can be provided on demand when required. The most suitable clinical decision support method involves providing well-

explained and objective classification confidence indices specific to the PI (e.g., the proximal classification space density for similar cases).

Figure 7 shows the comparative R.O.C. analytics between RILML systems, a consensus of three radiologists, and RILML system-Radiologist joint performance. We recorded that Radiologists' performance (three radiologists R1, R2, and R3) alone was always better than the stand-alone RILML system, as displayed in ROC graph Fig. 7. The performance of R1 (Sen 0.89 ± 0.04, Spe 0.86 ± 0.02, *K* 0.79 ± 0.03, and AUC 0.85), R2 (Sen 0.88 ± 0.03, Spe 0.85 ± 0.04, *K* 0.783 ± 0.03, and AUC 0.83), and R3 (Sen 0.90 ± 0.04; Spe 0.87 ± 0.03, *K* 0.80 ± 0.01, and AUC 0.91). Whereas we observe the joint operation, RILML system-Radiologist performs best in these three [3]. RILML Stand-alone system recorded lower performance indices (Sen 0.82 ± 0.05; Spe 0.84 ± 0.05, *K* 0.71 ± 0.03, and AUC 0.81). On the other hand, the joint venture of the RILML system-Radiologist shown notable figures (Sen0.93 ± 0.02; Spe 0.91 ± 0.01, *K* 0.87 ± 0.02, and AUC 0.93).

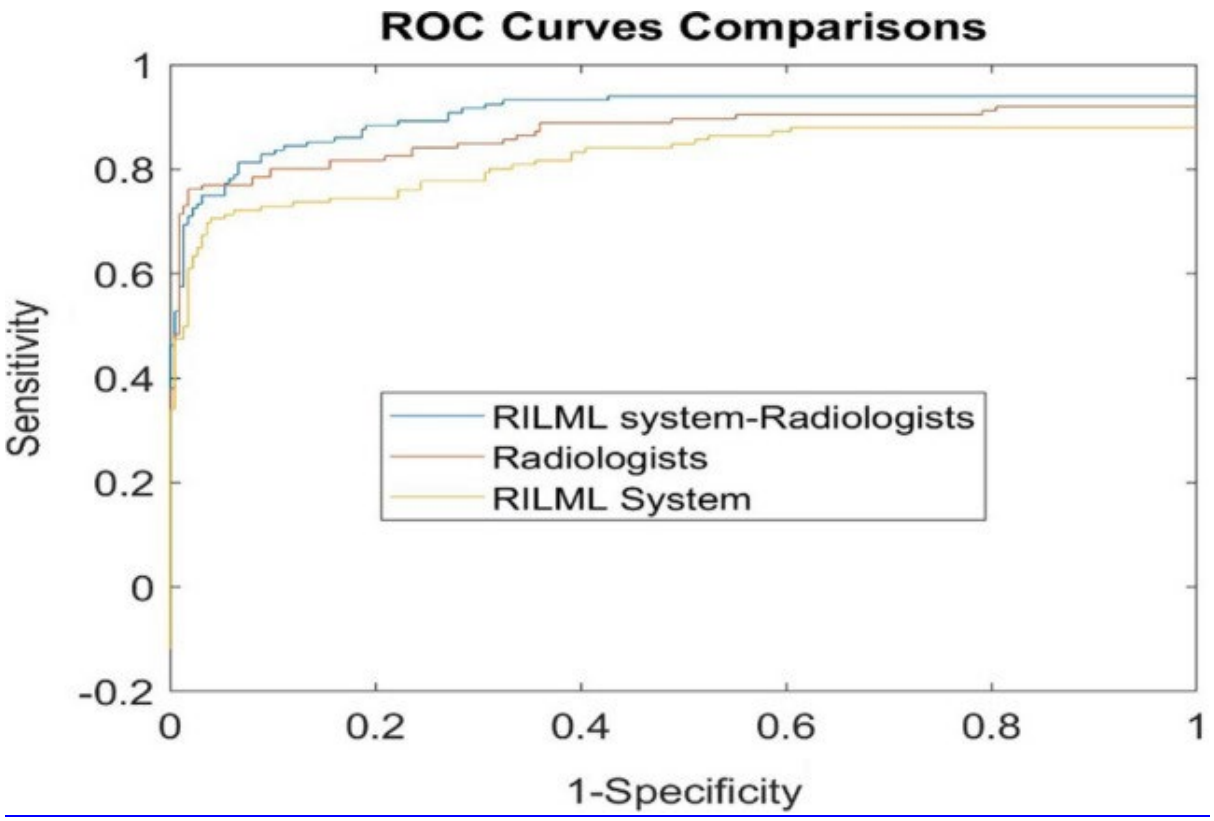

Fig. 7

Comparative ROC between RILML, Radiologists, and RILML with Radiologist

The proposed strategy was evaluated using two metrics. First, the time taken by the proposed model for labeling a CT scan was compared with the time required for manual contouring. Second, we assessed the proposed model's segmentation accuracy in different stages to determine whether an increase in the annotated training data increased the model accuracy. The manual contouring of COVID-19 infection regions on one CT scan required 187 ± 38.5 min. The contouring time decreased considerably to 59 ± 6.3 min when the proposed model was trained with 48 annotated CT scans and used. The contouring time further decreased to 31 ± 5.2, 17 ± 2.5, and 6.2 ± 0.58 min when the

proposed model was trained with 107, 168, and 275 annotated CT scans, respectively, and used. The proposed model's segmentation accuracy was evaluated by calculating the Dice similarity coefficient for the entire validation dataset (550 CT scans). The segmentation accuracy increased from 74 ± 15.4% when using 48 samples for training to 81 ± 9.7%, 85 ± 5.3%, and 90 ± 2.9% when using 107, 168, and 275 samples for training, respectively. The high segmentation accuracy considerably decreases human intervention and, thus, the time required for annotation and labeling.

Structure learning is crucial for successfully implementing probabilistic graphical schemes because the arrangement is the primary factor affecting the model's generalizability, and it defines the statistical relationships between concepts. Structure learning involves searching the space for all possible networks by exploiting the model's capacity to explain the observed (training) data while avoiding overfitting. Because the exploration space is generally huge, structure learning algorithms use heuristics to increase the speed of what would otherwise be a close search concerning a generic cost function. In this study, the B.I.C. score was used to achieve performing structure learning. The results presented in Fig. 8 indicate that complex structures with high human-in-the-loop performance, such as opening from a straightforward naïve Bayes model, can be learned through structure learning. Table Table77 shows the Performance Metrics for RILML model and Radiologists.

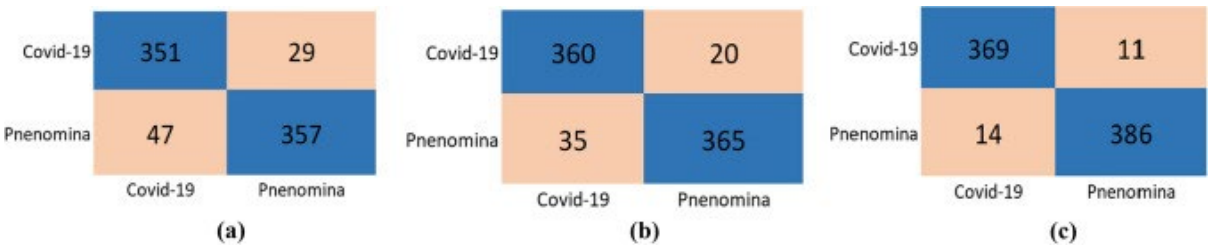

Fig. 8

Confusion Matrix of **a** RILML, **b** Radiologist, and **c** Radiologist + RILML

Table 7

Performance metrics for RILML model and radiologists

	Radiologists only	First iteration	Second iteration	Third iteration	Fourth iteration
Time (min)	187 ± 38.5	59 ± 6.3	31 ± 5.2	17 ± 2.5	6.2 ± 0.58
Dice coefficient (%)	Not applicable	74 ± 15.4%	81 ± 9.7%	86 ± 5.3%	90 ± 2.7%
Number of images	1	48	107	168	275

The aforementioned results suggest that the RILML model can replace high-cost pre-trained transfer learning networks that are prone to data overfitting if the training datasets are incomplete and sparse. Thus, the human-in-the-loop network can overcome the limitations caused by incomplete

domain representation or sparse training data. If rich data are available, the proposed structure learning approach can construct a useful RILML system. The pre-trained and ML networks exhibit a smaller performance gap after model learning than before model learning. This finding verifies the status of data learning ability and hidden variables. The results obtained for the pre-trained structure subsequently model learning indicate that the neural networks clustered texture, calcification, and proximity nodes to a great extent but distributed shape and location nodes more. The behavior of pre-trained networks or data-centric structure learning algorithms depends on the data used. Therefore, if scarce data are available, these networks are prone to overfitting. This finding should be confirmed on massive datasets.

Images reflected incomprehensible by any of the three radiologists were discarded from the investigation. Of the 810 images (390 COVID and 420 pneumonia), 780 (380 COVID and 400 pneumonia) were productively read by Radiologists at *Symbiosis University Hospital and Research Centre (SUHRC), SIU campus at Lavale, Pune, India*. Radiologists were not mandatory to list reasons for eliminating CT scan images, though, where remarks were delivered these associated with deprived image quality initiated by unskilled CT scan operators and inappropriate patient placing. Below, Fig. **8** presents the confusion matrix of Radiologists alone, with RILML and RILL alone. 360 images were COVID-19(+) read as COVID-19(+), 35 images those were pneumonia recognized as COVID-19(+), 20 images those were COVID-19(+) labeled as pneumonia, 365 subjects were pneumonia marked as pneumonia. This shows the mixed performance of Radiologists when they operate standalone. Whereas the performance of RILML model standalone was slightly poor than the Radiologists standalone, 351 images those were COVID-19(+) read as COVID-19(+), 29 images those were pneumonia recognized as COVID-19(+), 47 images those were COVID-19(+) labeled as pneumonia, 357 subjects were pneumonia marked as Pneumonia by RILML model standalone. The best performance recorded when the RILML model-Radiologists collaborated, 369 images those were COVID-19(+) read as COVID-19(+) correctly, 11 images those were pneumonia recognized as COVID-19(+), 14 images those were COVID-19(+) labeled as pneumonia, 386 subjects were pneumonia marked as pneumonia by the joint venture.

Assessment of the model outputs by the radiologists

This section presents the interpretations provided by expert radiologists and obtained with the proposed model. The RILML model can automatically detect COVID-19 infection by using CT scan images and without any feature extraction procedures. The proposed model provides a subsequent estimation to expert radiologists working in health centers. It can save diagnosis time; thus, specialists can provide more attention to severe cases. Moreover, the Grad-CAM heat map method was used to visualize the proposed model's decisions. The heat map indicates the areas focused on by the model in the CT scan analysis. Thus, we ensured that a radiologist could approve the model outcome.

The comments of the radiologists on the proposed model’s output were as follows:

- The proposed model exhibited an outstanding COVID-19 detection performance in the binary classification task.
- The proposed model successfully detected COVID-19.
- Images of patients infected with pneumonia were also included in this research. The proposed model diagnosed some subjects with COVID-19 as having pneumonia. However, COVID-19 is a type of pneumonia. Thus, in the aforementioned cases, the model diagnoses were correct, but the interpretations were wrong.

- The proposed RIML model is sensitive to pneumonia detection. Although it can positively forecast pneumonia and labeled as no discoveries in the dataset.
- The proposed scheme provided wrong forecasts for low-quality CT scan images and for a subject with ARDS, for whom the lung image is diffuse and considerable lung ventilation is missing.
- The proposed scheme can accurately notice COVID-19 with a heat map in standard subjects. However, its efficiency decreases for cases of ARDS and Pneumonia. The heat map exhibited a lower concentration area for the CT scan images of patients without COVID-19 than those with COVID-19
- The proposed model is adequate for assessing the effectiveness of the treatment based on the heat map. Moreover, it can assist experts in the diagnosis, isolation, treatment, and follow-up of patients.

Radiological imaging plays an essential character in the early diagnosis, isolation, and treatment stages of COVID-19. The proposed research can straightforwardly detect nodular opacities, consolidation areas, and GGO, which are pathognomic results obtained from the CT data of the subject with COVID-19. Peripheral, lower lobe, and bilateral involvement are observed in patients with COVID-19, and the planned research can distinguish lesion localization. These researches can be used to identify patients with the primary phases of COVID-19 [39]. Primary disease diagnosis can enable immediate treatment to be provided and disease transmission to be prevented. As mentioned above, the models are also crucial for identifying patients with COVID-19 who do not exhibit early symptoms. The proposed model may provide erroneous diagnoses for patients with significantly reduced lung ventilation and diffuse late lung parenchyma due to the low-quality CT scan images obtained for such patients. Radiologists find it challenging to evaluate low-quality CT scan images. The radiological and clinical images of patients with later-stage COVID-19 are well established, and experts can easily detect the disease in these patients. Deep learning models play an essential role in screening and diagnosis in the case of early-stage infections.

Conclusion

In conclusion, the research results indicate that highly accurate A.I. algorithms can be used for rapidly identifying COVID-19 patients, which would assist the combating of the current COVID-19 outbreak. The proposed model, which has the same accuracy as a radiologist, is useful for quickly diagnosing infectious diseases, such as COVID-19, without physical tests or radiologist inputs. The role of ML in COVID-19-related CT scanning is expanding. An increasing number of studies have shown that computer algorithms can outperform radiologists. However, the aim should not be to determine whether an ML tool can outperform a physician but to determine whether a radiologist can achieve better performance with an ML aid than without it. Interactive cooperation between automated systems and radiologists should be encouraged. Radiologists can integrate their clinical experience by visualizing well-labeled cases classified by the automatic system as proximal to the P.I. Such integration would enable a critical assessment of the automatic tool's performance. In this research, an ML-based model is proposed for detecting and classifying COVID-19 cases from CT scans. The proposed model has an end-to-end structure, is fully automated, and does not require manual feature extraction.

It can also serve as a valuable tool for testing the proposed model's generalizability during the processes of certification and development. Large-scale experiments on clinical and public datasets indicated that the proposed method could effectively alleviate the burden on radiologists for

annotating COVID-19 CT scan images. The proposed approach combines algorithmic effectiveness and clinical feasibility, and it can be directly used in clinics by radiologists. The developed system uses an SVM model and a Bayesian network. Patients diagnosed as COVID-19-positive by the proposed model can be directed to advanced centers for diagnosis confirmation. After diagnosis confirmation, these patients can be treated immediately. Moreover, patients diagnosed as COVID-19 negative by the proposed model can be prevented from unnecessarily undergoing RT-PCR tests and occupying health centers.

Future work

In future research, we will improve the developed annotation platform and algorithm in the following aspects. (1) First, we will design a user-friendly graphical interface. (2) We will collect additional adequate data for testing the proposed method. We plan to use the GAN network [39] to compare a large number of similar Covid-19 CT scans with the ground truth. (3) The annotation platform's accuracy will be improved through boundary marking out. (4) A large dataset with clinical information will be arranged because the spread of COVID-19 is increasing. Different ML approaches will be explored. The proposed model's generalizability will be evaluated at many health centers to validate its robustness. (5) The proposed model will be placed in a cloud so that it can provide instant diagnosis and immediately assist the rehabilitation of affected patients. This step would significantly decrease the clinician workload. Moreover, we will attempt to collect radiology images of COVID-19 patients from Denmark, India, and China and evaluate these data with the proposed model. After the improvements, as mentioned above, are conducted, attempts will be made to position the proposed model in local hospitals for screening.

Appendix

Terminologies	**Description**
AUC	Area Under the Curve
CT	Computed Tomography
ARDS	Acute Respiratory Distress Syndrome
SARS	Severe Acute Respiratory Syndrome
WHO	World Health Organization
SARS-CoV	Severe Acute Respiratory Syndrome- coronavirus
MERS-CoV	Middle East respiratory syndrome- coronavirus
V.S.T	Visual Semantic Terms
P.I	Present Images
L.I	Labeled Images
ML	Machine Learning

Terminologies	Description
AI	Artificial Intelligence
SMOTE	Synthetic Minority Over-sampling Technique
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SVM	Support Vector Machine
RILM	Radiologist-in-the-loop-machine
RILML	Radiologist-in-the-loop-machine learning system
G.D.P	Gradient directional pattern
GDP2	Gradient directional pattern 2
G.L.T.P	Geometric Local Textural Patterns
I.W.L.D	Improved Weber local descriptor
L.A.P	Localized angular phase
LBP	Local Binary Pattern
L.D.I.P	Local directional pattern
LDiPv	Local Directional Pattern Variance
IDN	Inverse difference moment normalized
L.D.N.P	local directional number pattern
L.G.I.P	Local gradient increasing pattern
L.G.P	Local gradient patterns
L.P.Q	Local phase quantization
LTeP	Local Ternary Pattern
LTrP	Local tetra pattern
M.B.C	Monogenic Binary Coding
L.F.C	Local Frequency Descriptor

Terminologies	Description
L.M.P	Local Mapped Pattern
B.I.C	Bayesian Information Criterion
RT-PCR	Reverse transcription polymerase chain reaction
COVID-19	A novel Coronavirus disease

Declarations

Conflict of interest

The authors have no conflict to declare.

Footnotes

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Aylward B, Liang W (2020) Report of the WHO-china joint mission on coronavirus disease 2019 (COVID-19)

2. WHO (2020) Coronavirus disease 2019 (COVID-19) situation Report-24

3. Hwang S, Chung G, Lee J, et al. Sleep/wake estimation using only anterior tibialis electromyography data. Biomed Eng Online. 2012;11:26. doi: 10.1186/1475-925X-11-26. [Europe PMC free article] [Abstract] [CrossRef] [Google Scholar]

4. Carley S, Horner D, Body R, Mackway-Jones K. Evidence-based medicine and COVID-19: what to believe and when to change. Emerg Med J. 2020;37:572–575. doi: 10.1136/emermed-2020-210098. [Abstract] [CrossRef] [Google Scholar]

5. WHO (2020) Laboratory testing strategy recommendations for COVID-19: interim guidance, 21 Mar 2020

6. Ahmad S. A review of COVID-19 (Coronavirus Disease-2019) diagnosis, treatments and prevention. Eurasian J Med Oncol. 2020 doi: 10.14744/ejmo.2020.90853. [CrossRef] [Google Scholar]

7. Jacobi A, Chung M, Bernheim A, Eber C. Portable chest X-ray in coronavirus disease-19 (COVID-19): a pictorial review. Clin Imaging. 2020;64:35–42. doi: 10.1016/j.clinimag.2020.04.001. [Europe PMC free article] [Abstract] [CrossRef] [Google Scholar]

8. Zhang J, Xie Y, Li Y, Shen C, Xia Y (2020) Covid-19 screening on chest x-ray images using deep learning based anomaly detection. arXiv preprint arXiv:2003.12338

9. Ozdemir O. Coronavirus disease 2019 (COVID-19): diagnosis and management. Erciyes Med J. 2020;42:242–248. [Google Scholar]

10. Watson J, Whiting PF, Brush JE. Interpreting a covid-19 test result. BMJ m1808. 2020 doi: 10.1136/bmj.m1808. [Abstract] [CrossRef] [Google Scholar]

11. Naudé W. Artificial intelligence vs COVID-19: limitations, constraints and pitfalls. AI Soc. 2020;35:761–765. doi: 10.1007/s00146-020-00978-0. [Europe PMC free article] [Abstract] [CrossRef] [Google Scholar]

12. Chen H, Ai L, Lu H, Li H. Clinical and imaging features of COVID-19. Radiol Infect Dis. 2020;7:43–50. doi: 10.1016/j.jrid.2020.04.003. [Europe PMC free article] [Abstract] [CrossRef] [Google Scholar]

13. Narin A, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. Pattern Anal Appl. 2021;24:1207–1220. doi: 10.1007/s10044-021-00984-y. [Europe PMC free article] [Abstract] [CrossRef] [Google Scholar]

14. Asif S, Wenhui Y, Jin H, et al. Classification of COVID-19 from chest X-ray images using deep convolutional neural networks. medRxiv. 2020 doi: 10.1101/2020.05.01.20088211. [CrossRef] [Google Scholar]

15. Minaee S, Kafieh R, Sonka M, et al. Deep-COVID: predicting COVID-19 from chest X-ray images using deep transfer learning. Med Image Anal. 2020;65:101794. doi: 10.1016/j.media.2020.101794. [Europe PMC free article] [Abstract] [CrossRef] [Google Scholar]

16. Başaran E, Cömert Z, Çelik Y. Convolutional neural network approach for automatic tympanic membrane detection and classification. Biomed Signal Process Control. 2020;56:101734. doi: 10.1016/j.bspc.2019.101734. [CrossRef] [Google Scholar]

17. Tang S, Wang C, Nie J, et al. EDL-COVID: ensemble deep learning for COVID-19 case detection from chest X-Ray images. IEEE Trans Ind Inf. 2021;17:6539–6549. doi: 10.1109/TII.2021.3057683. [CrossRef] [Google Scholar]

18. Wang S, Kang B, Ma J, et al. A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19) Eur Radiol. 2021;31:6096–6104. doi: 10.1007/s00330-021-07715-1. [Europe PMC free article] [Abstract] [CrossRef] [Google Scholar]

19. Chandra TB, Verma K, Singh BK, et al. Coronavirus disease (COVID-19) detection in Chest X-Ray images using majority voting based classifier ensemble. Expert Syst Appl. 2021;165:113909. doi: 10.1016/j.eswa.2020.113909. [Europe PMC free article] [Abstract] [CrossRef] [Google Scholar]

20. Shi F, Wang J, Shi J, et al. Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. IEEE Rev Biomed Eng. 2021;14:4–15. doi: 10.1109/RBME.2020.2987975. [Abstract] [CrossRef] [Google Scholar]

21. Lin H, Upchurch P, Bala K (2019) Block annotation: better image annotation with sub-image decomposition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp 5290–5300

22. Ghayvat H, Liu J, Mukhopadhyay SC, Gui X. Wellness sensor networks: a proposal and implementation for smart home for assisted living. IEEE Sens J. 2015;15:7341–7348. doi: 10.1109/JSEN.2015.2475626. [CrossRef] [Google Scholar]

23. Ghayvat H, Gope P. Smart aging monitoring and early dementia recognition (SAMEDR): uncovering the hidden wellness parameter for preventive well-being monitoring to categorize cognitive impairment and dementia in community-dwelling elderly subjects through AI. Neural Comput Appl. 2021 doi: 10.1007/s00521-021-06139-8. [CrossRef] [Google Scholar]

24. Chen Y-C, Lin Y-Y, Yang M-H, Huang J-B. Show, match and segment: joint weakly supervised learning of semantic matching and object co-segmentation. IEEE Trans Pattern Anal Mach Intell. 2020 doi: 10.1109/TPAMI.2020.2985395. [Abstract] [CrossRef] [Google Scholar]

25. Maia M, Pimentel JS, Pereira IS, et al. Convolutional support vector models: prediction of coronavirus disease using chest X-rays. Information. 2020;11:548. doi: 10.3390/info11120548. [CrossRef] [Google Scholar]

26. Alahmer H, Ahmed A. Computer-aided classification of liver lesions from CT images based on multiple ROI. Procedia Comput Sci. 2016;90:80–86. doi: 10.1016/j.procs.2016.07.027. [CrossRef] [Google Scholar]

27. Kumar VDA, Sharmila S, Kumar A, et al. A novel solution for finding postpartum haemorrhage using fuzzy neural techniques. Neural Comput Appl. 2021 doi: 10.1007/s00521-020-05683-z. [CrossRef] [Google Scholar]

28. Zhao J, Zhang Y, He X, Xie P (2020) COVID-CT-dataset: a CT scan dataset about COVID-19. arXiv Prepr arXiv arXiv:2003.13865

29. Iwendi C, Bashir AK, Peshkar A, et al. COVID-19 patient health prediction using boosted random forest algorithm. Front Public Heal. 2020 doi: 10.3389/fpubh.2020.00357. [Europe PMC free article] [Abstract] [CrossRef] [Google Scholar]

30. Chowdhury MEH, Rahman T, Khandakar A, et al. Can AI Help in screening viral and COVID-19 pneumonia? IEEE Access. 2020;8:132665–132676. doi: 10.1109/ACCESS.2020.3010287. [CrossRef] [Google Scholar]

31. Mehre SA, Mukhopadhyay S, Dutta A, Harsha NC, Dhara AK, Khandelwal N (2016) An automated lung nodule detection system for CT images using synthetic minority oversampling. In Medical Imaging 2016: Computer-Aided Diagnosis, Vol 9785. International Society for Optics and Photonics, p 97850H

32. Awais M, Ghayvat H, Krishnan Pandarathodiyil A, et al. Healthcare professional in the loop (HPIL): classification of standard and oral cancer-causing anomalous regions of oral cavity using textural analysis technique in autofluorescence imaging. Sensors. 2020;20:5780. doi: 10.3390/s20205780. [Europe PMC free article] [Abstract] [CrossRef] [Google Scholar]

33. Manivannan S, Li W, Zhang J, et al. Structure prediction for gland segmentation with hand-crafted and deep convolutional features. IEEE Trans Med Imaging. 2018;37:210–221. doi: 10.1109/TMI.2017.2750210. [Abstract] [CrossRef] [Google Scholar]

34. Ghayvat H, Awais M, Gope P, et al. ReCognizing suspect and predicting the spread of contagion based on mobile phone location data (COUNTERACT): a system of identifying COVID-19 infectious and hazardous sites, detecting disease outbreaks based on the internet of things, edge computing, and. Sustain Cities Soc. 2021;69:102798. doi: 10.1016/j.scs.2021.102798. [CrossRef] [Google Scholar]

35. Zhang P, Wang C, Kumar N, et al. Artificial intelligence technologies for COVID-19-like epidemics: methods and challenges. IEEE Netw. 2021;35:27–33. doi: 10.1109/MNET.011.2000741. [CrossRef] [Google Scholar]

36. Barnawi A, Chhikara P, Tekchandani R, et al. Artificial intelligence-enabled Internet of Things-based system for COVID-19 screening using aerial thermal imaging. Future Gener Comput

Syst. 2021;124:119–132. doi: 10.1016/j.future.2021.05.019. [Europe PMC free article] [Abstract] [CrossRef] [Google Scholar]

37. Marvasti NB, Yoruk E, Acar B. Computer-aided medical image annotation: preliminary results with liver lesions in CT. IEEE J Biomed Heal Inf. 2018;22:1561–1570. doi: 10.1109/JBHI.2017.2771211. [Abstract] [CrossRef] [Google Scholar]

38. Patel K, Mehta D, Mistry C, et al. Facial sentiment analysis using AI techniques: state-of-the-art, taxonomies, and challenges. IEEE Access. 2020;8:90495–90519. doi: 10.1109/ACCESS.2020.2993803. [CrossRef] [Google Scholar]

39. Ozturk T, Talo M, Yildirim EA, et al. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med. 2020;121:103792. doi: 10.1016/j.compbiomed.2020.103792. [Europe PMC free article] [Abstract] [CrossRef] [Google Scholar]