scientific reports

Check for updates

OPEN

Development and validation of a hybrid deep learning–machine learning approach for severity assessment of COVID-19 and other pneumonias

Doohyun Park[1], Ryoungwoo Jang[2], Myung Jin Chung[3,4], Hyun Joon An[2], Seongwon Bak[2], Euijoon Choi[5] & Dosik Hwang[1,6,7,8]✉

The Coronavirus Disease 2019 (COVID-19) is transitioning into the endemic phase. Nonetheless, it is crucial to remain mindful that pandemics related to infectious respiratory diseases (IRDs) can emerge unpredictably. Therefore, we aimed to develop and validate a severity assessment model for IRDs, including COVID-19, influenza, and novel influenza, using CT images on a multi-centre data set. Of the 805 COVID-19 patients collected from a single centre, 649 were used for training and 156 were used for internal validation (D1). Additionally, three external validation sets were obtained from 7 cohorts: 1138 patients with COVID-19 (D2), and 233 patients with influenza and novel influenza (D3). A hybrid model, referred to as Hybrid-DDM, was constructed by combining two deep learning models and a machine learning model. Across datasets D1, D2, and D3, the Hybrid-DDM exhibited significantly improved performance compared to the baseline model. The areas under the receiver operating curves (AUCs) were 0.830 versus 0.767 ($p = 0.036$) in D1, 0.801 versus 0.753 ($p < 0.001$) in D2, and 0.774 versus 0.668 ($p < 0.001$) in D3. This study indicates that the Hybrid-DDM model, trained using COVID-19 patient data, is effective and can also be applicable to patients with other types of viral pneumonia.

Since the outbreak of coronavirus disease 2019 (COVID-19) in 2019, the pandemic has had a profound and widespread impact, resulting in a significant increase in mortality globally. Particularly, elderly individuals and those have severe underlying medical conditions are at a higher risk of experiencing severe complications[1–5]. This alarming context highlighted the pressing need for a robust severity assessment system to guarantee appropriate care for severe patients. Fortuitously, as of now, COVID-19 is transitioning into the endemic phase. Nonetheless, it is crucial to remain mindful that pandemics related to infectious respiratory diseases (IRDs) can emerge unpredictably. This reality underscores the importance of not only managing the current situation but also preparing for other IRDs such as viral pneumonia (VP) and bacterial pneumonia (BP) to be better equipped for potential future outbreaks.

The presence of multifocal ground-glass opacity (GGO), consolidation, reticular opacity, and crazy-paving pattern in the lung fields is frequently observed in patients diagnosed with pneumonia, with GGO and consolidation being the most prevalent findings[6–9]. As evidenced by Park et al., the diagnostic performance for patients can be enhanced by taking various characteristics of lung abnormalities, such as GGO and consolidation, into consideration[10]. In effect, COVID-19 severity assessment algorithms primarily focus on pulmonary involvement when utilizing computed tomography (CT) scans, as the severity of the disease in patients with COVID-19 can

[1]School of Electrical and Electronic Engineering, Yonsei University, Seoul, Republic of Korea. [2]Vuno Inc., Seoul, Republic of Korea. [3]Medical AI Research Center, Samsung Medical Center, Seoul 06351, Republic of Korea. [4]Department of Radiology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Republic of Korea. [5]Department of Artificial Intelligence, Yonsei University, Seoul, Republic of Korea. [6]Center for Healthcare Robotics, Korea Institute of Science and Technology, 5, Hwarang-ro 14-gil, Seongbuk-gu, Seoul 02792, Republic of Korea. [7]Department of Oral and Maxillofacial Radiology, Yonsei University College of Dentistry, Seoul, Republic of Korea. [8]Department of Radiology and Center for Clinical Imaging Data Science (CCIDS), Yonsei University College of Medicine, Seoul, Republic of Korea. ✉email: dosik.hwang@yonsei.ac.kr

be determined by analyzing the extent of lung involvement[11]. Furthermore, numerous studies have reported a correlation between the quantitative measurement of lung involvement in CT scans and laboratory findings as well as clinical parameters, often using machine learning (ML) techniques or a stratified scoring system for the analysis[12–14]. For instance, Lessmann et al. reported that the severity of COVID-19 patients can be determined from CT images by calculating the percentage of affected lung tissue per lobe[15]. Wenli et al. demonstrated that texture features for lesion volume and non-lesion lung volume are instrumental in determining the severity of COVID-19[16]. In this study, we limit our focus to the most common patterns of lesions, GGO and consolidation, as previously reported in relevant literature.

To date, several studies have utilized deep learning (DL) for the severity assessment of COVID-19 patients. Zhang et al. demonstrated the application of two imaging biomarkers derived from lung field and lung abnormality segmentation models for the severity assessment of COVID-19 patients[17]. Similarly, Goncharov et al. leveraged DL to generate a segmentation mask and calculate the affected lung percentage for severity assessment purposes[18]. Chieregato et al. presented a method that combined laboratory and clinical data with imaging features[19]. They extracted imaging features from CT scans using a DL model and integrated them into a CatBoost ML model, along with tabular data. Gao et al. proposed a dual-branch combination network that leverages lesion segmentation information for DL model training, focusing on the lesion area while simultaneously performing lesion segmentation and COVID-19 prediction[20]. As such, most studies have designed a severity assessment model utilizing either lung-masked or lesion-masked CT images[21]. To our knowledge, no study has yet developed a severity assessment model using both types of masked CT images and combined the DL model with quantitative features obtained from the lung area.

Consequently, this study introduces a hybrid approach that combines two DL models and one ML model, trained using quantitative features. Notably, the proposed model underwent rigorous external validation that included not only COVID-19 patients but also other types of IRDs including influenza, novel influenza, and BP.

Methods

Patient population. In this study, all research was performed in accordance with relevant guidelines/regulations and all experimental protocols were approved by the Institutional Review Board of Samsung Medical Center, Pusan National University Hospital, Chonnam National University Hospital, Keimyung University Dongsan Medical Center, Chungnam National University Hospital, Gachon University Gil Medical Center, Kyungpook National University Hospital, and Chungnam National University Sejong Hospital, which also waived written informed consent for this study. This study included data from 1243 patients diagnosed with COVID-19, admitted to a referral hospital. Data from 438 patients who did not undergo CT imaging were excluded from the analysis. The remaining 805 patients were randomly divided into two groups, allocating 80% of the participants for training and 20% for internal validation (referred to as D1). Three additional external data sets were collected from 7 different external cohorts. The external validation sets consisted of 1138 patients diagnosed with COVID-19 (referred to as D2), 233 patients with influenza and novel influenza (referred to as D3), and 268 patients with BP (referred to as D4), respectively. In this study, each patient was divided into severe and non-severe cases based on admission to the intensive care unit, mortality, and whether they received at least one of four specific treatments: steroid injection, oxygen supply, mechanical ventilation, or extracorporeal membrane oxygenation. A data flow diagram and the clinical characteristics of the patients are detailed in Fig. 1 and Table 1, respectively. Representative CT images from both non-severe and severe COVID-19 cases are provided in Fig. 2.

CT acquisition parameters. The CT acquisition parameters for the dataset from one referral hospital were as follows: 34–782 mAs, 100–150 kVp, 0.625–5.00 mm slice thickness, and 0.35–0.97 mm pixel size. For the external validation sets, the CT acquisition parameters were unavailable due to data anonymization.
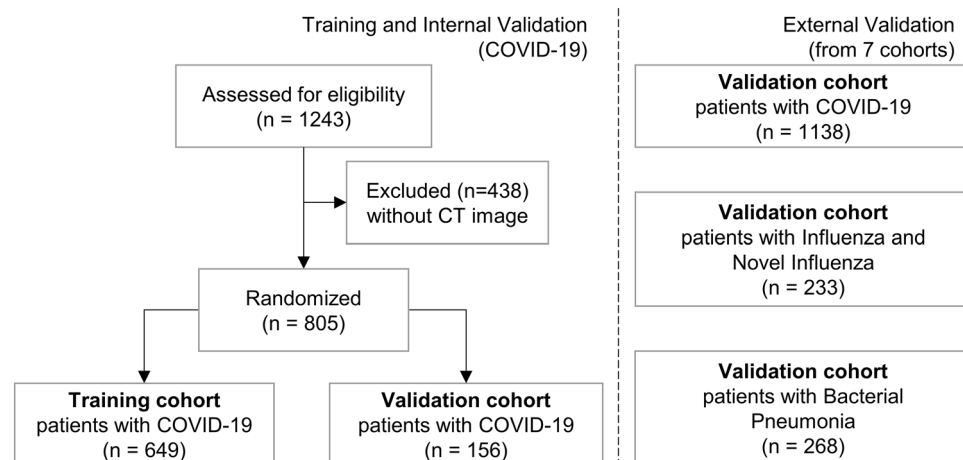

Figure 1. Data flow diagram of patients. COVID-19: coronavirus disease 2019. CT: computed tomography.

Characteristics	No. (%) for training and internal validation			No. (%) for external validation						
	Training Set (n = 649)	Validation Set (n = 156) (D1)	p value*	Patients with COVID-19 (n = 1138) (D2)	p value*	Patients with influenza and novel influenza (n = 233) (D3)	p value*	Patients with bacterial pneumonia (n = 268) (D4)	p value*
Mean age (year)†	50 ± 17	52 ± 16	0.183	63 ± 16	< 0.001	67 ± 19	< 0.001	69 ± 18	< 0.001
Sex			0.976		0.395		0.015		< 0.001
Men	326 (50%)	78 (50%)		588 (52%)		93 (40%)		100 (37%)	
Women	333 (50%)	78 (50%)		550 (48%)		140 (60%)		168 (63%)	
Treatment									
Steroid injection	59 (9%)	11 (7%)	0.417	397 (35%)	< 0.001	19 (8%)	0.666	48 (18%)	< 0.001
Oxygen supply	21 (3%)	7 (4%)	0.444	288 (25%)	< 0.001	57 (24%)	< 0.001	110 (41%)	< 0.001
Mechanical ventilation	10 (2%)	2 (1%)	0.811	71 (6%)	< 0.001	8 (8%)	0.080	32 (12%)	< 0.001
ECMO	43 (7%)	7 (4%)	0.321	16 (1%)	< 0.001	1 (0%)	< 0.001	0 (0%)	< 0.001
Outcome									
ICU admission	140 (22%)	33 (21%)	0.909	62 (5%)	< 0.001	31 (13%)	0.006	44 (16%)	0.077
Mortality	143 (22%)	38 (24%)	0.533	62 (5%)	< 0.001	10 (4%)	< 0.001	29 (11%)	< 0.001
Severity			0.651		< 0.001		0.448		< 0.001
Non-severe	426 (66%)	106 (68%)		626 (55%)		160 (69%)		136 (51%)	
Severe	223 (34%)	50 (32%)		512 (45%)		73 (31%)		132 (49%)	

Table 1. Demographic and pathological characteristics. COVID-19, coronavirus disease 2019; ECMO, Extracorporeal membrane oxygenation; ICU, intensive care unit. *p-values were calculated by comparing with the training set using the t-test or the chi-squared test. †Mean ± standard deviation.

a Non-severe cases with COVID-19
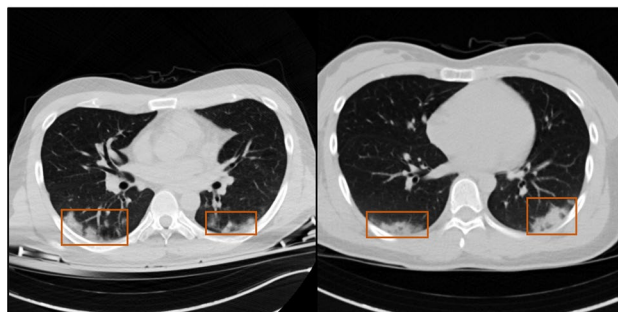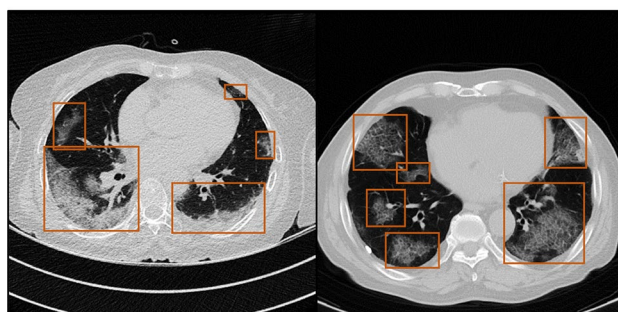

b Severe cases with COVID-19

Figure 2. Example CT images with COVID-19. (**a**) Non-severe cases. (**b**) Severe cases. COVID-19, coronavirus disease 2019.

DL model—segmentation.　We employed the lung segmentation DL model from a previous study[22] based on 2D U-Net[23] (**UNet_lung**) and the VUNO Med-LungQuant (Seoul, Republic of Korea) version 1.0.0.4. for lesion segmentation based on 2D U-Net (**UNet_lesion**). The input for the **UNet_lung** was axial CT images resized to $256 \times 256 \times 160$, and the output was segmentation masks of the left and right lung, respectively, with a size of $256 \times 256 \times 160$. The input for the **UNet_lesion** was lung-masked CT images, and the output was segmentation masks of GGO and consolidation, respectively. The size of the input and output for the **UNet_lesion** was $256 \times 256 \times 160$.

DL model—classification. Two DL models were constructed based on 3D ResNet-50[24] for the severity assessment. More details for ResNet-50 are available in Supplementary Fig. S1 online. 1) One DL model was trained with lung-masked CT images (DL_{lung}). 2) The other DL model was trained with lesion-masked CT images (DL_{lesion}). The overall architecture of the proposed DL model is presented in Fig. 3. input for DL_{lung} was the lung-masked CT image of size $192 \times 192 \times 160$ obtained by cropping, masking, and resizing the segmented lung region, which was obtained by the elementwise multiplication of the CT image and the output of $UNet_{lung}$. The output of DL_{lung} was an estimate of the severity. The input for DL_{lesion} was the lesion-masked CT image of size $192 \times 96 \times 160$ obtained by cropping, masking, and resizing the segmented lesion region, which was obtained by the elementwise multiplication of the CT image and the output of $UNet_{lesion}$. The output of DL_{lesion} was an estimate of the severity. Based on the same DL architecture, we constructed another DL model trained with lung-cropped CT images (baseline model) to analyze the effectiveness of lung or lesion masking. The input for the baseline model was the lung-cropped CT image of size $192 \times 192 \times 160$, obtained by cropping and resizing the segmented lung region from $UNet_{lung}$. The output of the baseline model was an estimate of the severity. The DL models were implemented using the Python Pytorch library. They were trained using the Adam optimizer[25] with a weight decay of 0.0005 and a learning rate of 0.0001 with a cosine annealing schedular[26] for 200 epochs. The DL models were trained using an Inter Xeon E5-2630 central processing unit and an NVIDIA GeForce GTX TITAN X graphics processing unit with CUDA version 11.4. Augmentation with a random axial rotation of −15 to 15 degrees was performed during training. The training set was separated into 80% training and 20% validation (not the same as D1) sets.

ML model—classification. Anatomically, there are 18–20 bronchopulmonary segments. Abnormal findings such as GGO and consolidation in each segment have different clinical significance[27,28]. However, segmenting individual bronchopulmonary segments is more challenging than segmenting bilateral lungs. Therefore, we used $UNet_{lung}$ to segment only bilateral lungs and divided them into 4 segments each, which are simplified bronchopulmonary segments. As a result, there are 11 lung areas: right upper, right upper-middle, right lower-middle, right lower, whole right, left upper, left upper-middle, left lower-middle, left lower, whole left, and whole lung. Furthermore, there are three types of lesions: GGO, consolidation, and both GGO and consolidation. In conclusion, a total of 33 features (number of lung areas × number of lesion types) are obtained by calculating the proportions of lesions in each lung area, as summarized in Supplementary Table S1. In this study, we defined this feature set as quantitative Lung Involvement Features (LIFe). Using the LIFe, we constructed a random forest (RF)[29] model. In this study, we defined this RF model as ML_{LIFe} because the RF is one of the ML algorithms.

Combination of DL and ML models. In this study, we proposed a hybrid model called Hybrid-DDM by the combination of two DL models (DL_{lung} and DL_{lesion}) and one ML model (ML_{LIFe}). The Hybrid-DDM was obtained by uniform averaging the estimates of the severity from the three models.
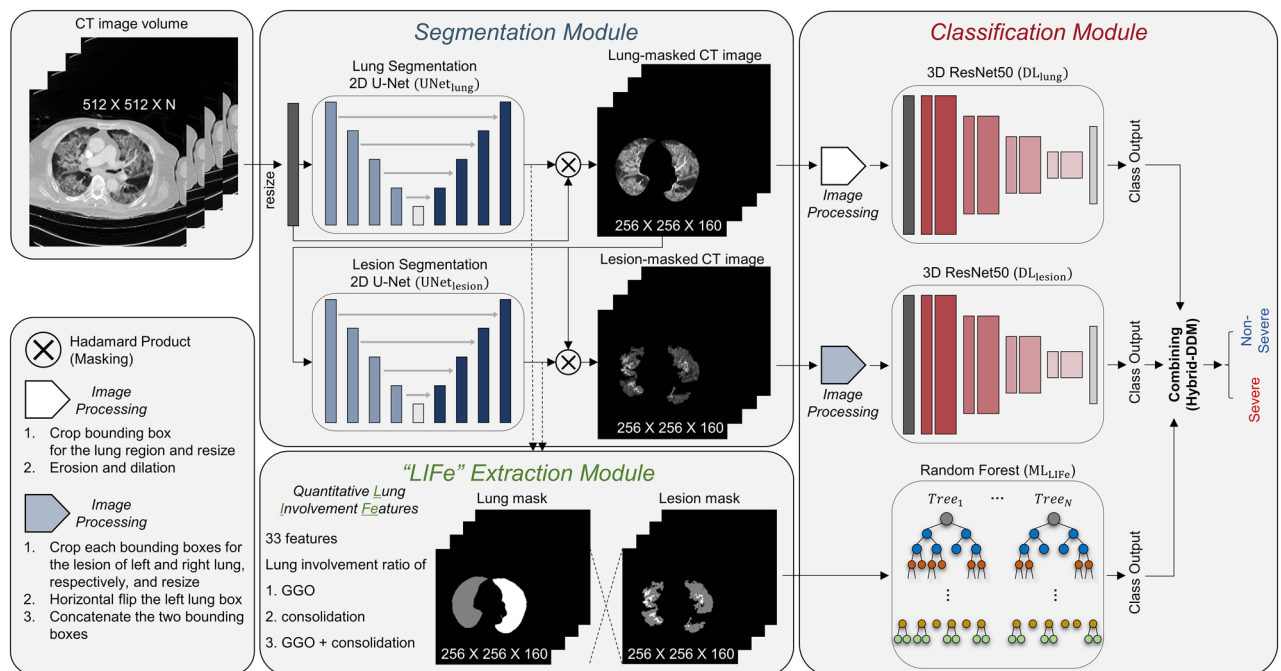

Figure 3. The overall flow of our proposed Hybrid-DDM model. CT, computed tomography. GGO, ground-glass opacity.

Statistical analysis. We used the following software: RF model construction (MATLAB. (2019). 9.7.0.1190202 (R2019a). Natick, Massachusetts: The MathWorks Inc.), statistical analysis (R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.). The severity assessment performance was evaluated with the area under the receiver operating characteristic (ROC) curve (AUC), and Delong's method[30] was used to compare two AUC values. A *p*-value lower than 0.05 was considered statistically significant.

Results

The AUC and ROC curves of the Hybrid-DDM and the baseline models in each validation set are shown in Fig. 4a,b. More detailed results were summarized in Table 2.

The baseline model showed the lowest performances in D1, D2, and D3 with AUCs (95% confidence interval [CI]) of 0.767 (0.688–0.846), 0.753 (0.725–0.782), and 0.668 (0.597–0.740), respectively. The combination of the DL_{lung} and the DL_{lesion} ($DL_{lung-lesion}$) showed the improved performance than each DL model with AUCs (95% CI) of 0.812 (0.735–0.888) versus 0.790 (0.711–0.869) and 0.807 (0.728–0.887) in D1, 0.794 (0.768–0.820) versus 0.756 (0.727–0.785) and 0.788 (0.761–0.814) in D2, and 0.765 (0.696–0.834) versus 0.755 (0.686–0.824) and 0.732 (0.657–0.808) in D3, respectively.

The ML_{LIFe} showed comparable performance to the $DL_{lung-lesion}$ with AUCs of 0.811 (0.737–0.885), 0.777 (0.750–0.804), and 0.754 (0.683–0.825) in D1, D2, and D3, respectively. Figure 5 shows the feature importance of 33 features used in the ML_{LIFe}. Feature importance in the RF model measures the contribution of each feature to the model's predictions. It is quantified by the degree to which the model's performance decreases when the values of a particular feature are randomly permuted, thereby disrupting the relationship between the feature and the label. Consequently, a feature is considered highly important if its random permutation leads to a significant decline in the model's performance, indicating that the model relies heavily on this feature for accurate predictions[29]. As a result, GGO-related features tended to have higher importance than consolidation-related features. Box plots of 33 features for severe and non-severe patients are shown in Fig. 6. The features are listed in descending order of feature importance from left to right. In all datasets, severe patients tend to have larger feature values (the proportion of lesions) than non-severe patients. There is also a tendency for feature values to be larger in the lower lobes (L3, L4, R3, and R4) than in the upper lobes (L1, L2, R1, and R2).

Lastly, the Hybrid-DDM showed the highest performance in D1, D2, and D3 with AUCs (95% CI, *p*-value compared with baseline model) of 0.830 (0.758–0.898, *p* = 0.036), 0.801 (0.776–0.827, *p* < 0.001.), and 0.774 (0.705–0.842, *p* < 0.001.). However, in the D4, which is a set of BP patients, there was no significant difference in
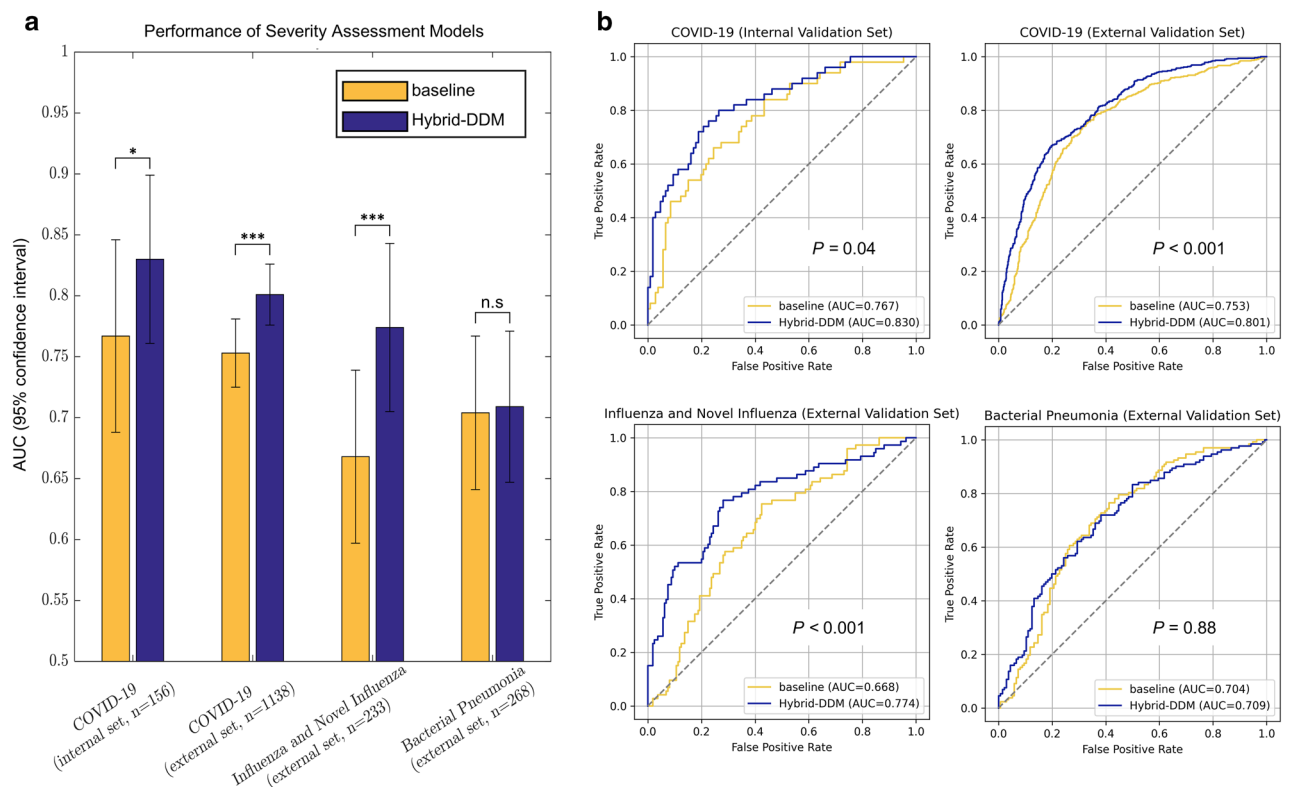

Figure 4. Performances of the severity assessment models. (**a**) Bar graph for model performance in each data set. (**b**) Receiver operating characteristic (ROC) curves of the severity assessment models. *: *p* < 0.05, **: *p* < 0.01, ***: *p* < 0.001. DL, deep learning. COVID-19, coronavirus disease 2019. AUC, area under the receiver operating characteristic curve.

Model	Internal validation				External validation											
	Patients with COVID-19 (n = 156) (D1)				Patients with COVID-19 (n = 1138) (D2)				Patients with influenza and novel influenza (n = 233) (D3)				Bacterial pneumonia (n = 268) (D4)			
	AUC (±95% CI)	p value*	SEN	SPE	AUC (±95% CI)	p value*	SEN	SPE	AUC (±95% CI)	p value*	SEN	SPE	AUC (±95% CI)	p value*	SEN	SPE
Baseline	0.767 (0.688–0.846)	Reference	0.660	0.755	0.753 (0.725–0.782)	Reference	0.740	0.687	0.668 (0.597–0.740)	Reference	0.753	0.575	0.704 (0.641–0.767)	Reference	0.765	0.588
DL$_{lung}$	0.790 (0.711–0.869)	0.497	0.840	0.651	0.756 (0.727–0.785)	0.846	0.701	0.732	0.755 (0.686–0.824)	0.006	0.767	0.681	0.678 (0.614–0.742)	0.424	0.667	0.662
DL$_{lesion}$	0.807 (0.728–0.887)	0.282	0.620	0.925	0.788 (0.761–0.814)	0.008	0.672	0.776	0.732 (0.657–0.808)	0.076	0.644	0.750	0.682 (0.618–0.745)	0.498	0.735	0.574
DL$_{lung-lesion}$	0.812 (0.735–0.888)	0.183	0.740	0.793	0.794 (0.768–0.820)	< 0.001	0.721	0.733	0.765 (0.696–0.834)	0.001	0.808	0.638	0.710 (0.648–0.772)	0.865	0.599	0.713
ML$_{LIFe}$	0.811 (0.737–0.885)	0.244	0.640	0.868	0.777 (0.750–0.804)	0.066	0.734	0.701	0.754 (0.683–0.825)	0.017	0.575	0.850	0.698 (0.636–0.761)	0.844	0.727	0.640
Hybrid-DDM	0.830 (0.758–0.898)	0.036	0.800	0.736	0.801 (0.776–0.827)	< 0.001	0.668	0.804	0.774 (0.705–0.842)	< 0.001	0.767	0.719	0.709 (0.647–0.771)	0.877	0.720	0.618

Table 2. Performances of the severity assessment model for patients with COVID-19 and other three types of pneumonia. Cut-off points for sensitivity and specificity were determined by the validation data of the training set (not the same as D1). COVID-19, coronavirus disease 2019; AUC, area under the receiver operating curve; CI, confidence interval; SEN, sensitivity; SPE, specificity. *p-values were calculated by Delong's method.
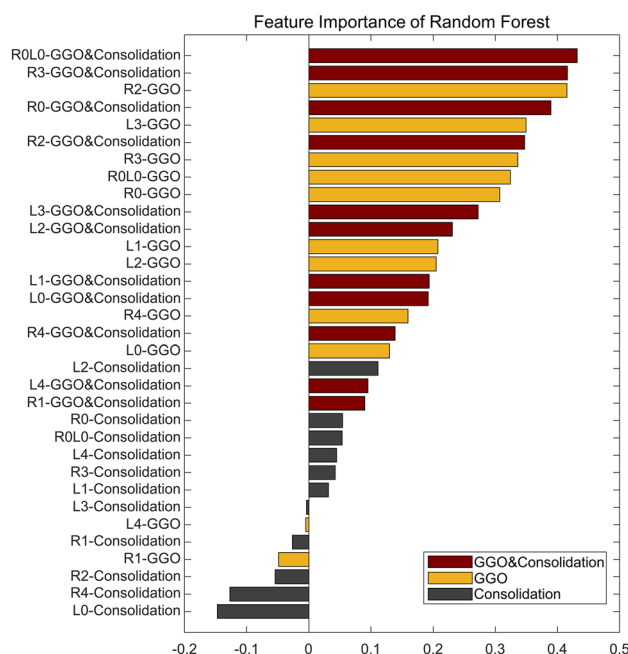

Figure 5. Feature importance of the random forest model (**ML$_{LIFe}$**). A total of 33 features were used for model training.

AUCs (95% CI) between the baseline model and the Hybrid-DDM (0.704 [0.641–0.767] vs 0.709 [0.647–0.771], $p = 0.877$).

Discussion

This objective of this study was to develop an automated severity assessment model for patients with COVID-19 and other IRDs using CT images. To this end, we developed a Hybrid-DDM model that combined two DL models and one ML model. Our investigation yielded three crucial insights. Firstly, training a model with lung-masked
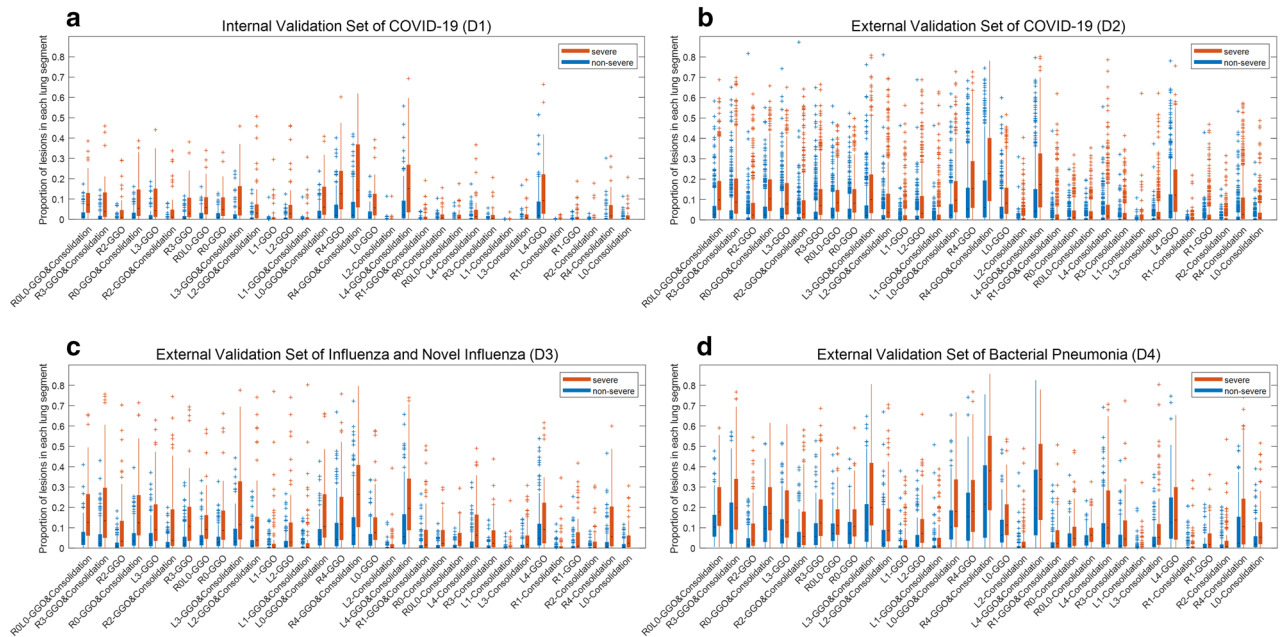
Figure 6. Visualization of the LIFe for severe and non-severe patients as a box plot. The LIFe is listed in descending order of feature importance from left to right. (**a**) Internal validation set of patients with COVID-19 (n = 156). (**b**) External validation set of patients with COVID-19 (n = 1138). (**c**) External validation set of patients with influenza and novel influenza (n = 233). (**d**) External validation set of patients with bacterial pneumonia (n = 268).

or lesion-masked CT images enhanced the efficiency of severity assessment of patients with COVID-19, as compared to training solely with lung-cropped CT images. Secondly, the integration of two DL models with an ML model improved the performance of the severity assessment model. Thirdly, while the Hybrid-DDM model demonstrated significant effectiveness for patients with VP, it was not similarly effective for patients with BP.

The expedient severity assessment of patients with COVID-19 constituted a vital component of patient care and mandated immediate attention. Utilizing CT imaging for the diagnosis of disease severity in patients with IRDs provides clinicians with critical insights into the progression of the disease and potential responses to treatment, thereby enabling timely and tailored therapeutic interventions to enhance patient outcomes. Although the COVID-19 pandemic is currently in decline and transitioning into an endemic phase, it remains essential to prepare for future pandemics given their unpredictable nature. Consequently, it is crucial to establish reliable severity assessment systems for other IRDs. Numerous studies have risen to this challenge, exploring an array of techniques for severity assessment of patients with COVID-19 and other IRDs[31]. Specifically, many studies have adopted DL techniques that utilize lung or lesion information as guiding features during model training[20,32,33]. Our study aimed to assess the merits of employing differently pre-processed CT images for training DL models. Leveraging lung-masked CT images can be beneficial as it incorporates perilesional areas, which may encompass opacity, ambiguous lesions, and bronchus regions that are not discernible in lesion-masked CT images. Models trained on such comprehensive characteristics could estimate the severity of IRDs based on overall texture-based features instead of solely on lesion volume. On the contrary, the use of lesion-masked CT images allows the model to concentrate on lesion volume, texture, and shape, thus facilitating the determination of disease severity based on the presence and extent of lesions within the lung.

The combination of DL_{lung} and DL_{lesion} demonstrated superior AUC compared to individual models for both the internal and external validation sets of patients with COVID-19 (0.812 vs 0.790 and 0.807 in D1, and 0.794 vs 0.756 and 0.788 in D2, respectively). This improvement was also evident in the external validation set of patients with influenza and novel influenza (0.765 vs 0.755 and 0.732 in D3), substantiating the effectiveness of combining DL_{lung} and DL_{lesion} in severity assessment of patients with VP. Furthermore, the baseline model trained without any disease-related guidance was found to be inferior to both DL_{lung} and DL_{lesion}.

In clinical practice, the severity of COVID-19 and other IRDs can often be determined by evaluating the extent of lung involvement[15,16]. In order to quantitatively assess disease severity, we formulated a set of quantitative features, termed LIFe, which we subsequently utilized to construct a ML model. LIFe calculates the proportion of distinct lesion types, including GGO and consolidations. An added benefit of LIFe is that it eliminates the need for a lobe-specific segmentation mask, as it can be derived directly from the comprehensive lung segmentation mask. Our investigation showed that an ML model trained with LIFe features showed comparable performance to those of DL models. It is noteworthy that most lesions in patients with COVID-19 are typically located in the lower lobe of the lung[34,35]. Consistently, our study also found that LIFe values for patients with IRDs were higher in the lower lobe than in the upper lobe. In particular, the difference in GGO-related LIFe values between severe and non-severe patients was more marked than the difference in consolidation-related values. This insight guided us towards the development of a severity assessment model premised on measuring

the proportion of lesions in bronchopulmonary segments, diverging from DL models that rely on nonlinear features extracted from the entire lung or specific lesions. However, a reliance solely on LIFe values might overlook other radiological features present in CT images. By incorporating DL models into the ML model, we were able to improve our severity assessment performance, as measured by the AUC, from 0.811 to 0.830 for D1, from 0.777 to 0.801 for D2, and from 0.754 to 0.774 for D3, respectively.

However, our approach did not prove effective for patients with BP. Even when incorporating prior information like lung masks, lesion masks, or LIFe values, the model's performance did not differ from DL models using lung-cropped CT images. This can be attributed to our training data being derived from patients with COVID-19, where one severe complication is VP[36]. Clinically, VP and BP exhibit distinct characteristics such as symptoms, disease severity, and radiological findings, which have been the subject of numerous studies[37,38]. In fact, we observed that the difference in LIFe values between severe and non-severe patients was smaller in BP patients than in VP patients. Thus, applying a severity assessment model trained on VP patients to BP patients would be unsuitable. Although our investigation featured a multi-institutional and multi-disease validation approach, the model fell short in enhancing performance for patients with BP. This underscores the imperative for cultivating more sophisticated and universally applicable methodologies that could be beneficial to patients with BP as well.

Our study had some limitations. Firstly, while our study includes data from multiple diseases, further validation should be performed on a variety of infectious respiratory diseases to ensure its applicability. Secondly, our research was conducted retrospectively, and prospective validation is required to strengthen its credibility. Thirdly, both internal and external datasets consist solely of patients from South Korea, so additional research is needed to extend its implications to other nations. Lastly, our model's practical utility could be improved if it were able to distinguish between VP and BP using CT images, laboratory findings, or clinicopathological information.

Despite these limitations, our findings have shown that training a DL model on CT images that were masked with either lung or lesion information, and subsequently integrating this with a ML model, significantly enhanced the performance of the severity assessment model for patients with VP. In the future, incorporating additional clinicopathological information such as age, gender, smoking history, or symptoms like cough, fever, and chills, could further improve the model's performance for the severity assessment of patients with IRDs.

Data availability
The datasets generated and/or analysed during the current study are not publicly available due to privacy issues but are available from the corresponding author on reasonable request.

References
1. Clark, C. E., McDonagh, S. T. J., McManus, R. J. & Martin, U. COVID-19 and hypertension: Risks and management. A scientific statement on behalf of the British and Irish Hypertension Society. *J. Hum. Hypertens.* **35**, 304–307. https://doi.org/10.1038/s41371-020-00451-x (2021).
2. Lim, S., Bae, J. H., Kwon, H. S. & Nauck, M. A. COVID-19 and diabetes mellitus: From pathophysiology to clinical management. *Nat. Rev. Endocrinol.* **17**, 11–30. https://doi.org/10.1038/s41574-020-00435-4 (2021).
3. Choi, G. J., Kim, H. M. & Kang, H. The potential role of dyslipidemia in COVID-19 severity: An umbrella review of systematic reviews. *J. Lipid Atheroscler.* **9**, 435. https://doi.org/10.12997/jla.2020.9.3.435 (2020).
4. Javanmardi, F., Keshavarzi, A., Akbari, A. & Emami, A. Prevalence of underlying diseases in died cases of COVID-19: A systematic review and meta-analysis. *PloS One* **15**, e0241265. https://doi.org/10.1371/journal.pone.0241265 (2020).
5. Kompaniyets, L. *et al.* Peer reviewed: Underlying medical conditions and severe illness among 540,667 adults hospitalized with COVID-19, March 2020–March 2021. *Prev. Chronic Dis.* https://doi.org/10.5888/pcd18.210123 (2021).
6. Hani, C. *et al.* COVID-19 pneumonia: A review of typical CT findings and differential diagnosis. *Diagn. Interv. Imaging* **101**, 263–268. https://doi.org/10.1016/j.diii.2020.03.014 (2020).
7. Garg, M., Prabhakar, N., Gulati, A., Agarwal, R. & Dhooria, S. Spectrum of imaging findings in pulmonary infections. Part 1: Bacterial and viral. *Pol. J. Radiol.* **84**, 205–213. https://doi.org/10.5114/pjr.2019.85812 (2019).
8. Koo, H. J. *et al.* Radiographic and CT features of viral pneumonia. *Radiographics* **38**, 719–739. https://doi.org/10.1148/rg.2018170048 (2018).
9. Vilar, J., Domingo, M. L., Soto, C. & Cogollos, J. Radiology of bacterial pneumonia. *Eur. J. Radiol.* **51**, 102–113. https://doi.org/10.1016/j.ejrad.2004.03.010 (2004).
10. Park, D. *et al.* Importance of CT image normalization in radiomics analysis: Prediction of 3-year recurrence-free survival in non-small cell lung cancer. *Eur. Radiol.* **32**, 1–10. https://doi.org/10.1007/s00330-022-08869-2 (2022).
11. Burian, E. *et al.* Intensive care risk estimation in COVID-19 pneumonia based on clinical and imaging parameters: Experiences from the Munich cohort. *J. Clin. Med.* **9**, 1514. https://doi.org/10.3390/jcm9051514 (2020).
12. Majrashi, N. A. *et al.* A systematic review of the relationship between chest CT severity score and laboratory findings and clinical parameters in COVID-19 pneumonia. *Diagnostics* **13**, 2223. https://doi.org/10.3390/diagnostics13132223 (2023).
13. Oi, Y. *et al.* Prediction of prognosis in patients with severe COVID-19 pneumonia using CT score by emergency physicians: A single-center retrospective study. *Sci. Rep.* **13**, 4045. https://doi.org/10.1038/s41598-023-31312-5 (2023).
14. Albataineh, Z., Aldrweesh, F. & Alzubaidi, M. A. COVID-19 CT-images diagnosis and severity assessment using machine learning algorithm. *Clust. Comput.* https://doi.org/10.1007/s10586-023-03972-5 (2023).
15. Lessmann, N. *et al.* Automated assessment of COVID-19 reporting and data system and chest CT severity scores in patients suspected of having COVID-19 using artificial intelligence. *Radiology* **298**, 298–239. https://doi.org/10.1148/radiol.2020202439 (2021).
16. Cai, D. W. *et al.* CT quantification and machine-learning models for assessment of disease severity and prognosis of COVID-19 patients. *Acad. Radiol.* **27**, 1665–1678. https://doi.org/10.1016/j.acra.2020.09.004 (2020).
17. Li, Z. *et al.* From community-acquired pneumonia to COVID-19: A deep learning–based method for quantitative analysis of COVID-19 on thick-section CT scans. *Eur. Radiol.* **30**, 6828–6837. https://doi.org/10.1007/s00330-020-07042-x (2020).
18. Goncharov, M. *et al.* CT-based COVID-19 triage: Deep multitask learning improves joint identification and severity quantification. *Med. Image Anal.* **71**, 102054. https://doi.org/10.1016/j.media.2021.102054 (2021).

19. Chieregato, M. *et al.* A hybrid machine learning/deep learning COVID-19 severity predictive model from CT images and clinical data. *Sci. Rep.* **12**, 1–15. https://doi.org/10.1038/s41598-022-07890-1 (2022).
20. Gao, K. *et al.* Dual-branch combination network (DCN): Towards accurate diagnosis and lesion segmentation of COVID-19 using CT images. *Med. Image Anal.* **67**, 101836. https://doi.org/10.1016/j.media.2020.101836 (2021).
21. Liu, Y. *et al.* Self-paced multi-view learning for CT-based severity assessment of COVID-19. *Biomed. Signal. Process. Control.* **83**, 104672. https://doi.org/10.1016/j.bspc.2023.104672 (2023).
22. Kim, G. B. *et al.* Comparison of shallow and deep learning methods on classifying the regional pattern of diffuse lung disease. *J. Digit. Imaging* **31**, 415–424. https://doi.org/10.1007/s10278-017-0028-9 (2018).
23. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional networks for biomedical image segmentation. *Med. Image Comput. Comput. Assist. Interv.* https://doi.org/10.1007/978-3-319-24574-4_28 (2015).
24. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* https://doi.org/10.1109/CVPR.2016.90 (2016).
25. Kingma D.P., Ba J. Adam: A method for stochastic optimization. arXiv preprint. 1412.6980. https://doi.org/10.48550/arXiv.1412.6980 (2014).
26. Loshchilov, I., Hutter, F. Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint. 1608.03983. https://doi.org/10.48550/arXiv.1608.03983 (2016).
27. Shan, F. *et al.* Abnormal lung quantification in chest CT images of COVID-19 patients with deep learning and its application to severity prediction. *Med. Phys.* **48**, 1633–1645. https://doi.org/10.1002/mp.14609 (2021).
28. Shan, F. *et al.* Lung infection quantification of COVID-19 in CT images with deep learning. arXiv preprint. 2003.04655. https://doi.org/10.48550/arXiv.2003.04655 (2020).
29. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32. https://doi.org/10.1023/A:1010933404324 (2001).
30. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* **44**, 837–845. https://doi.org/10.2307/2531595 (1988).
31. Mondal, M. R. H., Bharati, S. & Podder, P. Diagnosis of COVID-19 using machine learning and deep learning: A review. *Curr. Med. Imaging* **17**, 1403–1418. https://doi.org/10.2174/1573405617666210713113439 (2021).
32. Wang, L., Lin, Z. Q. & Wong, A. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Sci. Rep.* **10**, 1–12. https://doi.org/10.1038/s41598-020-76550-z (2020).
33. Wu, X. *et al.* Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: A multicentre study. *Eur. J. Radiol.* **128**, 109041. https://doi.org/10.1016/j.ejrad.2020.109041 (2020).
34. Song, F. *et al.* Emerging coronavirus 2019-nCoV pneumonia. *Radiology* **295**, 200274. https://doi.org/10.1148/radiol.2020200274 (2020).
35. Bernheim, A. *et al.* Chest CT findings in Coronavirus Disease-19 (COVID19): Relationship to duration of infection. *Radiology* **295**, 200463. https://doi.org/10.1148/radiol.2020200463 (2020).
36. Nassef, M., Shorrab, A. A., Buheji, M., Buheji, A. R. & Abosamak, M. F. Alleviation of antimicrobial therapy in ICU during COVID-19 second wave—A review paper. *Int. J. Manag.* **11**, 2163–2175. https://doi.org/10.34218/IJM.11.12.2020.204 (2020).
37. Virkki, R. *et al.* Differentiation of bacterial and viral pneumonia in children. *Thorax* **57**, 438–441. https://doi.org/10.1136/thorax.57.5.438 (2002).
38. Thomas, J., Pociute, A., Kevalas, R., Malinauskas, M. & Jankauskaite, L. Blood biomarkers differentiating viral versus bacterial pneumonia aetiology: A literature review. *Ital. J. Pediatr.* **46**, 1–10. https://doi.org/10.1186/s13052-020-0770-3 (2020).

Acknowledgements

Author contributions

Conceptualization: R.J., H.J.A., S.B.; Data Curation: M.J.C.; Formal Analysis: D.P.; Funding Acquisition: D.H.; Investigation: D.P., R.J., H.J.A., D.H.; Methodology: D.P., R.J., H.J.A.; Project Administration: D.H.; Resources: M.J.C., S.B.; Software: D.P., R.J., H.J.A., E.C.; Supervision: S.B., D.H.; Validation: D.P., H.J.A.; Visualization: D.P.; Writing-original draft: D.P.; Writing-review & Editing: D.P., R.J., H.J.A., E.C., D.H.

Competing interests

D Park, E Choi, and D Hwang received research grants from VUNO Inc. R Jang, HJ An, and S Bak are employees of VUNO Inc. MJ Chung declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-40506-w.

Correspondence and requests for materials should be addressed to D.H.

Reprints and permissions information is available at www.nature.com/reprints.

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