RESEARCH

BMC Pulmonary Medicine

Open Access



Unsupervised machine learning clustering approach for hospitalized COVID-19 pneumonia patients

Nuttinan Nalinthasnai¹, Ratchainant Thammasudjarit², Tanapat Tassaneyasin^{1,3}, Dararat Eksombatchai¹, Somnuek Sungkanuparph³, Viboon Boonsarngsuk¹, Yuda Sutherasan¹, Detajin Junhasavasdikul¹, Pongdhep Theerawit⁴ and Tananchai Petnak^{1*}

Abstract

Background Identification of distinct clinical phenotypes of diseases can guide personalized treatment. This study aimed to classify hospitalized COVID-19 pneumonia subgroups using an unsupervised machine learning approach.

Methods We included hospitalized COVID-19 pneumonia patients from July to September 2021. K-means clustering, an unsupervised machine learning method, was performed to identify clinical phenotypes based on clinical and laboratory variables collected within 24 hours of admission. Variables were normalized before clustering to ensure equal contribution to the analysis. The optimal number of clusters was determined using the elbow method and Silhouette scores. Cox proportional hazard models were used to compare the risk of intubation and 90-day mortality across the identified clusters.

Results Three clinically distinct clusters were identified among 538 hospitalized COVID-19 pneumonia patients. Cluster 1 (N= 27) consisted predominantly of males and showed significantly elevated serum liver enzymes and LDH levels. Cluster 2 (N= 370) was characterized by lower chest x-ray scores and higher serum albumin levels. Cluster 3 (N= 141) was characterized by older age, diabetes mellitus, higher chest x-ray scores, more severe vital signs, higher creatinine levels, lower hemoglobin levels, lower lymphocyte counts, higher C-reactive protein, higher D-dimer, and higher LDH levels. When compared to cluster 2, cluster 3 was significantly associated with increased risk of 90-day mortality (HR, 6.24; 95% Cl, 2.42–16.09) and intubation (HR, 5.26; 95% Cl 2.37–11.72). In contrast, cluster 1 had a 100% survival rate with a non-significant increase in intubation risk compared to cluster 2 (HR, 1.40, 95% Cl, 0.18–11.04).

Conclusions We identified three distinct clinical phenotypes of COVID-19 pneumonia patients, with cluster 3 associated with an increased risk of respiratory failure and mortality. These findings may guide tailored clinical management strategies.

Keywords COVID-19, Pneumonia, Clustering analysis, Machine learning, Mortality

*Correspondence: Tananchai Petnak tananchai.pet@gmail.com ¹Division of Pulmonary and Pulmonary Critical Care Medicine, Department of Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Rama VI Road, Ratchathewi, Bangkok 10400, Thailand



²Department of Computer Science, Faculty of Science, Srinakharinwirot University, Bangkok, Thailand

³Faculty of Medicine Ramathibodi Hospital, Chakri Naruebodindra Medical Institute, Mahidol University, Samutprakan, Thailand ⁴Division of Critical Care Medicine, Department of Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit ine to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creative.commons.org/licenses/by-nc-nd/4.0/.

Background

The coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), was first reported in Wuhan, China, in December 2019 [1]. Since then, COVID-19 has rapidly spread globally. The severity of COVID-19 significantly varies among individuals, with clinical presentations ranging from mild symptoms to respiratory failure and death [1]. This variability indicates the existence of different clinical phenotypes among individuals, which are potentially associated with different treatment responses and outcomes. Several factors have been demonstrated to be associated with increased risks of mortality and respiratory failure in COVID-19 patients, including advanced age, diabetes mellitus, heart disease, obesity, immunocompromised status, cancer, chronic kidney disease, high d-dimer levels, low lymphocyte counts, high C-reactive protein (CRP) levels, and elevated lactate dehydrogenase (LDH) [2-7].

Currently, artificial intelligence (AI) and machine learning (ML) methods have become one of the most frequent methods to identify clinical phenotypes across various diseases. The increasing number of studies on the use of AI and ML in medicine indicates a growing trend in this field [8]. Previous studies have identified distinct clinical phenotypes in COVID-19 patients using unsupervised machine learning approaches. For example, Sokolski et al. identified phenotypes in patients with cardiovascular comorbidities, Epsi et al. described early symptom clusters correlated with hospitalization and long-term outcomes, and Siepel et al. characterized phenotype evolution during ICU care [9–11].

In this study, we aimed to identify clinical phenotypes of hospitalized COVID-19 pneumonia using an unsupervised ML technique. We hypothesized that these findings provide actionable insights for prognosis and management by highlighting phenotypes that may benefit from intensified monitoring or targeted interventions.

Materials and methods

We conducted a retrospective study including data from Ramathibodi Hospital and Chakri Naruebodindra Medical Institute, which is an affiliated hospital of Ramathibodi Hospital. Ethical approval was obtained from the Ramathibodi Hospital Ethics Committee, which oversees research conducted across both institutions (MURA2022/606).

Participants

Electronic medical records (EMR) of patients with COVID-19 pneumonia who were hospitalized at Ramathibodi Hospital and Ramathibodi Chakri Naruebodindra Hospital, Thailand, between July 2021 and September 2021 were reviewed for eligibility. The diagnosis of COVID-19 pneumonia was made by a combination of a positive polymerase chain reaction (PCR) test for SARS-CoV-2 and the presence of pulmonary infiltrations identified by either chest radiographs or computed tomography (CT) scans. Patients requiring mechanical ventilation at admission or those with missing data for any of the variables listed below were excluded.

Data collection and outcomes

EMRs of eligible patients were reviewed. Variables were selected for clustering based on their known association with COVID-19 mortality and intubation, as well as COVID-19 pathophysiology [2-7]. The following variables obtained within 24 h of admission were collected: age, sex, height, body weight, clinical presentations, pre-existing comorbidities, smoking status, blood pressure, respiratory rate, body temperature, heart rate, oxygen saturation, complete blood count (CBC), blood urea nitrogen (BUN), serum creatinine, liver function tests, serum d-dimers, serum CRP, serum LDH, HbA1C, and the cycle threshold ratio of PCR for SARS-CoV-2. The worst value was selected to analyze variables with multiple measurements recorded within the first 24 hours after admission. This approach was chosen to reflect the highest severity of illness during the early phase of hospitalization.

The severity of pulmonary infiltrations was retrospectively assessed by two independent investigators using the established scoring system, ranging from 0 to 18 [12]. Discrepancies in scoring were resolved through discussion and consensus.

The number of COVID-19 vaccinations received was included as one of the clustering variables. In Thailand, vaccination began in March 2021. During the study period, vaccine availability was initially limited and prioritized for healthcare workers and individuals at high risk of COVID-19 mortality, such as those with diabetes mellitus, obesity, and other comorbid conditions.

The primary outcomes of interest were 90-day mortality and invasive mechanical ventilation required within 90 days of admission.

Unsupervised machine learning clustering analysis

Unsupervised ML clustering analysis was performed using the K-means clustering method to identify clinical phenotypes of COVID-19 pneumonia. A total of 48 variables (as shown in Table 1) were included in the ML K-means clustering algorithm. Prior to clustering, all variables were normalized by mean centering to 0 and scaling the standard deviation to 1. The Silhouette scores for each number of clusters (K ranges from 2 to 10) were calculated, and the elbow method was used to determine the optimal number of clusters. The K-means clustering analysis was conducted using the Orange Data Mining program (version 3.38.1, available at https://orange.biola

Table 1	Baseline characteristics of included	patients classified by	/ cluster

Variables	All (N=538)	Cluster 1 (N=27)	Cluster 2 (N = 370)	Cluster 3 (N=141)	P-value
Age, years	56.3±18.7	53.4±15	52.1±18.7	67.7±14.2	< 0.001
Male, N (%)	250 (46.5)	19 (70.4)	156 (42.2)	75 (53.2)	0.003
Ever smoking, N (%)	72 (13.4)	3 (11.1)	49 (13.2)	20 (14.2)	0.90
BMI, kg/m ²	27.0 ± 6.5	26.9 ± 5.4	27.7±6.9	25.1 ± 5.1	< 0.001
Vital signs					
Body temperature, °C	36.8 ± 0.5	36.8 ± 0.4	36.7±0.4	37.0 ± 0.7	< 0.001
Respiratory rate, times/min	22.4 ± 3.8	22.1 ± 2.5	21.4±2.7	24.8 ± 5.3	< 0.001
Heart rate, beats/min	88.4 ± 15.4	90.5 ± 16.4	88.4 ± 15.1	88.0 ± 15.8	0.75
Mean arterial pressure, mm Hg	100.7 ± 14.0	103.2 ± 15.0	102.5 ± 12.6	95.6±16.0	< 0.001
SpO ₂ % at ambient air	96.4 ± 2.9	95.9 ± 2.6	97.3±1.3	94.1 ± 4.3	< 0.001
Number of vaccinations, N (%)					0.16
No vaccination	334 (62.1)	20 (74.1)	227 (61.4)	87 (61.7)	
1 dose of vaccination	141 (26.2)	3 (11.1)	95 (25.7)	43 (30.5)	
2 dose of vaccination	63 (11.7)	4 (13.8)	48 (13.0)	11 (7.8)	
Comorbid disease, N (%)					
Diabetes mellitus	159 (29.6)	4 (14.8)	82 (22.2)	73 (51.8)	< 0.001
Ischemic heart disease	34 (6.3)	0 (0)	18 (4.9)	16 (11.3)	0.01
Chronic lung diseases	35 (6.5)	1 (3.7)	22 (5.9)	12 (8.5)	0.48
Cerebrovascular diseases	26 (4.8)	3 (11.1)	13 (3.5)	10 (7.1)	0.07
Immunocompromised patients	29 (5.4)	2 (7.4)	17 (4.6)	10 (7.1)	0.48
Cancer	28 (5.2)	2 (7.4)	15 (4.1)	11 (7.8)	0.20
Chronic renal failure	53 (9.9)	1 (3.7)	16 (4.3)	36 (25.5)	< 0.001
Symptoms					
Fever	309 (57.4)	14 (51.9)	213 (57.6)	82 (58.2)	0.83
Chest pain	25 (4.6)	2 (7.4)	18 (4.9)	5 (3.5)	0.64
Dyspnea	82 (15.2)	6 (22.2)	50 (13.6)	26 (18.4)	0.22
Cough	409 (76.0)	18 (66.7)	275 (74.3)	116 (82.3)	0.09
Headache	88 (16.4)	2 (7.4)	73 (19.7)	13 (9.2)	0.007
Sputum	179 (33.3)	8 (29.6)	117 (31.6)	54 (38.3)	0.33
Sore throat	148 (27.5)	3 (11.1)	107 (28.9)	38 (27.0)	0.13
Muscle ache	152 (28.3)	9 (33.3)	113 (30.5)	30 (21.3)	0.10
Rhinorrhea	169 (11.4)	5 (18.5)	131 (35.4)	33 (23.4)	0.01
Gastrointestinal symptom	102 (19.0)	7 (25.9)	66 (17.8)	29 (20.6)	0.50
Chest X-ray score	2.6 ± 1.6	2.9 ± 1.3	2.0 ± 0.9	4.2 ± 2.0	< 0.001
CT ratio of PCR for COVID – 19	20.38 ± 6.15	20.55 ± 5.70	19.76±6.01	21.96±6.33	0.001
CBC					
Hemoglobin, g/dl	13.1±2.0	13.6 ± 2.1	13.5 ± 1.7	11.9±2.2	< 0.001
Hematocrit, %	39.9 ± 5.9	41.0±6.3	41.2 ± 5.1	36.1±6.3	< 0.001
White blood cell counts, cells/mm ³	5823 ± 2807	5219 ± 2607	5633 ± 2374	6435 ± 3681	0.008
Lymphocyte counts, cells/mm ³	1561 ± 931	1276±743	1755±977	1108±617	< 0.001
Platelets count, 10 ³ /mm ³	246.7 ± 97.7	275.9±141.2	250.3±81.9	231.7±121.8	0.04
Liver function test					
AST, IU/L	45.4±36.8	142.8±68.9	34.6±18.0	54.8 ± 29.5	< 0.001
ALT, IU/L	41.0±38.4	156.3±76.7	33.7±21.0	38.2 ± 26.0	< 0.001
ALP, IU/L	83.7±43.5	172.8±92.2	47.8±25.6	90.0 ± 46.6	< 0.001
GGT, IU/L [†]	50.5 (28-100)	441 (246-615)	45 (27-80)	57 (29-111)	< 0.001
Total bilirubin, mg/dL [†]	0.4 (0.3–0.6)	0.6(0.5–0.8)	0.4(0.3–0.6)	0.5(0.4–0.8)	< 0.001
Direct bilirubin, mg/dL †	0.2 (0.2–0.3)	0.3(0.3-0.4)	0.2(0.2-0.3)	0.3(0.2-0.4)	< 0.001
Albumin, g/dL	4.2 ± 0.4	4.2 ± 0.5	4.3±0.3	3.7 ± 0.4	< 0.001
Total protein, g/dL	7.1±0.6	7.6±0.6	7.7±0.5	7.1±0.6	< 0.001
Renal function					
Creatinine, mg/dL [†]	0.87 (0.72-1.08)	0.85 (0.70-1.16)	0.83 (0.70-1.00)	1.01 (0.81-1.48)	< 0.001
Blood urine nitrogen mg/dl †	12 (9–16)	12 (9–15)	11 (9–14)	17 (12–29)	< 0.001

Table 1 (continued)

Variables	All (N=538)	Cluster 1 (N=27)	Cluster 2 (N = 370)	Cluster 3 (N = 141)	P-value
HbA1c, mg%	6.25 ± 1.68	5.76±0.76	6.14±1.63	6.64±1.87	0.003
D-dimers, ng/ml [†]	527.5 (321–870)	575 (340–1113)	433 (281–672)	850 (487–1519)	< 0.001
C-reactive protein, mg/dL [†]	18.8 (5.9–45.3)	23.9 (14.2–84)	11.9 (3.8–26.5)	61.2 (33.1–108)	< 0.001
Lactate dehydrogenase, U/L	246.1±96.9	340.7±134.9	209.5 ± 57.9	324.1±111.7	< 0.001

Continuous variables present as mean ± SD

Categorical variables present as frequency (%)

[†]Variables present as median (IQR)



Fig. 1 Study enrollment flow chart

b.si/), a free and open-source data visualization and analysis program.

Statistical analysis

Continuous variables were summarized as a mean with standard deviation (SD) or median with interguartile range (IQR), as appropriate. Categorical variables were presented as frequencies and percentages. Missing data were handled by excluding patients with missing values for key clustering variables. A standard mean difference of ±0.3 was used to identify key characteristics of each cluster. To compare the data differentiation between clusters, we used ANOVA or the Kruskal-Wallis test for continuous variables and the chi-square test for categorical variables. Cox proportional hazard models were performed to evaluate the association of each cluster with 90-day mortality and intubation rate. A P-value of < 0.05 was considered statistical significance. Statistical analyses, including survival analysis and other comparisons, were conducted using R version 4.0.3, accessed through the RStudio interface (RStudio, Inc., Boston, MA, USA).

Result

Of 790 COVID-19 patients admitted between July and September 2021, 252 patients were excluded due to the absence of pneumonia (n = 196), requiring mechanical ventilation upon admission (n = 16), or missing data (n = 40), as shown in Fig. 1. The remaining 538 hospitalized COVID-19 pneumonia patients were included in the K-means clustering analysis. Using the elbow method, 3 clinically distinct clusters were identified, including 27 patients (5%) in cluster 1, 370 patients (68.8%) in cluster 2, and 141 patients (26.2%) in cluster 3. The baseline characteristics of each cluster are described in Table 1. The standard mean difference plot was assessed to identify clinically distinct clusters (Fig. 2) and pairwise comparisons between clusters provided in the supplementary file (Table S1).

Phenotype characteristics

Cluster 1 consisted predominantly of males and was characterized by elevation of serum aspartate aminotransferase (AST), alanine aminotransferase (ALT), gamma-glutamyltransferase (GGT), alkaline phosphatase



Fig. 2 Mean standardized differences for each baseline variable were compared across the three clusters. The X-axis represents the standardized differences value, and the Y-axis represents baseline variables. The cut-off values of the mean standardized difference of <-0.3 or >0.3 were indicated by dashed vertical lines

(ALP), and LDH. The mean for serum AST, ALT, ALP, and LDH levels in cluster 1 were 142.8 ± 68.9 , 156.3 ± 76.7 , 172.8 ± 92.2 , and 340.7 ± 134.9 IU/L, respectively, while the median serum GGT level was 441 (IQR; 246–615) IU/L.

Cluster 2 was characterized by a lower chest x-ray score and higher serum albumin levels. The mean chest x-ray score and serum albumin levels in cluster 2 were 2.0 ± 0.9 and 4.31 ± 0.3 g/dl, respectively.

Finally, cluster 3 was more likely to have a higher mean age (67.7 ± 14.2 years), higher prevalence of diabetes mellitus (51.8%), higher mean chest x-ray score (4.2 ± 2.0 scores), higher mean respiratory rate (24.8 ± 5.3 /minute), higher mean body temperature (37.0 ± 0.7 °C), lower mean blood pressure (95.6 ± 16.0 mmHg), higher median serum creatinine levels (1.01, IQR; 0.81-1.48 mg/dl), lower mean hemoglobin levels (11.9 ± 2.2 g/dl), lower mean lymphocyte counts (1108 ± 617 cell/mm³), higher median CRP levels (61.2, IQR; 33.1-108 mg/dL), higher median D-dimer levels (324.1 ± 111.7 U/L).

Association between phenotype and clinical outcomes

Cluster 1 had a 90-day mortality rate of 0%, while cluster 2 and cluster 3 had rates of 1.6% and 10.6%, respectively. Compared to cluster 2, cluster 3 was significantly associated with an increased risk of 90-day mortality (HR 6.24, 95% CI 2.42–16.09, P<0.001) (Fig. 3).

The rate of 90-day intubation was 3.7%, 2.4%, and 12.8% in cluster 1, cluster 2, and cluster 3, respectively. When compared to cluster 2, cluster 3 was associated with a

significantly increased risk of 90-day intubation (HR 5.26, 95% CI 2.37–11.72, P<0.001), while cluster 1 was not significantly associated with an increased risk of 90-day intubation compared to cluster 2 (HR 1.40, 95% CI 0.18–11.04, P=0.75), as showed in Fig. 4.

Discussion

Our study identified distinct clinical phenotypes in hospitalized COVID-19 pneumonia. Three distinct clusters were identified upon hospital presentation among patients with COVID-19 pneumonia. The majority of patients were classified into cluster 2, associated with a lower severity, leading to an approximately 90-day mortality rate of 1.6% and an intubation rate of 2.4%. In contrast, a subset of patients with the more severe disease were classified into cluster 3, associated with a 90-day mortality rate of 10.6% and an intubation rate of 12%. Finally, cluster 1, characterized by elevated liver enzymes, demonstrated a 90-day mortality rate of 0% and an intubation rate of 3.7%.

A previous study of 1022 hospitalized COVID-19 patients, with or without pneumonia, were also classified into 3 distinct phenotypes [13]. One of their phenotypes, associated with low mortality, was characterized by lower D-dimer and CRP levels, comparable to our findings in cluster 2. Additionally, the high-mortality phenotype mainly consisted of patients with elevated inflammation markers, similar to those in cluster 3 of our study. Even though the previous study demonstrated that hepatocellular injury was associated with a worse prognosis, our study revealed that patients in cluster 1, who had



Fig. 3 Kaplan-Meier curves of 90-day mortality



Fig. 4 Kaplan-Meier curves of intubation

hepatocellular injury without other organ failures, had an excellent survival rate. However, given the limited sample size of cluster 1, these findings should be interpreted with caution and validated in larger cohorts.

A potential concern with clustering analyses is that smaller clusters may sometimes represent outliers or noise rather than meaningful subgroups, for instance, in cluster 1. However, our analysis demonstrated that cluster 1 exhibits distinct clinical characteristics that justify a meaningful subgroup. While both clusters 1 and 3 had elevated LDH levels, cluster 1 was uniquely characterized by significantly higher AST and ALT levels, suggesting distinct liver involvement. Furthermore, patients in cluster 1 presented with less severe clinical conditions compared to those in Cluster 3.

These findings are further supported by the mean standardized differences, which highlight the unique clinical features of each cluster, including those distinguishing cluster 1. The results of pairwise comparisons, included in the supplementary materials (Table S1), also reinforce the distinctiveness of cluster 1.

In our study, patients in cluster 2 demonstrated favorable outcomes, mostly patients with high albumin levels and low chest X-ray scores. High albumin levels indicate good nutritional status and lower systemic inflammation, while low chest X-ray scores suggest less severe pulmonary involvement. In contrast, cluster 3 was associated with the poorest clinical outcomes. This cluster mainly included patients of higher age, having comorbidities, such as diabetes mellitus and renal disease, having more pulmonary infiltrations on chest X-rays, higher CRP levels, higher D-dimer levels, lower lymphocyte counts, and an injury to multiple organs.

Several previous studies have demonstrated that certain risk factors are associated with a worse prognosis in COVID-19 pneumonia. For instance, a study conducted in China reported that older age, higher SOFA score, and D-dimers > 1.0 µg/mL upon admission were associated with increased risk of in-hospital death in COVID-19 [14]. Another retrospective study found that patients with severe COVID-19 and comorbid diabetes had increased leukocyte and neutrophil counts, as well as higher levels of CRP, D-dimers, and fibrinogen [15]. A systematic review and meta-analysis also revealed that biomarkers such as higher CRP, D-dimers, increased creatinine, and lower albumin levels were associated with increased mortality [16]. Our findings in cluster 3 were consistent with previous reports, confirming that patients with increased inflammatory marker levels and organ dysfunction are associated with increased mortality risk.

Phenotypic approaches provide more comprehensive information for patient prognosis compared to traditional risk factor analysis since this approach involves complex interactions between risk factors. Although individual patients might not have all the defining characteristics of a specific phenotype, they may still benefit from the treatment for the overall phenotype. Future research on phenotypic models should be conducted to improve personalized management of complex diseases.

Our study has several limitations to concern. Firstly, as a retrospective study, it is subject to potential biases

and residual confounding. Secondly, our study focused on 90-day mortality and intubation. Thus, long-term outcomes were not assessed. Additionally, the study did not identify the specific COVID-19 strain. The exclusion of patients with missing data for key clustering variables could theoretically reduce the diversity of the dataset and introduce selection bias. However, the number of excluded patients was small (40 patients, 5% of the total cohort), which minimizes the potential impact on the generalizability of our findings. The use of K-means clustering, which assumes spherical clusters and relies on Euclidean distance, makes it most suitable for continuous variables. While our dataset consisted exclusively of continuous variables, the predominance of continuous variables among the distinguishing features may partially reflect this methodological choice. Additionally, K-means clustering may not fully capture complex, non-spherical data structures, potentially limiting the phenotype identification. Alternative clustering techniques could provide complementary insights or refine the results. Future studies should explore these approaches to validate and expand upon our findings. Finally, our study was conducted in 2021, before widespread vaccination, advanced COVID-19 treatments, and the emergence of new variants. These factors may limit the direct applicability of our findings to current practice. Nonetheless, the phenotypic approach remains valuable as it highlights broader disease patterns and interactions that provide insights for personalized care.

Conclusion

In conclusion, our study identified three distinct clinical phenotypes among hospitalized patients with COVID-19 pneumonia, each associated with different risks of 90-day mortality and intubation.

Supplementary Information

The online version contains supplementary material available at https://doi.or g/10.1186/s12890-025-03536-w.

Supplementary Material 1

Author contributions

NN contributed to data collection, analysis, and manuscript drafting and is responsible for ensuring data accuracy. RT contributed to the study design, data analysis, and data interpretation. TT contributed to data collection and manuscript revision. DE contributed to the study conception, design, and manuscript revision. SS, VB, YS, DJ, and PT contributed to the study conception, data interpretation, and manuscript revision. TP contributed to the study conception, and manuscript revision. All authors reviewed and approved the final version of the manuscript for submission.

Funding

None.

Page 8 of 9

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

The study was approved by the Institutional Review Board of the Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Thailand (MURA2022/606). As this was a retrospective study, the requirement for obtaining informed consent from participants was waived by the ethics committee. The research was conducted in accordance with the ethical standards of the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 18 October 2024 / Accepted: 28 January 2025 Published online: 08 February 2025

References

- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395(10223):497–506.
- Bartoletti M, Giannella M, Scudeller L, Tedeschi S, Rinaldi M, Bussini L, Fornaro G, Pascale R, Pancaldi L, Pasquini Z. Development and validation of a prediction model for severe respiratory failure in hospitalized patients with SARS-CoV-2 infection: a multicentre cohort study (PREDI-CO study). Clin Microbiol Infect. 2020;26(11):1545–53.
- Fu L, Wang B, Yuan T, Chen X, Ao Y, Fitzpatrick T, Li P, Zhou Y, Lin Y-f, Duan Q. Clinical characteristics of coronavirus disease 2019 (COVID-19) in China: a systematic review and meta-analysis. J Infect. 2020;80(6):656–65.
- Gordon CJ, Tchesnokov EP, Woolner E, Perry JK, Feng JY, Porter DP, Götte M. Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. J Biol Chem. 2020;295(20):6785–97.
- Guan W, Liang W, Zhao Y, Liang H, Chen Z, Li Y, Liu X, Chen R, Tang C, Wang T. Comorbidity and its impact on 1590 patients with Covid-19 in China: a nationwide analysis Eur Respir J. press DOI 2020, 10:13993003.13900547-13992020.
- Qin C, Zhou L, Hu Z, Zhang S, Yang S, Tao Y, Xie C, Ma K, Shang K, Wang W. Dysregulation of immune response in patients with coronavirus 2019 (COVID-19) in Wuhan, China. Clin Infect Dis. 2020;71(15):762–8.
- Tang N, Li D, Wang X, Sun Z. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. J Thromb Haemost. 2020;18(4):844–7.
- MacEachern SJ, Forkert ND. Machine learning for precision medicine. Genome. 2021;64(4):416–25.
- Epsi NJ, Powers JH, Lindholm DA, Mende K, Malloy A, Ganesan A, Huprikar N, Lalani T, Smith A, Mody RM, et al. A machine learning approach identifies distinct early-symptom cluster phenotypes which correlate with hospitalization, failure to return to activities, and prolonged COVID-19 symptoms. PLoS ONE. 2023;18(2):e0281272.
- Siepel S, Dam TA, Fleuren LM, Girbes ARJ, Hoogendoorn M, Thoral PJ, Elbers PWG, Bennis FC, Dutch ICUDSC. Evolution of clinical phenotypes of COVID-19 patients during Intensive Care Treatment: an unsupervised machine learning analysis. J Intensive Care Med. 2023;38(7):612–29.
- Sokolski M, Trenson S, Reszka K, Urban S, Sokolska JM, Biering-Sorensen T, Hojbjerg Lassen MC, Skaarup KG, Basic C, Mandalenakis Z, et al. Phenotype clustering of hospitalized high-risk patients with COVID-19 - a machine learning approach within the multicentre, multinational PCHF-COVICAV registry. Cardiol J. 2024;31(4):512–21.
- Borghesi A, Maroldi R. COVID-19 outbreak in Italy: experimental chest X-ray scoring system for quantifying and monitoring disease progression. Radiol Med. 2020;125(5):509–13.

- Lusczek ER, Ingraham NE, Karam BS, Proper J, Siegel L, Helgeson ES, Lotfi-Emran S, Zolfaghari EJ, Jones E, Usher MG. Characterizing COVID-19 clinical phenotypes and associated comorbidities and complication profiles. PLoS ONE. 2021;16(3):e0248956.
- Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet. 2020;395(10229):1054–62.
- Ye W, Chen G, Li X, Lan X, Ji C, Hou M, Zhang D, Zeng G, Wang Y, Xu C. Dynamic changes of D-dimer and neutrophil-lymphocyte count ratio as prognostic biomarkers in COVID-19. Respir Res. 2020;21(1):1–7.
- Tian W, Jiang W, Yao J, Nicholson CJ, Li RH, Sigurslid HH, Wooster L, Rotter JI, Guo X, Malhotra R. Predictors of mortality in hospitalized COVID-19 patients: a systematic review and meta-analysis. J Med Virol. 2020;92(10):1875–83.

Publisher's note

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