Supplementary File

Methods

Appendix 1. Study design, setting, and outcome definition

This retrospective cohort included all consecutive adults (≥18 years) admitted to Baqiyatallah Hospital (October 2020–May 2021) with reverse transcription polymerase chain reaction (RT-PCR)–confirmed COVID-19 and a chest computed tomography (CT) within one day of admission. Exclusions: incomplete clinical data (e.g., missing CT or key labs) or inter-hospital transfers/early discharge within 24 hours. The binary outcome was peripheral oxygen saturation (SpO_2) <90% vs ≥90% using the first measurement within two hours of arrival and prior to high-flow oxygen or ventilatory support. Of 1,744 screened, 736 were excluded (incomplete data n=416—220 no CT, 196 missing labs/clinical; transfers/early discharge n=320—190 transferred, 130 <24-h discharge), yielding 1,008 patients (training n=706; validation n=302). Ethics approval: IR.BMSU.BAQ.REC.1400.079; all data were anonymized. The sample size (n=1,008) aligns with machine learning (ML) guidance of ~10–20 samples per retained feature; models retained 8–22 features.

Appendix 2. Clinical data collection and verification

Electronic health records extraction captured demographics (age, sex), exposure/contact history, comorbidities (hypertension, diabetes, coronary heart disease, prior surgery, hepatitis B, etc.), presenting symptoms (fever, cough, chills, dizziness, fatigue, myalgia), and routine admission labs (White blood cell [WBC], lymphocytes, eosinophils, neutrophils, C-reactive protein [CRP], D-dimer, lactate dehydrogenase [LDH], creatine phosphokinase [CK]-MB, alanine transaminase [ALT], aspartate aminotransferase [AST], creatinine, blood urea nitrogen [BUN], procalcitonin). In a validation subset (n=150), self-reported symptoms were cross-checked against clinician notes: fever concordance 93.4% (κ=0.81), cough κ=0.79, fatigue κ=0.75.

Appendix 3. CT acquisition protocol

CTs were acquired on a GE Revolution EVO 64-slice scanner (GE Healthcare, Milwaukee, WI) within one day of admission: supine position, 120 kVp, automatic tube current modulation, 0.725 mm collimation; reconstructions at 1-mm and 5-mm intervals. Coverage spanned thoracic inlet to upper abdomen. Native 12-bit images were mapped to 8-bit with lung windows (width 1000 Hounsfield units [HU]; level −500 HU).

Appendix 4. Segmentation workflow and quality assurance

A 2D U-Net (3×3 convolutions, batch normalization, ReLU) was trained in-house on this cohort for lung/lesion segmentation. Data augmentation included random rotations (≤20°), shear, and zoom (0.9–1.1). Automated masks were reviewed by four radiologists (≥2 years' chest imaging

experience); final adjudication by two seniors (6 and >11 years). Mean Dice (U-Net vs expert-adjusted) was 0.946 for lungs and 0.882 for lesions. Inter-rater reliability across 120 cases: mean Cohen's κ for binary CT features 0.84 (range 0.78–0.89); intraclass correlation coefficient (ICC) for continuous measures (e.g., lesion volume, non-lesion lung volume [NLLV]) 0.91 (95% CI 0.88–0.94).

Appendix 5. CT feature engineering and radiomics harmonization

Volumetric metrics included lesion volume (mL; integrated across slices with slice-thickness normalization), total lung volume, non-lesion lung volume (NLLV = total lung − lesion), and %NLLV (NLLV/total). Texture features (entropy, skewness, kurtosis, energy) were derived from 3D histograms over lesion and NLLV regions using SimpleITK and SciPy with image biomarker standardization initiative (IBSI)-conformant preprocessing. CT intensities were normalized to −1000 to 400 HU; all masks were resampled to isotropic 1-mm^3 voxels. Binary pattern features (e.g., ground-glass opacity, consolidation, crazy-paving, halo/reverse-halo, peripheral distribution, lower-zone predominance, traction bronchiectasis, vascular thickening, subpleural lines, air bronchograms, pleural effusion, interstitial thickening, lymphadenopathy, cavitation, fibrotic bands) were radiologist-verified. Crazy-paving was encoded as present/absent.

Appendix 6. Data integration, encoding, and imputation

Clinical and lab variables were standardized using z-scores $z=(x-\mu)/\sigma$ computed from the training set (μ, σ). Categorical variables were one-hot or binary encoded (e.g., nodule size/number/location; architectural distortion types). Overall missingness was 2.7% across the dataset, affecting 9.4% of patients; most missing values were in procalcitonin (6.1%), D-dimer (4.9%), and CRP (2.3%), largely reflecting early pandemic logistics. Little's missing completely at random (MCAR) test: $\chi^2=204.3$ (df=213, p=0.64), supporting MCAR. Multiple imputation by chained equations (20 iterations) was used; pooled estimates followed Rubin's rules.

Appendix 7. Feature selection and stability analyses

To balance bias–variance with n≈700/≈300 splits, we targeted 8–22 retained features depending on classifier. Methods: (i) Recursive feature elimination (RFE) for Linear support vector machine (SVM) and SVM-radial basis function (RBF) (typical optima 7–13 features); (ii) embedded importances for Random Forest/XGBoost using split frequency and impurity reduction (often 13–22 features); (iii) univariate screening for Logistic Regression/Naïve Bayes (Pearson for continuous; χ^2 for categorical; $p<0.05$); and (iv) Minimum redundancy maximum relevance (mRMR) to reduce redundancy (yielding ~10–15 features). Stability was assessed via subsampling: 100 random 70% training subsamples recorded selected sets; stability score = selection frequency. Features with stability ≥0.7 were considered robust; sensitivity analyses

repeated training using only stability ≥0.8 features and after trimming top/bottom 1% of continuous values. Performance changes were minimal (ΔAUC <0.02).

Appendix 8. Classifiers, hyperparameter tuning, and class imbalance handling

Models: Linear SVM (linear kernel, tuned C), SVM-RBF (tuned C, γ), Logistic Regression (L1/L2 with tuned C), Random Forests (tuned max_depth, min_samples_split, min_samples_leaf), Gaussian Naïve Bayes (tuned var_smoothing), and XGBoost (tuned learning_rate, max_depth, n_estimators, min_child_weight). Hyperparameters were optimized with Bayesian search (scikit-optimize BayesSearchCV), 4-fold stratified CV, 30 iterations/model, using balanced accuracy as the objective. Class imbalance mitigation included synthetic minority over-sampling technique (SMOTE) and adaptive synthetic sampling (ADASYN) resampling and cost-sensitive learning. Two preprocessing pipelines were compared: StandardScaler-normalized vs original scales. Models were not trained with defaults.

Appendix 9. Validation, metrics, and interpretability

Primary evaluation used 10-fold stratified cross-validation on training data and a held-out validation set (n=302). Metrics included area under the curve (AUC), balanced accuracy, sensitivity, specificity, precision, and F1—emphasizing minority-class performance. Interpretability combined: (i) normalized coefficients (Linear SVM, Logistic Regression); (ii) embedded importances (RF/XGBoost) with permutation importance; and (iii) Shapley additive explanations (SHAP) values for global rank and local attribution consistency. Convergence between SHAP distributions, embedded ranks, and stability scores was examined.

Appendix 10. Unsupervised analyses and visualization

Principal Component Analysis supported visualization (e.g., SVM-RBF decision landscapes). Component retention used eigenvalue >1 and scree-plot inspection to capture substantial variance while avoiding overfitting.

Appendix 11. Use of large language model assistance

Specialized configurations of GPT-4 were used for (i) guiding hyperparameter search setups and reporting templates, (ii) literature review support, and (iii) language refinement of the manuscript. Final methodological and analytical decisions, as well as all data handling and model training, were performed by the study team. All of AI assistances in writing, review of literature and other technical aspects were checked and supervised carefully by the authors.

Appendix 12. Clinical Features of COVID-19 Patients by Oxygen Saturation Levels

Clinical characteristics	O2 < 90%		O2 ≥ 90%	
	Training (N=224)	Validation (N=96)	Training (N=482)	Validation (N=206)
Age (years)	64.8 ± 12.3	66.2 ± 12.8	55.4 ± 10.9	54.6 ± 11.2
Gender (Female)	45.1 (101)	46.9 (45)	48.3 (233)	47.6 (98)
Body mass index (kg/m²)	28.47 ± 4.21	28.6 ± 4.5	26.7 ± 3.9	26.2 ± 3.7
Contact History	33.0 (74)	35.4 (34)	28.4 (137)	27.7 (57)
Clinical Examination				
Oxygen Saturation (%)	85.3 ± 3.9	85.5 ± 4.1	95.2 ± 3.2	94.7 ± 3.1
Diastolic Pressure (mmHg)	75.1 ± 10.4	76.3 ± 10.7	73.5 ± 9.4	72.4 ± 8.8
Respiratory Rate (breaths/min)	24.8 ± 5.7	25.4 ± 5.9	18.7 ± 4.6	18.4 ± 4.2
Systolic Pressure (mmHg)	129.7 ± 15.3	131.2 ± 15.7	120.8 ± 14.2	118.9 ± 14.0
Body Temperature (°C)	38.52 ± 1.19	38.4 ± 1.3	37.3 ± 0.9	37.2 ± 1.0
Comorbidities and Smoking				
Hypertension	55.0 (123)	57.3 (55)	45.2 (218)	44.7 (92)
Diabetes	20.1 (45)	21.9 (21)	15.6 (75)	16.5 (34)
Cardiovascular Disease	17.9 (40)	18.8 (18)	12.2 (59)	11.7 (24)
COPD	12.1 (27)	11.5 (11)	8.92 (43)	9.71 (20)
Chronic Liver Disease	8.04 (18)	9.38 (9)	7.26 (35)	6.80 (14)
Asthma	9.82 (22)	10.4 (10)	8.09 (39)	8.25 (17)
Emphysema	5.80 (13)	6.25 (6)	5.18 (25)	5.34 (11)
Cancer	7.14 (16)	6.25 (6)	4.56 (22)	5.83 (12)
Symptoms				
Fever	84.8 (190)	82.3 (79)	67.6 (326)	69.4 (143)
Cough	80.4 (179)	78.1 (75)	72.2 (348)	73.3 (151)
Chills	45.1 (101)	46.9 (45)	28.4 (137)	26.7 (55)
Fatigue	70.1 (157)	72.9 (70)	52.5 (253)	51.9 (107)

Body Aches	65.2 (146)	68.8 (66)	51.7 (249)	50.5 (104)
Dizziness	29.9 (67)	31.3 (30)	22.0 (106)	21.4 (44)
Loss of Taste/Smell	40.2 (90)	41.7 (40)	34.4 (166)	35.9 (74)

This table presents the clinical characteristics, comorbidity profiles, and symptomatology of COVID-19 patients, stratified by oxygen saturation levels (below 90% and at or above 90%) and further categorized into training and validation cohorts. Continuous variables are expressed as mean ± standard deviation, while categorical variables are presented as percentages with the number of patients affected. COPD: Chronic Obstructive Pulmonary Disease.

Appendix 13. Biological Features of COVID-19 Patients by Oxygen Saturation Levels

Biological Measures	O2 < 90%		O2 ≥ 90%	
	Training (N=224)	Validation (N=96)	Training (N=482)	Validation (N=206)
White Blood Cell Count (per µL)	10,530 ± 2,450	10,620 ± 2,550	9,400 ± 2,200	9,450 ± 2,100
Lymphocyte Count (per µL)	796 ± 152	820 ± 160	1,050 ± 230	1,060 ± 240
Eosinophil Count (per µL)	41.2 ± 19.8	43.7 ± 22.3	55.6 ± 25.1	56.4 ± 26.1
Neutrophil Count (per µL)	7,987 ± 1,989	8,080 ± 2,030	6,400 ± 1,800	6,460 ± 1,790
C-Reactive Protein (CRP, mg/L)	60.3 ± 20.7	61.2 ± 21.1	38.4 ± 15.6	37.7 ± 16.0
Platelet Count (per µL)	180,000 ± 29,700	181,500 ± 31,200	200,500 ± 25,600	201,000 ± 25,100
D-Dimer (µg/mL)	2.48 ± 1.03	2.65 ± 1.10	1.80 ± 0.80	1.75 ± 0.78
Lactate Dehydrogenase (LDH, U/L)	298 ± 98	306 ± 103	265 ± 79	262 ± 77
Alanine Aminotransferase (ALT, U/L)	40.1 ± 14.8	41.3 ± 15.2	32.7 ± 11.5	33.1 ± 12.4
Aspartate Aminotransferase (AST, U/L)	50.3 ± 20.2	52.1 ± 21.7	40.8 ± 14.3	41.2 ± 13.7
Procalcitonin (µg/mL)	0.20 ± 0.11	0.22 ± 0.13	0.10 ± 0.05	0.12 ± 0.06

This table presents the biological measures of COVID-19 patients, stratified by oxygen saturation levels (below 90% and at or above 90%) and further categorized into training and validation cohorts. Continuous variables are expressed as mean ± standard deviation.

Appendix 14. Computed Tomography Features of COVID-19 Patients by Oxygen Saturation Levels

Computed Tomography Features	O2 < 90%		O2 ≥ 90%	
	Training (N=224)	Validation (N=96)	Training (N=482)	Validation (N=206)
Lesion Volume (mL)	502 ± 101	508 ± 105	250 ± 75	248 ± 72
Non-Lesion Lung Volume (NLLV, mL)	2,005 ± 505	2,010 ± 520	2,700 ± 600	2,710 ± 610
Ground-glass opacity	75.0 (168)	78.1 (75)	45.0 (217)	43.2 (89)
Consolidation	59.8 (134)	62.5 (60)	25.6 (123)	26.2 (54)
Crazy Paving	50.0 (112)	52.1 (50)	30.1 (145)	29.1 (60)
Halo Sign	40.2 (90)	42.7 (41)	20.8 (101)	18.9 (39)
Reversed Halo Sign	35.3 (79)	36.5 (35)	17.4 (84)	16.5 (34)
Peripheral Topography	79.9 (179)	81.3 (78)	50.8 (245)	49.5 (102)
Lower Zone Predominance	70.1 (157)	71.9 (69)	45.6 (220)	44.7 (92)
Vascular Thickening	50.0 (112)	52.1 (50)	30.1 (145)	29.1 (60)
Subpleural Lines	55.0 (123)	56.3 (54)	36.7 (177)	37.9 (78)

This table presents the computed tomography features of COVID-19 patients, stratified by oxygen saturation levels (below 90% and at or above 90%) and further categorized into training and validation cohorts. Continuous variables are expressed as mean ± standard deviation, while categorical variables are presented as percentages with the number of patients affected.

Appendix 15. Complete set of features selected by the top-performing classifier in each model type.

Model Type	Feature Name	Importance	Stability
Clinical (Logistic Regression, n = 11)	Age (years)	0.51	0.89
	Gender (binary)	0.33	0.81
	Fever (yes/no)	0.31	0.73
	Cough	0.29	0.71
	Fatigue	0.28	0.69
	Chills	0.27	0.72
	Hypertension	0.25	0.70
	Body mass index (kg/m²)	0.23	0.68
	Diastolic Pressure (mmHg)	0.21	0.66
	Respiratory Rate (breaths/min)	0.20	0.72
	Loss of Taste/Smell	0.19	0.65
Laboratory (Linear SVM, n = 13)	WBC count (per µL)	0.53	0.88
	Lymphocyte count (per µL)	0.35	0.83
	Platelet count (per µL)	0.32	0.80
	Neutrophil count (per µL)	0.31	0.78
	C-reactive protein (mg/L)	0.30	0.76
	Lactate dehydrogenase (U/L)	0.29	0.75
	Alanine transaminase (U/L)	0.28	0.74
	Aspartate aminotransferase (U/L)	0.27	0.73
	D-Dimer (µg/mL)	0.26	0.72
	Procalcitonin (µg/mL)	0.25	0.70
	Creatinine (mg/dL)	0.24	0.68
	Blood Urea Nitrogen (mg/dL)	0.23	0.67
	Albumin (g/dL)	0.22	0.66
CT-Based (Random Forest, n = 20)	Mean Lesion Volume (mL)	0.24	0.90
	Lower Zone Predominance	0.20	0.85
	NLLV Skewness	0.16	0.80
	Crazy Paving Pattern	0.15	0.76
	Consolidation	0.14	0.74
	Ground Glass Opacity	0.13	0.72
	Subpleural Lines	0.12	0.70
	Peripheral Distribution	0.11	0.69
	Vascular Thickening	0.11	0.69
	Reversed Halo Sign	0.10	0.68
	Pleural Effusion	0.09	0.67
	Entropy (Lesion Texture)	0.09	0.66
	Kurtosis (NLLV Texture)	0.08	0.65
	Mean NLLV Volume (mL)	0.08	0.64
	Fibrotic Bands	0.07	0.63
	Cavitation	0.07	0.62
	Lymphadenopathy	0.06	0.61
	Architectural Distortion (any)	0.06	0.60
	Traction Bronchiectasis	0.06	0.60
	Air Bronchograms	0.06	0.59
Integrated (SVM-RBF, n = 22)	WBC count (per µL)	0.31	0.88
	Mean NLLV Volume (mL)	0.30	0.85

	Crazy Paving Pattern	0.22	0.72
	Age (years)	0.20	0.83
	Lesion Volume (mL)	0.19	0.81
	C-reactive protein (mg/L)	0.19	0.80
	Fever	0.18	0.78
	Subpleural Lines	0.18	0.77
	Ground Glass Opacity	0.17	0.75
	Platelet Count (per μL)	0.17	0.75
	Lower Zone Predominance	0.16	0.74
	Alanine transaminase (U/L)	0.15	0.73
	Aspartate aminotransferase (U/L)	0.15	0.72
	Hypertension	0.14	0.70
	Respiratory Rate (breaths/min)	0.13	0.69
	Vascular Thickening	0.13	0.68
	D-Dimer (μg/mL)	0.12	0.68
	Fatigue	0.12	0.67
	Entropy (Lesion Texture)	0.11	0.66
	Traction Bronchiectasis	0.11	0.65
	Body mass index (kg/m²)	0.10	0.64
	Cough	0.10	0.63

Importance values reflect raw, model-specific feature contributions. Stability indicates the proportion of 100 resampled training subsets in which the feature was retained. WBC: White blood cell; NLLV: Non-lesion lung volume; SVM: Support vector machine; RBF: Radial basis function.