International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I8P112

Volume 74 Issue 8, 182-196, August 2026


*Original Article*

# A Hybrid 3D CNN and Clinical Data Fusion Model for COVID-19 Mortality Prediction

Khaoula Echabbi[1], Mohammed Douimi[2], El Moukhtar Zemmouri[3], salsabil Hamdi[4]

[1,2,3,4]*Laboratory for Mathematical Modeling and Computer Science, Moulay Ismail University-ENSAM, Meknes, Morocco.*

[1]*Corresponding Author : echabbikhaoula@gmail.com*

Received: 09 February 2026 Revised: 18 June 2026 Accepted: 11 July 2026 Published: 29 August 2026

***Abstract*** *- The COVID-19 pandemic revealed acute flaws in hospitals' decision-making processes, exposing substantial deficiencies in how critically ill patients are managed or in one dimensionalizing complex multidimensional data in a timely manner. Many clinical scoring systems such as the Sequential Organ Failure Assessment (SOFA) and the Acute Physiology and Chronic Health Evaluation II (APACHE-II) are still very widely used based upon their implementation in routine clinical practice, however, both methods are limited by their reliance on a small number of predetermined clinical characteristics as well as not allowing for extraction of high dimensional, data from multiple complex data sources (e.g., volumetric CT). Significant advancements in deep learning techniques have provided significant advancements in terms of both the potential applications as well as how accurately and automatically images can be interpreted. However, most of these methods are limited by only using a single image modality (i.e., either imaging data or clinical data) and therefore provide limited performance since characteristics of either type of data represent only part of the overall information available from the combination of both types of data. To address these limitations, we propose a multimodal deep learning framework for assessing the probability of death in COVID-19 patients that utilizes volumetric CT imaging in conjunction with structured clinical data and integrates a 3D convolutional neural network for CT image feature representation and a dense neural network for clinical representation of clinical data. In order to ensure consistent and reliable data input for our models, we have developed a dedicated preprocessing pipeline for our data that includes lung segmentation and Hounsfield unit normalisation. Results of our experiments completed on Moroccan COVID-19 patient data show that our multimodal approach outperformed the unimodal approach, suggesting that data fusion with respect to predicting risk in a clinical context provides significant benefit. In addition, our proposed framework will be generalisable and may be applicable to many other areas of medicine requiring the integration of multimodal datasets.*

***Keywords*** *- Artificial intelligence, Computed tomography, Deep learning, Medical imaging, Multimodal learning.*

## 1. Introduction

Global Healthcare Systems have been severely tested by COVID-19, with the structure of hospitals and decisions made by healthcare professionals, especially in Intensive Care Units, being challenged [1]. High patient volume and lack of ICU capacity made these issues worse. Quickly and accurately identifying patient severity is important for the efficient use of resources and improved outcomes.

New developments in artificial intelligence and deep learning can provide additional perspectives on medical diagnosis and Prognosis. 3D convolutional neural networks have been shown to effectively determine spatial patterns from CT Imaging. However, there is a lack of research in combining multiple data types within the clinical and imaging domains. Most studies are focused on one data type, usually imaging; therefore, the clinical application of the current technology is limited.

This issue also exists in Morocco during the COVID-19 Crisis, as the disparity in the skill level of radiology and the unequal geographic distribution of imaging facilities led to delays in CT imaging interpretations and created stress on the healthcare system. Clinicians are presented with large amounts of complex radiology data that support the need for automated, reliable Decision Support Systems that can process large amounts of data.

These problems have led to this study that proposes to create a Multimodal Deep Learning Network (MDLN) to predict COVID-19 Patient Mortality using Volumetric CT Imaging and Structured Clinical Data. The proposed MDLN framework employs a Hybrid Architecture (3D CNN for imaging characteristics plus DN for clinical variables). Pre-processing, such as Lung Segmentation and the Normalization of Hounsfield Units will also be performed to enhance the quality of the data used in the training of the model.

The primary contribution of this work is to combine multiple complementary data sources to develop a more robust representation of the status of the patient.

The study will test the proposed framework on a cohort of Moroccan COVID-19 patients; however, it will also have the potential for the same application in other cases where Multimodal Integration of Data will have the potential to be impactful.

This study lays the foundation for future research into creating and validating multimodal deep learning decision support systems through multi-site studies and the development of a broad comparative dataset that supports the development of robust and clinically reliable decision support systems in a real-life environment.

## 2. Background and Related Work

Patients presenting with possible indicators of severe COVID-19 as a primary focus for both hospitals and hospitals during public health emergent situations wherein the health system is operatively limited.

Estimating mortality can assist hospitals in allocating critical care beds and developing treatment strategies for at-risk patients, as well as to make decisions in a timely manner when responding to emergencies [2].

Thus, finding ways of creating more accurate predictive models will continue to warrant further exploration into this topic.

While many studies have made significant advancements in terms of methodology, the current models employed have significant limitations in their applications due to a lack of generalizability; however, some studies have successfully employed multiple methods of prediction by combining structured data that is segmented by diagnostic procedure, and it was found that both methods demonstrated significant value in their ability to predict adverse outcomes.

Therefore, until the existing discrepancies between each predictive method are resolved in order to achieve an accurate representation of the patient population, improvement will not occur within the patient population.

The research provided in this document is a guide to developing more precise models of risk stratification of COVID-19 as suggested from the previous work that has been published in peer-reviewed literature and will further organize the topic of mortality prediction into three classifications: (i) prediction of mortality utilizing clinical Metadata, (ii) prediction of mortality utilizing imaging, and (iii) prediction of mortality with the use of multimodal data combining deep learning and traditional machine learning approaches and algorithms.

### *2.1. Estimates of Mortality based on Clinical Metadata*

Many researchers have concluded by studying the effect of demographic data such as race, age, and disability and co-morbidities such as hypertension, diabetes, and heart disease on the severity of COVID-19 disease [3].

Therefore, providing effective variables for estimating mortality risk. Age, by itself, as the most frequently used risk factor, has been studied extensively with regard to COVID-19 and the elderly. Generally, older adults suffer from higher rates of mortality than younger adults for a variety of reasons (e.g., immune senescence, presence of pre-existing illnesses, etc.); however, researchers have also concluded that men are generally more likely than women to die from increased rates of severe disease from COVID-19 due to an intrinsic difference in immune function attributed to the presence of the Angiotensin-Converting Enzyme 2 (ACE2) receptor on the surface of the body's cells. There are numerous co-morbidities that are known to be strong independent predictors of mortality (e.g., diabetes, hypertension, heart disease, kidney disease, and obesity); therefore, data from these co-morbidities is often presented as a simple binary score. However, in some instances, researchers have included weighted scoring when assessing the severity of these co-morbidities.

C-Reactive Protein (CRP), Lactate Dehydrogenase (LDH), ferritin, lymphocyte count, and the Neutrophil-To-Lymphocyte Ratio (NLR) are representative laboratory diagnostic markers of acute inflammation and/or coagulopathy and demonstrate high predictive correlations with both overall mortality and with advanced stages of COVID-19 disease. Logistic regression analysis, random forest classification, Support Vector Machine (SVM) analysis and gradient-boosting methods have been shown to provide results similar to one another with regard to the predictive capability of clinical metadata models, but each has limitations due to a variety of factors. Neural networks have also been proposed by some researchers as a way to use data from both metadata and radiology/imaging studies of patients' lung pathology to provide more accurate predictive values for COVID-19 mortality estimates based on non-linear associations seen through both traditional and advanced methods of data analysis.

Current limitations of the current clinical metadata-based models are a direct result of the lack of complete information provided when relying solely on either clinical Metadata or imaging data. When relying on imaging, the models fail to assign an accurate severity assessment of the patient's lung pathology as determined by Computed Tomography (CT) scan, which is a critical element in determining the potential severity of COVID-19. Another limitation when relying on metadata, data must be considered at the institutional level; therefore, models can be limited based on missing values, thereby diminishing the predictive capability of these models.

Therefore, using only metadata-based clinical information for predicting mortality from COVID-19 using clinical Metadata will only provide a partial mortality prediction.

### *2.2. Imaging-Based Mortality Prediction*

Medical imaging is an essential part of the pulmonary assessment of COVID-19 patients. Ground-glass opacities, consolidations, bilateral infiltrates and a peripheral distribution of lesions are some of the common radiological images that are strongly associated to disease severity. The reason why the Chest X-ray (CXR) is so popular is because it is fast, available and affordable.

Convolutional Neural Network (CNN) deep learning models, including ResNet, DenseNet, and EfficientNet, have demonstrated good performance in detecting abnormalities and predicting mortality. They are typically pretrained on ImageNet and fine-tuned on COVID-19 datasets. However, chest X-ray is still limited by its two-dimensional projection and lower sensitivity compared to CT scanning.

Chest CT has a much greater anatomic detail and can volumetrically evaluate lung lesions. 2D-slice-based CNNs and 3D-volumetric models such as 3D-ResNet, I3D, and C3D have enhanced severity assessment and mortality prediction. Also introduced are attention mechanisms and ensemble learning strategies that enhance robustness and classification performance.

Despite these developments, models based on imaging remain significant drawbacks. A large majority of them do not take into account the patient-specific clinical context, which is often required to stratify risks.

Irrespective of such developments, imaging-based models are a major limitation. Most of them fail to consider the patient-specific clinical context, which in most cases is needed to rank risks.

Moreover, estimating the 3D CT process is computationally intensive, and variations in acquisition procedures can limit a model's ability to predict across institutions. The second significant weakness is that it is not explainable, an issue that remains a significant barrier to clinical adoption.

### *2.3. Deep Learning Architectures for COVID-19 Mortality Prediction*

The use of deep learning models in predicting mortality has increased in recent years because they can deal with large, varied data and the relationships between those different types of data are often complex and not linearly related to each other [4]. CNN-based architectures have been successfully used to extract spatially-related features from CT images: for example, the texture of the lesions, the volume of the lesions, and where lesions are located with respect to other structures in the body. ResNet, DenseNet and 3D-ResNet to identify the volume of a lesion were all used as presumptive encapsulations [5].

RNNs and LSTM networks have been used to analyze the progression of diseases over time by using sequential clinical data (such as the patient's vital signs and the trends of the patient's laboratory work), allowing for improved monitoring of patients over a long period of time [6].

Hybrid CNN-Fully Connected Networks have produced superior results compared to unimodal systems in terms of fusing data from image encoding with structured clinical Metadata [7]. Recently, transformer-based models (including ViT) and transformer-based models mimicking BERT have been proposed to learn complex relationships between conceptual features with the aid of attention mechanisms [8].

The majority of studies using deep learning-based predictions for mortality continue to use basic methods of combining the features derived from different data types into a single prediction due to a lack of justification for designing the architecture used in the study. Studies with a limited amount of validation, including explainability, robustness and external validation studies, limit our confidence in implementing deep learning-based decision processes in real-world settings.

### *2.4. Multimodal Approaches and Research Gap*

Mortality forecasting is no longer able to depend solely on a specific type of information, recent research shows. Each patient’s outcome is influenced by both the visible lung damage demonstrated on their CT scans and the overall clinical picture presented through their other lab values and demographic characteristics.

The main study [9] developed a model based on an LSTM type of Machine Learning (ML) algorithm trained using 122 clinical/demographic characteristics with an accuracy of 90%. Although there are several good results from this study, there are many limitations to the model.

First, the sample size of participants was not identified which raises the issue of whether or not the model can generalize to patients not included in the sample. Second, the study did not provide external validation of their findings making it impossible to determine the generalizability of their results.

Third, the model was not analyzed for feature importance/interpretability within the context of clinical significance, so clinicians cannot fully understand how the predictions were made. Lastly, lung imaging data was excluded from analysis so that investigators could not directly evaluate lung damage, which would be regarded as one of the

most significant predictors of COVID-19 severity. Study [10] compared Several ML Models (SVM: 82% accuracy, random forests: 80%, decision trees: 76%) to determine mortality based on clinical/laboratory data and comprised 100 patients. While the comparison of the ML models provides interesting information, the number of subjects used for the analyses is extremely low. The authors also did not provide adequate information about the validation method, including the training-test split strategy/cross-validation, nor did they mention whether or how they addressed the issue of class imbalance. Additionally, biomarkers that are less specific, such as CRP, LDH, and ferritin, have limited ability to reliably assess lung involvement, which represents a key component of COVID-19 prognostication.

The study used inflammatory and coagulatory markers (CRP, D-dimer, ferritin) as predictors in the analysis using random forests, SVMs, and logistic regression [11]. There were several limitations with the study: 1) it utilized only static images of the patients and did not take into account the evolution of the disease over time, and 2) it did not use any type of imaging data in order to assess pulmonary injury, thus limiting the available methods to measure the injuries produced by the virus. In addition, as inflammatory markers could signify the presence of extra-pulmonary complications, they may also lead to an inaccurate assessment of the overall disease severity of the patients.

The authors proposed the use of multimodal approaches to combine clinical data and medical imaging data to improve predictive accuracy [12]. The results from their study showed that when using combined modalities to make predictions, the models utilized to make predictions on heterogeneously represented data types performed better than when unimodal predictions were made. However, there are still some limitations with this study, such as having an inadequate number of entries in the training datasets and not validating their models outside of the institution in order to assess the external generalizability of their models. In addition, the research did not include a dialogue on how each modality contributed to the outcome of the prediction model. A study conducted by the authors of [1] focused on the use of deep learning models for medical imaging data, specifically the use of CNNs to obtain the spatial characteristics from CT scans and their ability to determine the viability of lung lesions associated with COVID-19. Using only imaging data without any additional clinical metadata means that there is no patient or context-specific information available to support the real-world applicability of the model.

In another article, [13], the authors explored using hybrid and multimodal deep learning for mortality prediction using multiple data types; however, neither method is generally interpretable and does not provide an exact contribution of each data type. Moreover, variances in types of tested datasets and a lack of rigorous validation can influence the validity and generalization of the developed models.

As a result, this study has identified a clear gap in the existing literature. Most models focus exclusively on clinical Metadata or have utilized simplified MDAS fusion methods without sufficient validation. In conjunction with published studies, there has yet to be developed any rigorously defined frameworks that allow for the integration of non-aggregated volumetric CT scans and well-structured clinical data while maintaining an emphasis on interpretability and generalizability.

To address this deficit, this article presents a novel multimodal deep learning approach that integrates 3D volumetric CT imaging data with clinical Metadata (clinical observations) via a hybrid architecture of 3-Dimensional Convolutional Neural Network (CNN) and Dense Neural Network (DNN). Unlike currently existing frameworks that use only metadata, this architecture permits direct analysis of lung injury as well as the utilization of additional related clinical data. This method aims to improve mortality prediction accuracy and provide enhanced risk analysis with clinically meaningful, individualized probability scores that will facilitate the prioritization of high-risk patients in the inpatient hospital environment.

## 3. Methodology

### *3.1. Dataset Description*

Our data originates from a collaborative research project involving Sheikh Khalifa Hospital. The data are from a private group of COVID-19 patients who had Computed Tomography (CT) scans of their chest while hospitalized. The data include volumetric CT scans in DICOM format, and each patient's associated structured clinical Metadata. This dataset was developed based on actual clinical experiences in Moroccan hospitals, where the actual experience of variability of quality of image acquisition, lack of available data, and poor quality of clinical documentation were all frequent operational issues. The variables you will find in this dataset include demographic variables, clinical symptoms, radiologic results, quantitative results from imaging, and final outcome (death or survival). These variables were chosen because of their known clinical significance for forecasting COVID-19 severity and mortality.

The major variables are:

- Patient ID: Patient anonymized ID.
- Age: Patient's age (num.).
- Gender: Gender variable (binary) for categorical gender;
- Diabetes: Indicates presence of diabetes (binary).
- Fever at Admission: Indicates presence of fever at admission (binary).

- Consolidation on CT: represents the presence of consolidation on CT (category).
- Crazy Paving: represents the presence of the craziest paving pattern on CT (category).
- Ground Glass Opacity (GGO): represents the presence of GGO on CT (category).
- Volume Affected by Disease (%): Percentage of total lung volume affected by disease (cont).
- Diffuse Density (HU): Indicated via Hounsfield Units on CT.
- Consolidated Density: (HU): Indicated via Hounsfield Units on CT.
- Death: reflects the outcome of death (binary).

CT scans of the chest provide excellent imaging information describing pulmonary lesions and disease severity. However, Metadata provides important and relevant contextual information on patient condition and systemic risk factors. The combination of these complementary data sources of image and Metadata will produce a more thorough representation of patient status upon which mortality forecasting can be based.

### *3.2. Clinical Data Analysis and Feature Selection*

To determine which variables had the strongest relation to patient mortality, an Exploratory Data Analysis (EDA) of structured clinical Metadata was required to assist in reducing dimensionality, removing redundant predictors, and increasing interpretability prior to performing the multimodal fusion. The primary purposes of performing the EDA were to increase the predictive power of the resulting predictive models and to assist in finding clinically significant predictors of mortality risk.

#### *3.2.1. Image Preprocessing*

A dedicated preprocessing pipeline was used to ensure consistency, robustness and compatibility between images and clinical Metadata prior to model training [14].

*Handling of Missing Values*

The manner in which missing values were addressed depended on the variable type. Continuous variables were imputed using the sample median to limit sensitivity to extreme values (example, age, percentage of volume achieved, diffuse and condensed density (HU)). Categorical variables were addressed by using the mode as the metric of central tendency (example, sex, diabetes, fever, condensation, crazy paving and ground glass opacity). Patients that did not have DICOM scans available or had insufficient quality scans were excluded from the study to preserve the quality of imaging.

*Standardisation of Features*

The application of z-score normalisation to continuous variables ensured standardisation of the data: Figure 1 This transformation results in most numeric variables being placed upon a consistent scale and therefore precludes any one feature from dominating because of its relative size/value.

*Detection of Outliers*

The Interquartile Range (IQR) method was used to detect outliers, with any value outside of the IQR determined as an outlier. These outlier values were then removed for all continuous variables in order to mitigate noise and optimise model stability.

*Encoding of Categorical Variables*

The binary encoding for binary categorical variables was assigned as [0, 1], while categorical variables with more than two categories were encoded using one-hot encoding to satisfy the requirements of a neural network. Proper class distribution analysis was performed due to the potential for severe class imbalance within mortality prediction data sets; should you discover class imbalances, either the Synthetic Minority Oversampling Technique (SMOTE) or a class-weighted loss must be applied so that your model will remain unbiased by the majority.

#### *3.2.2. Variable Importance Analysis using XGBoost*

The feature importance evaluation utilized the Extreme Gradient Boosting technique (XGBoost) to identify the key features associated with mortality. XGBoost [15] is an ensemble learning tree-based parameterization with strong predictive power, the ability to model non-linear data robustly, and a good ability to model the interaction of features.

As such, compared to traditional linear models, XGBoost is a more flexible modelling tool for complex clinical variable relationships, thus making it a good option for modelling medical risk prediction.

The two key uses of the XGBoost algorithm in this study include:

- Determining the most effective predictors of mortality from clinical variables.

The feature predictive importance ranking can be used pre-multimodal fusion. To improve generalization of the model and reduce the potential for overfitting, L1 (Lasso) and L2 (Ridge) regularization were employed in the training of the model. The gain score (mean additional prediction accuracy across all trees for each variable) was used to measure the importance of each variable, as assessed by the XGBoost.

The findings of the study indicate that the Volume Achieved (%) variable was the strongest predictor of mortality. This result indicates that the volume of pulmonary involvement is the strongest predictor of mortality. This is also consistent with clinical knowledge that the development of severe respiratory failure.

Other important features from the analysis included Age and Diffuse Density (HU), both of which indicate the

diminished physiological reserve in the older patient population and the degree of deterioration in the lungs. Other variables with low importance, such as Fever and Condensation, are complementary information and will also be used with imaging variables generated by the 3D-CNN model.

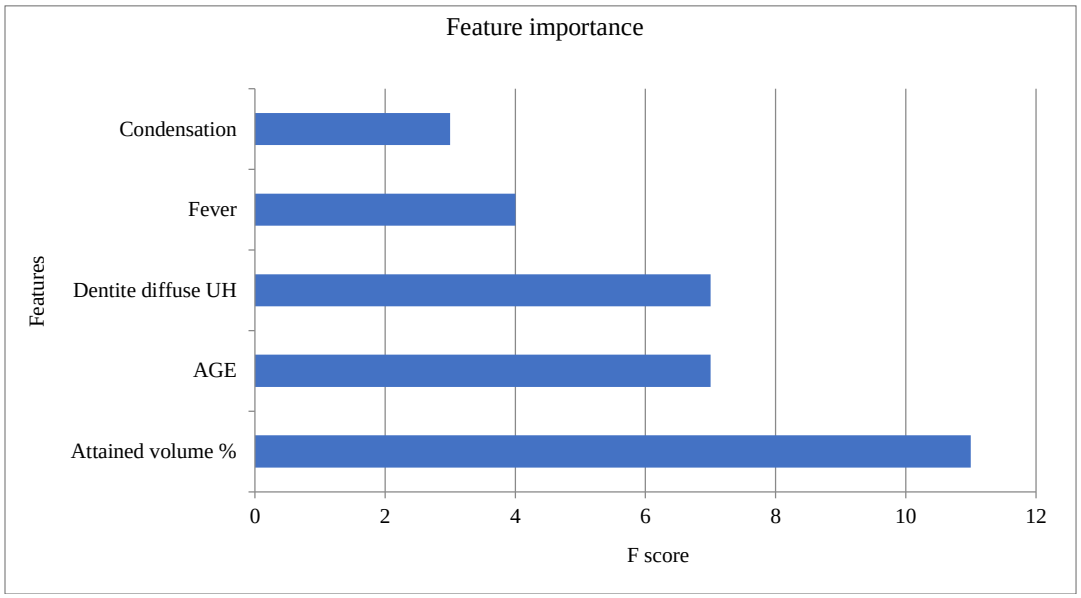



**Fig. 1 Feature importance**

### *3.3. CT Scan Processing and Lung Segmentation*

The successful use of AI-based automated clinical prognosis of COVID-19 requires accurate preprocessing of CT scans of the IVFs (primarily chest) in order to predict mortality based on volumetric analysis of individuals under investigation for Covid-19. Raw CT volumes typically include extraneous features which affect classification performance due to their presence; these include the ribs, mediastinum, heart tissue and surrounding soft tissue structures. In order to remedy this deficiency, a special preprocessing and segmentation pipeline was developed to remove all surrounding tissue prior to classification. The rationale was to train the deep learning model to only use clinically relevant features associated with the severity of COVID-19, such as GGO, consolidations and diffuse infiltrates. By implementing an ROI-based methodology for each of these cases, it was possible to increase robustness, reduce false positives and also improve classification accuracy by minimising the effect of background structures which do not correspond to the pathological process under investigation. Figure 2 depicts the entire preprocessing pipeline from image standardisation to Hounsfield Unit normalisation to lung segmentation to volumetric preparation for 3D-CNN analysis of the CT images.

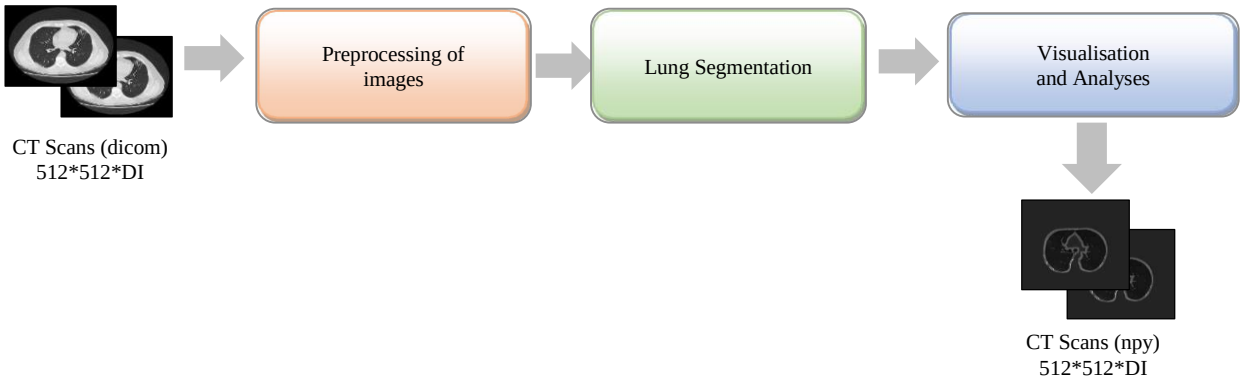



**Fig. 2 Proposed CT scan preprocessing operations**

#### *3.3.1. CT Image Preprocessing and Standardisation*

All CT images were obtained from Sheikh Khalifa Hospital and stored in DICOM format. In order to achieve spatial and intensity consistency throughout the CT dataset, each named image was preprocessed, as they were all from different acquisition protocols and scanner parameters.

*Voxel Resampling*

To standardize spatial resolution, all CT volumetric images were resampled to a common isotropic voxel spacing of 1 mm, using a trilinear interpolation method. This step removes any differences in slice thickness and in-plane resolution, thereby ensuring a homogeneous volumetric representation of patients.

*Hounsfield Unit (HU) Normalization*

The radiodensity values of all CT images were converted to Hounsfield Unit (HU) values, which is a standardized measurement of true tissue density on CT [16]. The goal of this conversion is achieved through:

The slope and intercept were taken from the DICOM Metadata TAGs (0028,1053) and (0028,1052); these values are typically set to 1 and -1024, respectively. To reduce the influence of vendors on changes in intensity and create similar densities of the same tissues across scanners, this conversion is required.

*3.3.2. CT Image Standardization and Windowing*

The CT images, all obtained from Sheikh Khalifa Hospital, were archived in Digital Imaging & Communication in Medicine (DICOM) format and acquired under varying protocols of acquisition as well as different scanner settings; thus, a preprocessing step was required to align the images. with respect to both spatial resolution and intensity for the purposes of achieving both spatial and intensity consistency of all scanned files, which required the normalisation of spatial resolution for all CT volumes by means of voxel resampling with trilinear interpolation into a common isotropic voxel with a uniform isotropic voxel size of, thereby normalising all slice thicknesses and in-plane resolutions from different scanners at all scanning sites so that a uniform volumetric representation exists for all individuals.

In addition, the radiodensities of each CT image are expressed in Hounsfield Units (HUs, i.e. Standardized Measures of Tissue Density in CT), and were converted prior to resampling using slopes/intercepts calculated from DICOM metadata tags (0028, 1053 and 0028, 1052) which are normally set at one and -1024, respectively. To achieve similar radiodensity measurements of tissue between vendors, and in order to better characterise the density of tissue, radiodensity measurements are converted to HUs using the above-referenced metadata tags to normalise the radiodensity measurements between different vendors.

For visualising the pulmonary abnormalities identified in the lungs, the following two standard CT window presets will be used:

- Lung Window - Centre: -600 HU, Width: 1500 HU
- Mediastinal Window - Centre: 40 HU, Width: 400 HU

The lung window visualises the pulmonary parenchyma (e.g. ground glass opacity - GGO, and lung consolidation) and the mediastinal window visualises blood vessels and lymph nodes.

The two sets of images will constitute the input to be used to develop the segmentation phase as multi-channel inputs.

*3.3.3. Lung Segmentation*

The segmentation of medical images continues to be a considerable challenge for several key reasons. First, there are large variations in both the anatomy of the lungs, discontinuities with the extent and severity of the disease that exists in the lungs, and lastly, within the various protocols used to create the imaging data. Further complicating the challenges with segmentation, traditional methods (e.g. thresholding, region-growing, atlas-based segmentation) may be satisfied with a set of criteria for normal subject segmentation but may be unsatisfactory for a particular imaging data set containing an acute COVID-19 patient because bilateral coalescences, pleural effusion, and diffuse infiltrates can each independently impact the validity of traditional imaging segmentation criteria. Therefore, the authors employed K-Means clustering as an alternative method of lung tissue segmentation because it is simple to implement, requires minimal resources to execute, and computationally has a high efficiency in a hospital environment where there may be limited resources available for execution of imaging. The K-Means [17] algorithm was executed to segment the voxels into 2 tissues types of major importance - low and high HUs.

The K-Means clustering algorithm was modified using K-Means++ initialization to produce stable and reproducible centroids. The maximum number of iterations allowed for convergence of the K-Means clustering algorithm was capped at 100 iterations with a tolerance of $\leq 0.00001$ HUs. The K-Means clustering provided for the calculation of a segmentation threshold by taking the two terminal cluster centroids from the K-Means clustering and defining the threshold that could be used to delineate the lung from the surrounding structures.The final step of the segmentation is to improve the quality of the segmented lung; therefore, morphological processing was applied to the clusters to improve the segmentation quality. Applying operations of morphological transformation after clustering should provide a marked improvement to the quality of the lung segmentation data set. The morphological transformation used spherical structuring elements with a 2-voxel radius to eliminate small (noisy) areas, vascular artefact (disruption of blood vessel contour), and small sections of trachea. The morphological transformation of trophic dilation used a 10-voxel radius spherical structuring element to restore the morphology at the borders of the lung while preserving the total volume of the lung.

Component analysis of thoracic structures superior to the diaphragm resulted in the isolation of the two largest left and

right lung structures. Component structures with fewer than 10,000 voxels were excluded. Moreover, the convex hull operations were executed to retrieve all areas of concavity associated with pathology. The final lung mask files created in NIfTI file format and reviewed for anatomical consistency by a radiology professional. Segmentation performance evaluation of the auto-generated segmented masks compared to semi-manually labeled masks appropriately assigned by the same two radiologists for a total of 100 randomly selected cases from the total database of cases; mean DSC and IoU were calculated to evaluate performance.

The results of the evaluation of the two segmentation processes were as follows (in Table A): The results of this study provide evidence of agreement of the two types of segmentation processes (automatic versus manual); therefore, the authors validate the image processing pipeline for segmentation of optical tissue lesions.

### *3.4. 3D Deep Learning Network for the Classification of CT Scans*

COVID-19 causes pulmonary lesions such as ground glass opacities, bilateral infiltrates, and consolidations. These kinds of pulmonary lesions can be represented volumetrically as 3D volumetric representations and appear on multiple slices. The geographic location of these lesions helps in determining the prognosis for morbidity and mortality severity.

In contrast to traditional 2D CNNs, the use of 3D Convolutional Neural Networks (3D-CNNs) is more effective in the processing of these pulmonary lesions because the entire volumetric representation is taken into account (as opposed to the individual slice) with a correlation to all of the slices.

Therefore, 3D architectures will be useful in the analysis of severe pneumonia resulting from viral infections using volume CT imaging.

Advantages of the 3D Convolutional Neural Networks (3D-CNNs) over conventional 2D CNNs include [18]:

1: The use of 3D convolution filters allows for simultaneous height, width, and depth scanning and evaluation, resulting in volumetric patterns of lesions being registered.
2: Because 3D max-pooling maintains spatial coherence within the area of four pixels in addition to providing a reduction in dimensionality, 2D pooling partitions the pixels of the slices, destroying spatial coherence.
3: The 3D-CNNs enable a recognition of lesions, the area of the lesions, and the degree of bilaterality with respect to volume CT imaging 2D CNNs.

#### *3.4.1. Proposed Multimodal Architecture*

The proposed two-parallel line approach is shown in Figure 3.

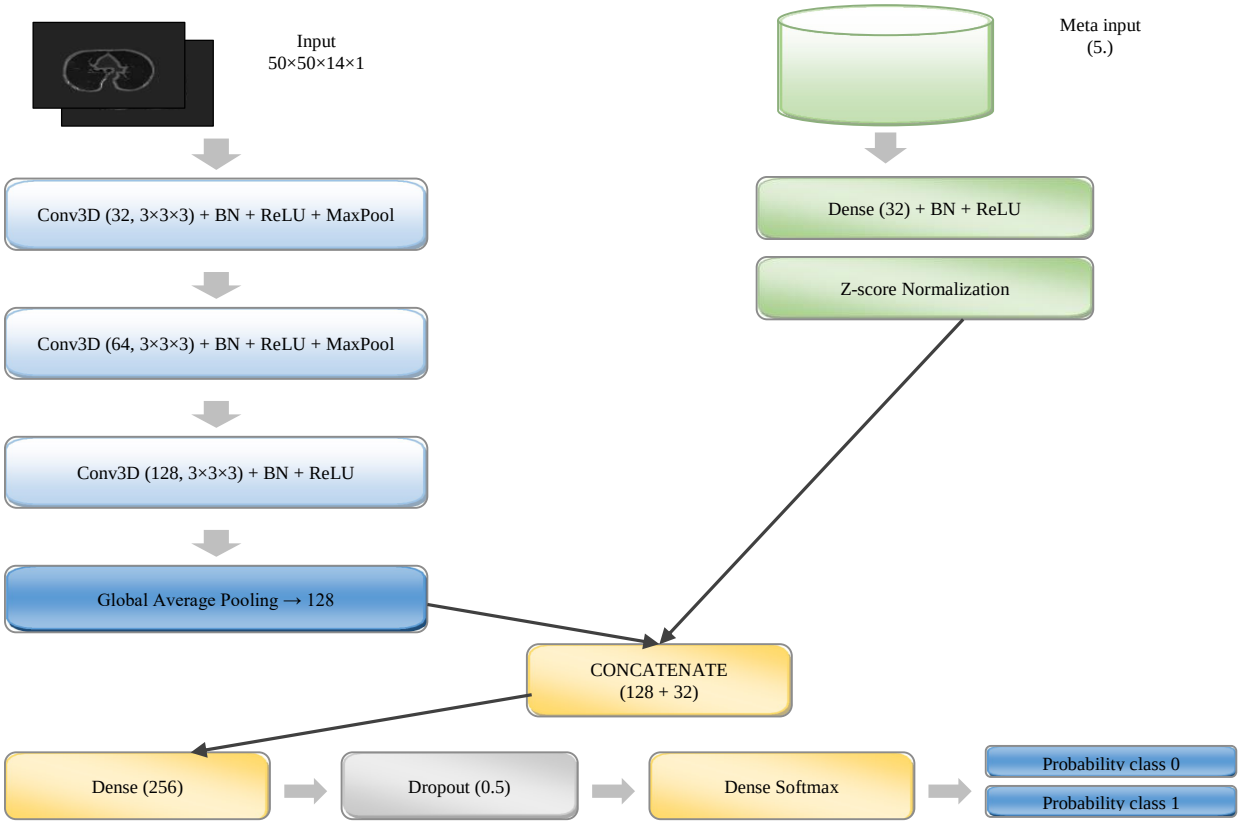



**Fig. 3 Proposed 3D CNN architecture**

Branch 1 - CT Imaging Branch (3D CNN)

Slice Selection: A total of 14 representative slices were sampled to represent the full anatomical region of both lungs (i.e. apex to base). The imaging branch contains three convolutional blocks that stack together sequentially. Each convolutional block contains:

- 3D Convolution
- Batch Normalization
- ReLU Activation Function
- 3D Max Pooling

As the features are learned progressively from layer to layer, their corresponding depth of texture patterns, representative structural features from intermediate anatomical levels, and identifiable pathological representation of high-level structures. The number of feature dimensions has been reduced through Global Average Pooling (GAP) to represent features that are representative of the lungs.

Branch 2 - Clinical Metadata Branch

This branch includes the implementation of clinical metadata identified through the use of XGBoost feature importance analysis for classification. Clinical Metadata was standardized to the mean by the use of z-scores. Principal Component Analysis (PCA) was considered as an optional dimensionality-reduction technique, but was not applied in the final experimental pipeline. The clinical branch is a fully connected, dense ANN with the following configurations:

- Dense layers
- Batch Normalization
- ReLU Activation Function

This will enable the successful classification of patients based upon the nonlinear relationship between the clinical characteristics of the patient population associated with the respective clinical data for predictive output.

*3.4.2. Late Fusion Strategy*

Merging features selected in two distinct data streams, one a collection of imaging data, and the second a collection of associated metadata, at the end of the analysis, instead of merging both data streams at the beginning of the analysis, allows for an increase in the likelihood of a successful classification of the data sets for the following three reasons:

1. The nature of the feature space associated with images and with Metadata is very different.
2. The use of late fusion provides a level of robustness to the failure of either modality, since classification is done on both features independently before being fused together.
3. The use of late fusion also allows us to have interpretable classification, as the classes generated using late fusion can be understood independently of one another, and thus can be combined in a modular fashion, as opposed to concatenating the data streams providing the features that were used to generate each respective class.

The hybrid features derived from combining both imaging and metadata were transmitted through a three-layer neural network comprised of:

- Dense layer (256 neurons)
- Dropout regularization (rate = 0.5)
- L2 regularization

with the output from the last classification layer providing a unique mortality probability score between 0 and 1. The motivation for using a multimodal feature set is that while metadata-only models ignore pulmonary injury, the imaging-only models ignore the systemic clinical component of disease.

Based on these two deficiencies, the proposed model is a more comprehensive and clinically relevant system for predicting mortality.

## 4. Experiments and Results

### *4.1. Experimental Setup and Data Preprocessing*

This study focuses on estimating the mortality rates of patients who have contracted COVID-19. A cohort of 100 patients matched to this research was collected at Sheikh Khalifa Hospital. In order to accomplish this study's aims, both structured clinical phenotypes and chest CT scans are assessed utilizing a multimodal deep learning model. To better predict mortality based on individual patient factors such as age, fever, diabetes and radiographic findings, a group of segmented lung regions (based on volumetric CT scan data) were combined with individual clinical characteristics as a way to analyze improved predictions of mortality.

Due to the relatively small size of the sample population, a strict validation plan using the given data was incorporated to maximize the efficiency of the results and allow for an accurate assessment of performance. Additionally, using stratified sampling allowed for the maintenance of balance between the two classes of patients (ie, surviving/non-surviving) in training and validation sets.

The problem of predicting the outcome of a patient (ie, surviving/non-surviving) was formulated as a binary classification problem:

1) Class 0 - surviving (non-mortal)
2) Class 1 - non-surviving (mortal)

In addition, to provide accurate estimates of the number of patients presenting with COVID-19 and to limit overfitting of the data used during training and validation, data augmentation was accomplished only for training data. Data augmentation procedures included:

1) Random rotation of images (±15°)
2) Random zoom (±10%)
3) Random spatial translation of the volume.

The changes made by applying these augmentation techniques resulted in greater variability in the training data and did not negatively impact the independent test or validation data since there were no data leaks resulting from the augmentation of training data.

In addition, the distribution of each class was evaluated during the entire training process to reduce potential bias that may have been present due to the distribution of patients between classes.

### *4.2. Architecture of Model and Training System*

#### *4.2.1. Multimodal Architecture Design*

An interface architecture that incorporates both volumetric Computed Tomography (CT) images and structured clinical data was created as a hybrid deep learning approach. The architecture relies on a combination of an image-analysis architecture, namely a three-dimensional convolutional neural network (3D-CNN), and a patient metadata architecture, namely a fully connected neural network.

Using these two separate sources of input, the intended result is to apply both image data and patient-specific metadata within the model to improve the prediction of mortality risk for the patient. Advantages of multimodal architecture that utilize spatial pulmonary patterns obtained from CT images, as well as clinical variables related to the patient, allow for contextual information to be available to the model and ultimately enhance the accuracy of the predicted risk of mortality.

The system has two parallel input streams:

1. Input Image - Volumetric CT images with dimensions 14x50x50x1 (14 CT slices selected from the lung region and resized to 50x50 spatial resolution, providing a single channel in grayscale).
2. Input Metadata - Structured normalization of clinical variables for analysis, including age, fever, diabetes, radiologic changes, and quantitative indicators provided via CT analysis.

Evaluation of multicollinearity and potential overfitting prior to the use of input metadata was considered. However, Principal Component Analysis (PCA) was not applied in the final experimental pipeline prior to entry into the fully connected neural network for analysis.

#### *4.2.2. 3 D-CNN Branch*

In the imaging pipeline, there are three sequential convolutional blocks defined by:

- 3-dimensional Convolutional layer (Conv3D)
- Batch Normalization layer
- ReLU activation function
- MaxPooling3D operation

This configuration allows for step-wise extraction of:

- Low-level Textural Features,
- Middle-level Anatomical Structures,
- High-level Pathological Patterns, using a series of adjacent CT scans.

A Global Averaging Pooling (GAP) layer is subsequently applied after the last convolutional layer, and this helps to reduce dimensionality while keeping only the most distinguishing volumetric features as opposed to flattening.

This architecture also limits the potential growth of parameters and maximizes generalization.

#### *4.2.3. Metadata Branch Clinical*

A dense neural network has been designed to form a metadata pathway that incorporates the following components:

- Fully Connected Input Layers
- Batch Normalization
- ReLU Activation Functions

The metadata pathway is a method where clinical risk factor interventions (predicted causes of adverse patient outcomes) are transformed into a latent representation (non-linear combination) of clinical risk factors. This is important because the clinical risk associated with mortality is more often due to an intervention (a variable) performed on a patient rather than by using only predictors; therefore, using only predictor variables will yield unreliable estimates of mortality risk.

#### *4.2.4. Feature Fusion and Classification Head*

Using the late fusion method, the outputs from each branch are merged by concatenating their feature sets. This combined feature set will then be passed through the following layers before reaching the final classification (output) layer:

- 256 Neuron Dense Layer
- Dropout Regularization at 50% Rate
- L2 Weight Regularization

The final (output) layer is a softmax function that produces a probability distribution (between 0 and 1) for both "surviving" and "deceased" potential outcomes.

The result is a robust system that will help maintain consistent performance from the models if either (or both) of the modalities has partial noise or poor quality information.

*4.2.5. Training Configuration*

The architecture was trained for 100 epochs using the Adam optimizer with categorical cross-entropy loss.

The Python programming language was selected to create the architecture using Keras & TensorFlow [19]. The architecture was trained by using the Early Stopping mechanism and conducting Validations to prevent overfitting as well as to provide a stable solution.

***4.3. Experimental Results***

*4.3.1. Comparative Performance Evaluation*

To determine the value of clinical metadata, we assessed two configurations of the proposed architecture. We created two configurations:

Configuration A (CT Image only): Used CT images on the 3D-CNN branch.

Configuration B (All Data): CT images with clinical Metadata in the integrated multimodal architectural hierarchy.

Four standard classification metrics were used to evaluate each configuration's performance:

Precision: Proportion of true positives/(true positives + false positives) [20].

Recall: Proportion of true positives/actual positives in dataset [21].

F1 Score: Harmonic mean of precision and recall [22].

Accuracy: Proportion of correctly classified cases/all cases tested [20].

Metrics were evaluated on the validation set after training convergence.

Quantitative results for the two configurations are summarized in Table 1.

The baseline (CT Images Only) achieved:

- Accuracy 80%
- Precision 64%
- Recall 80%
- F1 Score 71%

There were significant improvements across all metrics after adding Metadata;

Hybrid Architecture was able to achieve:

- Accuracy 86%
- Precision 73%
- Recall 86%
- F1 Score 79%

This corresponds to:

- 6% accuracy improvement,
- 9% precision improvement,
- 6% recall improvement,
- 8% F1 Score improvement.

These results demonstrate that Metadata provides prognostic information beyond what is available from just CT images and enhances the ability to distinguish between mortality and survival cases.

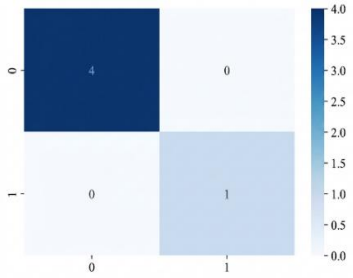
**Table 1. Performance metrics of the 3D CNN model**

| Model | Precision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|
| 3D CNN without Metadata | 64.00 | 80.00 | 71.00 | 80.00 |
| 3D CNN with Metadata | 73.00 | 86.00 | 79.00 | 86.00 |

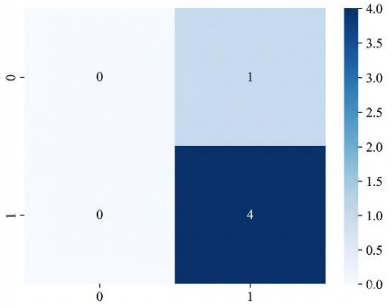
*4.3.2. Confusion Matrix Analysis*

The confusion matrices [23] of the two configurations of the two models, as seen in Figure 4, demonstrate that the model based solely on images generated an above-average number of false positives, as some of the patients that actually were living had been incorrectly classed to be at high risk of mortality. By comparison, the hybrid model showed better class separation, with a significant reduction in false positives.

Metadata characteristics of age, fever, diabetes, & radiological parameters helped the model correctly identify true negative cases and prevent unnecessary mortality notifications. Clinically, it is important that an excessive number of false-positive notifications, coupled with the geographic location of a patient, may lead to incorrect ICU priority assignment and inefficient wastage of limited hospital resources in routine allocation.




**(a) Using only CT images**




**(b) Using CT images combined with Metadata**
**Fig. 4 Confusion matrix**

*4.3.3. Effect of Metadata Integration*

The continual rise in the performance of all tests shows that multimodal learning can be helpful for predicting the prognosis of diseases in medicine. CT images of the lung show the degree of damage from disease, and the clinical Metadata has patient-specific information that cannot be gained through the use of CT images.

The increase in accuracy (by 6%) and precision (by 9%) shows that the clinical Metadata that has been added provides clinically relevant and assistive prognostic variables to the other types of information obtained from CT Images. In clinical practice, doctors would not just use CT images to determine how their patients would do for the rest of their lives. They would use symptoms, history, and the results of lab tests when predicting how the patient will do in the future.

*4.3.4. Statistical Significance Analysis*

Using statistical analysis, a Student t-test [24] was conducted to assess the statistical significance of the performance increase observed through the cross-validated accuracy scores. Each fold of the accuracy scores was performed on a pairwise basis, comparing each accuracy score to one another to find out if they were statistically different from one another. The null hypothesis used to test were that the classification of the images with Metadata was not statistically better than only using images to classify them.

The results from the statistical test showed a t-statistic of 1.00 with 2 degrees of freedom and the one-sided p-value was p-value = 0.21. The significance level $\alpha = 0.05$ did not reject the null hypothesis, showing no statistical differences, and therefore, the hybrid model using Metadata does not provide a significant improvement over just using images, even though the hybrid model had a statistically insignificant increase of 6%.

The main contributing factor to the insubstantial conclusion regarding Metadata is mainly attributed to the limited sample size, n = 100 patients (low statistical power and high variability). The p-value = 0.21 shows a positive trend supporting the integration of Metadata into the evaluation process; however, there is insufficient evidence to state with confidence an established finding at this time.

In addition, the positive outcome of this system was evidenced by all the analyses of the systems performance and not only the accuracy of the system. Therefore, it can be stated that the proposed multivariate strategy is applicable and wielding in this case study.

In summary, this work provides an encouraging but uncertain foundation from which to advance the proposed framework into larger multicenter cohorts with sufficient statistical power, thereby verifying and generalizing the application of the framework in the future.

## 5. Discussion

Our study's results reinforce the hypothesis that CT imaging and clinical metadata provide two different but complementary assessments of the severity of COVID-19. In particular, volumetric chest CT scans can provide more precise spatial detail about various ways COVID-19 damages the lungs, including water in the lungs (ground-glass opacities), fluid accumulation in the lungs (consolidation), lung infection on both sides (bilateral infiltrates), and lung damage from fluid (interstitial changes). These findings are highly correlated with the increased severity of COVID-19 illness.

On the other hand, clinical metadata (i.e., age, the presence of diabetes, fever, and other chronic diseases) describe only the systemic effects of the patient's age and physiologic weakness, and therefore do not reflect the damage to the patient's lungs caused by the virus.

The combination of the two assessment methods (i.e., CT images and clinical Metadata) is much closer to actual clinical decision-making, as radiological information is generally viewed in conjunction with the patient's history, symptoms, laboratory findings, etc., rather than on the basis alone of the radiologic findings. Thus, multimodal learning methods improve the ability to differentiate between survival and death and result in improved prognostic predictions.

A substantial advantage of this study over prior studies is that we have included an evaluation of lung injury using volumetric CT scan analysis as opposed to only using data derived from clinical or laboratory variables. In fact, most models using only clinical or laboratory variables would include systemic inflammatory processes and/or other associated risk factors. Pulmonary injury is one of the most significant contributing variables for COVID-19 mortality; however, conventional means of evaluating the extent of injury are not adequately effective. 3D-CNN technology can maintain inter-slice volumes and volumetrically consistent features better than 2D imaging methods, which rely solely on

the analysis of images sliced into two-dimensional sections. COVID-19 lung pathology is manifest in a bilateral distribution and, often times, it will be manifest across multiple adjacent slices of the lung - the slices will often not completely capture the volume represented by each slice by the 2D analysis alone; therefore, 3D Convolution would help in illustrating the actual anatomy of the pulmonary disease appeared in the model and would provide better visualization of the pulmonary diseased anatomy.

The differences between the two research studies make it difficult to have a simple numerical comparison, as there are also significant differences within the datasets being used, validation methods being utilized, and both included and excluded patient populations between the studies; however, the findings from both researchers demonstrate an overall consistency with multimodal volumetric methods for predicting mortality and the indications received from both studies are in agreement with findings published to date and indicate that there were several architectural choices intended to build robustness of the architecture; however, the total number of images in the dataset is limited to prevent sample bias.

For instance, by using Global Average Pooling, there were fewer parameters available for training and fewer issues related to overfitting the architecture compared to traditional flattening methods; and both L2 weight penalties and dropout regularization increased the ability to generalize. The model was designed using lungs as its exclusive interest because the lung was defined beforehand, allowing the model to operate solely on lung parenchyma by removing extraneous anatomy (e.g., bones, soft tissues, etc.) and thereby reducing the computational burden on the device. As a result of the aforementioned design features, the suggested methodology offers greater feasibility for application within real-world clinical practice settings like those encountered in hospitals where computational resources may be limited.

Statistical analysis of the Student t-test yielded a p-value of 0.21, which is greater than the conventional significance threshold of 0.05. Therefore, the observed performance improvements cannot be considered statistically significant. This issue can largely be attributed to the small number of observations, which is associated with creating minimal statistical power and maximization of variance across the multiple validation folds of the model; therefore, the ease with which one could continue to see improved evaluation metrics (e.g., accuracy, precision, recall, and F1-score) due to metadata integration suggests that the advantages derived from incorporating this type of metadata are unlikely to occur by chance and indicate an actual beneficial effect that is the result of integrating the two types of Metadata into the training of the model. Therefore, future studies will require additional, larger-scale, diverse multicentre external validation cohorts before they can be considered for use in clinical practice.

This dataset reflects the standard practice of care for COVID-19 within a hospital setting in Morocco during the early pandemic; therefore, while generalization may be possible, the lack of multicenter external validation limits the ability to generalize these findings further across population groups and imaging protocols.

There are, however, limitations to this work that require consideration. The first is the relatively small size of the sample population, which limits our confidence in making generalizations. The second is the study design, which is single-centre, thereby limiting extrinsic validity. The third is that explainability methods (e.g., Grad-CAM or SHAP) were not applied in the present study; hence, the interpretability of the final model could not be further assessed. The fourth is that there was relatively little consideration of hyperparameter optimization or exploration of other fusion techniques (e.g., attention-based fusion or transformer-based architecture) employed in this study; these limitations provide an excellent opportunity for future work and reveal the need for larger, more diverse, multicentre cohorts with external validation to be conducted prior to implementing any applications within the clinical setting.

## 6. Conclusion and Future Work

The multi-modal deep learning model presented in the current work develops a framework for predicting mortality due to COVID-19 by utilizing both volumetric Computed Tomography (CT) images and a structured collection of Metadata (clinical data). The architecture proposed in this paper combines a 3D convolutional neural network that extracts spatially relevant pulmonary features from chest CT images with a parallel dense neural network that encodes clinical variables unique to each patient. Next, a late fusion process combines the two modalities and produces a final prediction of mortality.

This design leverages the complementary capabilities of imaging and Metadata while also allowing for computationally efficient application within the limited resources available in many healthcare environments. The overall results from the experiments demonstrated that clinical metadata-based predictor models combined with CT imaging provide greater accuracy, precision, recall, and F1 score than models based on either modality alone. Therefore, multimodal models have a higher likelihood of producing reliable and usable predictions of patient-specific clinical variables to show visual evidence of lung abnormalities. In addition, the results from the confusion matrix analysis showed a significant decrease in false positive results relative to previous studies, particularly in identifying patients who survived, thus reducing unnecessary capacity preference in Intensive Care Units (ICUs) and leading to improved resource allocation in the hospital. Although the statistical significance test had a value below the conventional threshold due to the small number of subjects in this study, the values for all

evaluations were positive with a monotonic trend, indicating that these results should be interpreted as practical relevance rather than an absolute or definitive form for an initial study to prove the viability of deep learning using multiple forms of modality data with clinical relevance to be developed into large-scale validation. This paper broadly asserts that volumetric medical imaging combined with structured clinical data can improve predictive modeling in healthcare.The model presented in this work presents a scalable, flexible base for the inclusion of additional medical devices using multi-modal data fusion across multiple centers. Future studies will work towards conducting larger sample size studies with greater diversity across multiple hospitals to improve generalizability and statistical validity. Further incorporation of methods to produce explainable AI (e.g., Grad-CAM, SHAP) should enhance the interpretability of predictors and improve clinician confidence in automated prediction results. Exploration of other methods of implementing fusion methods, transformer-based multimodal architectures, and extensive hyperparameter optimization will be important targets for future research to maximize the performance and robustness of this model. Ultimately, innovations from AI may yield decision-support tools for physicians working in the challenging environment of healthcare, which will enhance patient outcomes by facilitating early and accurate consideration of risk.

## References


[1] I Ketut Parindra, Menap Menap, and Lalu Abdul Khalik, "Optimization of Hospital Health Services: Insights from Recent Literature," *Bioscientist: Scientific Journal of Biology*, vol. 12, no. 2, pp. 1636-1655, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Tao Ai et al., "Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases," *Radiology*, vol. 296, no. 2, pp. 32-40, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[3] Laure Wynants et al., "Prediction Models for Diagnosis and Prognosis of COVID-19 Infection: Systematic Review and Critical Appraisal," *BMJ*, vol. 369, pp. 1-22, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] O.S. Albahri et al., "Systematic Review of Artificial Intelligence Techniques in the Detection and Classification of COVID-19 Medical Images in Terms of Evaluation and Benchmarking: Taxonomy Analysis, Challenges, Future Solutions and Methodological Aspects," *Journal of Infection and Public Health*, vol. 13, no. 10, pp. 1381-1396, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[5] Viswajeetsinh Jadeja et al., "Convolutional Neural Networks: An Extensive Review of Architectures and their Applications," *2023 6th International Conference on Contemporary Computing and Informatics*, Gautam Buddha Nagar, India, pp. 460-467, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Marta Villegas et al., "Predicting the Evolution of COVID-19 Mortality Risk: A Recurrent Neural Network Approach," *Computer Methods and Programs in Biomedicine Update*, vol. 3, pp. 1-9, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Feng Shi et al., "Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19," *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 4-15, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[8] Yao-Hung Hubert Tsai et al., "Multimodal Transformer for Unaligned Multimodal Language Sequences," *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Florence, Italy, pp. 6558-6569, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[9] Mahbubunnabi Tamal et al., "Artificial Intelligence (AI) Based Prediction of Mortality, ICU Admission and Ventilation Support Requirement for COVID-19 Patients Using 122 Clinical and Demographic Parameters," *medRxiv*, pp. 1-17, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Marwah Ahmed Halwani, and Manal Ahmed Halwani, "Prediction of COVID-19 Hospitalization and Mortality Using Artificial Intelligence," *Healthcare*, vol. 12, no. 17, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] Saeed Shakibfar et al., "Artificial Intelligence-Driven Prediction of COVID-19-Related Hospitalization and Death: A Systematic Review," *Frontiers in Public Health*, vol. 11, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Kagan Tur et al., "Multi-Modal Machine Learning Approach for COVID-19 Detection using Clinical Biomarkers and Chest X-Ray Images," *Diagnostics*, vol. 14, no. 24, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Albatoul S. Althenayan et al., "COVID-19 Hierarchical Classification Using a Deep Learning Multi-Modal," *Sensors*, vol. 24, no. 8, pp. 1-30, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Shih-Cheng Huang et al., "Fusion of Medical Imaging and Electronic Health Records using Deep Learning: A Systematic Review and Implementation Guidelines," *npj Digital Medicine*, vol. 3, no. 1, pp. 1-9, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Akshaya Karthikeyan et al., "Machine Learning based Clinical Decision Support System for Early COVID-19 Mortality Prediction," Frontiers in Public Health, vol. 9, pp. 1-13, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[16] Tami D. DenOtter, and Johanna Schubert, *Hounsfield Unit*, StatPearls, 2023. [Google Scholar]

[17] Joaquín Pérez-Ortega, Nelva Nely Almanza-Ortega, and David Romero, "Balancing Effort and Benefit of *K*-Means Clustering Algorithms in Big Data Realms," *PLOS One*, vol. 13, no. 9, pp. 1-19, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[18] Dinggang Shen, Guorong Wu, and Heung-Il Suk, "Deep Learning in Medical Image Analysis," *Annual Review of Biomedical Engineering*, vol. 19, pp. 221-248, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[19] Jason Brownlee, "Machine Learning Mastery with Python," *Machine Learning Mastery*, 2019. [Google Scholar]
[20] Davide Chicco, and Giuseppe Jurman, "The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binarys Classification Evaluation," *BMC Genomics*, vol. 21, no. 1, pp. 1-13, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[21] Jesse Davis, and Mark Goadrich, "The Relationship Between Precision-Recall and ROC Curves," *Proceedings of the 23rd International Conference on Machine Learning*, Association for Computing Machinery, New York, NY, United States, pp. 233-240, 2006. [CrossRef] [Google Scholar] [Publisher Link]
[22] Sebastian Raschka, "Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning," *arXiv preprint*, pp. 1-49, 2018. [CrossRef] [Google Scholar] [Publisher Link]
[23] Pushpa Singh et al., *Diagnosing of Disease using Machine Learning*, Machine Learning and the Internet of Medical Things in Healthcare, Academic Press, pp. 89-111, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[24] Tae Kyun Kim, "T Test as a Parametric Statistic," *Korean Journal of Anesthesiology*, vol. 68, no. 6, pp. 540-546, 2015. [CrossRef] [Google Scholar] [Publisher Link]