Multi-Dataset Multi-Task Learning for COVID-19 Prognosis

Filippo Ruffini[1], Lorenzo Tronchin[1], Zhuoru Wu[2], Wenting Chen[3], Paolo Soda[1], Linlin Shen[2], and Valerio Guarrasi[1]⋆

[1] Unit of Computer Systems & Bioinformatics, Department of Engineering, Università Campus Bio-Medico di Roma, Rome, Italy
[2] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
[3] Department of Electrical Engineering, City University of Hong Kong, China

Abstract. In the fight against the COVID-19 pandemic, leveraging artificial intelligence to predict disease outcomes from chest radiographic images represents a significant scientific aim. The challenge, however, lies in the scarcity of large, labeled datasets with compatible tasks for training deep learning models without leading to overfitting. Addressing this issue, we introduce a novel multi-dataset multi-task training framework that predicts COVID-19 prognostic outcomes from chest X-rays (CXR) by integrating correlated datasets from disparate sources, distant from conventional multi-task learning approaches, which rely on datasets with multiple and correlated labeling schemes. Our framework hypothesizes that assessing severity scores enhances the model's ability to classify prognostic severity groups, thereby improving its robustness and predictive power. The proposed architecture comprises a deep convolutional network that receives inputs from two publicly available CXR datasets, AIforCOVID for severity prognostic prediction and BRIXIA for severity score assessment, and branches into task-specific fully connected output networks. Moreover, we propose a multi-task loss function, incorporating an indicator function, to exploit multi-dataset integration. The effectiveness and robustness of the proposed approach are demonstrated through significant performance improvements in prognosis classification tasks across 18 different convolutional neural network backbones in different evaluation strategies. This improvement is evident over single-task baselines and standard transfer learning strategies, supported by extensive statistical analysis, showing great application potential.

Keywords: Chest X-rays · Deep Learning · CNN · Transfer Learning

1 Introduction

Numerous studies have applied deep learning techniques to COVID-19 for both diagnosis and prognosis. While most focus on image-based AI solutions for

---

⋆ Corresponding author: Valerio Guarrasi, E-mail: valerio.guarrasi@unicampus.it

COVID-19 diagnosis [28], prognosis prediction receives less attention [6]. Many prognostic studies employ single-task learning (STL) methods to predict outcomes like intensive care unit (ICU) admission or mortality [2, 21, 23, 24, 26]. Alternative approaches predict patient severity progression [14, 17, 25, 27], using datasets like AIforCOVID [30], to classify patients into mild and severe based on treatment outcomes. Some STL models assess COVID-19 pneumonia severity from CXR using structured severity scores [8, 9, 34], like the Brixia-score [29] that segments the lungs into six areas for severity grading [4], with these scores statistically linked to treatment outcomes, underlining their prognostic value [5].

Alternately, multi-task learning (MTL) trains models on multiple related tasks simultaneously to enhance reciprocal learning [7]. In healthcare, studies show that combining two relevant tasks can boost model performance, utilizing various task combinations, architectures for parameter sharing, or loss functions [33]. Specifically for COVID-19, MTL has been applied to diagnostic predictions and lesion segmentation on imaging data [1, 15], with some studies exploring the synergy between diagnosis and prognosis for a more integrated disease perspective [3, 31]. While MTL exhibits promising performance, its primary application remains within datasets having multiple labels per instance. Integrating different data sources for distinct tasks is challenging due to the need for universal annotations across datasets, requiring substantial efforts in data harmonization.

Research in MTL is evolving towards multi-dataset multi-task (MDMT) learning approaches, aiming to leverage diverse data sources to uncover indirect task connections beyond direct data instance links [16, 18, 19, 32]. This strategy, akin to transfer learning principles, seeks to enrich the learning process by integrating varied datasets. Nonetheless, much of this research does not specifically address medical imaging. For instance, in [32] it was implemented an MDMT framework on EEG datasets for action recognition with partially overlapping labels, while in [19] it was developed a MDMT framework for video-based action recognition, focusing on closely related activities.

Our MDMT learning model aims to predict COVID-19 severity by integrating two distinct datasets, AIforCOVID [30] and BRIXIA [29], each highlighting different facets of the same medical condition. BRIXIA focuses on radiological signs of the disease, showcasing the patient's condition shortly after diagnosis, whereas AIforCOVID forecasts future health status based on clinical evolution. By merging these tasks, we intend to show that combining severity assessments can provide deeper insights into patient outcome predictions, thereby enhancing severity classification accuracy. Furthermore, the use of datasets with different labeling schemes offers an innovative approach in healthcare, allowing for efficient multi-source data integration without the need for extensive relabeling, often required in traditional data merging methods. Our main contributions are as follows: i) we introduce a MDMT model for two publicly available COVID-19 CXR datasets, each assigned to a different task; ii) we implement a multi-task loss function incorporating an indicator function to facilitate multi-task optimization across data sources; iii) through an extensive evaluation using 18 CNN
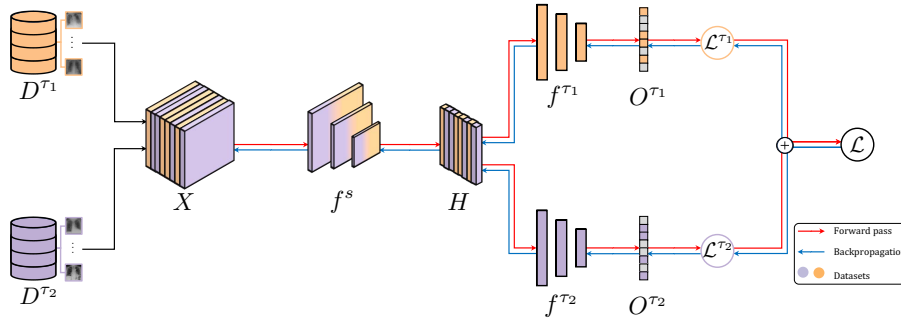
Fig. 1. Overview of the MDMT model architecture

architectures, we demonstrate significant performance gains compared to conventional transfer learning and STL approaches.

2 Methods

In the context of the proposed MDMT model, we propose a novel framework designed to leverage shared and task-specific features across multiple datasets, thereby enhancing the model's ability to generalize and perform distinct tasks simultaneously. This approach not only capitalizes on the inherent diversity and complementarity of multi-source data, but also introduces a flexible architecture that adapts to the unique requirements of each task through dedicated heads and a unified backbone for feature extraction. In Figure 1, we report an overview of the proposed MDMT model.

Given two datasets, D^{τ_1} and D^{τ_2}, each referring to a task τ_1 and τ_2, respectively, a single batch X is formed by randomly selecting elements such that $X_i \in D^{\tau_1} \cup D^{\tau_2}$ for any element X_i in X. A shared backbone feature extraction network, denoted as f^s, processes X to yield a set of shared features H. This shared representation is then fed into two task-specific fully connected network heads, f^{τ_1} and f^{τ_2}, for tasks τ_1 and τ_2, respectively, producing outputs O^{τ_1} and O^{τ_2}. The model employs an indicator function $\mathcal{I}(X_i, \tau_j)$ to determine the association of a sample X_i from batch X to a task τ_j, crucial for task-specific loss computations. The losses for the tasks τ_1 and τ_2, represented as $\mathcal{L}^{\tau_1}$ and $\mathcal{L}^{\tau_2}$, respectively, are computed using O^{τ_1}, Y^{τ_1} and O^{τ_2}, Y^{τ_2}, where Y^{τ_1} and Y^{τ_2} denote the true labels of X for the respective tasks.

In the following, we delve deeper into the specifics of each component within the network architecture, encompassing the shared backbone, the task-specific heads and the MDMT learning mechanism.

In our model, the shared backbone consists of CNN layers that act as a common trainable module for the flow of information regardless of the tasks. The output of this shared network is represented by the function $H = f^s(X; \theta^s)$, where X denotes the input batch from the combined datasets D^{τ_1} and D^{τ_2}, and θ^s symbolizes the trainable parameters of the CNN backbone. This unified

feature extraction mechanism is crucial for capturing generalized representations that are beneficial across multiple tasks.

In our architecture, we enhance the shared CNN backbone by appending two task-specific heads of fully connected layers, which identify distinct task-specific modules for D^{τ_1} and D^{τ_2} and their respective tasks. The output vector for each task-specific branch is defined as:

$$O^{\tau_j} = f^{\tau_j}(H, \theta^{\tau_j}), \quad j \in \{1, 2\} \tag{1}$$

Where H is the output from the shared backbone and θ^{τ_j} are the parameters of the task-specific fully connected networks, ensuring a dedicated head for processing the features according to the specific requirements of each task.

The uniqueness of our approach lies in the integration of the two loss functions $\mathcal{L}^{\tau_1}$ and $\mathcal{L}^{\tau_2}$ through the adoption of an indicator function, inspired by Lee et al.'s approach to handling censored data in overall survival analysis [22]. This adaptation is crucial given the context where, within a mixed batch, no samples bear labels for both tasks concurrently, making it impractical to directly apply a loss function across different datasets. Hence, the proposed total loss function, $\mathcal{L}$, is crafted as:

$$\mathcal{L} = \sum_{i=1}^{|X|} \mathcal{I}(X_i, \tau_1) \cdot \mathcal{L}^{\tau_1}(O_i^{\tau_1}, Y_i^{\tau_1}) + \mathcal{I}(X_i, \tau_2) \cdot \mathcal{L}^{\tau_2}(O_i^{\tau_2}, Y_i^{\tau_2}) \tag{2}$$

where $|X|$ is the number of instances in X.

The indicator function $\mathcal{I}(X_i, \tau_j)$, plays a pivotal role in the MDMT model by allowing the integration of multi-source data into a unified framework, especially when computing task-specific losses. This function is designed to identify the association of a given sample with its specific task, which is crucial for models that handle data from multiple datasets to perform multiple task optimization simultaneously. It can be defined as:

$$\mathcal{I}(X_i, \tau_j) = \begin{cases} 1 & \text{if sample } X_i \text{ is associated with task } \tau_j \\ 0 & \text{otherwise.} \end{cases}$$

Since the batch X contains elements from both datasets D^{τ_1} and D^{τ_2}, and each dataset may be associated with different tasks, $\mathcal{I}(X_i, \tau_j)$ helps to selectively compute the loss for each task by including only those samples related to the task in question. The use of an indicator function in this manner adds flexibility and scalability to the model. It allows for the seamless addition of new tasks or datasets by simply adjusting the indicator function to include new elements in the loss computation process without needing significant architectural changes. In the proposed MDMT model, the training process is designed to be end-to-end, indicating that all model components are trained simultaneously during the training phase. This approach ensures that the model learns to optimize its parameters in a cohesive manner, leveraging gradients derived from the total loss to update each component of the architecture.

For the two task-specific loss functions, we selected $\mathcal{L}^{\tau_1}$ and $\mathcal{L}^{\tau_2}$ considering each task characteristics and labeling schemes, as detailed in Section 3. Given that τ_1, for the severity group prediction, and τ_2, for the severity assessment, are both classification tasks, we selected the Cross-Entropy Loss, defined as follows:

$$\mathcal{L}^{\tau_j}(O_i^{\tau_j}, Y_i^{\tau_j}) = -\sum_{k=1}^{c^{\tau_j}} Y_{i,k}^{\tau_j} \log(O_{i,k}^{\tau_j}) \tag{3}$$

where $Y_i^{\tau_j}$ is the real label, $O_i^{\tau_j}$ the predicted probabilities for the sample X_i and c^{τ_j} is the number of class labels for the considered task τ_j. These loss functions are tailored to the distinct characteristics of each task within the model, allowing for precise adjustments to the model parameters based on the nature of the tasks at hand. The ability to compute task-specific losses accurately is crucial for the model to balance learning across tasks, preventing the dominance of any single-task over the others and promoting a synergistic improvement in performance.

3 Experimental Setup

This section presents our evaluation framework for the MDMT model against conventional approaches, detailing the datasets, preprocessing steps and experimental configurations.

3.1 Datasets and Tasks

Our study leverages two distinct and publicly available CXR datasets, namely AIforCOVID [30] and BRIXIA [29], each assigned to a distinct task: severity prognosis and severity assessment, respectively. These datasets were specifically chosen for their complementary characteristics, enabling a comprehensive assessment of the model's ability to generalize across different tasks and data distributions.

The AIforCOVID dataset D^{τ_1} encompasses 1586 CXR examinations from COVID-19 positive adult patients, confirmed via RT-PCR tests, from six different centers. This dataset has been compiled into three releases, each expanding the dataset's size: the initial release included 820 patients, followed by an additional 284 patients, and concluded with an additional 482 patients. Patients are classified into two categories based on treatment outcomes: mild (home isolation or hospitalization without ventilatory support) and severe (requiring non-invasive ventilation, ICU admission, or resulting in death), making the task a binary classification problem, with the target variable $Y_i^{\tau_1} \in \{0, 1\}$ representing these outcomes.

The BRIXIA dataset D^{τ_2} consists of 4707 CXR images from COVID-19 positive patients treated in sub-intensive and intensive care units, obtained from various machine manufacturers and varied acquisition parameters. A noteworthy feature of this dataset is the Brixia score, which grades lung opacity across six regions on a scale of 0 to 3. To simplify the scoring system, we sum the regional

scores into a global score G_i for each image X_i. This score is then categorized into four severity levels through a threshold function:

$$Y_i^{\tau_2} = \begin{cases} G_i < 5, & \text{Category 0} \\ 5 \leq G_i < 9, & \text{Category 1} \\ 9 \leq G_i < 14, & \text{Category 2} \\ G_i \geq 14, & \text{Category 3} \end{cases} \tag{4}$$

This categorization yields a multi-class target variable $Y_i^{\tau_2} \in \{0, 1, 2, 3\}$, facilitating the model's task of predicting the severity of lung involvement in COVID-19 patients.

To ensure the data from both AIforCOVID and BRIXIA datasets are coherent and prepared for our model, we employed a preprocessing pipeline that includes lung segmentation via a U-NET model, bounding-box extraction, resizing and normalization. For ease of reproducibility, these steps are the same used in [10, 12–14].

3.2   Experimental Configurations

To comprehensively validate the efficacy of our MDMT model, we designed three distinct experimental configurations, aimed at demonstrating the model's superiority over traditional STL and fine-tuning approaches.

We establish a baseline by training the CNN backbone on each dataset and task separately. This step serves as a foundation for comparison, highlighting the potential enhancements brought by MTL. The experiments are denoted as STL^{τ_1} and STL^{τ_2} for D^{τ_1} and D^{τ_2} datasets, respectively. Each backbone utilized in this phase is pre-trained on the ImageNet dataset [20].

As a comparative benchmark, we explore a conventional fine-tuning strategy where the model pre-trained on the BRIXIA dataset STL^{τ_2} is fine-tuned for the AIforCOVID task D^{τ_1}. This experiment, denoted as FT, investigates the effectiveness of transferring learned features from one task to another, contrasting it against the integrated learning approach of the MDMT model.

The proposed MDMT Learning model employs as pre-training the average of the weights from the baseline models STL^{τ_1} and STL^{τ_2}. This experiment, denoted $MDMT$, allows simultaneous training on both tasks, leveraging shared and task-specific features, as detailed in Section 2.

To evaluate the robustness of the proposed methodology, all the aforementioned experiments are conducted on 18 CNN backbones across 6 major architectural families: DenseNet, EfficientNet, GoogLeNet, MobileNet, ResNet and ShuffleNet. For all experiments and networks, uniform training configurations were applied. Optimization was carried out using the Adam optimizer, with an initial learning rate of 0.001, momentum of 0.9, and weight decay of 0.0001. Training included a 40-epoch warm-up period and was limited to a maximum of 300 epochs, incorporating early stopping after 40 epochs to prevent overfitting. All training sessions were conducted on four NVIDIA TESLA A100 GPUs, utilizing a batch size of 128.

Experiment	CV			LOCO		
	ACC	F1	GM	ACC	F1	GM
STL^{τ_1}	66.0	63.9	65.9	61.2	57.3	60.1
FT	67.3	63.1	66.9	64.8	58.9	63.8
$MDMT$	**68.6**	**66.6**	**68.5**	**65.7**	**64.3**	**66.0**

Table 1. Average model performance metrics for task τ_1, calculated across different backbone architectures.

Statistic	Test	CV			LOCO		
		ACC	F1	GM	ACC	F1	GM
μ	$MDMT$ vs STL^{τ_1}	∗ ∗ ∗	∗ ∗ ∗	∗ ∗ ∗	∗ ∗ ∗	∗ ∗ ∗	∗ ∗ ∗
	$MDMT$ vs FT	∗∗	∗ ∗ ∗	∗∗			
σ	$MDMT$ vs STL^{τ_1}	∗	∗	∗	∗		∗
	$MDMT$ vs FT				∗		∗

Table 2. Statistical comparison of $MDMT$ vs STL^{τ_1} and $MDMT$ vs FT. Significance levels are marked with ∗ for a p-value < 0.05, ∗∗ for a p-value < 0.01 and ∗ ∗ ∗ for a p-value < 0.001.

To ensure the reliability and generalizability of our findings, we employ both 5-fold stratified cross-validation (CV) and leave-one-center-out (LOCO) validation strategies. By evaluating the Accuracy (ACC), F1-score (F1) and G-mean (GM), we perform a comprehensive evaluation of the model's performance across different partitions of the data, highlighting its ability to generalize to unseen data and mitigate bias towards specific data distributions.

4   Results and Discussions

To evaluate the effectiveness of our proposed MDMT method, we compared our MDMT framework with each experimental configuration, validation strategy and CNN backbone employed, as shown in Table 1 and in more detail in Table S1. We opted to discuss the results obtained on τ_1 task, given its unique relationship with the task τ_2. The task τ_2 not only temporally precedes τ_1 but also possesses strong predictive power for prognostic prediction, therefore incorporating τ_2 can lead to performance improvement on τ_1.

To provide a comprehensive overview of our results, Table 1 displays the average performance metrics across all experimental configurations and CNN architectures tested in our study. Our analysis reveals a clear superiority of the MDMT approach over the conventional methodologies, including both the STL and fine-tuning strategies. In Table S1, all the performances for all the shared backbones are reported. These results not only confirm MDMT's superior average performance, but also demonstrate that our method achieves the highest absolute performance over backbone architectures, experimental scenarios and validation strategies.

To rigorously evaluate the effectiveness of the proposed MDMT method, we employed a single-tail t-test for statistical comparisons across the various

experimental configurations. This analysis involved assessing both the average performance metrics μ and their standard deviations σ across test folds. Our goal was to determine whether there was a statistically significant increase in μ and a decrease in σ for $MDMT$ in comparison to the STL^{τ_1} and the FT settings. Notably, an increase in μ indicates a significant increase in performance and a reduction of σ indicates enhanced model robustness. The findings, detailed in Table 2, show the statistical significance achieved by MDMT.

Analyzing the results in CV, we observed that the $MDMT$ method consistently outperformed the STL^{τ_1} and the FT approaches in terms of average performance metrics across the CNN architectures, further validated through statistical comparisons. When compared to STL^{τ_1}, $MDMT$ not only demonstrates superior performance across all metrics but also shows a reduction in standard deviation, indicating both higher performance and increased robustness. In contrast, while $MDMT$ exhibits statistically significant performance enhancements over FT, the improvement in robustness is not statistically significant. To clarify, we conducted additional tests on σ, confirming that $MDMT$'s robustness is statistically comparable to that of the fine-tuning strategies.

In the LOCO validation scenario, $MDMT$ demonstrated superior performance relative to both the STL^{τ_1} and the FT methods. This performance superiority underscores the robustness of $MDMT$ across various backbone architectures and its enhanced ability to generalize across different clinical settings. The statistical analysis highlights a significant improvement in performance metrics for $MDMT$ compared to STL^{τ_1}. Regarding the standard deviation, the reduction in variability for ACC and F1 metrics suggests increased reliability of $MDMT$ over the compared methods. To address any potential concerns about non-significant differences, we conducted one-tail t-tests, which confirmed that $MDMT$'s performance is not statistically inferior to that of the FT approach.

For brevity, we presented results only from the final release, though all experiments were conducted across all three releases of the AIforCOVID dataset, yielding consistent results and conclusions. Furthermore, all experiments and analyses were similarly conducted on τ_2, showing no significant change in performance with respect to the STL^{τ_2} and the corresponding FT, underscoring that while τ_2 benefits τ_1, the converse is not necessarily true.

5   Conclusion

In this study, we introduced a MDMT approach aimed at enhancing the generalization capabilities of deep neural networks for COVID-19 prognosis prediction. Our approach demonstrates thorough an extensive analysis that performance improvements are achievable by extracting prognostic information from multiple datasets. Our method not only showcases higher performance across various experimental benchmarks but also establishes itself as a robust solution within the biomedical field for the integration of diverse datasets for different tasks into a unified training framework, leading to a multi-perspective analysis of the disease.

However, the application of this approach is limited by the need to select datasets and tasks that follow a well-determined logic, ensuring that the different prognostic labelling schemas correlate with each other. In future work, we plan to develop a model capable of processing prognostic information from a broader range of heterogeneous datasets, together with the adoption of explainable AI techniques [11] to understand the importance of each task and the integration of multimodal data aimed to achieve a holistic understanding of the disease's aspects, all with the aim to achieve more precise, trustworthy and complete systems.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Amyar, A., et al.: Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation. Computers in biology and medicine **126**, 104037 (2020)
2. Bae, J., , et al.: Predicting mechanical ventilation and mortality in COVID-19 using radiomics and deep learning on chest radiographs: a multi-institutional study. Diagnostics **11**(10), 1812 (2021)
3. Bao, G., et al.: COVID-MTL: Multitask learning with Shift3D and random-weighted loss for COVID-19 diagnosis and severity assessment. Pattern Recognition **124**, 108499 (2022)
4. Borghesi, A., et al.: COVID-19 outbreak in Italy: experimental chest X-ray scoring system for quantifying and monitoring disease progression. La radiologia medica **125**(5), 509–513 (2020)
5. Borghesi, A., et al.: Chest X-ray versus chest computed tomography for outcome prediction in hospitalized patients with COVID-19. La radiologia medica **127**(3), 305–308 (2022)
6. Buttia, C., et al.: Prognostic models in COVID-19 infection that predict severity: a systematic review. European journal of epidemiology **38**(4), 355–372 (2023)
7. Caruana, R.: Multitask learning. Machine learning **28**, 41–75 (1997)
8. Cohen, J.P., et al.: Predicting covid-19 pneumonia severity on chest x-ray with deep learning. Cureus **12**(7) (2020)
9. Danilov, V.V., et al.: Automatic scoring of COVID-19 severity in X-ray imaging based on a novel deep learning workflow. Scientific reports **12**(1), 12791 (2022)

10. Guarrasi, V., et al.: A multi-expert system to detect COVID-19 cases in X-ray images. In: 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS). pp. 395–400. IEEE (2021)

11. Guarrasi, V., et al.: Multimodal explainability via latent shift applied to COVID-19 stratification. arXiv preprint arXiv:2212.14084 (2022)

12. Guarrasi, V., et al.: Optimized fusion of CNNs to diagnose pulmonary diseases on chest X-Rays. In: International Conference on Image Analysis and Processing. pp. 197–209. Springer (2022)

13. Guarrasi, V., et al.: Pareto optimization of deep networks for COVID-19 diagnosis from chest X-rays. Pattern Recognition **121**, 108242 (2022)

14. Guarrasi, V., et al.: Multi-objective optimization determines when, which and how to fuse deep networks: An application to predict COVID-19 outcomes. Computers in Biology and Medicine **154**, 106625 (2023)

15. He, K., et al.: Synergistic learning of lung lobe segmentation and hierarchical multi-instance classification for automated severity assessment of COVID-19 in CT images. Pattern recognition **113**, 107828 (2021)

16. Hosseini, S., et al.: Distill-2MD-MTL: Data distillation based on multi-dataset multi-domain multi-task frame work to solve face related tasksks, multi task learning, semi-supervised learning. arXiv preprint arXiv:1907.03402 (2019)

17. Jiao, Z., et al.: Prognostication of patients with COVID-19 using artificial intelligence based on chest x-rays and clinical data: a retrospective study. The Lancet Digital Health **3**(5), e286–e294 (2021)

18. Kaiser, L., et al.: One model to learn them all. arXiv preprint arXiv:1706.05137 (2017)

19. Kapidis, G., et al.: Multi-dataset, multitask learning of egocentric vision tasks. IEEE Transactions on Pattern Analysis and Machine Intelligence (2021)

20. Ke, A., et al.: CheXtransfer: performance and parameter efficiency of ImageNet models for chest X-Ray interpretation. In: Proceedings of the conference on health, inference, and learning. pp. 116–124 (2021)

21. Kulkarni, A.R., et al.: Deep learning model to predict the need for mechanical ventilation using chest X-ray images in hospitalised patients with COVID-19. BMJ innovations pp. bmjinnov–2020 (2021)

22. Lee, C., et al.: Deephit: A deep learning approach to survival analysis with competing risks. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)

23. Lee, J.H., et al.: Development and Validation of a Multimodal-Based Prognosis and Intervention Prediction Model for COVID-19 Patients in a Multicenter Cohort. Sensors **22**(13), 5007 (2022)

24. Li, H., et al.: Predicting intensive care need for COVID-19 patients using deep learning on chest radiography. Journal of Medical Imaging **10**(4), 044504 (2023)

25. Li, Z., et al.: A multistage multimodal deep learning model for disease severity assessment and early warnings of high-risk patients of COVID-19. Frontiers in public health **10**, 982289 (2022)

26. Rahman, T., et al.: BIO-CXRNET: a robust multimodal stacking machine learning technique for mortality risk prediction of COVID-19 patients using chest X-ray images and clinical data. Neural Computing and Applications pp. 1–23 (2023)

27. Schöning, V., et al.: Development and validation of a prognostic COVID-19 severity assessment (COSA) score and machine learning models for patient triage at a tertiary hospital. Journal of translational medicine **19**(1), 1–11 (2021)

28. Shi, F., et al.: Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. IEEE reviews in biomedical engineering **14**, 4–15 (2020)
29. Signoroni, A., et al.: BS-Net: Learning COVID-19 pneumonia severity on a large chest X-ray dataset. Medical Image Analysis **71**, 102046 (2021)
30. Soda, P., et al.: AIforCOVID: Predicting the clinical outcomes in patients with COVID-19 applying AI to chest-X-rays. An Italian multicentre study. Medical image analysis **74**, 102216 (2021)
31. Wang, S., et al.: A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. European Respiratory Journal **56**(2) (2020)
32. Xie, Y., et al.: Cross-dataset transfer learning for motor imagery signal classification via multi-task learning and pre-training. Journal of Neural Engineering **20**(5), 056037 (2023)
33. Zhao, Y., et al.: Multi-task deep learning for medical image computing and analysis: A review. Computers in Biology and Medicine **153**, 106496 (2023)
34. Zhu, J., et al.: Deep transfer learning artificial intelligence accurately stages COVID-19 lung disease severity on portable chest radiographs. PloS one **15**(7), e0236621 (2020)