Predicting mortality in SARS-COV-2 (COVID-19) positive patients in the inpatient setting using a Novel Deep Neural Network.

Maleeha Naseem[1], Hajra Arshad[2], Syeda Amrah Hashimi[2], Furqan Irfan[3], Fahad Shabbir Ahmed[4, 5, 6]

[1] Department of Community Health Sciences, Aga Khan University, Karachi, Pakistan 74900.

[2] Medical College, Aga Khan University, Karachi, Pakistan 74900.

[3] College of Osteopathic Medicine, Michigan State University, East Lansing, MI 48824.

[4] Clinicaro Machine Learning Group, New Haven, CT, 06510.

[5] Department of Pathology, Wayne State University, Detroit, MI 48201.

[6] Corresponding author

Email Address: Fahad Shabbir Ahmed, fahadshabbirahmed@gmail.com

Corresponding author: Fahad Shabbir Ahmed, Lead Scientist, Clinicaro Machine Learning group; Department of Pathology, Wayne State University / Detroit Medical Center. Harper Professional Building, 4160 John R St, Detroit, MI 48201. Phone: 631-644-3981; Email; fahadshabbirahmed@gmail.com.

Conflict of Interest: None.

Disclosures: None

Abstract words count: 389

Manuscript word count: 2870

ABSTRACT

Background

The second wave of COVID-19 pandemic is anticipated to be worse than the initial one and will strain the healthcare systems even more during the winter months. Our aim was to develop a machine learning-based model to predict mortality using the deep learning Neo-V framework. We hypothesized this novel machine learning approach could be applied to COVID-19 patients to predict mortality successfully with high accuracy.

Methods

The current Deep-Neo-V model is built on our previously statistically rigorous machine learning framework [Fahad-Liaqat-Ahmad Intensive Machine (FLAIM) framework] that evaluated statistically significant risk factors, generated new combined variables and then supply these risk factors to deep neural network to predict mortality in RT-PCR positive COVID-19 patients in the inpatient setting. We analyzed adult patients (≥18 years) admitted to the Aga Khan University Hospital, Pakistan with a working diagnosis of COVID-19 infection (n=1228). We excluded patients that were negative on COVID-19 on RT-PCR, had incomplete or missing health records. The first phase selection of risk factor was done using Cox-regression univariate and multivariate analyses. In the second phase, we generated new variables and tested those statistically significant for mortality and in the third and final phase we applied deep neural networks and other traditional machine learning models like Decision Tree Model, k-nearest neighbor models and others.

Results

A total of 1228 cases were diagnosed as COVID-19 infection, we excluded 14 patients after the exclusion criteria and (n=)1214 patients were analyzed. We observed that several clinical and laboratory-based variables were statistically significant for both univariate and multivariate analyses while others were not. With most significant being septic shock (hazard

ratio [HR], 4.30; 95% confidence interval [CI], 2.91-6.37), supportive treatment (HR, 3.51; 95% CI, 2.01-6.14), abnormal international normalized ratio (INR) (HR, 3.24; 95% CI, 2.28-4.63), admission to the intensive care unit (ICU) (HR, 3.24; 95% CI, 2.22-4.74), treatment with invasive ventilation (HR, 3.21; 95% CI, 2.15-4.79) and laboratory lymphocytic derangement (HR, 2.79; 95% CI, 1.6-4.86). Machine learning results showed our DNN (Neo-V) model outperformed all conventional machine learning models with test set accuracy of 99.53%, sensitivity of 89.87%, and specificity of 95.63%; positive predictive value, 50.00%; negative predictive value, 91.05%; and area under the curve of the receiver-operator curve of 88.5.

Conclusion

Our novel Deep-Neo-V model outperformed all other machine learning models. The model is easy to implement, user friendly and with high accuracy.

Keywords: COVID-19, deep learning, mortality prediction, Neo-V Framework

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-COV 2) has caused 60 million infections and 1.4 million deaths worldwide (1) and 800, 000 deaths in the United States (2) . Despite strict measures being deployed and special instructions given to the mass public, the second wave is anticipated to be far worse than the first one (3). Some countries have started to ease those earlier restrictions because of economic implications from the initial lockdown, which may create a further deepening of the current crisis, as cases continue to rise. This could overwhelm the already strained healthcare systems across the United States and the world.

Machine learning has been extensively used in the automotive, defense and fin-tech industry over the past couple of years with great success. The use of these systems to predict health outcomes have been limited. Epidemic Renormalization Group (eRG) has used a machine learning framework to predict the time evolution of the first and second wave based on the data from the first wave in Europe (4). We have previously developed algorithms that predict mortality in the clinical setting and performed better than most clinical scales utilized currently to predict mortality (5-7) During the current COVID-19 pandemic crisis, the aim was to develop a mortality prediction tool that can predict death in COVID-19 patients at admission. This would help the already strained healthcare systems and physicians around the world in crucial clinical decision making, resource management and family-counselling. We hypothesize that machine learning, specifically deep-learning could be applied to COVID-19 patients with high accuracy. Using deep-learning to predict mortality in these patients may assist in clinical decision making, risk stratification and planning strategies in future for such pandemics at a larger scale. Not much work has been done in mortality prediction in COVID-19 patients in with lower socio-economic countries (LMIC) or developed countries. Our hypothesis is that machine learning could be applied to COVID-19 patients with high

accuracy. Hence, predicting mortality and clinical outcomes using ML algorithms may assist in clinical decision making, risk stratification and planning strategies in future for such pandemics at a larger scale.

METHODS

Clinical setting and dataset

Clinical data was conducted at the Aga Khan University Hospital (AKUH). All patients' records were completely anonymous, and the data collected has received Institutional Review Board/Ethical Review Committee (IRB/ERC) approval from Aga Khan University Hospital (AKUH), Pakistan. The dataset was de-identified and our study complied with the ethical principles recommended by Helsinki declaration (1964) and its amendments. We retrospectively collected data from AKUH – electronic medical record (EMR) that were admitted with a primary diagnosis of COVID-19 infection to the hospital between February 2020 and September 2020.

Patient data collection and selection criteria.

We included adult patients (>18 years of age) that were admitted to the hospital with a diagnosis of COVID-19 or were tested positive during their admission on reverse transcriptase polymerase chain reaction (RT-PCR) based on Center for Disease Control and Prevention (CDC) and College of American Pathologist (CAP) guidelines (8, 9). Data was collected on demographics, and comorbidities at admission, the first 24-hours of laboratory investigations (hematological and blood biochemistry, (**Table 1.**), imaging and complete clinical characteristics (history, examination, treatment, hospital course and outcomes). We excluded all patients that had RT-PCR negative tests for COVID-19 and incomplete records or inaccurate medical record information.

The Neo-V Framework

Neo-V is a tri-phase bio-statically rigorous machine learning that builds on our previous framework that had better accuracy then the currently used clinical scoring systems in predicting mortality in the intensive care unit (ICU) patients(5, 6).

Phase I: Also known as the statistical-phase; in which data was analyzed by univariate and multivariate Cox-regression analysis (X) using IBM SPSS (version 24.0.0.0) (X) for outcome assessment with hazard ratio and confidence intervals. A *p*-value of <0.05 was considered statistically significant. We also did demographic data frequency analysis. Statistical analysis was carried out on all the variables included in **Table 1.**

Phase II: In contrast to our previously published model we created new variables for the existing dataset called neo-variables. These variables included a combination of two clinically relevant labs that were significant in both the univariate and the multivariate analysis **(Table 1.)**. These variables also underwent univariate and multivariate analysis for outcome assessment with hazard ratio and confidence intervals

Phase III: Biological datasets are highly imbalanced with respect to the outcomes (i.e. more people were survived, then those who didn't) and machine learning models are very sensitive to imbalanced data and can produce variable and non-reproducible results. To address this, we optimized the dataset using Synthetic Minority Over-sampling Technique (SMOTE) algorithm during the training process (10). In the machine learning phase, we used all variables that were statistically significant in phase I and II in both the univariate and multivariate analysis (non-significant risk factors were excluded). The final dataset was randomized and divided into a training and testing set with a 70/30 percent split respectively (30% data left out to test the models). After partitioning the data, we allocated feature vectors of the training instances by X_train with corresponding outcome label as Y_train. Similarly,

for the test set we allocated X_test and Y_test as testing vector instances and corresponding outcomes, respectively. The models trained on X_train and Y_train. The models tried to learn the behavior/distribution of the data and generate a hypothesis/fitting function. Once the training is concluded the model will then test the X_test and produce an output (prediction) called Y_pred. A comparison is done between Y_pred and Y_test. We had previously discussed that reduction of the number of irrelevant risk factors can produce better performances and significantly improve classifications. In this study we used conventional machine learning models including Random Trees (CART), K-Nearest Neighbor (KNN), Support Vector Classifier - Radial Basis Function (SVC - RBF), Ada-Boost-Classifier (ABC) and Quadratic Discriminant Analysis (QDA) and a deep neural network (DNN).

Deep-FLAIM

FLAIM framework only has phase I and III and we used it to compare it with Neo-V Framework. The Deep-FLAIM model is a 4 layered model and details have been reported previously(5).

Performance, Primary and Secondary Outcomes Analyses.

Performance of all models was evaluated by comparing their accuracies and area under the receiver-operator curves (AUROC). Primary outcomes included sensitivity and specificity, while secondary outcomes included positive predictive values (PPV) and negative predictive values (NPV).

RESULTS

From a total of 1228 patients we selected 1214 patients that were adult patients with complete data and RT-PCR proven COVID-19 infections. Demographics of this population showed a

median age of 55 years (range 19-96 years), around 28% (n=336) of the admitted population being elderly (>65 years of age). Median length of stay 5 days (range 1-54 days), most patients admissions were male (61.5%).

The clinical characteristics of these patients included hypertension (43%, n=521) and diabetes (36%, n=441) being the most common comorbid. Presenting symptoms ranged widely from being asymptomatic to shortness. The most significant clinical risk factors for death during the hospital admission included chronic kidney disease (n=71, HR=2.74, 95%CI=1.63-4.61), ischemic heart disease (n=158, HR=1.80, 95%CI=1.19-2.73), other comorbidities (n=571, HR=1.72, 95%CI=1.21-2.45), shortness of breath (n=570, HR=1.96, 95%CI=1.35-2.86), other symptoms (non-respiratory and non-gastrointestinal symptoms, n=321, HR=1.66, 95%CI=1.16-2.38), acute kidney injury (n=179, HR=2.47, 95%CI=1.74-3.51), acute respiratory distress syndrome / respiratory failure (n=147, HR=2.53, 95%CI=1.77-3.63), septic shock (n=49, HR=4.30, 95%CI=2.91-6.37), intensive care unit (ICU) admission (n=106, HR=3.24, 95%CI=2.22-4.74), number of ICU admissions during current hospitalization (range=0-2, HR=2.01, 95%CI=1.49-2.87), invasive ventilation (n=74, HR=3.21, 95%CI=2.15-4.79), non-invasive ventilation (BiPAP/CPAP) (n=243, HR=1.87, 95%CI=1.32-2.67), supportive treatment (n=617, HR=1.87, 95%CI=1.32-2.67) and blood group AB+ (n=48, HR=2.48, 95%CI=1.26-4.93); while the utilization of systemic steroids (n=430, HR=0.66, 95%CI=0.45-0.97) and presence of fever (n=768, HR=0.57, 95%CI=0.40-0.81) were associated with better overall all survival. The most significant (statistical association with mortality) laboratory abnormalities for these patients included white cell count (n=455, HR=2.09, 95%CI=1.47-2.59), lymphocyte counts(n=378, HR=2.79, 95%CI=1.60-4.86), neutrophil count (n=1044, HR=2.67, 95%CI=1.24-5.70), platelets count (n=363, HR=1.41, 95%CI=1.00-2.01), hematocrit (n=505, HR=1.81 , 95%CI=1.26-2.61), hemoglobin (n=879, HR=1.63, 95%CI=1.05-2.55), blood urea nitrogen (BUN, n=382,

HR=2.78, 95%CI=1.94-4.01), creatinine (n=373, HR=2.73, 95%CI=1.90-3.94), international normalized ratio (INR, n=198, HR=3.24, 95%CI=2.28-4.63), prothrombin time (PT, n=378, HR=1.79, 95%CI=1.24-2.59). There were some risk factors that were significant for mortality but had too few patients including; shock liver (n=3, HR=11.47, 95%CI=2.77-47.68), blood group AB- (n=6, HR=7.63, 95%CI=2.39-24.48), Rhinorrhea (n=7, HR=0.001 , 95%CI=0.01-1.7e200), treatment with intravenous IgG (n=6, HR=3.95, 95%CI=1.25-12.49) and pneumothorax (clinical or radiological, n=22, HR=2.01 , 95%CI=1.18-3.46) . **Table 1.** shows results from univariate and multivariate analysis (hazard ratios, confidence intervals and *p*-values) of all the clinical and laboratory data.

In phase II (new variable phase) we created 11 new variables that included; total number of comorbidities (range 0-6, HR=1.30, 95%CI=1.14-1.49) , more than 2 comorbidities (n=499, HR=1.79, 95%CI=1.23-2.61), total number of symptoms(range 0-6, HR=1.07, 95%CI=0.91-1.26), total types of treatments received (range 0-6, HR=1.27, 95%CI=1.11-1.46), abnormal labs: creatinine or blood urea nitrogen (CR or BUN, (n=517, HR=0.36, 95%CI=0.36-0.89)), hemoglobin or hematocrit (HB or HCT, (n=883, HR=0.56, 95%CI=0.36-0.89)), platelets or INR (PLT or INR, (n=460, HR=2.09, 95%CI=1.47-2.98)), PT or INR (PT or INR, (n=502, HR=0.414, 95%CI=)), total leukocyte counts of lymphocytes (TLC or LYMP, (n=0.19, HR=, 95%CI=0.09-0.43)), total leukocyte counts or neutrophils (TLC or NEU, (n=1085, HR=0.18, 95%CI=0.06-0.57)), total number of abnormal labs(range 0-17, HR=1.22, 95%CI=1.15-1.30), univariate and multivariate statistical analysis can be seen in **Table 2**. The new variables in phase II referred to as ‘neo-variables’ were statistically significant in the uni and the multivariate analysis except the total number of symptoms.

The performance of our previously designed Deep-FLAIM model was compared to the Neo-V framework (including Deep-Neo-V and other conventional machine learning algorithms) see **Table 3.** Performance results show Deep-FLAIM (training accuracy = 86.7%, testing

accuracy = 84.7%, sensitivity = 68.9, specificity = 86.9, PPV = 42.5, NPV = 95.2, FPR =13.1 and AUROC = 86.9); while conventional machine learning show: Random Forest (RF, training accuracy = 94.9%, testing accuracy = 85.8%, sensitivity = 28.9, specificity = 93.4, PPV = 32.4, NPV = 90.4, FPR = 6.25 and AUROC = 69.5), k- Nearest Neighbors (k-NN, training accuracy = 92.3%, testing accuracy = 77.8%, sensitivity = 42.2, specificity = 82.8, PPV = 25.7, NPV = 91.1, FPR = 17.2 and AUROC = 66.5), Simple Vector Classifier – Radial Basis Function (SVC – RBF, training accuracy = 69.0%, testing accuracy = 65.8%, sensitivity = 75.6, specificity = 64.4, PPV = 23.0, NPV = 94.9, FPR = 35.6 and AUROC = 80.8), Decision Trees (DT, training accuracy = 98.9%, testing accuracy = 85.2%, sensitivity = 42.2, specificity = 91.3, PPV = 40.4, NPV = 91.8, FPR = 8.8 and AUROC = 66.7), Adaptive Boosted Classifier (ABC, training accuracy = 90.1%, testing accuracy = 85.5%, sensitivity = 48.9, specificity = 90.6, PPV = 42.3, NPV = 82.7, FPR = 9.4 and AUROC = 77.65) and Quadratic discriminant analysis (QDA, training accuracy = 94.6%, testing accuracy = 87.4%, sensitivity = 60.0, specificity = 91.3, PPV = 49.1, NPV = 94.2, FPR = 8.74 and AUROC = 80.84). While our best model was a deep neural network (Deep-Neo-V) with training accuracy = 98.7%, testing accuracy = 87.7%, sensitivity = 33.3, specificity = 95.3, PPV = 50.0, NPV = 91.1, FPR = 4.69 and AUROC = 88.5. All receiver operator curves are shown in **Supplemental Figure X.**

DISCUSSION

As the second wave of COVID-19 has started to unfold the already stained healthcare systems globally are being pushed to the limit with hospital and intensive care unit (ICU) beds reaching full-capacity. Impact of the virus has been global with developed countries even struggling with infection rates and hospitalization (11). The second wave is anticipated

to be much tougher than the first one (12). With new vaccines on the horizon infection rates in the United states have skyrocketed to 1.34 million cases being diagnosed in the second week of December 2020 and more than 1100 deaths (13). The vaccines have just received FDA approval for widespread use of the vaccine for prevention of the disease while the actual logistics and distribution plans are still unknown (14). However, there is a need for the existence of clinical biomarkers and predictive models for mortality in these patients. machine learning has been used to predict mortality in cancer(15), cardiac disease (16); while our own work on mortality prediction has been on trauma patients, postoperative ileus cases in the ICU (5, 6) and diverticulitis in the inpatient setting (17) with good accuracies. A lot of epidemiological studies reporting clinical, laboratory and mortality outcomes have been done worldwide including developed and developing countries, but very few actually reported or developed a machine learning model for predicting the outcomes with a set accuracy.

In the current study we developed a new method of machine learning (Neo-V Framework) that uses a smaller amount of cases to train a deep neural network to give better predictions. This model is different from our previous FLAIM Framework (two-phase) and has a tri-phase structure (**Figure 1.**) Clinically we used just 1214 hospitalized patients for mortality prediction in RT-PCR positive COVID-19 cases using only data from the first 24-hours after admission. Clinical data analysis showed that with increasing age the patients' mortality also increases. There are number of clinical risk factors that were associated with worse outcomes and documented in the clinical literature like chronic obstructive lung disease [COPD](18), chronic kidney disease [CKD](19), ischemic heart disease [IHD](20), pneumothorax (radiological or clinical diagnosis)(21), acute respiratory syndrome [ARDS](22), septic shock(23), shortness of breath(24), ICU admission(25), AB+ Blood group(26) and recurrent admission to the ICU. Hematological labs that were associated with mortality (previously presented in the results) were also seen in other studies(27). Biochemical laboratory

abnormalities like creatinine, blood urea nitrogen(28), INR and PT(29) were also associated with mortality. Patients managed with invasive ventilation(30) and non-invasive ventilation (31) which actually signifies that the patients that were not able to maintain normal respiratory physiology had worse outcomes. Having fever(32) and the use of systemic steroids(33) as an early symptom of COVID-19 had better prognosis in our patient population.

Machine learning has been used to predict mortality in the inpatient and the ICU setting in a number of different clinical conditions including our prior papers. The methods need to continue to evolve to have better outcome predictions, previously we developed the FLAIM-Framework approach, which was an attempt to build a workflow pipeline to produce more accurate results. The Neo-V Framework builds on our previous work and has the statistical power of FLAIM, but it can be used to apply deep learning to smaller datasets. We call this technique "horizontal expansion" of the dataset in which we horizontally expand the data by combining two or more variables to create new-variables (Neo-V). The combination was conditional that the variables were clinically relevant e.g. BUN and creatinine. In contrast, vertical expansion of the dataset is adding new patients.

We wanted to do a complete and thorough analysis of the Deep-Neo-V algorithm, so we compared it to other deep-learning models out there on COVID-19 mortality prediction currently available in medical literature(34). Our Deep-Neo-V model outperforms all our conventional models and our Deep-FLAIM model. It also outperformed the currently available Deep-learning model by Zhu J.S. *et al* in terms of training set accuracy, testing set accuracy, AUROC, Precision, specificity and positive predictive value(34). However, the Deep-Neo-V model underperformed in terms of sensitivity and slightly with negative predictive value. The Deep-Neo-V will continue to improve and develop and will potentially be replaced by a model with better performance parameters (accuracy, PPV and NPV). This

model in its current configuration can be used to predict mortality after day-1 (considers labs and clinical characteristics in the first 24 hours) of hospital admission and can help in stratification of patients. It can help clinicians answer a number of questions and aid in decision-making like triaging patients, should an elderly patient receive aggressive treatment, or a younger patient receive life supportive management. These are tough questions and the model will give clinicians clarity about the course and plan for the patient. This model can also help clinicians with family counselling, appropriate decision making, limit excessive intervention or aggressive treatments and effective resource management. Like most digital tools this current algorithm is user friendly, can provide results instantaneously and easy to use. After further validation this model can be incorporated into hospital patient management systems and ready for clinical use.

The Deep-Neo-V model has some limitations in terms of the available dataset, retrospective nature of the dataset and data form a single hospital, analyzed at admission and day-one data, other observational study confounders may exist and are unaccounted for. In the immediate future we actively look to validate these findings in an external dataset. In the longer term we will continue to develop an algorithm built on the Neo-V Framework approach that has the potential to be implemented, initially in future pandemics because of its ability to accurately predict outcomes using smaller datasets.

CONCLUSION

Deep-Neo-V is a statistically robust machine learning model that is developed for clinical use to predict mortality risk in patients admitted with RT-PCR proven COVID-19 infection. The mortality prediction was modeled based on clinically relevant variables (patient associated risk factors and the first 24-hours labs. Our experimental results show that with a high

accuracy and specificity it has the potential to develop as a test of choice for predicting mortality in COVID-19 patients. These findings need further external validation.

AUTHORS CONTRIBUTION:

Maleeha Naseem: Study design, methodology, protocol design, IRB/ERC approval, data collection supervision, data collection tool and quality control, statistical analysis review, medical literature review and article writing.

Hajra Arshad: Data collection, protocol design, data collection tool, statistical analysis review, medical literature review and article writing.

Syeda Amrah Hashmi: Data collection, statistical analysis review, medical literature review and article writing.

Furqan Irfan. Medical literature review, statistical analysis review, and article writing.

Fahad Shabbir Ahmed. MD: Original concept, study design, methodology, data collection tool and construction of FLAIM, Deep-FLAIM and Neo-V frameworks, statistical analysis and review, python-coding for machine learning and article writing, editing, review.

CONFLICT OF INTEREST

None

FUNDING STATEMENT

None

FIGURE LEGENDS

Table 1. Demographics with Univariate and multivariate analysis of clinical variables as part of Phase I of the Neo-V and FLAIM machine learning frameworks.

Table 2. Derivative variables and there univariate and multivariate analysis with mortality during hospital stay.

Table 3. Primary and secondary outcomes machine learning results. Test set accuracy, training set accuracy, precision, sensitivity, specificity, positive predictive value, negative predictive value and AUROC.

Table 4. Comparison of currently available Deep learning models for mortality in the inpatient setting.

Figure 1. Experimental design of the study and the machine frameworks

Figure 2. Receiver Operating Curve for Deep-Neo-V with AUROC

Supplemental Figure 1. Receiver Operating Curve with AUROC for a) RF, b) kNN, c) SVC-RBF, d) DT, e) ABC, f) QDA and g) Deep-FLAIM

REFERENCES

1. Hsiang S, Allen D, Annan-Phan S, Bell K, Bolliger I, Chong T, et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. Nature. 2020;584(7820):262-7.
2. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. 2020;20(5):533-4.
3. Hunter DJ. Trying to “Protect the NHS” in the United Kingdom. New England Journal of Medicine. 2020.
4. Cacciapaglia G, Cot C, Sannino F. Second wave COVID-19 pandemics in Europe: a temporal playbook. Sci Rep. 2020;10(1):15514.
5. Ahmed FS, Ali L, Joseph BA, Ikram A, Ul Mustafa R, Bukhari SAC. A statistically rigorous deep neural network approach to predict mortality in trauma patients admitted to the intensive care unit. J Trauma Acute Care Surg. 2020;89(4):736-42.
6. Ahmad FS, Ali L, Raza Ul M, Khattak HA, Hameed T, Wajahat I, et al. A hybrid machine learning framework to predict mortality in paralytic ileus patients using electronic health records (EHRs). Journal of Ambient Intelligence and Humanized Computing. 2020.
7. Naseem M, Akhund R, Arshad H, Ibrahim MT. Exploring the Potential of Artificial Intelligence and Machine Learning to Combat COVID-19 and Existing Opportunities for LMIC: A Scoping Review. J Prim Care Community Health. 2020;11:2150132720963634.
8. Pathologist CoA. Guidance for COVID-19 Testing for CAP-Accredited Laboratories. CAP Guidelines, caporg. 2020.
9. Prevention CfDCa. CDC’s Diagnostic Test for COVID-19 Only and Supplies. CDC Guidelines, cdcgov. 2020.
10. N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer. SMOTE: Synthetic Minority Over-sampling Technique. J Artif Int Res. 2002;16(1):321–57.
11. Keni R, Alexander A, Nayak PG, Mudgal J, Nandakumar K. COVID-19: Emergence, Spread, Possible Treatments, and Global Burden. Front Public Health. 2020;8:216.
12. Xu S, Li Y. Beware of the second wave of COVID-19. Lancet. 2020;395(10233):1321-2.
13. (CDC) CfDCaP. CDC COVID Data Tracker: Maps, charts, and data provided by the CDC. wwwcdcgov. 2020.
14. (FDA) FDA. COVID-19 Vaccines: FDA has rigorous scientific and regulatory processes in place to facilitate development and ensure the safety, effectiveness and quality of COVID-19 vaccines. Online. 2020.
15. Parikh RB, Manz C, Chivers C, Regli SH, Braun J, Draugelis ME, et al. Machine Learning Approaches to Predict 6-Month Mortality Among Patients With Cancer. JAMA Netw Open. 2019;2(10):e1915997.
16. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. Eur Heart J. 2017;38(7):500-7.
17. Ahmed FS, Raza-Ul-Mustafa, Ali L, Imad-ud-Deen, Hameed T, Ikram A, et al. Machine Learning Can Predict Deaths in Patients with Diverticulitis During their Hospital Stay. medRxiv. 2020:2020.02.04.20020222.
18. Pranata R, Soeroto AY, Huang I, Lim MA, Santoso P, Permana H, et al. Effect of chronic obstructive pulmonary disease and smoking on the outcome of COVID-19. Int J Tuberc Lung Dis. 2020;24(8):838-43.
19. Ozturk S, Turgutalp K, Arici M, Odabas AR, Altiparmak MR, Aydin Z, et al. Mortality analysis of COVID-19 infection in chronic kidney disease, haemodialysis and renal transplant patients compared with patients without kidney disease: a nationwide analysis from Turkey. Nephrol Dial Transplant. 2020;35(12):2083-95.

20. Hendren NS, Drazner MH, Bozkurt B, Cooper LT, Jr. Description and Proposed Management of the Acute COVID-19 Cardiovascular Syndrome. Circulation. 2020;141(23):1903-14.
21. Wang XH, Duan J, Han X, Liu X, Zhou J, Wang X, et al. High incidence and mortality of pneumothorax in critically Ill patients with COVID-19. Heart Lung. 2020.
22. Li X, Ma X. Acute respiratory failure in COVID-19: is it "typical" ARDS? Crit Care. 2020;24(1):198.
23. Coz Yataco AO, Simpson SQ. Coronavirus Disease 2019 Sepsis: A Nudge Toward Antibiotic Stewardship. Chest. 2020;158(5):1833-4.
24. Shi L, Wang Y, Wang Y, Duan G, Yang H. Dyspnea rather than fever is a risk factor for predicting mortality in patients with COVID-19. J Infect. 2020;81(4):647-79.
25. Zhao Z, Chen A, Hou W, Graham JM, Li H, Richman PS, et al. Prediction model and risk scores of ICU admission and mortality in COVID-19. PLoS One. 2020;15(7):e0236618.
26. Hultström M, Persson B, Eriksson O, Lipcsey M, Frithiof R, Nilsson B. Blood type A associates with critical COVID-19 and death in a Swedish cohort. Critical Care. 2020;24(1):496.
27. Henry BM, de Oliveira MHS, Benoit S, Plebani M, Lippi G. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis. Clin Chem Lab Med. 2020;58(7):1021-8.
28. Cheng Y, Luo R, Wang K, Zhang M, Wang Z, Dong L, et al. Kidney disease is associated with in-hospital death of patients with COVID-19. Kidney Int. 2020;97(5):829-38.
29. Altschul DJ, Unda SR, Benton J, de la Garza Ramos R, Cezayirli P, Mehler M, et al. A novel severity score to predict inpatient mortality in COVID-19 patients. Sci Rep. 2020;10(1):16726.
30. Becher T, Frerichs I. Mortality in COVID-19 is not merely a question of resource availability. The Lancet Respiratory medicine. 2020;8(9):832-3.
31. Rahim F, Amin S, Noor M, Bahadur S, Gul H, Mahmood A, et al. Mortality of Patients With Severe COVID-19 in the Intensive Care Unit: An Observational Study From a Major COVID-19 Receiving Hospital. Cureus. 2020;12(10):e10906-e.
32. Leung C. Risk factors for predicting mortality in elderly patients with COVID-19: A review of clinical data in China. Mech Ageing Dev. 2020;188:111255-.
33. Sterne JAC, Murthy S, Diaz JV, Slutsky AS, Villar J, Angus DC, et al. Association Between Administration of Systemic Corticosteroids and Mortality Among Critically Ill Patients With COVID-19: A Meta-analysis. Jama. 2020;324(13):1330-41.
34. Zhu JS, Ge P, Jiang C, Zhang Y, Li X, Zhao Z, et al. Deep-learning artificial intelligence analysis of clinical variables predicts mortality in COVID-19 patients. J Am Coll Emerg Physicians Open. 2020:10.1002/emp2.12205.

Table 1. Demographics with Univariate and multivariate analysis of clinical variables as part of Phase I of the Neo-V and FLAIM machine learning frameworks.

			Univariate			Multivariate		
		n = 1214 (n, effected) and Ranges	p-value	HR	95%CI	p-value	HR	95%CI
Demographics	AGE	range (19-96 years)	<0.001*	1.04	1.03-1.06	-	-	-
	Age >50 years	755	<0.001*	3.15	1.81-5.50	-	-	-
	Age >65 years	336	<0.001*	2.43	1.72-3.44	-	-	-
	Female	467	0.489	0.88	0.61-1.27	-	-	-
	Readmission	46	0.147	0.05	0.01-2.95	0.951	0.001	0.01-3.40E+164
	Mortality	130 (10.7%)	-	-	-	-	-	-
Blood Grouping	A-	6	0.289	2.91	0.41-20.90	0.553	1.822	0.26-13.18
	A+	132	0.642	0.88	0.51-1.51	0.561	0.852	0.5-1.47
	B-	19	0.332	1.64	0.60-4.45	0.858	1.098	0.4-3.02
	B+	233	<0.05*	1.5	1.03-2.18	0.051	1.459	1-2.13
	O-	13	0.27	0.04	0.00-11.71	0.947	0.001	0.01-2.07E+130
	O+	183	0.224	1.28	0.86-1.90	0.25	1.262	0.85-1.88
	AB-	6	<0.05*	3.88	1.23-12.27	<0.001*	7.634	2.39-24.48
	AB+	48	<0.05*	2.55	1.29-5.04	<0.001*	2.483	1.26-4.93
Comorbidities	Chronic kidney disease	71	<0.001*	3.45	2.08-5.72	<0.001*	2.737	1.63-4.61
	Chronic liver disease	13	0.15	2.32	0.74-7.33	0.125	2.467	0.78-7.82
	Chronic obstructive lung disease	14	<0.05*	2.93	1.19-7.21	0.163	1.918	0.77-4.79
	Diabetes	441	<0.05*	1.45	1.03-2.05	0.261	1.224	0.87-1.74
	Hypertension	521	0.58	1.41	0.99-2.01	0.493	1.137	0.79-1.64
	Ischemic heart disease	158	<0.001*	2.24	1.62-3.62	<0.05*	1.801	1.19-2.73

	Other Comorbidities	571	<0.05*	1.76	1.24-2.48	<0.05*	1.721	1.21-2.45
Symptoms	Asymptomatic	87	0.177	0.38	0.12-1.27	0.269	0.497	0.15-1.72
	Chest pain (non-cardiac)	56	0.252	1.63	0.71-3.71	0.434	1.395	0.61-3.21
	Cough	522	0.110	0.76	0.54-1.07	0.28	0.823	0.58-1.18
	Fatigue / lethargy	52	0.793	1.15	0.42-3.12	0.63	0.78	0.29-2.14
	Fever	768	<0.001*	0.56	0.39-0.79	<0.05*	0.568	0.4-0.81
	Gastrointestinal symptoms	93	0.875	0.94	0.42-2.14	0.626	0.813	0.36-1.88
	Headache	16	0.367	0.05	0.01-34.53	0.953	0.001	0.01-4.24E+150
	Hemoptysis	9	0.609	0.05	0.01-5037.35	0.961	0	0-1.48E+166
	Malaise	83	0.976	0.99	0.41-2.43	0.537	0.752	0.31-1.87
	More than 2 symptoms	901	0.975	1.01	0.66-1.56	0.817	0.95	0.62-1.47
	Myalgia	28	0.284	1.88	0.6-5.94	0.221	2.06	0.65-6.54
	Nasal obstruction	2	0.788	0.05	0.01-157108055.1	0.969	0.001	0.01-4.22E+198
	Other symptoms	321	<0.001*	1.97	1.38-2.8	<0.05*	1.66	1.16-2.38
	Pneumothorax (clinical / radiological)	22	<0.05*	2.00	1.17-3.43	<0.05*	2.014	1.18-3.46
	Rhinorrhea	7	0.502	0.05	0.01-327.09	0.964	0.001	0.01-1.70E+200
	Sore throat	58	0.240	0.51	0.16-1.59	0.637	0.758	0.24-2.41
	Sputum	29	0.070	2.03	0.95-4.35	0.247	1.576	0.73-3.41
Clinical	Acute Kidney Injury	179	<0.001*	2.95	2.08-4.18	<0.001*	2.465	1.74-3.51
	ARDS / Respiratory failure	147	<0.001*	2.81	1.96-4.03	<0.001*	2.527	1.77-3.63
	Septic Shock	49	<0.001*	5.05	3.43-7.43	<0.001*	4.299	2.91-6.37
	Shock Liver	3	<0.001*	11.63	2.84-47.73	<0.001*	11.476	2.77-47.68
	Shortness of breath	570	<0.001*	1.81	1.24-2.62	<0.001*	1.96	1.35-2.86

Point of Care	Special care unit	308	0.355	1.18	0.84-1.68	0.577	1.106	0.78-1.58
	Days spent in the ICU	Range (0-20 days)	0.187	1.03	0.99-1.07	0.133	1.029	1-1.07
	Intensive care unit (ICU)	106	<0.001*	2.71	1.88-3.92	<0.001*	3.241	2.22-4.74
	Number of admissions ICU during current hospitalization	Range (0-2)	<0.001*	1.78	1.3-2.43	<0.001*	2.066	1.49-2.87
Radiology	Bilateral chest X-ray abnormalities	753	0.484	1.14	0.8-1.64	0.34	1.195	0.83-1.72
	Unilateral chest X-ray abnormalities	131	0.911	0.97	0.56-1.7	0.997	0.999	0.58-1.75
Day One labs	Abnormal blood lymphocyte count	378	<0.001*	3.02	1.74-5.26	<0.001*	2.786	1.6-4.86
	Abnormal blood neutrophil count	1044	<0.05*	2.5	1.17-5.36	<0.05*	2.657	1.24-5.7
	Abnormal platelets count	363	<0.05*	1.43	1.01-2.03	<0.05*	1.413	1-2.01
	Abnormal serum albumin	49	0.296	1.45	0.73-2.87	0.207	1.552	0.79-3.07
	Abnormal serum ALT	946	0.380	1.23	0.78-1.95	0.475	1.183	0.75-1.88
	Abnormal serum APTT	347	0.217	1.26	0.88-1.81	0.137	1.32	0.92-1.9
	Abnormal serum bilirubin	69	0.587	0.82	0.4-1.69	0.33	0.699	0.34-1.42
	Abnormal serum BUN	382	<0.001*	3.15	2.19-4.52	<0.001*	2.784	1.94-4.01
	Abnormal serum calcium	387	0.596	0.91	0.62-1.33	0.633	0.91	0.62-1.35
	Abnormal serum creatinine	373	<0.001*	3.21	2.24-4.6	<0.001*	2.73	1.9-3.94
	Abnormal serum hematocrit	505	<0.001*	1.87	1.33-2.65	<0.05*	1.808	1.26-2.61
	Abnormal serum hemoglobin	879	<0.05*	1.67	1.08-2.58	<0.05*	1.63	1.05-2.55
	Abnormal serum INR	198	<0.001*	3.38	2.38-4.8	<0.001*	3.243	2.28-4.63
	Abnormal serum LDH	751	0.597	1.11	0.77-1.58	0.806	1.047	0.74-1.5
	Abnormal serum magnesium	268	0.535	0.88	0.57-1.35	0.841	0.957	0.63-1.48
	Abnormal serum phosphorus	32	0.893	1.08	0.4-2.91	0.58	0.752	0.28-2.07
	Abnormal serum potassium	241	0.886	0.97	0.64-1.49	0.678	0.914	0.6-1.41
	Abnormal serum procalcitonin	523	0.312	1.2	0.85-1.69	0.485	1.132	0.8-1.61

	Abnormal serum PT	378	<0.05*	1.5	1.05-2.15	<0.05*	1.787	1.24-2.59
	Abnormal serum sodium	458	0.250	1.23	0.87-1.73	0.501	1.127	0.8-1.6
	Abnormal white cell count	455	<0.001*	2.08	1.45-2.96	<0.001*	2.089	1.47-2.98
Treatment	Anti-malarial	2	0.385	2.4	0.34-17.27	0.587	1.73	0.24-12.51
	Anti-viral drugs	99	0.715	0.9	0.51-1.6	0.538	0.835	0.47-1.49
	CRRT	7	<0.05*	2.72	1.19-6.23	0.054	2.287	0.99-5.3
	Hydroxychloroquine	77	0.230	0.61	0.27-1.38	0.336	0.667	0.3-1.53
	Intravenous IgG	6	<0.05*	3.87	1.23-12.23	<0.05*	3.945	1.25-12.49
	Invasive ventilation	74	<0.001*	2.79	1.89-4.1	<0.001*	3.208	2.15-4.79
	Lopinavir/Ritonavir	1	0.616	0.05	0.01-6488.77	0.962	0.001	0.01-6.44E+175
	Non-invasive ventilation (BiPAP/CPAP)	243	<0.001*	1.96	1.38-2.8	<0.001*	1.872	1.32-2.67
	Plasmapheresis	96	0.317	1.25	0.82-1.91	0.234	1.294	0.85-1.98
	Supportive treatment	617	<0.001*	3.89	2.22-6.81	<0.001*	3.507	2.01-6.14
	Symptomatic treatment	1056	<0.05*	0.6	0.36-0.98	0.058	0.613	0.37-1.02
	Systemic steroids	430	0.238	0.8	0.55-1.17	<0.05*	0.657	0.45-0.97

Table 2. Derivative variables and there univariate and multivariate analysis with mortality during hospital stay.

Derivatives for ML	Total No. Comorbidities	Range (0-6)	<0.001*	1.44	1.27-1.64	<0.001*	1.301	1.14-1.49
	More than 2 comorbidities	499	<0.001*	2.28	1.59-3.28	<0.001*	1.788	1.23-2.61
	Total No. symptoms	Range (0-6)	0.379	1.08	0.92-1.27	0.439	1.066	0.91-1.26
	Total No. Treatments received	Range (0-6)	<0.001*	1.31	1.14-1.51	<0.05*	1.268	1.11-1.46
	CR or BUN	517	<0.001*	0.31	0.21-0.46	<0.001*	0.356	0.24-0.54
	HB or HCT	883	<0.05*	0.55	0.35-0.87	<0.05*	0.562	0.36-0.89
	PLT or INR	460	<0.001*	2.21	1.55-3.16	<0.001*	2.087	1.47-2.98
	PT or INR	502	<0.001*	0.46	0.33-0.66	<0.001*	0.414	0.29-0.6
	TLC or LYMP	882	<0.001*	0.18	0.08-0.4	<0.001*	0.186	0.09-0.43
	TLC or NEU	1085	<0.05*	0.2	0.07-0.61	<0.05*	0.18	0.06-0.57
	Total No. laboratory abnormalities	Range (0-17)	<0.001*	1.22	1.16-1.3	<0.001*	1.216	1.15-1.3

Table 3. Primary and secondary outcomes machine learning results

Test set (n = 365)	FLAIM Framework Deep-FLAIM	Deep-Neo-V	RF (d=10, e=2)	KNN (k=3)	SVC - RBF	DT (d=10)	ABC	QDA
Test set accuracy	84.66	87.67	85.75	77.81	65.75	85.21	85.48	87.4
Training set accuracy	86.69	98.70	94.94	92.34	69.02	98.94	90.11	94.58
Precision	0.44	0.61	0.37	0.38	0.46	0.46	0.46	0.49
Sensitivity	68.89	33.33	28.89	42.22	75.56	42.22	48.89	60
Specificity	86.88	95.31	93.75	82.81	64.38	91.25	90.63	91.25
Positive predictive value	42.47	50.00	39.40	25.68	22.97	40.43	42.31	49.09
Negative predictive value	95.21	91.05	90.36	91.07	94.93	91.82	92.65	94.19
Area Under Receiver-Operator Curve	86.90	88.50	69.50	66.50	80.80	66.74	77.65	80.84

Table 4. Comparison of currently available Deep learning models for mortality in the inpatient setting.

	Current	Li X et al, 2020
Sample Size (n)	1214	1108
Clinical setting	In-patient	In-patient
Training and Test (set) split	70/30	90/10
Datapoint / Variables	47	15
Testing accuracy	87.67	85.3
Training accuracy	98.70	89.2
AUROC	0.885	0.844
Sensitivity	33.33	75
Specificity	95.31	87.2
PPV	50.00	52.2
NPV	91.05	94.9

Figure 1. Experimental design of the Framework

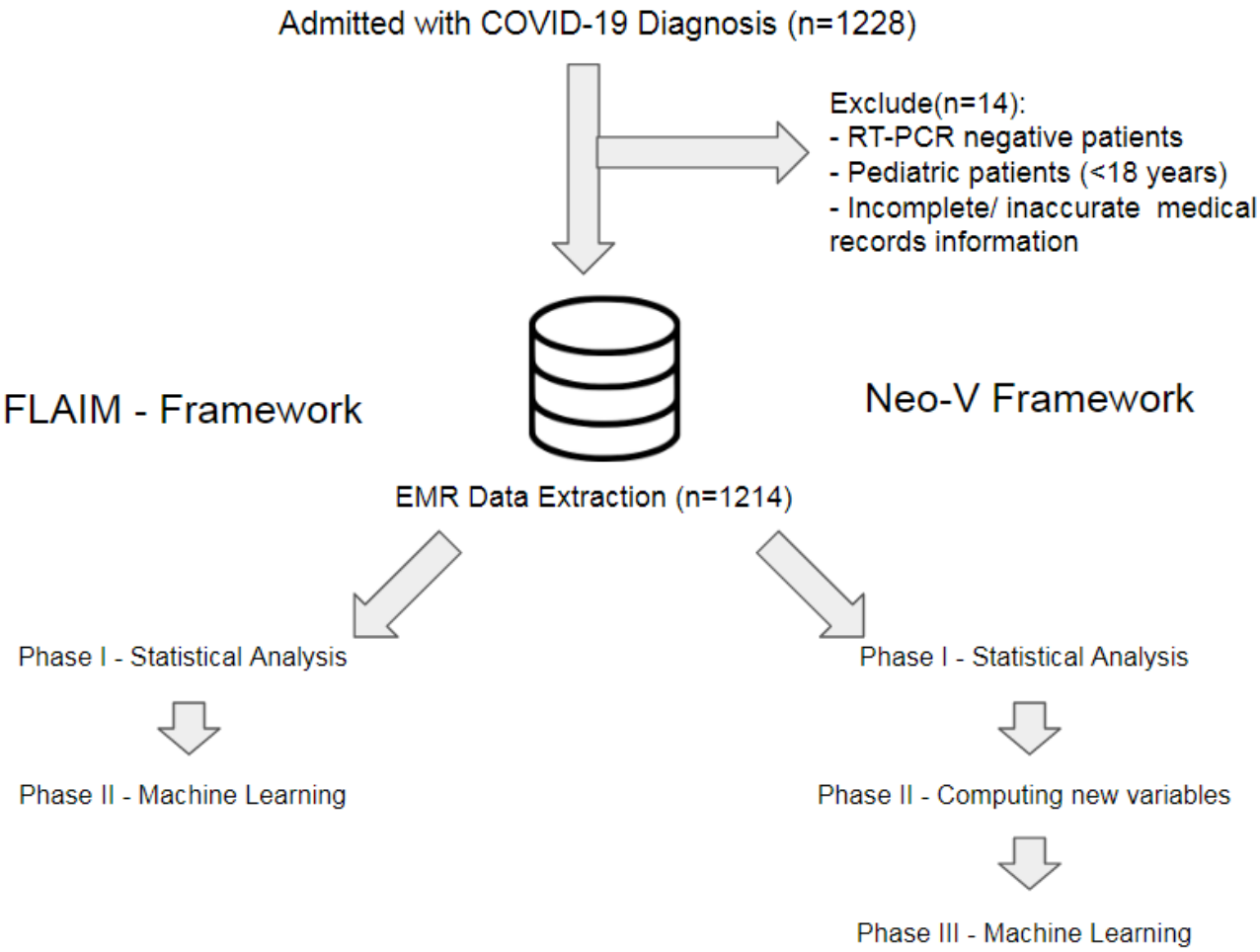
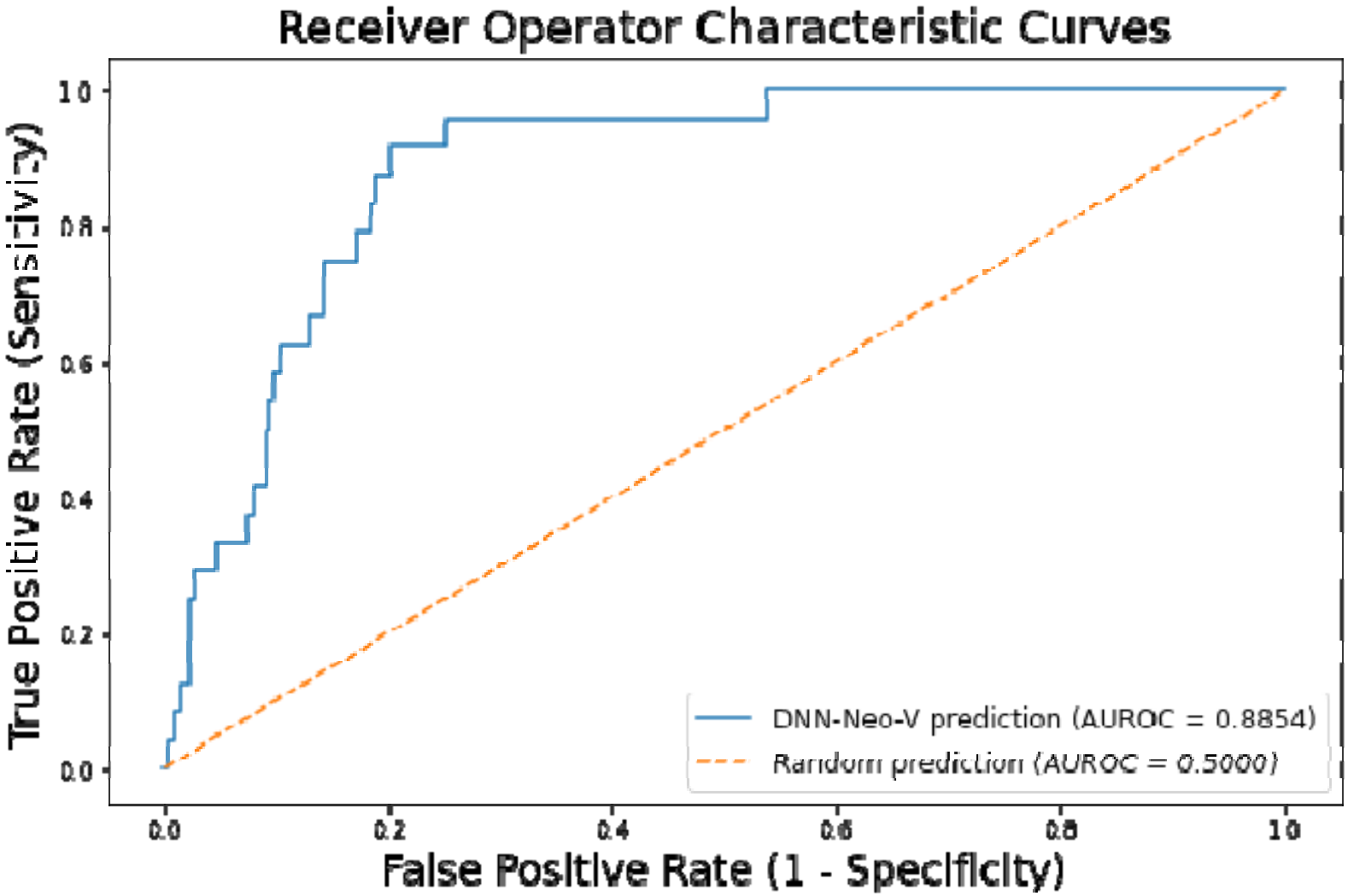
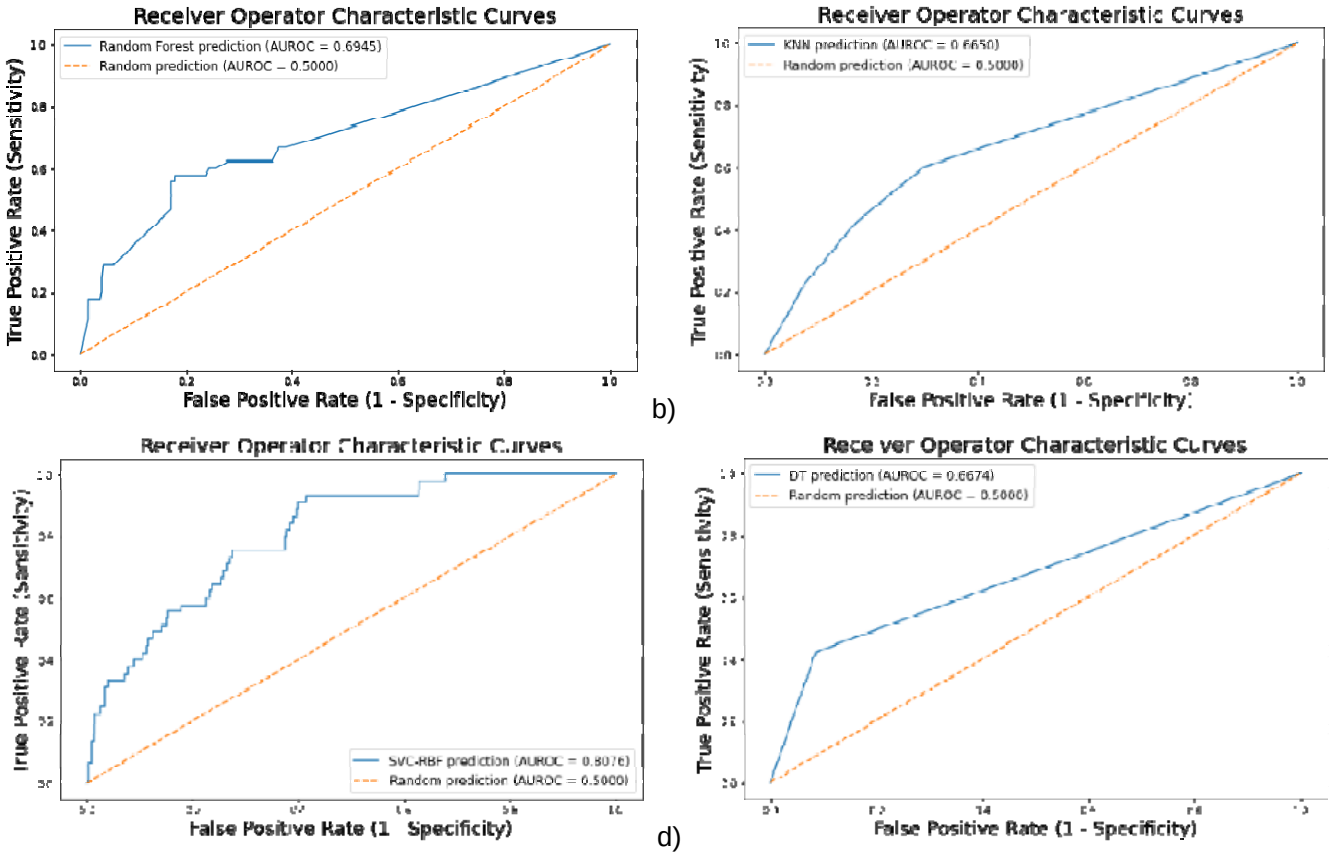
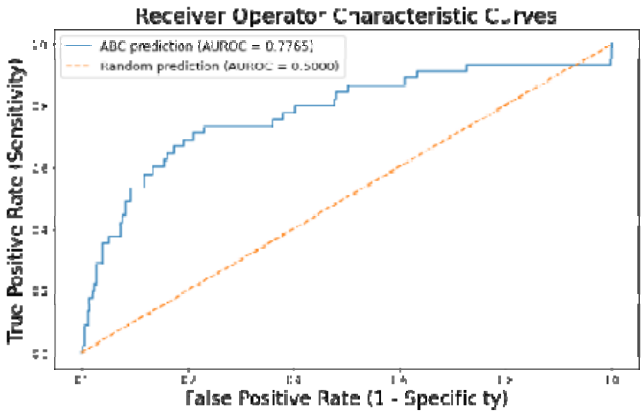


Figure 2. Receiver Operating Curve for Deep-Neo-V with AUROC

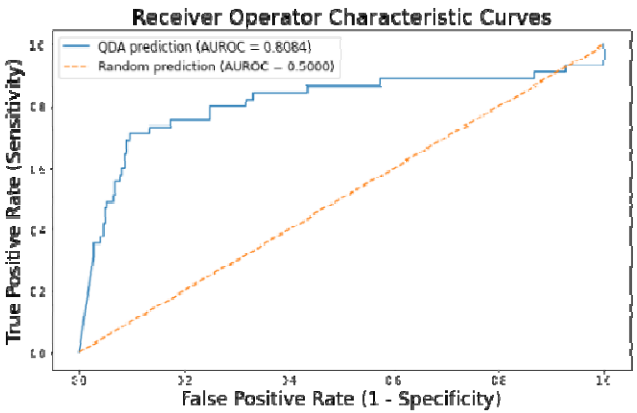


Supplemental Figure 1. Receiver Operating Curve with AUROC for a) RF, b) kNN, c) SVC-RBF, d) DT, e) ABC , f) QDA and g) Deep-FLAIM

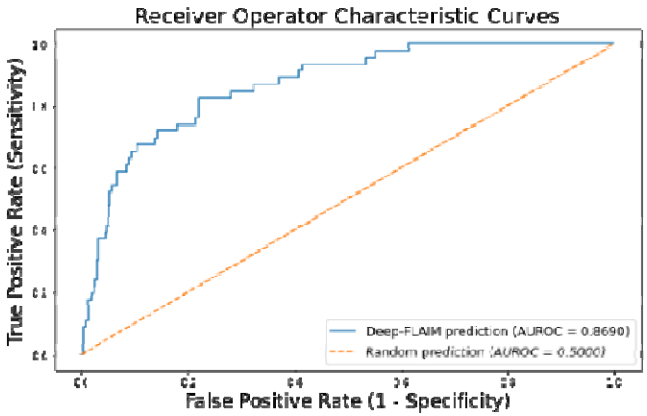



Receiver Operator Characteristic Curves
ABC prediction (AUROC = 0.7765)
Random prediction (AUROC = 0.5000)
True Positive Rate (Sensitivity)
False Positive Rate (1 - Specific ty)

e)


Receiver Operator Characteristic Curves
QDA prediction (AUROC = 0.8084)
Random prediction (AUROC = 0.5000)
True Positive Rate (Sensitivity)
False Positive Rate (1 - Specificity)

f)


Receiver Operator Characteristic Curves
Deep-FLAIM prediction (AUROC = 0.8690)
Random prediction (AUROC = 0.5000)
True Positive Rate (Sensitivity)
False Positive Rate (1 - Specificity)

g)