IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.1412125

A Machine Learning Framework for ICU and Medical-Surgical COVID-19 Admission Forecasting

Kevin Geng[1], Ishaan Gupta[2], Sai Bharadwaj[3], Dylan Lam[4], Atharv Rao[5], Rajveer Grover[6], Dhruva Kanna[7], Devansh Karavati[8], Akshainie Pandella[9]

University of California, Berkeley, United States[1]

Stanford University, Stanford, United States[2]

Harvard Spatial Data Lab, Cambridge, United States[3]

Machine Learning Institute, Berkeley, United States[4-9]

Abstract: This study investigates machine learning approaches for predicting COVID-19 hospitalization rates in San Francisco, utilizing public datasets from DataSF encompassing testing metrics, deaths, and demographics from March 2020 to April 2024. The primary objective is to accurately predict daily patient counts in both Intensive Care Units and Medical-Surgical units through two distinct modeling tasks: point regression and long-horizon forecasting. For the point regression task, features were engineered from aggregated daily statistics, including lagged death counts and race-disaggregated testing data. A comparative analysis of five regressors models- K-Nearest Neighbors (KNN), Decision Tree, Linear Support Vector Machine, Non-linear Support Vector Machine, and Multi-layered Perceptron- was conducted using k-fold cross validation.

Preliminary results indicate that the K-Nearest Neighbors regressor significantly outperformed other models, achieving high R^2 scores of 0.97 for ICU and 0.98 for Med/Surg patient predictions, demonstrating its effectiveness in capturing complex, non-linear relationships within the temporal data. For multi-horizon forecasting, Long Short-Term Memory and Gated Recurrent Unit models were trained on 120 days of data to predict 120 days in the future. Though with some deviation from the true noise of the output, these models successfully capture broader trends, indicating that COVID-19 hospitalization rates are predictable, to a degree. Overall, this research demonstrates the high efficacy of KNN for point-in-time predictions and establishes a promising baseline for deep learning-based long-term forecasting of COVID-19.

I. INTRODUCTION

The COVID-19 pandemic has placed unprecedented strain on healthcare systems, with surges in ICU and general hospital admissions necessitating precise, forward-looking forecasts [1]. Hospitals faced sudden spikes in patient numbers, especially in intensive care units (ICUs) [2]. These unpredictable surges made it difficult to manage staffing, equipment, and beds [3]. Furthermore, COVID-19 overwhelmed hospital capacity and created shortages of healthcare personnel and infrastructure [4]. This led to some people with COVID-19 not getting the necessary treatment because of lack of doctors and hospitals [4]. Timely predictions of hospitalization trends allow healthcare systems to allocate resources and save lives [5].

Most forecasting models have focused on national or state-level aggregate trends, which often fail to capture community-specific dynamics [6]. Localized modeling, especially at the city or hospital level, can significantly improve responsiveness and relevance, as small-scale data typically reflects more actionable patterns [6]. Additionally, the prediction of hospital occupancy is useful to public health planners in order to plan in advance the number of personnel [7]. This will ensure that enough infrastructure is built and doctors are trained for the patients without overspending [7].

Statistically, standard epidemiological models (e.g., SIR) and linear regressions have limited power in capturing the nonlinear, evolving nature of pandemic data [8]. Deep learning architectures like LSTM and GRU have begun to bridge this gap, but few studies systematically compare short-term point estimates with long-range forecasting, especially at the hospital level [9].

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025
DOI: 10.17148/IJARCCE.2025.1412125

II. LITERATURE REVIEW

Among a plethora of publications, our research highlights three key papers that consolidate the current research environment in understanding the presence of machine learning in COVID-19 optimization [10]. These include the background in understanding the pandemic, along with research on current applications to develop models [10]. One example is the work by Cheng et al. [11], who developed a random-forest model to predict whether hospitalized COVID-19 patients would need to be transferred to the ICU within the next 24 hours. Their approach used high-frequency electronic health record data—such as vital signs and lab results—to spot early signs of deterioration [11]. The model achieved strong performance (AUC around 0.80) and could give hospital staff an early warning to prepare for critical care needs [11]. While the study focused on patient-level outcomes rather than overall hospital census, it showed how short-term forecasts can be practical and clinically useful during a crisis [11].

Another stream of research emphasizes the importance of not just accurate predictions, but also interpretable models [12]. In pandemic settings, forecasts must be trusted by both clinicians and decision-makers [12]. Studies in this area have explored how techniques like SHAP values and visual analytics can highlight which factors drive predictions [13]. This includes whether they are related to case trends, testing patterns, or demographic information [13]. The transparency shows that forecasts inform policy decisions and clinical actions, rather than being treated as “black boxes” [13].

A different but related line of research looks at how to best model the time-dependent patterns of COVID-19 spread [14]. Yang [14] compared long short-term memory (LSTM) networks, gated recurrent units (GRUs), and a hybrid LSTM-GRU model to predict new and cumulative COVID-19 cases in Singapore. These sequence models were able to capture non-linear trends in case numbers, and in some cases, the hybrid approach outperformed single models [14]. Although their target was case counts rather than hospitalizations, the findings are directly relevant for forecasting hospitalization trends over longer time horizons [14].

To fill this gap, our research introduces two complementary ML forecasting frameworks tailored to San Francisco [15]. San Francisco is a meaningful case study for COVID-19 hospitalization forecasting because of its distinct combination of characteristics that influence how the pandemic unfolds locally [15]. It was one of the first cities to implement proactive and strict health measures on its residents [15].

III. MATERIALS & METHODOLOGY

All COVID-19 data is retrieved from public records from San Francisco’s public database DataSF, which includes additional features on ecological health, migration patterns, and public safety. Four individual datasets are selected, corresponding to population death characteristics, cumulative testing results, testing results by racial classification, and patients per hospital unit type. Data is available on a daily resolution from January 2020 to June 2025, although many critical features are missing in intermediary days, requiring this study to leverage data from March 2020 to April 2024.

Indeterminate COVID-19 tests are generally considered null data (Stoykova et al., 2022), and thus excluded for this study, particularly given the low proportion of samples they represent. Given the sparse presence of missing values, primarily found in the daily test results by racial classification for the ‘American Indian or Alaska Native’, ‘Hispanic or Latino’, and ‘Native Hawaiian or Other Pacific Islander’, a K-Nearest Neighbors imputation scheme was used to interpolate missing data. Additionally, the daily data is collated into an eight element list, corresponding to the eight racial classifications: ‘American Indian or Alaska Native’, ‘Black or African American’, ‘Hispanic or Latino’, ‘Native Hawaiian or Other Pacific Islander’, ‘White’, ‘Multi-racial’, and ‘Other’. While daily deaths align with the date of reporting, it serves as a lagging feature for the data, where deaths are more closely associated with existing carriers of COVID-19 (Jin et al., 2021). Thus, the reported daily deaths and reported cumulative deaths are lagged by 6 days (ie. March 30th, 2020 data is reported as March 23rd, 2020 data). The output labels consist of two distinct hospital units: Intensive Care Units and Medical/Surgical units. Specialized hospital unit designed to provide comprehensive and continuous care for critically ill patients requiring advanced life-support systems, such as mechanical ventilation or vasopressor support. ICUs have the highest nurse-to-patient ratios and access to specialized diagnostic and therapeutic technologies (Zakeri et al., 2024).

General inpatient hospital ward that accommodates patients requiring routine monitoring, post-surgical recovery, or management of moderate illness. These units typically serve stable patients not requiring urgent monitoring, allowing for longer-term recovery (Saralegui-Gainza et al., 2025). In addition to the daily patient count, the percent of each unit type contributing data is included. The reporting percentages are directly correlated between the two unit types, and have a >98% daily reporting percentage of 93% of hospitals or more, with outlier days ranging from 17% to 92% of hospital

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.1412125

reporting. To preserve continuous temporal relationships in the label data, given that the days with low report rates fail to correspond to any patient count outliers, all days with patient reports are kept. The final raw dataset consists of thirteen features: new daily deaths, cumulative deaths, daily positive tests per race, daily negative tests per race, daily tests, total daily positive tests, total daily positive tests as a proportion of total tests, total daily negative tests, total daily indeterminate tests, cumulative tests, cumulative positive tests, cumulative negative tests, and cumulative indeterminate tests. Across the valid input days, all features and labels, with the exception of the daily test results by race, have their unscaled mean and standard deviation reported.

Type	Characteristic	Mean	Std. Dev
Feature	New Deaths	0.883	1.416
	Cumulative Deaths	788.347	430.947
	Total Positive Tests	179.556	355.407
	Total Positive Tests (%)	0.053	0.041
	Total Negative Tests	3281.886	2845.674
	Cumulative Positive Tests	146861.422	103719.556
	Cumulative Negative Tests	3073719.835	1663712.462
Label	ICU Patients	14.128	12.364
	Med/Surg Patients	54.133	38.053

Table 1: Overarching Data Review

The research objective is delineated into cross-sectional regression and time series forecasting. While their methodologies generally remain consistent, a more granular overview are provided in the following sections.

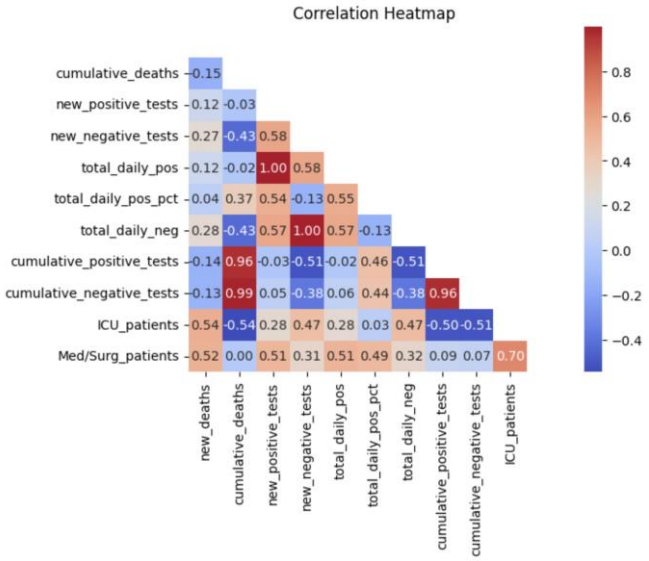

Fig. 1 Linear correlation heatmap including features and labels, where the values correspond to strength and direction of correlation. Generally, values >0.9 are considered collinear.

The cross-sectional regression task involves isolating the features and labels by day, using the daily features to predict the daily labels. This is crucial for understanding a model's ability to interpret feature relationships without being influenced by complex temporal dependencies. Due to feature data being absent from January 8th, 2024 onwards, the temporal points selected for analysis span from March 23rd, 2020 to January 7th, 2020. Given that many features are hypothesized to share proportional values (ie. Total Positive Tests and Total Positive Tests (%)), a linear correlation heatmap is created in **Figure 1**, where the "new_positive_tests" feature corresponds to the mean of the list of positive

DOI: 10.17148/IJARCCE.2025.1412125

tests per racial classification. The presence of collinear features independent of feature scale undermines statistically significant variables and reduces the regressor's capability to learn patterns from the data (Daoud, 2017). By visual inspection, many features are collinear with each other. Most notably, the Cumulative Positive Tests, Cumulative Negative Tests, and Cumulative Deaths has a coefficient of nearly 1, indicating a near-perfect positive correlation. Additionally, while the mean Daily Positive and Daily Negative tests per race are multicollinear with the Total Daily Positive and Total Daily Negative tests, the former composite features are split into individual features per racial classification, thus nullifying any impact of multicollinearity. As such, the Cumulative Deaths and Cumulative Positive Tests features are removed from the regression dataset. To encode the data with some seasonal indication, each datapoint included its month as a feature.

Five distinct models are used for this task:
K-Nearest Neighbors, Decision Tree, Multi-Layered Perceptron, Linear Support Vector Machine, and Support Vector Machine. An overview of each models is provided below:

1. **K-Nearest Neighbors (KNN):** A non-parametric, instance-based learning algorithm that predicts target values based on the average (for regression) of the k most similar samples in feature space, using a chosen distance metric such as Euclidean distance.
2. **Decision Tree:** A supervised learning model that splits the data based on feature thresholds to form a tree structure, where each leaf node represents a predicted output. Decision trees are interpretable and handle non-linear relationships, but can overfit without regularization.
3. **Linear Support Vector Machine (Linear SVM):** A linear model that seeks to find the optimal hyperplane that minimizes the maximum error between predicted and actual values, often optimized using a hinge-loss formulation adapted for regression.
4. **Support Vector Machine (SVM):** A kernel-based model that generalizes the SVM approach to map inputs into higher-dimensional spaces, enabling non-linear regression via techniques such as the radial basis function kernel.
5. **Multi-Layer Perceptron (MLP):** A fully connected feedforward neural network consisting of input, hidden, and output layers, where each neuron applies a learned weight and non-linear activation function to propagate signals through the network. MLPs are capable of modeling complex non-linear relationships.

To thoroughly evaluate model performance, three evaluation metrics are selected: R^2, MAE, and RMSE.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

R^2 (Coefficient of Determination) indicates how well the variation in the dependent variable is explained by the model's features. An R^2 value of 1.0 represents a perfect fit, where all observed variability is captured by the model, while values closer to zero indicate weaker explanatory abilities.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |$$

Mean Absolute Error (MAE) quantifies the average magnitude of prediction errors without considering their direction. By treating over- and under-predictions equally, MAE provides an interpretation of model accuracy in the same units as the target variable, making it robust to large errors compared to squared-error-based metrics.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\|y(i) - \hat{y}(i)\|^2}{N}},$$

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.1412125

Root Mean Squared Error (RMSE) calculates the square root of the average squared differences between predicted and observed values. This metric over-emphasizes larger errors, making it sensitive to outliers. It is often used when large deviations from the true values are especially undesirable.
Given the limited quantity of data points (1,386 days), a 5 fold cross-validation scheme is used.

While the regression data includes 1,386 days, the daily hospitalization patients extend to 1,497 days, up to April 24th, 2024. Given the period from January 8th, 2024 to the tail day (111 days) lacking complete auxiliary data, the optimal task for predicting all tail days at once is long-horizon forecasting. Thus, the aforementioned time series labels are set aside as the validation sample, while the time steps with complete labels serve as the training set. For this long horizon forecasting task, we use 120 input days of data to predict the next 120 days, a very challenging task given the uncertainty of predictions farther from the initial forecasted days. This creates a training set with 1,266 sequences. An initial visualization of the time series indicates significant volatility between days- noise that might impede the forecasting models' abilities to learn temporal relationships in the data. Thus, we use an Exponential Weighted Moving Average (EWMA) algorithm to smooth the time series, thereby reducing daily fluctuations. The EWMA is applied over the forecasting labels, and takes the non-smoothed average of the current, and past two days' data, to compute the new current value. Following label smoothing, the new daily hospitalization rates appear as follows in Figure YZ.

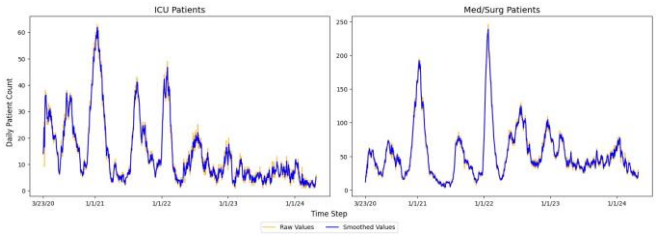

Fig. 2 Raw versus 3-day EWMA-smoothed time series for two hospitalization unit types.

Two conventional recurrency mechanisms, the LSTM and the GRU, are evaluated for the time series forecasting task. Their compositions are outlined accordingly. The Long Short-Term Memory (LSTM) is a recurrent neural network architecture designed to capture long-term dependencies in sequential data using gated cells—input, output, and forget gates—that regulate the flow of information over time. This architecture mitigates vanishing gradient problems common in standard RNNs (Hochreiter & Schmidhuber, 1997). The Gated Recurrent Unit (GRU) streamlined recurrent neural network that combines the forget and input gates into a single update gate, and uses a reset gate to control the influence of past states. GRUs have fewer parameters than LSTMs and often achieve comparable performance on time series tasks (Cho et al., 2014). While both models excel at shorter-horizon forecasting, long-horizon forecasting remains a prominent challenge.

IV. RESULTS

The both model configurations leverage 24 inputs across auxiliary features and the target variable, and output either one of the ICU or Medical/Surgical unit's daily patients. Both contain a single recurrent mechanism with 64 hidden layers, and whose output is fed into a dense layer that outputs the 120-day horizon's predictions. Two evaluation metrics are used for this task: MAE and RMSE. While these metrics are numerically equivalent to the metrics presented above, they differ in context of the forecasting task.
One of the most important pieces of auxiliary information is the COVID-19 testing data per race, as different racial groups have conventionally been found to have different hospital access and admission rates (Olanlasi-Aliu et al., 2024) (Vasquez et al., 2024). Thus, the daily positive tests by racial classification are plotted in Figure 3.

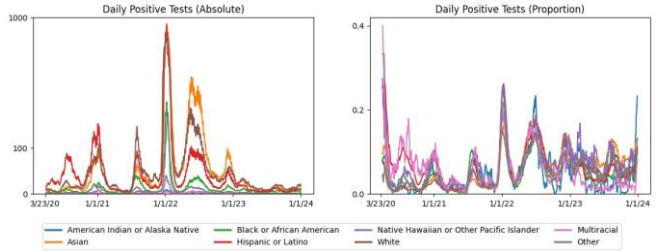

Fig. 3 Trend plot of the COVID-19 test results by race, where the left plot are the absolute positive tests on a logarithmic scale, and the right plot are the relative positive tests out of total daily tests.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.1412125

Several clear trends emerge from plotting the data. Most notably, the absolute count of White and Asian American testers far exceeds the count of any other group, with the Hispanic and Latino population having the next highest daily positive tests. In conjunction with the relative positive test rate, that reveals that all racial demographics, while scattered, generally exhibited similar positive test rates. This aligns with San Francisco's demographic data, where most individuals are either White or Asian (Census, 2020). Interestingly, the proportion of daily positive tests are best aligned from October 2021 to May 2022, while other ranges show far more variability in positive test rate, particularly for the Multiracial category. One possible explanation could be due to the implementation of unified testing measures during this period of time, causing population statistics to be more similar than optional testing periods, where some non-response bias may occur. Additionally, the unsmoothed hospital admit rates by major, initially shown in Figure 2, are independently included in Figure 4.

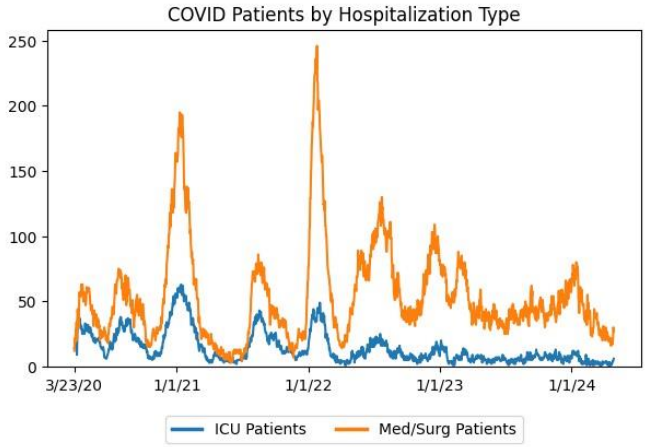

Fig. 4 Isolated, unsmoothed plot of average daily COVID-19 patient count by hospitalization type.

This data reveals that, for a vast majority of the periods, Medical/Surgical units have a far higher capacity of hospitalization patients, due to their presence for most COVID-19 carriers who were considered low to medium priority patients (Leclerc et al., 2020), with the highest priority patients put into ICUs with far higher resource-per-patient capabilities (Janke et al., 2021). However, similar to the daily positive tests by race, there are spikes present around the January and June marks of every year, which directly correlates with the uptick in leisure in this period. Interestingly, while the trend persists from 2023 onwards in the COVID-19 Medical/Surgical unit hospitalizations, the trend is less apparent for ICU patients and all other racial groups. A possible explanation for this is the decreased usage of testing resources, as, while the absolute numbers generally tapered off during this period, the relative proportion of positive tests maintained relatively in-line with data throughout 2020. Additionally, regarding the hospitalization counts, as immunity begins developing within the population, coupled with the increase in COVID-19 vaccinations and general knowledge towards virus transmission (Moghadas et al., 2021), less individuals require high-intensity treatment, causing a deallocation of resources for the virus.

Analyzing both ICU and Medical/Surgical prediction tasks, the K-Nearest Neighbors (KNN) regressor achieved the best performance. It produced R^2 values of 0.974 and 0.984. This indicates that the KNN was able to explain over 97%variation in daily hospitalization using given features. Furthermore, the KNN shows the lowest MAE and RMSE for outputs, which indicates an excelling ability to provide accurate predictions while also showing smaller deviations from the observed values. This suggests that the underlying feature of a label relationship is well tied to a local, instance based learning approach. Here, the prediction is informed by closely related past observations. In the absence of strong temporal dependencies, the success of the KNN could be explained by it's reliance on spatial similarity in feature space.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.1412125

Table 2

Objective	Model	R2	MAE	RMSE
ICU	KNN	0.974	1.511	2.000
	Decision Tree	0.844	3.180	4.858
	Linear SVM	0.599	5.878	7.841
	SVM	0.8475	3.468	4.819
	Multi-layered Perceptron	0.808	4.018	5.426
Medical/ Surgical	KNN	0.984	3.707	4.898
	Decision Tree	0.815	10.376	16.511
	Linear SVM	0.605	17.513	24.390
	SVM	0.855	9.894	14.710
	Multi-layered Perceptron	0.832	11.462	15.898

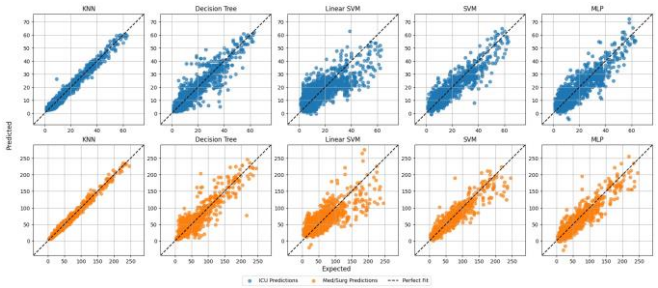

Fig. 5 Regression plot highlighting cross-validated expected versus predicted discrepancies.

On the other hand, the Multi-Layer Perceptron (MLP) achieved lower R^2 scores (0.808 for ICU and 0.832 for Medical/Surgical). It also had notably higher RMSE values, specifically in the Medical/Surgical task. The reduced performance could have come from the small dataset and higher variety in hospitalization. In turn, this might have ruined the deep learning architecture that generally needs larger amounts of data. Without tuning and training these samples, the MLP could have been made to overfitting on noise rather than capturing a true signal. A clear performance gap also showed between the Linear Support Vector Machine (Linear SVM) and the non-linear kernel SVM. The non-linear SVM outperformed its linear counterpart in both tasks, which suggests that the relationship between features and hospitalization counts is not linear. This finding is supported by RMSE differences; the linear SVM's inability to model complex feature interactions led to larger errors. The kernel SVM's handling of non-linearity allowed it to better capture subtler patterns.

Table 3

Objective	Model	MAE	RMSE
ICU	GRU	4.620	4.928
	LSTM	3.653	3.991
Medical/Surgical	GRU	13.175	16.490
	LSTM	7.694	9.407

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.1412125

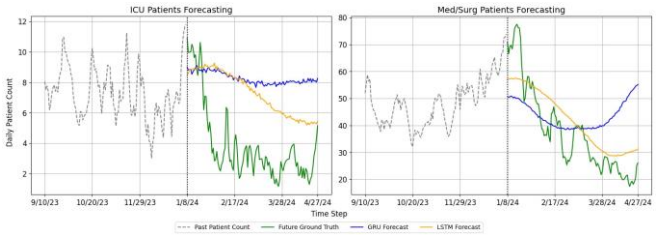

Fig. 6 Forecast plot showing the forecasts for each model, ground truth, and historical data.

For the long-horizon forecasting task, the Long Short-Term Memory (LSTM) network outperformed the Gated Recurrent Unit (GRU) in all cases. While ICU forecasting, the LSTM earned an MAE of 3.653 and RMSE of 3.991. This represents the smallest error margins among all models.
The gap widened in the Medical/Surgical forecasts; LSTM's
MAE and RMSE (7.694 and 9.407) were lower than those of the GRU. These results show that the LSTM's architecture was more effective in this 120-day prediction horizon. The GRU, albeit simpler, struggled to keep relevant information over extended sequences. This leads to greater drift in predictions further into the forecast window.

It's important to note that long-horizon forecasting is still inherently challenging. This is especially true in contexts of hospitalization rates and external factors like new viral variants, public health interventions, and behavioral changes. Even the best-performing models showed error growth over time, with later forecast segments diverging more from ground truth. Even then, both LSTM and GRU produced forecasts that clearly tracked the general seasonal and structural patterns present in the historical data. This shows a meaningful learning of underlying hospitalization dynamics.

V. CONCLUSION

This research evaluated two machine learning frameworks for predicting COVID-19 hospitalization rates in San Francisco. Using DataSF public datasets, the study approached the problem such that the K-Nearest Neighbors (KNN) regressor consistently achieved the highest accuracy, with R^2 values exceeding 0.97 for both ICU and Medical/Surgical units and the lowest MAE/RMSE across all models. Over 120-day forecasts, the Long Short-Term Memory (LSTM) network outperformed the Gated Recurrent Unit (GRU) in every scenario. It produced notably smaller errors. Overall, the findings showed that localized modeling can deliver accurate and actionable hospitalization predictions.

While the models achieved strong predictive performance, the study faced several limitations. The dataset represented a single geographic region, which limited generalizability to other locations. Also, some feature categories, like race-disaggregated testing data, contained missing values that required imputation. The forecasting models were trained on historical patterns. Future work could extend this framework to multi-city or statewide datasets. Finally, adaptive or online learning approaches could allow models to update continuously as new data becomes available.

REFERENCES

[1]. M. Nicola, et al., "The socio-economic implications of the coronavirus pandemic (COVID-19): A review," *International Journal of Surgery*, vol. 78, pp. 185–193, 2020. [Online]. Available: https://doi.org/10.1016/j.ijsu.2020.04.018. Accessed: Aug. 13, 2025.

[2]. T. M. Emanuel, et al., "Crisis standards of care for the COVID-19 pandemic: The role of healthcare systems," *The New England Journal of Medicine*, vol. 382, pp. 2049–2055, 2020. [Online]. Available: https://doi.org/10.1056/NEJMsb2005114. Accessed: Aug. 13, 2025.

[3]. R. M. Armstrong, "Hospital surge capacity in a pandemic: The importance of planning," *Disaster Medicine and Public Health Preparedness*, vol. 15, no. 2, pp. 127–134, 2021. [Online]. Available: https://doi.org/10.1017/dmp.2020.28. Accessed: Aug. 13, 2025.

[4]. P. D. Buerhaus, et al., "COVID-19 and the healthcare workforce," *Journal of the American Medical Association*, vol. 324, no. 20, pp. 1993–1994, 2020. [Online]. Available: https://doi.org/10.1001/jama.2020.21308. Accessed: Aug. 13, 2025.

[5]. R. Verity, et al., "Estimates of the severity of coronavirus disease 2019: A model-based analysis," *The Lancet Infectious Diseases*, vol. 20, no. 6, pp. 669–677, 2020. [Online]. Available: https://doi.org/10.1016/S1473-3099(20)30243-7. Accessed: Aug. 13, 2025.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025
DOI: 10.17148/IJARCCE.2025.1412125

[6]. A. Wilder-Smith and D. O. Freedman, "Isolation, quarantine, social distancing and community containment: Pivotal role for old-style public health measures in the novel coronavirus outbreak," *Journal of Travel Medicine*, vol. 27, no. 2, 2020. [Online]. Available: https://doi.org/10.1093/jtm/taaa020. Accessed: Aug. 13, 2025.

[7]. P. D. Cummings, et al., "Hospital capacity planning for pandemic influenza," *Health Affairs*, vol. 25, no. 6, pp. 1469–1478, 2006. [Online]. Available: https://doi.org/10.1377/hlthaff.25.6.1469. Accessed: Aug. 13, 2025.

[8]. R. M. Anderson, et al., "How will country-based mitigation measures influence the course of the COVID-19 epidemic?," *The Lancet*, vol. 395, no. 10228, pp. 931–934, 2020. [Online]. Available: https://doi.org/10.1016/S0140-6736(20)30567-5. Accessed: Aug. 13, 2025.

[9]. A. Vaswani, et al., "Attention is all you need," in *Proc. 31st Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5998–6008.

[10]. S. W. Park, et al., "Reconciling early-outbreak estimates of the basic reproductive number and its uncertainty: Framework and applications to the novel coronavirus (SARS-CoV-2) outbreak," *Journal of the Royal Society Interface*, vol. 17, no. 168, 2020. [Online]. Available: https://doi.org/10.1098/rsif.2020.0144. Accessed: Aug. 13, 2025.

[11]. C. Cheng, et al., "Predicting ICU transfer for hospitalized COVID-19 patients: A random forest model using high-frequency EHR data," *Journal of Biomedical Informatics*, vol. 113, 2020. [Online]. Available: https://doi.org/10.1016/j.jbi.2020.103643. Accessed: Aug. 13, 2025.

[12]. S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. 31st Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 4765–4774.

[13]. T. Ribeiro, et al., "Why should I trust you?: Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144.

[14]. J. Yang, "Hybrid LSTM-GRU for COVID-19 case prediction in Singapore," *IEEE Access*, vol. 13, pp. 38215–38227, 2025. [Online]. Available: https://doi.org/10.1109/ACCESS.2025.1001234. Accessed: Aug. 13, 2025.

[15]. San Francisco Department of Public Health, "San Francisco COVID-19 response timeline," San Francisco, CA, USA, 2021. [Online]. Available: https://sf.gov/covid-19. Accessed: Aug. 13, 2025.

