Bayesian LSTM Forecasting of COVID-19 ICU Occupancy With Uncertainty Estimations

Katarzyna Nieszporek[1], Giulia Capecchi[1] and Maria Grazia Bocci[1]

[1]*Department of Artificial Intelligence, Czestochowa University of Technology, Czestochowa, Poland*
[2]*Intensive Care Unit, National Institute for Infectious Diseases "Lazzaro Spallanzani" IRCCS, Rome, Italy*

Abstract

The COVID-19 pandemic has exposed the fragility of healthcare systems, especially in the management of Intensive Care Unit (ICU) resources. Accurate forecasting of ICU occupancy is essential to support public health decisions and to prevent saturation during epidemic waves. In this work, we propose a predictive model based on Long Short-Term Memory (LSTM) neural networks, combined with Monte Carlo dropout to estimate model uncertainty. This approach allows us to generate probabilistic forecasts of ICU demand, including confidence intervals that help quantify prediction reliability.

We apply the model to real-world data from California counties, using historical ICU occupancy records collected during the pandemic. We show that the model can anticipate trends up to several weeks in advance, maintaining good accuracy and consistent uncertainty calibration. To assess robustness, we compare the proposed LSTM model with simpler architectures, including GRU-based and feedforward neural networks, confirming the superior performance of LSTM in capturing complex temporal patterns.

Our results highlight the importance of integrating uncertainty estimates into forecasting systems, particularly in high-risk domains such as healthcare. The method is computationally efficient, easy to implement, and adaptable to other time series prediction tasks where uncertainty awareness is required.

Keywords

ICU occupancy prediction, COVID-19, LSTM networks, time series forecasting, healthcare resource planning

1. Introduction

The COVID-19 pandemic has revealed the limitations and vulnerabilities of healthcare systems across the world [1, 2]. One of the most critical challenges faced during the emergency phase was the saturation of Intensive Care Units (ICUs), which rapidly became unable to accommodate the growing number of patients requiring urgent and life-saving treatments [3, 4]. This situation has demonstrated the importance of predictive tools that can support healthcare systems in advance planning and resource allocation [5, 6].

In many countries, governments and hospitals were forced to make complex decisions regarding the distribution of limited resources under uncertain and rapidly evolving conditions [7, 8]. Often, these decisions had to be taken without the support of reliable forecasts, leading to either overestimations, with underutilized resources, or underestimations, with tragic consequences for patient care. This lack of foresight exposed how fragile and reactive most infrastructures still are when it comes to managing exceptional pressure over extended periods of time [9].

Within this critical context, it has become evident that the ability to forecast future resource needs is not merely a technical problem, but a strategic component of health governance [10]. Forecasting the number of ICU beds required, even with a moderate level of uncertainty, can allow health authorities to organize staff, reallocate equipment, delay non-urgent procedures, or coordinate patient transfers more effectively [11]. The accuracy and trustworthiness of such predictions directly influence the capacity to reduce avoidable mortality, limit system overload, and optimize healthcare delivery under crisis conditions.

From this context, the need for robust models capable of forecasting ICU occupancy became evident. However, the nature of the data involved in pandemic scenarios is deeply uncertain. The transmission rate of the virus, the effects of governmental restrictions, individual behaviors, and emerging virus variants all contribute to creating a system with high variability and limited predictability [12, 13]. Therefore, classical deterministic models are not sufficient. It becomes essential to account for the uncertainty associated with predictions.

One of the central problems in predictive modeling is not only generating a forecast, but also knowing how much that forecast can be trusted [14]. Traditional models usually produce a single estimate for each future value, implicitly assuming a high level of confidence. However, in situations where data is sparse, noisy, or rapidly changing, this assumption is unrealistic and potentially dangerous. In the case of ICU demand during a pandemic, overconfident predictions can lead to insufficient prepa-

ICYRIME 2025: 10th International Conference of Yearly Reports on Informatics, Mathematics, and Engineering. Czestochowa, January 14-16, 2025

mariagrazia.bocci@inmi.it (M. G. Bocci)

ration, while excessively conservative estimates can lead to the waste of precious resources [15].

To address this, we adopt the concept of model uncertainty as formulated by Gal and Ghahramani [16, 17], who demonstrated that dropout, typically used for regularization in deep learning, can also be interpreted as an approximate Bayesian inference method. This insight allows us to construct neural models that do not just provide point predictions, but also quantify the uncertainty associated with those predictions. This probabilistic information is crucial in high-stakes contexts, where decisions must often be made even when the available data is incomplete or ambiguous [18].

In this work, we apply this methodology to a Long Short-Term Memory (LSTM) network [19], a class of recurrent neural networks particularly suited to model temporal dependencies in sequential data. These networks are able to retain and exploit long-term dependencies in time series data, making them particularly effective for modeling the evolution of ICU admissions over time [20, 21]. By applying Monte Carlo (MC) dropout during both training and inference, our model is capable of generating probabilistic forecasts of ICU occupancy that include confidence intervals, enabling decision-makers to interpret the model outputs with greater caution and awareness.

We test this approach on real-world data concerning ICU occupancy across counties in California during the COVID-19 pandemic, with the goal of providing a flexible and interpretable forecasting system. The results indicate that the model is able to provide meaningful forecasts, even several weeks in advance, and that the uncertainty estimation can serve as a key element in the process of healthcare planning. In addition to evaluating prediction accuracy, we also assess the calibration of the uncertainty estimates, that is, how well the predicted confidence intervals match the true variability of future data.

In the following sections, we will review related works (Section 2), present the theoretical foundations of LSTM networks and model uncertainty (Section 3.1), describe our proposed model in detail (Section 4), report on experimental results and model performance (Section 5), and finally discuss the implications, limitations, and future directions of our research (Section 6).

2. Related Works

The prediction of ICU occupancy during health emergencies, such as the COVID-19 pandemic, has received increasing attention from the scientific community. This interest is motivated by the urgent need to support healthcare systems in optimizing the use of limited resources, especially in times of crisis.

Several studies have explored the application of machine learning techniques for forecasting hospital admissions, ICU transfers, and mortality risk. These approaches typically rely on the analysis of clinical data, epidemiological curves, and time series models. For instance, Chamola and Sikdar [22] provided a broad review of artificial intelligence methods applied to disaster and pandemic management, including early warning systems, resource allocation strategies, and decision support tools. Their work highlights the potential of AI to improve preparedness and response capacity in large-scale emergencies.

A specific example of this approach is presented by Cheng et al. [23], who developed a risk prediction tool using Random Forests to estimate the probability of ICU transfer within 24 hours. This tool is based on electronic health record data and allows physicians to identify high-risk patients in advance. Similarly, Ruyssinck et al. [24] proposed a model for ICU bed prediction using Random Survival Forests. In their study, the Sequential Organ Failure Assessment (SOFA) score was used as a key input, and the model outperformed traditional machine learning methods for survival analysis in critical care. Another relevant contribution is the work by Li et al. [25], who employed a Deep Neural Network (DNN) combined with a feature selection method (Boruta algorithm) to build a risk score for ICU admission and patient mortality. Their model showed good performance, especially in identifying the most relevant clinical predictors.

Beyond direct ICU prediction, many authors have proposed machine learning systems designed to detect early warning signals and support preventive decision-making. For instance, studies during the COVID-19 pandemic have explored how to integrate clinical markers with psychological, behavioral, and societal factors. A growing number of works has highlighted the potential of combining multimodal data, including physiological signals and high-level cognitive features, to support healthcare response strategies.

Notable efforts have also focused on the detection and classification of disinformation related to COVID-19, which may indirectly affect the behavior of the population and the load on hospital systems. De Magistris et al. [26] proposed an explainable fake news detection system combining named entity recognition and stance classification, showing that misinformation during a pandemic can propagate uncertainty and reduce adherence to preventive measures, thus indirectly affecting hospital saturation patterns.

Complementary studies explored the psychological and neurocognitive consequences of the pandemic, especially in relation to post-COVID stress syndromes. In particular, Russo et al. [27] proposed an innovative approach using remote EMDR therapy to treat long-COVID-related traumatic disorders. Such works are relevant be-cause they reveal how the pandemic impacted not only physical but also psychological health, both of which can influence the demand on healthcare facilities.

From a methodological point of view, advanced clustering and statistical learning techniques have been used to analyze both behavioral data and physiological signals in the context of pandemic-related stress. For example, Ponzi et al. [28] used Expectation Maximization and Gaussian Mixture Models to investigate the differences in psychodiagnostic profiles before and after the pandemic, using Rorschach test data. These types of investigations, though not focused directly on ICU occupancy, enrich the broader understanding of pandemic impacts on healthcare systems.

Moreover, the use of computer vision methods for surveillance and prevention has seen widespread experimentation. De Magistris et al. [29] developed an automatic CNN-based system for face mask detection, tested in real-world scenarios during the COVID-19 emergency. Monitoring compliance with mask-wearing policies is another aspect that, indirectly, affects the spread of infection and therefore the load on ICU infrastructures.

While these models and applications provide important insights, most of them focus on patient-level prediction using static clinical data, or address secondary aspects related to prevention and communication. In contrast, our approach focuses on a population-level prediction of ICU bed usage, considering temporal dynamics and variability over time. This aspect is crucial in a pandemic, where the number of new infections and hospitalizations can change rapidly due to social behavior, public policies, and virus mutations.

The closest works to our approach are the studies by Gal and Ghahramani [16, 17], who introduced the concept of model uncertainty in neural networks using dropout as a Bayesian approximation. Their framework allows the estimation of predictive uncertainty without requiring a full Bayesian treatment, which would be computationally expensive. This idea is especially valuable in the healthcare domain, where making overconfident predictions can lead to serious consequences, such as underestimating ICU demand or delaying interventions.

Therefore, our work builds upon these contributions by integrating the dropout-as-Bayesian approach with LSTM networks for time series forecasting. This combination enables us to provide not only accurate predictions of ICU occupancy, but also to associate each prediction with a quantitative measure of uncertainty. This feature is fundamental in supporting cautious and informed decisions in critical contexts such as healthcare planning during a pandemic.

3. Background

In this section we present the main theoretical foundations upon which our work is built. The aim is to provide the conceptual and methodological background needed to understand the structure and rationale of the model we propose. Forecasting ICU occupancy during a pandemic presents a unique combination of challenges: on one hand, the system evolves over time in a non-linear and context-dependent way; on the other hand, any predictive model must be able to represent and communicate its own uncertainty to support decision-making processes in high-risk environments.

To address these requirements, our work is based on two fundamental components. The first is the use of Long Short-Term Memory (LSTM) networks, a class of recurrent neural networks specifically designed to process sequential data. LSTMs are particularly well suited to time series forecasting tasks where past values influence future observations, even across long temporal gaps. Their internal structure allows them to capture dependencies over time more effectively than traditional models, making them ideal for modeling ICU admission patterns, which often follow delayed and seasonally modulated trends.

The second key component is the use of Monte Carlo dropout, a technique that allows us to estimate the predictive uncertainty of deep learning models in a computationally efficient way. Instead of relying on fully Bayesian neural networks, which are difficult to implement and often computationally prohibitive, Monte Carlo dropout enables approximate Bayesian inference through stochastic forward passes in standard architectures. This method allows us to associate each forecast with a confidence interval, which is essential in a healthcare context where predictions cannot be blindly trusted, and caution is required in interpreting the results.

These two elements—temporal modeling through LSTM networks and uncertainty estimation through dropout interpreted in a Bayesian framework—are then combined in our architecture to construct a robust, flexible, and interpretable forecasting model. The following subsections describe each component in detail.

3.1. Long Short-Term Memory (LSTM) Networks

LSTM networks are a special type of Recurrent Neural Networks (RNNs), specifically designed to handle sequences and temporal dependencies. Traditional RNNs suffer from problems such as the vanishing or exploding gradient, which make it difficult to learn long-term dependencies in time series. LSTMs solve this limitation by introducing a more sophisticated memory unit that includes several internal gates: the input gate, the forget gate, and the output gate [19].

Each LSTM cell contains a memory unit that can maintain information over long periods of time. The input gate controls how much new information is stored, the forget gate decides what information to discard, and the output gate determines how much of the memory is exposed to the next layers of the network. This internal mechanism allows LSTMs to learn long-range patterns and to keep important signals in memory while ignoring irrelevant data.

Due to these characteristics, LSTM networks are widely used in many real-world applications involving sequential data, such as speech recognition, financial forecasting, and medical monitoring [30]. In our study, we use a multi-layer LSTM network to model the temporal evolution of ICU occupancy in different regions, allowing the system to capture the non-linear dependencies and seasonalities typical of epidemic curves.

3.2. Model Uncertainty and Monte Carlo Dropout

In traditional machine learning, models produce point estimates: they predict a single value for each input. However, in high-stakes contexts like healthcare, it is essential not only to predict a value, but also to know how confident the model is in its prediction. This concept is known as model uncertainty.

A promising method to estimate uncertainty in neural networks is the technique proposed by Gal and Ghahramani [16, 17]. They demonstrated that applying dropout during both training and testing phases can be interpreted as a form of approximate Bayesian inference. This allows the model to simulate a posterior distribution over its weights and produce a distribution of outputs rather than a single point estimate.

In practice, this approach is implemented using Monte Carlo dropout (MC dropout). During inference, multiple stochastic forward passes are executed through the same network with dropout activated, and the results are averaged. This procedure generates both the expected output and a measure of variance, which reflects the model's confidence. Formally, if we perform T forward passes, each with a different dropout mask, the predictive mean is estimated as:

$$\mathbb{E}[y^*] = \frac{1}{T}\sum_{t=1}^{T} \hat{y}_t^*(x) \tag{1}$$

and the predictive variance as:

$$\mathbb{V}[y^*] = \frac{1}{T}\sum_{t=1}^{T} \hat{y}_t^{*2}(x) - (\mathbb{E}[y^*])^2 \tag{2}$$

In our case, this allows us to not only forecast ICU bed usage but also to associate each prediction with a confidence interval. This is particularly useful in guiding decisions regarding the allocation of resources, where the cost of a false prediction can be extremely high. Using dropout in this way enables a form of Bayesian modeling that is computationally efficient and compatible with modern deep learning frameworks.

4. Proposed Model

The core of our approach is the design of a forecasting model based on Long Short-Term Memory (LSTM) networks, enhanced with a method to quantify uncertainty using Monte Carlo dropout. The goal is to build a predictive system that can provide both the expected number of ICU beds occupied in the near future and the associated confidence intervals. This dual output is essential to support decision-makers in high-risk and high-variability environments such as healthcare.

4.1. Motivation and Design Rationale

During the pandemic, the evolution of ICU occupancy followed complex temporal patterns. These patterns are not only influenced by biological and medical factors (e.g., virus spread, severity of cases), but also by non-linear external factors such as lockdown policies, vaccination campaigns, or population movements. To capture these dynamics, it is necessary to adopt a model that is able to learn temporal dependencies and nonlinearities from past sequences.

Moreover, the data used for such predictions are subject to noise, inconsistencies, and rapid changes in trends. Hence, our model must also be able to express its own uncertainty: this means providing not only a prediction, but also an estimate of how reliable that prediction is. A decision made on a forecast with high uncertainty should be treated differently than one based on a highly confident output.

4.2. Model Architecture

We construct a deep LSTM model consisting of four recurrent layers, each followed by a dropout layer. The use of stacked LSTM layers allows the model to learn increasingly abstract temporal patterns, while the dropout layers serve both as regularization (during training) and as a Bayesian approximation (during inference).

Each model takes as input a univariate time series x_t representing the number of ICU patients at time t. Optionally, an encoded representation of the county is also included if multiple counties are modeled together. The model outputs a prediction $\hat{x}_{t+1}$, i.e., the estimated number of ICU beds that will be occupied in the next time step.

4.3. Dropout in Recurrent Layers

The novelty of our approach is in the use of **recurrent dropout** with fixed masks, as proposed by Gal and Ghahramani [17]. In contrast to traditional dropout, where different neurons are dropped at each time step, here the same dropout mask is applied consistently across all time steps for the recurrent weights. This strategy ensures a proper approximation of the variational inference process in RNNs.

During training, dropout is applied both to input and recurrent connections. The same is done during inference, which transforms the deterministic prediction into a stochastic one. Each forward pass through the network generates a slightly different result, depending on the random dropout masks.

4.4. Monte Carlo Dropout for Uncertainty Estimation

To estimate the uncertainty, we use the Monte Carlo (MC) dropout method. Specifically, we perform T stochastic forward passes at test time, each time using a different dropout mask. This process yields a set of predictions $\{\hat{x}_{t+1}^{(1)}, \hat{x}_{t+1}^{(2)}, \ldots, \hat{x}_{t+1}^{(T)}\}$.

From this set, we compute:

- The predictive mean:

$$\bar{x}_{t+1} = \frac{1}{T}\sum_{i=1}^{T}\hat{x}_{t+1}^{(i)}$$

- The predictive variance:

$$\mathbb{V}[x_{t+1}] = \frac{1}{T}\sum_{i=1}^{T}\left(\hat{x}_{t+1}^{(i)}\right)^2 - (\bar{x}_{t+1})^2$$

These statistics allow us to construct prediction intervals at various confidence levels (e.g., 68%, 95%, 99%). The wider the interval, the higher the uncertainty. This is especially important for hospital administrators: a sharp increase in uncertainty may signal unusual trends, requiring increased attention or human intervention.

4.5. Advantages of the Proposed Architecture

The model we propose offers several advantages: it is based on well-established deep learning techniques (LSTM, dropout), but reinterpreted in a Bayesian framework; it can handle multiple time series (e.g., different counties) either separately or jointly using encoded identifiers; it provides interpretable uncertainty estimates without requiring a full probabilistic model or computationally expensive Bayesian methods; it is compatible with any modern deep learning framework and can be deployed on standard hardware.

In the next section, we describe the experimental protocol, including the dataset, the preprocessing steps, and the empirical evaluation of our model.

4.6. Theoretical Justification and Bayesian Framing

The model presented in this work relies on a theoretical framework that interprets dropout as an approximation of Bayesian inference. This idea, introduced by Gal and Ghahramani, allows us to treat standard neural networks as approximate Bayesian models, without requiring changes in the model architecture or complex probabilistic methods.

In classical Bayesian inference, the goal is to estimate the posterior distribution of the model parameters given the observed data. Formally, we aim to compute:

$$p(W \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid W)\cdot p(W)}{p(\mathcal{D})} \tag{3}$$

where W represents the weights of the model and $\mathcal{D}$ is the training dataset. However, this posterior is typically intractable in neural networks, due to the high dimensionality of the parameter space and the non-linear nature of the model.

To overcome this difficulty, variational inference can be used. The idea is to approximate the true posterior distribution $p(W \mid \mathcal{D})$ with a simpler distribution $q(W)$, and to find the parameters of q that minimize the Kullback-Leibler divergence between q and the true posterior. This is equivalent to maximizing the evidence lower bound (ELBO), defined as:

$$\mathcal{L} = \mathbb{E}_{q(W)}[\log p(\mathcal{D} \mid W)] - \mathrm{KL}(q(W)\|p(W)) \tag{4}$$

In the interpretation proposed by Gal, the application of dropout during training and inference corresponds to sampling from a variational distribution $q(W)$, where the weights are randomly masked by a binary matrix drawn from a Bernoulli distribution. Specifically, each weight matrix is redefined as:

$$W_i = M_i \cdot \mathrm{diag}(z), \quad z_j \sim \mathrm{Bernoulli}(p) \tag{5}$$

This formulation allows the use of dropout not only as a regularization method, but as a way to simulate multiple network configurations drawn from the approximate posterior distribution. Each stochastic forward pass corresponds to a sample from the variational approximation.

The predictive distribution for a new input x^* is obtained by marginalizing over the weight distribution:

$$p(y^* \mid x^*, \mathcal{D}) = \int p(y^* \mid x^*, W) \cdot q(W) \, dW \quad (6)$$

Since this integral cannot be computed analytically, it is estimated by Monte Carlo sampling. In practice, we perform multiple forward passes through the network using different dropout masks and compute the empirical mean and variance of the predictions:

$$\mathbb{E}[y^*] \approx \frac{1}{T} \sum_{t=1}^{T} \hat{y}_t^*, \quad \mathrm{Var}[y^*] \approx \frac{1}{T} \sum_{t=1}^{T} (\hat{y}_t^*)^2 - (\mathbb{E}[y^*])^2 \quad (7)$$

This approach makes it possible to estimate the epistemic uncertainty of the model without requiring the use of fully Bayesian neural networks, which are often difficult to implement and computationally expensive.

In summary, the Bayesian interpretation of dropout provides a principled way to incorporate model uncertainty into deep learning. This is particularly important in healthcare applications, where decisions based on predictions must also take into account the confidence in those predictions. By applying Monte Carlo dropout, our model can offer both the expected evolution of ICU occupancy and a reliable measure of its own uncertainty.

5. Experiments

In this section, we describe the experimental framework used to validate our approach. We first present the dataset and the criteria adopted for its selection. Then we explain the preprocessing operations necessary to train the model. Finally, we report the training strategies and the performance metrics used to evaluate both accuracy and uncertainty.

5.1. Dataset Description

We used an open-access dataset provided by the California Department of Public Health (CDPH) [31]. The dataset includes daily data on the COVID-19 pandemic collected at the county level, covering the period from March 2020 to May 2021. Specifically, we focused on the number of ICU beds occupied by confirmed COVID-19 patients in each county.

The dataset includes measurements for 58 counties. However, not all of them have complete or stable records. To ensure data reliability, we performed a quality control phase and excluded counties with excessive missing values or inconsistent time series. The final dataset included 36 counties, which represent a good balance between geographic coverage and data quality.

5.2. Time Series Preprocessing

Since our objective is to predict ICU occupancy over time, we treated each county's record as a univariate time series. The raw data are noisy and affected by local fluctuations (e.g., reporting delays, corrections). Therefore, we applied the following preprocessing steps:

1. Missing Data Handling: Missing values were interpolated using a linear method, as long as the proportion of missing points was below 10% in a time series. Counties with too many missing values were excluded.

2. Normalization: To make the training process more stable, we scaled the data. Two scaling strategies were used: *Standard Scaler*: applied to grouped counties (mean 0, standard deviation 1); *MinMax Scaler*: applied to single-county models, with upper bound equal to the mean ICU occupancy N_{mean} (rather than 1), to reduce saturation effects.

3. Sliding Window: To generate training sequences, we applied a sliding window of fixed size $w = 14$ days. Each input sample is a sequence of ICU values over 14 days, and the target is the ICU value on the 15th day.

5.3. Grouping Similar Counties

To avoid building a separate model for each county, we explored the possibility of grouping counties with similar ICU occupancy profiles. We applied the Dynamic Time Warping (DTW) algorithm to compute pairwise distances between the normalized time series. Counties with low DTW distance were grouped together, resulting in six distinct clusters.

This grouping reduces the number of models to be trained. and improves model generalization by sharing statistical patterns across counties with similar epidemic curves.

5.4. Dropout Rate Tuning

We empirically evaluated several values of dropout probability p in the range [0.05, 0.3]. We observed that lower dropout rates (e.g., $p = 0.08$) yield more stable predictions, especially when training on small datasets. Higher dropout values resulted in wider uncertainty bands, but sometimes degraded the mean prediction.

This analysis confirmed the need to carefully tune dropout rates when using MC dropout for uncertainty estimation.

5.5. Monte Carlo Estimation

At test time, we performed $T = 1000$ stochastic forward passes through the trained model. For each input window, we obtained the predictive mean $\bar{y}_t$ as the average of outputs, the standard deviation σ_t as a measure of uncertainty, and the confidence intervals at 68%, 95%, and 99% using standard Gaussian quantiles:

$$\bar{y}_t \pm z \cdot \sigma_t, \quad \text{with } z = 1, 1.96, 2.58$$

This allowed us to associate each forecast with a visual band representing the model's confidence.

5.6. Evaluation Metrics

We evaluated model performance using two **Root Mean Square Error (RMSE):** to assess the accuracy of point forecasts

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Moreover we also used **Calibration of Uncertainty** as the percentage of true values falling within the predicted 95% confidence interval.

Across all counties, the average RMSE was 1.7 on the training set and 3.0 on the test set, confirming the model's ability to generalize across time.

These results demonstrate the model's ability to produce reliable and interpretable forecasts, even several weeks ahead. The use of uncertainty estimates increases trustworthiness and allows for more cautious resource planning.

5.7. Comparison with Alternative Architectures

To better assess the effectiveness of the proposed LSTM-based model with Monte Carlo dropout, we conducted a comparative study with two alternative neural architectures. The objective was to evaluate the importance of temporal memory, recurrent structure, and uncertainty estimation by analyzing models with different levels of complexity and expressiveness.

5.7.1. GRU-based model

Gated Recurrent Units (GRUs) are a simplified variant of LSTM networks. They have fewer parameters and use only two gates: an update gate and a reset gate. The reduced complexity makes GRUs faster to train, especially when computational resources are limited.

In our experiments, we implemented a GRU model with the same structure as the LSTM model (four recurrent layers), applying Monte Carlo dropout in the same way. The performance was slightly lower than the LSTM model, particularly in time series with strong seasonality or abrupt changes. However, training time was reduced by approximately 30%.

Table 1
Comparison of model performance

Model	RMSE (Test)	95% C.I.
LSTM + MC Dropout	3.00	94.1%
GRU + MC Dropout	3.25	91.3%
Feedforward + MC Dropout	4.02	88.6%

5.7.2. Feedforward model with sliding window

As a baseline, we tested a simple feedforward neural network with no recurrent connections. The input to the network was a fixed-size sliding window of the previous 14 days of ICU occupancy, and the output was the prediction for the following day.

This model was also trained with dropout, and uncertainty estimation was performed using Monte Carlo sampling. Despite its simplicity, the feedforward model performed reasonably well on counties with very regular trends. However, it failed to capture long-term dependencies and reacted poorly to abrupt changes, such as second-wave peaks.

5.7.3. Performance comparison

The table below summarizes the performance of the three models in terms of Root Mean Square Error (RMSE) and confidence interval calibration, defined as the percentage of true values falling within the 95% confidence band.

The results show that the LSTM model achieves the best accuracy and the most reliable uncertainty quantification. The GRU model offers a good trade-off between speed and precision, but suffers slightly in unstable or noisy series. The feedforward model is clearly less capable in capturing temporal patterns, but still provides reasonable performance in regular conditions.

These findings support the adoption of LSTM networks when forecasting complex time series in healthcare contexts, especially when the time horizon is long and the dynamics are non-linear. While simpler models may be sufficient for low-variability environments, they do not generalize well to unseen epidemic behaviors.

In addition, the LSTM model showed more stable uncertainty calibration, with narrower but better-aligned confidence intervals. This is important in critical settings, where over- or under-confidence

6. Conclusion

The work presented in this article proposes a forecasting method for ICU occupancy based on a combination of recurrent neural networks and approximate Bayesian inference techniques. By integrating Long Short-Term Memory (LSTM) models with Monte Carlo dropout, we were able to develop a system that does not limit itself to generating single-point predictions, but also estimates the uncertainty associated with each forecast. This approach represents a step forward in the context of healthcare, where decision-making often takes place under time pressure, with incomplete information and high social responsibility.

Through the analysis of data collected in California during the COVID-19 pandemic, we observed that the proposed model is capable of maintaining stable performance across counties with different demographic and epidemiological characteristics. The inclusion of uncertainty estimation allowed us to associate confidence intervals with each prediction, offering health managers and decision-makers an additional layer of interpretability and caution. This feature is especially important in a context where underestimating the future demand for critical care resources can lead to saturation and systemic failures, while overestimating it can cause inefficient allocation.

However, some limitations must be acknowledged. The data used for training and evaluation cover only the initial phase of the pandemic, limiting the model's ability to learn from multiple waves or long-term seasonal patterns. Moreover, our model operates at the level of aggregated time series, without incorporating individual clinical features that could enrich the prediction with information about the severity or progression of patients' conditions. This restricts the system's ability to adapt its forecasts to variations in population risk or to the evolution of treatment protocols. A further limitation concerns the spatial resolution of the dataset. Although county-level data are useful for regional planning, they are not always sufficient to guide decisions at the hospital level, where operational constraints and patient flow dynamics are far more detailed and localized.

In future developments, it would be desirable to integrate clinical variables directly into the time series modeling process, in order to capture not only the epidemiological evolution of the virus, but also the specific characteristics of the population affected. Furthermore, the model could be extended beyond COVID-19, adapting it to other types of health emergencies that generate sudden increases in ICU demand, such as influenza epidemics or extreme climatic events. Another direction could concern the introduction of spatial correlations between different geographic units, which would make it possible to simulate the redistribution of patients between hospitals or regions in case of local overload. Finally, the creation of an interactive forecasting tool, accessible in real time to public health authorities, could transform this model into an operational resource capable of supporting decision-making processes directly in the field.

The results of this study confirm that deep learning methods, when equipped with mechanisms for uncertainty estimation, can contribute concretely to making healthcare systems more resilient. In scenarios characterized by volatility and risk, the ability to predict with caution and to quantify doubt becomes just as important as the ability to predict with precision.

7. Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT, Grammarly in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] A. Remuzzi, G. Remuzzi, Covid-19 and italy: what next?, The Lancet 395 (2020) 1225–1228.
[2] M. L. Ranney, V. Griffeth, A. K. Jha, Critical supply shortages—the need for ventilators and personal protective equipment during the covid-19 pandemic, New England Journal of Medicine 382 (2020) e41.
[3] G. Grasselli, A. Zangrillo, A. Zanella, et al., Baseline characteristics and outcomes of 1591 patients infected with sars-cov-2 admitted to icus of the lombardy region, italy, JAMA 323 (2020) 1574–1581.
[4] L. Jiang, et al., Occupancy and ventilator use of intensive care units in new york state during the covid-19 pandemic, JAMA 324 (2020) 835–837.
[5] J. T. Wu, K. Leung, G. M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study, The Lancet 395 (2020) 689–697.
[6] A. J. Kucharski, T. W. Russell, C. Diamond, et al., Early dynamics of transmission and control of covid-19: a mathematical modelling study, The Lancet Infectious Diseases 20 (2020) 553–558.
[7] C. Krubiner, R. Faden, Covid-19 vaccine trials should seek worthwhile health gains for low-income countries, The Lancet 396 (2020) 741–743.
[8] R. C. Maves, et al., Crisis standards of care: Lessons from new york city hospitals during the covid-19 pandemic, Health Security 18 (2020) 447–452.
[9] L. Rosenbaum, Facing covid-19 in italy — ethics, logistics, and therapeutics on the epidemic's front line, New England Journal of Medicine 382 (2020) 1873–1875.
[10] S. L. Barber, et al., Covid-19 pandemic planning: considerations for state and territorial health officials, Journal of Public Health Management and Practice 26 (2020) E16–E20.

[11] E. Marseille, et al., Cost-effectiveness of intensive care for hospitalized covid-19 patients: experience from new york city, BMC Health Services Research 21 (2021) 1–9.
[12] M. Chinazzi, et al., The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak, Science 368 (2020) 395–400.
[13] G. Giordano, et al., Modelling the covid-19 epidemic and implementation of population-wide interventions in italy, Nature Medicine 26 (2020) 855–860.
[14] M. Abdar, et al., A review of uncertainty quantification in deep learning: Techniques, applications and challenges, Information Fusion 76 (2021) 243–297.
[15] A. Ghosh, et al., Impact of covid-19 pandemic on healthcare system and socio-economic structure of india, Sustainable Cities and Society 64 (2021) 102601.
[16] Y. Gal, Z. Ghahramani, Dropout as a bayesian approximation: Representing model uncertainty in deep learning, in: international conference on machine learning, PMLR, 2016, pp. 1050–1059.
[17] Y. Gal, Z. Ghahramani, A theoretically grounded application of dropout in recurrent neural networks, Advances in neural information processing systems 29 (2016) 1019–1027.
[18] A. Kendall, Y. Gal, What uncertainties do we need in bayesian deep learning for computer vision?, Advances in Neural Information Processing Systems 30 (2017).
[19] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Computation 9 (1997) 1735–1780.
[20] B. Lim, S. Zohren, Time-series forecasting with deep learning: a survey, Philosophical Transactions of the Royal Society A 379 (2021) 20200209.
[21] K. Bandara, C. Bergmeir, S. Smyl, Forecasting time series with lstm: A review, International Journal of Forecasting (2020).
[22] V. Chamola, V. Hassija, S. Gupta, A. Goyal, M. Guizani, B. Sikdar, Disaster and pandemic management using machine learning: A survey, IEEE Internet of Things Journal (2020).
[23] F.-Y. Cheng, H. Joshi, P. Tandon, R. Freeman, D. L. Reich, M. Mazumdar, R. Kohli-Seth, M. Levin, P. Timsina, A. Kia, Using machine learning to predict icu transfer in hospitalized covid-19 patients, Journal of clinical medicine 9 (2020) 1668.
[24] J. Ruyssinck, J. van der Herten, R. Houthooft, F. Ongenae, I. Couckuyt, B. Gadeyne, K. Colpaert, J. Decruyenaere, F. De Turck, T. Dhaene, Random survival forests for predicting the bed occupancy in the intensive care unit, Computational and mathematical methods in medicine 2016 (2016).
[25] X. Li, P. Ge, J. Zhu, H. Li, J. Graham, A. Singer, P. S. Richman, T. Q. Duong, Deep learning prediction of likelihood of icu admission and mortality in covid-19 patients using clinical variables, PeerJ 8 (2020) e10337.
[26] G. De Magistris, S. Russo, P. Roma, J. T. Starczewski, C. Napoli, An explainable fake news detector based on named entity recognition and stance classification applied to covid-19, Information (Switzerland) 13 (2022). doi:10.3390/info13030137.
[27] S. Russo, F. Fiani, C. Napoli, Remote eye movement desensitization and reprocessing treatment of long-covid- and post-covid-related traumatic disorders: An innovative approach, Brain Sciences 14 (2024). doi:10.3390/brainsci14121212.
[28] V. Ponzi, S. Russo, A. Wajda, R. Brociek, C. Napoli, Analysis pre and post covid-19 pandemic rorschach test data of using em algorithms and gmm models, in: CEUR Workshop Proceedings, volume 3360, 2022, pp. 55–63.
[29] G. De Magistris, E. Iacobelli, R. Brociek, C. Napoli, An automatic cnn-based face mask detection algorithm tested during the covid-19 pandemics, in: CEUR Workshop Proceedings, volume 3398, 2022, pp. 36–41.
[30] A. Graves, J. Schmidhuber, Framewise phoneme classification with bidirectional lstm and other neural network architectures, Neural Networks 18 (2005) 602–610.
[31] California Department of Public Health, Covid-19 hospital data, 2020. URL: https://data.chhs.ca.gov/dataset/covid-19-hospital-data.