Anticipating Tomorrow: Predicting Radiologist Case Volumes

University of California, Los Angeles

HMS_MGH_1

Pin-Han Chen Yi-An (Annie) Chen Seamus Gallivan Kuei-Chun Huang Jo-Hua Wu

1. Problem Definition

1.1. Challenges in medical imaging

In the domain of medical imaging, radiology holds a dominant position, offering deep insights into the health of patients. This field enables doctors to observe and decipher internal irregularities. Radiologists, positioned at the forefront of this effort, interpret a wide range of imaging scans, including Computed Tomography (CT) and Magnetic Resonance Imaging (MRI), as well as conventional X-rays. Their discerning analyses directly steer patient care decisions and shape treatment pathways.

Yet, as the pulse of healthcare demand varies, radiology departments face a pressing challenge: accurately forecasting patient exam volumes. This need stems from a desire to optimize healthcare delivery, but the stakes are multifaceted. Operational challenges emerge when unpredicted volumes lead to resource mismatches and equipment underutilization. Radiologists, caught offguard by fluctuating workloads, may grapple with increased fatigue and declining morale. Furthermore, patients bear the brunt of these inefficiencies, facing potential care delays that might hinder their diagnosis and treatment. From an institutional perspective, such property can strain financial resources and complicate reimbursement processes.

1.2. Radiology with Machine Learning

Understanding these challenges, the conceptual foundation of our study is rooted in healthcare resource optimization, honing in on radiology. We recognize the imperative of forecasting patient exam volumes, not only for the immediate future, but also up to a horizon of 3-4 days. This volume reflects both captured imaging scans and upcoming scheduled ones, influenced by diverse factors like institutional scheduling systems, historical trends, and even cyclical healthcare demands. The term "physician workload"

encapsulates the comprehensive responsibilities shouldered by radiologists, with a spotlight on the interpretation of diagnostic images.

Given the burgeoning emphasis on valuebased care, the timeliness and precision of these interpretations become even more paramount, impacting patient outcomes, institutional reputation, and reimbursement dynamics. With this backdrop, our method embarks on a mission to employ machine learning in crafting a predictive model adept at forecasting patient exam volumes in radiology. By integrating historical data with real-time scheduling intricacies, our model administrators equips healthcare and radiologists with predictive insights. facilitating well-informed planning and distribution. effective resource This convergence ultimately advances the pinnacle of patient care and operational aiding proficiency, administrators and radiologists in strategic decision-making and resource allocation, thereby ensuring the in patient care utmost quality and departmental efficiency.

2. Literature Review

As part of a literature review on radiology examination volume prediction we discovered three studies with particular relevance to our project.

2.1. Radiological Exam Volumes Prediction with Prophet Algorithm

In 2021, Becker, Chaim, Vargas *et al.* [1] evaluated the use of the Prophet algorithm, an

open-source Bayesian structural time series model proposed by Meta, to predict radiology examination volumes. They found that the algorithm captures weekly, seasonal, and overall trends to allow for better radiologist planning and labor allocation. While our project focuses on predicting one, two or three days in advance predictions, this paper details a useful open-source algorithm and promising results for weekly volume prediction.

2.2. Artificial Intelligence Predictive Analytics MRI Appointment Prediction

Chong, Tsai, and Lee *et al.* [2] used outpatient MRI appointments data to train a no-show prediction model. A predictive model developed with XGBoost, a decision tree-based ensemble machine learning algorithm that uses a gradient boosting framework, was deployed after various machine learning algorithms were evaluated. The simple intervention measure of using telephone call reminders for patients with the top 25% highest risk of an appointment noshow as predicted by the model was implemented over 6 months.

A key difference between this paper and our project problem definition is that this paper investigates predicting no-shows instead of examination volume. However, the model architectures, inputs, and outputs of the models are similar to those which will be evaluated in our project.

2.3. Patient Volume Prediction with Deep Learning and Statistical Models

In 2023, Pala et al. [3] proposed a new method to predict the multi-month patient examination volume. For prediction processes, both deep learning models such as LSTM, MLP, NNAR and ELM, as well as statistical based prediction models such as ARIMA, SES, TBATS, HOLT and THETAF were used. The results showed that the LSTM model outperformed the other models in estimating the monthly number of radiological case images. The prediction window for these models was evaluated at 12-month, 24-month, and 30-month intervals. The goal of this paper was similar to our project goal and the Prophet algorithm paper [1], but the prediction windows it uses were far larger than our anticipated project's windows.

In conclusion, the Prophet algorithm paper [1] most closely resembles our project's intended methods and prediction output. Thus, while the variety of our literature was valuable in assessing a variety of models and research goals, the Prophet algorithm paper [1] serves as the main foundation to our project's methods and evaluation.

3. Methods

3.1. Overview

For this study, we leveraged four primary datasets: A1, A2, T1, and T2. These datasets span a duration from 2027/07/06 to 2028/08/08, providing a comprehensive

overview of patient exam volumes during this period. Our exploration of the datasets began with an intent to decipher underlying patterns, potential anomalies, and inherent structures. The analysis was centered around the key temporal parameters:

- **T0**: Represents the order time.
- **T1**: Denotes the time when the task was added to the hospital system schedule.
- **T2**: Marks the scheduled time for interpretation.
- **T3**: Marks the task arrival time.
- **T4**: Signifies the actual commencement of the interpretation.
- **T5**: Indicates task completion, serving as our target variable.

3.2. Exploratory Data Analysis

3.2.1. Daily Distribution Insights

An inspection of daily distribution patterns for the years 2027 and 2028 revealed consistent trends across both years for all datasets.

Figure 3.2.1.1: Daily distribution for A1





* Since the daily distribution for all datasets are similar, the graphs for A2 and T2 are placed in the appendix for your reference.

3.2.2. Weekly Distribution Insights

All datasets demonstrated a higher volume of tasks during weekdays. Most of the exams were conducted during weekdays. Only small portions of them were done on weekends.





* Since the weekly distribution for all datasets are similar, the graphs for A2 and T2 are placed in the appendix for your reference.

3.2.3. Ratio of Different Days

Before the modeling, we want to inspect how many of the tasks are done before 1 or 7 days. This information helps us determine the accuracy of the predictions of our model.

First, we inspect the ratio of different days by at least 1 day, which should tell us how many propositions of the cases are useful for the model. Since the accuracy of prediction might be terrible if the majority of the cases are finished within 1 day.

• A1 and A2 datasets

For the A1 / A2 dataset, there are approximately **74%** / **77%** of the cases showing that T1 is **at least 1 day** before T2.

• T1 and T2 datasets

For the T1 / T2 dataset, there are approximately **52%** / **55%** of the cases showing that T1 is **at least 1 day** before T2.

Figure 3.2.3.1: A1 ratio of different days [1 day]

AI	то	Π	T2	тз	Т4	T5
то	-	56.03%	81.80%	82.10%	82.28%	82.41%
ті	0.00%	-	73.82%	74.23%	74.43%	74.61%
Т2	6.47%	6.56%	-	7.04%	7.28%	7.72%
Т3	0.00%	0.00%	0.03%	-	0.34%	0.88%
т4	0.00%	0.01%	0.03%	0.05%	-	0.55%
T5	0.00%	0.01%	0.01%	0.06%	0.00%	-

Figure 3.2.3.2: A2 ratio of different days [1 day]

A2	то	ті	T2	тз	Т4	T5
то	-	56.74%	84.58%	84.64%	84.63%	84.70%
ті	0.00%	-	76.57%	76.64%	76.65%	76.77%
T2	5.32%	5.35%	-	5.48%	5.50%	5.75%
тз	0.00%	0.00%	0.04%	-	0.04%	0.31%
Т4	0.00%	0.02%	0.04%	0.03%	-	0.27%
T5	0.00%	0.02%	0.02%	0.03%	0.00%	-

Figure 3.2.3.3: T1 ratio of different days [1 day]

ті	то	Π	T2	тз	T 4	T5
то	-	47.88%	63.25%	63.47%	63.54%	63.69%
ті	0.00%	-	51.43%	51.65%	51.74%	51.93%
T2	4.62%	4.73%	-	5.12%	5.23%	5.74%
Т3	0.00%	0.00%	0.07%	-	0.15%	0.74%
T4	0.00%	0.01%	0.05%	0.03%	-	0.59%
	0.00%	0.02%	0.04%	0.04%	0.01%	-

Figure 3.2.3.4: T2 ratio of different days [1 day]

Т2	то	ті	T2	тз	Т4	T5
то	-	49.07%	67.93%	67.99%	68.03%	68.19%
ті	0.00%	-	55.49%	55.55%	55.62%	55.88%
Т2	4.95%	5.03%	-	5.18%	5.24%	5.68%
тз	0.00%	0.01%	0.09%	-	0.12%	0.65%
T4	0.00%	0.05%	0.07%	0.05%	-	0.54%
Т5	0.00%	0.06%	0.02%	0.06%	0.01%	-

Second, we inspect the ratio of different days by at least 7 days as well. It could give us a glance of the distribution for the days, which also helps us evaluate the effect of outliers.

• A1 and A2 datasets

For the A1 / A2 dataset, there are approximately **55%** / **56%** of the

cases showing that T1 is **at least 1 day** before T2.

- T1 and T2 datasets
 - For the T1 / T2 dataset, there are approximately **36%** / **42%** of the cases showing that T1 is **at least 1 day** before T2.

Figure 3.2.3.5: A1 ratio of different days [7 days]

Al	то	Π	T2	тз	Т4	Т5
то	-	30.60%	64.09%	64.11%	64.10%	64.11%
ті	0.00%	-	54.50%	54.52%	54.50%	54.53%
Т2	6.47%	6.56%	-	5.11%	5.11%	5.14%
тз	0.00%	0.00%	0.03%	-	0.01%	0.03%
T4	0.00%	0.01%	0.03%	0.05%	-	0.03%
Т5	0.00%	0.01%	0.01%	0.06%	0.00%	-

Figure 3.2.3.6: A2 ratio of different days [7 days]

A2	то	ті	T2	тз	Т4	
то	-	29.95%	65.14%	65.14%	65.14%	65.14%
ті	0.00%	-	55.58%	55.59%	55.58%	55.59%
Т2	5.32%	5.35%	-	4.42%	4.42%	4.42%
тз	0.00%	0.00%	0.04%	-	0.00%	0.01%
Т4	0.00%	0.02%	0.04%	0.03%	-	0.01%
T5	0.00%	0.02%	0.02%	0.03%	0.00%	-

Figure 3.2.3.7: T1 ratio of different days [7 days]

ті	то	Π	T2	тз	T4	
то	-	29.12%	46.25%	46.25%	46.25%	46.27%
ті	0.00%	-	35.43%	35.43%	35.43%	35.46%
Т2	4.62%	4.73%	-	3.51%	3.51%	3.55%
тз	0.00%	0.00%	0.07%	-	0.00%	0.05%
T4	0.00%	0.01%	0.05%	0.03%	-	0.05%
Т5	0.00%	0.02%	0.04%	0.04%	0.01%	-

Figure 3.2.3.8: T2 ratio of different days [7 days]

T2	то	ті	T2	Т3	т4	T5
то	-	27.90%	51.37%	51.37%	51.37%	51.38%
ті	0.00%	-	41.97%	41.97%	41.97%	41.97%
Т2	4.95%	5.03%	-	3.62%	3.62%	3.63%
тз	0.00%	0.01%	0.09%	-	0.00%	0.01%
T4	0.00%	0.05%	0.07%	0.05%	-	0.01%
Т5	0.00%	0.06%	0.02%	0.06%	0.01%	-

3.3. Model Architecture and Training

3.3.1. Model

After comprehensive review а of algorithms, LightGBM potential the algorithm was identified as the most suitable model for this study. LightGBM, a gradient boosting framework, has gained recognition in the data science community for its efficiency in handling voluminous datasets and its robustness against overfitting. This gradient boosting framework, specifically designed for speed and performance, was tailored in two distinctive variants for our modeling endeavor:

• Standard Models:

We implement both a regression model and a classification model to yield optimal prediction results:

• LightGBM Regressor

Predict the upcoming volume for the next 1 to 7 days; $max_depth = 8$

• LightGBM Classifier

Predict whether the next 1 to 7 days will be a *busy day* or not; $max_depth = 8$

max_depth: hyperparameter, maximum distance between the root node of each tree and a leaf node. *busy day*: a day with top 50% highest volumes

• Advanced Models:

To obtain more accurate predictions for the busiest days, we add two advancements for our models:

• Weighted Regressor

We incorporate the target value (T5 counts) as a sample weighted to the model, which should generate an increased precision in the tasks which have higher volumes.

• Ensemble Model

We develop a formula to ensemble the predicted result from the regressor (*y_pred*) and classifier (*p*).

 $y_pred * (1 + (p - 0.5) * 5\%)$ where y_pred is the predicted volume from the regressor, pis the predicted probability from the classifier

3.3.2. Features

The essence of any predictive model's effectiveness lies in the information it takes in. In our project, we combined past data with time-related characteristics to extract and formulate:

- 1. Normalized T5 counts, offering a glance into the past week's trends.
- 2. The weekday of the target date, capturing cyclic weekly patterns.
- 3. The month of the target date, encapsulating broader monthly fluctuations.

4. Scheduled exam counts for the target date, directly influencing the tasks' progression.

Normalized T5 Counts serve as a beacon into recent trends, offering the model a glance at the past week's rhythms. By normalizing the completion counts from the preceding seven days, the model is equipped to discern short-term patterns, ensuring it remains sensitive to immediate changes or anomalies in task completions.

Simultaneously, **the Weekday of the Target Date** infuses the model with an understanding of cyclic weekly patterns. Radiology tasks, akin to many hospital operations, exhibit a rhythmic ebb and flow throughout the week. Whether it's a surge due to scheduled procedures or a quiet day, this feature ensures the model is attuned to these cyclical nuances.

Broadening the temporal horizon, the **Month of the Target Date** captures overarching patterns that play out over longer time frames. Whether it's seasonal health trends, hospital administrative decisions, or equipment maintenance schedules, this feature positions the model to anticipate broader monthly fluctuations.

Lastly, the **Scheduled Exam Counts for the Target Date** act as a direct barometer of the tasks' progression. As the number of scheduled exams for a day swells or recedes, it directly impacts the volume of radiology tasks. This feature provides the model with a real-time pulse of the day's potential workload, allowing for informed and accurate forecasts.

3.3.3. Procedure

Our training paradigm was underpinned by an accumulated rolling window strategy. In this approach, the model was trained on a continually expanding window of data, initiating predictions from the 31st day onwards. This temporal consistency in the training strategy ensures that the model is continually updated with the most recent data, thus enhancing its predictive accuracy.

3.4. Evaluation Metrics

To gauge the model's performance, we embraced a triad of evaluation metrics:

• Mean Absolute Error (MAE):

A straightforward metric, MAE quantifies the average magnitude of errors between the predicted and actual values, offering an unambiguous view of the model's accuracy.

• Mean Absolute Percentage Error (MAPE): Providing a relative perspective,

MAPE computes the prediction error as a percentage, offering insights into the model's accuracy in relative terms.

• Symmetric Mean Absolute Percentage Error (SMAPE): An evolution of MAPE, SMAPE corrects its inherent asymmetry, ensuring uniform treatment of errors, irrespective of direction.

4. Analysis

4.1. Standard Model vs. Weighted Model

4.1.1. A1 Dataset

Implementing the LightGBM regressor model trained with an accumulating rolling window with or without weighted loss rendered promising results for predicting the next 1st ~ 7th days. SMAPE results all fell within 7.30% ~ 9.09% Notably, the performance of the top 10% cases already performed better without weighted loss.

We found that with weighted loss, the performance for top 10% cases improved but overall performance would be slightly worsened.

Figure 4.1.1.1: A1 all data results [weighted]

All	MAE		MAPE (%)		SMAPE (%)	
Weighted Loss	N	Y	N	Y	N	Y
Day=1	21.51	23.01	8.26	9.81	7.30	7.92
Day=2	23.99	25.07	9.05	10.28	8.06	8.48
Day=3	23.79	25.01	8.90	10.46	7.88	8.41
Day=4	24.34	25.07	8.95	10.46	8.10	8.50
Day=5	24.25	26.24	9.14	10.81	8.19	8.89
Day=6	25.45	26.50	9.67	11.15	8.65	8.98
Day=7	24.61	26.61	9.35	11.22	8.35	9.09

Figure 4.1.1.2: A1	top 109	6 data results	[weighted]
--------------------	---------	----------------	------------

Тор 10%	M	٩E	МАР	E (%)	SMAF	PE (%)
Weighted Loss	N	Y	N	Y	N	Y
Day=1	21.36	20.51	4.14	3.97	4.28	4.12
Day=2	26.31	27.24	5.10	5.29	5.30	5.51
Day=3	27.57	26.88	5.34	5.21	5.55	5.40
Day=4	28.09	26.30	5.44	5.09	5.66	5.29
Day=5	27.77	25.68	5.39	4.98	5.61	5.18
Day=6	27.41	26.02	5.31	5.04	5.53	5.25
Day=7	27.19	26.02	5.27	5.05	5.47	5.25

4.1.2. T1 Dataset

Results for implementing the model on the T1 data were similar to those of the A1 data model results. With or Without weighted loss, for predicting the next 1st ~ 7th day, SMAPEs all fell within $6.56\% \sim 7.99\%$ With weighted loss, the performance for top 10% cases were improved but overall performance was slightly worsened.

Figure 4.1.2.1: T1 all data results [weighted]

All	M	٩E	МАР	E (%)	SMAF	PE (%)
Weighted Loss	N	Y	N	Y	N	Y
Day=1	25.85	26.73	6.81	7.29	6.56	6.81
Day=2	29.94	30.32	7.63	8.03	7.44	7.60
Day=3	30.69	29.66	7.75	7.74	7.53	7.35
Day=4	31.36	31.21	7.92	8.10	7.74	7.73
Day=5	29.65	30.94	7.62	8.23	7.44	7.79
Day=6	29.01	30.95	7.47	8.23	7.25	7.78
Day=7	29.85	31.73	7.67	8.44	7.46	7.99

Figure 4.1.2.2: T1	top 10%	data results	[weighted]
			L

-		-	-				
Тор 10%	м	AE	МАР	MAPE (%)		SMAPE (%)	
Weighted Loss	N	Y	N	Y	N	Y	
Day=1	36.62	35.32	6.22	5.60	6.49	6.25	
Day=2	42.08	40.91	7.14	6.93	7.51	7.28	
Day=3	43.97	41.80	7.46	7.09	7.86	7.45	
Day=4	43.86	43.24	7.43	7.33	7.83	7.71	
Day=5	41.03	40.04	6.96	6.79	7.29	7.10	
Day=6	40.14	42.15	6.81	7.15	7.13	7.52	
Day=7	40.52	40.61	6.88	6.90	7.20	7.24	

4.1.3. A2 Dataset

On the A2 data, predicting the next 1st ~ 7th days returned SMAPEs all within 8.40% ~ 11.8% With weighted loss, both the performance of top 10% cases and overall performance worsened.

Figure 4.1.3.1: A2 all data results [weighted]

All	м	٩E	МАР	E (%)	SMAF	PE (%)
Weighted Loss	N	Y	N	Y	N	Y
Day=1	17.44	20.98	10.78	15.01	8.40	10.42
Day=2	18.88	22.94	11.27	15.79	8.93	11.18
Day=3	18.89	22.66	11.06	15.72	8.91	11.02
Day=4	19.44	22.20	11.21	15.46	9.17	10.85
Day=5	19.66	22.97	11.77	15.96	9.53	11.18
Day=6	20.54	24.60	12.03	17.43	9.81	12.07
Day=7	20.80	24.21	12.11	17.21	10.03	11.80

Figure 4.1.3.2: A2 top 10% data results [weighted]

Тор 10%	м	٩E	МАР	E (%)	SMAF	PE (%)
Weighted Loss	N	Y	N	Y	N	Y
Day=1	21.52	22.10	5.11	5.25	5.40	5.62
Day=2	24.36	24.42	5.79	5.81	6.09	6.20
Day=3	25.19	24.16	5.99	5.74	6.30	6.03
Day=4	26.12	25.42	6.20	6.04	6.49	6.32
Day=5	26.09	26.47	6.19	6.30	6.52	6.64
Day=6	24.83	25.85	5.90	6.14	6.20	6.53
Day=7	25.76	26.44	6.11	6.28	6.43	6.66

4.1.4. T2 Dataset

Results for the T2 data were very similar to those of T1. Predicting the next 1st ~ 7th days returned SMAPEs all within 7.43% ~ 8.92%, with or without weighted loss. With weighted loss, the performance for top 10% cases were improved but overall performance was slightly worsened.

Figure 4.1.4.1:	T2 all	data results	[weighted]
-----------------	--------	--------------	------------

All	м	٩E	МАР	E (%)	SMAR	PE (%)
Weighted Loss	N	Y	N	Y	N	Y
Day=1	26.45	29.64	7.98	9.51	7.43	8.37
Day=2	28.83	30.82	8.59	9.92	8.08	8.74
Day=3	26.73	30.52	7.93	9.81	7.50	8.63
Day=4	27.45	29.75	8.37	9.76	7.77	8.59
Day=5	27.88	30.99	8.46	10.03	7.92	8.86
Day=6	27.11	31.30	8.10	10.03	7.62	8.92
Day=7	27.49	30.50	8.05	9.71	7.66	8.77

Figure 4.1.4.2: T2 top 10% data results [weighted]

Top 10%	M	٩E	МАР	E (%)	SMAF	PE (%)
Weighted Loss	N	Y	N	Y	N	Y
Day=1	38.07	35.93	6.64	6.27	6.95	6.55
Day=2	50.20	48.59	8.78	8.49	9.26	8.95
Day=3	47.82	47.72	8.36	8.34	8.82	8.80
Day=4	47.03	45.17	8.21	7.89	8.65	8.30
Day=5	47.90	47.45	8.38	8.30	8.83	8.75
Day=6	47.82	48.81	8.36	8.54	8.80	9.01
Day=7	50.85	47.49	8.90	8.31	9.39	8.75

4.2. Standard Model vs. Ensemble Model

4.2.1. A1, T1, and A2 Datasets

For A1, T1, and A2 datasets, we found that the performance for top 10% cases were improved by using ensemble method. But the overall performance would be slightly worsened.

It was what we expected since the ensemble learning method would intensify the prediction of data with high weight. Therefore, the model would perform better on those busiest days (top 10% cases) but lose the accuracy for overall prediction.

All	MAE		МАР	MAPE (%)		SMAPE (%)	
Ensembled	N	Y	N	Y	N	Y	
Day=1	21.51	22.77	8.26	8.50	7.30	7.62	
Day=2	23.99	24.69	9.05	9.17	8.06	8.26	
Day=3	23.79	24.43	8.90	9.05	7.88	8.09	
Day=4	24.34	25.26	8.95	9.15	8.10	8.37	
Day=5	24.25	25.39	9.14	9.40	8.19	8.54	
Day=6	25.45	27.01	9.67	10.02	8.65	9.08	
Day=7	24.61	25.98	9.35	9.66	8.35	8.74	

Figure 4.2.1.1: A1 all data results [ensemble]

Figure 4.2.1.2: A1 top 10% data results [ensemble]

Тор 10%	м	٩E	МАР	E (%)	SMAF	PE (%)
Ensembled	N	Y	N	Y	N	Y
Day=1	21.36	16.07	4.14	3.11	4.28	3.18
Day=2	26.31	18.86	5.10	3.65	5.30	3.75
Day=3	27.57	20.23	5.34	3.91	5.55	4.02
Day=4	28.09	21.65	5.44	4.19	5.66	4.30
Day=5	27.77	20.30	5.39	3.94	5.61	4.05
Day=6	27.41	20.72	5.31	4.02	5.53	4.13
Day=7	27.19	19.40	5.27	3.76	5.47	3.86

Figure 4.2.1.3: T1 all data results [ensemble]

All	M	AE	МАР	E (%)	SMAF	PE (%)
Ensembled	N	Y	N	Y	N	Y
Day=1	25.85	26.53	6.81	6.95	6.56	6.74
Day=2	29.94	30.77	7.63	7.84	7.44	7.68
Day=3	30.69	30.78	7.75	7.83	7.53	7.63
Day=4	31.36	31.93	7.92	8.08	7.74	7.93
Day=5	29.65	30.15	7.62	7.75	7.44	7.61
Day=6	29.01	29.74	7.47	7.67	7.25	7.49
Day=7	29.85	30.31	7.67	7.82	7.46	7.65

Figure 4.2.1.4. T1	top 109	6 data results	[ensemble]
1 15010 1.2.1. 1. 11	top 10/	o aata rebaito	[ensemole]

Тор 10%	М	AE	МАР	E (%)	SMAF	PE (%)
Ensembled	N	Y	N	Y	N	Y
Day=1	36.62	26.80	6.22	4.55	6.49	4.70
Day=2	42.08	32.22	7.14	5.46	7.51	5.70
Day=3	43.97	33.63	7.46	5.69	7.86	5.96
Day=4	43.86	33.36	7.43	5.64	7.83	5.90
Day=5	41.03	30.15	6.96	5.10	7.29	5.30
Day=6	40.14	29.27	6.81	4.96	7.13	5.15
Day=7	40.52	29.46	6.88	4.99	7.20	5.19

* Since the results for A1, T1, and A2 datasets are similar, we leave the table for dataset A2 in the appendix for your reference.

4.2.2. T2 Dataset

For the T2 dataset, we still found that the performance for the top 10% cases were improved by using ensemble method. But some predictions in all cases even have better performance when using ensemble method.

All	м	AE	MAPE (%)		SMAPE (%)	
Ensembled	N	Y	N	Y	N	Y
Day=1	26.45	26.59	7.98	7.99	7.43	7.49
Day=2	28.83	28.21	8.59	8.51	8.08	8.05
Day=3	26.73	26.38	7.93	7.90	7.50	7.52
Day=4	27.45	27.03	8.37	8.30	7.77	7.75
Day=5	27.88	27.56	8.46	8.38	7.92	7.90
Day=6	27.11	26.54	8.10	7.96	7.62	7.53
Day=7	27.49	26.99	8.05	7.97	7.66	7.62

•	•				-	
Top 10%	м	٩E	MAPE (%)		SMAPE (%)	
Ensembled	N	Y	N	Y	N	Y
Day=1	38.07	26.93	6.64	4.69	6.95	4.86
Day=2	50.20	38.78	8.78	6.78	9.26	7.07
Day=3	47.82	36.81	8.36	6.43	8.82	6.71
Day=4	47.03	35.96	8.21	6.26	8.65	6.54
Day=5	47.90	37.25	8.38	6.50	8.83	6.79
Day=6	47.82	37.19	8.36	6.49	8.80	6.79
Day=7	50.85	39.59	8.90	6.92	9.39	7.23

Figure 4.2.2.2: T2 top 10% data results [ensemble]

4.3. Weighted Model vs. Ensemble Model

4.3.1. All Datasets

For All datasets, we found that the performance was much better on both the top 10% cases and all cases when using ensemble method. It was expected since the ensemble learning method would improve the accuracy by aggregating the regressor prediction and classifier prediction.

SMAPE (%)	All		Top 10%	
Method	W E		w	E
Day=1	7.92	7.62	4.12	3.18
Day=2	8.48	8.26	5.51	3.75
Day=3	8.41	8.09	5.40	4.02
Day=4	8.50	8.37	5.29	4.30
Day=5	8.89	8.54	5.18	4.05
Day=6	8.98	9.08	5.25	4.13
Day=7	9.09	8.74	5.25	3.86

Figure 4.3.1: A1 comparison results [W vs. E]

igure 4.3.2: T1 comparison results [W vs. E]

SMAPE (%)	All		Тор	10%
Method	W E		w	E
Day=1	6.81	6.74	6.25	4.70
Day=2	7.60	7.68	7.28	5.70
Day=3	7.35	7.63	7.45	5.96
Day=4	7.73	7.93	7.71	5.90
Day=5	7.79	7.61	7.10	5.30
Day=6	7.78	7.49	7.52	5.15
Day=7	7.99	7.65	7.24	5.19

* Since the results for A2 and T2 datasets are similar, we leave the table for A2 and T2 datasets in the appendix for your reference.

4.4. Summary

Considering the goal of correctly predicting radiology appointments with little notice (ideally correctly predicting the number of visits 1-7 days in advance), the results of our LightGBM model are very promising.

For all cases, predicting the next 1st ~ 7th days from our ensemble model returned SMAPEs all less than **10%**

For the top 10% cases (busiest days), predicting the next 1st ~ 7th days from our ensemble model returned SMAPEs all less than **7.5%**

This intriguing experiment result opens up an avenue for us to assist the radiologist schedule healthcare resources effectively in the future.

5. Findings and Recommendations

5.1. Findings

From our experiments, we find that there are several things can be referred for future endeavor:

5.1.1. General findings

- Influential features:
 - **Counts of scheduled exams on target date**: It's expected since the counts of scheduled exams on target date is the most direct factor to impact our prediction.

Counts of ending exams in the previous 7 days: It's expected that the counts of ending exams in the previous 7 days also impact our models significantly. It might be related to the period of follow up appointments and the routine schedule of the exams.

- Basic regressor can provide us with good enough predictions.
- The interpretability of the basic classifier is enough to predict whether a day is busy or not accurately.

5.1.2. Findings with weighted models

- After applying sample weight, the predictions for the top 10% cases in A1, T1, and T2 datasets were improved.
- The predictions for the A2 dataset were worsened instead.
- The predictions of the top 10% cases were worse than all cases for the T2 dataset.

5.1.3. Findings with ensemble models

- For all datasets, the predictions of the top 10% cases were improved significantly.
- In contrast, the performance of all cases for all datasets dropped slightly.

5.2. Recommendations

In conclusion, we provide several steps for the future researchers to do, which should advance the performance of models if needed.

- Include more features
- Solve cold-start issue
- Optimize the threshold of the classifier
- Optimize the parameters of ensemble models

6. References

[1] Anton S. Becker, Joshup Chaim, and H.
Alberto Vargas *et al.* "Automatic
Forecasting of Radiology Examination
Volume Trends for Optimal Resource
Planning and Allocation" (2021).
<u>https://link.springer.com/article/10.1007/s10</u>
<u>278-021-00532-4</u>

[2] Le R. Chong, Koh T. Tsai, and Lee L. Lee *et al.* "Artificial Intelligence Predictive Analytics in the Management of Outpatient MRI Appointment No-Shows" (2020). https://pubmed.ncbi.nlm.nih.gov/32901567/

[3] Z. Pala *et al.* "Forecasting Future Monthly Patient Volume using Deep Learning and Statistical Models" (2023). https://link.springer.com/article/10.1007/s11 277-023-10341-3

7. Appendix

7.1. Other graphs for Exploratory Data Analysis



Figure 3.2.1.4: Daily distribution for T2



Figure 3.2.2.3: Weekly distribution for A2





7.2. Other graphs for Standard Model vs. Ensemble Model

Figure 4.2.1.5: A2 all data results [ensemble]

All	м	٩E	МАР	E (%)	SMAR	PE (%)
Ensembled	N	Y	N	Y	N	Y
Day=1	17.44	18.27	10.78	10.93	8.40	8.66
Day=2	18.88	19.20	11.27	11.28	8.93	9.05
Day=3	18.89	19.19	11.06	11.09	8.91	9.04
Day=4	19.44	19.90	11.21	11.27	9.17	9.34
Day=5	19.66	20.25	11.77	11.84	9.53	9.73
Day=6	20.54	21.40	12.03	12.14	9.81	10.04
Day=7	20.80	21.33	12.11	12.14	10.03	10.20



Тор 10%	М	AE	MAPE (%)		SMAPE (%)	
Ensembled	N	Y	N	Y	N	Y
Day=1	21.52	14.74	5.11	3.51	5.40	3.70
Day=2	24.36	16.23	5.79	3.86	6.09	4.05
Day=3	25.19	16.29	5.99	3.87	6.30	4.07
Day=4	26.12	17.35	6.20	4.11	6.49	4.29
Day=5	26.09	18.23	6.19	4.32	6.52	4.54
Day=6	24.83	17.13	5.90	4.07	6.20	4.28
Day=7	25.76	18.03	6.11	4.28	6.43	4.48

7.2. Other graphs for Weighted Model vs. Ensemble Model

SMAPE (%)	All		Top 10%	
Method	W E		w	E
Day=1	10.42	8.66	5.62	3.70
Day=2	11.18	9.05	6.20	4.05
Day=3	11.02	9.04	6.03	4.07
Day=4	10.85	9.34	6.32	4.29
Day=5	11.18	9.73	6.64	4.54
Day=6	12.07	10.04	6.53	4.28
Day=7	11.80	10.20	6.66	4.48

Figure 4.3.3: A2 comparison results [W vs. E]

Figure 4.3.4: T2 comparison results [W vs. E]

SMAPE (%)	All		Тор	10%
Method	W E		w	E
Day=1	8.37	7.49	6.55	4.86
Day=2	8.74	8.05	8.95	7.07
Day=3	8.63	7.52	8.80	6.71
Day=4	8.59	7.75	8.30	6.54
Day=5	8.86	7.90	8.75	6.79
Day=6	8.92	7.53	9.01	6.79
Day=7	8.77	7.62	8.75	7.23

7.4. The prediction curve of top 10% cases for each dataset

7.4.1. A1 dataset

Figure 7.4.1.1: A1 predictions for top 10% cases with



Figure 7.4.1.2: A1 predictions for top 10% cases with weighted model



Figure 7.4.1.3: A1 predictions for top 10% cases with ensemble model



7.4.2. A2 dataset





Figure 7.4.2.2: A2 predictions for top 10% cases with weighted model



Figure 7.4.2.3: A2 predictions for top 10% cases with ensemble model



7.4.3. T1 dataset

Figure 7.4.3.1: T1 predictions for top 10% cases with standard model



Figure 7.4.3.2: T1 predictions for top 10% cases with



Figure 7.4.3.3: T1 predictions for top 10% cases with ensemble model



7.4.4. T2 dataset





Figure 7.4.4.2: T2 predictions for top 10% cases with weighted model



Figure 7.4.4.3: T2 predictions for top 10% cases with ensemble model

