

RetCare: Towards Interpretable Clinical Decision Making through LLM-Driven Medical Knowledge Retrieval



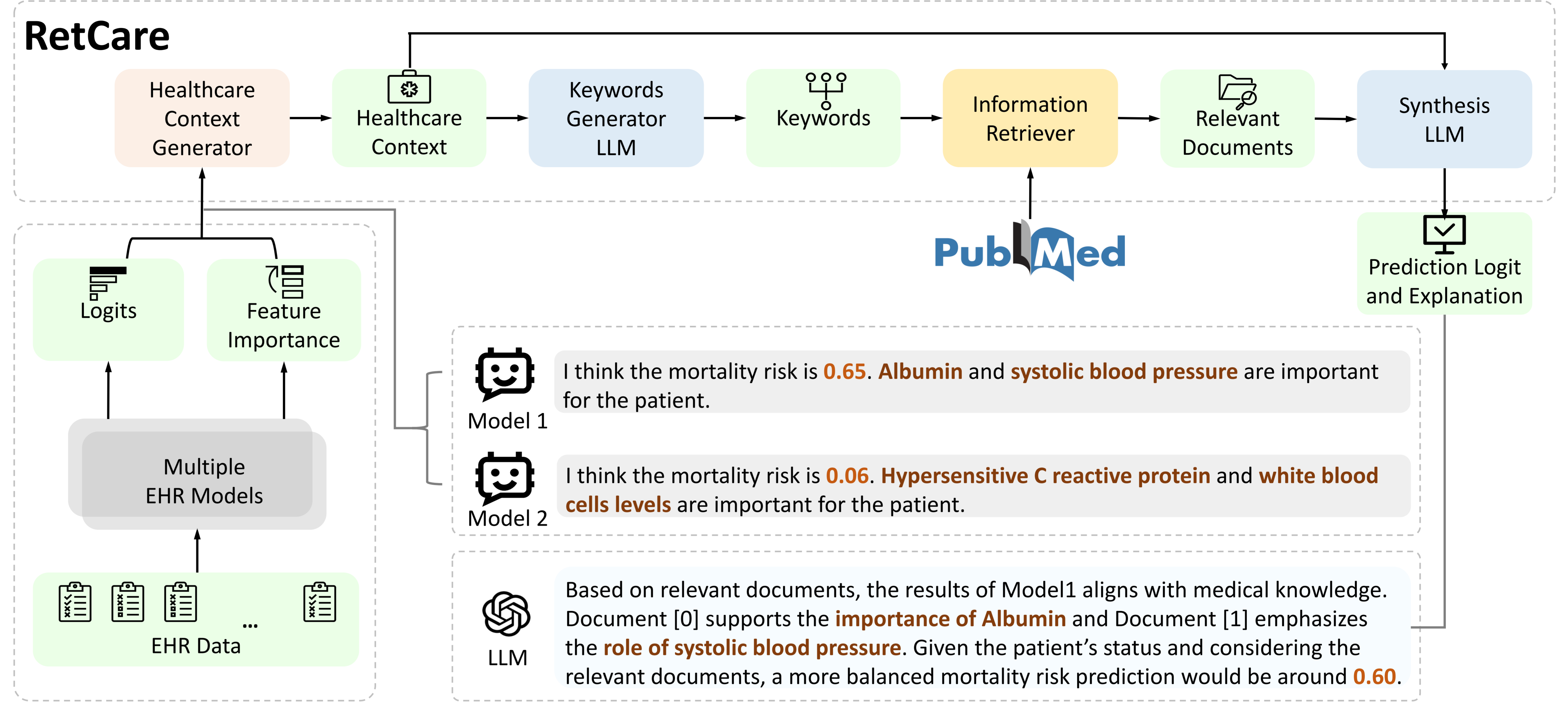
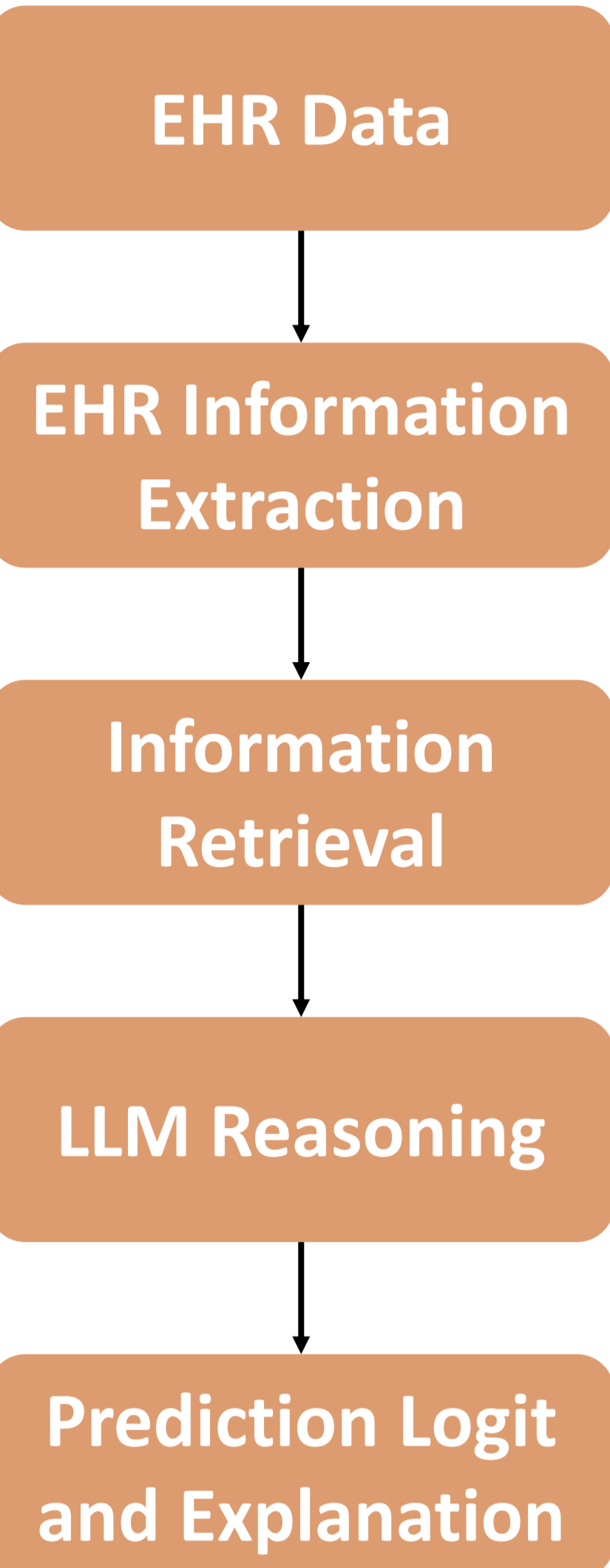
GitHub Code

Zixiang Wang*, Yinghao Zhu*, Junyi Gao, Xiaochen Zheng, Yuhui Zeng, Yifan He, Bowen Jiang, Wen Tang, Ewen M. Harrison, Chengwei Pan, Liantao Ma†, Ling Wang†



Challenge: How to enhance the **interpretability** and **trustworthiness** of AI model outputs for clinicians?
Objective: Develop an AI workflow that provides clinicians with **interpretable** and **reliable decision-making support** that aligns with their **clinical reasoning processes**.

Motivation



Pipeline

Datasets

- Dataset 1 (Peritoneal Dialysis (PD)):**
 - 656 patients, 13,091 records, 24.84% (163) mortality rate
- Dataset 2 (Covid Data Save Lives (CDSL)):**
 - 4,255 patients, 42,204 records, 45.98% (540) mortality rate
- PubMed:**
 - Over two million medical literature titles and abstracts

Models

- Machine learning model:** Logistic Regression
- EHR-specific predictive model:** ConCare
- Large language model:** GPT-4o
- Retriever model:** MedCPT

Methods	PD			
	ACC. (↑)	AUPRC (↑)	AUROC(↑)	F1 (↑)
LR	0.75	0.90	0.87	0.67
ConCare	0.75	0.87	0.85	0.76
RetCare (LR + Concare)	0.90	0.94	0.92	0.89

Methods	CDSL			
	ACC. (↑)	AUPRC (↑)	AUROC(↑)	F1 (↑)
LR	0.60	0.89	0.87	0.43
ConCare	0.70	0.89	0.90	0.67
RetCare (LR + Concare)	0.80	0.88	0.89	0.80

Conclusion: RetCare achieves the **highest accuracy and F1** across both datasets, underscoring the potential for **practical application in clinical decision-making scenarios**.

Experiments

Case Study

Role
You are an expert and experienced from the healthcare and biomedical domain with extensive medical knowledge and practical experience. I will provide you with **electronic health data** of an **End-Stage Renal Disease (ESRD)** patient and some analysis results including **mortality risk** and **feature importance weights** from our models. Your task is to analyze if the model's analysis results are reasonable **using the relevant documents**.

Relevant Documents
Document[0]: (Title: Can we predict when to start renal replacement therapy in patients with chronic kidney disease using 6 months of clinical data?)

Healthcare Context
This **male** patient, **aged 70.6**, is an **End-Stage Renal Disease(ESRD)** patient with original disease **Chronic glomerulonephritis**, and basic disease **Lung infection**.
Here is complete medical information from multiple visits of a patient:
- Blood chlorine: "109.0, 107.0, 103.0, ..., 102.0"
.....
We have 2 models LR, ConCare to predict the mortality risk and estimate the feature importance weight for the patient:
The mortality prediction risk for the patient from ConCare model is **0.4** out of 1.0, which means the patient is at a **low** level of death risk. Our model especially pays great attention to following features: Hemoglobin: **64.0 g/L**, with importance weight of **0.348** out of 1.0. The average value is 114.98 g/L for survival patients (**lower than the average** value by 44.0%) ...
.....
The AI model has found similar patients to the patient, including:
Patient 1: **male, 29.46** years old, with original disease **Chronic glomerulonephritis, recover and discharge**.
.....

Output Indicator
Please analyze by following the template below based on relevant documents:
Primary Summary
Please describe your task, summarize the patient's basic information and health status, and restate the AI model's prediction results along with the feature importance and partial statistic information.
Important Features Analysis
Please analyze the important features identified by the model, determine whether the identified features are reasonable. If they are reasonable, provide and cite relevant literature, include quotations from the sources, and explain the reasoning. If the features are not reasonable, provide and rank important features in your analysis, cite relevant documents and explain. Ensure that the features you identify are present in the dataset.
Prediction Analysis
Please evaluate the AI model's prediction result: too low, too high, or reasonable? If it's not reasonable, please provide your own prediction results, represented as a float number between 0 and 1.

Primary Summary
My task is to analyze the electronic health data of a **70.6-year-old male End-Stage Renal Disease (ESRD)** patient with **Chronic Glomerulonephritis** and a basic condition of **lung infection**. The AI models, LR and ConCare, predict the patient's mortality risk at **0.29 and 0.4**, respectively, indicating a **low** level of death risk. The LR model emphasizes **blood chlorine** with a SHAP value of -0.155, while the ConCare model highlights **hemoglobin** (64.0 g/L), **hypersensitive C-reactive protein** (0.78 mg/L), and **diastolic blood pressure** (56.0 mmHg) as important features.

Important Features Analysis
Blood Chlorine
• Relevance: Blood chlorine levels are **within the normal range, slightly higher** than the average for surviving ESRD patients (102.0 mmol/L vs. 98.21 mmol/L).
• Document Support: Documents do not specifically mention blood chlorine as a significant predictor for CKD progression or mortality in ESRD patients.
• Conclusion: The importance of blood chlorine in the LR model is **not strongly supported** by the provided documents.
Hemoglobin
• Relevance: Hemoglobin levels are **critically low** (64.0 g/L), which is **significantly below the average** for both surviving and deceased ESRD patients.
• Document Support: Document[4] and Document[10] highlight that low hemoglobin is a significant predictor of mortality in hemodialysis patients. Document[4] states, "**Low hemoglobin levels are associated with increased mortality in ESRD patients.**"
• Conclusion: Hemoglobin is a **reasonable and well-supported** feature in predicting mortality.
Hypersensitive C-Reactive Protein (CRP)
• Relevance: The patient's CRP level (0.78 mg/L) is **within the normal range** but **lower than the average** for both surviving and deceased ESRD patients.
• Document Support: Document[13] mentions inflammation markers, including CRP, as significant predictors of mortality in hemodialysis patients. However, lower CRP levels typically indicate better outcomes, which may **conflict with the model's weight**.
• Conclusion: The importance of CRP in ConCare's model is **partially supported**, but the low value suggesting lower risk is reasonable.

Prediction Results Analysis
Revised Prediction
The LR model's mortality prediction risk of **0.29 seems too low**, given the patient's critical low hemoglobin and diastolic blood pressure. The ConCare model's prediction of **0.4 is more plausible** but still appears to **underestimate** the risk considering the patient's severe anemia and hypotension. Based on the analysis of important features such as hemoglobin, diastolic blood pressure, and serum albumin, a more reasonable mortality risk prediction would be **around 0.6**.
Conclusion
While the AI models highlight relevant features, their predictions slightly underestimate the patient's mortality risk. The revised prediction better aligns with the clinical data and literature.