

# 利用注意力時序機制滾動式預測急性腎衰竭風險

## Rolling prediction of acute kidney injury using Hierarchical attention mechanism and Time series model

指導教授: 蔣榮先

專題成員: 郭冠良、蔡宇暘

開發工具: Anaconda、Google Colab

Django

測試環境: Python 3.7

### (一)、動機與目標

根據美國腎臟登錄系統(USRDS)的統計資料顯示，台灣洗腎率多年來仍維持世界第一高，在 2018 年急性和慢性腎病患者共花掉健保 513.78 億元，國內洗腎人數更超過 9 萬人也創下了歷年新高。

輕微的急性腎損傷在接受妥善的治療後，腎功能可以部份恢復；然而嚴重的急性腎損傷不僅會影響腎功能惡化，還可能因為尿毒症影響到全身的器官，導致慢性腎臟病，甚至需要永久洗腎以及死亡。

以重症加護病房治療的病人為例，發生急性腎衰竭的比率為 30~40%，需要洗腎的比例高達 50%以上，不僅伴隨高昂醫療支出，達到 50~80%以上的死亡率更是令人無法接受。

因此我們希望能透過機器學習的方式，並在與成大醫院醫師合作下，訓練出一套能對 AKI(急性腎衰竭)進行時序性的預測，並觀察病患在時間的推移下發生 AKI 的機率曲線分布，協助醫師做出判斷，達到爭取治療時間、降低治療成本以及提升整體醫療品質的目的。

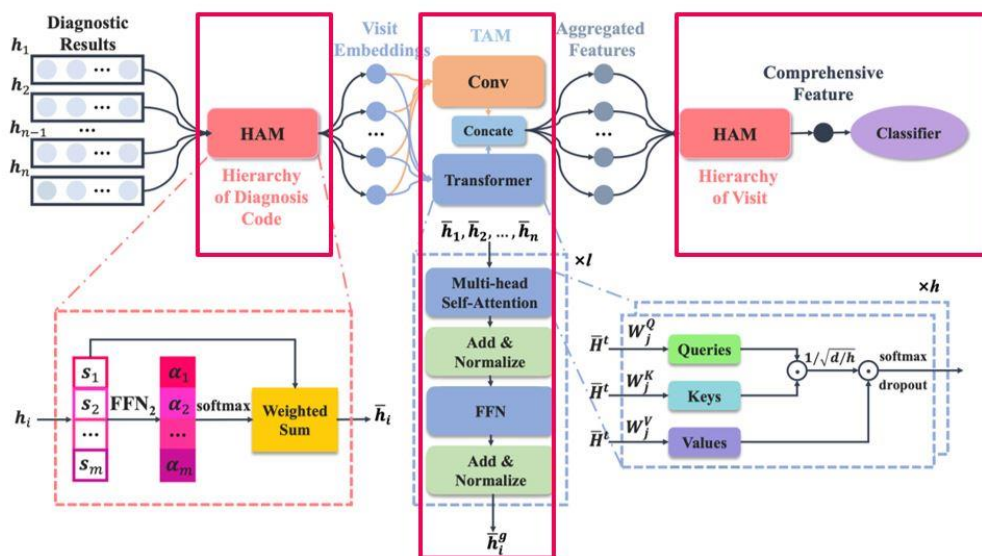
### (二)、資料集介紹

本次所採用的資料集全部來自成大醫院重症醫學科，並選用 2013 年至 2020 年間的資料，所選擇之病人年齡皆高於 18 歲，並使用了基本 11 特徵+額外 39 個特徵(特徵和樣本筆數列於附表 1 和 2)。

### (三)、模型介紹

在進行人工智慧相關的醫學任務中，常會發現醫學相關資料常具有病人間資料筆數不均和檢測時間間隔不相同等問題，以往學界所使用的 XGBoost、LSTM 等等模型因為對於輸入的資料有較嚴格的限制，因此會使用添補缺失值的方式來克服這些問題，雖然在測試集能取得不差的表現，但所填補的資料也往往會因為其真實性和合理性遭受醫師們的質疑。

而為了達成即使不填補缺失值也能進行預測，我們研究了諸如 **attention**、**transformer** 等對輸入要求較為彈性的機制，並在經過篩選和測試比較後選擇使用基於 **attention** 機制所建構的 **LSAN** 模型為基礎，並且對其結構稍作修改，使其能和本次任務目標相匹配。

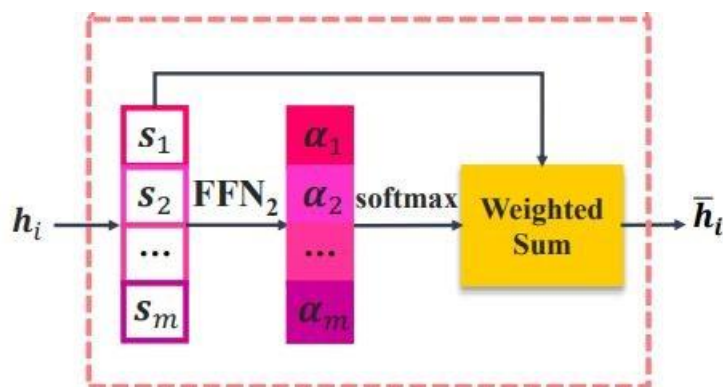


圖一、模型架構圖



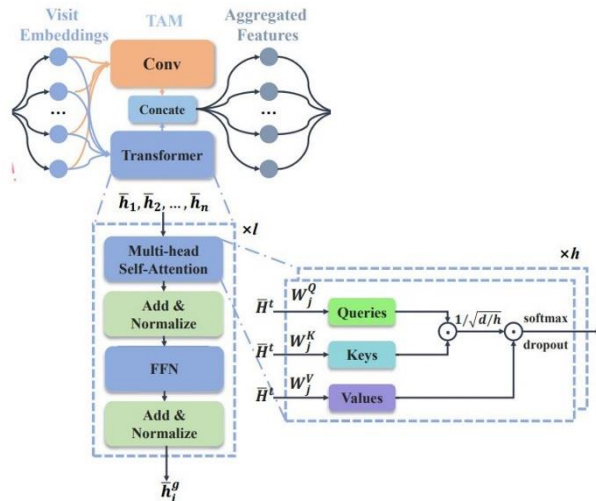
圖二、模型簡化圖

首先，第一層 **HAM**(Hierarchical Attention Module)可在訓練時透過三層的前饋神經網路去學習病人同個時間點中不同檢測項目對病情的影響，並將其轉化為模型權重。在獲得病人某個時間點上的所有檢測項目時，模型便能透過權重將所有檢測項目整理成單一數據。



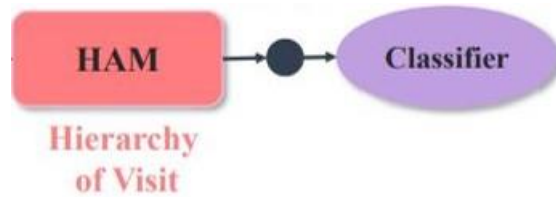
圖三、HAM 架構

而第二層 **TAM**(Temporal Aggregation Module)又可分為 **Long-term Dependencies** 和 **Short-term Dependencies**，其中前者可以透過 **Positional encoding** 的方式給予病人不同時間點的到訪資料不同的權重，對病人病況做長時間整體的評估判斷；而後者可以透過對相鄰時間點的到訪資料進行 **correlation**，捕捉病人病況在短時間內的發展。



圖四、TAM 架構

在將上層 TAM 所獲得的兩個結果相加，得到代表病人每一時間點的資訊後，再經由最後一層 HAM 學習病人不同時間點的狀態對病情的影響，並將其轉化為最終的模型權重，當模型在使用時獲得病人所有時間點的資訊後，此層 HAM 就能藉此產生單一結果，判斷病人未來是否會發病。



圖五、HAM 架構

#### (四)、實驗與結果

##### 1. 和時間長度相關的實驗

為了選出給予哪個模型 X 病人 Y 小時內的資料去預測病人 1 小時候會不會發生 AKI 會有最理想的結果，遂進行了此實驗。

時間長度	AUC	Precision	Recall	F1 Score
5	0.81	0.90	0.57	0.70
5	0.60	0.70	0.83	0.76
11	0.85	0.86	0.86	0.87
11	0.71	0.77	0.90	0.83
23	0.89	0.90	0.78	0.82
23	0.77	0.83	0.85	0.84

□ : LSAN

□ : [HiTANet](#) (Transformer based 的 model)

在進行實驗後，發現兩個模型給予 5 小時的資料均無法準確預測，11 及 23 小時表現優良，因 23 小時樣本數較少，最後選擇採用 11 小時預測，並選用在於此時間表現較好的 LSAN 模型。

## 2. 間隔長度(只使用基本 11 個特徵)的實驗

為了能準確預測多個小時後的病況，幫助急診醫師及早治療病人，遂測試在給予模型病人 11 小時內的資料時，能否準確預測病人 X 小時候會不會發生 AKI。

時間長度	AUC	Precision	Recall	F1 Score
3	0.81	0.82	0.80	0.82
6	0.74	0.75	0.86	0.80
3	0.68	0.78	0.88	0.82
6	0.66	0.75	0.85	0.80
3	0.78	0.69	0.68	0.69
6	0.72	0.52	0.57	0.54

□ : LSAN      □ : HiTANet      □ : T-LSTM(LSTM based)

在進行實驗後，發現 T-LSTM 表現顯著低於其他兩者，而 LSAN 和 HiTANet 兩者雖然皆有不錯表現，但 LSAN 仍略勝一籌。

## 3. 間隔長度(加入使用基本 39 個特徵)的實驗

為了能準確預測多個小時後的病況，幫助急診醫師及早治療病人，遂測試在給予模型病人 11 小時內的資料時，能否準確預測病人 X 小時候會不會發生 AKI。

時間長度	AUC	Precision	Recall	F1 Score
3	0.81	0.85	0.84	0.82
6	0.78	0.84	0.86	0.82
3	0.69	0.79	0.87	0.83
6	0.65	0.76	0.86	0.81
3	0.90	0.89	0.61	0.72
6	0.91	0.86	0.66	0.75

□ : LSAN      □ : HiTANet      □ : T-LSTM(LSTM based)

在進行實驗後，發現 T-LSTM 雖在 AUC 單一指標上取得不錯表現，但其他指標皆顯著低於其他兩者，而 LSAN 和 HiTANet 兩者在各指標表現平均，但 LSAN 仍略勝一籌。

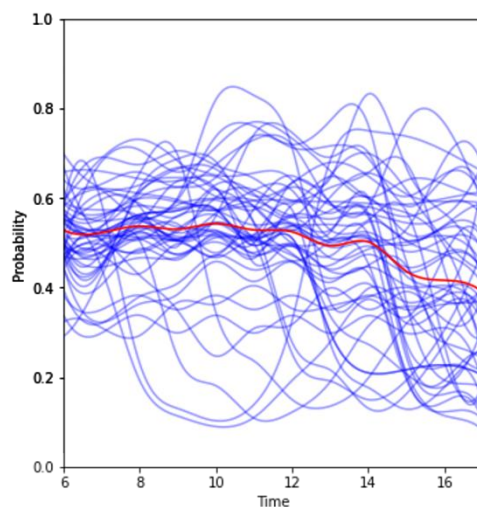
#### 4. 特徵數量的實驗

為了確認增加特徵確實有助於模型預測，遂進行此實驗

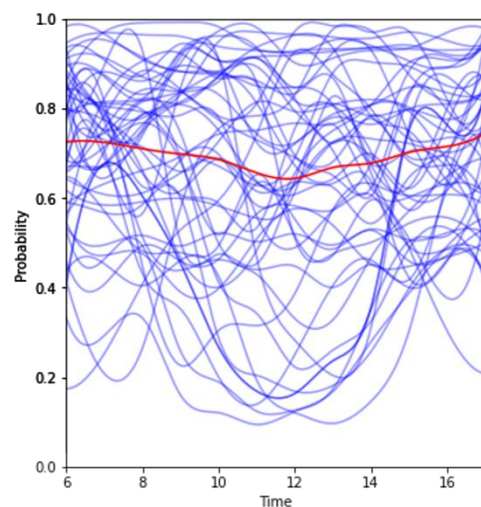
時間長度	特徵數量	AUC	Precision	Recall	F1 Score
6	11	0.74	0.75	0.86	0.80
6	50	0.78	0.84	0.86	0.82
3	11	0.81	0.82	0.80	0.82
3	50	0.81	0.85	0.84	0.82

在進行實驗後，可發現加入新特徵後對模型表現確實有著小幅度的提升。

#### 5. Rolling Prediction



圖六、無發生 AKI



圖七、有發生 AKI

左圖為最終無發生 AKI 的病人模型在各個時間點判斷的發病機率，右圖則為最終發生 AKI 之病人，可看出大致上不會發生 AKI 的病人模型皆會預測出較低的機率，而會發生 AKI 的病人模型預測出的機率會隨著時間越來越高。

在和醫師討論後認為此結果還有待改善，若是能在訓練時告知模型最終會發生 AKI 的患者還未發生 AKI 時的病況特徵，將能更完善此模型在預測 AKI 上的能力。

#### (五)、網站



圖七、網站首頁



×

## Task Introduction

根據台北榮民總醫院學訓的資料顯示，以重症加護病房治療的病人為例，發生急性腎衰竭的比率為30~40%，需要洗腎的比例高達50%以上，不僅伴隨高昂醫療支出，高達50~80%以上的死亡率更是令人無法接受。急性腎損傷後的倖存者常常腎功能無法完全恢復，使得慢性腎臟病與末期腎病需長期洗腎者之人數越來越多。

因此我們希望能夠藉由機器學習的方式，訓練出一套模型應用於預測急性腎衰竭的發生，並會隨著時間的推進與資料的不斷輸入而能夠顯示出病人發生急性腎衰竭的可能性，而醫生或是醫護人員便可透過機率分佈圖觀察到病人的狀況，從而協助醫生提早治療病患或是進行緊急處理，提升整體醫療品質。

Dataset採用成大醫院重症醫學科所提供的ICU臨床檢驗資料，並且與臨床醫師多次討論後選擇共50項特徵

► 特徵類別

模型訓練的方式選擇以小時為單位切割資料，並使用0~11小時的資料去預測第17小時的結果，從而讓模型可以學習到病人11小時內檢測數據的變化從而預測一小時後發生急性腎衰竭的機率。最終的預測結果圖是以每11小時為單位滑動，且不斷預測6小時後的機率所形成的機率分佈圖。最後藉由架設網站方便讓醫師使用並視覺化模型預測結果。

Advisor

圖八、網站專題題目介紹

×

## Model Introduction

病人整體的醫學檢測資料包含了病人(1)不同時間點的資料及(2)同一時間點下的不同檢測資料2種層層。現存的模型大部分專注於(1)對病人的影響，但卻往往忽略(2)的重要性，若不對(2)的資料作先行處理，容易導致許多和目標不相關的雜訊在處理(1)時被使用。

因此LSAN提出HAM和TAM兩module，嘗試先利用HAM處理(2)中的雜訊問題，再由TAM對病人整體病況同時做長時間和短時間的aggregation，藉由這兩個module，LSAN能達到超越許多現今所使用之模型表現。

範例下載

範例下載 2

範例下載 3

圖九、模型架構介紹

## Experiments

針對三種時間區段(11個特徵)  
時間長度:用病人X小時內的資料去預測病人1小時後會不會發生AKI

- ▶ HETANet
- ▶ LSAN

針對六種預測時間間隔(11個特徵)  
時間間隔(Gap):給予模型病人11小時的資料，預測病人X小時後會不會發生AKI

▼ T-LSTM

Gap	AUC	Prediction	Recall	F1 Score
1	0.72	0.60	0.65	0.56
2	0.64	0.44	0.79	0.56
3	0.78	0.69	0.68	0.69
4	0.74	0.51	0.62	0.56
5	0.68	0.45	0.72	0.56
6	0.83	0.74	0.54	0.63

- ▶ HETANet
- ▶ LSAN

針對六種預測時間間隔(30、40、50個特徵)  
時間間隔(Gap):給予模型病人11小時的資料，預測病人X小時後會不會發生AKI

- ▶ T-LSTM(50 features)
- ▶ HETANet(50 features)
- LSAN

圖十、結果介紹

×

UPLOAD FILE(PICKLE FILES):

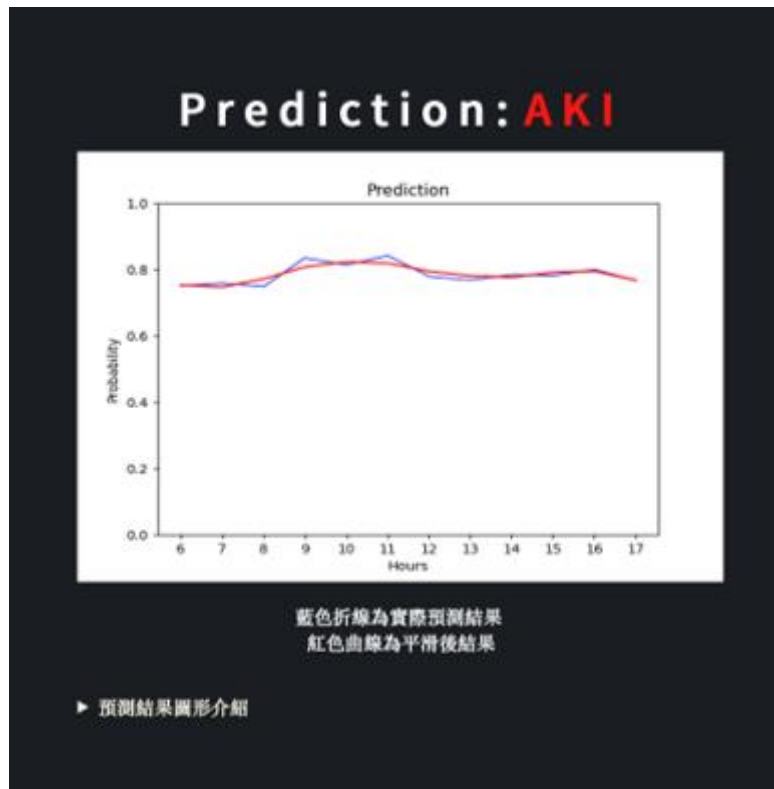
Please Read the [Model Introduction](#) Before you Upload File !

選擇檔案

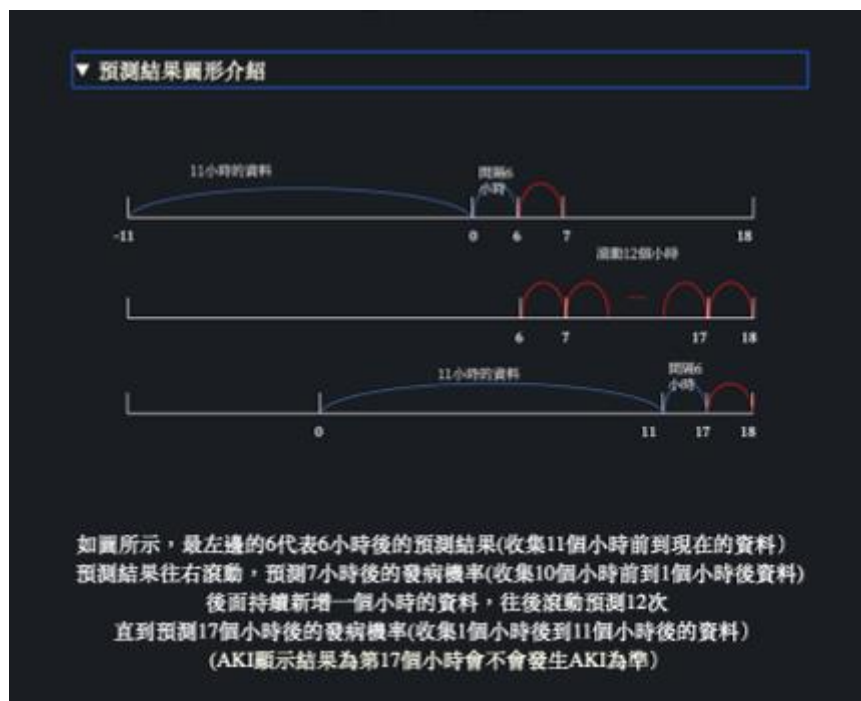
patient2.pickle

LSAN Prediction

圖十一、上傳病人資料



圖十二、預測結果



圖十三、預測結果說明



(五)、附表

附表 1、特徵

基礎 11	基本資料	年齡、性別
	血液檢測資料	HCT、Creatinine、eGFR、pH、K、BUN
	血壓資料	Art BP Mean、Art BP Systolic
	尿量資料	尿量
額外 39	血液檢測資料	Hb、NA、Plt、WBC、Baso、Eos、Lymph、Mono、Seg、ALT、AST、CA、MPV、MNAPTT、APTT、PT(MNPT)、PT、RBC、MCV、MCHC、MCH、RDW、PT(INR)、ALBUMIN、MG、Glucose (random)、P、TRIG、CHOL、GLU.A.C.、Aty-lym、Band、Blast、Meta、Pro、Myelo、NRBC、HbA1c
	尿液檢測資料	NIT、Cast、Crystal、Others

附表 2、樣本數量

時間長度	正樣本數	負樣本數
5	15169	3089
11	8829	3043
17	6258	2831
24	4901	2612
29	3908	2467