

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 教授

合作醫師：張育誌 醫師

Group 6-8. 郭冠良 蔡宇曉

1.

Introduction

Introduction – 目標

- ▷ 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。



Introduction – 合作醫師

張育誌 / Yu-Tzu Chang

臨床副教授兼主治醫師

經歷：

國立成功大學醫學院醫學系

國立成功大學醫學院臨床醫學研究所碩士

國立成功大學醫學院臨床醫學研究所博士

成大醫院腎臟內科主治醫師

成大醫學院醫學系內科學科臨床助理教授

成大醫院醫療品質委員會副召集人

與醫學系五年級、六年級導師



2.

Motivation

Motivation

- ▷ 根據美國腎臟登錄系統 (USRDS) 的統計資料顯示，台灣洗腎率仍維持世界第一高，在2018年急、慢性腎病患者共花掉健保513.78億元（國內洗腎人數超過9萬人創歷年新高）。



圖源：美國腎臟登錄系統 (USRDS)

Motivation

- ▷ 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，高達50~80%以上的死亡率更是令人無法接受。

急性腎衰竭(Acute kidney failure，又稱為Acute kidney injury或Acute renal failure)，是指腎臟突然失去最主要的過濾功能，導致血液中有過多的鹽、水分、廢棄物堆積。

Motivation

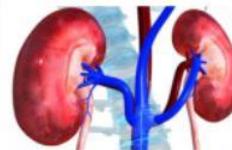
- ▷ KDIGO 全名為：Kidney Disease Improving Global Outcomes (改善全球腎臟病預後組織)，也是本次計畫採用的標準。按照危險程度分為三級，並有不同符合的條件。

急性腎衰竭的定義

Stage	SCr 血清肌酐 (mg/dl)			排尿量 (ml/kg/hour)
	RIFLE	AKIN	KDIGO	
時間	<7 天	<48 h	↑ 0.3: <48 h ↑ 50%: <7天	
Risk (1)	↑ 50%	↑ 0.3 or 50%	↑ 0.3 or 50%	<0.5 持續 6 h
Injury (2)	↑ 100%	↑ 100%	↑ 100%	<0.5 持續 12 h
Failure (3)	↑ 200%	↑ 200%	↑ 200% or 透析 or SCr ≥ 4	<0.3 持續 24 h 無尿 持續 12 h
Loss	透析 >4週			
End-stage	透析 >3月			



RIFLE: risk, injury, failure, loss, end-stage (ADQI 2004), h: hour
 ADQI: Acute Dialysis Quality Initiative
 AKIN: Acute Kidney Injury Network (2007)
 KDIGO: Kidney Disease Improving Global Outcomes (2012)



圖片來源

3. Dataset

Dataset

▷ 樣本數:

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

▷ 資料來源: 成大醫院重症醫學科。

▷ 資料時間: 選用位於2014至2020年間之病人資料。

▷ 所選擇之病人年齡皆高於18歲

Dataset – Features

基礎特徵：

- ▷ 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
- ▷ 尿液資料: 尿量

選用原因: 以上11個特徵是參考論文—[Automated Continuous Acute Kidney Injury Prediction and Surveillance: A Random Forest Model](#)後，再和醫師討論後所篩選而出。

Dataset – Features

後續增加特徵：

- ▷ 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

選用原因: 為Dataset病人中最多人擁有的50個特徵，經過篩選後剔除不合適的39個

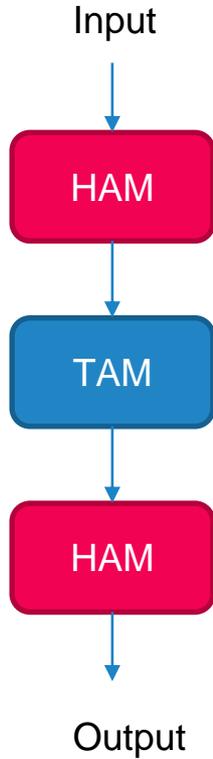
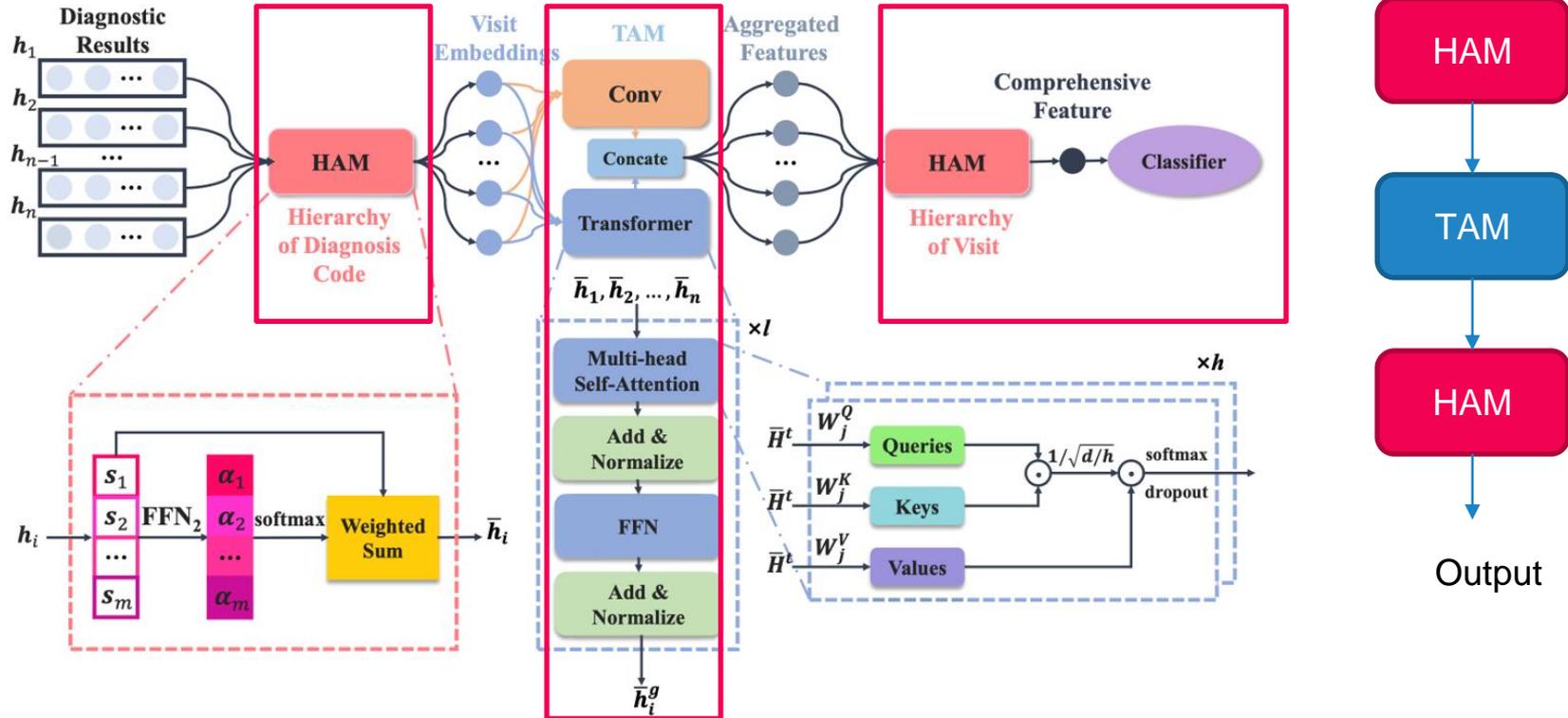
4. Model

Model - Reason

▷ 因為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。

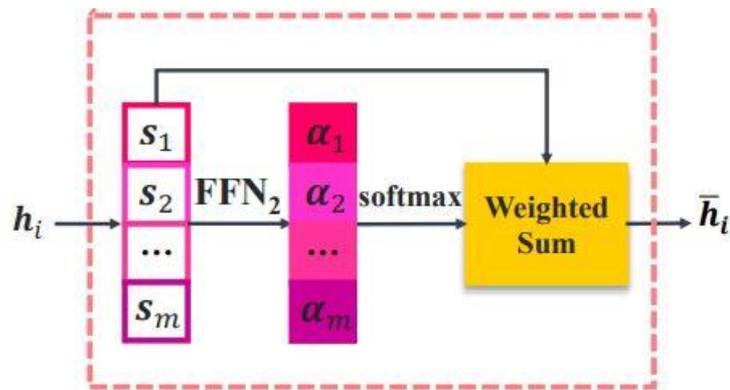
▷ 而為了達成即使不填補缺失值也能進行預測，我們研究了諸如attention、transformer等對輸入要求較為彈性的機制，並在經過篩選和測試比較後選擇使用基於attention機制所建構的LSAN模型為基礎，並且對其結構稍作修改，使其能和本次任務目標相匹配。

Model - Structure

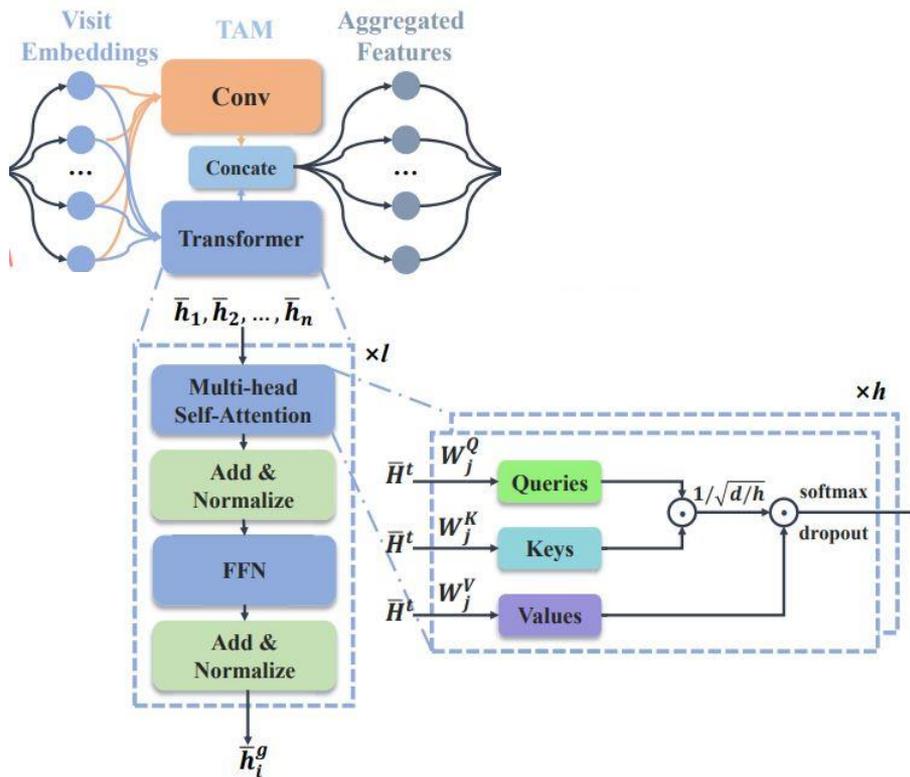


Hierarchical Attention Module(HAM)

- ▷ HAM可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，模型能透過權重將所有檢測項目整理成單一數據。



Temporal Aggregation Module(TAM)



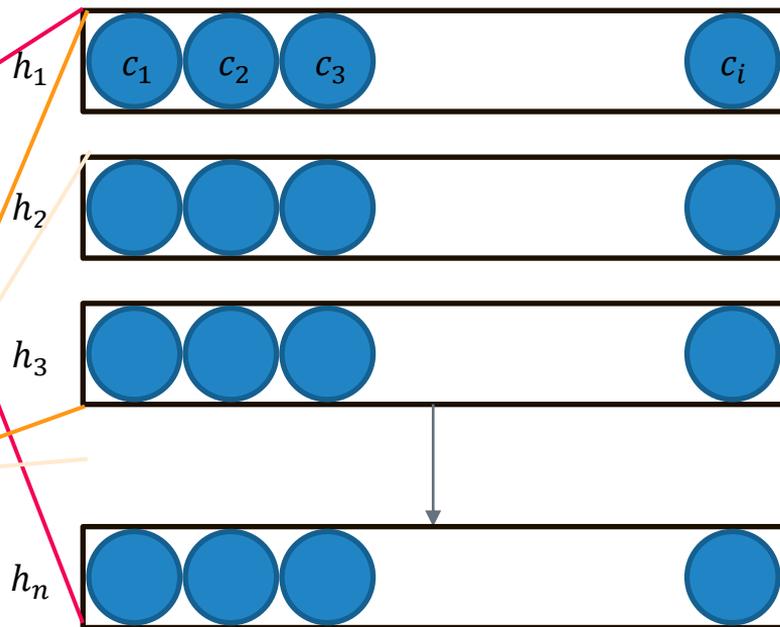
Temporal Aggregation Module(TAM)

Long-term Dependencies:

- ▷ 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。

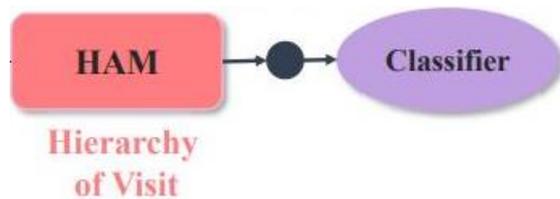
Short-term Dependencies:

- ▷ 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。



Hierarchical Attention Module(HAM)

- ▷ 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，並將其轉化為模型權重。
- ▷ 當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。



5. Experiments & Results

Experiments – 時間長度

- ▷ 時間長度: 用病人X小時內的資料去預測病人1小時候會不會發生AKI。
- ▷ 實驗目的: 選出模型能較準確預測的時間長度以及所適合使用的模型。

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

- ▷ 5小時無法準確預測
- ▷ 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，並選用在此時間表現較好的LSAN模型

□ : LSAN

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

Experiments – 間隔長度(只使用基本特徵)

- ▷ 時間長度: 用病人11小時內的資料去預測病人X小時候會不會發生AKI。
- ▷ 實驗目的: 若能準確預測多個小時後的病況，將有助於急診醫師及早治療病人。

間隔長度	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

- ▷ T-LSTM表現顯著低於其他兩者
- ▷ LSAN和HiTANet兩者皆有不錯表現，但LSAN仍略勝一籌。

□ : LSAN

□ : HiTANet (Transformer based的model)

□ : T-LSTM(LSTM based)

Experiments – 間隔長度(加入增加特徵)

間隔長度	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

- ▷ 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。
- ▷ LSAN和HiTANet兩者各指標表現平均，但LSAN仍略勝一籌。

□ : LSAN

□ : HiTANet (Transformer based的model)

□ : T-LSTM(LSTM based)

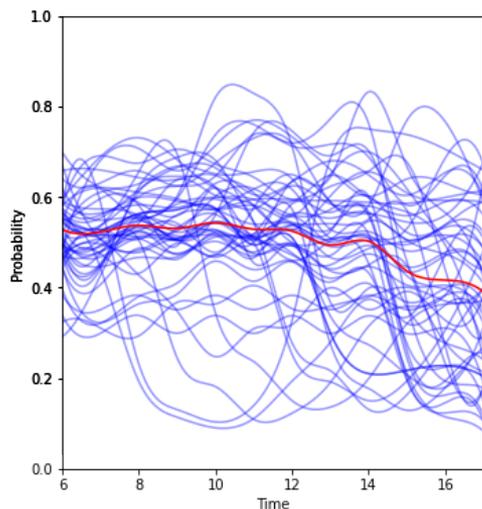
Experiments – 特徵數量

- ▷ 時間長度: 用病人11小時內的資料去預測病人X小時候會不會發生AKI。
- ▷ 實驗目的: 確認增加特徵數量確實有助於模型預測能力。

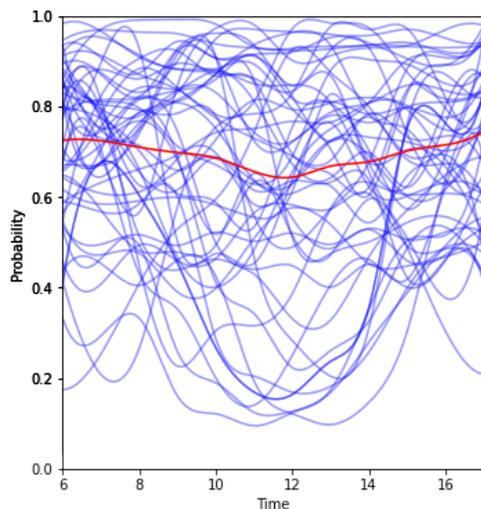
間隔長度	特徵數量	AUC	Precision	Recall	F1 Score
6	11	0.74	0.75	0.86	0.80
6	50	0.78	0.84	0.86	0.82
3	11	0.81	0.82	0.80	0.82
3	50	0.81	0.85	0.84	0.82

- ▷ 加入新特徵後對模型表現有著小幅度的提升。

Rolling Prediction



無發生AKI



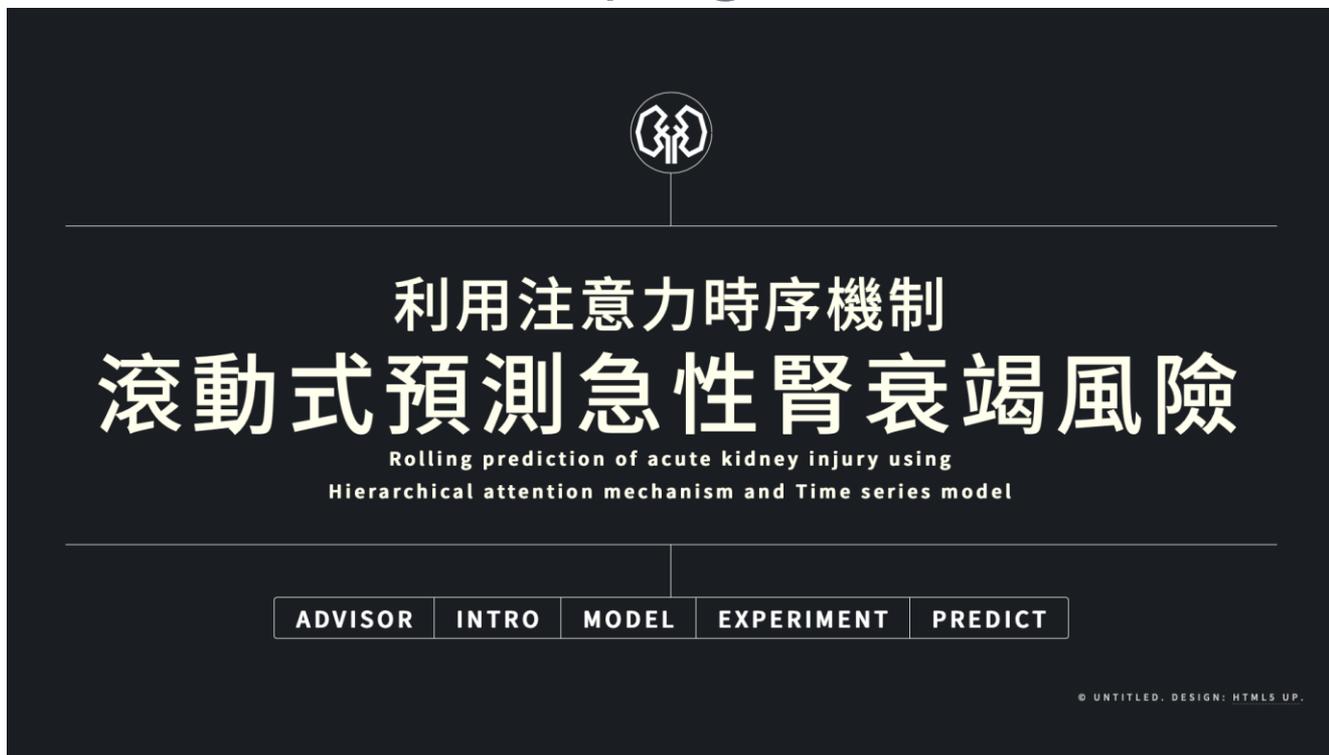
有發生AKI

- ▷ 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。

6.

Website

Website - Homepage



The image shows a dark-themed website homepage. At the top center is a circular logo with a stylized Chinese character. Below the logo is a horizontal line. The main title is in large white Chinese characters: '滾動式預測急性腎衰竭風險' (Rolling prediction of acute kidney injury risk). Above this title is the subtitle '利用注意力時序機制' (Using attention mechanism). Below the main title is the English subtitle 'Rolling prediction of acute kidney injury using Hierarchical attention mechanism and Time series model'. At the bottom, there is a horizontal line above a navigation menu with five buttons: 'ADVISOR', 'INTRO', 'MODEL', 'EXPERIMENT', and 'PREDICT'. In the bottom right corner, there is a small copyright notice: '© UNTITLED. DESIGN: HTML5 UP.'

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

ADVISOR | INTRO | MODEL | EXPERIMENT | PREDICT

© UNTITLED. DESIGN: HTML5 UP.

Website - Advisor

張育誌醫師

臨床副教授兼主治醫師

經歷：

國立成功大學醫學院醫學系

國立成功大學醫學院臨床醫學研究所碩士

國立成功大學醫學院臨床醫學研究所博士

成大醫院腎臟內科主治醫師

成大醫學院醫學系內科學科臨床助理教授

成大醫院醫療品質委員會副召集人

醫學系導師



蔣榮先教授

特聘教授

成大醫院健康數據資源中心執行長

經歷：

成大醫院健康數據資源中心執行長

美國系統生物學研究院 (ISB) 特聘教授

成功大學資訊工程學系教授

成功大學醫學院附設醫院資訊室主任

成功大學醫學院附設醫院資訊長

成功大學醫學院老年學研究所兼任教授

成功大學計算機與網路中心主任

成功大學圖書館副館長

工研院\電通所\研究員



Website - Introduction

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

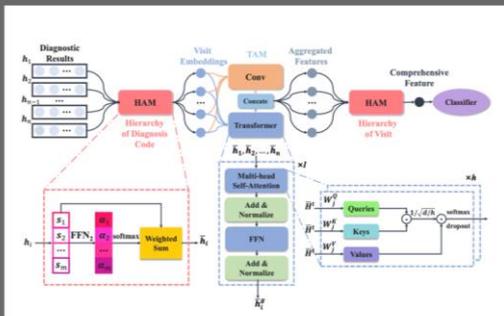
模型訓練的方式選擇以小時為單位切割資料，並使用0~11小時的資料去預測第17小時的結果，從而讓模型可以學習到病人11小時內檢測數據的變化從而預測一小時後發生急性腎衰竭的機率。最終的預測結果圖是以每11小時為單位滑動，且不斷預測6小時後的機率所形成的機率分佈圖。最後藉由架設網站方便讓醫師使用並視覺化模型預測結果。

Website - Model

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種階層，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

▶ 可從紅圈處下載範例病人資料，形式為pickle檔

Website - Experiment

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

- ▶ HiTANet
- ▶ 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

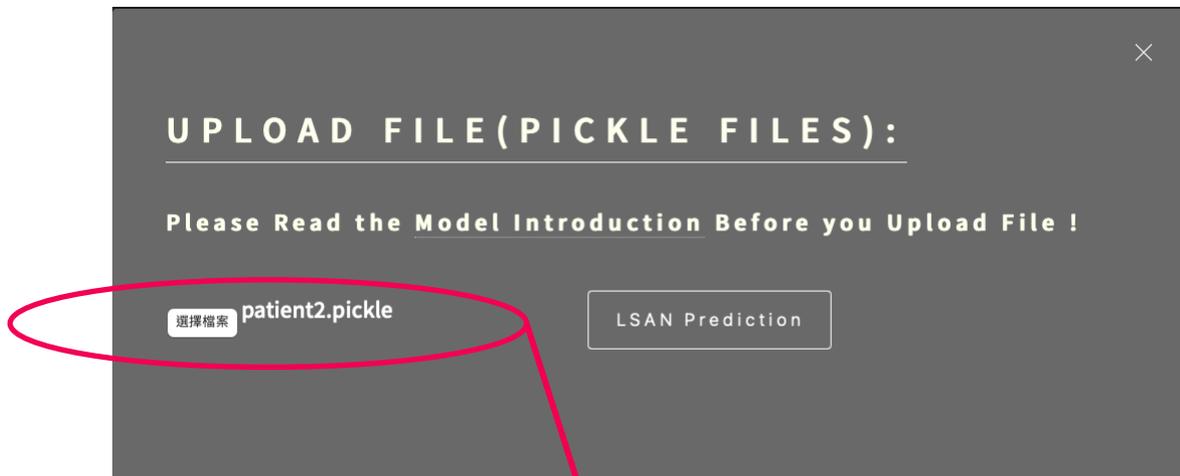
- ▶ HiTANet
- ▶ LSAN

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-LTSM(50 features)
- ▶ HiTANet(50 features)
- LSAN

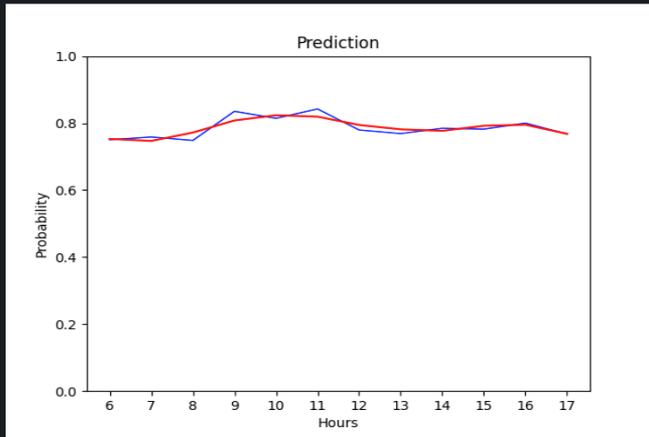
Website - Prediction



▶ 可從此處上傳網站提供的範例測資

Website - Prediction

Prediction: AKI



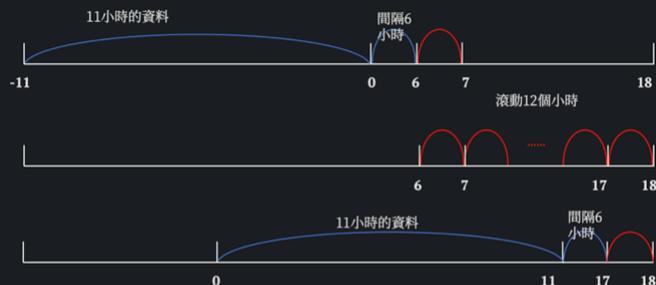
藍色折線為實際預測結果
紅色曲線為平滑後結果

► 預測結果圖形介紹

► 點開此處會有預測結果圖的判讀方式解說

Website - Prediction

▼ 預測結果圖形介紹



如圖所示，最左邊的6代表6小時後的預測結果(收集11個小時前到現在的資料)
預測結果往右滾動，預測7小時後的發病機率(收集10個小時前到1個小時後資料)
後面持續新增一個小時的資料，往後滾動預測12次
直到預測17個小時後的發病機率(收集1個小時後到11個小時後的資料)
(AKI顯示結果為第17個小時會不會發生AKI為準)

在此特別感謝
合作醫師 張育誌 醫師
指導教授 蔣榮先 教授
一路上的幫助