國立政治大學商學院風險管理與保險學系 專題研究報告

Department of Risk Management and Insurance College of Commerce

National Chengchi University (NCCU) Project Research Report

基於機器學習的肺癌預測模型於保險領域之應用
Lung Cancer Prediction in Insurance based on Machine
Learning

研究生: 葉乃瑋 Nai-Wei Yeh

中華民國 112 年 2 月 February, 2023

摘要

本研究以肺癌作為疾病標的,建立疾病預測模型,藉此解決現行保險實務上面臨之痛點,探討其在保險領域及醫療領域上的價值。本文利用資料集建立七種不同的模型,並利用在疾病預測方面最為適合的召回率 (Recall) 來進行模型的評估,選擇達到召回率 97.73% 的 Neural Network Architecture,以此模型算出的疾病預測機率,結合保險實務及對肺癌治療過程的研究,為不同條件的潛在肺癌病患打造不同險種及保險額度的規劃。本文期望可利用兩者之結合,除了幫助保險銷售人員提升保險規劃能力,協助病患選擇更適合的險種規劃,避免受龐大醫療費之苦,亦可使潛在肺癌病患提高警覺,更加注意身體狀況,以及了解保險規劃。

關鍵字:機器學習、深度學習、肺癌預測、保險、監督式學習、分類 演算法

Abstract

The purpose of this study is to develop a disease prediction model using lung cancer as the disease target, in order to solve the pain points of the current insurance practice and to explore its value in the insurance and medical fields. In this paper, seven different models were developed and the most suitable recall for disease prediction was used to evaluate the models. The Neural Network Architecture with a recall of 97.73% was selected. The disease prediction result calculated by this model was combined with insurance practice and research on the treatment process of lung cancer to create different types and amounts of insurance coverage for potential lung cancer patients with different conditions. This paper hopes to use the combination of the two to help insurance salespersons improve their insurance planning skills and help patients choose more suitable insurance plans to avoid suffering from huge medical expenses, and to make potential lung cancer patients more alert, pay more attention to their health conditions, and understand insurance planning.

Keywords: machine learning, deep learning, lung cancer prediction, insurance, supervised learning, classification algorithm.

目錄

摘	要		i
Ab	stract		ii
目:	錄		iii
昌	目錄		iv
表	目錄		iv
第	一章	研究背景與動機	1
第	二章	文獻探討	3
	2.1	疾病預測之演算法	3
	2.2	模型之評估準則	3
	2.3	商業健康保險概述	3
	2.4	肺癌及家族病史之關聯	4
第	三章	方法	5
	3.1	資料集	5
	3.2	模型說明e.ng.c.n	5
	3.3	評估準則	8
	3.4	商業保險介紹	10
	3.5	肺癌常見治療及花費介紹	10
第	四章	結果與討論	14
	4.1	資料預處理	14
	4.2	模型結果	15
	4.3	保險之應用	15
第	五章	結論	21
參	考文獻	k	22

圖目錄

3.1	類神經網路模型架構	8
	表目錄	
	政治	
3.1	資料集	5
	41/20	
3.2	Confusion Matrix	9
3.3	肺癌常見手術花費表	11
3.4	肺癌常見放射治療花費表	12
3.5	肺癌常見化學治療花費表	12
3.6	肺癌常見標靶治療花費表	13
4.1	One-hot encoding	14
4.2	Label encoding	14
4.3	不同模型之學習結果	15
4.4	肺癌常見治療花費及對應險種	16
4.5	險種優先順序	16
4.7	不同額度之規劃	17
4.6	險種必要給付項目與建議額度	18
4.8	疾病預測模型與保險規劃之應用	19

第一章 研究背景與動機

根據世界癌症基金會(World Cancer Research Fund International)發佈的統計報告[7],氣管、支氣管與肺癌為2020年全球最常見癌症的第二名。此外,根據台灣衛生福利部發表的110年國人死因統計結果報告[12],癌症為國人十大死因之首,其中肺癌更是名列第一。

另外,由[5]的研究可知,肺癌相對其他癌症,難以靠自我檢查就可於早期發現。若是肺癌的腫瘤長在肺的周邊部位,早期幾乎沒有症狀,通常等到腫瘤長大至中晚期,開始擠壓到周遭的器官,才會出現胸痛、肋膜積水、呼吸困難等症狀。若腫瘤生長在氣管或大支氣管內及週遭,雖容易因刺激支氣管壁,造成咳嗽、咳血痰或阻塞性肺炎等症狀,但常被誤認為感冒,而沒有積極進行檢查及治療。因此,多數肺癌患者發覺肺部健康有異狀而就醫時,癌細胞已擴散並發展至晚期,錯失黃金的治療時間。

除了難以即時發現早期病灶,後續的治療花費也是許多病患之壓力來源。根據台北癌症中心於民國 110 年所發行的「肺癌診療指引」[9] 中可發現,癌症屬於須長期抗戰的疾病,通常一次療程便是數個月不等。無論是常見的手術治療、化學治療、放射治療,或是近年來較新興的標靶治療、免疫治療等,雖有些藥物可透過健保給付,長期下來仍會成為許多家庭的資金缺口。

最後,根據衛福部公布的「健保有條件給付之癌症治療標靶藥品」[11]及「全民健康保險醫療服務給付項目及支付項目」[10]中含適應症之高價診療項目得知,由於現行醫療環境及健保制度,健保給付對病患來說日漸不足,商業保險之重要性逐漸提高。商業保險根據不同的險種及保額搭配,對於肺癌病患常使用到的醫療資源有非常大的幫助,例如:實支實付型、日額型、癌症險、重大傷病險等。此外,再輔以專業經理人對於市場概況的建議,可以客製保險規劃,幫助病患在治療方式選擇時,不再因為龐大的金錢壓力而錯失良好的治療機會。

綜上所述,本文以機器學習與深度學習的方式建立肺癌之疾病預測模型,加以分析,並結合肺癌治療與保險應用,用以解決保險業之痛點。由於保險銷售人員為客戶規劃時,並不一定具備足夠的醫學知識能為不同體況、背景的被保險人做出有差異化的規劃,故此研究透過兩者之結合,建立輔佐潛在病患選擇保險規劃的依據,幫助保險銷售人員提供更完善的保障建議,體現保險的價值。另外,亦可協助找到潛在肺癌病患,使其對自己的身體提高警覺,在病狀剛出現時能及時進行檢查並接受治療,以及可以更了解如何規劃險種,預防未來可能因龐大治療費用而造成的資金缺口。



第二章 文獻探討

2.1 疾病預測之演算法

關於處理疾病預測分類問題之演算法,Ganta 等作者(2022)[6]使用了SVM、Decision Trees、Logistic Regression、Random Forest 和 Naïve Bayes 模型,以建立三種不同癌症的預測;而 Shraddha 等作者(2020)[4]則以 KNN、Decision Trees 和 Logistic Regression 方法建立癌症預測模型。參考上述文獻,本文也將在後續使用類似演算法建立肺癌之預測模型,處理分類問題。

2.2 模型之評估準則

針對模型的評估準則,Miroslav (2017)[3]指出,Accuracy 準確率雖為最常被選用的指標,但較適合用在沒有特別關注的類別,只想盡量提升模型預測能力時。在醫療領域上常面臨數據不平衡的問題,若單純使用 Accuracy 準確率來衡量模型結果可能會造成誤導,使模型無法偵測出陽性樣本。

而 Recall 召回率在醫療領域上也是非常重要的指標,相較於整體準確率,當 我們更重視陽性樣本的預測好壞時,Recall 召回率會是不錯的衡量依據。因此, 本文將使用 Recall 召回率作為模型評估的指標。

2.3 商業健康保險概述

健康保險(Health Insurance)係指以被保險人身體或生命為保險標的,使被保險人在遭受疾病或意外事故,而產生傷害所導致之直接費用或間接損失獲得補償的保險。本文參考中華民國保險經營學會於西元 2013 年發表之文章 [8],及金

管會規定之住院醫療費用保險單示範條款(實支實付型)[14]、住院醫療費用保險單示範條款(日額型)[15]及人壽保險單示範條款[13],選擇醫療費用保險、癌症險及重大傷病險等多項肺癌可用到的險種進行介紹,並在本文後續針對上述險種進行險種優先順序分析及保額設定。

2.4 肺癌及家族病史之關聯

現今有許多研究指出,家族病史(Family Medical History)為肺癌的常見危險 因子之一。根據[2]的研究指出,具有肺癌病史之患者,其直系親屬肺癌的發生 率明顯增加。不過,目前並沒有發現哪些基因會經由遺傳直接造成肺癌,家族遺 傳表現僅為一種傾向。

在本研究中,家族病史為資料集所缺少的欄位,但在保險實務上,通常會利用家族病史來評估保險額度及險種,故會在後續的保險應用中用以協助保單規劃。

Zon Chengchi Univer

第三章 方法

3.1 資料集

本文所使用的資料集來源為 Hong, Z.Q. 與 Yang, J.Y. 於西元 1991 年發佈 [1], 共 309 筆數據,原始資料共有 16 個欄位。變數名稱、敘述及範圍,詳見下列表 3.1。

No.	欄位名稱	欄位代碼	敘述
-		GENDER	
1	Gender(性別)		M= 男性, F= 女性
2	Age in years(年龄)	AGE	[21 to 87]
3	Smoking(抽菸習慣)	SMOKING	2= 是, 1= 否
4	Yellow Fingers(黃指甲症候群)	YELLOW_FINGERS	2= 是, 1= 否
5	Anxiety(焦慮傾向)	ANXIETY	2= 是, 1= 否
6	Peer Pressure(同儕壓力)	PEER_PRESSURE	2= 是, 1= 否
7	Chronic Disease(慢性疾病)	CHRONIC DISEASE	2= 是, 1= 否
8	Fatigue(疲勞)	FATIGUE	2= 是, 1= 否
9	Allergy(過敏)	ALLERGY	2= 是, 1= 否
10	Wheezing(喘息)	WHEEZING	2= 是, 1= 否
11	Alcohol Consuming(酗酒)	ALCOHOL CONSUMING	2= 是, 1= 否
12	Coughing(咳嗽)	COUGHING	2= 是, 1= 否
13	Shortness of Breath(呼吸短促)	SHORTNESS OF BREATH	2= 是, 1= 否
14	Swallowing Difficulty(吞嚥困難)	SWALLOWING DIFFICULTY	2= 是, 1= 否
15	Chest Pain(胸痛)	CHEST PAIN	2= 是, 1= 否
16	Lung Cancer(罹患肺癌)	LUNG_CANCER	Yes= 罹患肺癌, No= 未罹患

表 3.1: 資料集

3.2 模型說明

機器學習為人工智慧的次領域,為利用過往的資料和經驗學習,找到其運行規則並建立模型,以進行預測;而深度學習為機器學習中的次領域,是更進階的分支,同樣都是為了建立迴歸/分類模型,唯一的差異在於深度學習使用人工類神經網路,而非其他統計模型。以下會介紹本文所使用之機器學習和深度學習之演

算法。

3.2.1 Logistic Regression 邏輯迴歸

邏輯回歸是一個最基本的二元線性分類器,由線性迴歸變化而來,其目標是要找出一條直線能夠將所有數據清楚地分成兩類。為了避免分類錯誤,加上一個對數函數 Sigmoid 函數,使得輸出更有彈性。邏輯回歸用到的 Sigmoid 函數值域在(0,1)之間,使用分類器將 0 和 1 之間的值轉換為 0 或 1。其 Sigmoid 的形式如下:

$$\sigma(f(x)) = \frac{1}{1 + e^{-f(x)}} = \frac{e^{f(x)}}{1 + e^{f(x)}}$$
(3.1)

3.2.2 Gaussian Naive Bayes 高斯貝氏分類器

貝氏分類器根據貝氏定理(Bayes'theorem)為基礎,透過機率統計來判斷 未知的資料類別,其中因應不同的資料集會有不一樣的訓練架構。

而高斯貝氏分類器便為其中一種架構,主要用於特徵為連續變數時,例如特徵長度為幾公分、重量為幾公斤等。於此模型下假設特徵資料為常態分配。首先,會先對資料進行分類,假設訓練及中有一個連續屬性x,然後計算每個類別中x的均值和變異數。使 μ_c 表示x在c類上的均值,令 σ_c^2 為x在c類上的變異數。機率計算公式如下:

$$P(x = v|c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{-\frac{(v-\mu_c)^2}{2\sigma_c^2}}$$
(3.2)

3.2.3 Bernoulli Naive Bayes 伯努力貝式分類器

伯努力貝式分類器也是以貝氏定理為基礎的其中一種分類器架構,適用於二元的特徵,亦即非 0 即 1,例如特徵是否出現、特徵大小、特徵長短等。其機率計算公式為:

$$P(x_i|y) = P(i|y)x_i + (1 - P(i|y))(1 - x_i)$$
(3.3)

3.2.4 Support Vector Machine 支持向量機

支持向量機是一種基於統計學習的監督式演算法,既強大且用途廣泛,能夠 執行線性或非線性分類,針對小樣本、非線性、高維度與局部最小點等問題具有 相對的優勢。

其運作方式為將低微度空間線性不可分的樣本映射到高維度空間去,找到一個超平面將這些樣本做有效的切割,並使這個超平面兩邊的樣本都與此超平面有 最遠距離。

3.2.5 Random Forest 隨機森林

隨機森林屬於集成學習(Ensemble Learning)的一種。集成學習是指將多種學習算法,通過適當的形式組合起來完成同一項任務。隨機森林是使用 Bagging 算法,加上隨機特徵採樣以生成多個不同數據集,藉由在每個數據集上訓練一個決策樹(Decision Trees),最終結合多棵不同決策樹之預測結果,使結果較不容易過擬合,以提升預測能力。

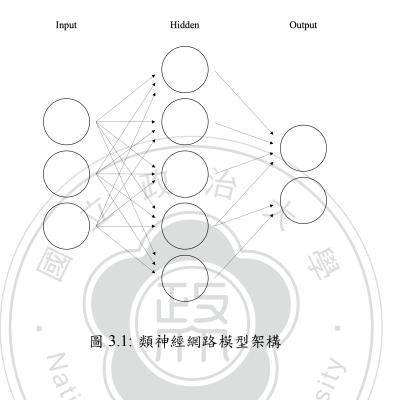
3.2.6 K Nearest Neighbors K-近鄰演算法

最近鄰居演算法屬於機器學習中監督式學習的其中一種算法,透過一群已經標記好類別的資料,針對未分類的資料做分類,用途廣泛,適用於離散型資料, 也適用於連續型資料。

其分類的標準是由鄰居「多數表決」決定,若身邊多數屬於某類別資料,則便把自己歸類於此類別。

3.2.7 Neural Network Architecture 類神經網路

類神經網路是模仿生物神經系統的結構,透過神經元(Neuron)互相連結, 建立層層傳導之模型,最後作出反應。類神經網路涉及數學統計學以及人工智慧 領域,通常透過統計學的學習方法將模型最佳化,並結合人工智慧學的人工感 知,使得類神經網路能像人一樣具有簡單的決定能力和判斷力,這種方法較機器 學習演算法有優勢。 資料分析中的類神經網路由大量的神經元相連組成,每個節點代表一種特定的輸出函數,稱為啟動函數。每兩個節點之連結都代表一個權重,也就是模型要求解的參數,求解之後會得到公式,未來只要輸入資料,即可經由層層傳導至輸出層,推算結果。



3.3 評估準則

不管是在機器學習或是深度學習中,分類任務評價指標都很重要,一個好的 評價指標對於訓練一個好的模型極其關鍵,一般使用準確率作為評價指標。然 而,若訓練資料的目標變數不平衡則需要藉找出其他適合評估的指標進行分析, 又根據上述文獻探討所提到之評估準則,因此,本文使用 Recall (召回率)來進 行模型間的比較。

為了評估機器學習模型分類器的結果,本文使用混淆矩陣(Confusion Matrix)以及模型評估報告(Classification Report)來計算測試及每一筆樣本被預估之結果。以下為混淆矩陣及模型評估報告介紹。

3.3.1 Confusion Matrix 混淆矩陣

混淆矩陣是用於計算預測值和實際值類別標籤組合的表格,本文將使用其中的值計算 Recall 召回率。其中,混淆矩陣中的值意義分別如下:True Positive 代表實為「陽性 (Positive)」,也正確判斷為「陽性」者;False Positive 代表實為「陰性 (Negative)」,但被誤判為「陽性」者;True Negative 代表實為「陰性 (Nagetive)」,也正確判斷為「陰性」者;False Negative 代表實為「陽性 (Positive)」,但被誤判為「陰性」者。

表 3.2: Confusion Matrix

 $\begin{array}{c|ccccc} & & & & & & & & & \\ \hline \textbf{True value} & & & & & & \\ P & & N & & & & & \\ \hline \textbf{Positive} & & & & & & \\ \hline \textbf{Positive} & & & & & \\ \hline \textbf{Positive} & & & & & \\ \hline \textbf{Positive} & & & & & \\ \hline \textbf{False} & & & & \\ \hline \textbf{True} & & & & \\ \hline \textbf{Negative} & & & & \\ \hline \textbf{Negative} & & & & \\ \hline \end{array}$

3.3.2 Classification Report 模型評估報告

模型評估報告用於顯示主要幾個分類指標的文字報告,在報告中顯示每個類的精確度、召回率、F值等資訊。其中,模型評估報告中的各項分類指標意義如下:

· Accuracy(準確度): 準確度代表無論結果為陰性或是陽性,判斷正確的比例。

$$\frac{N_{TP} + N_{TN}}{N_{FP} + N_{FN} + N_{TP} + N_{TN}} \tag{3.4}$$

• Precision(精確度):判斷為陽性的資料中,真實為陽性的比例。

$$\frac{N_{TP}}{N_{FP} + N_{TP}} \tag{3.5}$$

• Recall(召回率):實際陽性的資料中,正確被判斷為陽性的比例。

$$\frac{N_{TP}}{N_{FN} + N_{TP}} \tag{3.6}$$

• F1-score(F值):精確率和召回率的調和平均數。

$$F_{\beta} = \frac{(1+\beta^2)Precision \times Recall}{\beta^2 Precision + Recall}$$
(3.7)

3.4 商業保險介紹

本段分項介紹肺癌治療中會用到之險種,為後續保險應用之依據。

3.4.1 醫療費用保險

醫療費用保險並不僅限於肺癌的理賠,只要有醫療行為產生,即可啟動,理 賠範圍最為廣泛,通常包括實支實付型及日額型兩種。

實支實付型為損害補償契約,保險人在被保險人支付醫療費用之範圍內,負給付保險金責任。而日額型保單屬定額給付,保險人依據住院天數及每日住院之保險金額給付保險金。常見給付項目通常包含:每日病房費用保險金、住院醫療費用保險金、手術費用保險金和出院療養保險金等。

3.4.2 癌症險

癌症險中癌症之定義範圍採世界衛生組織(WHO)之惡性腫瘤,只要病患符合此定義,即可理賠。

通常癌症險可分為療程型及整筆給付型,常見給付項目包括:初次罹患癌症 保險金、癌症手術保險金、癌症住院保險金、癌症放射性治療保險金、癌症化學 治療保險金、移植或重建保險金。

3.4.3 重大傷病險

當被保險人初次罹患條款約定的重大傷病(健保重大傷病範圍)並取得重大傷病證明,即可給付一筆保險金,但其中有八大類與先天性疾病或職業病相關的疾病除外不賠。由於肺癌屬於重大傷病的一種,故可申請理賠。

3.5 肺癌常見治療及花費介紹

本段列出肺癌常見治療以及不同療程花費,為後續保險額度規劃之依據。

3.5.1 手術治療

手術治療的目的為將肺部原發性腫瘤的病灶及局部淋巴結切除,對於原發性 非小細胞肺癌是目前最佳的治療方式,依病灶情形和侵犯程度可分為楔形或肺小 節切除、肺葉切除和全肺切除。手術治療花費細項詳見表 3.3。

表 3.3: 肺癌常見手術花費表

手術	名稱	治療用法	平均花費
傳統開	胸手術	於病灶胸側劃開長達 25 公分的傷口, 剪斷肋骨,再以支撐器撐開。出血多 且併發症多、恢復期長。	健保給付
胸腔鉤	竞手術	切開 3-5 公分的傷口,輔以 3 個小於 1 公分的傷口,將手術器械伸入操作。 出血少、傷口小、恢復期短。	自費 3-5 萬
2D 單孔胸	腔鏡手術	為平面影像,只需 3-5 公分之傷口。 出血少、傷口較小、恢復期短。	自費 6-7 萬
	正常手術	影像更立體,分辨氣管、血管更清 楚。出血少、傷口小、恢復期短。	自費 8-9 萬
3D 單孔 胸腔鏡手術	免插管	胸腔鏡手術須全身麻醉,讓手術側 肺臟停止呼吸。插管易損傷氣管、喉 嚨。免插管利用注射少量麻醉使病人 睡著或局部麻醉,以降低病人恐懼。	自費 9-10 萬
	術前影像定位	出血少、傷口小、恢復期短。 用於較深的肺癌組織。出血少、傷口 小、恢復期短。	自費 12-13 萬
達文西	与手術	傷口較胸腔鏡大且多,但可進行更精 細、複雜操作,切除較乾淨。	自費 25 萬

3.5.2 放射治療

放射治療利用高能量放射線,穿過體表將輻射劑量集中於腫瘤部位以精確的 殺死癌細胞。通常會搭配手術作為輔助療法,可在手術前照射使腫瘤縮小,或殺 死手術後殘餘、肉眼不可見的癌細胞。另外,也可對晚期患者進行緩解性治療。 一般療程為每天照射一次,每週連照五天,持續數週,可於門診接受治療,不需 住院。放射治療花費細項詳見表 3.4。

表 3.4: 肺癌常見放射治療花費表

放射治療名稱	平均花費
影像導航	自費 3-6 萬
弧形刀	自費 3-6 萬
真光刀	自費 3-6 萬
電腦刀	自費 20 萬
螺旋刀	自費 15-35 萬
質子刀	自費一次2萬,完整療程約自60萬
氩氦刀	自費 15-20 萬
伽瑪刀	自費 14 萬

3.5.3 化學治療

化學治療是以口服或注射抗癌藥物的方式進行治療,藉血液流至身體各部位。當癌細胞轉移至身體其它部位時,手術和放射治療無法完全控制腫瘤,須採用這種全身性治療,以減緩腫瘤的生長來減輕症狀及增加生活品質。

因癌細胞生長速度較正常細胞快,對化學藥物更加敏感,故可以抑制或殺死全身癌細胞。然而,生長較快速的正常細胞也會同時受到影響,因此便產生常見的副作用,如血球下降、噁心、嘔吐。化學治療花費細項詳見表 3.5。

表 3.5: 肺癌常見化學治療花費表

藥物名稱	藥物類別	平均花費
卡鉑 Paraplatin	鉑類化合物	健保給付
順鉑 Cisplatin	鉑類化合物	健保給付
剋癌易 Taxotere	化學藥物	健保給付
好復 5-FU	化學藥物	健保給付
健擇 Gemzar	化學藥物	健保給付
汰癌勝 Taxol	化學藥物	健保給付
愛寧達 Alimta	化學藥物	健保給付
優富多 UFT	化學藥物	健保給付
敏畢瘤 Vinblastine	化學藥物	健保給付
溫諾平 Navelbin	化學藥物	健保給付

3.5.4 標靶治療

標靶治療為近年來新興的治療方式。相對於傳統化學治療的非選擇性細胞毒 殺作用,副作用較大且難避免,標靶治療為針對特定腫瘤基因發展藥物以作用在 癌細胞內,因此不會有類似化療的副作用。

目前常見治療肺癌的標靶藥物有三種類型,其中 EGFR-TKLs 類型之藥物為 表皮生長因子接受器酪胺酸激脢抑制劑,而 ALKI 類型之藥物為間變性淋巴瘤激 脢抑制劑。目前,多數的標靶藥物仍非常昂貴,使用前需和醫師詳細討論是否適 合。標靶治療花費細項詳見表 3.6。

施业 为 60	滋 业 ** 口,1	亚 14 计 典
藥物名稱	藥物類別	平均花費
愛瑞莎 Iressa	EGFR-TKLs 口服藥	健保給付
得舒緩 Tarceva	EGFR-TKLs 口服藥	健保給付
妥復克 Gilotrif	EGFR-TKLs 口服藥	健保給付
泰格莎 Tagrisso	EGFR-TKLs 口服藥	健保有條件,未達條件自費一個月30-40萬
截剋瘤 Xakori	ALKI 口服藥	健保給付
立克癌 Zykadia	ALKI口服藥	健保給付
安立適 Alecensa	ALKI 口服藥	健保給付
癌思停 Avastin	抗血管新生注射針劑	自費一個月 5-6 萬
欣銳擇 Cyramza	抗血管新生注射針劑	自費一個月 4-6 萬

表 3.6: 肺癌常見標靶治療花費表

3.5.5 免疫治療

免疫治療對於標準治療無效的肺癌病患,將自體免疫細胞抽取出來,經由體外的藥物刺激培養,蛻變成具備較強追獵肺癌能力的細胞,再回輸體內。部分非小細胞癌病人,經醫師評估後可使用免疫治療,目前多種免疫治療藥物仍在研究中。

若符合條件,免疫治療可以申請健保給付。反之,則以第一線免疫藥物加上抑制血管新生的標靶藥物,依病人體重決定藥物劑量,治療1次之費用合計約在20至22萬元間。而一般療程每三週一次,至少四次,基本花費約80、90萬元。

第四章 結果與討論

4.1 資料預處理

首先,這份資料集有309 筆資料,有16個欄位,並無缺失值,共有33 筆重複的資料,移除重複值後剩下276 筆資料。接著,針對種類變數,在運用機器學習的方法時,為了方便處理,我們將其轉換為數值。本文先針對性別(GENDER)欄位中的敘述(M=男性、F=女性)取代為(Male=男性、Female=女性),再使用One-hot Encoding的方式來處理,下列表4.1為處理過後的欄位資訊。

表 4.1: One-hot encoding

No.	欄位名稱	欄位代碼(前)	敘述	欄位代碼(後)	
1	Gender	GENDER		GENDER_Male GENDER_Female	
nengchi					

接著,再針對是否罹患肺癌(LUNG_CANCER)欄位,使用 Label Encoding的方式,將類別資料轉換為數值,下列表 4.2為經過處理後的欄位資訊。

表 4.2: Label encoding

No.	欄位名稱	欄位代碼	敘述(前)	敘述(後)
16	Lung Concer	LUNC CANCED	Yes= 罹患肺癌	1= 罹患肺癌
10	Lung Cancer	LUNG_CANCER		0=未罹患

4.2 模型結果

本文使用 80% 的資料做為訓練集,20% 的資料為測試集來分割。接著本文使用七種模型來做訓練,最後再使用 Recall(召回率)作為評估準則。由下列表4.3之結果,為類神經網路 (Neural Network Architecture)模型表現較佳,故本文之後續疾病預測,將以此模型更深入探討。

模型	Recall 召回率
Logistic Regression	83.33%
Gaussian Naive Bayes	82.20%
Bernoulli Naive Bayes	82.20%
Support Vector Machine	87.50%
Random Forest	70.83%
K Nearest Neighbors	69.70%
Neural Network Architecture	97.73%

表 4.3: 不同模型之學習結果

4.3 保險之應用

4.3.1 肺癌常見治療花費及對應險種

本段綜合上述對肺癌治療方法及花費的了解,並根據前述文獻探討中肺癌可用到的險種,加以探討。下表 4.4 列出其治療花費相對應的險種,了解實務上保險在醫療領域之應用,幫助後續在險種規劃更精確。

4.3.2 險種重要程度排序(針對肺癌治療)

保險之於病患,給付項目是否可以涵蓋多數治療行為,以及是否能夠讓病患 在適當的時機拿到理賠金,皆極其重要。根據上述文獻探討之險種介紹,結合市 場實務理賠經驗,針對罹患肺癌病人可能遇到的醫療行為及治療方式,於下表 4.5 分別以實用性及理賠立即性來評估險種的優先順序。

以實用性來說,醫療費用保險優先,療程型之癌症險次之,最後則為整筆給 付型癌症險及重大傷病險。醫療費用保險的理賠範圍,不論是任何疾病或是意

	項目	平均費用	對應險種
	病房費用	5,500 元 / 日	實支實付型醫療費用保險日額型醫療費用保險
	手術治療費用	13 萬元 / 次	實支實付型醫療費用保險療程型癌症險
	化療費用	健保給付	實支實付型醫療費用保險療程型癌症險
住院中	放療費用	3,000 元 / 次	實支實付型醫療費用保險療程型癌症險
	標靶治療費用	6萬元/次	實支實付型醫療費用保險療程型癌症險(少數)
	免疫治療費用	20 萬元 / 次	實支實付型醫療費用保險
	病房膳食費用	150元/日	實支實付型醫療費用保險日額型醫療費用保險
門診中	掛號費	500 元 / 次	日額型醫療費用保險
	營養品	1.2 萬元 / 月	整筆給付癌症險重大傷病險
出院療養	輔具	5,000 元 / 次	整筆給付癌症險重大傷病險
	看護	5萬元/月	整筆給付癌症險重大傷病險
隱性成本	薪資補償	5萬元/月	整筆給付癌症險重大傷病險

表 4.4: 肺癌常見治療花費及對應險種

外,皆有涵蓋,故給付包含之範圍最為廣泛;而療程型的癌症險是針對「癌症」 所產生的醫療行為來進行理賠,故理賠範圍較醫療費用保險少;最後的整筆給付 型癌症險和重大傷病險,皆只理賠一筆罹癌保險金,只要確定罹患癌症即可啟 動,並沒有針對化學治療、放射治療或是住院病房費等醫療行為細項之理賠,故 理賠範圍最少,實用性最低。

然而,以理賠立即性來說,整筆給付型癌症險及重大傷病險為最優先,其餘皆為次之。以理賠依據作為比較,整筆給付型癌症險及重大傷病險只要有醫生診斷證明或是病理切片報告等確定罹癌相關文件,即可申請,即使病患還在接受治療,理賠作業仍可繼續,幫助病患彌補最一開始的資金缺口,如緊急需要進行手術的花費或第一次住院之醫藥費等,故理賠立即性為優先。至於其餘險種,皆須等到療程結束或出院批價後,才有辦法申請理賠,病患還是須自備一筆醫藥費先行負擔,後續才能拿到理賠金,故理賠立即性次之。

綜上所述,結合實用性與理賠立即性之排序,列出綜合優先順序排名:醫療 費用保險為優先,療程行癌症險次之,而整筆給付癌症險及重大傷病險為最後。

表 4.5: 險種優先順序

险種名稱	實用性	理賠立即性	綜合優先順序
醫療費用保險	優先	次之	優先
療程型癌症險	次之	次之	次之
整筆給付型癌症險	最後	優先	最後
重大傷病險	最後	優先	最後

4.3.3 不同額度之保險規劃

首先,本段先將上述險種列舉出其必要理賠項目,以及在保險實務上一般大 眾規劃的建議額度,如表 4.6。

再根據上述之險種介紹、肺癌常見花費及對應險種與險種重要程度排序,利 用必要理賠項目,配置出基礎、一般、完整三種不同額度的規劃,方便後續不同 疾病預測機率及其他因素之病患做選擇,如表 4.7。

险種名稱		必要理賠項目	基礎規劃	一般規劃	完整規劃
醫療費用保險	實支實付型	住院醫療費用保險金	10-20 萬/次	20-30 萬/次	30-40 萬/次
		住院手術限額	5-10 萬/次	10-20 萬/次	20-25 萬/次
		門診手術限額	5-10 萬/次	10-20 萬/次	20-25 萬/次
	日額型	每日病房費用保險金	2,000 元/日	3,000 元/日	4,000 元/日
癌症險	療程型	癌症每日病房費用保險金	無	2,000 元/日	3,000 元/日
		癌症手術醫療保險金	無	5-10 萬/次	10-20 萬/次
	整筆給付型	重度癌症給付	無	無	200 萬
重大傷病險		重度癌症給付	無	無	150 萬

表 4.7: 不同額度之規劃

值得注意的是,在規劃險種時,一個險種雖可能涵蓋多個給付項目,但保險額度會是相同的,因此通常會選擇最必要的理賠項目計算需要的保額,故在此利用每個險種之必要項目來進行規劃。此外,此規劃的設計僅針對治療中的花費進行評估,不包含死亡及其後續可能產生的費用,也不會在終身及定期險種的選擇多加著墨。

4.3.4 影響規劃之其他因素

根據上述文獻探討之肺癌常見危險因子以及實務規劃經驗,在此列出影響規劃依據的其他因素,方便後續選擇方案時可以進行調整。由於其他因素並無出現在資料集中,故須藉由詢問的方式得知。

不吸菸者家族病史 除了上述文獻探討之引述,根據衛生福利部指出,對於不吸 菸卻罹患肺癌之病患,肺癌家族史是目前所知最重要的危險。另在「台灣

表 4.6: 險種必要給付項目與建議額度

险 種 名 稱	名稱	必要理賠項目	建議額度	選擇原因
殿 张 弗 田 亿 1公	實支實付型	住院醫療費用保險金	20-30 萬/次	大筆的藥費及耗材支出皆由此理賠,如:化放療、標靶藥物、免疫療法。
西 / / / / / / / / / / / / / / / / / / /		住院手術限額	10-20 萬/次	手術為癌症常見治療方式,恢復期短、傷口小的微創手術花費較高。
		門診手術限額	10-20 萬/次	越來越多手術不需住院即可治療。
	日額型	每日病房費用保險金	3,000 元/日	須長期治療
		n:		病房費的補貼,可讓病患在選擇治療環境時更無壓力。
	成 任 任	癌症每日病房費用保險金	3,000元/日	除一般病房費外,多了癌症住院日額可讓選擇更多樣。
癌症險	次在 ビ	癌症手術醫療保險金	10-20 萬/次	癌症病患身體非常虛弱,針對手術若是有另一筆理賠,更可
		ni		幫助病患有多餘的資金購買補品。
	整筆給付型	重度癌症給付	200 萬	通常大眾不一定有一筆緊急預備金,而一般醫療險須先行負
) (擔、療程結束後才能申請理賠,故整筆給付可幫助彌補一開
				始的資金缺口。加上可能接受新型的治療如標靶藥物,故將
		0	. (保額拉高。
重大傷病險	病險	重度癌症給付	月50 萬	原因同整筆給付之癌症險,而重大傷病險除了理賠範圍更加
				廣泛外,理賠不用依比例計算,原位癌以外不管第幾期的癌
				症都可給付。

不吸菸肺癌高危險群之研究(TALENT)」計畫之統計數據指出,在有肺癌家族病史與沒有肺癌家族病史的受試者中,罹患肺癌的機率分別為3.3%及2.0%,表示有肺癌家族病史族群相較沒有家族病史之患者,罹癌風險較高。

居住地點 居住於不同縣市,其醫院的病房費也會不同。根據各大醫院官網之單人房價錢,簡單分為北、中、南、東四個區域做比較。北部醫院單人房平均6,500元/日;中部醫院單人房平均5,000元/日;南部醫院單人房平均5,500元/日;東部醫院單人房平均3,000元/日。

家庭責任 有家庭責任者若是生病的話,家庭恐遭受劇變,龐大的經濟壓力隨之而來,因此,可將整筆給付型癌症險及重大傷病險補強,彌補不能工作之資金缺口。不論是作為薪資補償或是支付一開始龐大的醫藥費,對於有家庭責任的患者來說,幫助都非常大。

4.3.5 保險應用結論

本文研究旨在建立疾病預測模型,透過生活習慣與身體狀況評估罹患肺癌之 機率,並藉由此機率結合保險實務規劃經驗,選擇適合自己的保額及險種依據, 未來在購買保單時可以作為參考。綜上所述,本文最後的分析步驟如下:

首先,先利用建立好的疾病預測模型,輸入自己的資料,得到罹患肺癌之機率。接著,利用此機率分級判斷適合何種保額之規劃,若是有家族病史,根據前面的結論須調整機率的分級,故使用另外的判斷式作為依據。

機率分級	無家族病史	有家族病史	適合保額
低機率	0-25%	0-10%	不特別建議
中機率	26-50%	11-40%	基礎規劃
中高機率	51-75%	41-70%	一般規劃
高機率	76-100%	71-100%	完整規劃

表 4.8: 疾病預測模型與保險規劃之應用

最後,根據居住地點以及家庭責任資訊將選擇好的方案進行微調。居住地點 會影響未來可能居住的醫院單人房價錢,故若選擇的方案中,住院日額保額加上 癌症住院日額保額,並不足以支付其居住地點之單人房價錢,則建議在住院日額 保額上補齊差額。其微調公式如下:

$$P_N' = \begin{cases} P_N + P_L - P_C & \text{if } P_N + P_C < P_L \\ P_N & \text{otherwise.} \end{cases}$$

$$(4.1)$$

建議住院日額保額為 P_N ,建議癌症住院日額保額為 P_C ,居住地點所對應的單人房價格為 P_L ,調整後住院日額保額為 P_N'

至於家庭責任,則是以依賴人口來衡量。根據政府之定義,0-14歲及65歲以上人口為依賴人口。除原選規劃外,以一年薪資計算,若家中有依賴人口,則1個依賴人口,整筆給付型癌症險及重大傷病險之保額各多增加1倍之年薪,依此類推。其微調公式如下:

$$A'_{C} = A_{C} + R \cdot D$$

$$A'_{M} = A_{M} + R \cdot D$$

$$(4.2)$$

年薪為 R,依賴人口為 D,建議整筆給付癌症險保額為 A_C ,建議重大傷病保額為 A_M ,調整後整筆給付癌症險保額為 A_C' ,調整後重大傷病保額為 A_M'

Chengchi Univer

第五章 結論

本文利用機器學習及深度學習的方法建立疾病預測模型,以肺癌作為疾病標的,並結合保險規劃的實務經驗,探討疾病預測模型之於保險實務及醫療領域上的應用價值。期許此模型能在保險銷售人員對於醫療知識較不足的情況下,協助險種規劃更加精確;亦幫助肺癌病患對自己的身體狀況提高警覺,並解決其對於保險知識缺乏的痛點。

針對本文的資料集,使用召回率作為選擇模型的評估指標,利用七種不同的模型分析,選擇表現最好的深度學習之類神經網路模型 Neural Network Architecture 進行後續疾病預測。最後,利用預測出的罹患肺癌機率來判斷保單及保額的選擇,更結合保險實務上可能影響保單規劃的其他因子,打造更精細的保單,幫助找對最適合的險種規劃以及保額配置。

整體而言,疾病預測不僅可幫助保險業體現保險之價值,提升保險銷售人員的競爭力,更可為客戶端提供價值,協助找出潛在的肺癌病患及適合的保險規劃選擇依據,使病患免於承受龐大醫藥費之壓力。

本研究針對肺癌預測,原因在於肺癌是一個高度嚴重且具有挑戰性的疾病,對於個人和社會的健康都有重大的影響。本研究結果證明,機器學習方法在肺癌預測方面的有效性,對於保險應用及醫療領域具有重要意義。然而,除了肺癌之外,還有其他重要疾病值得進一步探討和研究,並應用於保險實務中。因此,我們期望未來能進一步擴大研究範圍,從多維度進行疾病預測,以更好地滿足保險實務的需要,這些研究成果的應用將對個人和社會健康帶來更大的價值和貢獻。

參考文獻

- [1] Z.-Q. Hong and J.-Y. Yang. Optimal discriminant plane for a small number of samples and design method of classifier on the plane. *Pattern Recognition*, 24(4):317–324, 1991.
- [2] J. ichi Nitadori, M. Inoue, M. Iwasaki, T. Otani, S. Sasazuki, K. Nagai, and S. Tsugane. Association between lung cancer incidence and family history of lung cancer: Data from a large-scale population-based cohort study, the jphc study. *Chest*, 130(4):968–975, 2006.
- [3] M. Kubat. An Introduction to Machine Learning. Springer, 9 2017.
- [4] S. S. More, V. B. Lobo, R. M. Laban, S. Panchal, M. Patil, and G. Pathak. Cancer prediction and insurance eligibility using machine learning techniques. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pages 725–730, 2020.
- [5] B. Reem, N. & Horacio. Current and future development in lung cancer diagnosis. *International Journal of Molecular Sciences*, 22(16), 2021.
- [6] G. Sruthi, C. L. Ram, M. K. Sai, B. P. Singh, N. Majhotra, and N. Sharma. Cancer prediction using machine learning. In 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), volume 2, pages 217–221, 2022.
- [7] World Cancer Research Fund International. Lung cancer statistics. Retrieved from https://www.wcrf.org/cancer-trends/lung-cancer-statistics/, 2021.

- [8] 中華民國保險經營學會. 壽險業經營健康保險業務. 取自 http://www.insop. org.tw/resources/forum/?p=23, 2013.
- [9] 李 岡 遠. 肺 癌 診 療 指 引. 取 自 https://www.cancertaipei.tw/wp-content/uploads/2022/03/03%E8%82%BA%E7%99%8C-%E4%B8%AD.pdf, 2021.
- [10] 衛生福利部中央健康保險署. 健保有條件給付之癌症治療標靶藥品. 取 自 https://www.nhi.gov.tw/Content_List.aspx?n=7C9F4EE71A4D5751& topn=787128DAD5F71B1A, 2022.
- [11] 衛生福利部中央健康保險署. 全民健康保險醫療服務給付項目及支付項目. 取 自 https://www.nhi.gov.tw/Content_List.aspx?n=7C9F4EE71A4D5751& topn=787128DAD5F71B1A, 2022.
- [12] 衛生福利部統計處. 110 年國人死因統計結果報告. 取自 https://dep.mohw. gov.tw/DOS/lp-5069-113-xCat-y110.html, 2022.
- [13] 金融監督管理委員會. 人壽保險單示範條款修正條文對照表. 取自 https://law.fsc.gov.tw/LawContent.aspx?id=GL002670, 2019.
- [14] 金融監督管理委員會. 住院醫療費用保險單示範條款(實支實付型)修正條文對照表. 取自 https://law.fsc.gov.tw/LawContent.aspx?id=GL002670, 2019.
- [15] 金融監督管理委員會. 住院醫療費用保險單示範條款(日額型)修正條文對 照表. 取自 https://law.fsc.gov.tw/LawContent.aspx?id=GL002670, 2019.