

癌种预测模型建立系统及建立方法、癌种预测系统

技术领域

本揭露是有关于一种癌种预测模型建立系统及建立方法、癌种预测系统。

背景技术

癌症可能发生在人体的许多器官，如肝脏、肾脏、肠胃道、大脑等。癌症可透过定期身体检查，早期发现早期治疗，以提高治愈效果。因此，如何通过设备实现癌种检测是本技术领域业者努力目标之一。

发明内容

因此，本揭露提出一种癌种预测模型建立系统及建立方法、癌种预测系统，可改善前述习知问题。

本揭露一实施例提出一种癌种预测模型建立系统。癌种预测模型建立系统包括第一层学习模型建立单元及第二层第一学习模型建立单元。第一层学习模型建立单元用以使用机器学习技术，依据多个第一学习检体的各者的第一拷贝数变异、第一检体癌种及第一性别，建立第一层学习模型；及，以多个第二学习检体的各者的第二性别及第二拷贝数变异做为第一层学习模型的输入，使第一层学习模型输出各第二学习检体的多个第一输出癌种。第二层第一学习模型建立单元用以使用机器学习技术，依据各第二学习检体的此些第一输出癌种及第二检体癌，建立第二层第一学习模型。

本揭露另一实施例提出一种癌种预测系统。癌种预测系统包括储存器及预测单元。储存器用以储存前述的第一层学习模型及第二层第一学习模型。预测单元用以输入待测检体的检体拷贝数变异及一检体性别于第一层学习模型，以取得待测检体的第一预测癌种；判断待测检体的检体性别是否为第二性别；以及，当待测检体的检体性别为第二性别，输入待测检体的第一预测癌种于第二层第一学习模型，以取得待测检体的第二预测癌种。

本揭露另一实施例提出一种癌种预测模型的建立方法。癌种预测模型的建立方法包括以下步骤：使用机器学习技术，依据多个第一学习检体的各者的第一拷贝数变异、第一检体癌种及第一性别，建立第一层学习模型；以多个第二学习检体的各者的一第二性别及第二拷贝数变异做为第一层学习模型的输入，第一层学习模型输出各第二学习检体的多个第一输出癌种；以及，使用机器学习技术，依据第二学习检体的此些第一输出癌种及一第二检体癌种，建立第二层第一学习模型。

为了对本揭露的上述及其他方面有更佳的了解，下文特举实施例，并配合所附图式详细说明如下：

附图说明

图 1 绘示依照本揭露一实施例的癌种预测模型建立系统的功能模块图。

图 1 绘示依照本发明一实施例的癌种预测系统的功能模块图。

图 3 绘示图 1 的癌种预测模型建立系统的癌种预测模型的建立方法的流程图。

图 4 绘示图 2 的癌种预测系统的癌种预测方法的流程图。

具体实施方式

请参照图 1，其图绘示依照本揭露一实施例的癌种预测模型建立系统 100 的功能模块图。癌种预测模型建立系统 100 包括第一层学习模型建立单元 110、第二层第一学习模型建立单元 120 及第二层第二学习模型建立单元 130。第一层学习模型建立单元 110、第二层第一学习模型建立单元 120 及/或第二层第二学习模型建立单元 130 例如是采用半导体制程所形成的实体电路。在一实施例中，第一层学习模型建立单元 110、第二层第一学习模型建立单元 120 与第二层第二学习模型建立单元 130 中至少二者可整合成单一单元。在一实施例中，第一层学习模型建立单元 110、第二层第一学习模型建立单元 120 与第二层第二学习模型建立单元 130 中至少一者可整合于一控制器(controller)或一处理器(processor)中。

如图 1 所示，第一层学习模型建立单元 110 用以：(1). 使用一

机器学习技术,依据多个第一学习检体 P1(未标示)的各者的第一拷贝数变异(Copy number variation, CNV) V_{P1} 、第一检体癌种(学习检体所属者的罹患癌种) $C1_{P1}$ 及第一性别(学习检体所属者的性别) S_{P1} , 建立第一层学习模型 M1; (2). 以多个第二学习检体 P2(未标示)的各者的第二性别 S_{P2} 及第二拷贝数变异 V_{P2} 做为第一层学习模型 M1 的输入,使第一层学习模型 M1 输出各第二学习检体 P2 的多个第一输出癌种 $C2_{P2}$ 。第二层第一学习模型建立单元 120 用以使用机器学习技术,依据各第二学习检体 P2 的此些第一输出癌种 $C2_{P2}$ 及第二检体癌种 $C1_{P2}$, 建立第二层第一学习模型 M21。在本实施例中, 第一层学习模型 M1 为混合型学习模型(例如, 不限性别), 因此可避免有些样本数少癌种所造成的学习(训练)误差。

如图 1 所示, 在一实施例中, 第一层学习模型建立单元 110 更用以: 使用机器学习技术, 依据多个健康检体 P4(未标示)的各者的第四拷贝数变异 V_{P4} 、健康类别 H_{P4} 及第四性别 S_{P4} , 建立第一层学习模型 M1。如此, 所建立的第一层学习模型 M1 更包含健康类别 H_{P4} , 可判断属于健康类别 H_{P4} 的待测检体。第四性别 S_{P4} 不限男性或女性。健康检体 P4 的基因组片段的拷贝数并未异常增加或减少(未有突变基因), 即, 其第四拷贝数变异 V_{P4} 是正常的。

此外, 第二层第一学习模型 M21 的建立除了可依据第一输出癌种 $C2_{P2}$ 及第二检体癌种 $C1_{P2}$ 外, 亦可依据其它检体信息建立, 例如是学习检体的年龄。举例来说, 如图 1 所示, 第二层第一学习模型建立单元 120 可依据各第二学习检体 P2 的此些第一输出癌种 $C2_{P2}$ 、第二检体癌种 $C1_{P2}$ 及年龄 G_{P2} , 建立第二层第一学习模型 M21。

如图 1 所示, 除了可依据第二性别 S_{P2} 建立第二层第一学习模型 M21 外, 亦可依据其它性别建立其它学习模型。举例来说, 第一层学习模型建立单元 110 更用以: 以多个第三学习检体 P3(未标示)的各者的第三性别 S_{P3} 及第三拷贝数变异 V_{P3} 做为该第一层学习模型的输入, 使第一层学习模型 M1 输出各第三学习检体 P3 的多个第二输出癌种 $C2_{P3}$ 。第二层第二学习模型建立单元 130 用以使

用机器学习技术，依据各第三学习检体 P_3 的此些第二输出癌种 C_{2P_3} 及第三检体癌种 C_{1P_3} ，建立第二层第二学习模型 M_{22} 。

此外，第二层第二学习模型 M_{22} 的建立除了可依据第二输出癌种 C_{2P_3} 外，亦可依据其它检体信息建立，例如是学习检体的年龄。举例来说，如图 1 所示，第二层第二学习模型建立单元 130 可依据各第三学习检体 P_3 的此些第二输出癌种 C_{2P_3} 、第三检体癌种 C_{1P_3} 及的年龄 G_{P_3} ，建立第二层第二学习模型 M_{22} 。

此外，本文的输出癌种(例如，第一输出癌种 C_{2P_2} 及/或第二输出癌种 C_{2P_3})例如可以机率型态表示。下表 1 列出第一层学习模型 M_1 所输出的多个第一输出癌种 C_{2P_2} 的机率。表 1 仅以 5 笔学习检体为例说明，然亦可多于 5 笔。如下表 1 所示，不同编号代表不同学习检体。此些第一输出癌种 C_{2P_2} 为不同癌种，例如是肝癌、摄护线癌、乳癌等。以编号#1 的检体来说，肝癌(第一输出癌种 C_{2P_2})的机率为 9.958%，摄护线癌的机率为 77.335%，而乳癌的机率为 0.011%，其余第一输出癌种 C_{2P_2} 并未列出，但都会有一机率数值。表 1 的数值仅为示例，并非用以限定本发明实施例。其它编号的解读方式类似编号#1，于此不再赘述。第二层第一学习模型建立单元 120 可依据表 1 的机率信息，建立第二层第一学习模型 M_{21} 。此外，第二层第二学习模型建立单元 130 也可依据类似表 1 的机率信息(第二输出癌种 C_{2P_3})，建立第二层第二学习模型 M_{22} 。

表 1

类别 检体	肝癌	摄护线癌	乳癌	...	健康
#1	0.09958	0.77335	0.000111	...	0.0002
#2	0.01061	0.20688	0.122995	...	0.0001
#3	0.12278	0.09811	0.057845	...	0.0001
#4	0.007043	0.36396	0.050012	...	0.0001
#5	0.007040	0.02467	0.000051	...	0.0001

此外，本文的机器学习技术例如是支持矢量机(Support Vector

Machine, SVM)。然，本发明实施例并不限定机器学习技术的种类，只要是能学习本文所要求的输入信息而能建立癌种学习模型者，皆可做为本文机器学习技术的应用。

此外，本文的学习检体例如是循环肿瘤细胞(Circulating Tumor Cell, CTC)。循环肿瘤细胞是指脱离肿瘤组织进入血液中的肿瘤细胞，不同器官来源的循环肿瘤细胞会偏好带有特定类型的基因突变。因此，透过检测循环肿瘤细胞带有哪些特定的基因突变类型，则可以推断此循环肿瘤细胞来自器官(癌种)。使用循环肿瘤细胞做为检体具有以下几个优点：(1). 抽血即可检查(不需物理切片、辐射摄像)，成本及风险都低；(2). 适于长期监控癌症复发；(3). 克服肿瘤的异质性；(4). 提早预测出远端转移；以及，(5). 快速地反应当下肿瘤状况。

此外，拷贝数变异是一种现象，其中基因组的各部分重复，并且基因组中的重复数在个体之间变化。具体而言，拷贝数变异是一种影响大量碱基对的复制或缺失事件。整个人类基因组的大约三分之二可能由重复组成，并且人类基因组的 4.8%~9.5% 可以归类为拷贝数变异。当基因组片段的拷贝数增加时，表示 CNV 异常。

此外，本文的学习检体例如是取自于美国癌症基因体图谱(The Cancer Genome Atlas, TCGA)，其记录各样本的检体癌症名称、所属年龄、CNV、性别等。TCGA 收录大约 34 种癌症名称。在一实施例中，34 种癌症名称可归纳或整并为例如 12 个癌种或更少，例如是脑癌、食道癌、肺癌、肾癌、男性癌症(癌症名称例如包含睪丸癌、摄护腺癌)、妇癌(癌症名称例如包含卵巢癌、子宫颈癌、子宫癌、子宫内膜癌)、肝癌、膀胱癌、前纵膈腔癌(癌症名称例如包含甲状腺癌、胸腺癌)、头颈癌、乳癌及/或胃肠道癌(癌症名称例如包含大肠癌、直肠癌、胰脏癌、胃癌)等。此外，本文的第一性别 S_{P1} 不限男性或女性，例如，这些第一学习检体 P1 所属者的第一性别可为男性与女性的组合。第二性别与第三性别不同。第二性别 S_{P2} 及第三性别 S_{P3} 限男性或女性，例如，全部第二学习检体 P2 所属者的第二性别 S_{P2} 是男性，而全部第三学习检体 P3 所属者的第

三性别 S_{P3} 是女性；或例如，全部第二学习检体 $P2$ 所属者的第二性别 S_{P2} 是女性，而全部第三学习检体 $P3$ 所属者的第三性别 S_{P3} 是男性。

请参照下表 2，在本实施例中，用于第一层学习模型 $M1$ 的样本数量、用于第二层第一学习模型 $M21$ 的样本数量、用于第二层第二学习模型 $M22$ 的样本数量与用于验证的样本数量的总和以 100%来说，其中用于学习模型的样本数量例如占 80%，而用于验证的样本数量例如占 20%。用于第一层学习模型 $M1$ 的样本数量、用于第二层第一学习模型 $M21$ 的样本数量、用于第二层第二学习模型 $M22$ 的样本数量的总和以 100%来说，当中的第一层学习模型 $M1$ 的样本数量占 80%，而用于第二层第一学习模型 $M21$ 的样本数量与用于第二层第二学习模型 $M22$ 的样本数量的总和占 20%。此些样本数量可取自于 TCGA、医院及/或政府单位等。上述 80%、20%等比例仅为示例，非用以限定本发明实施例。

表 2

第一层模型 $M1$ 的样本数	第二层第一学习模型 $M21$ 的样本数 + 第二层第二学习模型 $M22$ 的样本数	用于验证的样本数
100%		
80%		20%
80%	20%	-

在取得第一层学习模型 $M1$ 、第二层第一学习模型 $M21$ 及第二层第二学习模型 $M22$ 后，可使用此些学习模型，取得至少一待测检体的预测癌种，以下进一步举例说明。

请参照图 2，其绘示依照本发明一实施例的癌种预测系统 200 的功能模块图。癌种预测系统 200 包括储存器 210 及预测单元 220。储存器 210 及预测单元 220 例如是采用半导体制程所形成的实体电路。具体而言，储存器 210 可以是存储器，其可整合于预测单元

220 或与预测单元 220 分开配置。在一实施例中，预测单元 220 可整合于一控制器或一处理器中。

储存器 210 用以储存第一层学习模型 M1 及第二层第一学习模型 M21。预测单元 220 用以：(1). 输入待测检体 PT 的检体拷贝数变异 V_{PT} 及检体性别 S_{PT} 于第一层学习模型 M1，以取得待测检体 PT 的第一预测癌种 $C1_{PT}$ ；(2). 判断待测检体 PT 的检体性别 S_{PT} 是否为第二性别 S_{P2} ；以及，(3). 当待测检体 PT 的检体性别为第二性别 S_{P2} ，输入待测检体 PT 的第一预测癌种 $C1_{PT}$ 于第二层第一学习模型 M21，以取得待测检体 PT 的第二预测癌种 $C2_{PT}$ 。

前述检体性别 S_{PT} 不限男性或女性，换言之，检体性别 S_{PT} 可以是男性或女性。

此外，储存器 210 更用以储存前述第二层第二学习模型 M22。预测单元 220 更用以：(1). 判断待测检体 PT 的检体性别 S_{PT} 是否为第三性别 S_{P3} ；以及，(2). 当待测检体 PT 的检体性别 S_{PT} 为第三性别 S_{P3} 时，输入待测检体 PT 的第一预测癌种 $C1_{PT}$ 于第二层第二学习模型 M22，以取得待测检体 PT 的第二预测癌种 $C2_{PT}$ 。

以下说明多个待测检体的第二预测癌种 $C2_{PT}$ 的准确性。如下表 3-1 及 3-2，其列出多个待测检体的第二预测癌种 $C2_{PT}$ 与实际罹患癌种的数据分析，如灵敏度 Sens、特异度 Spec、阳性预测值 (Positive predictive value, PPV) 及阴性预测值 (Negative Predictive Value, NPV)。针对食道癌，由于其位置特殊(刚好介于头颈和肠胃之间)，故只要预测结果是头颈癌或胃肠道癌，都属正确预测。

表 3-1

类别	男性检体验证结果							
	Sens		Spec		PPV		NPV	
	T1	T3	T1	T3	T1	T3	T1	T3
健康	1	1	0.93	0.93	0.68	0.68	1	1
前纵膈癌	0.19	0.44	0.99	0.99	0.45	0.67	0.97	0.98

胃肠道癌	0.68	0.92	0.96	0.96	0.71	0.77	0.95	0.99
肾癌	0.83	0.88	0.97	0.97	0.78	0.79	0.98	0.98
肝癌	0.66	0.9	0.99	0.99	0.87	0.9	0.98	0.99
肺癌	0.74	0.89	0.93	0.93	0.59	0.63	0.96	0.98
男性癌症	0.81	0.96	0.97	0.97	0.78	0.81	0.97	0.99
脑癌	0.87	0.96	0.99	0.99	0.91	0.92	0.98	0.99
乳癌	0	0.5	1	1	Null	1	1	1
头颈癌	0.49	0.85	0.97	0.97	0.55	0.68	0.96	0.99
膀胱癌	0.65	0.74	1	1	0.97	0.97	0.98	0.98
女性癌症	Null	Null	1	1	Null	Null	1	1
食道癌	0.83	0.92	1	1	1	1	0.99	1

表 3-2

类别	女性检体验证结果							
	Sens		Spec		PPV		NPV	
	T1	T3	T1	T3	T1	T3	T1	T3
健康	1	1	0.91	0.91	0.5	0.5	1	1
前纵膈癌	0.16	0.89	0.99	0.99	0.68	0.92	0.94	0.99
胃肠道癌	0.77	0.91	0.97	0.97	0.72	0.75	0.98	0.99
肾癌	0.61	0.74	0.99	0.99	0.77	0.8	0.98	0.99
肝癌	0.53	0.62	1	1	0.75	0.78	0.99	0.99
肺癌	0.57	0.85	0.96	0.96	0.56	0.65	0.96	0.99
男性癌症	Null	Null	1	1	Null	Null	1	1

脑 癌	0.9	0.91	0.99	0.99	0.92	0.93	0.98	0.99
乳 癌	0.79	0.95	0.96	0.96	0.82	0.85	0.95	0.99
头颈癌	0.32	0.74	1	1	0.69	0.83	0.98	0.99
膀胱癌	0.23	0.38	1	1	0.75	0.83	0.98	0.99
女性癌症	0.8	0.98	0.89	0.89	0.73	0.77	0.93	0.99
食道癌	0.43	0.43	1	1	1	1	0.99	1

灵敏度 Sens、特异度 Spec、阳性预测值 PPV 及阴性预测值 NPV 分别以下式(1)~(4)表示，其中参数 TP、FP、FN 及 TN 的定义说明于表 4。

$$\text{Sens}=\text{TP}/(\text{TP}+\text{FN})\dots(1)$$

$$\text{Spec}=\text{TN}/(\text{FP}+\text{TN})\dots(2)$$

$$\text{PPV}=\text{TP}/(\text{TP}+\text{FP})\dots\dots(3)$$

$$\text{NPV}=\text{TN}/(\text{FN}+\text{TN})\dots(4)$$

表 4

实际情况 预测结果	有罹癌	无罹癌
阳	检测结果为 A 癌症， 且实际情况有 A 癌症 (真阳性，TP)	检测结果为 A 癌症， 但实际情况并没有 A 癌症 (伪阳性，FP)
阴	检测结果没有 A 癌症， 但实际情况有 A 癌症 (伪阴性，FN)	检测结果没有 A 癌症， 且实际情况也没有 A 癌症 (真阴性，TN)

此外，上表 3-1 及 3-2 的 T1(Top 1)指得是：以 TP 来说，当实际情况也是 A 癌症且检测结果中机率最高的也是 A 癌症时，则 TP 值加 1；当实际情况是 A 癌症但检测结果中机率最高者并非 A 癌症时，则 TP 的值加 0。T3(Top 3)指的是：以 TP 来说，当实际情况是 A 癌症且检测结果中机率最高的前 3 名也包含 A 癌症时，则 TP 值加 1；当实际情况是 A 癌症但检测结果中机率最高的前 3 名都不包含 A 癌症时，则 TP 值加 0。

以表 3-1 及 3-2 的分析数据来看，对于一些癌种(例如胃肠道癌、肝癌、男性癌症、脑癌、乳癌、女性癌症等)，灵敏度 Sens 高于 90%。在早筛的应用里，建议灵敏度要摆在最高优先度。

如下表 5 所示，以第一层学习模型 M1 来说，预测男性罹癌的准确度达 68.5%，而预测女性罹癌的准确度达 69.8%。以第二层学习模型来说，预测男性罹癌的准确度提高至 72.91%，而女性罹癌的准确度提高至 71.21%。

表 5

性别	第一层学习模型 M1	第二层学习模型
男性	68.5 %	72.91 %
女性	69.8 %	71.21 %

请参照图 3，其绘示图 1 的癌种预测模型建立系统 100 的癌种预测模型的建立方法的流程图。

在步骤 S110 中，请同时参照图 1，第一层学习模型建立单元 110 使用机器学习技术，依据多个第一学习检体 P1 的各者的第一拷贝数变异 V_{P1} 、第一检体癌种 $C1_{P1}$ 及第一性别 S_{P1} ，建立第一层学习模型 M1。在一实施例中，请同时参照图 1，第一层学习模型建立单元 110 更用以：使用机器学习技术，依据多个健康检体 P4(未标示)的各者的第四拷贝数变异 V_{P4} 、健康类别 H_{P4} 及第四性别 S_{P4} ，建立第一层学习模型 M1。

在步骤 S120 中，请同时参照图 1，以多个第二学习检体 P2 的各者的第二性别 S_{P2} 及第二拷贝数变异 V_{P2} 做为第一层学习模型

M1 的输入，第一层学习模型 M1 输出各第二学习检体 P2 的多个第一输出癌种 $C2_{P2}$ 。

在步骤 S130 中，请同时参照图 1，第二层第一学习模型建立单元 120 使用机器学习技术，依据各第二学习检体 P2 的此些第一输出癌种 $C2_{P2}$ 及第二检体癌种 $C1_{P2}$ ，建立第二层第一学习模型 M21。在一实施例中，第二层第一学习模型建立单元 120 更可依据各第二学习检体 P2 的年龄 G_{P2} ，建立第二层第一学习模型 M21。

在步骤 S140 中，请同时参照图 1，第二层第二学习模型建立单元 130 以多个第三学习检体 P3 的各者的第三性别 S_{P3} 及第三拷贝数变异 V_{P3} 做为第一层学习模型 M1 的输入，第一层学习模型 M1 输出对应的第二输出癌种 $C2_{P3}$ 。

在步骤 S150 中，请同时参照图 1，第二层第二学习模型建立单元 130 使用机器学习技术，依据各第三学习检体 P3 的此些第二输出癌种 $C2_{P3}$ 及第三检体癌种 $C1_{P3}$ ，建立第二层第二学习模型 M22。在一实施例中，第二层第二学习模型建立单元 130 更可依据各第三学习检体 P3 的年龄 G_{P3} ，建立第二层第二学习模型 M22。

请参照图 4，其绘示图 2 的癌种预测系统 200 的癌种预测方法的流程图。

在步骤 S210 中，请同时参照图 1，预测单元 220 输入待测检体 PT 的检体拷贝数变异 V_{PT} 及检体性别 S_{PT} 于第一层学习模型 M1，以取得待测检体 PT 的第一预测癌种 $C1_{PT}$ 。

在步骤 S220 中，请同时参照图 1，预测单元 220 判断待测检体 PT 的检体性别 S_{PT} 是否为第二性别 S_{P2} 。若待测检体 PT 的检体性别 S_{PT} 为第二性别 S_{P2} ，流程进入步骤 S230；若否，流程进入步骤 S240。

在步骤 S230 中，请同时参照图 1，预测单元 220 输入待测检体 PT 的第一预测癌种 $C1_{PT}$ 于第二层第一学习模型 M21，以取得待测检体 PT 的第二预测癌种 $C2_{PT}$ 。

在步骤 S240 中，预测单元 220 判断待测检体 PT 的检体性别 S_{PT} 是否为第三性别 S_{P3} 。若待测检体 PT 的检体性别 S_{PT} 为第三性

别 S_{P3} ，流程进入步骤 S250。

在步骤 S250 中，预测单元 220 输入待测检体 PT 的第一预测癌种 $C1_{PT}$ 于第二层第二学习模型 M22，以取得待测检体 PT 的第二预测癌种 $C2_{PT}$ 。

在实施例中，检体性别 S_{PT} 不是第二性别 S_{P2} 就是第三性别 S_{P3} ，因此流程亦可省略步骤 S240。

综上，本揭露一实施例揭露癌种预测模型建立系统、癌种预测系统及应用其的方法。癌种预测模型建立系统利用机器学习技术建立二层学习模型。癌种预测系统使用此二层学习模型，可简便、快速地预测待测检体的癌种。在一实施例中，第一层学习模型例如是混合型学习模型(例如，不限性别)，因此可避免有些癌种的样本数少所造成的学习(训练)误差。第二层学习模型可针对特定性别预测癌种。

综上所述，虽然本揭露已以实施例揭露如上，然其并非用以限定本揭露。本揭露所属技术领域中具有通常知识者，在不脱离本揭露的精神和范围内，当可作各种的更动与润饰。因此，本揭露的保护范围当视权利要求书所界定者为准。

权 利 要 求 书

1、一种癌种预测模型建立系统，其特征在于，包括：

第一层学习模型建立单元，用以：

5 使用机器学习技术，依据多个第一学习检体的各者的第一拷贝数变异、第一检体癌种及第一性别，建立第一层学习模型；

 以多个第二学习检体的各者的第二性别及第二拷贝数变异做为该第一层学习模型的输入，使该第一层学习模型输出各该第二学习检体的多个第一输出癌种；以及

10 第二层第一学习模型建立单元，用以：

 使用该机器学习技术，依据各该第二学习检体的该些第一输出癌种及第二检体癌种，建立第二层第一学习模型。

 2、如权利要求1所述的癌种预测模型建立系统，其特征在于，该些第一学习检体的该些第一性别为男性与女性的组合，而该些第二学习检体的该些第二性别全部为男性或女性。

 3、如权利要求1所述的癌种预测模型建立系统，其特征在于，该第一层学习模型建立单元更用以：

 以多个第三学习检体的各者的第三性别及第三拷贝数变异做为该第一层学习模型的输入，使该第一层学习模型输出各该第三学习检体的多个第二输出癌种；

 其中，该癌种预测模型建立系统更包括：

 第二层第二学习模型建立单元，用以：

 使用该机器学习技术，依据各该第三学习检体的该些第二输出癌种及第三检体癌种，建立第二层第二学习模型；

25 其中该些第二学习检体的该些第三性别全部为男性或女性，且该第二性别与该第三性别不同。

 4、如权利要求3所述的癌种预测模型建立系统，其特征在于，该第二层第二学习模型建立单元更用以：

 依据各该第三学习检体的年龄，建立该第二层第二学习模型；

30 其中，该第二层第一学习模型建立单元更用以：

依据各该第二学习检体的年龄，建立该第二层第一学习模型。

5、一种癌种预测系统，其特征在于，包括：

存储器，用以：储存如权利要求1所述的该第一层学习模型及该第二层第一学习模型；

5 预测单元，用以：

输入待测检体的检体拷贝数变异及检体性别于该第一层学习模型，以取得该待测检体的多个第一预测癌种；

判断该待测检体的检体性别是否为该第二性别；以及

10 当该待测检体的该检体性别为该第二性别，输入该待测检体的该第一预测癌种于该第二层第一学习模型，以取得该待测检体的第二预测癌种。

6、如权利要求5所述的癌种预测系统，其特征在于，该存储器更用以储存如权利要求3所述的该第二层第二学习模型；该预测单元更用以：

15 判断该待测检体的该检体性别是否为第三性别；以及

当该待测检体的该检体性别为该第三性别时，输入该待测检体的该第一预测癌种于该第二层第二学习模型，以取得该待测检体的该第二预测癌种。

7、一种癌种预测模型的建立方法，其特征在于，包括：

20 使用机器学习技术，依据多个第一学习检体的各者的第一拷贝数变异、第一癌种及第一性别，建立第一层学习模型；

以多个第二学习检体的各者的第二性别及第二拷贝数变异做为该第一层学习模型的输入，该第一层学习模型输出各该第二学习检体的多个第一输出癌种；以及

25 使用该机器学习技术，依据各该第二学习检体的该些第一输出癌种及第二检体癌种，建立第二层第一学习模型。

8、如权利要求7所述的建立方法，其特征在于，更包括：

30 以多个第三学习检体的各者的第三性别及第三拷贝数变异做为该第一层学习模型的输入，使该第一层学习模型输出对应的第三癌种；以及

使用该机器学习技术，依据该些第三癌种，建立第二层第二学习模型；

其中，该第二性别与该第三性别不同。

9、如权利要求8所述的建立方法，其特征在于，于依据该些第三癌种建立该第二层第二学习模型的步骤更包括：

依据各该第三学习检体的年龄，建立该第二层第二学习模型；

其中，于依据各该第二学习检体的该些第一输出癌种及该第二检体癌种建立该第二层第一学习模型的步骤更包括：

依据各该第二学习检体的年龄，建立该第二层第一学习模型。