



(12) 发明专利申请

(10) 申请公布号 CN 116848548 A

(43) 申请公布日 2023. 10. 03

(21) 申请号 202180092579.7

(22) 申请日 2021.11.30

(30) 优先权数据

63/120102 2020.12.01 US

(85) PCT国际申请进入国家阶段日

2023.08.01

(86) PCT国际申请的申请数据

PCT/IB2021/061125 2021.11.30

(87) PCT国际申请的公布数据

W02022/118190 EN 2022.06.09

(71) 申请人 强生企业创新公司

地址 美国新泽西州

(72) 发明人 G·小沃斯科 R·S·J·埃斯特帕

C·M·金赛 C·S·史蒂文森

(74) 专利代理机构 中国专利代理(香港)有限公司 72001

专利代理师 陈新娟 吕传奇

(51) Int.Cl.

G06T 7/00 (2017.01)

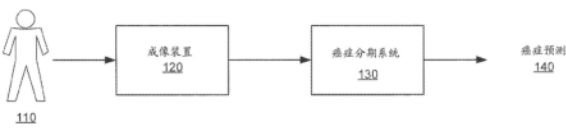
权利要求书9页 说明书24页 附图30页

(54) 发明名称

预测淋巴结的癌累及的机器学习

(57) 摘要

本文公开了用于确定转移癌的受试者水平风险的方法,其包括训练和/或部署模型以确定: 1) 个体淋巴结累及的淋巴结水平风险;以及/或者2) 淋巴结累及的受试者水平风险。因此,该方法可识别具有结节病的高风险或低风险的患者,并且任选地使得能够例如经由治疗来指导癌症患者的干预。



100

图1A

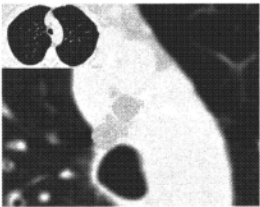


图2B

1. 一种确定受试者的转移癌的受试者水平风险的方法,所述方法包括:
获得从所述受试者捕获的包括所述受试者的多个淋巴结的一个或多个图像;以及
通过将风险模型应用于所获得的一个或多个图像的提取的特征来预测转移癌的所述受试者水平风险,所述风险模型被训练用于针对图像中的淋巴结预测转移癌的淋巴结(LN)水平风险,
其中所述风险模型是至少使用从针对参考个体的淋巴结集合的转移癌的插补LN水平风险导出的标签来训练的,所述淋巴结集合的所述插补LN水平风险是至少使用区分训练图像的癌性淋巴结和非癌性淋巴结的插补模型来确定的。
2. 根据权利要求1所述的方法,其中预测转移癌的所述受试者水平风险还包括:
基于由所述风险模型预测的所述多个淋巴结的转移癌的LN水平风险,选择所述多个淋巴结中的一个或多个淋巴结;以及
使用针对所述一个或多个淋巴结预测的转移癌的所述LN水平风险来确定转移癌的所述受试者水平风险。
3. 根据权利要求2所述的方法,其中选择所述多个淋巴结中的一个或多个淋巴结包括识别具有最高概率的LN水平风险的所述淋巴结,并且其中确定转移癌的所述受试者水平风险包括将所述最高概率的LN水平风险指定为转移癌的所述受试者水平风险。
4. 根据权利要求1至3中任一项所述的方法,其中所述风险模型以比由所述插补模型预测的所述插补LN水平风险更大范围的概率预测转移癌的LN水平风险。
5. 根据权利要求1至4中任一项所述的方法,其中通过以下步骤选择参考个体的所述淋巴结集合:
确定中位风险值;以及
将所述参考个体的所述淋巴结的插补LN水平风险与所述中位风险值进行比较。
6. 根据权利要求5所述的方法,其中进一步通过以下方式选择参考个体的所述淋巴结集合:
在所述淋巴结集合中包括具有大于所述中位风险值的LN水平风险的一个或多个淋巴结。
7. 根据权利要求5或6所述的方法,其中将所述淋巴结集合中的具有小于所述中位风险值的LN水平风险的一个或多个淋巴结从所述淋巴结集合中排除。
8. 根据权利要求5至7中任一项所述的方法,其中所述中位风险值是所述参考个体的中位受试者内LN水平风险。
9. 根据权利要求1至8中任一项所述的方法,其中所述一个或多个图像包括计算机断层扫描(CT)图像。
10. 根据权利要求1至9中任一项所述的方法,其中所述一个或多个图像是从胸廓CT扫描获得的。
11. 根据权利要求1至10中任一项所述的方法,还包括根据转移癌的所预测的受试者水平风险来确定癌症的分期。
12. 根据权利要求1至11中任一项所述的方法,还包括根据转移癌的所预测的受试者水平风险针对所述受试者选择诊断或治疗。
13. 根据权利要求12所述的方法,其中针对所述受试者选择诊断或治疗包括当转移癌

的所预测的受试者水平风险低于阈值风险值时,选择手术肿瘤切除或组合支气管镜或内窥镜诊断和治疗。

14.根据权利要求12所述的方法,其中针对所述受试者选择诊断或治疗包括当转移癌的所预测的受试者水平风险高于阈值风险值时,从执行PET或PET-CT扫描或执行淋巴结活检中的一者或两者选择附加诊断测试。

15.根据权利要求12所述的方法,其中针对所述受试者选择诊断或治疗包括当转移癌的所预测的受试者水平风险低于阈值风险值时,选择PET或PET-CT扫描的附加诊断测试。

16.根据权利要求1至15中任一项所述的方法,其中使用从胸廓CT扫描获得的训练图像来训练所述插补模型。

17.根据权利要求16所述的方法,其中从胸廓CT扫描获得的所述训练图像包括一个或多个淋巴结。

18.根据权利要求1至17中任一项所述的方法,其中使用国家肺癌筛查试验(NLST)的训练图像来训练所述插补模型。

19.根据权利要求1至17中任一项所述的方法,其中使用定制数据集的训练图像来训练所述插补模型。

20.根据权利要求16至19中任一项所述的方法,其中通过从所述训练图像中的至少一个训练图像生成两个或更多个放射组学组来训练所述插补模型。

21.根据权利要求20所述的方法,其中所述两个或更多个放射组学组包括所述训练图像的高斯-拉普拉斯变换和所述训练图像的小波变换。

22.根据权利要求20或21所述的方法,其中通过以下方式来进一步训练所述插补模型:从所述两个或更多个放射组学组中的每个放射组学组提取特征;以及至少使用所提取的特征来训练所述插补模型。

23.根据权利要求20或21所述的方法,其中通过以下方式进一步训练所述插补模型:
在所述放射组学组中限定感兴趣区域(ROI),所限定的感兴趣区域包括淋巴结;
从所述放射组学组中的所述ROI中提取特征;以及
至少使用所提取的特征来训练所述插补模型。

24.根据权利要求22或23所述的方法,其中所提取的特征包括一阶特征、形状特征、灰度共生矩阵1(GLCM)特征、灰度游程长度矩阵(GLRLM)特征、灰度区域大小矩阵(GLSZM)特征和邻域灰度调差分矩阵(NGTDM)特征的一个或多个特征类别。

25.根据权利要求16至24中任一项所述的方法,其中使用在训练图像中指示转移的存在或不存在的参考基础真值来进一步训练所述插补模型。

26.根据权利要求25所述的方法,其中所述参考基础真值是从指示所述癌症的所述分期的Nstage可协变值导出的,其中“0”或“1”癌症分期指示不存在转移,并且其中“2”或“3”癌症分期指示存在转移。

27.根据权利要求26所述的方法,其中所述参考基础真值指示个体淋巴结是癌性的还是非癌性的。

28.根据权利要求27所述的方法,其中基于所述肿瘤或淋巴结的特性从所述Nstage可协变值反向计算指示个体淋巴结是癌性的还是非癌性的所述参考基础真值。

29.根据权利要求1至28中任一项所述的方法,其中所述插补模型是随机森林分类器或

LASSO分类器。

30. 根据权利要求1至29中任一项所述的方法,其中所述风险模型是随机森林分类器或LASSO分类器。

31. 根据权利要求30所述的方法,其中使用监督学习技术来训练所述风险模型。

32. 根据权利要求1至31中任一项所述的方法,其中所述多个淋巴结包括2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个或21个淋巴结。

33. 根据权利要求1至31中任一项所述的方法,其中所述多个淋巴结包括8个至20个淋巴结、10个至19个淋巴结、12个至18个淋巴结或14个至16个淋巴结。

34. 根据权利要求1至33中任一项所述的方法,其中所述转移癌是转移肺癌。

35. 根据权利要求1至34中任一项所述的方法,其中所述风险模型表现出至少 $AUC = 78\%$ 、灵敏度 $= 82\%$ 或特异性 $= 60\%$ 的性能。

36. 一种训练权利要求1至35中任一项的所述风险模型的方法,所述方法包括:

应用插补模型以针对训练图像中的多个淋巴结生成插补LN水平风险;

选择淋巴结集合,其中所述集合中的所述淋巴结具有大于中位风险值的LN水平风险;

使用所述淋巴结集合中的所述淋巴结的所述LN水平风险作为用于训练所述风险模型的参考基础真值,所述风险模型能够以比由所述插补模型预测的所述插补LN水平风险更大范围的概率预测LN水平风险。

37. 根据权利要求36所述的方法,其中通过以下方式来进一步训练所述风险模型:从一个或多个放射组学组提取特征;以及至少使用所提取的特征来训练所述风险模型。

38. 根据权利要求37所述的方法,其中从两个或更多个放射组学组提取特征包括:

在所述一个或多个放射组学组中限定感兴趣区域(ROI),所限定的感兴趣区域包括淋巴结;

从所述一个或多个放射组学组中的所述ROI中提取特征。

39. 根据权利要求37或38所述的方法,其中所提取的特征包括一阶特征、形状特征、灰度共生矩阵1 (GLCM) 特征、灰度游程长度矩阵 (GLRLM) 特征、灰度区域大小矩阵 (GLSZM) 特征和邻域灰色调差分矩阵 (NGTDM) 特征的一个或多个特征类别。

40. 一种用于确定受试者的转移癌的受试者水平风险的系统,所述系统包括:

成像装置,所述成像装置被配置成捕获所述受试者的一个或多个图像;和

计算装置,所述计算装置被配置成执行以下步骤:

获得从所述受试者捕获的包括所述受试者的多个淋巴结的一个或多个图像;以及

通过将风险模型应用于所获得的一个或多个图像的提取的特征来预测转移癌的所述受试者水平风险,所述风险模型被训练用于针对图像中的淋巴结预测转移癌的淋巴结(LN)水平风险,

其中所述风险模型是至少使用从针对参考个体的淋巴结集合的转移癌的插补LN水平风险导出的标签来训练的,所述淋巴结集合的所述插补LN水平风险是至少使用区分训练图像的癌性淋巴结和非癌性淋巴结的插补模型来确定的。

41. 根据权利要求40所述的系统,其中预测转移癌的所述受试者水平风险还包括:

基于由所述风险模型预测的所述多个淋巴结的转移癌的LN水平风险,选择所述多个淋

巴结中的一个或多个淋巴结;以及

使用针对所述一个或多个淋巴结预测的转移癌的所述LN水平风险来确定转移癌的所述受试者水平风险。

42. 根据权利要求41所述的系统,其中选择所述多个淋巴结中的一个或多个淋巴结包括识别具有最高概率的LN水平风险的所述淋巴结,并且其中确定转移癌的所述受试者水平风险包括将所述最高概率的LN水平风险指定为转移癌的所述受试者水平风险。

43. 根据权利要求40至42中任一项所述的系统,其中所述风险模型以比由所述插补模型预测的所述插补LN水平风险更大范围的概率预测转移癌的LN水平风险。

44. 根据权利要求40至43中任一项所述的系统,其中通过以下步骤选择参考个体的所述淋巴结集合:

确定中位风险值;以及

将所述参考个体的所述淋巴结的插补LN水平风险与所述中位风险值进行比较。

45. 根据权利要求44所述的系统,其中进一步通过以下方式选择参考个体的所述淋巴结集合:

在所述淋巴结集合中包括具有大于所述中位风险值的LN水平风险的一个或多个淋巴结。

46. 根据权利要求44或45所述的系统,其中将所述淋巴结集合中的具有小于所述中位风险值的LN水平风险的一个或多个淋巴结从所述淋巴结集合中排除。

47. 根据权利要求44至46中任一项所述的系统,其中所述中位风险值是所述参考个体的中位受试者内LN水平风险。

48. 根据权利要求40至47中任一项所述的系统,其中所述成像装置是计算机断层扫描(CT)扫描仪、磁共振成像(MRI)扫描仪、正电子发射断层扫描(PET)扫描仪、X射线扫描仪或超声成像装置中的一者。

49. 根据权利要求40至48中任一项所述的系统,其中所述成像装置是CT扫描仪,并且其中所述一个或多个图像包括计算机断层扫描(CT)图像。

50. 根据权利要求40至49中任一项所述的系统,其中所述一个或多个图像包括胸廓CT图像。

51. 根据权利要求40至50中任一项所述的系统,其中所述计算装置被进一步配置为执行以下步骤:根据转移癌的所预测的受试者水平风险来确定癌症的分期。

52. 根据权利要求40至51中任一项所述的系统,其中所述计算装置被进一步配置为执行以下步骤:根据转移癌的所预测的受试者水平风险针对所述受试者选择诊断或治疗。

53. 根据权利要求52所述的系统,其中针对所述受试者选择诊断或治疗包括当转移癌的所预测的受试者水平风险低于阈值风险值时,选择手术肿瘤切除或组合支气管镜或内窥镜诊断和治疗。

54. 根据权利要求53所述的系统,其中针对所述受试者选择诊断或治疗包括当转移癌的所预测的受试者水平风险高于阈值风险值时,选择执行PET或PET-CT扫描或执行淋巴结活检中的一者或两者的附加诊断测试。

55. 根据权利要求53所述的系统,其中针对所述受试者选择诊断或治疗包括当转移癌的所预测的受试者水平风险低于阈值风险值时,选择PET或PET-CT扫描的附加诊断测试。

56. 根据权利要求40至54中任一项所述的系统,其中使用从胸廓CT扫描获得的训练图像来训练所述插补模型。

57. 根据权利要求56所述的系统,其中从胸廓CT扫描获得的所述训练图像包括一个或多个淋巴结。

58. 根据权利要求40至57中任一项所述的系统,其中使用国家肺癌筛查试验(NLST)的训练图像来训练所述插补模型。

59. 根据权利要求40至57中任一项所述的系统,其中使用定制数据集的训练图像来训练所述插补模型。

60. 根据权利要求56至59中任一项所述的系统,其中通过从所述训练图像中的至少一个训练图像生成两个或更多个放射组学组来训练所述插补模型。

61. 根据权利要求60所述的系统,其中所述两个或更多个放射组学组包括所述训练图像的高斯-拉普拉斯变换和所述训练图像的小波变换。

62. 根据权利要求60或61所述的系统,其中通过以下方式来进一步训练所述插补模型:从所述两个或更多个放射组学组中的每个放射组学组提取特征;以及至少使用所提取的特征来训练所述插补模型。

63. 根据权利要求60或61所述的系统,其中通过以下方式进一步训练所述插补模型:

在所述放射组学组中限定感兴趣区域(ROI),所限定的感兴趣区域包括淋巴结;

从所述放射组学组中的所述ROI中提取特征;以及

至少使用所提取的特征来训练所述插补模型。

64. 根据权利要求62或63所述的系统,其中所提取的特征包括一阶特征、形状特征、灰度共生矩阵1(GLCM)特征、灰度游程长度矩阵(GLRLM)特征、灰度区域大小矩阵(GLSZM)特征和邻域灰色调差分矩阵(NGTDM)特征的一个或多个特征类别。

65. 根据权利要求56至64中任一项所述的系统,其中使用在训练图像中指示转移的存在或不存在的参考基础真值来进一步训练所述插补模型。

66. 根据权利要求65所述的系统,其中所述参考基础真值是从指示所述癌症的所述分期的Nstage可协变值导出的,其中“0”或“1”癌症分期指示不存在转移,并且其中“2”或“3”癌症分期指示存在转移。

67. 根据权利要求66所述的系统,其中所述参考基础真值指示个体淋巴结是癌性的还是非癌性的。

68. 根据权利要求67所述的系统,其中基于所述肿瘤或淋巴结的特性从所述Nstage可协变值反向计算指示个体淋巴结是癌性的还是非癌性的所述参考基础真值。

69. 根据权利要求40至68中任一项所述的系统,其中所述插补模型是随机森林分类器或LASSO分类器。

70. 根据权利要求40至69中任一项所述的系统,其中所述风险模型是随机森林分类器或LASSO分类器。

71. 根据权利要求70所述的系统,其中使用监督学习技术来训练所述风险模型。

72. 根据权利要求40至71中任一项所述的系统,其中所述多个淋巴结包括2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个或21个淋巴结。

73. 根据权利要求40至72中任一项所述的系统,其中所述多个淋巴结包括8个至20个淋巴结、10个至19个淋巴结、12个至18个淋巴结或14个至16个淋巴结。

74. 根据权利要求40至73中任一项所述的系统,其中所述转移癌是转移肺癌。

75. 根据权利要求40至74中任一项所述的系统,其中所述风险模型表现出至少 $AUC=78\%$ 、灵敏度 $=82\%$ 或特异性 $=60\%$ 的性能。

76. 根据权利要求40至75中任一项所述的系统,其中通过以下方式训练所述风险模型:

应用插补模型以针对训练图像中的多个淋巴结生成插补LN水平风险;

选择淋巴结集合,其中所述集合中的所述淋巴结具有大于中位风险值的LN水平风险;

使用所述淋巴结集合中的所述淋巴结的所述LN水平风险作为用于训练所述风险模型的参考基础真值,所述风险模型能够以比由所述插补模型预测的所述插补LN水平风险更大范围的概率预测LN水平风险。

77. 根据权利要求76所述的系统,其中通过以下方式来进一步训练所述风险模型:从一个或多个放射组学组提取特征;以及至少使用所提取的特征来训练所述风险模型。

78. 根据权利要求77所述的系统,其中从两个或更多个放射组学组提取特征包括:

在所述一个或多个放射组学组中限定感兴趣区域(ROI),所限定的感兴趣区域包括淋巴结;

从所述一个或多个放射组学组中的所述ROI中提取特征。

79. 根据权利要求77或78所述的系统,其中所提取的特征包括一阶特征、形状特征、灰度共生矩阵1 (GLCM) 特征、灰度游程长度矩阵 (GLRLM) 特征、灰度区域大小矩阵 (GLSZM) 特征和邻域灰度调差分矩阵 (NGTDM) 特征的一个或多个特征类别。

80. 一种非暂态计算机可读介质,所述非暂态计算机可读介质包括指令,所述指令在由处理器执行时使所述处理器:

获得从受试者捕获的包括所述受试者的多个淋巴结的一个或多个图像;以及

通过将风险模型应用于所获得的一个或多个图像的提取的特征来预测转移癌的受试者水平风险,所述风险模型被训练用于针对图像中的淋巴结预测转移癌的淋巴结(LN)水平风险,

其中所述风险模型是至少使用从针对参考个体的淋巴结集合的转移癌的插补LN水平风险导出的标签来训练的,所述淋巴结集合的所述插补LN水平风险是至少使用区分训练图像的癌性淋巴结和非癌性淋巴结的插补模型来确定的。

81. 根据权利要求80所述的非暂态计算机可读介质,其中使所述处理器预测转移癌的所述受试者水平风险的指令还包括当由所述处理器执行时使所述处理器进行以下操作的指令:

基于由所述风险模型预测的所述多个淋巴结的转移癌的LN水平风险,选择所述多个淋巴结中的一个或多个淋巴结;以及

使用针对所述一个或多个淋巴结预测的转移癌的所述LN水平风险来确定转移癌的所述受试者水平风险。

82. 根据权利要求81所述的非暂态计算机可读介质,其中使所述处理器选择所述多个淋巴结中的一个或多个淋巴结的指令还包括当由所述处理器执行时使所述处理器识别具有最高概率的LN水平风险的所述淋巴结的指令,并且其中使所述处理器确定转移癌的所述

受试者水平风险的指令包括当由所述处理器执行时使所述处理器将所述最高概率的LN水平风险指定为转移癌的所述受试者水平风险的指令。

83. 根据权利要求80至82中任一项所述的非暂态计算机可读介质,其中所述风险模型以比由所述插补模型预测的所述插补LN水平风险更大范围的概率预测转移癌的LN水平风险。

84. 根据权利要求80至83中任一项所述的非暂态计算机可读介质,其中通过以下步骤选择参考个体的所述淋巴结集合:

确定中位风险值;以及

将所述参考个体的所述淋巴结的插补LN水平风险与所述中位风险值进行比较。

85. 根据权利要求84所述的非暂态计算机可读介质,其中进一步通过以下方式选择参考个体的所述淋巴结集合:在所述淋巴结集合中包括具有大于所述中位风险值的LN水平风险的一个或多个淋巴结。

86. 根据权利要求84或85所述的非暂态计算机可读介质,其中将所述淋巴结集合中的具有小于所述中位风险值的LN水平风险的一个或多个淋巴结从所述淋巴结集合中排除。

87. 根据权利要求84至86中任一项所述的非暂态计算机可读介质,其中所述中位风险值是所述参考个体的中位受试者内LN水平风险。

88. 根据权利要求80至87中任一项所述的非暂态计算机可读介质,其中所述一个或多个图像包括计算机断层扫描(CT)图像。

89. 根据权利要求80至88中任一项所述的非暂态计算机可读介质,其中所述一个或多个图像是从胸廓CT扫描获得的。

90. 根据权利要求80至89中任一项所述的非暂态计算机可读介质,还包括在由所述处理器执行时使所述处理器根据转移癌的所预测的受试者水平风险来确定癌症的分期的指令。

91. 根据权利要求80至90中任一项所述的非暂态计算机可读介质,还包括在由所述处理器执行时使所述处理器根据转移癌的所预测的受试者水平风险来针对所述受试者选择诊断或治疗的指令。

92. 根据权利要求91所述的非暂态计算机可读介质,其中使所述处理器针对所述受试者选择诊断或治疗的所述指令还包括在由所述处理器执行时使所述处理器在转移癌的所预测的受试者水平风险低于阈值风险值时选择手术肿瘤切除或组合支气管镜或内窥镜诊断和治疗的指令。

93. 根据权利要求91所述的非暂态计算机可读介质,其中使所述处理器针对所述受试者选择诊断或治疗的所述指令还包括在由所述处理器执行时使所述处理器在转移癌的所预测的受试者水平风险高于阈值风险值时选择执行PET或PET-CT扫描或执行淋巴结活检中的一者或两者的附加诊断测试的指令。

94. 根据权利要求91所述的非暂态计算机可读介质,其中使所述处理器针对所述受试者选择诊断或治疗的所述指令还包括在由所述处理器执行时使所述处理器在转移癌的所预测的受试者水平风险低于阈值风险值时选择PET或PET-CT扫描的附加诊断测试的指令。

95. 根据权利要求80至94中任一项所述的非暂态计算机可读介质,其中使用从胸廓CT扫描获得的训练图像来训练所述插补模型。

96. 根据权利要求95所述的非暂态计算机可读介质, 其中从胸廓CT扫描获得的所述训练图像包括一个或多个淋巴结。

97. 根据权利要求80至96中任一项所述的非暂态计算机可读介质, 其中使用国家肺癌筛查试验(NLST)的训练图像来训练所述插补模型。

98. 根据权利要求80至96中任一项所述的非暂态计算机可读介质, 其中使用定制数据集的训练图像来训练所述插补模型。

99. 根据权利要求95至98中任一项所述的非暂态计算机可读介质, 其中通过从所述训练图像中的至少一个训练图像生成两个或更多个放射组学组来训练所述插补模型。

100. 根据权利要求99所述的非暂态计算机可读介质, 其中所述两个或更多个放射组学组包括所述训练图像的高斯-拉普拉斯变换和所述训练图像的小波变换。

101. 根据权利要求99或100所述的非暂态计算机可读介质, 其中通过以下方式来进一步训练所述插补模型: 从所述两个或更多个放射组学组中的每个放射组学组提取特征; 以及至少使用所提取的特征来训练所述插补模型。

102. 根据权利要求99或100所述的非暂态计算机可读介质, 其中通过以下方式进一步训练所述插补模型:

在所述放射组学组中限定感兴趣区域(ROI), 所限定的感兴趣区域包括淋巴结;

从所述放射组学组中的所述ROI中提取特征; 以及

至少使用所提取的特征来训练所述插补模型。

103. 根据权利要求101或102所述的非暂态计算机可读介质, 其中所提取的特征包括一阶特征、形状特征、灰度共生矩阵1 (GLCM) 特征、灰度游程长度矩阵 (GLRLM) 特征、灰度区域大小矩阵 (GLSZM) 特征和邻域灰度调差分矩阵 (NGTDM) 特征的一个或多个特征类别。

104. 根据权利要求95至103中任一项所述的非暂态计算机可读介质, 其中使用在训练图像中指示转移的存在或不存在的参考基础真值来进一步训练所述插补模型。

105. 根据权利要求104所述的非暂态计算机可读介质, 其中所述参考基础真值是从指示所述癌症的所述分期的Nstage可协变值导出的, 其中“0”或“1”癌症分期指示不存在转移, 并且其中“2”或“3”

癌症分期指示存在转移。

106. 根据权利要求105所述的非暂态计算机可读介质, 其中所述参考基础真值指示个体淋巴结是癌性的还是非癌性的。

107. 根据权利要求106所述的非暂态计算机可读介质, 其中基于所述肿瘤或淋巴结的特性从所述Nstage可协变值反向计算指示个体淋巴结是癌性的还是非癌性的所述参考基础真值。

108. 根据权利要求80至107中任一项所述的非暂态计算机可读介质, 其中所述插补模型是随机森林分类器或LASSO分类器。

109. 根据权利要求80至108中任一项所述的非暂态计算机可读介质, 其中所述风险模型是随机森林分类器或LASSO分类器。

110. 根据权利要求109所述的非暂态计算机可读介质, 其中使用监督学习技术来训练所述风险模型。

111. 根据权利要求80至110中任一项所述的非暂态计算机可读介质, 其中所述多个淋

巴结包括2个、3个、4个、5个、6个、7个、8个、

9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个或21个淋巴结。

112. 根据权利要求80至111中任一项所述的非暂态计算机可读介质,其中所述多个淋巴结包括8个至20个淋巴结、10个至19个淋巴结、

12个至18个淋巴结或14个至16个淋巴结。

113. 根据权利要求80至112中任一项所述的非暂态计算机可读介质,其中所述转移癌是转移肺癌。

114. 根据权利要求80至113中任一项所述的非暂态计算机可读介质,其中所述风险模型表现出至少 $AUC=78\%$ 、灵敏度 $=82\%$ 或特异性 $=60\%$ 的性能。

115. 根据权利要求80至114中任一项所述的非暂态计算机可读介质,其中通过以下方式训练所述风险模型:

应用插补模型以针对训练图像中的多个淋巴结生成插补LN水平风险;

选择淋巴结集合,其中所述集合中的所述淋巴结具有大于中位风险值的LN水平风险;

使用所述淋巴结集合中的所述淋巴结的所述LN水平风险作为用于训练所述风险模型的参考基础真值,所述风险模型能够以比由所述插补模型预测的所述插补LN水平风险更大范围的概率预测LN水平风险。

116. 根据权利要求115所述的非暂态计算机可读介质,其中通过以下方式来进一步训练所述风险模型:从一个或多个放射组学组中提取特征;以及至少使用所提取的特征来训练所述风险模型。

117. 根据权利要求116所述的非暂态计算机可读介质,其中从两个或更多个放射组学组提取特征包括:

在所述一个或多个放射组学组中限定感兴趣区域(ROI),所限定的感兴趣区域包括淋巴结;

从所述一个或多个放射组学组中的所述ROI中提取特征。

118. 根据权利要求115或116所述的非暂态计算机可读介质,其中所提取的特征包括一阶特征、形状特征、灰度共生矩阵1 (GLCM) 特征、灰度游程长度矩阵 (GLRLM) 特征、灰度区域大小矩阵 (GLSZM) 特征和邻域灰色调差分矩阵 (NGTDM) 特征的一个或多个特征类别。

预测淋巴结的癌累及的机器学习

[0001] 相关申请的交叉引用

[0002] 本申请要求2020年12月1日提交的美国临时专利申请63/120,102号的权益和优先权,该临时专利申请的全部公开内容据此全文以引用方式并入以用于所有目的。

背景技术

[0003] 胸部的计算机断层扫描(CT)成像已经成为用于评估处于肺癌风险中的那些人的临床标准。这些努力集中在肺结节的检测上,肺结节可能代表疾病的最早期和最可治愈的分期。此类肺结节的检测触发了更广泛的临床评估,以确定其是否是癌性的,并且如果是癌性的,则其是否扩散到胸部中的其他结构或甚至身体的远侧区域。这样做是为了确定肺癌的分期,该分期又决定了治疗干预的计划。

[0004] 肺癌分期的一个方面是确定其在胸部淋巴结中的存在或不存在(结节状态)。当前的护理标准涉及分期的方法,其中首先确定结节是良性的还是恶性的,然后进行附加测试,诸如正电子发射断层扫描(PET)和基于外科手术、内窥镜或支气管镜的淋巴结取样。组织取样程序增加了后处理的风险、费用和时间。需要新的非侵入性方法来减轻这些风险并缩短诊断和治疗之间的时间。如本文进一步详细描述,应用于用于肺癌筛查的初始胸廓CT扫描的基于机器学习和深度学习的技术可以解决这些需要。

[0005] 肺癌检测、分期和治疗计划是可能花费数周或数月的多步骤过程。除了此类临床工作的风险和费用之外,该过程所需的额外时间导致显著的患者焦虑和提高的间期转移机会。事实上,较长的诊断到治疗时间与较差的总体存活率有关。减轻这些问题的示例性方法是允许几乎同时进行癌症诊断、分期以及在支气管镜疗法的情况下允许治疗的方法。

发明内容

[0006] 本文公开的本发明的实施方案基于用于在高风险人群中筛查肺癌的胸廓CT扫描的分析。在一些方面,其识别具有转移到胸部中的至少一个淋巴结(LN)并且因此需要附加测试(诸如PET扫描和淋巴结活检)的癌症高风险的患者。这允许临床提供者优先对那些最需要的人进行测试。相反,淋巴结转移风险低的疑似肺癌患者可直接进行即时治疗,诸如手术切除或组合支气管镜诊断和治疗。这将消除疾病诊断和治疗之间的时间。具体地,本文公开的本发明的实施方案使得能够识别具有结节病的高风险或低风险的患者,从而允许临床医生优先在高风险个体中进行附加的测试或使具有低转移风险的肺癌患者直接进行确认支气管镜活检,随后立即进行支气管镜递送治疗。

[0007] 本文公开了确定受试者的转移癌的受试者水平风险的方法,该方法包括:获得从受试者捕获的一个或多个图像,该一个或多个图像包括受试者的多个淋巴结;以及通过将风险模型应用于所获得的一个或多个图像的提取的特征来预测转移癌的受试者水平风险,该风险模型被训练用于针对图像中的淋巴结预测转移癌的淋巴结(LN)水平风险,其中风险模型是至少使用从针对参考个体的淋巴结集合的转移癌的插补LN水平风险导出的标签来训练的,该淋巴结集合的插补LN水平风险是至少使用区分训练图像的癌性淋巴结和非癌性

淋巴结的插补模型来确定的。在各种实施方案中,预测转移癌的受试者水平风险还包括:基于由风险模型预测的多个淋巴结的转移癌的LN水平风险,选择该多个淋巴结中的一个或多个淋巴结;以及使用针对该一个或多个淋巴结预测的转移癌的LN水平风险来确定转移癌的受试者水平风险。

[0008] 在各种实施方案中,选择该多个淋巴结中的一个或多个淋巴结包括识别具有最高概率的LN水平风险的淋巴结,并且其中确定转移癌的受试者水平风险包括将最高概率的LN水平风险指定为转移癌的受试者水平风险。在各种实施方案中,风险模型以比由插补模型预测的插补LN水平风险更大范围的概率预测转移癌的LN水平风险。在各种实施方案中,通过以下方式选择参考个体的淋巴结集合:确定中位风险值;以及将参考个体的淋巴结的插补LN水平风险与该中位风险值进行比较。在各种实施方案中,进一步通过以下方式选择参考个体的淋巴结集合:在淋巴结集合中包括具有大于中位风险值的LN水平风险的一个或多个淋巴结。在各种实施方案中,将淋巴结集合中的具有小于中位风险值的LN水平风险的一个或多个淋巴结从淋巴结集合中排除。

[0009] 在各种实施方案中,中位风险值是参考个体的中位受试者内LN水平风险。在各种实施方案中,该一个或多个图像包括计算机断层扫描(CT)图像。在各种实施方案中,该一个或多个图像是从胸廓CT扫描获得的。在各种实施方案中,本文公开的方法还包括根据转移癌的预测的受试者水平风险确定癌症的分期。在各种实施方案中,本文公开的方法还包括根据转移癌的预测的受试者水平风险针对受试者选择诊断或治疗。在各种实施方案中,针对受试者选择诊断或治疗包括当转移癌的预测的受试者水平风险低于阈值风险值时,选择手术肿瘤切除或组合支气管镜或内窥镜诊断和治疗。在各种实施方案中,针对受试者选择诊断或治疗包括当转移癌的预测的受试者水平风险高于阈值风险值时,选择执行PET或PET-CT扫描或执行淋巴结活检中的一者或两者的附加诊断测试。

[0010] 在各种实施方案中,使用从胸廓CT扫描获得的训练图像来训练插补模型。在各种实施方案中,从胸廓CT扫描获得的训练图像包括一个或多个淋巴结。在各种实施方案中,使用国家肺癌筛查试验(NLST)的训练图像来训练插补模型。在各种实施方案中,使用定制数据集的训练图像来训练插补模型。

[0011] 在各种实施方案中,通过从训练图像中的至少一个训练图像生成两个或更多个放射组学组来训练插补模型。在各种实施方案中,两个或更多个放射组学组包括训练图像的高斯-拉普拉斯变换和训练图像的小波变换。在各种实施方案中,通过以下方式来进一步训练插补模型:从两个或更多个放射组学组中的每个放射组学组提取特征;以及至少使用所提取的特征来训练插补模型。在各种实施方案中,通过以下方式进一步训练插补模型:在放射组学组中限定感兴趣区域(ROI),该限定的感兴趣区域包括淋巴结;从放射组学组中的ROI中提取特征;以及至少使用所提取的特征来训练插补模型。在各种实施方案中,提取的特征包括一阶特征、形状特征、灰度共生矩阵1(GLCM)特征、灰度游程长度矩阵(GLRLM)特征、灰度区域大小矩阵(GLSZM)特征和邻域灰度调差分矩阵(NGTDM)特征的一个或多个特征类别。

[0012] 在各种实施方案中,使用在训练图像中指示转移的存在或不存在的参考基础真值来进一步训练插补模型。在各种实施方案中,参考基础真值是从指示癌症的分期的Nstage可协变值导出的,其中“0”或“1”癌症分期指示不存在转移,并且其中“2”或“3”癌症分期指

示存在转移。在各种实施方案中,参考基础真值指示个体淋巴结是癌性的还是非癌性的。在各种实施方案中,基于肿瘤或淋巴结的特性从Nstage可协变值反向计算指示个体淋巴结是癌性的还是非癌性的参考基础真值。

[0013] 在各种实施方案中,插补模型是随机森林分类器或LASSO分类器。在各种实施方案中,风险模型是随机森林分类器或LASSO分类器。在各种实施方案中,使用监督学习技术来训练风险模型。在各种实施方案中,多个淋巴结包括1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个或21个淋巴结。在各种实施方案中,多个淋巴结包括8个至20个淋巴结、10个至19个淋巴结、12个至18个淋巴结或14个至16个淋巴结。在各种实施方案中,转移癌是转移肺癌。在各种实施方案中,风险模型表现出至少AUC=78%、灵敏度=82%或特异性=60%的性能。

[0014] 本文还公开了一种训练权利要求1至35中任一项的风险模型的方法,该方法包括:应用插补模型以针对训练图像中的多个淋巴结生成插补LN水平风险;选择淋巴结集合,其中该集合中的淋巴结具有大于中位风险值的LN水平风险;使用淋巴结集合中的淋巴结的LN水平风险作为用于训练风险模型的参考基础真值,该风险模型能够以比由插补模型预测的插补LN水平风险更大范围的概率预测LN水平风险。在各种实施方案中,通过以下方式来进一步训练风险模型:从一个或多个放射组学组提取特征;以及至少使用所提取的特征来训练风险模型。在各种实施方案中,从两个或更多个放射组学组提取特征包括:在该一个或多个放射组学组中限定感兴趣区域(ROI),该限定的感兴趣区域包括淋巴结;从该一个或多个放射组学组中的ROI中提取特征。在各种实施方案中,提取的特征包括一阶特征、形状特征、灰度共生矩阵1 (GLCM) 特征、灰度游程长度矩阵 (GLRLM) 特征、灰度区域大小矩阵 (GLSZM) 特征和邻域灰度调差分矩阵 (NGTDM) 特征的一个或多个特征类别。

[0015] 本文另外公开了一种用于执行上述方法的系统。在各种实施方案中,一种用于确定受试者的转移癌的受试者水平风险的系统包括:成像装置,该成像装置被配置成捕获受试者的一个或多个图像;以及计算装置,该计算装置被配置成执行以下步骤:获得从受试者捕获的包括受试者的多个淋巴结的一个或多个图像;以及通过将风险模型应用于所获得的一个或多个图像的提取的特征来预测转移癌的受试者水平风险,该风险模型被训练用于针对图像中的淋巴结预测转移癌的淋巴结 (LN) 水平风险,其中风险模型是至少使用从针对参考个体的淋巴结集合的转移癌的插补LN水平风险导出的标签来训练的,该淋巴结集合的插补LN水平风险是至少使用区分训练图像的癌性淋巴结和非癌性淋巴结的插补模型来确定的。

[0016] 本文另外公开了一种用于执行上述方法的非暂态计算机可读介质。在各种实施方案中,一种非暂态计算机可读介质包括指令,该指令在由处理器执行时致使处理器:获得从受试者捕获的一个或多个图像,该一个或多个图像包括受试者的多个淋巴结;以及通过将风险模型应用于所获得的一个或多个图像的提取的特征来预测转移癌的受试者水平风险,该风险模型被训练用于针对图像中的淋巴结预测转移癌的淋巴结 (LN) 水平风险,其中风险模型是至少使用从针对参考个体的淋巴结集合的转移癌的插补LN水平风险导出的标签来训练的,该淋巴结集合的插补LN水平风险是至少使用区分训练图像的癌性淋巴结和非癌性淋巴结的插补模型来确定的。

附图说明

[0017] 根据以下描述和附图,本发明的这些和其它特征、方面和优点将变得更好理解。

[0018] 图 (FIG.) 1A 描绘了根据一个实施方案的通过确定受试者的转移癌的受试者水平风险进行癌症分期的系统环境概览。

[0019] 图 1B 描绘了根据一个实施方案的癌症分期系统的框图。

[0020] 图 2A 描绘了根据一个实施方案的识别淋巴结站的示例性淋巴结标测图。

[0021] 图 2B 是根据一个实施方案的示例性淋巴结标注。

[0022] 图 3A 描绘了根据一个实施方案的插补模型和风险模型的示例性实施方式。

[0023] 图 3B 根据一个实施方案更详细地描绘了风险模型的实施方式。

[0024] 图 4 示出了用于实现图 1A、图 1B、图 2A、图 2B、图 3A 和图 3B 所示的实体的示例性计算机。

[0025] 图 5A 至图 5R 描绘了不同淋巴结放射组学特征的示例性相关性热图。

[0026] 图 6 描绘了插补模型的淋巴结水平性能。

[0027] 图 7A 和图 7B 描绘了使用由 LN 水平分类器预测的最高概率得分的受试者水平性能。

[0028] 图 8 描绘了在两步自助方法中的插补模型的淋巴结水平性能。

[0029] 图 9A 和图 9B 描绘了使用自助方法的受试者水平性能。

[0030] 应当指出的是,只要可行,相似或相同的参考标号可用于附图并且可表示相似或相同的功能。例如,在参考标号之后的字母诸如“个体 315A”指示该文本具体指代具有该特定参考标号的元素。文本中没有后面的字母的参考标号诸如“个体 315”指的是图中具有该参考标号的任何或所有元素(例如,文本中的“个体 315”指的是图中的参考标号“个体 315A”、“个体 315B”、“个体 315C”和“个体 315D”)。

具体实施方式

[0031] I. 定义

[0032] 除非另外指明,否则在权利要求和说明书中使用的术语按如下所阐述的那样定义。

[0033] 术语“受试者”包括细胞、组织或生物体、人或非人,而不论是体内、离体还是体外,男性或是女性。

[0034] 术语“哺乳动物”涵盖人和非人两者,并且包括但不限于人、非人灵长类、犬、猫、鼠、牛、马和猪。

[0035] 术语“样品”或“测试样品”可包括单个细胞或多个细胞或细胞碎片或体液的等分试样(诸如血液样品),其通过包括静脉穿刺、排泄、射精、按摩、活检、针抽吸、灌洗样品、刮痧、手术切口或干预或本领域已知的其他手段的手段从受试者获取。体液的等分试样的示例包括羊水、房水、胆汁、淋巴液、母乳、间质液、血液、血浆、耵聍(耳垢)、考珀液(预射精液)、乳糜、食糜、女性潮射、月经、粘液、唾液、尿液、呕吐物、泪液、阴道润滑液、汗液、血清、精液、皮脂、脓、胸膜液、脑脊液、滑液、细胞内液和玻璃体液。

[0036] 术语“参考个体”是指具有已知癌症分期的个体。例如,参考个体可以包括健康个体(例如,对照个体)。作为另一个示例,参考个体可包括先前诊断患有癌症并分期的个体。作为另一个示例,参考个体包括先前诊断患有已转移至淋巴结的癌症(例如,IV 期癌症)的

个体。

[0037] 术语“获得一个或多个图像”涵盖获得从受试者捕获的一个或多个图像或获得从取自受试者的样品捕获的一个或多个图像。获得一个或多个图像可涵盖执行从受试者或从取自受试者的样品捕获该一个或多个图像的步骤。该短语还可涵盖例如从第三方接收一个或多个图像，该第三方已执行从受试者或从取自受试者的样品捕获该一个或多个图像的步骤。本领域技术人员可经由多种已知方式（包括存储在存储内存上）获得该一个或多个图像。

[0038] 术语“淋巴结累及”、“淋巴结转移”或“淋巴结癌”是指在受试者的一个或多个淋巴结中存在或不存在转移的癌。

[0039] 术语“淋巴结转移风险”、“淋巴结水平风险”和“LN水平风险”可互换使用，并且通常是指在特定淋巴结中的转移癌的风险。

[0040] 短语“转移癌的受试者水平风险”或“转移癌的患者水平风险”可互换使用，并且是指受试者中的癌症已经转移的可能性。在各种实施方案中，受试者中的癌症是已经在受试者的一个或多个淋巴结中转移的肺癌。

[0041] 术语“癌症分期”或“癌症的分期”是指受试者中癌症的分期。癌症分期可以指包括0-4期的TNM分期系统。另外，癌症分期可指结节性癌症0-3期。结节性癌症0期指示无结节累及，结节性癌症1期指示肺门结节累及，并且结节2期或3期指示纵隔累及。

[0042] 必须注意，除非上下文明确地指明，否则如本说明书所用，单数形式“一个”、“一种”、“该”和“所述”包括复数指代物。

[0043] II. 系统环境概览

[0044] 图1A描绘了根据一个实施方案的通过确定受试者的转移癌的受试者水平风险进行癌症分期的系统环境概览100。系统环境100提供上下文以便介绍受试者110、成像装置120和用于确定癌症预测140的癌症分期系统130。在各种实施方案中，由癌症分期系统130确定的癌症预测140用作受试者110的癌症分期。在各种实施方案中，由癌症分期系统130确定的癌症预测140是受试者110的预测的风险水平并且充当用于确定受试者110的癌症分期的基础。

[0045] 在各种实施方案中，受试者是健康的。因此，本文所述用于癌症分期的方法可有益于健康受试者的癌症的早期检测。在各种实施方案中，受试者先前被诊断为患有癌症。在各种实施方案中，受试者是健康的，但怀疑患有一种形式的癌症（例如，已经表现出与癌症相关联的症状的受试者）。在特定实施方案中，受试者的癌症的类型为肺癌。因此，本文描述的用于癌症分期的方法可以有益于识别和区分处于转移癌（包括淋巴结累及）的高风险的受试者和处于具有淋巴结转移的低风险的疑似肺癌的受试者。处于高转移癌风险下的受试者可以经历另外的测试诸如PET扫描和淋巴结活检，而处于淋巴结转移低风险下的受试者可以进行立即治疗，诸如肺癌的手术切除或组合支气管镜诊断和治疗。

[0046] 在各种实施方案中，成像装置120捕获来自受试者110的图像用于进一步分析。在各种实施方案中，成像装置120从取自受试者110的测试样品捕获图像。图像和/或样品可由个体或由第三方例如医疗保健提供者或医疗专业人员获得。医疗专业人员的示例包括医师、急诊技师、护士、现场急救员、心理学家、采血师、医学物理学人员、执业护士、外科医生、牙医以及本领域技术人员已知的任何其他公认医疗专业人员。

[0047] 在一些实施方案中,成像装置120捕获受试者110的解剖位置的图像。受试者的示例性解剖位置可包括肺、胸腔、肾、肝、胰腺、脑、胃、肠、臀部、膝盖、腿、手臂和面部。在各种实施方案中,成像装置120捕获受试者110的胸腔的图像。在各种实施方案中,成像装置120捕获受试者110的一个或多个淋巴结的图像。在各种实施方案中,成像装置120捕获包括受试者110的一个或多个淋巴结的胸腔的图像。

[0048] 在各种实施方案中,成像装置120是计算机断层扫描(CT)扫描仪、磁共振成像(MRI)扫描仪、正电子发射断层扫描(PET)扫描仪、X射线扫描仪或超声成像装置中的一者。在特定实施方案中,成像装置120是捕获受试者110的一个或多个图像的CT扫描仪。在特定实施方案中,成像装置120是CT扫描仪,该CT扫描仪捕获包括受试者110的一个或多个淋巴结的胸腔的一个或多个CT图像。

[0049] 一般来讲,癌症分期系统130分析一个或多个图像(例如,由成像装置120捕获的图像)并生成癌症预测140。在各种实施方案中,癌症预测140可以是受试者110中癌症的预测分期。在各种实施方案中,癌症预测140可以是受试者110中癌症淋巴结累及的预测可能性。在各种实施方案中,癌症预测140可以包括推荐的临床方法(例如,附加测试诸如PET扫描和淋巴结活检)或治疗(例如,手术切除和/或药物治疗)。癌症分期系统130可包括一个或多个计算机,该一个或多个计算机体现为如下文相对于图4所讨论的计算机系统400。因此,在各种实施方案中,在计算机上(in silico)执行参考癌症分期系统130所述的步骤。

[0050] 在各种实施方案中,成像装置120和癌症分期系统130由不同方采用。例如,第一方操作成像装置120以从受试者110捕获一个或多个图像,然后将所捕获的一个或多个图像提供给第二方,该第二方实现癌症分期系统130以确定癌症预测140。

[0051] 现在参考图1B,该图描绘了根据实施方案的示出癌症分期系统130的计算机逻辑部件的框图。具体地,癌症分期系统130可以包括淋巴结特征模块145、插补模块150、风险训练模块155、风险部署模块160、癌症预测模块165和训练数据存储库170。

[0052] 一般来讲,淋巴结特征模块145从源自受试者或训练受试者的图像中提取特征。插补模块150训练并部署插补模型。插补模型分析提取的特征并针对图像中的每个淋巴结确定癌症的淋巴结水平风险。风险训练模块155使用训练数据来训练风险模型,该训练数据至少包括由插补模型确定的癌症的淋巴结水平风险。风险部署模块160实施风险模型以分析从获自受试者(例如,图1A中的受试者110)的图像中提取的特征,以确定受试者110的癌症预测,诸如受试者水平转移风险或癌症分期的预测。换句话说,癌症分期系统130训练并实施至少两个模型(例如,插补模型和风险模型)。因此,通过涉及该至少两个模型的该多步骤实施方式,癌症分期系统130针对受试者生成癌症预测,诸如受试者水平转移风险或癌症分期的预测。

[0053] 插补模型和风险模型两者的实施导致改善的癌症预测。具体地,插补模型从训练数据的受试者水平N期学习,并且应用于从淋巴结提取的特征,以定义哪些淋巴结可能有助于阳性N期。然后,风险模型从插补的数据学习更精细的淋巴结转移风险概率。这里,风险模型以比由插补模型预测的插补LN水平风险更大范围的概率预测转移癌的LN水平风险。这使得能够更准确地确定转移癌的受试者水平风险。

[0054] 下文参考如下两个阶段描述癌症分期系统130的部件:1) 训练阶段和2) 部署阶段。更具体地,训练阶段是指基于训练数据诸如从参考个体(例如,具有已知癌症分期的个体)

捕获的训练图像建立和训练一个或多个模型。因此,使用训练数据来训练模型,使得在部署阶段期间,模块的实施使得能够预测受试者(例如,图1A中的受试者110)的癌症分期。在各种实施方案中,还可实施插补模块150以训练和部署模型,该模型在下文称为插补模型。因此,插补模块150在训练阶段期间训练插补模型,并且在部署阶段期间部署插补模型。

[0055] 在一些实施方案中,在训练阶段和部署阶段中的一者期间应用癌症分期系统130的部件。例如,在训练阶段期间应用风险训练模块155和训练数据存储库170以训练风险模型。另外,在部署阶段期间应用风险部署模块160。在各种实施方案中,癌症分期系统130的部件可由不同方执行,这取决于是否在训练阶段或部署阶段期间应用这些部件。在此类场景中,预测模型的训练和部署由不同方执行。例如,在训练阶段期间应用的风险训练模块155和训练数据存储库170可由第一方采用(例如,以训练风险模型),并且在部署阶段期间应用的风险部署模块160可由第二方执行(例如,以部署风险模型)。类似地,插补模块150可被应用于由第一方训练插补模型,并且插补模块150可被应用于由第二方部署插补模型。

[0056] III. 用于癌症分期的方法

[0057] 本文描述的实施方案包括用于通过分析从受试者捕获的一个或多个图像并且应用一个或多个模型以确定涉及淋巴结的转移癌风险的受试者水平进行癌症分期的方法。此类方法可由图1B中所述的癌症分期系统130执行。

[0058] 参照图1B中的癌症分期系统130的模块,淋巴结特征模块145分析一个或多个图像并从该一个或多个图像提取特征。在各种实施方案中,淋巴结特征模块145从获自参考个体的训练图像提取特征。从训练图像中提取的这些特征可用于训练一个或多个模型,诸如下文进一步详细描述插补模型。在各种实施方案中,淋巴结特征模块145提取获自受试者(例如,图1A中的受试者110)的特征图像。可以通过部署模型诸如风险模型来分析从获自受试者的图像中提取的这些特征,如下面进一步详细描述的。

[0059] 在各种实施方案中,一个或多个图像或训练图像包括图像中的一个或多个淋巴结。例如,图像或训练图像可以是包括一个或多个淋巴结的胸腔的图像。在各种实施方案中,图像包括至少2个淋巴结。在各种实施方案中,图像包括1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个或21个淋巴结。在各种实施方案中,图像包括1-10个淋巴结、1-15个淋巴结、2-10个淋巴结、2-15个淋巴结、2-20个淋巴结、4-10个淋巴结、4-15个淋巴结、4-20个淋巴结、5-10个淋巴结、5-15个淋巴结、5-20个淋巴结、8-15个淋巴结、8-20个淋巴结、10-15个淋巴结、10-20个淋巴结、10-15个淋巴结、12-15个淋巴结、12-20个淋巴结或15-20个淋巴结。在各种实施方案中,从单个受试者或参考个体捕获多个图像。该多个图像可以捕获受试者或参考个体的不同淋巴结。在各种实施方案中,该多个图像包括至少2个淋巴结。在各种实施方案中,该多个图像包括2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个或21个淋巴结。在各种实施方案中,该多个图像包括8个至20个淋巴结。在各种实施方案中,该多个图像包括10个至19个淋巴结、12个至18个淋巴结或14个至16个淋巴结。

[0060] 图2A描绘了根据一个实施方案的识别淋巴结站的示例性淋巴结标测图。更具体地,图2A是根据美国胸腔协会映射方案修改的局部淋巴结分类的描绘。这里,淋巴结站被标注以描绘每个淋巴结的位置和分类。具体地,标注如下:(1) 锁骨上结节,(2R/2L) 上气管旁结节,(3A) 前血管结节,(3P) 椎前结节,(4R/4L) 下气管旁结节,(5) 主动脉下结节,(6) 主动

脉旁结节, (7) 脑下结节, (8) 食管旁结节, (9) 肺韧带结节, (10) 肺门结节, (11) 叶间结节, (12) 肺叶结节, (13) 节状结节, 和 (14) 亚段结节。

[0061] 淋巴结特征模块145对图像中的一个或多个淋巴结进行标注。在各种实施方案中, 淋巴结特征模块145通过限定淋巴结周围的区域来对一个或多个淋巴结进行标注。在各种实施方案中, 淋巴结周围的限定区域是球形区域。在各种实施方案中, 球形区域具有Xmm的半径。在各种实施方案中, X为5mm。在各种实施方案中, X为7.5mm。在一些实施方案中, X为0.5mm、1mm、1.5mm、2mm、2.5mm、3mm、3.5mm、4mm、4.5mm、5.5mm、6mm、6.5mm、7mm、8mm、8.5mm、9mm、9.5mm、10mm、10.5mm、11mm、11.5mm、12mm、12.5mm、13mm、13.5mm、14mm、14.5mm或15mm。在各种实施方案中, X在1mm与15mm之间。在各种实施方案中, X在2mm与12mm之间。在各种实施方案中, X在3mm与10mm之间。在各种实施方案中, X在4mm与8mm之间。在各种实施方案中, X在5mm与7.5mm之间。在各种实施方案中, X在4mm与6mm之间。在各种实施方案中, X在6mm与8mm之间。在各种实施方案中, X在7mm与8mm之间。在各种实施方案中, X在10mm与15mm之间。

[0062] 在各种实施方案中, 淋巴结特征模块145通过限定淋巴结周围的两个或更多个球形区域来对一个或多个淋巴结进行标注。例如, 淋巴结特征模块145通过限定淋巴结周围5mm的第一球形区域和淋巴结周围7.5mm的第二球形区域来对一个或多个淋巴结进行标注。图2B是根据一个实施方案的示例性淋巴结标注。具体地, 图2B描绘了围绕两个2R上气管旁结节的两个限定的球形区域。

[0063] 在各种实施方案中, 淋巴结特征模块145从淋巴结周围的每个限定区域提取特征。在各种实施方案中, 淋巴结特征模块145执行原始输入的变换以生成从中提取附加特征的变换的图像。在一个实施方案中, 淋巴结特征模块145执行原始图像的小波变换。在一个实施方案中, 淋巴结特征模块145执行原始图像的高斯-拉普拉斯 (LoG)。因此, 淋巴结特征模块145可以从变换的图像中的淋巴结周围的限定区域提取特征。在各种实施方案中, 淋巴结特征模块145从原始图像和变换的图像中的淋巴结周围的限定区域提取特征。

[0064] 在特定实施方案中, 淋巴结特征模块145执行原始图像的两个变换 (例如, 原始图像的小波变换和原始图像的高斯-拉普拉斯 (LoG) 变换两者)。因此, 淋巴结特征模块145可以从第一变换图像和第二变换图像两者中的淋巴结周围的限定区域提取特征。在各种实施方案中, 淋巴结特征模块145从原始图像、第一变换图像和第二变换图像中的淋巴结周围的限定区域提取特征。

[0065] 来自原始图像、第一变换图像和/或第二变换图像的示例性特征类别可以包括一阶特征、形状特征、灰度共生矩阵1 (GLCM) 特征、灰度游程长度矩阵 (GLRLM) 特征、灰度区域大小矩阵 (GLSZM) 特征和邻域灰度调差分矩阵 (NGTDM) 特征。一阶特征是指强度值的一阶统计的特征。GLCM特征从图像上给定偏移处的共生灰度值的直方图导出。GLRLM特征涉及灰度游程长度矩阵的统计。GLSZM特征涉及灰度区域大小矩阵的统计。NGTDM特征涉及相邻灰度调差分矩阵的统计。不同特征类别中的特征的示例包括相关特征、LMC2、MCC、熵差、联合熵、熵和、聚类显著性、平均差、联合平均、平均和、自相关、方差差、对比度、聚类倾向、平方和、LDMN、LDN、聚类阴影、LMC1、逆方差、LD、LDM、联合能量和最大概率。

[0066] 在各种实施方案中, 淋巴结特征模块145从图像中提取至少50个特征。在各种实施方案中, 淋巴结特征模块145从图像中提取至少100个特征、至少150个特征、至少200个特征、至少250个特征、至少300个特征、至少350个特征、至少400个特征、至少450个特征、至少

500个特征、至少550个特征、至少600个特征、至少650个特征、至少700个特征、至少750个特征、至少800个特征、至少850个特征、至少900个特征、至少950个特征、至少1000个特征、至少1100个特征、至少1200个特征、至少1300个特征、至少1400个特征、至少1500个特征、至少1600个特征、至少1700个特征、至少1800个特征、至少1900个特征、至少2000个特征、至少3000个特征、至少4000个特征、至少5000个特征、至少6000个特征、至少7000个特征、至少8000个特征、至少9000个特征或至少10,000个特征。在各种实施方案中,淋巴结特征模块145提取10个特征和100.00个特征之间的特征。在各种实施方案中,淋巴结特征模块145提取100个特征和1000个特征之间的特征。在各种实施方案中,淋巴结特征模块145提取300个特征和900个特征之间的特征。在各种实施方案中,淋巴结特征模块145提取500个特征和1000个特征之间的特征。在各种实施方案中,淋巴结特征模块145提取500个特征和900个特征之间的特征。在各种实施方案中,淋巴结特征模块145提取700个特征和1000个特征之间的特征。在各种实施方案中,淋巴结特征模块145提取700个特征和900个特征之间的特征。在各种实施方案中,淋巴结特征模块145提取800个特征和1000个特征之间的特征。在各种实施方案中,淋巴结特征模块145提取800个特征和900个特征之间的特征。

[0067] 在各种实施方案中,淋巴结特征模块145将所提取的特征提供给插补模块150以用于训练或部署插补模型。例如,如果所提取的特征来源于从参考个体捕获的训练图像,则淋巴结特征模块145将所提取的特征提供给插补模块150。在各种实施方案中,淋巴结特征模块145向风险训练模块155提供所提取的特征以训练风险模型。例如,如果所提取的特征来源于从参考个体捕获的训练图像,则淋巴结特征模块145将所提取的特征提供给风险训练模块155以训练风险模型。在各种实施方案中,淋巴结特征模块145将所提取的特征提供给风险部署模块160以部署风险模型。例如,如果所提取的特征来源于从受试者(例如,图1A中的受试者110)捕获的图像,则淋巴结特征模块145将所提取的特征提供给风险部署模块160。

[0068] III.A. 训练插补模型

[0069] 插补模块150训练并部署插补模型。一般来讲,该插补模型被构造成使得其分析从包括一个或多个淋巴结的图像提取的特征,并且针对该一个或多个淋巴结预测该淋巴结是癌性的概率。因此,基于从包括一个或多个淋巴结的图像提取的特征,插补模型针对淋巴结中的每个淋巴结插补概率,该插补的概率指示淋巴结是癌性的(例如,转移癌)的可能性。换句话说,通过在淋巴结水平处确定转移癌的风险,插补模型区分癌性淋巴结和非癌性淋巴结。

[0070] 在各种实施方案中,插补模型是以下中的任一者:回归模型(例如,线性回归、逻辑回归或多项式回归)、决策树、随机森林、支持向量机、朴素贝叶斯模型、k均值聚类或神经网络(例如,前馈网络、卷积神经网络(CNN)、深度神经网络(DNN)、自编码神经网络、生成对抗网络或递归网络(例如,长短期记忆网络(LSTM)、双向递归网络、深度双向递归网络)或它们的任何组合。在特定实施方案中,插补模型是随机森林分类器或最小绝对值收敛和选择算子(LASSO)分类器。在各种实施方案中,该插补模型是机器学习模型。在各种实施方案中,该插补模型是提升梯度机器学习模型(例如,xgboost)。

[0071] 插补模型可使用机器学习实现的方法来训练,该机器学习实现的方法诸如为以下的任一者:线性回归算法、逻辑回归算法、决策树算法、支持向量机分类、朴素贝叶斯分类、K

最近邻分类、随机森林算法、深度学习算法、梯度提升算法以及降维技术诸如流形学习、主成分分析、因子分析、自动编码器正则化和独立成分分析或它们的组合。在各种实施方案中,插补模型使用监督学习算法、无监督学习算法、半监督学习算法(例如,部分监督)、弱监督、迁移、多任务学习或它们的任何组合来训练。

[0072] 在各种实施方案中,插补模型具有一个或多个参数,诸如超参数或模型参数。通常在训练之前建立超参数。超参数的示例包括学习率、决策树的深度或叶子、深度神经网络中的隐藏层数量、k均值聚类中的聚类数量、回归模型中的惩罚以及与代价函数相关的正则化参数。通常在训练期间调节模型参数。模型参数的示例包括与神经网络的层中的节点相关联的权重、支持向量机中的支持向量以及回归模型中的系数。使用训练数据来训练(例如,调整)插补模型的模型参数,以提高插补模型的预测能力。

[0073] 插补模块150使用训练数据来训练插补模型。可从训练数据存储库170存储和/或检索训练数据。在各种实施方案中,训练数据包括从训练图像提取的特征,该训练图像包括从参考个体获得的一个或多个淋巴结。在各种实施方案中,参考个体包括健康个体(例如,对照个体)。在各种实施方案中,参考个体包括先前诊断患有癌症的个体。在各种实施方案中,参考个体包括先前诊断患有已经转移至淋巴结的癌症的个体。因此,从此类个体捕获的图像包括一个或多个癌性淋巴结。

[0074] 在各种实施方案中,用于训练插补模型的训练数据包括指示淋巴结是否是癌性的(在下文中也被称为“阳性”或“+”)或者淋巴结是否是非癌性的(在下文中也被称为“阴性”或“-”)的参考基础真值。在各种实施方案中,训练数据中的参考基础真值为二进制值,诸如“1”或“0”。例如,可以在训练数据中识别具有“1”值的癌性的阳性淋巴结,而可以在训练数据中识别具有“0”值的非癌性的阴性淋巴结。总之,插补模块150使用训练数据来训练插补模型,以使损失函数最小化,使得插补模型可以基于输入(例如,淋巴结的提取的特征)更好地预测结果(例如,淋巴结的淋巴结转移癌的概率)。在各种实施方案中,为最小绝对值收敛和选择算子(LASSO)回归、岭回归或弹性网络回归中的任一者构建损失函数。

[0075] 在各种实施方案中,可从公共可用数据库获得和/或得出训练数据。例如,训练数据可以从国家肺癌筛查试验(NLST)获得和/或导出,该国家肺癌筛查试验包括参考个体的结节分期(N期)癌信息。在一些实施方案中,训练数据可独立于公共可用数据库来获得和收集,例如通过从多个参考个体捕获一个或多个淋巴结的图像。例如,此类训练数据可以是定制数据集。定制数据集的训练数据可以在患者护理期间获取,并且因此可以具有淋巴结的对应外科取样信息。训练数据可以包括参考个体的N期癌症信息,例如,先前由临床医生基于参考个体的成像和/或活检结果诊断的N期信息。

[0076] 在各种实施方案中,指示淋巴结是否是癌性的参考基础真值是从临床分期数据导出的结节分期(N期)信息。例如,公开可用的NLST数据集包括参考个体的临床分期信息(例如,分期1a、1b、2a、2b、3a等),但不包括淋巴结特定信息。换句话讲,NLST数据集不区分个体淋巴结是癌性的还是非癌性的。在此类实施方案中,指示淋巴结是否是癌性的参考基础真值可以基于参考个体的癌症的N期信息和其他特性(例如,肿瘤大小、肿瘤位置、肿瘤形状和/或肿瘤密度)和/或淋巴结的N期信息和其他特性(例如,淋巴结大小、淋巴结密度和/或淋巴结形状)来反向计算。作为第一示例,N期为0或1的参考个体还没有遭遇癌症转移,并且因此该参考个体的淋巴结被指定为“阴性”或非癌性的参考基础真值。作为另一个示例,N期

为2或3的参考个体已遭遇癌症转移。因此,基于肿瘤和/或淋巴结的特性,淋巴结被指定为“阳性”或“阴性”的参考基础真值。

[0077] III.B. 部署插补模型

[0078] 插补模块150部署训练的插补模型以分析从包括一个或多个淋巴结的图像提取的特征,并且针对该一个或多个淋巴结预测该淋巴结是癌性的概率。在各种实施方案中,针对测试数据集来部署训练的插补模型。在各种实施方案中,用于训练插补模型的测试数据集和训练数据集可以从公共数据集导出。例如,公共数据集可以被拆分成两个子集以用于训练和测试插补模型。在一些实施方案中,公共数据集经历50:50的训练数据集:测试数据集拆分。在一些实施方案中,公共数据集经历60:40的训练数据集:测试数据集拆分。在一些实施方案中,公共数据集经历80:20的训练数据集:测试数据集拆分。

[0079] 训练的插补模型的部署针对淋巴结产生概率(例如,在0和1之间的连续概率),每个概率指示对应的淋巴结是否可能是癌性的。在各种实施方案中,插补模块150提供由插补模型预测的淋巴结的淋巴结转移风险中的一个或多个淋巴结转移风险,以用于训练风险模型。在各种实施方案中,插补模块150提供由插补模型预测的淋巴结的全部淋巴结转移风险,以用于训练风险模型。在各种实施方案中,插补模块150执行选择过程以识别由插补模型预测的淋巴结的淋巴结转移风险的子集。因此,随后将淋巴结转移风险的该子集用于训练风险模型。

[0080] 在一些实施方案中,插补模块150通过将淋巴结转移风险与阈值进行比较来选择淋巴结转移风险的子集。如果淋巴结的淋巴结转移风险高于阈值,则选择淋巴结的该淋巴结转移风险以包括在用于训练风险模型的子集中。如果淋巴结的淋巴结转移风险低于阈值,则将淋巴结的该淋巴结转移风险从该子集中排除,并且因此不用于训练风险模型。

[0081] 在各种实施方案中,阈值是固定值。在各种实施方案中,阈值是50%的转移癌风险。在各种实施方案中,阈值是30%、35%、40%、45%、55%、60%、65%、70%、75%、80%、85%、90%、95%或99%的转移癌风险。在一些实施方案中,阈值在30%和70%之间。在一些实施方案中,阈值在40%和60%之间。在一些实施方案中,阈值在45%和55%之间。在一些实施方案中,阈值在48%和52%之间。在一些实施方案中,阈值在49%和51%之间。在各种实施方案中,阈值是可变的。在各种实施方案中,阈值是患有癌症的中位受试者内概率。

[0082] 插补模块150将淋巴结转移风险的子集提供给风险训练模块155以训练风险模型。

[0083] III.C. 训练风险模型

[0084] 风险训练模块155使用训练数据来训练风险模型,该训练数据包括由插补模型预测的淋巴结的淋巴结转移风险中的至少一个或多个淋巴结转移风险。这里,该一个或多个淋巴结转移风险可以是如上所述的由插补模块150选择的淋巴结转移风险的子集。用于训练风险模型的淋巴结转移风险的子集代表与用于训练插补模型的训练数据集相比纯化的训练数据集。具体地,纯化的训练数据集包括更可能是癌性的淋巴结的子集,而从纯化的训练数据集中去除不太可能是癌性的淋巴结(但是否则可能代表假阳性或对于训练风险模型而言是混杂的)。因此,通过用该纯化的训练数据集训练风险模型,该风险模型预测那些患有转移性疾病的受试者中所有淋巴结的更大范围的概率。更具体地,风险模型更可能将可能患有癌症的淋巴结指定为具有更高的淋巴结转移风险,而将不太可能患有癌症的淋巴结指定为具有更低的淋巴结转移风险。

[0085] 在各种实施方案中,风险模型分析从包括一个或多个淋巴结的图像提取的特征,并且针对该一个或多个淋巴结中的每个淋巴结预测该淋巴结是癌性的连续概率。因此,基于从包括一个或多个淋巴结的图像提取的特征,风险模型确定淋巴结中的每个淋巴结的概率,所确定的概率指示淋巴是癌性的(例如,转移癌)的可能性。这里,风险模型在淋巴结水平确定转移癌的风险。在各种实施方案中,风险模型类似于插补模型,因为两个模型都分析从图像提取的特征并且在淋巴结水平预测转移癌的风险。然而,风险模型是使用插补模型的预测来训练的,例如,由插补模型预测的一个或多个淋巴结水平风险。总之,这使得风险模型能够以比由插补模型预测的插补LN水平风险更大范围的概率预测转移癌的LN水平风险。

[0086] 使用训练数据来训练该风险模型,该训练数据包括由插补模型预测的至少一个或多个淋巴结转移风险。这里,该一个或多个淋巴结转移风险可以作为用于训练风险模型的参考基础真值。在各种实施方案中,训练数据可另外包括从训练图像提取的特征,该训练图像包括从参考个体获得的一个或多个淋巴结。在一些实施方案中,用于训练风险模型的训练图像可以与先前用于训练插补模型的训练图像相同。在一些实施方案中,用于训练风险模型的训练图像不同于先前用于训练插补模型的训练图像。

[0087] 在各种实施方案中,训练数据包括单独的训练示例,使得使用每个训练示例来训练该风险模型。例如,每个训练示例可以包括特定训练图像的一个或多个淋巴结的提取特征以及参考基础真值(例如,特定训练图像的相同的一个或多个淋巴结的淋巴结转移风险)。因此,对于每个训练示例,风险模型被训练以最小化损失函数(例如,最小绝对值收敛和选择算子(LASSO)回归、岭回归或弹性网络回归中的任一者的损失函数)。

[0088] 在各种实施方案中,风险模型是以下中的任一者:回归模型(例如,线性回归、逻辑回归或多项式回归)、决策树、随机森林、支持向量机、朴素贝叶斯模型、k均值聚类或神经网络(例如,前馈网络、卷积神经网络(CNN)、深度神经网络(DNN)、自编码神经网络、生成对抗网络或递归网络(例如,长短期记忆网络(LSTM)、双向递归网络、深度双向递归网络)或它们的任何组合。在特定实施方案中,风险模型是随机森林分类器或最小绝对值收敛和选择算子(LASSO)分类器。在一些实施方案中,风险模型是机器学习模型。

[0089] 风险模型可使用机器学习实现的方法来训练,该机器学习实现的方法诸如为以下的任一者:线性回归算法、逻辑回归算法、决策树算法、支持向量机分类、朴素贝叶斯分类、K最近邻分类、随机森林算法、深度学习算法、梯度提升算法以及降维技术诸如流形学习、主成分分析、因子分析、自动编码器正则化和独立成分分析或它们的组合。在各种实施方案中,风险模型使用监督学习算法、无监督学习算法、半监督学习算法(例如,部分监督)、弱监督、迁移、多任务学习或它们的任何组合来训练。

[0090] 在各种实施方案中,风险模型具有一个或多个参数,诸如超参数或模型参数。通常在训练之前建立超参数。超参数的示例包括学习率、决策树的深度或叶子、神经网络中的隐藏层数量、k均值聚类中的聚类数量、回归模型中的惩罚以及与代价函数相关的正则化参数。通常在训练期间调节模型参数。模型参数的示例包括与神经网络的层中的节点相关联的权重、支持向量机中的支持向量以及回归模型中的系数。使用训练数据来训练(例如,调整)风险模型的模型参数以提高风险模型的预测能力。

[0091] III.C. 部署风险模型

[0092] 风险部署模块160部署风险模型以确定受试者(例如,上文在图1A中描述的受试者110)的转移癌的受试者水平风险。这里,风险部署模块160提供包括从受试者捕获的一个或多个淋巴结的图像的提取特征作为训练的风险模型的输入。训练的风险模型分析所提取的特征并且输出淋巴结的概率,该概率指示个体淋巴结是否可能是癌性的。

[0093] 在各种实施方案中,在部署风险模型之后,风险部署模块160从由风险模型预测的转移癌的个体淋巴结水平风险转变为受试者的转移癌的受试者水平风险。在各种实施方案中,风险部署模块160组合转移癌的个体淋巴结水平风险以确定转移癌的受试者水平风险。在各种实施方案中,风险部署模块160跨转移癌的淋巴结水平风险执行统计操作。统计操作可识别转移癌的个体淋巴结水平风险中的最大值、最小值、平均值、中值或众数值。在特定实施方案中,统计操作是转移癌的个体淋巴结水平风险中的最大值。因此,将具有转移性疾病的最高淋巴结风险水平的淋巴结指定为受试者的受试者的转移性疾病的受试者水平风险。

[0094] 癌症分期模块165基于受试者的转移性疾病的受试者水平风险确定受试者的癌症分期。在各种实施方案中,癌症分期模块165通过将受试者的转移性疾病的受试者水平风险与一个或多个阈值风险值进行比较来确定受试者的癌症分期。这些阈值风险值可以充当用于确定受试者中的癌症是否可能处于0期、1期、2期或3期中的任一者中的截止值。

[0095] 在一些实施方案中,如果受试者水平风险大于30%,则将受试者分类为具有至少3期癌症。在一些实施方案中,如果受试者水平风险大于40%、大于50%、大于60%、大于70%、大于80%或大于90%风险,则将受试者分类为具有至少3期癌症。

[0096] 在各种实施方案中,阈值风险值定义对应于一个或多个癌症分期的风险范围。因此,阈值风险值设置风险范围的上限和下限。在一些实施方案中,如果受试者水平风险在30%和100%之间,则受试者被分类为具有结节3期癌症。在一些实施方案中,如果受试者水平风险在40%和100%之间,则受试者被分类为具有结节3期癌症。在一些实施方案中,如果受试者水平风险在50%和100%之间,则受试者被分类为具有结节3期癌症。

[0097] 在一些实施方案中,如果受试者水平风险在10%和40%之间,则受试者被分类为具有结节2期癌症。在一些实施方案中,如果受试者水平风险在15%和40%之间,则受试者被分类为具有结节2期癌症。在一些实施方案中,如果受试者水平风险在20%和40%之间,则受试者被分类为具有结节2期癌症。在一些实施方案中,如果受试者水平风险在30%和40%之间,则受试者被分类为具有结节2期癌症。在一些实施方案中,如果受试者水平风险在15%和30%之间,则受试者被分类为具有结节2期癌症。在一些实施方案中,如果受试者水平风险在20%和30%之间,则受试者被分类为具有结节2期癌症。

[0098] 在一些实施方案中,如果受试者水平风险在5%和20%之间,则受试者被分类为具有淋巴结0期或淋巴结1期癌症。在一些实施方案中,如果受试者水平风险在10%和20%之间,则受试者被分类为具有淋巴结0期或淋巴结1期癌症。在一些实施方案中,如果受试者水平风险在5%和15%之间,则受试者被分类为具有结节0期或结节1期癌症。在一些实施方案中,如果受试者水平风险在8%和12%之间,则受试者被分类为具有结节0期或结节1期癌症。

[0099] IV. 用于癌症分期的示例性方法

[0100] 图3A描绘了根据一个实施方案的插补模型和风险模型的示例性实施方式。这里,

参考个体305A和305B可以被识别为具有特定的N期。例如,参考个体305A被识别为N期阴性,从而指示参考个体305A不表现出淋巴结癌累及。参考个体305B被识别为N期阳性,从而指示参考个体305B表现出淋巴结癌累及。然而,受试者水平N期不区分特定淋巴结以及个体淋巴结是癌性的还是非癌性的。

[0101] 插补模型310插补每个淋巴结的转移风险(例如,淋巴结转移风险),从而使得能够区分具有转移癌的淋巴结和非癌性的淋巴结。在各种实施方案中,插补模型310预测每个淋巴结的概率,该概率表示淋巴结是癌性的概率。

[0102] 如图3A所示,插补模型310的输出可以识别个体305C(对应于参考个体305A)没有表现出淋巴结累及。换句话说,个体305C的淋巴结中没有一个可能是癌性的。这确认了N期阴性识别。另外,插补模型310的输出可识别个体305D(对应于参考个体305B)表现出一个或多个淋巴结的淋巴结累及。例如,插补模型310可以插补概率,使得个体305D中恰好一个淋巴结被识别为癌性的,而个体305D中的其他淋巴结是非癌性的。

[0103] 在各种实施方案中,使用由插补模型310预测的淋巴结的淋巴结转移风险中的一个或多个淋巴结转移风险来训练风险模型320。在各种实施方案中,由插补模型310预测的淋巴结的所有淋巴结转移风险用于训练风险模型320。在各种实施方案中,对应于淋巴结的子集的淋巴结转移风险的子集用于训练风险模型320,从而使得风险模型能够针对具有转移性疾病的那些受试者中的所有淋巴结预测更大范围的概率。

[0104] 一旦训练,就部署风险模型320以预测一个或多个受试者315的转移癌的受试者水平风险。因此,风险模型320被部署为从转移性疾病的淋巴结水平风险转变为转移性疾病的受试者水平风险。如图3A所示,部署风险模型320以分析从个体315捕获的图像中提取的特征。风险模型320预测转移癌到达个体315的受试者水平风险。例如,风险模型320预测个体315A具有20%的转移癌风险,个体315B具有30%的转移癌风险,个体315C具有50%的转移癌风险,并且个体315D具有60%的转移癌风险。因此,个体315A和315B可被分类为早期癌症(例如,0期或1期),而个体315C和315D可被分类为晚期癌症(例如,2期、3期或4期)。

[0105] 个体315可以根据他们的转移性疾病的受试者水平风险接受不同的治疗。例如,具有低淋巴结转移风险的个体315A和315B可以直接进行立即治疗,诸如手术切除或组合支气管镜诊断和治疗。作为另一示例,处于具有淋巴结转移的较高风险下的个体315C和315D可以接收附加的测试,诸如PET扫描和淋巴结活检。

[0106] 图3B根据一个实施方案更详细地描绘了风险模型320的实施方式。这里,风险模型320分析从受试者350捕获的图像中提取的特征,并且输出受试者的一个或多个淋巴结中的每个淋巴结的概率。如图3B所示,风险模型320输出三个淋巴结的概率。在其他实施方案中,风险模型320可以输出受试者350的附加的或更少的淋巴结概率。

[0107] 如图3B所示,风险模型320预测第一淋巴结具有20%的癌性概率,并且相反地,具有80%的非癌性概率。另外,风险模型320预测第二淋巴结具有30%的癌性概率,并且相反地,具有70%的非癌性概率。另外,风险模型320预测第三淋巴结具有60%的癌性概率,并且相反地,具有40%的非癌性概率。分析个体淋巴结概率以识别转移癌的受试者水平风险。在该示例中,转移癌的最大淋巴结概率被看作是转移癌的受试者水平概率。具有60%癌性概率的第三淋巴结是淋巴结中最高的概率。因此,受试者350的转移癌的受试者水平概率是60%。

[0108] V. 癌症

[0109] 本文描述的方法涉及通过确定受试者的癌症的受试者水平风险的癌症分期。在各种实施方案中,受试者的癌症可包括以下中的一者或多者:淋巴瘤、B细胞淋巴瘤、T细胞淋巴瘤、蕈样肉芽肿病、霍奇金氏病、骨髓性白血病、膀胱癌、脑癌、神经系统癌、头颈癌(head and neck cancer)、头颈鳞状细胞癌、肾癌(kidney cancer)、肺癌(lung cancer)、成神经细胞瘤/成胶质细胞瘤、卵巢癌、胰腺癌、前列腺癌、皮肤癌、肝癌、黑色素瘤、口腔、喉部、咽部和肺部的鳞状细胞癌、结肠癌、宫颈癌(cervical cancer)、宫颈癌(cervical carcinoma)、乳腺癌、上皮癌、胃肠道癌、肾癌(renal cancer)、泌尿生殖系统癌、肺癌(pulmonary cancer)、食管癌、胃癌、甲状腺癌、头颈癌(head and neck carcinoma)、大肠癌、造血系统癌、睾丸癌、结肠和/或直肠癌、子宫癌或前列腺癌。在各种实施方案中,该癌症可以是肺细支气管肺泡癌(BAC)、膀胱癌、女性生殖道恶性肿瘤(例如,子宫浆液性癌、子宫内膜癌、外阴鳞状细胞癌和子宫肉瘤)、卵巢表面上皮癌(例如,卵巢的透明细胞癌、上皮卵巢癌、输卵管癌和原发性腹膜癌)、乳腺癌、非小细胞肺癌(NSCLC)、男性生殖道恶性肿瘤(例如,睾丸癌)、腹膜后或腹膜癌、胃食管腺癌、食管胃结合部腺癌、干细胞癌、食管和食管胃接合癌、宫颈癌、胆管癌、胰腺癌、肝外胆管腺癌、小肠恶性肿瘤、胃腺癌、未知原发性癌(CUP)、结肠直肠腺癌、食管癌、前列腺腺癌、肾癌、头颈鳞状细胞癌、胸腺癌、非黑色素瘤皮肤癌、甲状腺癌(例如乳头状癌)、头颈癌、肛门癌、非黑色素瘤非神经细胞癌(非EOC)、葡萄膜黑色素瘤、恶性胸膜间皮瘤、小细胞肺癌(SCLC)、中枢神经系统癌症和神经内分泌肿瘤中的任一种。在某些实施方案中,癌症是乳腺癌、非小细胞肺癌、膀胱癌、肾癌、结肠癌和黑色素瘤。在一些实施方案中,受试者的癌症可为转移癌,包括膀胱癌、乳腺癌、结肠癌、肾癌、肺癌、黑色素瘤、卵巢癌、胰腺癌、前列腺癌、直肠癌、胃癌、甲状腺癌或子宫癌中的任一者。在特定实施方案中,癌症是肺癌。在特定实施方案中,癌症是一种类型的肺癌,包括小细胞肺癌、非小细胞肺癌、腺癌、鳞状细胞癌、大细胞癌、小细胞癌、复合性小细胞癌、肺肉瘤、肺淋巴瘤、支气管类癌中的任一者。

[0110] VI. 指导干预

[0111] 本文所述的实施方案涉及通过确定受试者中转移癌的受试者水平风险来确定癌症分期。在各种实施方案中,基于针对受试者预测的转移癌的癌症分期和/或受试者水平风险向受试者提供干预。此类干预(下文称为指导干预)可以是以下中的任一种:诊断施加、治疗剂施加、受试者的后续测试、后续活检(例如,癌症活检或淋巴结活检、淋巴结的支气管镜取样)、后续图像扫描(例如,PET扫描、MRI扫描、超声成像或X射线成像)、肿瘤切除、支气管镜诊断、治疗药物的选择和/或施加、药物组合物的选择/施加、临床试验中的招募、临床试验中的非招募、或者它们的任何组合。

[0112] 在各种实施方案中,针对受试者预测的转移癌的受试者水平风险或预测的癌症分期指示受试者处于高转移癌风险中。在各种实施方案中,处于高转移癌风险的受试者具有高于阈值得分的转移癌的预测受试者水平风险。在各种实施方案中,阈值得分为高于20%的转移癌风险。在各种实施方案中,阈值得分为高于25%的转移癌风险。在各种实施方案中,阈值得分为高于30%、高于35%、高于40%、高于45%、高于50%、高于55%、高于60%、高于65%、高于70%、高于75%、高于80%、高于85%、高于90%、高于95%或高于99%的转移癌风险。在各种实施方案中,阈值得分在20%和80%之间。在一些实施方案中,阈值在30%

和70%之间。在一些实施方案中,阈值在40%和60%之间。在一些实施方案中,阈值在45%和55%之间。在一些实施方案中,阈值在20%和60%之间。在一些实施方案中,阈值在30%和55%之间。在一些实施方案中,阈值在35%和50%之间。在一些实施方案中,阈值在40%和45%之间。

[0113] 在各种实施方案中,如果受试者的预测癌症分期为至少2期(例如,2期或3期),则受试者处于高转移癌风险中。在此类实施方案中,指导干预涉受试者的后续测试、受试者的后续活检或受试者的后续图像扫描中的一者。这允许临床提供者优先考虑处于最高风险的受试者的后续测试、活检或图像扫描。

[0114] 在各种实施方案中,如果预测受试者处于高转移癌风险中,则对于受试者可以排除不适合晚期转移癌的规程。被排除的此类规程的示例包括肿瘤切除、肺切除、放射疗法(例如,外照射放疗)、化学疗法(例如,局部或全身化学疗法)、免疫调节疗法、或甚至全身麻醉。

[0115] 在特定实施方案中,如果预测受试者处于高转移癌风险中,则受试者经历后续图像扫描以确认该预测。例如,如果预测受试者处于高转移癌风险中,则受试者进一步经历PET扫描。在此类实施方案中,如果PET扫描返回阴性结果(例如,没有淋巴结累及),则受试者可以经历进一步测试。例如,受试者可进一步经历后续活检以明确地对淋巴结进行取样。在一些实施方案中,后续活检是淋巴结的支气管镜取样。

[0116] 在特定实施方案中,如果预测受试者处于高转移癌风险中,则受试者直接经历后续活检以确认该预测。例如,受试者不需要经历后续图像扫描来确认该预测,而是直接经历后续活检。例如,如果预测受试者处于高转移癌风险中,则受试者进一步经历淋巴结的支气管镜取样。

[0117] 如本文所述,转移癌的受试者水平风险可由跨淋巴结的个体淋巴结水平风险确定。因此,对于基于受试者风险水平预测具有高转移癌风险的受试者,可识别导致高受试者水平风险的特定淋巴结。因此,那些特定的淋巴结可以作为后续测试或治疗的目标。在特定实施方案中,那些特定淋巴结可经历后续图像扫描(例如,PET扫描)以确认该预测。在特定实施方案中,那些特定淋巴结可经历后续活检以明确地对特定淋巴结进行取样。因此,后续活检可以确认在那些特定淋巴结中存在或不存在转移癌。

[0118] 在各种实施方案中,如果预测受试者处于高转移癌风险中,则可将受试者招募进临床试验中。例如,此类临床试验可以是测试针对转移癌的治疗功效,并且因此其可以在其合格性标准中指定招募者将表现出转移癌。在各种实施方案中,如果预测受试者处于高转移癌风险中,则将停止将受试者招募进临床试验中。例如,此类临床试验可以是测试针对非转移形式的癌症的治疗的治疗功效。

[0119] 在各种实施方案中,针对受试者预测的转移癌的受试者水平风险指示受试者处于低转移癌风险下。在多个实施方案中,处于低转移癌风险的受试者具有低于阈值得分的预测的受试者水平风险。在各种实施方案中,阈值得分低于20%的转移癌风险。在各种实施方案中,阈值得分低于25%的转移癌风险。在各种实施方案中,阈值得分为低于30%、低于35%、低于40%、低于45%、低于50%、低于55%、低于60%、低于65%、低于70%、低于75%、低于80%、低于85%、低于90%、低于95%或低于99%的转移癌风险。在一些实施方案中,阈值在5%和40%之间。在一些实施方案中,阈值在10%和30%之间。在一些实施方案中,阈值

在15%和25%之间。在一些实施方案中,阈值在18%和22%之间。在各种实施方案中,如果受试者的预测癌症分期低于2期(例如,0期或1期),则受试者处于低转移癌风险下。在此类实施方案中,指导干预涉及肿瘤切除、支气管镜诊断、治疗药物的选择和/或施加、和/或药物组合物的选择/施加中的一者。这使得能够快速识别和治疗尚未发生转移的受试者。

[0120] 在各种实施方案中,可基于转移癌的受试者水平风险来选择治疗剂和/或将治疗剂施用给受试者,所选择的治疗剂可能表现出抗癌疗效。示例性治疗剂包括化学疗法、放射疗法、抗原特异性单克隆抗体或免疫疗法。在各种实施方案中,向处于高转移癌风险的受试者提供治疗剂。在各种实施方案中,处于高转移癌风险的受试者具有高于阈值得分的转移癌的预测受试者水平风险。在各种实施方案中,阈值得分为高于20%的转移癌风险。在各种实施方案中,阈值得分为高于25%的转移癌风险。在各种实施方案中,阈值得分为高于30%、高于35%、高于40%、高于45%、高于50%、高于55%、高于60%、高于65%、高于70%、高于75%、高于80%、高于85%、高于90%、高于95%或高于99%的转移癌风险。在各种实施方案中,如果受试者的预测癌症分期为至少2期(例如,2期、3期或4期),则受试者处于高转移癌风险中。

[0121] 在各种实施方案中,向处于低转移癌风险的受试者提供治疗剂。在多个实施方案中,处于低转移癌风险的受试者具有低于阈值得分的预测的受试者水平风险。在各种实施方案中,阈值得分低于20%的转移癌风险。在各种实施方案中,阈值得分低于25%的转移癌风险。在各种实施方案中,阈值得分为低于30%、低于35%、低于40%、低于45%、低于50%、低于55%、低于60%、低于65%、低于70%、低于75%、低于80%、低于85%、低于90%、低于95%或低于99%的转移癌风险。在各种实施方案中,如果受试者的预测癌症分期低于2期(例如,0期或1期),则受试者处于低转移癌风险下。

[0122] 在各种实施方案中,如果预测受试者处于低转移癌风险下,则可以将受试者引导至早期癌症治疗。早期癌症治疗的示例包括肿瘤切除、肺切除、化学疗法(例如,局部或全身化学疗法)、局部递送治疗(例如,经支气管性针滴注或基于导管的消融)或放射。

[0123] 在特定实施方案中,如果预测受试者处于低转移癌风险中,则受试者经历后续图像扫描以确认该预测。例如,如果预测受试者处于低转移癌风险中,则受试者进一步经历PET扫描。在此类实施方案中,如果PET扫描返回阴性结果(例如,没有淋巴结累及),则PET扫描结果与预测对准。然而,如果PET扫描返回阳性结果(例如,淋巴结累及),则受试者可以经历进一步测试。例如,受试者可进一步经历后续活检以明确地对淋巴结进行取样。在一些实施方案中,后续活检是淋巴结的支气管镜取样。

[0124] 在各种实施方案中,治疗剂是生物制剂,例如细胞因子、抗体、可溶性细胞因子受体、反义寡核苷酸、siRNA等。此类生物制剂涵盖生物制剂的突变蛋白和衍生物,所述衍生物可包括例如融合蛋白、聚乙二醇化衍生物、胆固醇缀合的衍生物及本领域已知的类似衍生物。还包括细胞因子和细胞因子受体的拮抗剂,例如诱捕网(trap)和单克隆拮抗剂。还包括与本文所阐述的活性剂具有生物类似性或生物等效性的药物。

[0125] 肺癌的治疗剂可包括化疗药物,诸如多西他赛、顺铂、卡铂、吉西他滨、白蛋白结合型紫杉醇、紫杉醇、培美曲塞、吉非替尼、厄洛替尼、布加替尼(Alunbrig®)、卡马替尼(Tabrecta®)、塞尔帕替尼(Retevmo®)、恩曲替尼(Rozlytrek®)、劳拉替尼(Lorbrena®)、拉罗替

尼 (Vitrakvi[®])、达可替尼 (Vizimpro[®]) 和长春瑞滨。肺癌的治疗剂可包括抗体疗法, 诸如度伐利尤单抗 (Imfinzi[®])、纳武利尤单抗 (Opdivo[®])、帕博利珠单抗 (Keytruda[®])、阿替利珠单抗 (Tecentriq[®]) 和雷莫芦单抗。

[0126] 在各种实施方案中, 可基于受试者转移癌风险水平来选择药物组合物和/或将药物组合物施用给受试者, 所选择的治疗剂可能表现出抗癌疗效。施用给个体的药物组合物包括活性剂, 诸如上述治疗剂。活性成分以治疗有效量 (即, 在施用时足以治疗由此介导的疾病或医学病症的量) 存在。组合物还可包含增强递送和疗效 (例如, 增强活性成分的递送和稳定性) 的各种其它药剂。因此, 例如, 根据所需的剂型, 组合物还可包含药学上可接受的无毒载体或稀释剂, 它们被定义为常用于配制供动物或人类施用的药物组合物的溶媒。选择稀释剂以便不影响该组合的生物活性。此类稀释剂的示例为蒸馏水、缓冲水、生理盐水、PBS、林格氏溶液、葡萄糖溶液和汉克斯溶液。此外, 药物组合物或制剂可包含其它载体、佐剂或无毒、非治疗性、无致免疫性稳定剂、赋形剂等。组合物还可包含用于接近生理条件的附加物质, 诸如 pH 调节和缓冲剂、毒性调节剂、润湿剂和去垢剂。组合物还可包含多种稳定剂中的任一种, 诸如抗氧化剂。

[0127] 可按多种不同方式施用本文所述的药物组合物。示例包括经由口、鼻内、模块内、直肠、局部、腹膜内、静脉内、肌内、皮下、真皮下、经皮、鞘内、颅内方法施用包含药学上可接受的载体的组合物。在各种实施方案中, 药物组合物可经由基于导管的部署 (例如, 使用基于柔性或刚性机器人的支气管镜) 施用。

[0128] 此类药物组合物可以为了治疗 (例如, 在诊断患有肺癌的患者之后) 目的而施用。本发明的上下文中使用的疾病或病症的防止、预防或防范是指施用组合物以防止肺癌或肺癌的症状中的一些或全部症状的发生、发作、进展或复发或降低肺癌发作的可能性。肺癌的治疗 (treating)、处理 (treatment) 或疗法 (therapy) 应意指通过施用根据本发明的治疗来减缓、停止或逆转癌症的进展。在优选的实施方案中, 治疗肺癌意指逆转癌症的进展, 在理想情况下达到消除癌症自身的程度。

[0129] VII. 计算机实现

[0130] 在一些实施方案中, 本发明的方法 (包括通过确定受试者的转移癌的受试者水平风险进行癌症分期的方法) 在一个或多个计算机上执行。

[0131] 例如, 可以以硬件或软件或两者的组合来实现插补模型和/或风险模型以及数据库存储装置的构建和部署。在本发明的一个实施方案中, 提供了一种机器可读存储介质, 该介质包括用机器可读数据编码的数据存储材料, 当使用由使用所述数据的指令编程的机器时, 该机器可读数据能够显示任何数据集以及本文所述模型的执行和结果。本发明可在可编程计算机上执行的计算机程序中实现, 所述可编程计算机包括处理器、数据存储系统 (包括易失性和非易失性存储器和/或存储元件)、图形适配器、指向装置、网络适配器、至少一个输入装置和至少一个输出装置。显示器联接到图形适配器。将程序代码应用于输入数据以执行上述功能并生成输出信息。以已知方式将输出信息应用于一个或多个输出装置。计算机可以是例如常规设计的个人计算机、微型计算机或工作站。

[0132] 每个程序可以高级过程语言或面向对象的编程语言来实现以与计算机系统通信。然而, 这些程序可根据需要以汇编语言或机器语言来实现。在任何情况下, 语言可为编译或

解释语言。每个此类计算机程序优选地被存储在由通用或专用可编程计算机可读的存储介质或装置(例如,ROM或磁盘)上,以便在计算机读取存储介质或装置时配置和操作计算机以执行本文所述的过程。该系统还可被视为被实现为配置有计算机程序的计算机可读存储介质,其中如此配置的存储介质使得计算机以特定且预定义的方式操作以执行本文所述的功能。

[0133] 可在多种介质中提供签名模式及其数据库以促进其使用。“介质”是指包含本发明的签名模式信息的制品。可将本发明的数据库记录在计算机可读介质(例如,可由计算机直接读取和访问的任何介质)上。此类介质包括但不限于:磁存储介质,诸如软盘、硬盘存储介质和磁带;光学存储介质,诸如CD-ROM;电存储介质,诸如RAM和ROM;以及这些类别的混合,诸如磁/光学存储介质。本领域技术人员可容易地理解,当前已知的计算机可读介质中的任一者可如何用于创建包含本发明数据库信息的记录的制品。“记录的”是指用于使用本领域已知的任何此类方法在计算机可读介质上存储信息的过程。可基于用于访问所存储的信息的装置来选择任何方便的数据存储结构。多种数据处理器程序和格式可用于存储,例如文字处理文本文件、数据库格式等。

[0134] 在一些实施方案中,本发明的方法(包括通过确定受试者的转移癌的受试者水平风险进行癌症分期的方法)在分布式计算系统环境中(例如,在云计算环境中)的一个或多个计算机上执行。在本说明书中,“云计算”被定义为用于实现对可配置计算资源的共享集合的按需网络访问的模型。可采用云计算来提供对可配置计算资源的共享集合的按需访问。可配置计算资源的共享集合可经由虚拟化来快速提供,并且以低管理工作量或服务提供商交互来发布,然后相应地进行缩放。云计算模型可由诸如例如按需自助服务、广泛网络接入、资源池化、快速弹性、可计量服务等各种特性组成。云计算模型还可暴露各种服务模式,诸如例如软件即服务(“SaaS”)、平台即服务(“PaaS”)和基础设施即服务(“IaaS”)。还可使用诸如私有云、社区云、公共云、混合云等的不同部署模型来部署云计算模型。在本说明书和权利要求书中,“云计算环境”是其中采用云计算的环境。

[0135] 图4示出了用于实现图1A、图1B、图2A、图2B、图3A和图3B所示的实体的示例性计算机。计算机400包括联接到芯片组404的至少一个处理器402。芯片组404包括存储器控制器集线器420和输入/输出(I/O)控制器集线器422。存储器406和图形适配器412联接到存储器控制器集线器420,并且显示器418联接到图形适配器412。存储装置408、输入装置414和网络适配器416联接到I/O控制器集线器422。计算机400的其它实施方案具有不同架构。

[0136] 存储装置408是非暂态计算机可读存储介质,诸如硬盘驱动器、光盘只读存储器(CD-ROM)、DVD或固态存储器装置。存储器406保存由处理器402使用的指令和数据。输入接口414是触摸屏界面、鼠标、跟踪球或其他类型的指向装置、键盘或它们的某种组合,并且用于将数据输入到计算机400中。在一些实施方案中,计算机400可被配置成经由来自用户的手势从输入接口414接收输入(例如,命令)。图形适配器412在显示器418上显示图像和其它信息。网络适配器416将计算机400联接到一个或多个计算机网络。

[0137] 计算机400适于执行计算机程序模块以便提供本文所述的功能。如本文所用,术语“模块”是指用于提供指定功能的计算机程序逻辑。因此,模块可在硬件、固件和/或软件中实现。在一个实施方案中,程序模块被存储在存储装置408上,被加载到存储器406中,并且由处理器402执行。

[0138] 图1A或图1B的实体所使用的计算机400的类型可根据实体所需的实施方案和处理能力而改变。例如,癌症分期系统130可在单个计算机400中运行或在彼此通过网络(诸如服务器农场)通信的多个计算机400中运行。计算机400可缺少上述部件中的一些部件,诸如图形适配器412和显示器418。

[0139] VIII. 系统

[0140] 本文还公开了通过确定受试者的转移癌的受试者水平风险来进行癌症分期的系统。在各种实施方案中,此类系统可包括至少上文在图1A中所描述的癌症分期系统130。在各种实施方案中,癌症分期系统130被体现为计算机系统,诸如具有图4中所述的示例性计算机400的计算机系统。

[0141] 在各种实施方案中,该系统包括成像装置,诸如上文在图1A中所描述的成像装置120。在各种实施方案中,该系统包括癌症分期系统130(例如,计算机系统)和成像装置两者。在此类实施方案中,癌症分期系统130可与成像装置120通信地耦接以接收从受试者捕获的图像(例如,CT扫描)。该计算机系统在计算机上实现插补模型和风险模型中的一者或两者,以分析图像并确定受试者的结节病的受试者水平风险。

[0142] 实施例

[0143] 下面是用于执行本发明的特定实施方案的实施例。这些实施例仅提供用于说明目的,并不旨在以任何方式限制本发明的范围。已努力确保关于所使用的数字(例如量、温度等)的准确性,但应容许一些实验误差和偏差。

[0144] 实施例1:研究人群和特征提取

[0145] 使用来自国家肺癌筛查试验(NLST)的CT臂的数据来执行该项研究。NLST的该亚群组包括来自658名研究参与者的数据,301名患有良性结节,并且357名患有肺癌。检查658个研究参与者的每个CT扫描,并且标注多达21个淋巴结站(每个人16+/-6个)(标注9969个LN;对照组为4631个LN,并且肺癌组为5318个LN)。从该初始样品中,288个患有肺癌的人具有可用于分析的分期数据。为了建立平衡的初始群组,识别287个对照受试者。淋巴结站描绘在图2A中并详述于表1中。

[0146] 表1:按站点的淋巴结标注

	淋巴站	#LN (每个受试者平均)	
		对照组	癌症组
[0147]	1L	34	22
	1R	29	13
	2L	191	186
	2R	732	879
	3A	203	263
	3P	133	139
	4L	430	459
	4R	837	987
	5	386	457
	6	189	132
	7	477	541
	8L	212	230
	8R	140	215
	9L	29	98
	9R	8	14
	10L	181	188
	10R	92	60
	11L	74	106
	11R	57	59
	12L	5	4
	12R	5	8

[0148] 对于每个LN站,限定具有5mm和7.5mm半径的球形区域。球心由标注标签定义。图2B中示出了两个LN标注的示例。

[0149] 这些较小(5mm)和较大(7.5mm)的球形ROI(感兴趣区域)用于探索用于模型构建的图像的范围。从这些球形ROI中,提取3个单独的放射组学组,每个放射组学组具有6个特征类别。每个组在它们所利用的度量的数量上是不同的,并且其是基于CT的不同变换。组1使用原始图像作为特征提取的输入。组2和组3分别采用输入图像的小波变换和高斯-拉普拉斯(LoG)变换。特征类别包括一阶特征、形状特征、灰度共生矩阵1(GLCM)特征、灰度游程长度矩阵(GLRLM)特征、灰度区域大小矩阵(GLSZM)特征和邻域灰度调差分矩阵(NGTDM)特征。特征的细目如表3所示。

[0150] 表3:来自CT扫描图像的特征

[0151]

特征的数量						
特征类别						
图像类别	一阶	形状*	GLCM	GLRLM	GLSZM	NGTDM
原先的	22	14	24	16	16	5
小波	144	0	192	128	128	40
LoG	18	0	24	16	16	5

[0152] *当每个站用固定球形区域分析时不包括在LN分析中的形状特征

[0153] 图5A至图5R描绘了GLCM特征类别的不同组上的不同淋巴结放射组学特征的示例性相关性热图。另外,图5A至图5R各自示出了特定特征。图5A至图5R中所示的此类特征的示例包括相关特征、LMC2、MCC、熵差、联合熵、熵和、聚类显著性、平均差、联合平均、平均和、自

相关、方差差、对比度、聚类倾向、平方和、LDMN、LDN、聚类阴影、LMC1、逆方差、LD、LDM、联合能量和最大概率。图5C至图5R示出了关于不同滤波器的特征。具体地，图5C至图5R提供了“XYZ”的名称，其中“X”、“Y”和“Z”各自指应用于图像的X、Y或Z方向的滤波器。“H”是指高通滤波器，而“L”是指低通滤波器。因此，名称“HHH”指的是从已经在X、Y和Z方向中的每个方向上经历高通滤波的3D图像导出的特征。名称“HHL”是指从已经在X和Y方向上经历高通滤波并且已经在Z方向上经历低通滤波的3D图像导出的特征。

[0154] 实施例2：用于癌症分期的插补模型

[0155] 建模开始于对每个淋巴结的癌累及的预测。为此，将（在NLST中报告的）癌症分期插补到每个标注的LN站（在NLST中未报告单独的LN数据，因此LN累及从癌症分期“收回”）。然后将数据以50:50拆分用于训练：测试，这些数据的分布在表2中提供。

[0156] 表2：实施例50:50数据拆分

N 期	训练		测试	
	对照	癌症	对照	癌症
0	2337	2179	2263	1713
1	0	127	0	73
2	0	125	0	342
3	0	40	0	42

[0158] 该模型的结果是二元的（使用NLST数据集中的Nstage协变量，Nodal转移是/否。“否”是N=0或N=1，并且“是”是N=2或N=3）。归一化度量（z得分）并训练和测试随机森林和LASSO分类器。对于每个LN，生成其是癌性的连续概率。例如，可以存在具有20%癌性概率的LN（通过减法意味着其被分配80%没有癌症的概率）。研究参与者内和跨研究参与者的所有LN被认为是独立的数据。初始模型的性能如图6所示。具体地，该模型实现了曲线下面积（AUC）=65.3%，准确度（ACC）=96.6%，灵敏度=61.2%，以及特异性=60.8%。

[0159] 对NLST数据的限制是LN特定信息不可用，并且平均16个LN被标注并在每个人中进行基于放射学的评估。因此，该模型假定当在临床医学中时患有转移癌的患者中的所有LN都是“+”，通常只有1个LN具有癌细胞（癌症不同时扩散到所有LN，而通常扩散到仅一个或两个LN）。因此，从受试者水平癌症分期插补个体LN累及引入假阳性，这混淆LN水平的模型性能（再次，因为在该人的CT扫描内的所有LN被假定为“+”）。

[0160] 然后，对于给定受试者的所有LN站，使用由LN水平分类器预测的最高概率得分来确定N期可能性。图7A至图7B描绘了基于N期可能性的受试者水平性能（AUC=69.1%，灵敏度=72.7%，并且特异性=61.6%）。得分的分布（例如，最大概率）也跨N个期绘制，如图7B所示。这里，对于3期癌症的病例观察到较高的最大概率（左组）。

[0161] 实施例3：递升淋巴结分类方法：自助患者内LN

[0162] 执行两步自助方法以从转移性疾病的LN水平风险转变为转移性疾病的受试者水平风险。特别地，两步自助方法包括如以上参考图3A所描述的插补模型和风险模型的实现。这里，挑战是受试者的N期被插补到所有LN站以训练分类器。考虑到大多数LN是无癌症的，这种插补在LN站引入错误。

[0163] 两个步骤如下：

[0164] 1. 插补：对于患有已知转移性疾病的每个受试者，识别最可能涉及癌症的淋巴结。

这是通过选择具有高于中位受试者内值的癌性概率(来自上一个模型的输出)的LN(包括图3A中所示的插补模型的插补过程)对每个受试者进行的。将低于中位受试者内概率的LN归类为阴性。

[0165] 2. 分类: 群组被拆分成50:50训练: 测试。然后使用LN结果的插补池来训练第二分类器(例如, 图3A中的风险模型320)以预测每个受试者的结节病的存在。

[0166] 该模型的输出是每个LN为癌性的连续概率, 类似于上述实施例2中的模型的输出。然而, 附加的插补和分类步骤的益处是现在有更大范围的概率被分配给患有转移性疾病的那些受试者中的所有淋巴结。更具体地, 那些可能患有癌症的LN具有更高的成为“+”的概率, 而那些不太可能患有癌症的LN具有更低的成为“+”的概率。对于预测步骤, 基于具有最高概率的LN生成LN累及的受试者水平得分(图3B)。具体地, 这涉及采用LN特定概率并提供具有其肺癌的LN累及的受试者特定概率。在图3B所示的测试受试者中, 底部LN被分配了60%的患癌症的机会(例如, $\text{Prob}(+) = 0.6$)。这是三个LN中具有最高“+”概率的LN。因此, 受试者水平概率被分配值0.6。

[0167] 在使用随机森林分类器训练递升方法之后, 自助模型的受试者水平性能达到 $\text{AUC} = 77.7\%$ 、灵敏度 $= 81.8\%$ 和特异性 $= 59.9\%$, 如图9A中所示。因此, 两步自助模型的性能与实施例2中描述的方法相比是改进的。使用LASSO分类器实现类似的性能。最后, 该模型可用于绘制跨N期的得分(或最大概率)的分布(图9B)。有趣的是, 注意到对于具有较高N期的那些情况, 该模型产生较高的最大概率。该概率得分可以在不同的临床上下文中使用, 以实现不同的分类决策制定工具。用于训练和测试的附加数据可以提供将N预测为分类协变量而不是二进制协变量的模型。

[0168] 本申请中引用的所有出版物、专利、专利申请和其他文献出于所有目的据此全文以引用方式并入, 其程度如同每个单独的出版物、专利、专利申请或其他文献被单独地指示为出于所有目的以引用方式并入。

[0169] 虽然已经示出和描述了各种特定实施方案, 但上述说明书不是限制性的。应当理解, 在不脱离本公开的精神和范围的情况下, 可进行各种改变。在查看本说明书后, 许多变型形式对于本领域技术人员将变得显而易见。

[0170] 实施例4: 使用基于CT图像的预测装置的示例性指导干预

[0171] 由于涉及PET扫描的常规方法的较差性能, 实现本文所述方法(例如, 部署用于预测淋巴结的转移癌的淋巴结(LN)水平风险的风险模型)的基于图像的装置将影响临床护理。最近对219名进行了系统性淋巴结解剖/取样的NSCLC患者进行的检查检查了术前PET分期的性能(例如, Li等人, 通过18F-FDGPET/CT. PLoS ONE 8(10): e78552对NSCLC的淋巴结分期中的假阴性和假阳性诊断的暗示, 其全部内容通过引用整体并入)。用于检测肺门和纵隔淋巴结转移的PET的灵敏度、特异性、PPV和NPV为74.2%、54.4%、86.8%和73.5%, 其中假阴性率为13.2%, 并且假阳性率为45.5%。因此, 本文所述的方法代表用于检测已导致淋巴结累及的转移性疾病的非侵入性技术, 其将改善决策制定和临床护理。该装置将对决策制定和临床护理产生的影响的示例包括:

[0172] 提示在肿瘤大小低转移风险的受试者中进行快速附加测试。由于在那些具有小肺肿瘤的患者中没有使用术前PET扫描或淋巴结活检的护理标准, 基于由CT装置所捕获的CT图像的淋巴结累及预测促使随后经由PET扫描或经由受试者的淋巴结的支气管镜取样进行

测试。相比之下,常规方法将指导受试者进行手术并且其中更晚期的疾病分期将仅在手术中被识别。

[0173] 提示获得阴性PET结果的那些中的淋巴结的基于组织的取样。尽管普遍意识到PET具有假阳性和假阴性失败率,但是诊断测试受到PET扫描结果的严重影响。具有较小肿瘤大小和阴性PET的患者通常进行手术而不需要附加的LN取样。该患者的子集将受益于手术之前的附加LN取样。

[0174] 对具有假阳性PET扫描的患者进行分期并证明其适于外科干预。目前,基于PET扫描强烈怀疑患有晚期疾病的患者继续进行非治愈性全身/非外科治疗而不需要附加的确认性LN取样。接受淋巴结累及的预测(例如,基于淋巴结的CT图像)的患者和阴性PET结果患者将改为在做出关于手术的决定之前进行更明确的淋巴结取样。因此,这些患者可以是手术切除的候选者,并且将增加他们治愈的机会。

[0175] 告知哪个LN最可能患有转移性疾病。处于高转移性疾病风险的患者进行淋巴结的支气管镜或手术取样。典型地,该策略将其最强的焦点置于接近肿瘤的那些LN上,使得更远的LN未被取样或被欠取样。这里,本文所述的方法使得能够预测淋巴结累及,并且还能够识别累及的淋巴结。这指导临床团队适当地集中他们的取样策略以获得组织、最大化产量并且最小化遗漏诊断的风险。

[0176] 集中于不具有LN转移的患者的临床试验的优化招募。患有局部疾病的患者经常寻求集中于局部递送(化学疗法或消融的经支气管性针滴注)的临床试验。不幸的是,在进行淋巴结取样的规程期间,经常发现此类患者患有更晚期的癌症。这延迟了适当的护理。因此,本文所述方法的实施方式(例如,部署用于预测淋巴结的转移癌的淋巴结(LN)水平风险的风险模型)使得能够识别表现出淋巴累及的患者,并且因此这些患者不需要招募到集中于不具有转移的患者的临床试验中。

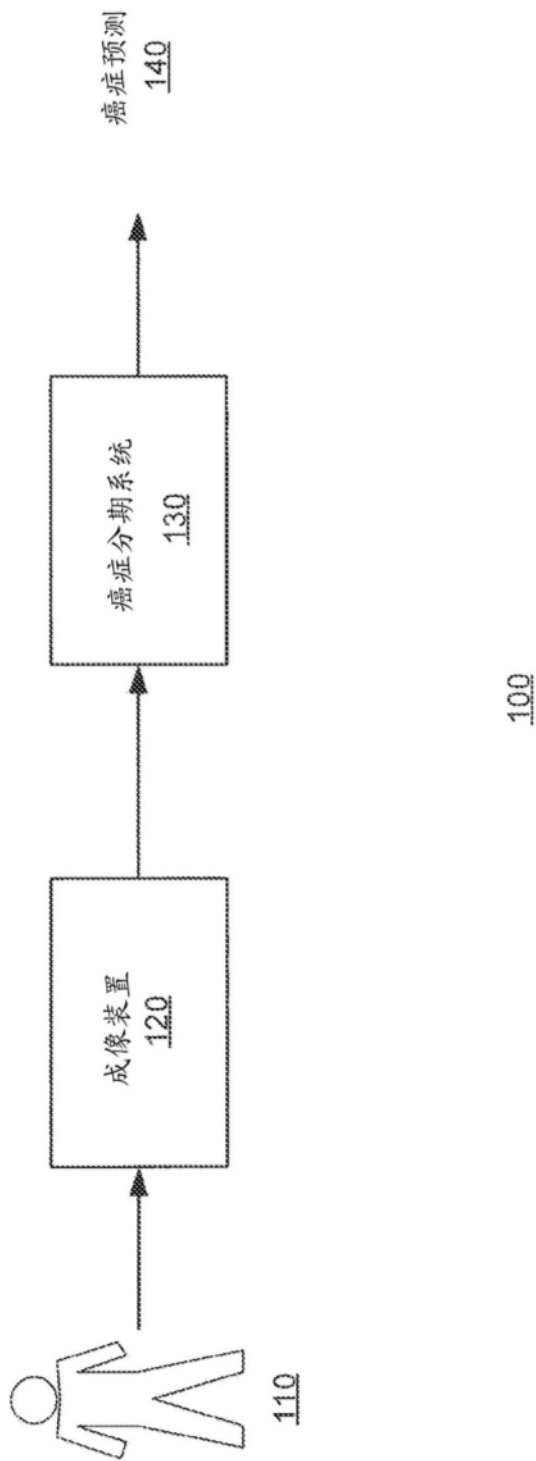


图1A

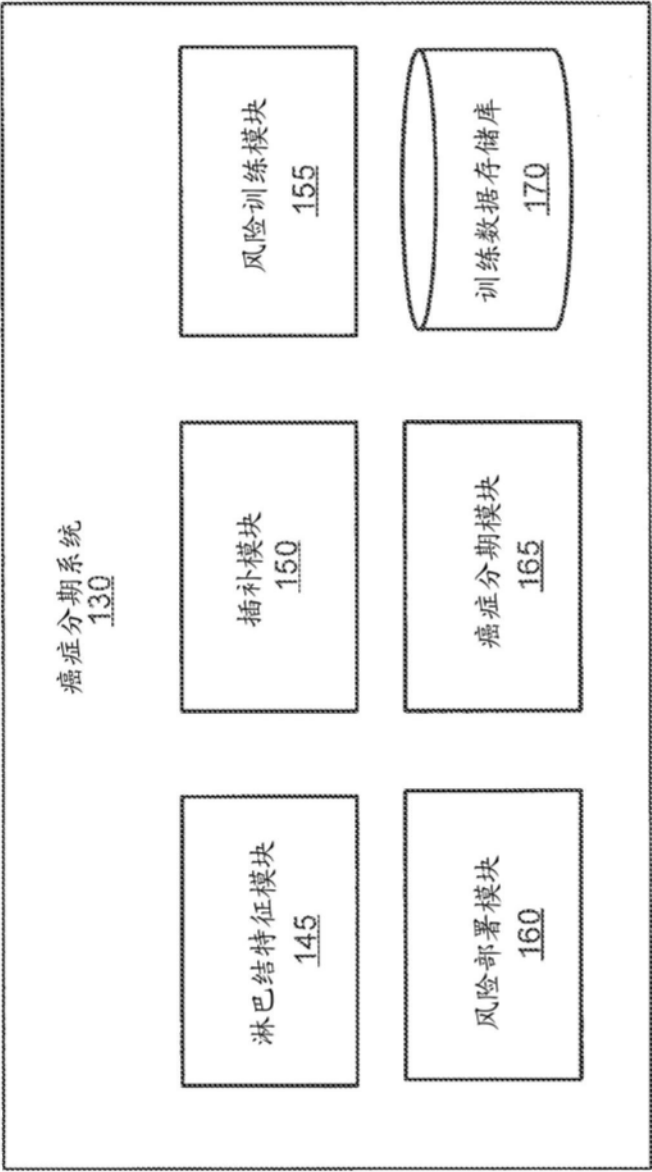


图1B

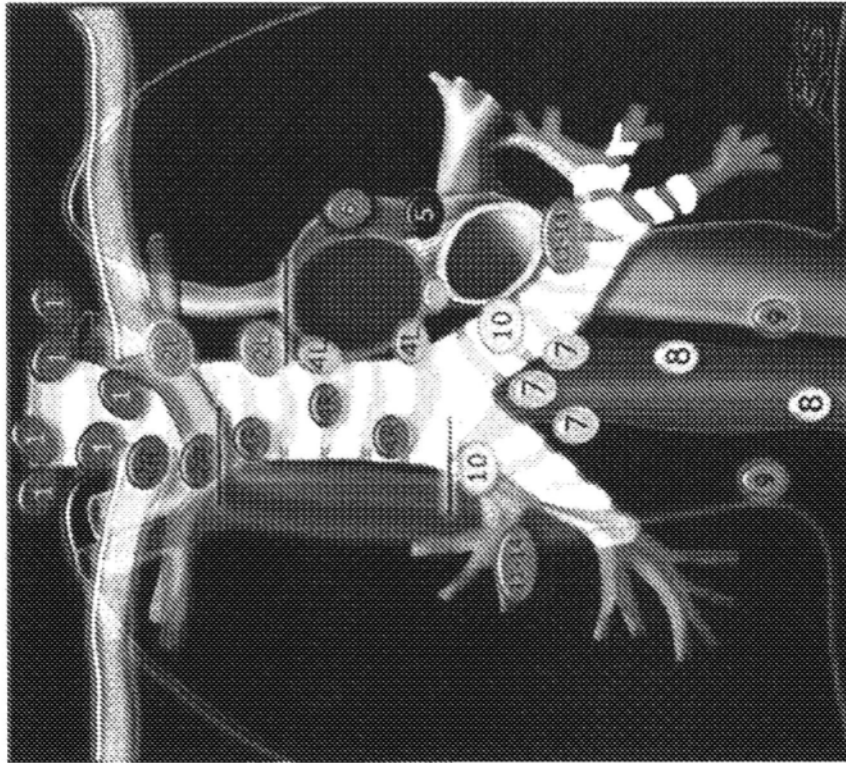


图2A

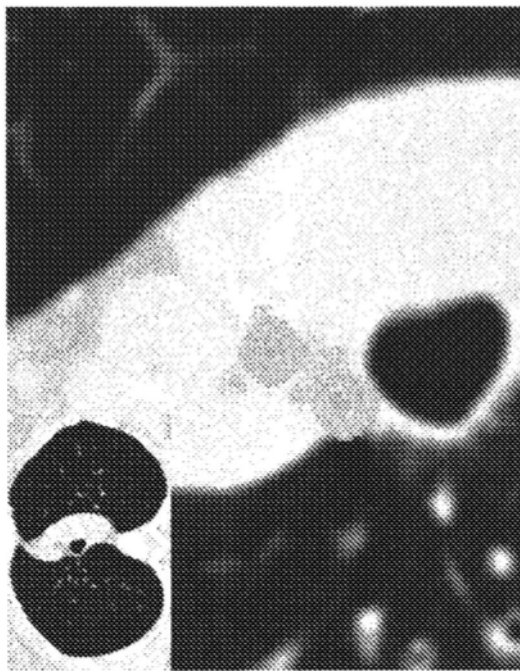


图2B

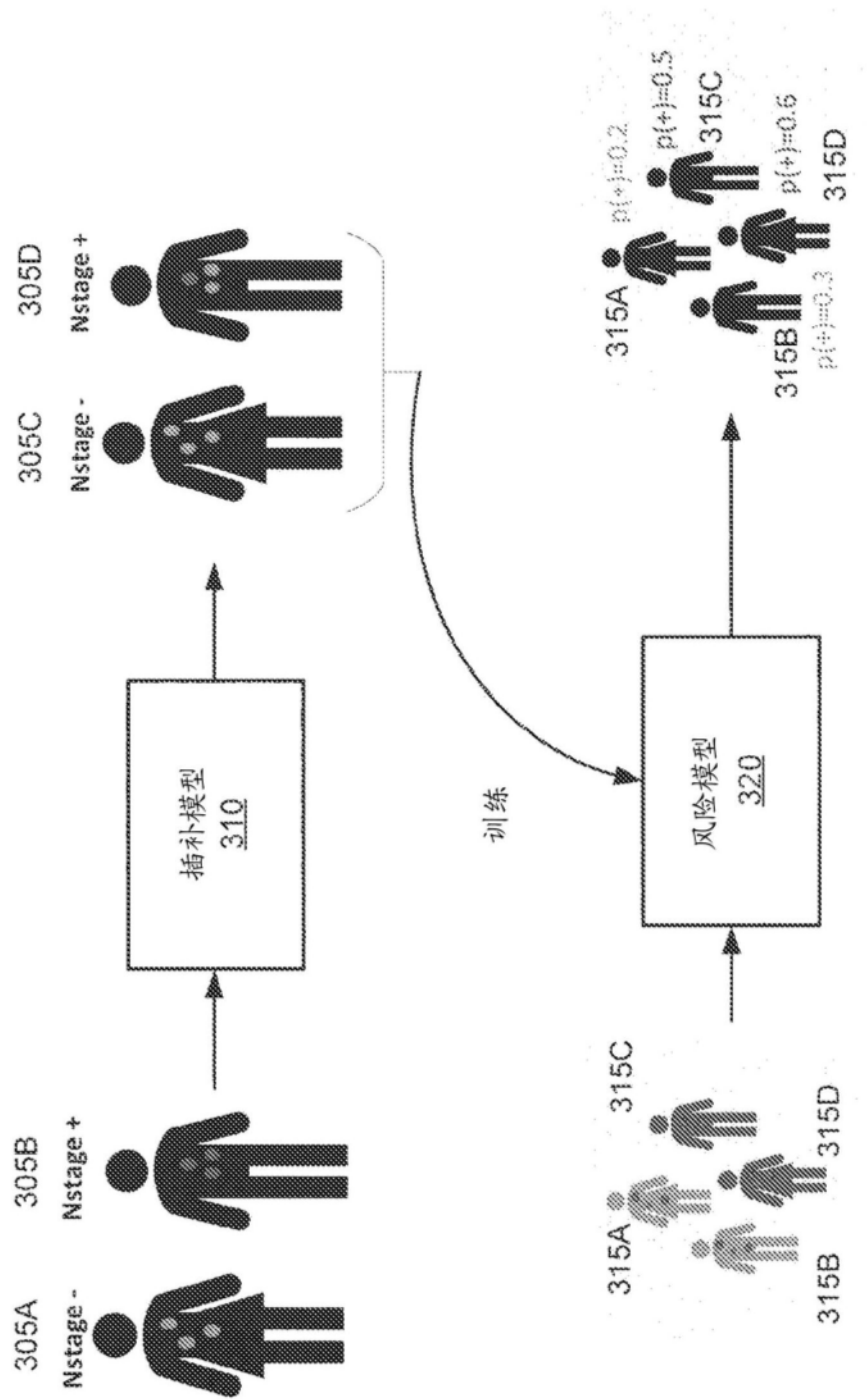


图3A

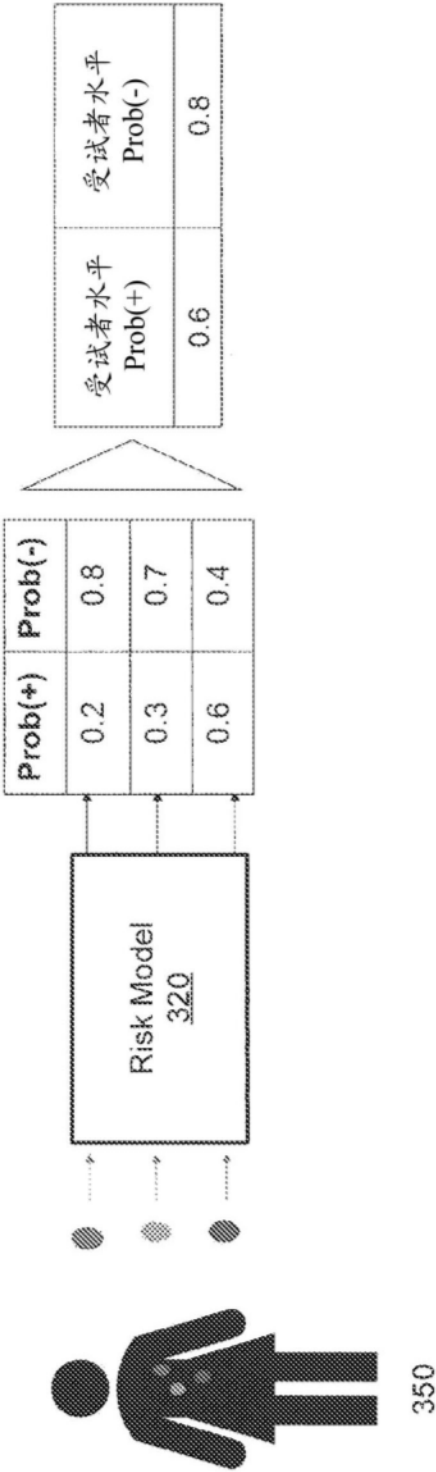


图3B

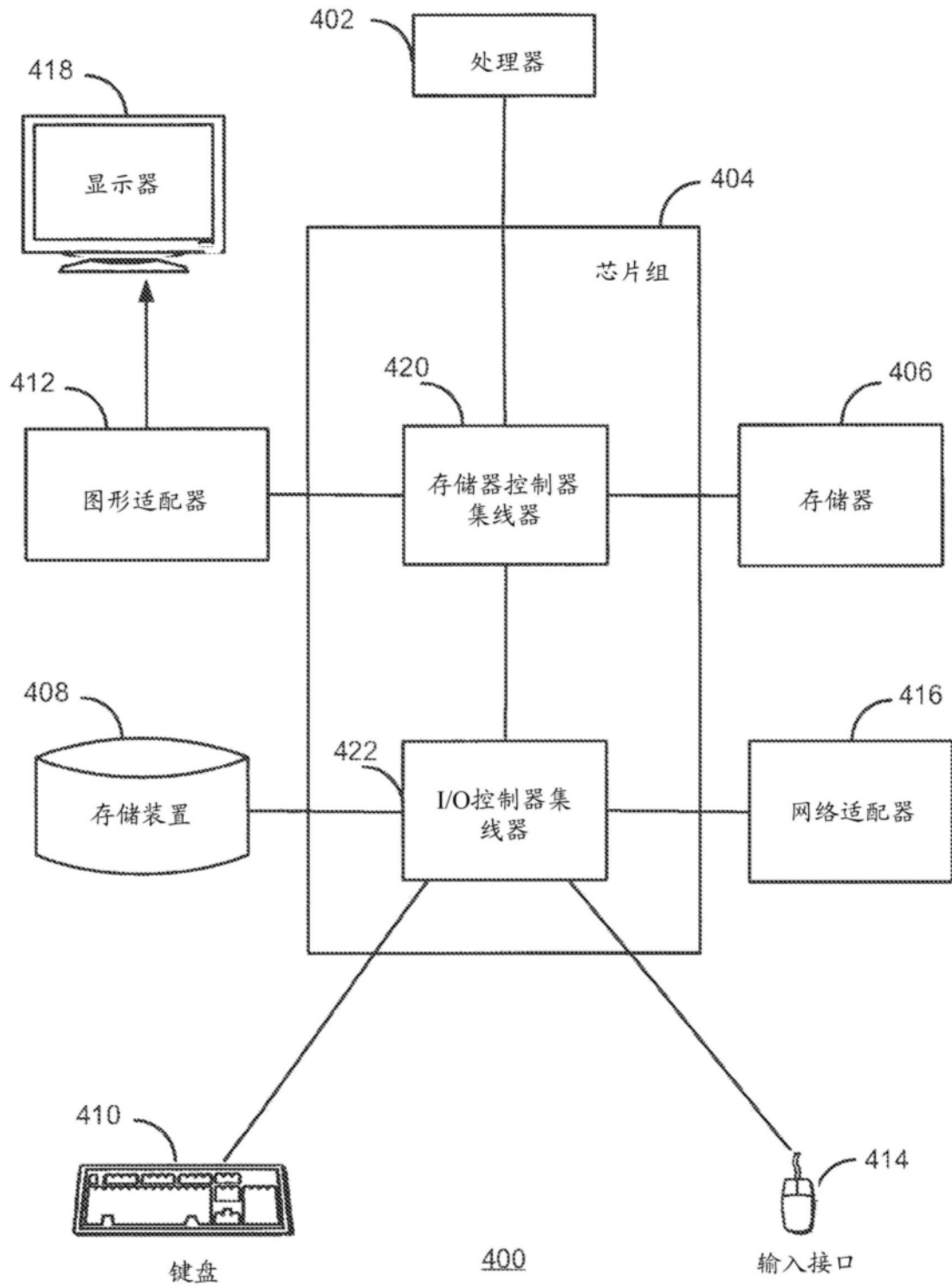


图4



图5A

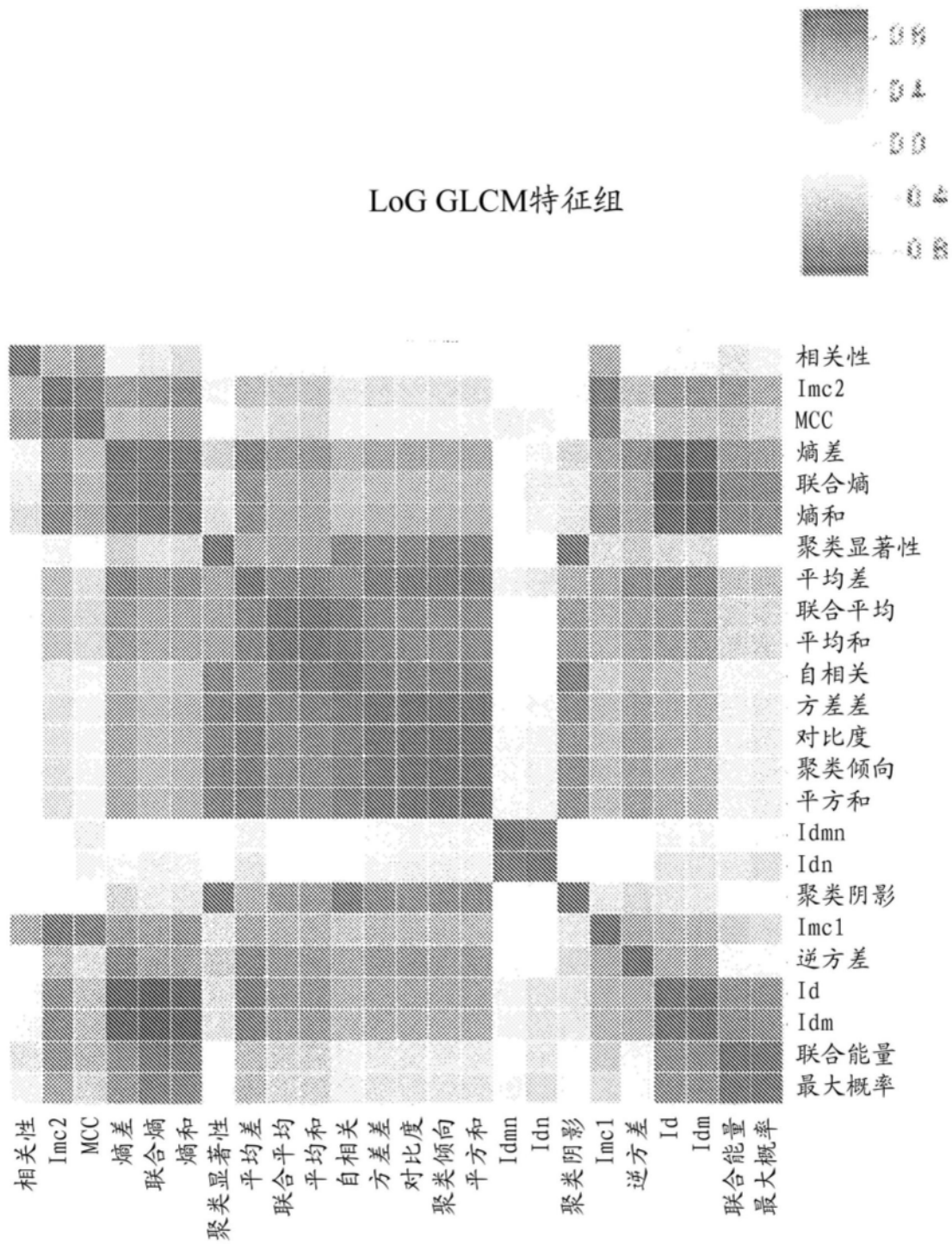


图5B

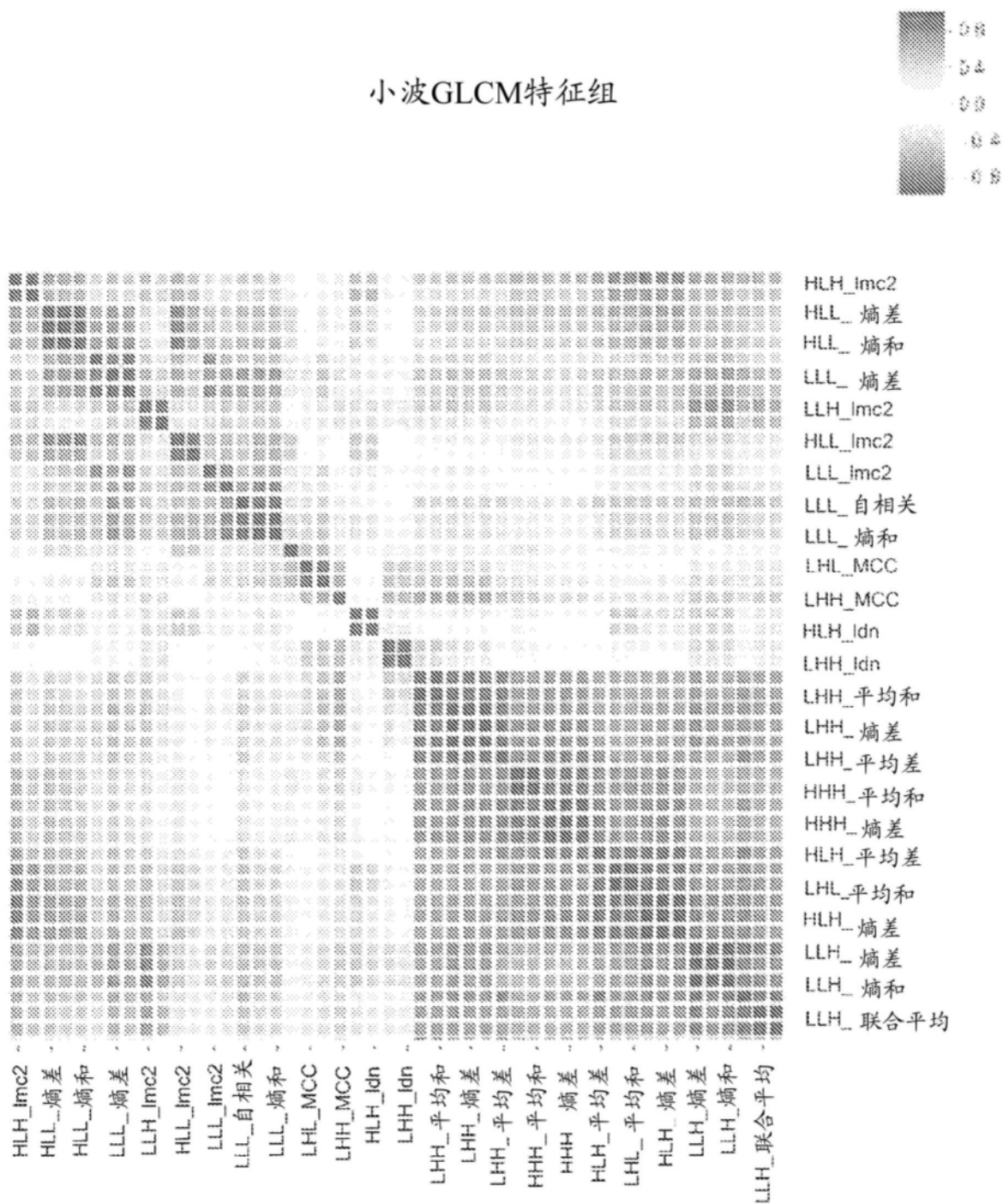


图5C

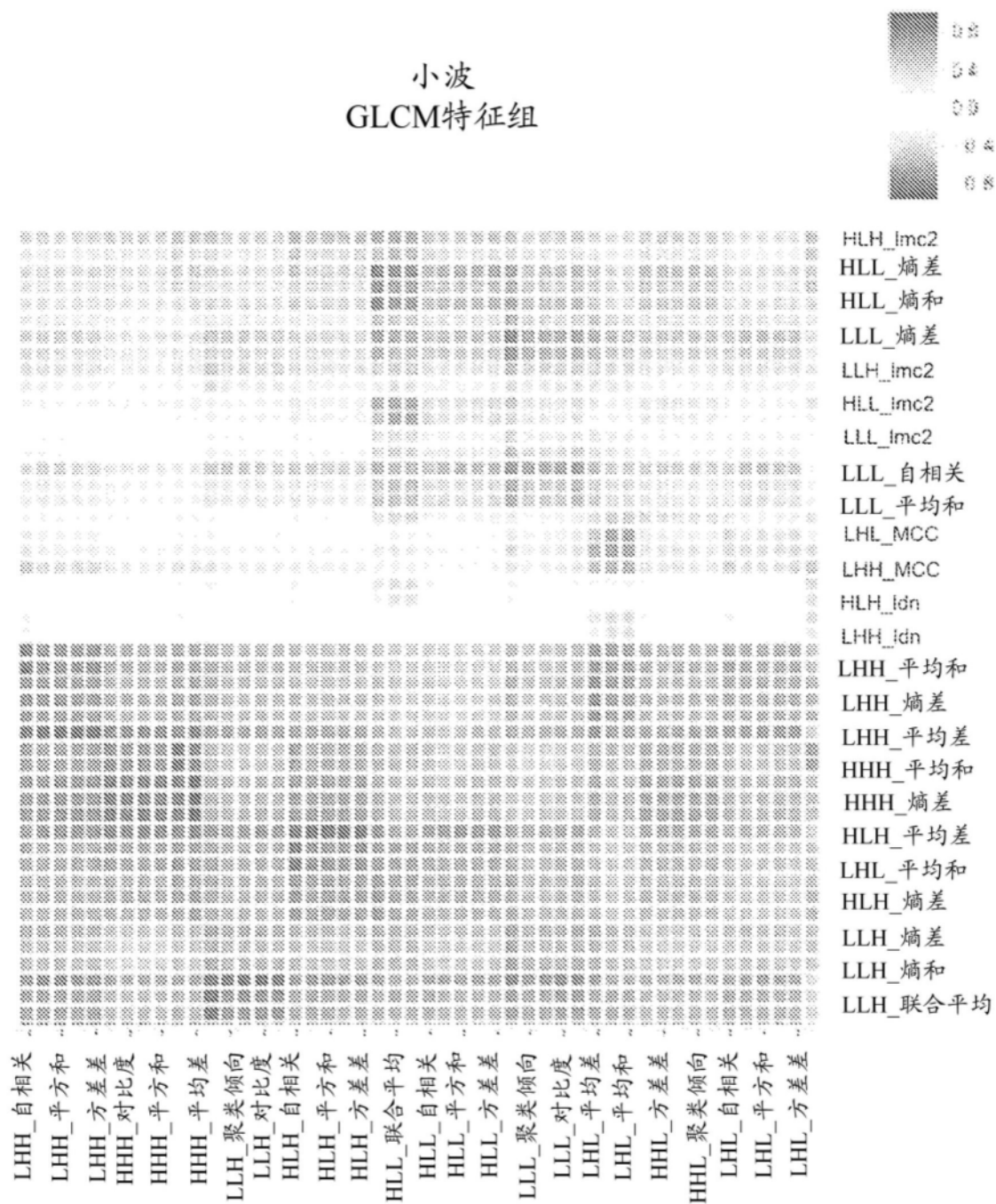


图5D

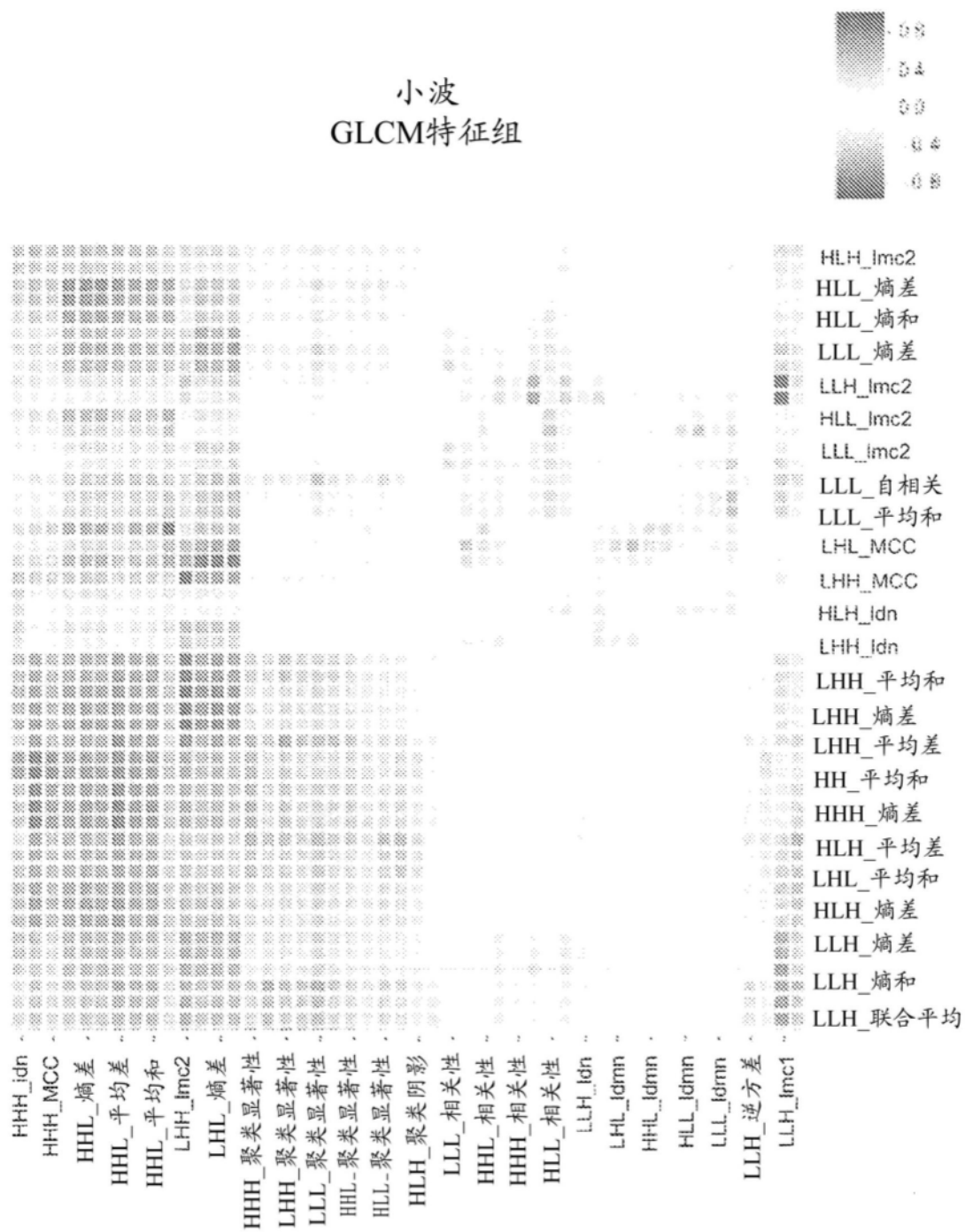


图5E

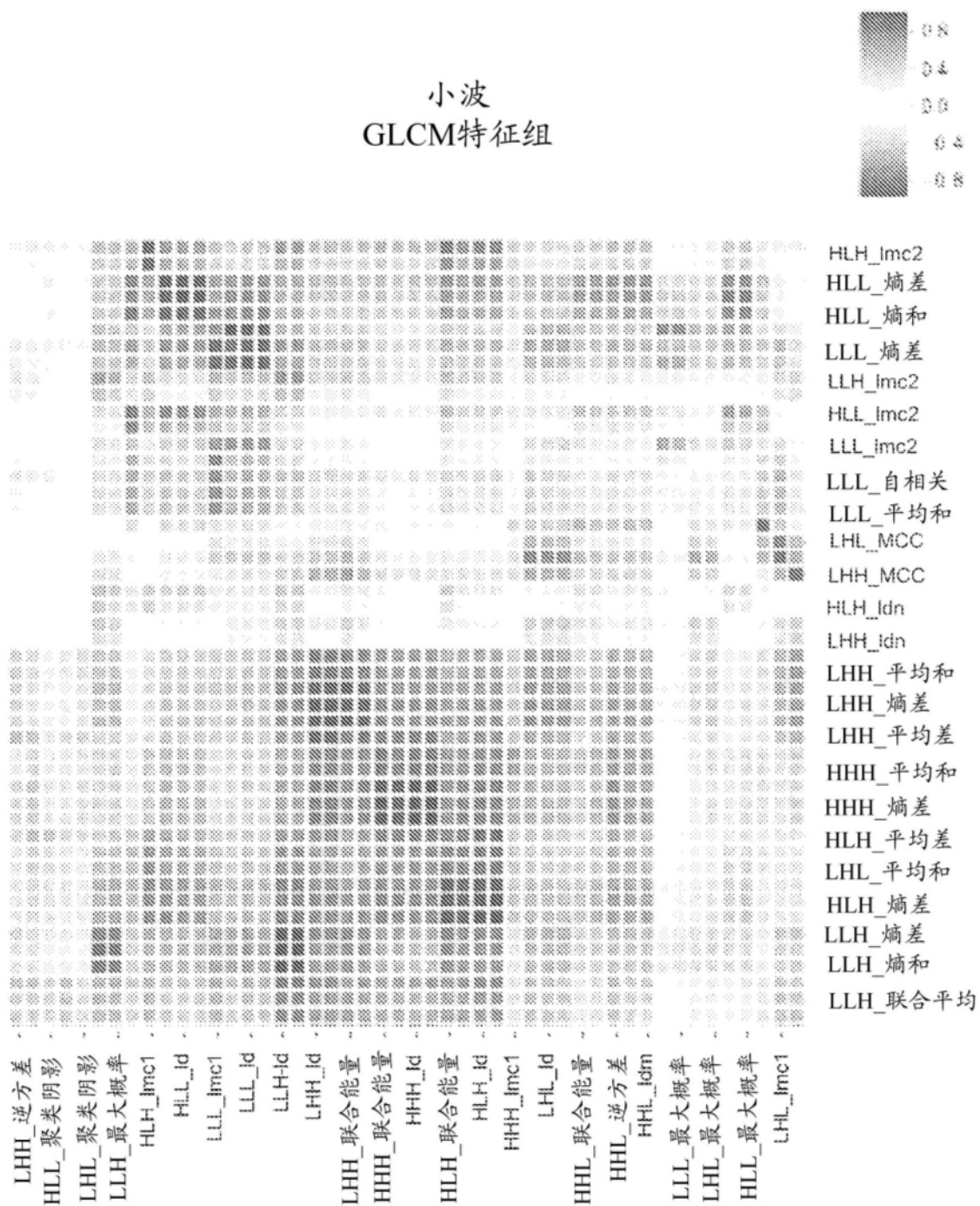


图5F

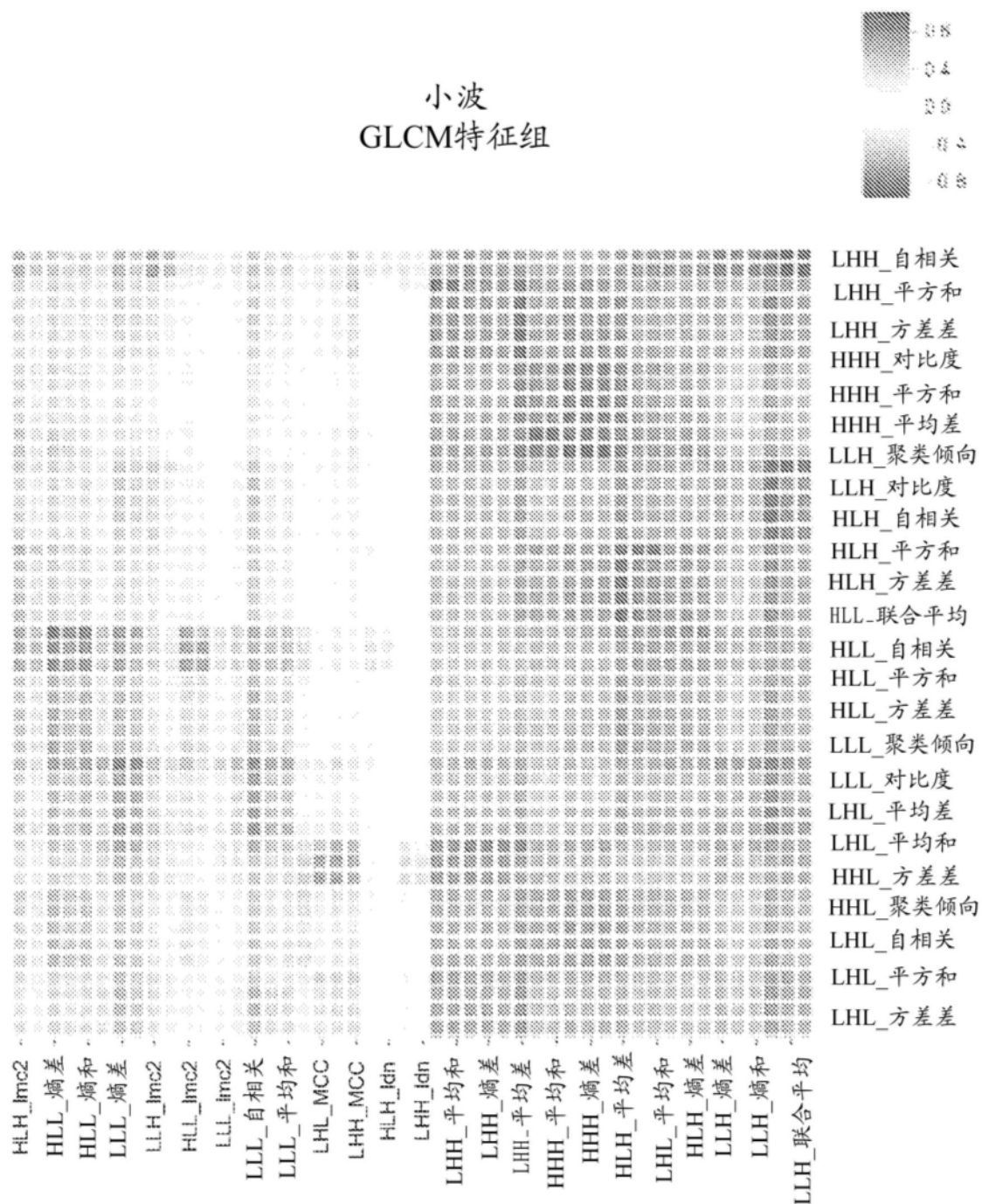


图5G

小波 GLCM特征组



LHH_自相关	LHH_自相关
LHH_平方和	LHH_平方和
LHH_方差差	LHH_方差差
HHH_对比度	HHH_对比度
HHH_平方和	HHH_平方和
HHH_平均差	HHH_平均差
LLH_聚类倾向	LLH_聚类倾向
LLH_对比度	LLH_对比度
HLH_自相关	HLH_自相关
HLH_平方和	HLH_平方和
HLH_方差差	HLH_方差差
HLL_联合平均	HLL_联合平均
HLL_自相关	HLL_自相关
HLL_平方和	HLL_平方和
HLL_方差差	HLL_方差差
LLL_聚类倾向	LLL_聚类倾向
LLL_对比度	LLL_对比度
LHL_平均差	LHL_平均差
LHL_平均和	LHL_平均和
HHL_方差差	HHL_方差差
HHL_聚类倾向	HHL_聚类倾向
LHL_自相关	LHL_自相关
LHL_平方和	LHL_平方和
LHL_方差差	LHL_方差差

图5H

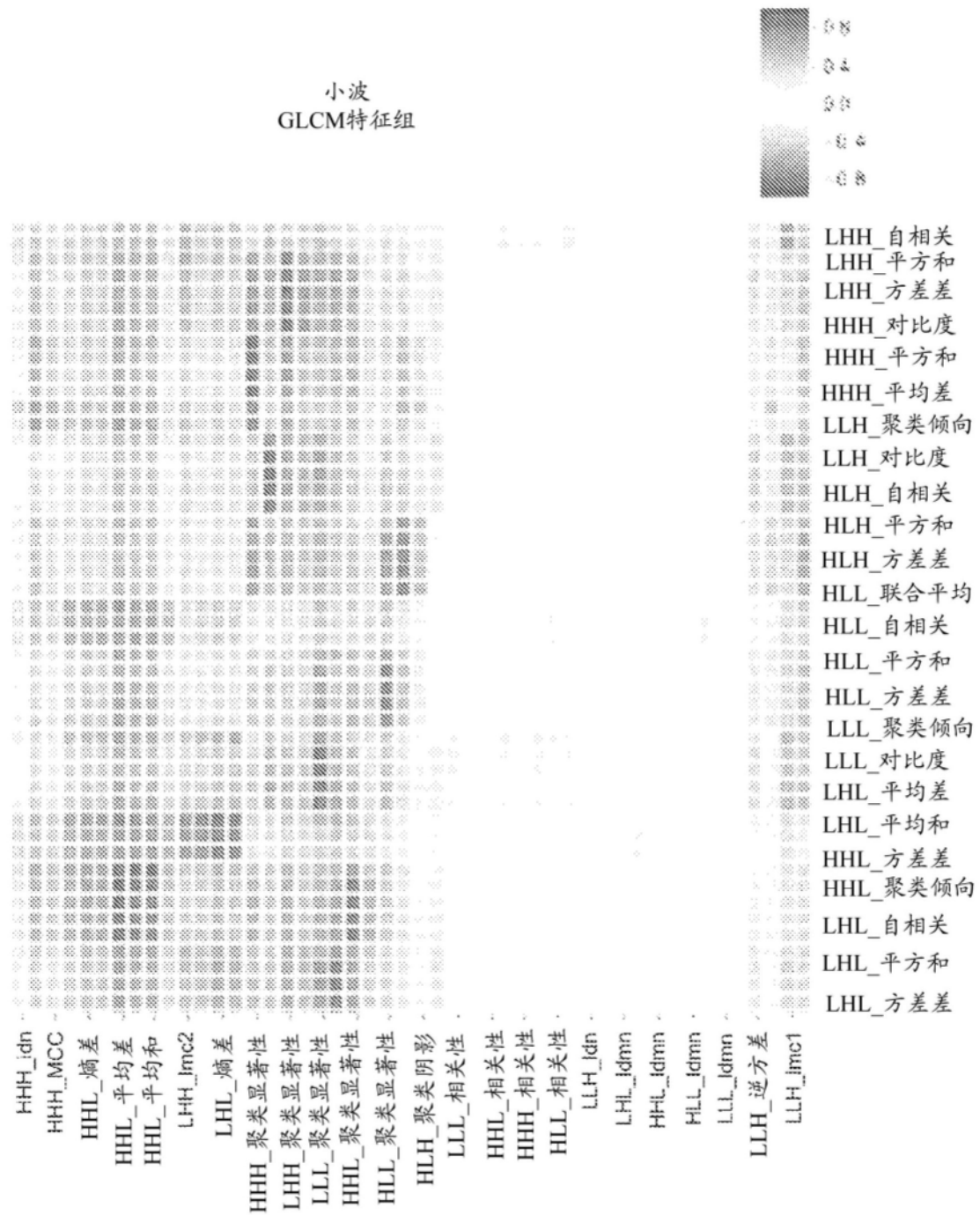


图5I

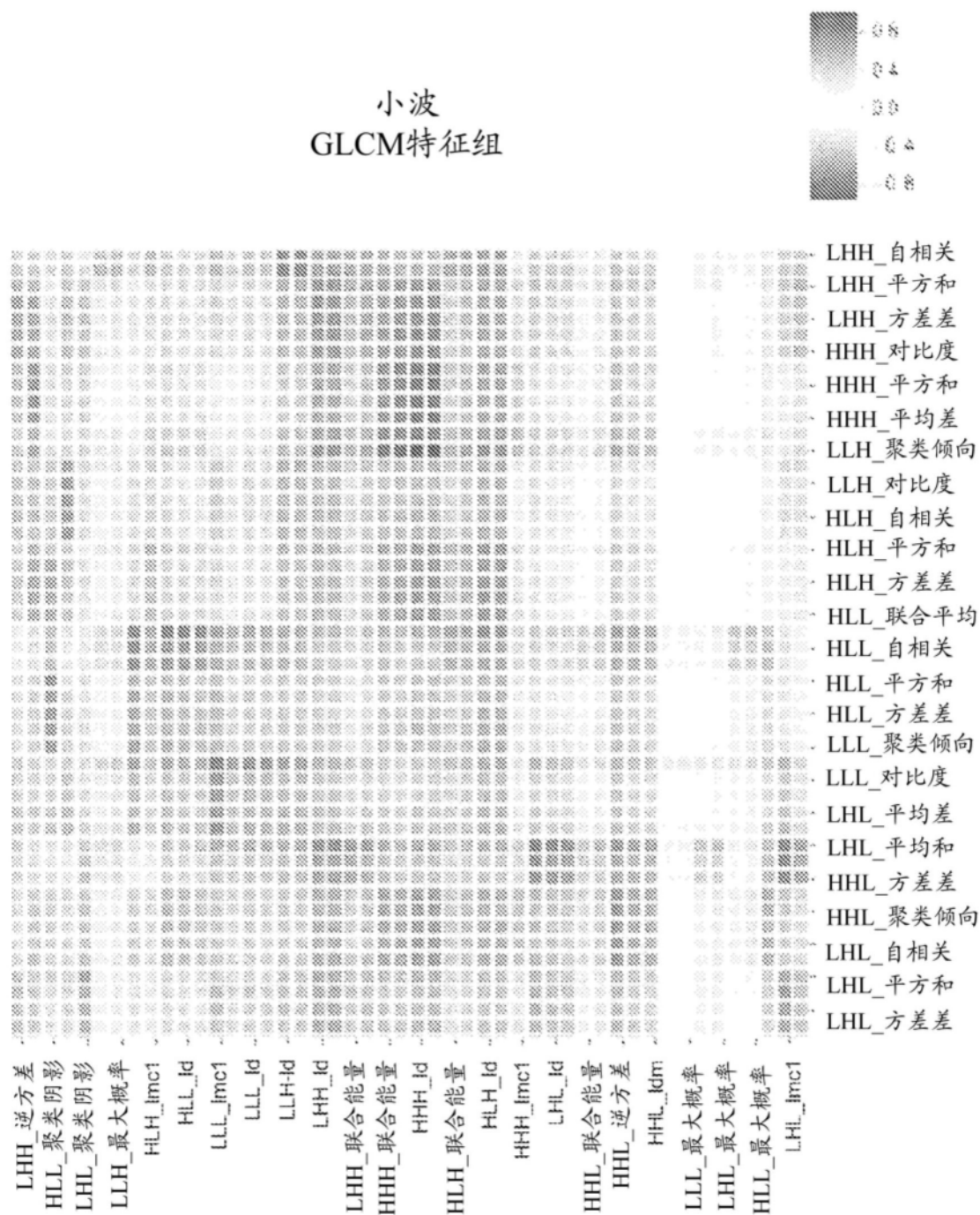


图5J

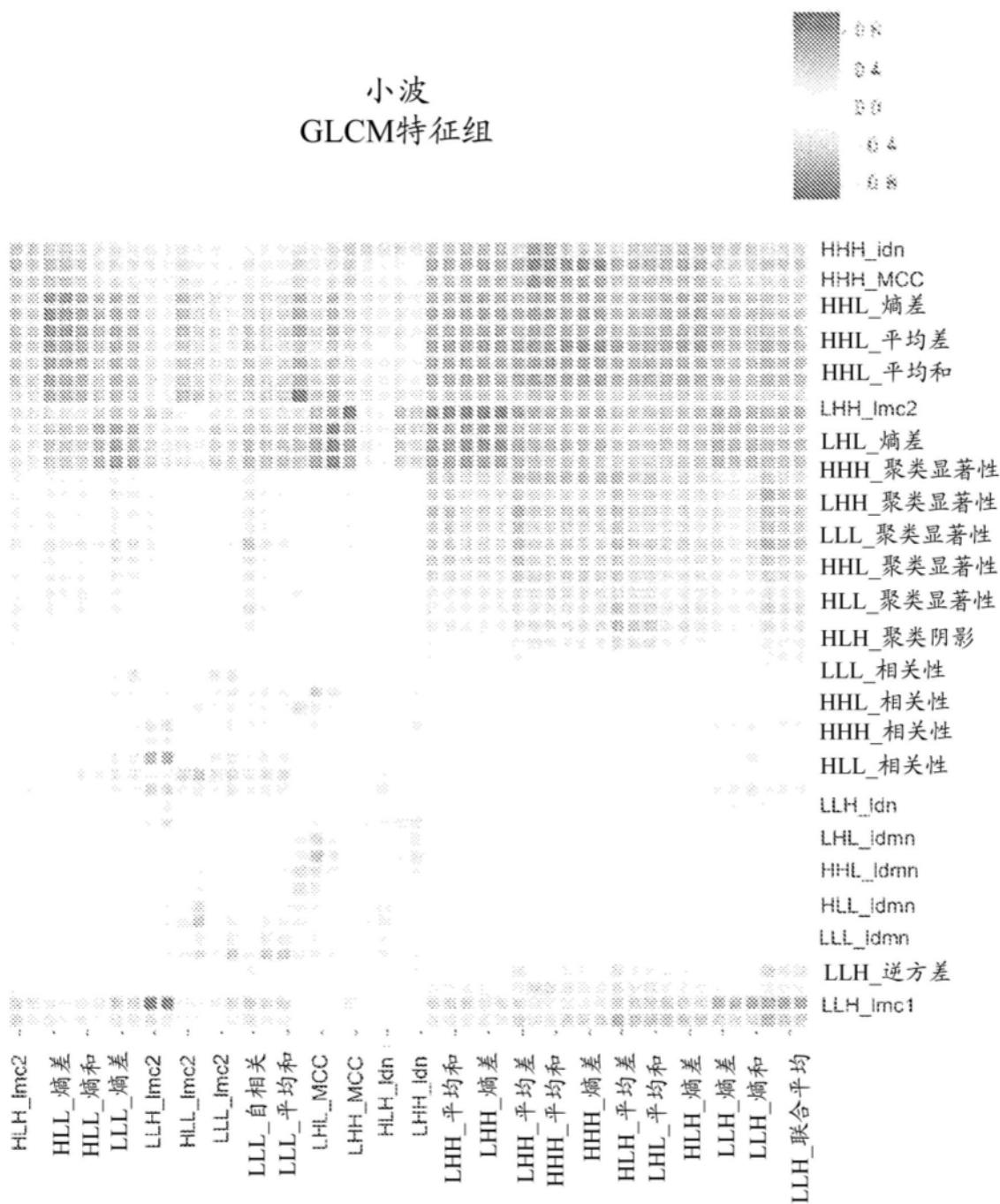


图5K



图5L

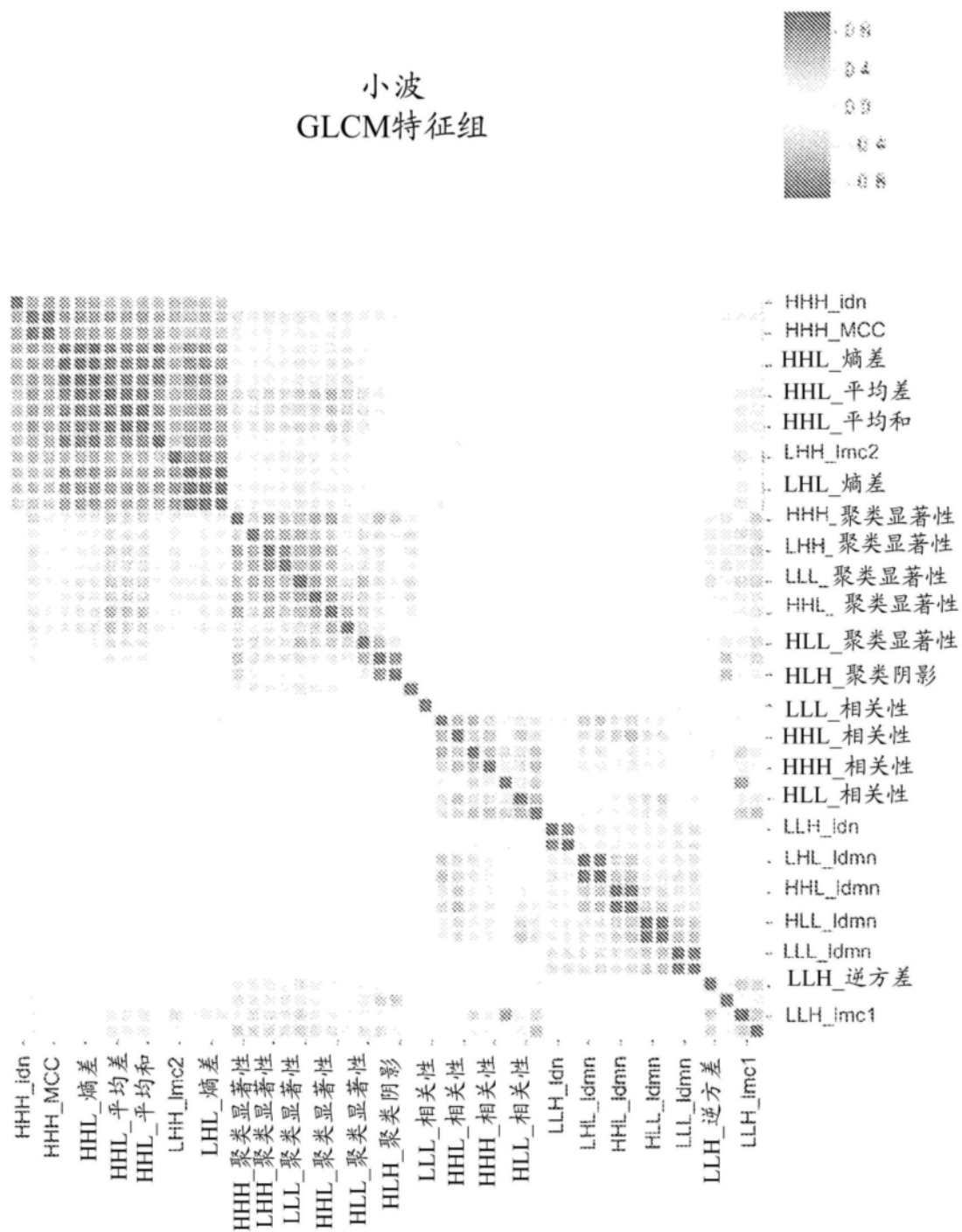


图5M



图5N

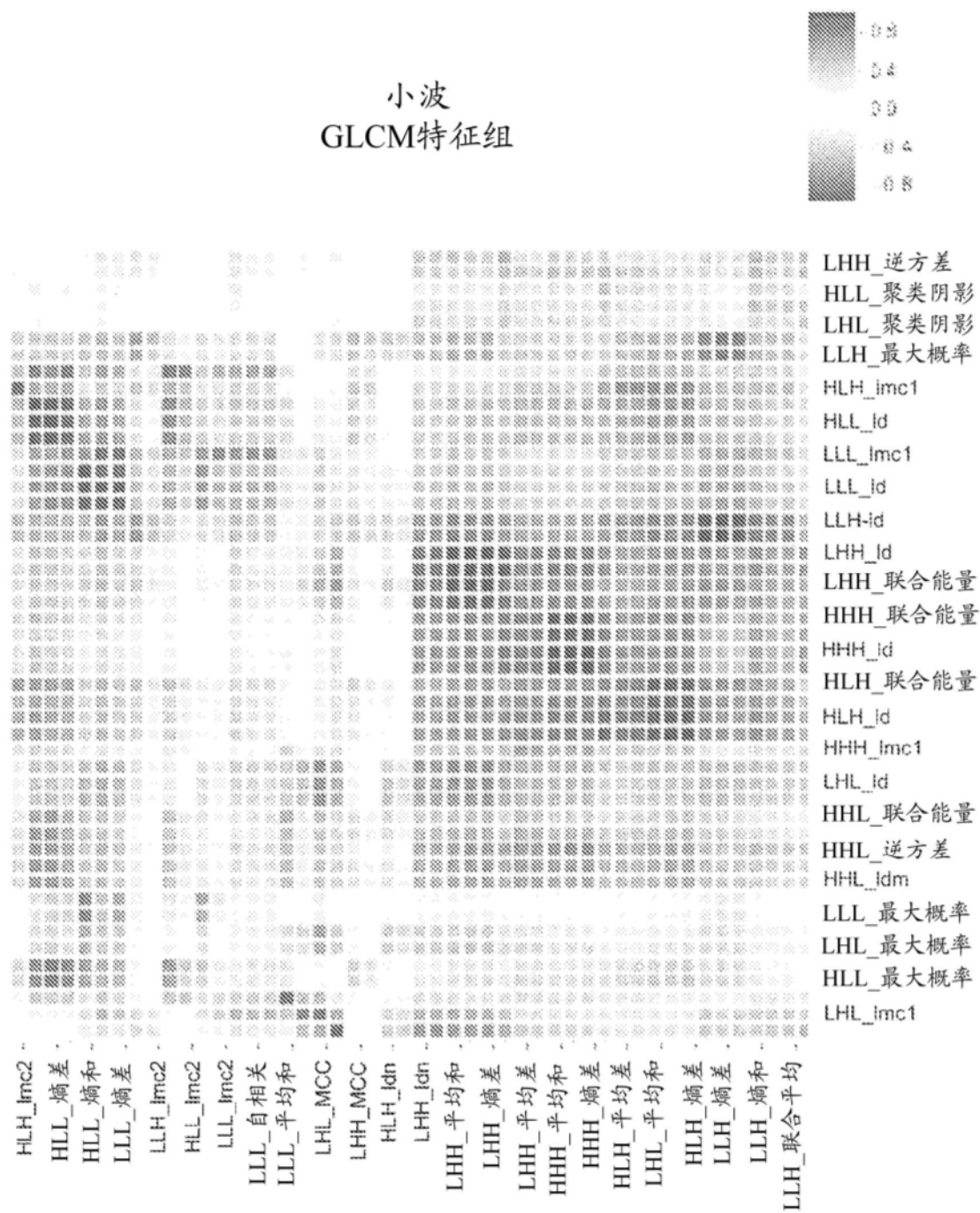


图50

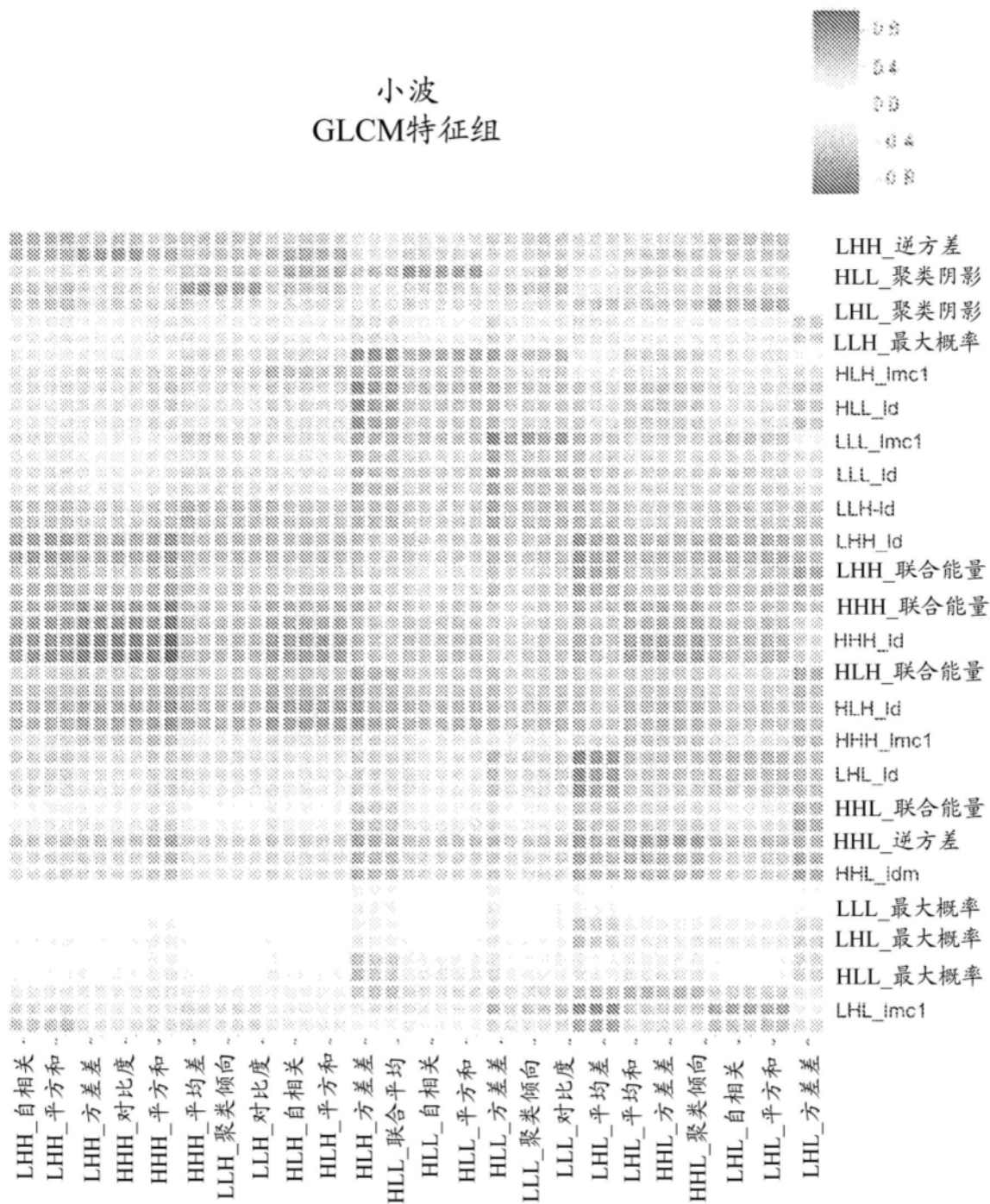


图5P

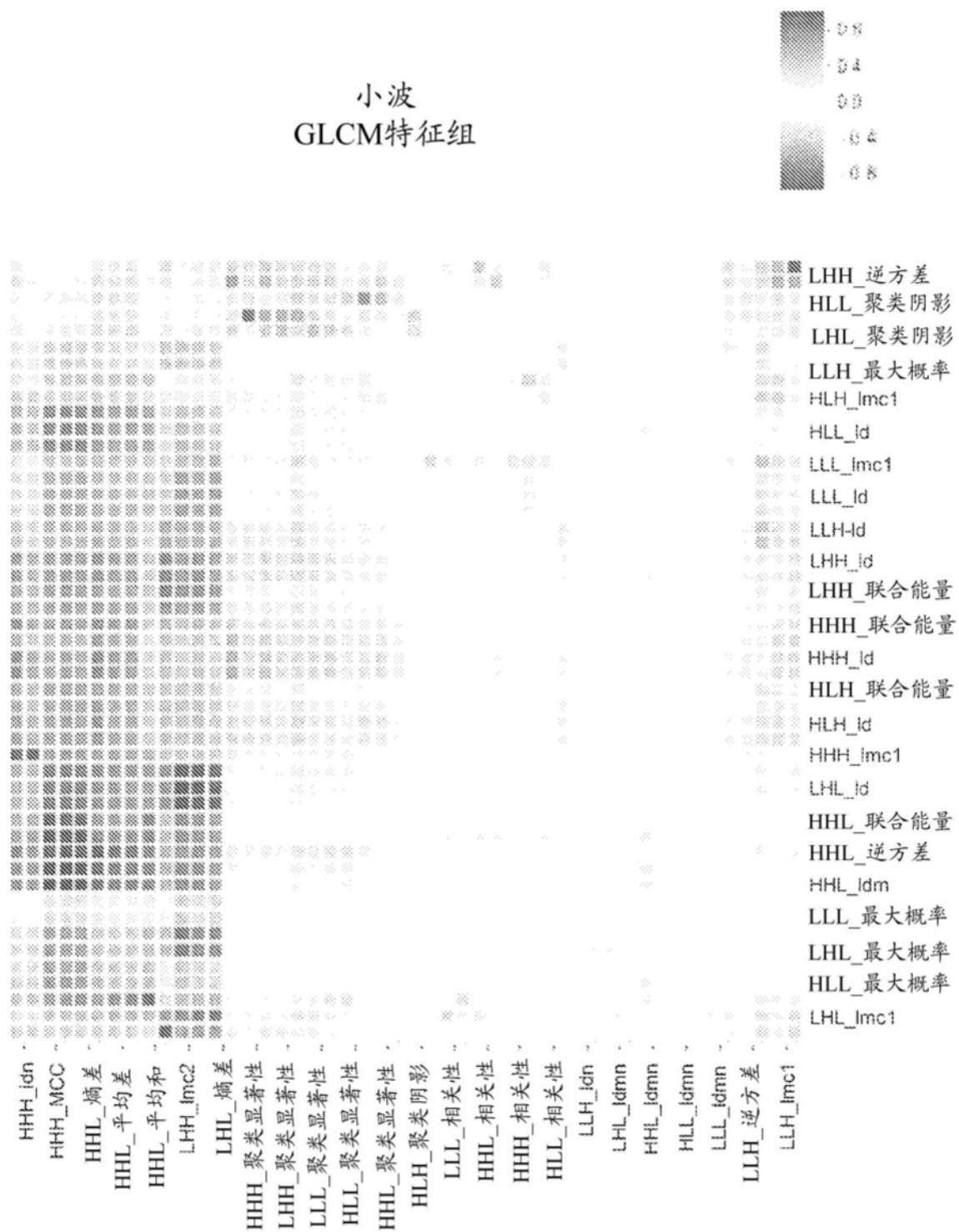


图5Q

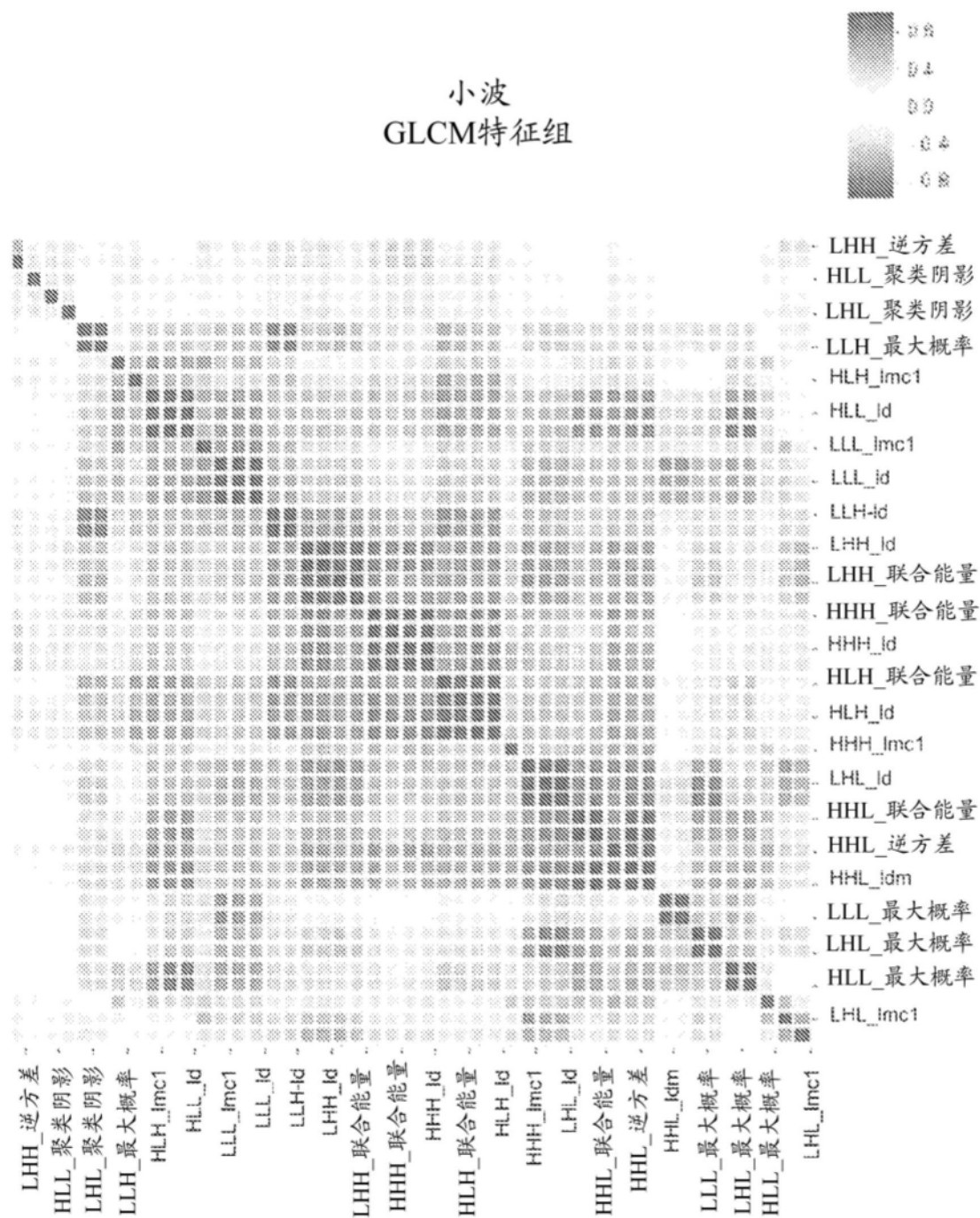


图5R

50-50拆分:LN水平性能

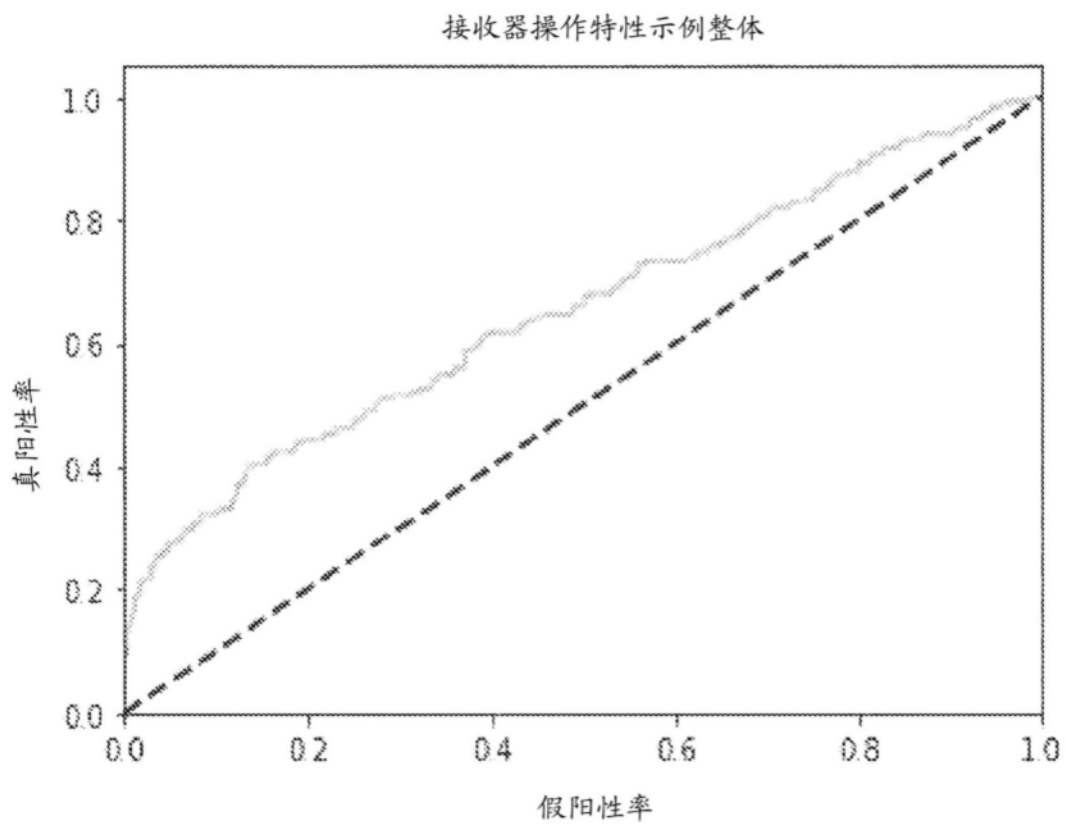


图6

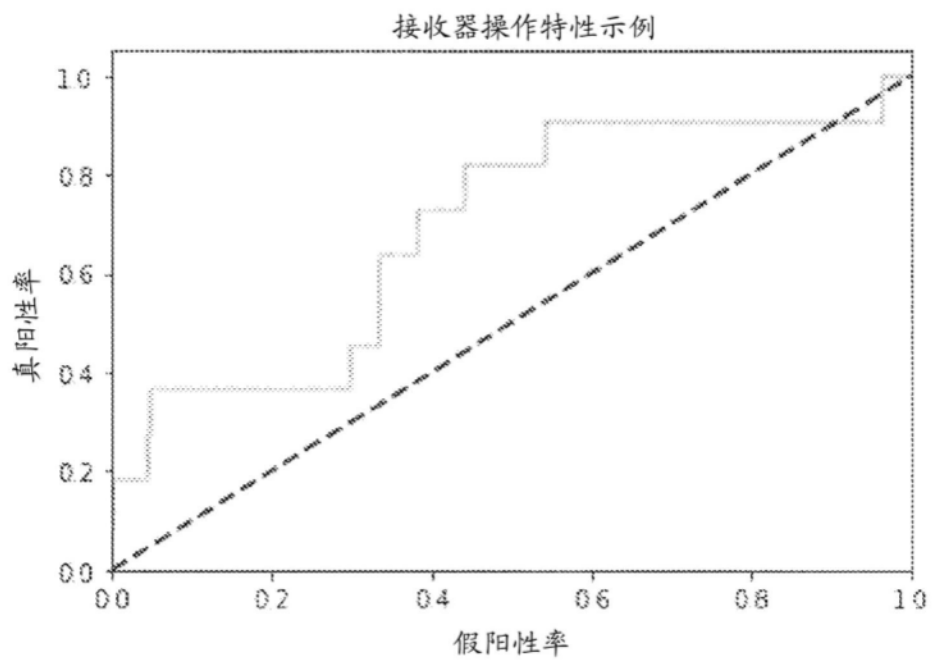


图7A

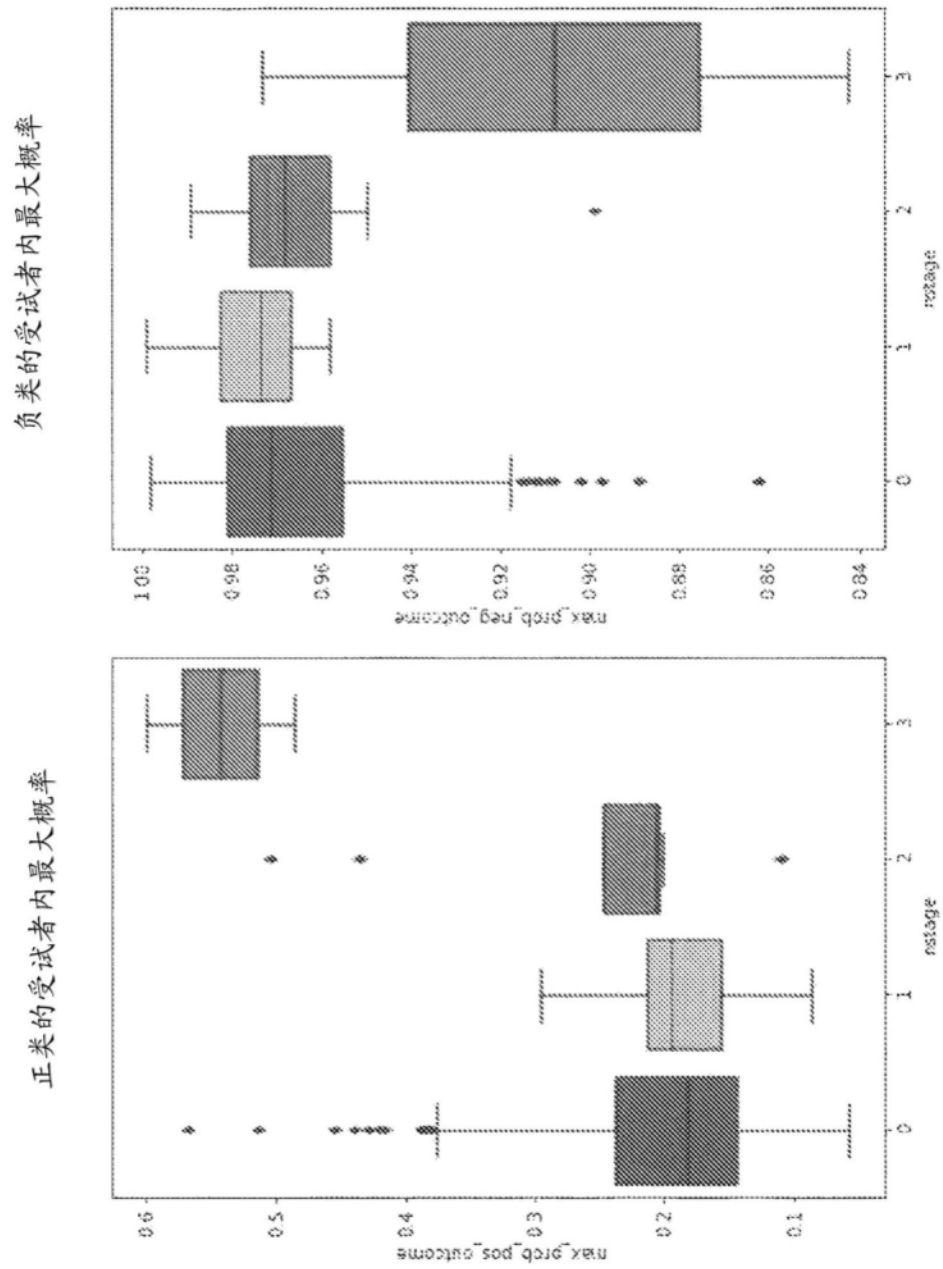


图7B

50-50拆分LN水平性能

接收器操作特性示例整体

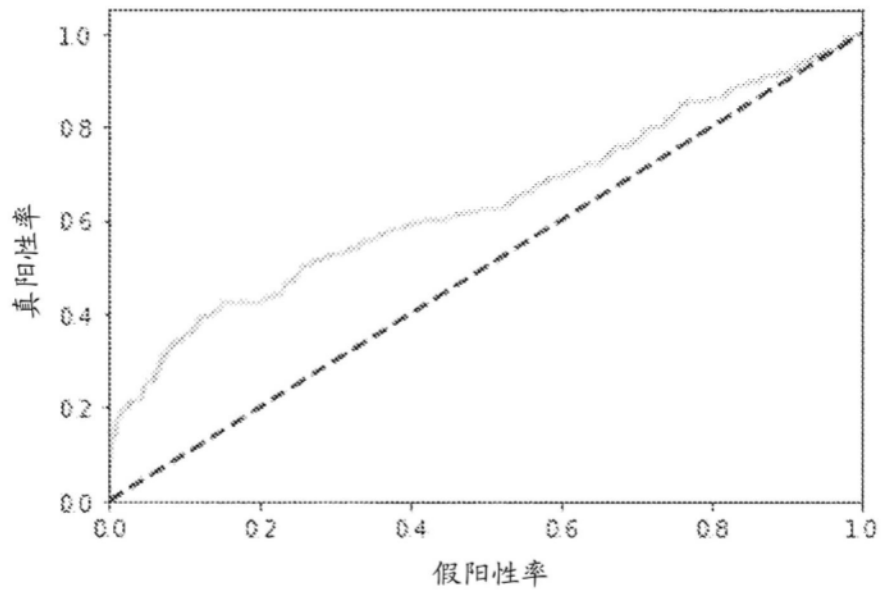


图8

自助方法：
受试者水平性能 (50-50拆分)

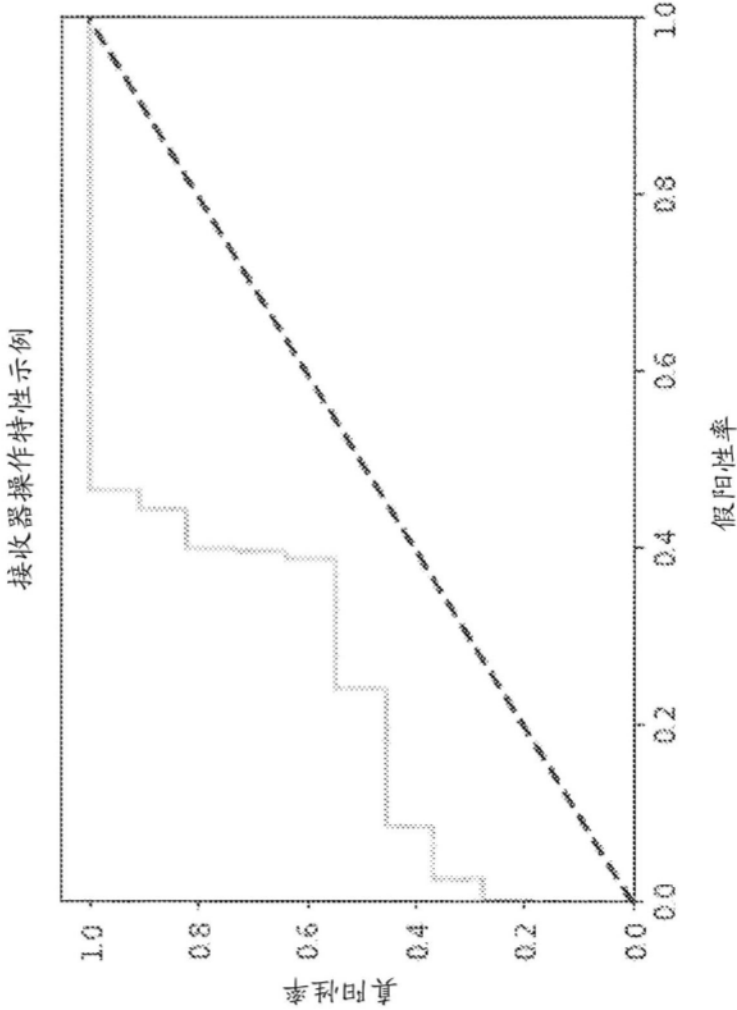


图9A

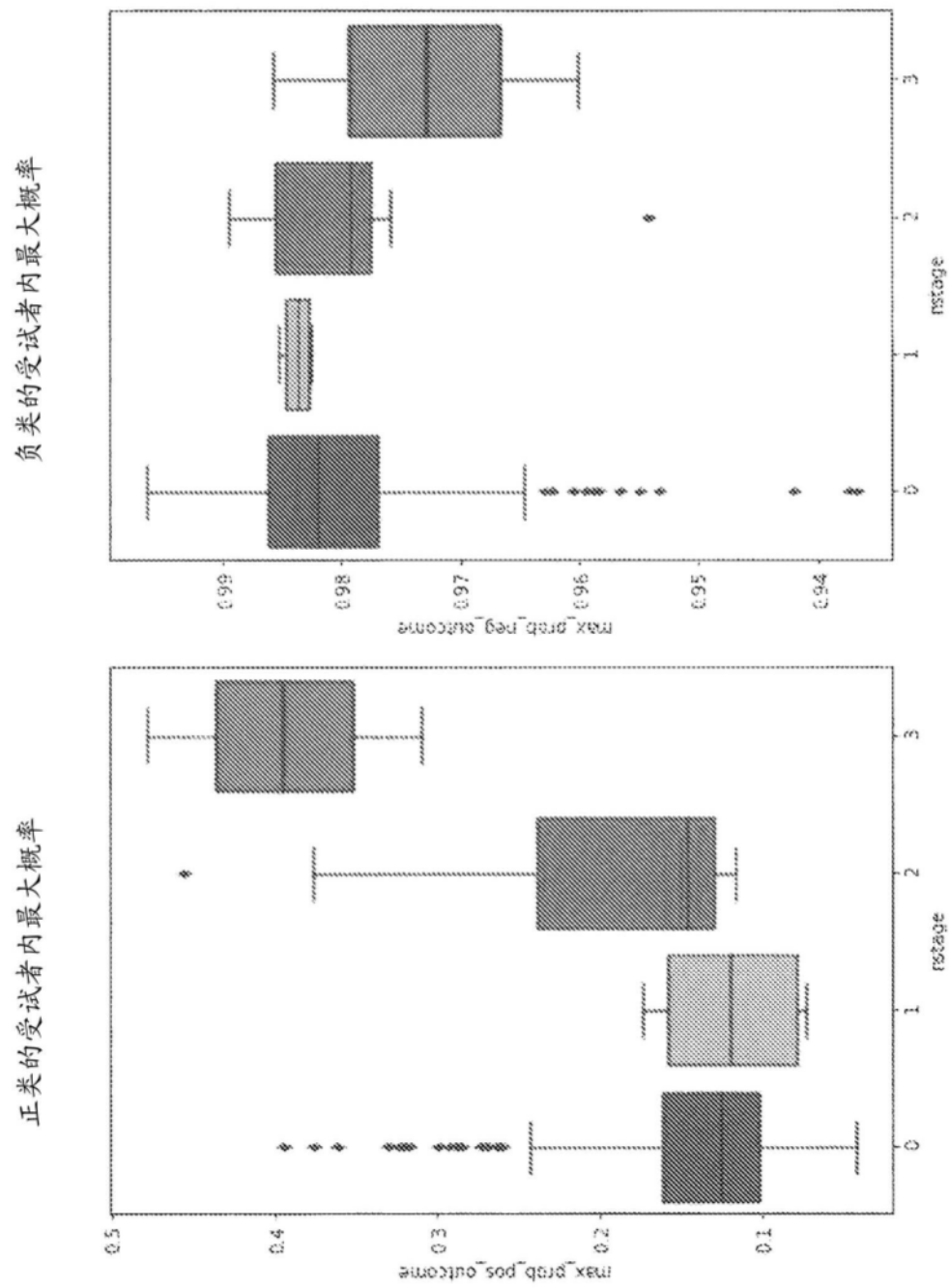


图9B