



(12) 发明专利申请

(10) 申请公布号 CN 115485740 A

(43) 申请公布日 2022. 12. 16

(21) 申请号 202180029759.0

(74) 专利代理机构 北京市铸成律师事务所
11313

(22) 申请日 2021.04.22

专利代理师 王珺 李文颖

(30) 优先权数据

63/013,737 2020.04.22 US

(51) Int.Cl.

G06V 10/764 (2022.01)

(85) PCT国际申请进入国家阶段日

G06V 10/70 (2022.01)

2022.10.20

G06V 10/82 (2022.01)

(86) PCT国际申请的申请数据

PCT/US2021/028563 2021.04.22

(87) PCT国际申请的公布数据

W02021/216822 EN 2021.10.28

(71) 申请人 PDF决策公司

地址 美国加利福尼亚州

(72) 发明人 本田智纪 R·伯奇 朱青

J·D·大卫

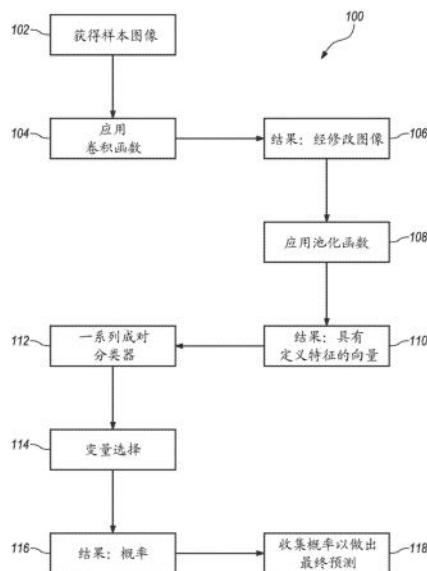
权利要求书3页 说明书6页 附图13页

(54) 发明名称

异常晶片图像分类

(57) 摘要

一种半导体图像分类器。应用卷积函数来修改晶片图像,以便提取关于图像的关键信息。经修改图像被压缩,然后通过一系列成对分类器进行处理,每个分类器被配置为确定图像比所述对中的一者更像另一者。收集来自每个分类器的概率以形成对每个图像的预测。



1. 一种方法,其包括:

获得形成在半导体基板上的半导体晶片的第一多个样本图像,所述第一多个样本图像包括多个不同的图像类型;

将至少第一卷积函数应用于所述第一多个图像中的每一者,以生成多个经修改图像,所述多个经修改图像各自对应于所述多个样本图像中的相应一者;

将池化函数应用于所述经修改图像中的每一者以形成多个对应向量,所述多个对应向量各自定义每个相应经修改图像的多个特征;

对每个对应向量执行变量选择,以便从所述多个特征中选择在确定所述向量的图像类型时更重要的较小的特征集;

在多个基于机器学习的成对分类器模型中处理每个对应向量,所述成对分类器模型中的每一者被配置为确定所述向量更像所述多个不同的图像类型的唯一多个对中的一对的对对应概率;

将来自处理步骤的所确定概率输入到最终的基于机器学习的分类器模型中,所述分类器模型被配置为基于来自所述处理步骤的所确定概率将所述多个图像类型中的一者分配给每个对应向量;以及

运行所述最终的基于机器学习的分类器模型。

2. 如权利要求1所述的方法,应用至少第一卷积函数的步骤还包括:

识别具有值矩阵的内核;以及

将所述第一多个图像乘以所述内核以生成所述多个经修改图像。

3. 如权利要求2所述的方法,进一步地其中所述值矩阵具有与和所述多个不同的图像类型中的一者相关联的特征大小成比例的大小。

4. 如权利要求2所述的方法,其还包括:

通过尝试各自具有不同的值矩阵的多个不同内核来识别所述内核。

5. 如权利要求1所述的方法,其还包括:

在生成所述多个经修改图像之前多次执行所述第一卷积函数。

6. 如权利要求1所述的方法,应用至少第一卷积函数的所述步骤还包括:

将多个不同的卷积函数应用于所述第一多个图像中的每一者以生成多组经修改图像,每组与所述多个不同的卷积函数中的一者相关联,每个经修改图像对应于所述多个样本图像中的相应一者;以及

确定所述多个卷积函数中的至少一者在示出与所述多个图像类型中的至少一者相关联的特征方面是有效的。

7. 如权利要求6所述的方法,其还包括:

将所述多个不同的卷积函数的组合应用于所述第一多个图像中的每一者,以生成所述多组经修改图像。

8. 如权利要求1所述的方法,应用池化函数的步骤还包括:

将统计量度应用于所述第一多个图像中的每一者中的多个点以生成所述多个对应向量中的每一者。

9. 如权利要求8所述的方法,其还包括:

针对跨所述第一多个图像中的每一者的第一方向上的第一多个点获取第一统计量度,

以形成第一多个对应向量;以及

针对跨所述第一多个图像中的每一者的第二方向上的第二多个点获取第二统计量度,以形成第二多个对应向量。

10. 如权利要求1所述的方法,执行变量选择的步骤还包括:

构建多个预测模型,所述多个预测模型各自对应于所述多个成对分类器模型中的一者;

对于每个预测模型:

按等级对与对应向量相关联的所述多个特征进行排序;

多次运行所述预测模型以进行前向变量选择,每次按等级顺序将所述多个特征中的一者添加到所述预测模型中,并且当所添加的一个特征提高所述预测模型的准确度时将所添加的一个特征并入到所述预测模型中;以及

多次运行所述预测模型以进行后向变量选择,每次按反向等级顺序去除所述多个特征中的一者,并且当所述预测模型的准确度因所去除的一个特征的去除而降低时将所去除的一个特征重新并入到所述预测模型中。

11. 一种方法,其包括:

获得形成在半导体基板上的半导体晶片的第一多个样本图像,所述第一多个样本图像包括出现在所述样本图像上的多个不同类别的缺陷;

将多个数学运算符单独或以各种组合应用于所述第一多个图像中的每一者,从而生成多个经修改图像,每个经修改图像对应于所述多个样本图像中的相应一者;

将所述多个经修改图像压缩成多个单维表示,每个单维表示定义每个对应经修改图像的多个特征;

对每个单维表示执行变量选择,以便从所述多个特征中选择在确定所述单维表示的缺陷类别时更重要的较小的特征集;

在多个基于机器学习的成对分类器模型中处理每个对应的单维表示,所述成对分类器模型中的每一者被配置为确定所述单维表示更指示所述多个不同类别的缺陷的唯一多个对中的一对的对应概率;以及

运行最终的基于机器学习的分类器模型,所述分类器模型被配置为基于来自处理步骤的所确定概率将所述多个类别的缺陷中的一者分配给每个对应的单维表示。

12. 如权利要求11所述的方法,执行变量选择的步骤还包括:

构建多个预测模型,所述多个预测模型各自对应于所述多个成对分类器模型中的一者;

对于每个预测模型:

按等级对与对应的单维表示相关联的所述多个特征进行排序;

多次运行所述预测模型以进行前向变量选择,每次按等级顺序将所述多个特征中的一者添加到所述预测模型中,并且当所添加的一个特征提高所述预测模型的准确度时将所添加的一个特征并入到所述预测模型中;以及

多次运行所述预测模型以进行后向变量选择,每次按反向等级顺序去除所述多个特征中的一者,并且当所述预测模型的准确度因所去除的一个特征的去除而降低时将所去除的一个特征重新并入到所述预测模型中。

13. 如权利要求11所述的方法,其中所述数学运算符是卷积函数。

14. 如权利要求11所述的方法,压缩步骤还包括:

将池化函数应用于所述多个单维表示。

15. 一种方法,其包括:

获得样本图像;

通过一系列成对分类器处理所述样本图像,每个成对分类器被配置为确定所述样本图像比多个类别的缺陷的唯一对中一者更像另一者的概率;以及

处理所确定概率以生成用于将所述多个类别的缺陷中的一者分配给所述样本图像的最终预测。

16. 如权利要求15所述的方法,处理所述样本图像的步骤还包括:

在所述成对分类器之前将卷积函数和池化函数应用于所述样本图像以将所述样本图像降为定义多个特征的至少一个单维表示,使得所述成对分类器确定对应的单维表示的概率。

17. 如权利要求16所述的方法,其还包括:

多次运行每个成对分类器以进行前向变量选择,每次将所述多个特征中的一者添加到所述成对分类器中,并且当所添加的一个特征提高所述成对分类器的准确度时,将所添加的一个特征并入到所述成对分类器中;以及

多次运行每个成对分类器以进行后向变量选择,每次去除所述多个特征中的一者,并且当所述成对分类器的准确度因所去除的一个特征的去除而降低时,将所去除的一个特征重新并入到所述成对分类器中。

18. 如权利要求17所述的方法,其还包括:

运行最终分类器模型,所述最终分类器模型被配置为基于来自所述多个成对分类器中的每一者的结果将所述多个类别的缺陷中的一者分配给所述单维表示。

19. 如权利要求19所述的方法,其还包括:

针对多个样本图像重复方法步骤。

20. 一种方法,其包括:

获得形成在半导体基板上的半导体晶片的第一多个样本图像,所述第一多个样本图像包括多个不同的图像类型;

计算多个单维度量,所述多个单维度量各自对应于所述多个样本图像中的相应一者,每个度量各自定义相应样本图像的多个特征;

对每个度量执行变量选择,以便从所述多个特征中选择在确定所述度量的图像类型时更重要的较小的特征集;

在多个基于机器学习的成对分类器模型中处理每个对应度量,所述成对分类器模型中的每一者被配置为确定所述度量更指示所述多个不同的图像类型的唯一多个对中的一对的对应概率;以及

运行最终的基于机器学习的分类器模型,所述分类器模型被配置为基于来自处理步骤的所确定概率将所述多个图像类型中的一者分配给每个对应向量。

异常晶片图像分类

[0001] 交叉引用

[0002] 本申请要求2020年4月22日提交的名称为“Defect Image Classifier”的美国临时申请号63/013737的优先权,该申请以引用方式整体并入本文。

技术领域

[0003] 本申请涉及半导体制造工艺,并且更具体地,涉及用于使用图像分析对半导体晶片难题/问题进行分类的系统和方法。

背景技术

[0004] 一般来讲,例如通过诸如Google、Facebook等基于网络的努力,图像分类已得到高度发展,它们按其性质可访问数百万张图片,这些图片可用作基于机器的图像分类方案的训练集。然而,典型的半导体公司具有小得多的数据集,可能只有几百张图像用于训练集。因此,期望调整和开发使用这些较小数据集对半导体晶片进行基于机器的分析和分类的技术。

附图说明

[0005] 图1是示出使用图像处理技术对晶片难题进行分类的过程的流程图。

[0006] 图2A至图2D示出四种不同类型或类别的原始晶片图像,包括正常、瑕疵、未指明缺陷和条纹。

[0007] 图3A至图3D分别是各自通过高斯模糊函数修改的图2A至图2D的四个不同晶片图像。

[0008] 图4是用于实现图3A至图3D所示的高斯模糊函数的示例7x7内核。

[0009] 图5A至图5D分别是通过边缘检测函数修改的图2A至图2D的四个不同晶片图像。

[0010] 图6A至图6C是用于实现边缘检测函数的示例3x3内核,其中图6C的边缘检测函数实现图5A至图5D所示的边缘检测函数。

[0011] 图7A至图7D分别是通过锐化函数修改的图2A至图2D的四个不同晶片图像。

[0012] 图8是用于实现图7A至图7D所示的锐化函数的示例3x3内核。

[0013] 图9是示出可用于特征选择和成对分类的分层建模过程的流程图。

[0014] 图10是更详细地示出图9的成对分类步骤的流程图。

[0015] 图11是更详细地示出图10的变量选择步骤的流程图。

[0016] 图12A和图12B是示例成对分类器的准确度的图形图解。

[0017] 图13A和图13B是示出从各个成对分类器聚合的预测结果的准确度的表格。

具体实施方式

[0018] 本公开涉及用于基于异常类型对半导体晶片的图像进行分类的方法和系统。半导体加工应用程序的机器学习解决方案的核心是通过特征工程和特征选择开发的。特征工程

是从原始数据生成特征以更好地表示预测机器学习模型的解决方案的问题的过程。特征选择(也称为变量选择)是选择对理解和预测建模的特定问题贡献最大的特征并丢弃那些贡献不显著的特征的过程。然而,针对任何特定应用程序实现特征工程和特征选择的方式继续发展,并且每个应用程序当然依赖于基于客户端的上下文,包括半导体处理环境的细节。特征工程常常被认为包括特征选择步骤。

[0019] 在一种方法中,可通过首先将一个或多个卷积函数应用于晶片图像的原始图像来识别并生成特征,以通过图像观察和分析来尝试且更好地识别特定晶片问题。每个卷积的结果是经修改图像,其(希望)更详细地示出特定晶片问题。然而,因为不同的卷积函数或卷积函数的组合可揭示不同的问题,并且因为每个制造设施通常具有操作和环境差异,所以可存在不同的函数或函数组合,其作为在不同位置处的相同类型问题(诸如瑕疵、缺陷或条纹)的解决方案集。

[0020] 在卷积步骤之后,应用全局池化函数来将经修改图像压缩成单维向量结果。向量结果在许多方面更容易处理:更简单的表示;更少的存储和处理要求;数字度量可用于逻辑电路等。因此,池化步骤用于通过对图像应用各种池化函数来定义可能感兴趣的特征。在一个实施方案中,可在跨图像的各种点处应用统计函数,尽管也可以采用其他类型的函数。例如,模糊积分(诸如Choquet积分)已用于基于模糊数学压缩数据集,快速傅里叶变换已用于频率分析,应用了简单的线性聚合等。

[0021] 然后使用分层建模方法来选择特征,以关注用于区分类别对的变量。例如,针对每对定义的问题类别,构建单独的预测分类器模型。此外,每个预测分类器模型执行其自己的变量选择步骤,以便将作为模型输入的变量的数量减少并将这些变量限制为该预测模型的重要变量。因此,每个预测分类器模型的结果指示晶片问题是否更像第一问题类别或第二问题类别,而不是常规分类器的方法,常规分类器的方法试图将类别分配给向量,而不是像本文,一系列成对比较。

[0022] 最终分类器预测模型基于作为来自单独的成对预测分类器中的每一者的输入的预测概率对问题进行最终预测。

[0023] 图1是示出基于图像处理对晶片进行分类的简单过程100的流程图。在步骤102中,获得样本图像。然后在步骤104中应用一个或多个卷积函数以在步骤106中生成经修改图像。一般来讲,应针对新的或更新的客户端应用程序单独和组合尝试多个不同的卷积函数,以寻找将对来自该应用程序的图像有用的最有效的图像处理技术。

[0024] 然后在步骤108中应用池化函数,所述池化函数用于将二维图像减少成步骤110中具有多个定义的特征的一维表示。例如,可通过在跨图像的维度的各种点处应用统计量度来产生向量表示,例如,特征可以是在 $x=1$ 、 $x=2$ 等处获取的标准偏差。

[0025] 在步骤112中,通过一系列成对分类器处理样本图像的每个一维表示(即向量),包括作为变量选择过程114的一部分迭代运行每个分类器,所述变量选择过程将每个分类器模型的特征数量减少到两个至四个特征。步骤116中的结果是概率或可能性:向量更像唯一对中的第一个还是唯一对中的第二个?在步骤118处收集来自所有成对分类器的概率以对每个成对比较进行最终预测。

[0026] 呈现了几个示例来帮助说明这些概念。图2A至图2D是在半导体工艺中使用常规光学成像设备获取的晶片(在基板上)的四个不同的原始图像。每个图像可作为200kB JPG文

件或类似文件存储。图2A是基板222的图像220的示例,所述图像具有形成在基板上的看起来“正常”的圆形晶片224,即从图像看没有明显问题。图2B是基板242的图像240的示例,所述基板具有形成在基板上的圆形晶片244,所述图像包括显现为在图像的右上角中的大黑圈246的突出的黑色“瑕疵”以及基板上与晶片相邻的若干其他瑕疵。图2C是具有圆形晶片264的基板262的图像260的示例,所述图像具有显现为晶片边缘上的若干小黑圈的未指明“缺陷”,诸如该图像的右下角上的缺陷246。图2D是在基板上具有圆形晶片284的基板282的图像280,所述图像具有在整个基板和晶片图像上出现的一系列垂直边缘或“条纹”286。这些类型的问题(瑕疵、缺陷、条纹)中的每一种具有不同的根本原因,并且需要不同的分析和关注。当然,可存在可为此处描述的同类分析的对象的其他问题或缺陷。

[0027] 卷积

[0028] 在获得样本晶片图像(“原始图像”)之后,将一个或多个卷积函数应用于样本。这些技术可应用于黑白或灰度图像以及彩色图像。根据与特定晶片组相关联的问题,不同的卷积函数或函数组对于感兴趣的特定问题可能是必要的或重要的。事实上,可能需要一些试验和错误来确定用于在特定晶圆厂位置处捕获特定类型问题的信号的卷积和聚合(池化)函数的正确组合。

[0029] 例如,高斯模糊是一种图像软化技术,其使用低通滤波器通过去除极端异常值来使不均匀的像素值平滑,从而使图像模糊。这种技术看起来对检测诸如图2B所示的圆形瑕疵有用。图3A至图3D分别示出在已应用高斯模糊滤波器之后处理图像220、240、260和280中的每一者从而产生图像320、340、360和380的结果。被选择作为高斯函数应用于图像的卷积矩阵或内核应当与缺陷或瑕疵的大小相关,例如,内核大小应当与缺陷的半径相关。如果存在小缺陷,小内核(诸如 2×2 或 3×3)将足够。如果缺陷较大,则内核应当更大;例如,如果缺陷大小为10,则应用 10×10 内核。缺陷的大小通常通过x、y像素计数(例如, 2×2 或 3×3)而不是缺陷的实际尺寸长度和宽度来测量。因此,如果相同的缺陷具有以不同缩放比获取的图像,则将相应地调整相关内核的大小。存在已被研究并表明对图像分析有效的许多众所周知的卷积矩阵,并因此,可以选择应当对捕获感兴趣缺陷有效的卷积函数。

[0030] 在这种情况下,如图4所示的 7×7 内核被应用于原始图像。期望的是,较亮图像被过滤掉,而较暗图像仍然保留,从而为区分不同的图像提供基础。例如,经修改图像340(从图2B中的图像240转换而来)中的瑕疵246在图3B中仍然显现为突出的黑圈。同样,最突出的小缺陷266仍然出现在图3C中。然而,在寻找像瑕疵246的更大的项目时,可过滤掉像缺陷266的小项目。

[0031] 在一个实施方案中,将多次应用卷积函数,因为该步骤可使某些特征显示得更好。例如,通过进行高斯模糊函数的多次卷积,较暗区域保持较暗,而其他区域变亮,从而实现可检测的信号差异,预测机器学习模型可将所述信号差异用作用于预测/检测此类型瑕疵的输入。

[0032] 由图2B和图3B中的瑕疵的示例引起的一个可能困难是它接近图像的边缘,这会产生信号检测的困难;并且事实上,多次卷积可能使这个问题变得更糟。

[0033] 现在参考图5A至图5D,分别示出在已应用边缘检测函数之后处理原始图像220、240、260、280中的每一者以生成图像520、540、560和580的结果。边缘检测算法将似乎对检测图2D所示的条纹图案286有意义。然而,当边缘可见但不是很明显时,那么可能仍然难以

检测到诸如在图5D的示例中的条纹图案。

[0034] 在这种情况下,将各种3x3内核(图6A至图6C所示)应用于原始图像220、240、260、280。图6C的内核似乎在这个特定示例中表现最好,产生了图5A至图5D的图像;但是根据被采样的图像的数量和质量以及要检测的问题,内核值的其他变化可能产生好的结果。另外,通过选择要在单个方向(x或y)上或在多个方向(x和y)上应用的内核,可获得不同的结果。

[0035] 在图7A至图7D中,已将锐化函数应用于原始图像,并且结果表明,在锐化函数之后条纹286更加明显,如图7D所示。

[0036] 可将来自图像处理和计算机视觉领域的许多其他函数应用于图像,以观察它们对检测半导体中各种不同类型缺陷的关键信号的影响,所述函数包括诸如Canny边缘检测、Scharr滤波器、Prewit滤波器、Sobel滤波器等的函数。特别地,已证明当Scharr滤波器产生较低的最大梯度值(即相对于相邻像素的像素差异的量度)时,Scharr滤波器在预测正常图像时是有用的。

[0037] 此外,可采用其他类型的技术来获得对训练模型有用的结果。代替其中将图像转换为另一个图像的卷积函数,可基于图像分析计算度量以将图像转换为作为表示图像的度量的单个数字。例如,香农熵是一种用于量化图像分布的技术,并且可用于定位纯随机缺陷。参见,例如,Pal,Nikhil R.和Sankar K.Pal,Entropy:A new definition and its applications,IEEE Transactions on Systems,Man and Cybernetics,第21卷,第5期,第1260-1270页(1991)。其他度量技术包括Hu的矩集,如果缺陷在图像中是对称的,则所述矩集可用于示出强矩体。参见,例如,Huang,Zhihu和Jinsong Leng,Analysis of Hu's Moment Invariants on Image Scaling and Rotation,2010 2nd International Conference On Computer Engineering and Technology,第7卷,第V7-476页(2010)。另一种技术是正交矩形多项式分解。参见,例如,Ye等人,Comparative Assessment of Orthogonal Polynomials for Wavefront Reconstruction over the Square Aperture,JOSA A 31,第10期,第2304-11页(2014)。

[0038] 在另一个示例中,图像的低维表示可用作模型(诸如离散余弦变换、奇异值分解等)的度量输入,包括表示来自图像处理的压缩误差的度量。

[0039] 度量可作为输入直接提供给机器学习模型,但仍然需要特征选择。卷积结果必须被压缩成一维表示,诸如向量,以便输入到机器学习模型并进行特征选择。

[0040] 卷积技术和度量技术中的任一种可根据需要单独使用或组合使用以对特定客户应用程序有效,以便做出在不同类别或种类的图像之间进行区分的决定。

[0041] 池化

[0042] 如上所述,在卷积之后应用池化函数,以将二维修改图像降为单维表示或向量。这可通过在每个方向上的各个点处(例如,在 $x=1$ 、 $x=2$ 等和 $y=1$ 、 $y=2$ 等处)获取经修改图像的统计量度来完成,所述统计量度包括但不限于均值、中值、最大值、最小值、标准偏差、百分位数等。经修改图像的向量的一个示例是x方向上的标准偏差值的集合,而同一经修改图像的另一个向量可以是y方向上的标准偏差值的集合。因此,向量的各个统计值中的每一者是“工程化的”特征或变量,如果被确定为重要,则可选择所述特征或变量,如下所讨论。

[0043] 不同于试图确定向量表示哪种特定类型或类别的常规多类分类器,本技术针对每对图像类型使用分类器,以确定向量是更像第一图像类型还是更像第二图像类型。结果被

取为单一数字,其表示目标图像是一种类型或另一种类型的可能性。然后将来自所有成对分类器的所有概率结果馈送到最终模型中,以做出最终的预测确定。

[0044] 建模

[0045] 分层建模允许特征选择关注对区分和分离每对类别贡献最显著的变量。另一方面,鉴于所需的成对分类器的数量与图像种类数量的平方成比例地增加,分层建模在计算上更加密集。图9示出用于分层建模的过程900的一个实施方案,所述过程仅由步骤902和步骤904组成,步骤902用于晶片图像的成对分类,步骤904用于预测每个成对图像比较的最可能类别。

[0046] 图10更详细地示出成对分类步骤902。在步骤1002中,针对每对类别构建单独的预测模型。在上面的示例中,存在四种不同的类别(正常、瑕疵、缺陷、条纹),并因此有六个不同的成对分类器:正常与瑕疵;正常与缺陷;正常与条纹;瑕疵与缺陷;瑕疵与条纹;以及缺陷与条纹。因此,在此示例中,从成对分类器计算了六个概率分数。在步骤1004中,对于每个成对分类器模型,从有限的图像子集中选择图像样本,例如,选择仅两个子集用于训练集。需注意,缺陷图像不应用于训练分类器来区分正常与瑕疵。在步骤1006中执行变量(特征)选择,并且在步骤1008中执行每个成对分类器的最终预测。

[0047] 在图11中更详细地示出了变量选择步骤1006的一个示例。首先,在步骤1102中,例如使用随机森林算法对针对每个图像在卷积和池化步骤中生成的特征(或度量)进行等级排序以确定变量重要性。通过使用k倍交叉验证(例如,20倍)构建每个模型并获取变量重要性的总和来获得该变量重要性。

[0048] 在步骤1104中执行前向变量选择。基于来自步骤1102的等级排序,可使用20倍交叉验证来运行每个模型,每次添加一个特征,如果所述特征提高了模型的准确度,则保留所述特征。也可使用其他度量,如FI分数或平均AUC(曲线下面积)。

[0049] 在步骤1106中执行后向变量选择以用于修剪所生成特征。每个模型再次运行,从最后一个到第一个每次去掉一个添加的特征;并且如果通过去除所述特征而降低了模型的准确度,则返回所述特征。通常,这个变量选择步骤1006将特征的数量减少到两个至四个。

[0050] 也可采用或尝试其他已知的变量选择技术,包括但不限于计算变量重要性并去掉低于阈值的所有变量的“贪婪”优化算法,或设计用于去除冗余变量的降维方法。

[0051] 在给定可用计算资源的情况下,图像种类的数量对于有效建模来说可能太大。如果是这样,则可以与成对图像比较相同的方式来执行晶片图像与相似晶片图像难题组的成对比较,以便确定晶片图像与哪个组更相似,而不是与哪个晶片图像更相似。然后,在可能的组内,晶片图像经历如上所讨论的成对图像分类以预测最终分类。

[0052] 成对分类器的准确度已如图12A至图12B所示进行展示。图12A中的图表1210包括六个ROC曲线,它们是四种不同图像类型A至D的六个成对分类器的结果。曲线图1211是A-B分类器的结果;曲线图1212是A-C分类器的结果;曲线图1213是A-D分类器的结果;曲线图1214是B-C分类器的结果;曲线图1215是B-D分类器的结果;并且曲线图1216是C-D分类器的结果。

[0053] 从曲线图1215可看出,B-D对在该示例中是最难分类的,具有最小的曲线下面积和大量的可能假阳性,而A-B对的曲线图1211具有非常大的曲线下面积,但也迅速向上到尖峰,指示非常少的假阳性,并因此是有效且可靠的分类器。

[0054] 类似地,图12B的图表1220包括六个ROC曲线,它们是四种不同图像类型(瑕疵、缺陷、正常和条纹)的六个成对分类器的结果。曲线图1221是瑕疵-缺陷分类器的结果;曲线图1222是瑕疵-正常分类器的结果;曲线图1223是瑕疵-条纹分类器的结果;曲线图1224是缺陷-正常分类器的结果;曲线图1225是缺陷-条纹分类器的结果;并且曲线图1226是正常-条纹分类器的结果。

[0055] 从曲线图1222可看出,在图12B的示例中,根据最小的曲线下面积和大量的可能假阳性,瑕疵-正常对是最难分类的。图12B中的其余曲线图指示非常少的假阳性,并且特别地,缺陷-正常对的曲线图1224显然指示没有假阳性,并因此在缺陷图像与正常图像之间分类时是100%有效的。

[0056] 一旦特征已被剔除至每个模型的最重要贡献特征,通过使用20倍交叉验证和随机森林技术,仅使用重要贡献特征作为模型的输入变量来运行每个模型以减少过拟合,来确定对每个成对建模步骤的最终预测(图10,步骤1008)。在此步骤中,对所有模型的所有晶片进行预测。例如,用于预测缺陷图像的模型使用“正常与瑕疵”分类器来在寻找缺陷时帮助排除那些可能的可能性。

[0057] 一旦已确定每个成对建模步骤的最终预测,就构建二级模型(图9,步骤904)以基于来自成对分类器的结果预测每个模型的最可能类别。每个成对模型的预测概率是这个二级模型的输入,并且应当由 $N * (N-1) / 2$ 个不同的概率值组成。最终的分类器可以是任何分类算法,包括逻辑回归、支持向量机(SVM)、随机森林、梯度提升机、XGBoost、人工神经网络等。可选地,可通过使用N倍分层交叉验证来实现参数调整。

[0058] 分层方法的替代方案是直接建模,直接建模尝试直接使用输入参数计算每个类别的概率。可利用在分层模型中使用的特征选择和参数调整来提高直接模型性能。如果存在足够的训练数据和种类,则这种方法是有意义的。一种已表明以这种方式表现良好的模型是卷积神经网络。然而,在半导体行业中获得足够的准确数据很困难,因此决定了分层建模方案的发展。

[0059] 图13A和图13B示出对应于图12A和图12B中针对成对分类示出的示例的最终图像分类的结果。对于本文所描述的技术的两个示例,预测模型的准确率为大约95%。这些结果也表明成对分类并不完美。然而,许多图像也可能包含多个缺陷,从而增加了分类的复杂性。可通过包括置信度分数来提供预测的更多置信度。

[0060] 前述书面描述旨在使普通技术人员能够制作和使用本文所描述的技术,但是普通技术人员将理解所述描述不是限制性的并且还将理解存在本文所描述的具体实施方案、方法和示例的变化、组合和等同物。

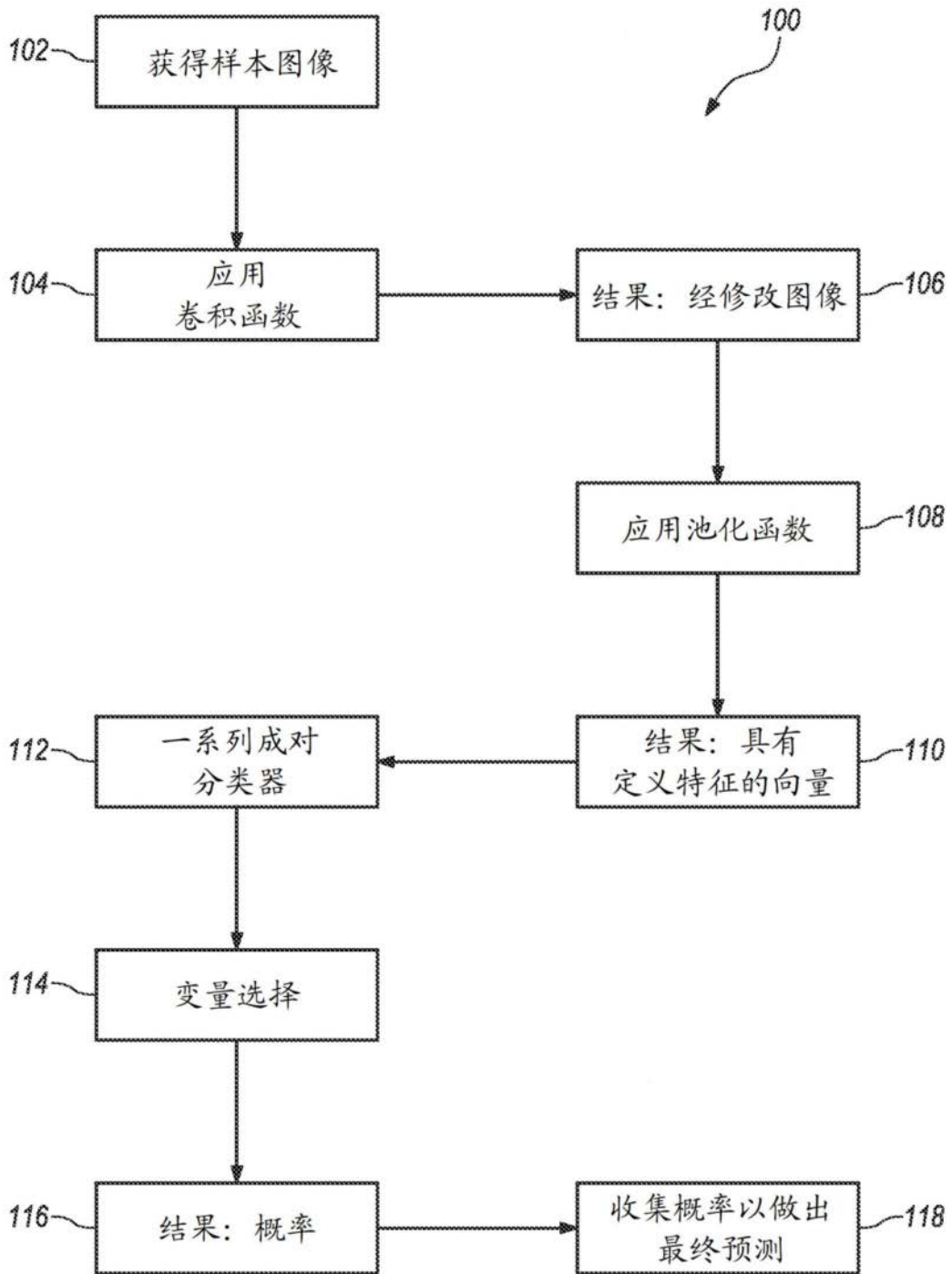


图1

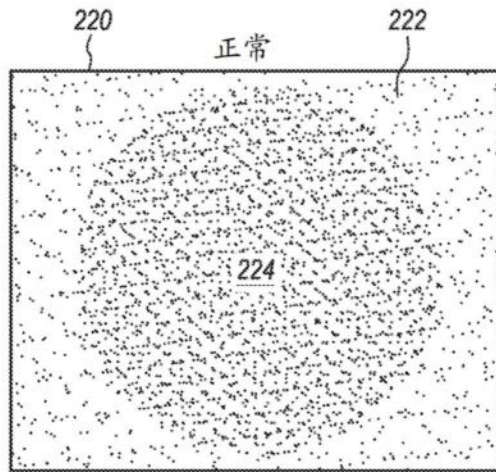


图2A

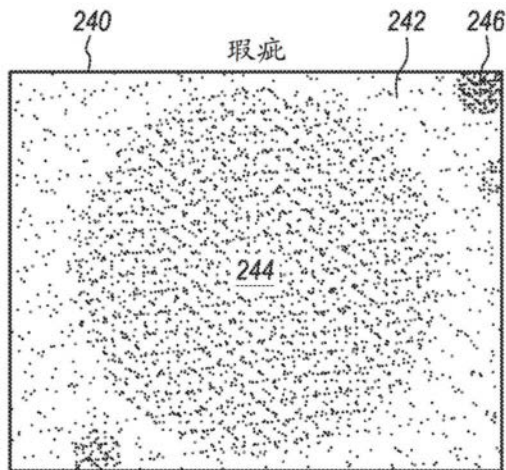


图2B

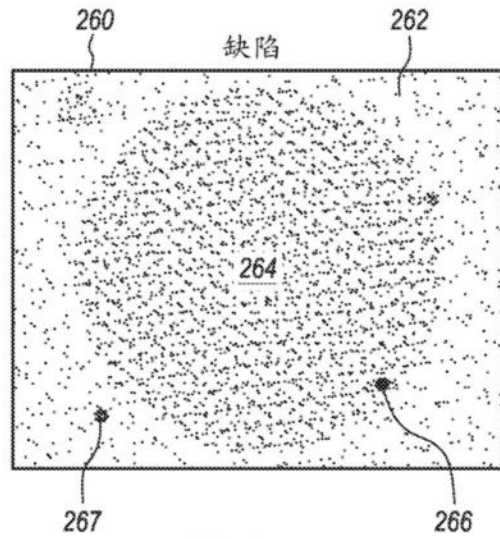


图2C

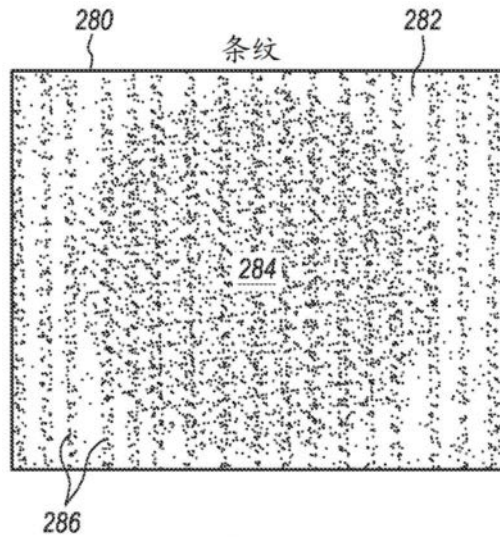


图2D

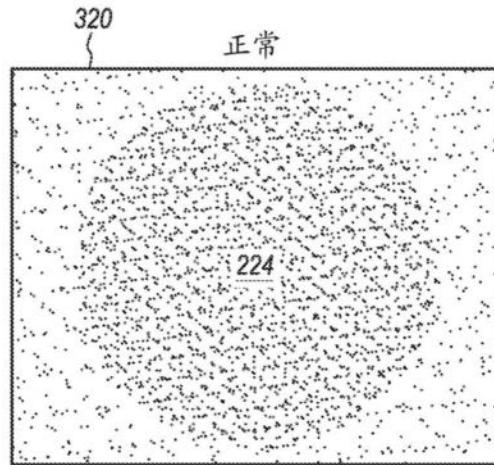


图3A

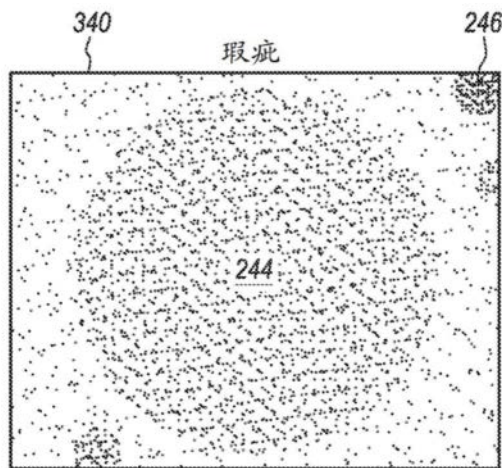


图3B

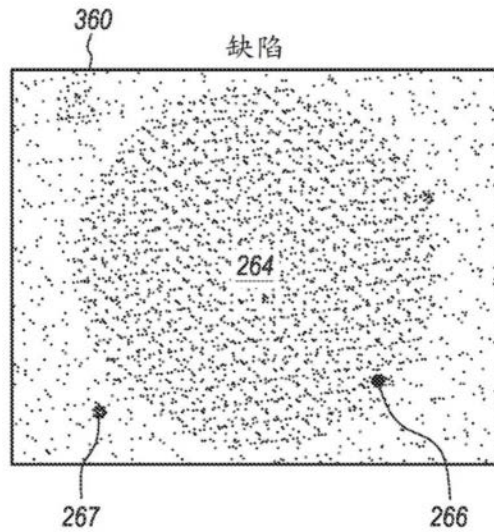


图3C

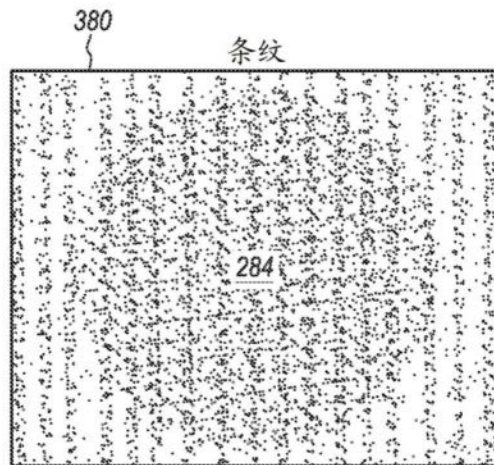


图3D

| | | | | | | |
|-------------------|------------|-------------------|-------------------|-------------------|------------|-------------------|
| 0.00000067 | 0.00002292 | 0.00019117 | 0.00038771 | 0.00019117 | 0.00002292 | 0.00000067 |
| 0.00002292 | 0.00078633 | 0.00655965 | 0.01330373 | 0.00655965 | 0.00078633 | 0.00002292 |
| 0.00019117 | 0.00655965 | 0.05472157 | 0.11098164 | 0.05472157 | 0.00655965 | 0.00019117 |
| 0.00038771 | 0.01330373 | 0.11098164 | 0.22508352 | 0.11098164 | 0.01330373 | 0.00038771 |
| 0.00019117 | 0.00655965 | 0.05472157 | 0.11098164 | 0.05472157 | 0.00655965 | 0.00019117 |
| 0.00002292 | 0.00078633 | 0.00655965 | 0.01330373 | 0.00655965 | 0.00078633 | 0.00002292 |
| 0.00000067 | 0.00002292 | 0.00019117 | 0.00038771 | 0.00019117 | 0.00002292 | 0.00000067 |

图4

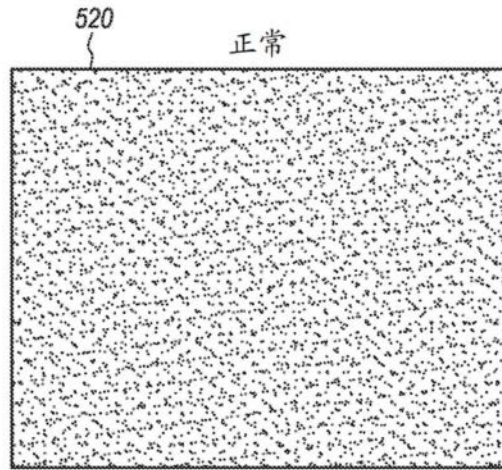


图5A

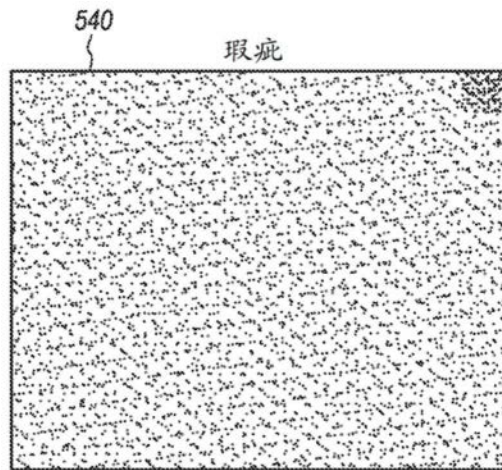


图5B

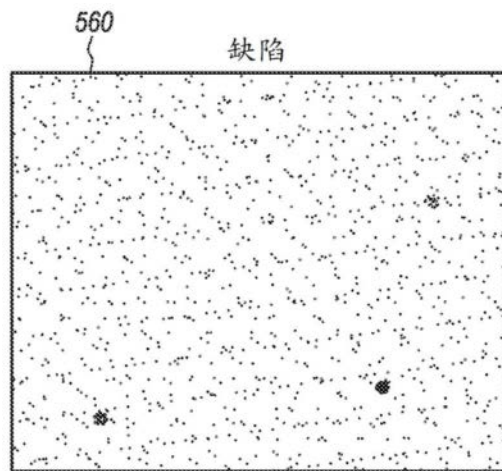


图5C

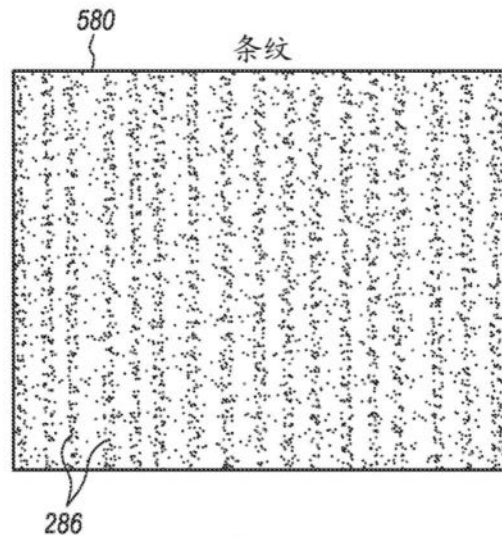


图5D

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

图6A

$$\begin{bmatrix} 0 & 0 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

图6B

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

图6C

正常

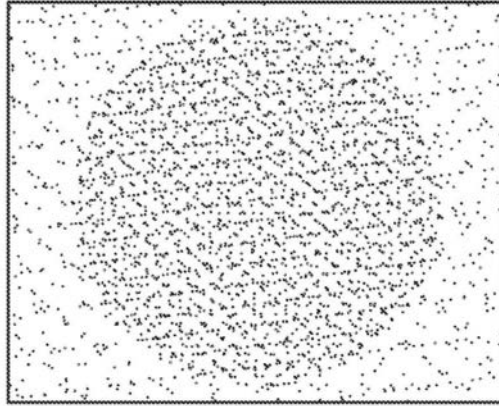


图7A

瑕疵

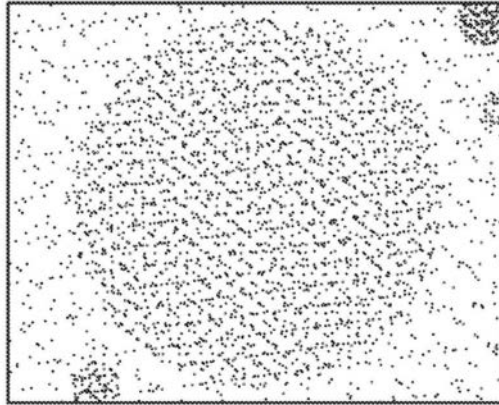


图7B

缺陷

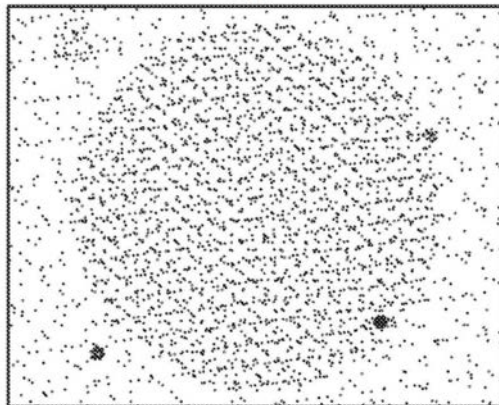


图7C

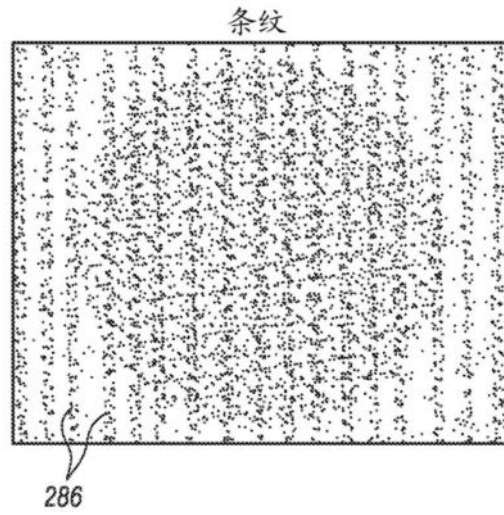


图7D

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

图8



图9

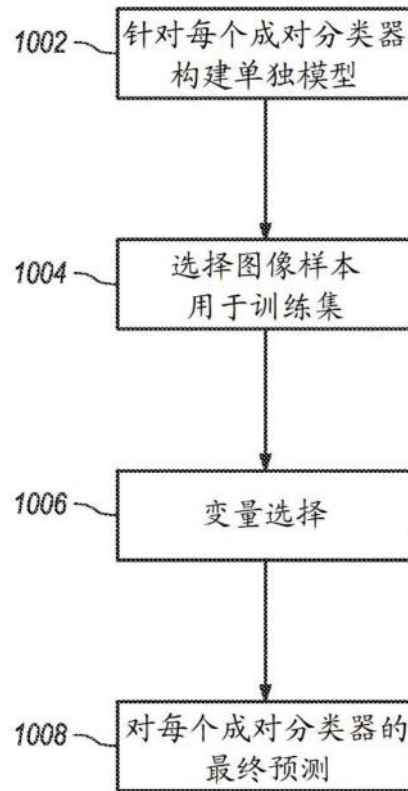


图10

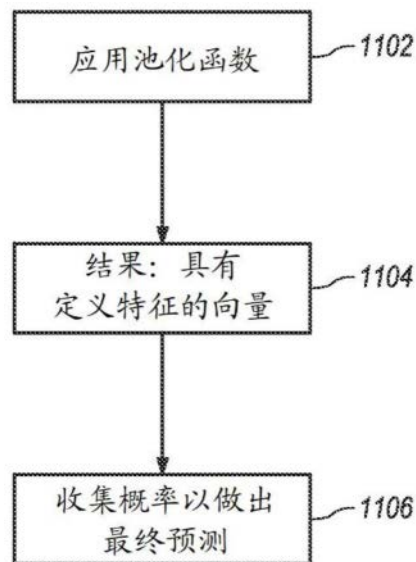


图11

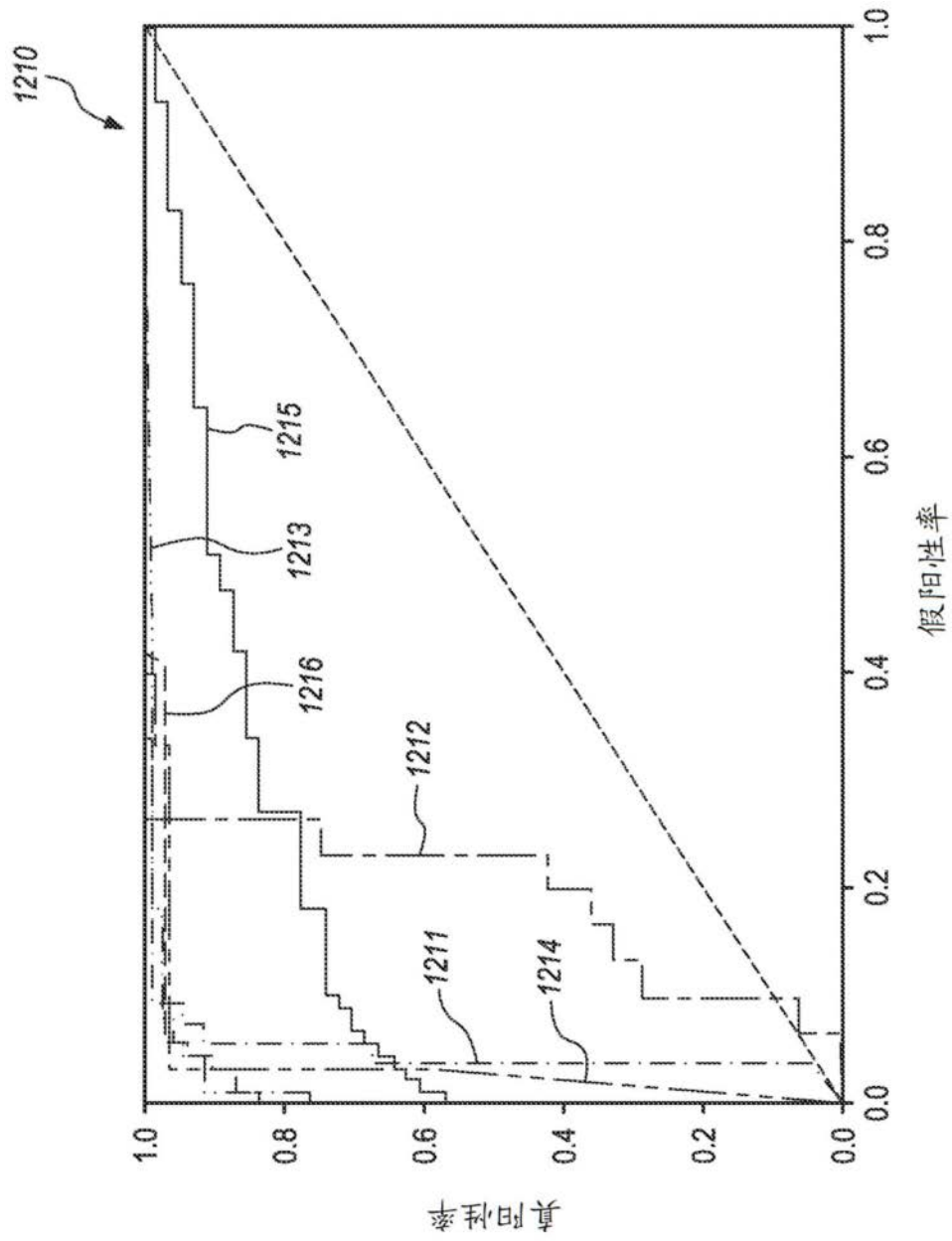


图12A

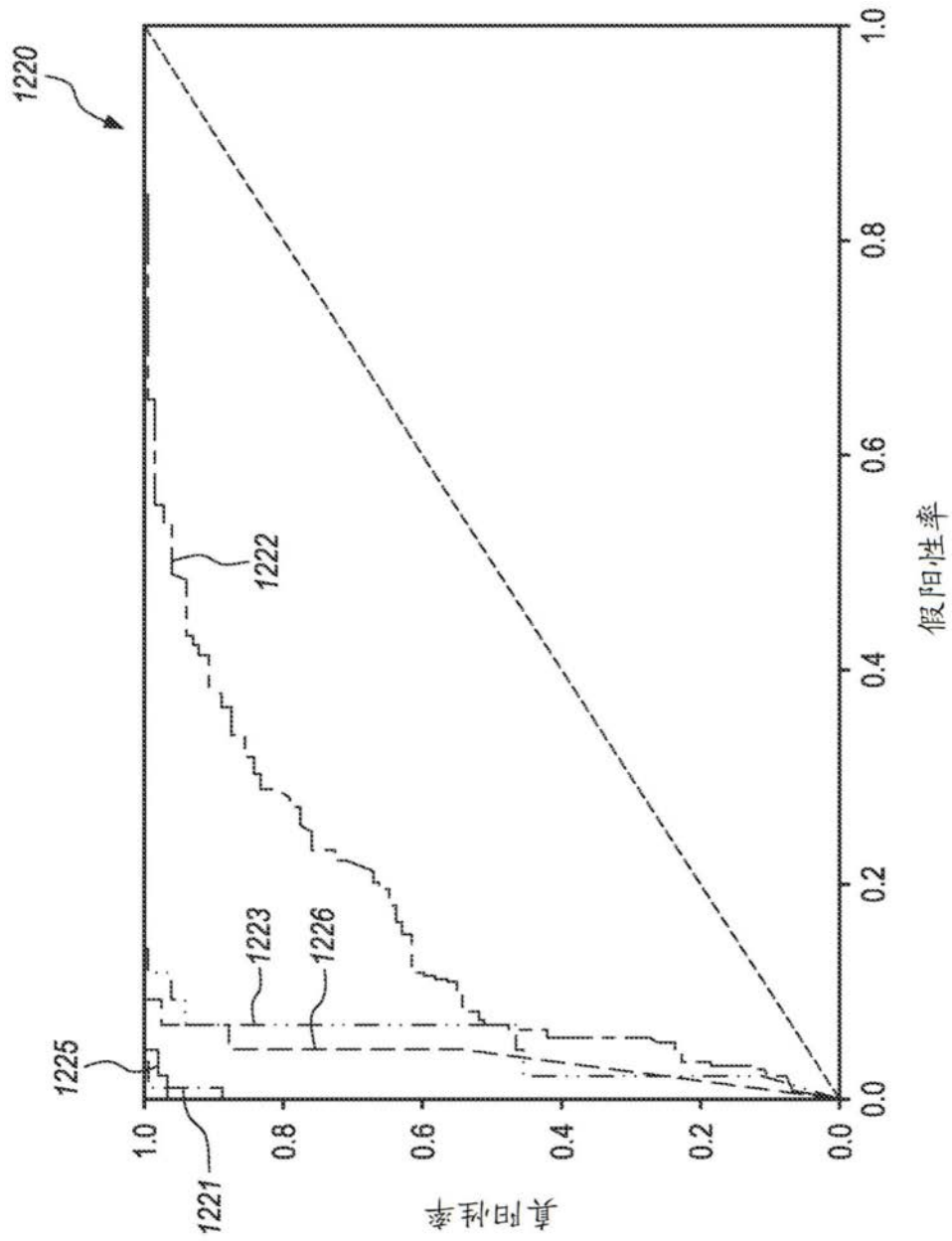


图12B

| | | 预测 | | | |
|----|---|-----|----------|----|----|
| | | A | B | C | D |
| 实际 | A | 28 | 0 | 0 | 2 |
| | B | 0 | 48 | 1 | 4 |
| | C | 0 | 0 | 67 | 0 |
| | D | 0 | 3 | 0 | 85 |
| | | 准确度 | 0.957983 | | |

图13A

| | | 预测 | | | |
|----|----|-----|----------|----|-----|
| | | 缺陷 | 条纹 | 瑕疵 | 正常 |
| 实际 | 缺陷 | 77 | 0 | 0 | 2 |
| | 条纹 | 1 | 40 | 0 | 1 |
| | 瑕疵 | 0 | 0 | 78 | 15 |
| | 正常 | 0 | 0 | 11 | 300 |
| | | 准确度 | 0.942857 | | |

图13B