基于机器学习的城市区域双极化天线GPS信号接收条件分类

Sanghyun Kim延世大学融合技术学院韩国仁川
sanghyun.kim@yonsei.ac.kr

Jiwon Seo延世大学融合技术学院韩国仁川
jiwon.seo@yonsei.ac.kr

arXiv:2305.03956v1 [cs.LG] 6 May 2023

摘要—**在城市区域，密集的建筑物经常遮挡并反射全球定位系统（GPS）信号，导致能够接收到的可见卫星数量较少，同时存在大量多径信号。这是导致城市区域定位不可靠的一个重大问题。如果能够检测出某颗卫星的信号接收条件，就可以通过排除受到多路径污染的卫星信号或降低其权重来提升定位性能。因此，我们利用双极化天线开发了一种基于机器学习的GPS信号接收条件分类方法。我们采用决策树算法进行分类，使用三个特征，其中一个特征只能从双极化天线获取。我们利用从不同地点采集的GPS信号训练机器学习模型。当将从GPS原始信号中提取的特征作为输入时，生成的机器学习模型会输出三种信号接收条件之一：仅非视距（NLOS）、仅视距（LOS）或LOS+NLOS。我们使用多个测试数据集分析分类准确率，并将其与采用双单极化天线的现有方法进行比较。因此，当测试数据集是在与训练数据集不同的地点采集时，分类准确率达到64.47%，略高于采用双单极化天线的现有方法的准确率。因此，双极化天线解决方案比双单极化天线解决方案更具优势，因为其外形尺寸更紧凑，且性能与另一种解决方案相近。**

索引词—**信号接收条件分类、全球定位系统、机器学习、双极化天线**

I. 引言

随着城市空中交通等新型城市交通方式的兴起，预计全球导航卫星系统（GNSS），尤其是全球定位系统（GPS）在城市区域的应用将显著增加 [1]–[9]。然而，在城市区域，GNSS 的多路径效应信号可能导致定位结果不可靠 [10]–[19]。这是由于城市密集建筑造成的两个主要问题。首先，建筑物可能阻挡视距（LOS）信号，从而大幅减少可见卫星的数量。其次，GNSS 信号可能被建筑物反射，导致接收到的非视距（NLOS）信号增加。第一个问题可以借助多个 GNSS 星座得到改善，但第二个问题仍然是一个重大挑战。

在城市区域，来自某颗 GNSS 卫星的信号可能在三种不同条件下被接收：仅接收到 LOS 信号（即，仅视距条件）、仅接收到 NLOS 信号（即，仅非视距条件），或同时接收到 LOS 和 NLOS 信号（即，LOS+NLOS 条件）。需要注意的是，仅非视距条件与 LOS+NLOS 条件被视为不同的现象，因为它们可能产生不同的测距误差 [20]。在 LOS+NLOS 条件下，由于相关处理中会抑制延迟超过 1.5 个码片的反射信号，测距误差是有界的 [21],；而在仅非视距条件下，测距误差则是无界的。如果检测到多径条件（即，仅非视距或 LOS+NLOS 条件），则可以通过排除或降低受多路径污染信号的权重来改善定位性能 [22], [23]。因此，人们一直在积极研究各种多路径检测方法 [24]–[26]。

一种方法使用鱼眼相机识别天空图中被建筑物遮挡的卫星，并将其判定为仅非视距条件 [27]。另有研究提出了一种利用三维城市模型生成天空图的方法 [28]。然而，这些方法的局限在于只能将信号分为两种情况：仅非视距和其他情况（即仅视距或视距+非视距）条件。另一种方法利用三维城市模型和射线追踪技术，对所有三种情况（即仅视距、视距+非视距和仅非视距）进行分类 [29]。尽管该方法能够对处于这三种情况下的卫星信号进行分类，但其局限在于计算量大，而且建立和加载三维城市模型需要投入大量工作 [20]。

此外，还有几种可用的统计方法

本研究得到韩国科学技术信息通信部资助的韩国国家研究基金会（NRF）和无人车辆先进研究中心（UVARC）“无人车辆核心技术研发计划”的支持（2020M3C1C1A01086407）。本研究还作为韩国海洋水产部资助的“综合 R-Mode 导航系统开发[PMS4440]”项目的一部分开展（20200450）。

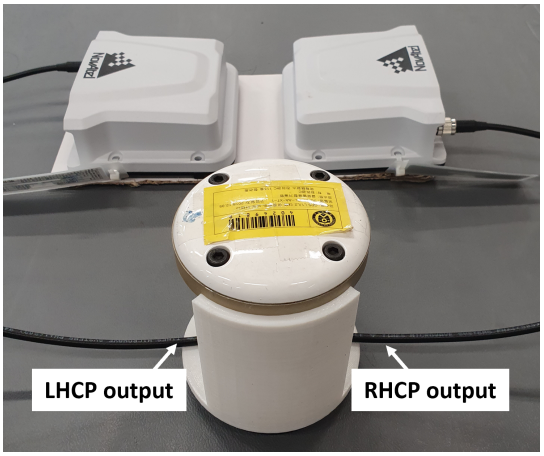


Fig. 1. Dual-连接了两个接收机的双极化天线 .

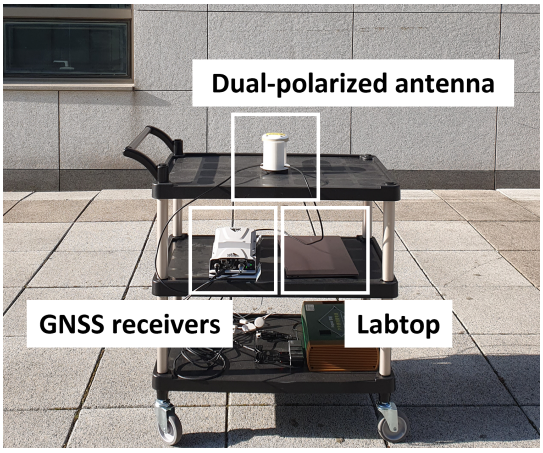


图 2. GPS 信号采集的硬件设置。

用于在测量层面检测多路径。一种此类方法通过比较三个频率下的载噪密度比（C/N_0）测量值来检测多径[30]。其原理是，在多路径环境中，频率之间的C/N_0 会发生变化。类似地，研究人员也使用了双极化天线的右旋圆极化（RHCP）信号与左旋圆极化（LHCP）信号之间的 C/N_0 差异[31], [32]。当 GNSS 信号发生反射时，其极化方式会从 RHCP 变为 LHCP。因此，可以通过观测 RHCP 信号与 LHCP 信号之间的 C/N_0 差异来检测多路径状态，该差异会受到多路径环境的影响。其他方法使用码减载波（CMC）和多普勒频移等测量值 [33], [34]。此外，越来越多的研究正在采用机器学习方法，以 GNSS 测量值作为机器学习特征来开发信号接收状态分类器[35]–[37]。

由于机器学习模型的性能会受到所用特征类型的显著影响，因此从 GNSS 测量值中提取并应用适当的特征至关重要。一种基于支持向量机（SVM）的机器学习分类方法将 GNSS 信号的相关输出用作特征之一 [38]。然而，该方法只能将信号分为两种状态：仅视距和多路径（即仅非视距或 LOS+NLOS）状态。还有一种方法使用梯度提升决策树（GBDT）算法和三个特征（即 C/N_0、伪距残差和卫星仰角） [39]，可将信号分为三种状态（即仅非视距、仅视距和 LOS+NLOS）。然而，在与训练环境不同的环境中进行测试时，其分类准确率约为 55%。另一种基于机器学习的信号接收分类方法比较了四种机器学习算法，包括 GBDT 算法 [40]。该方法利用五个特征，包括来自两副单极化天线的双差伪距残差，以提高分类性能。

表 I 各数据集的数据样本数量

GPS 数据集	T0[1]	T1[2]	T2[3]	T3[4]
总样本数	7500	7500	6311	11575
仅非视距样本	2500	2500	2038	3608
仅视距样本	2500	2500	2195	4033
LOS+NLOS样本	2500	2500	2078	3934

[1]在地点 A、B 和 C 收集的训练数据集 [2]在地点 A、B 和 C 收集的测试数据集 [3]在地点 D 收集的测试数据集 [4] 在地点 E 收集的测试数据集

在与训练环境不同的环境中进行测试时，分类准确率为 48%–63%。

本文提出一种基于机器学习的方法，使用双极化天线对GPS信号接收条件进行分类，并将其性能与使用双单极化天线的现有方法 [40]进行比较。由于外形尺寸更小，双极化天线方案比双单极化天线方案更具优势。使用双极化天线时，可以同时接收右旋圆极化和左旋圆极化信号，从而提取用于多路径检测的有效特征。我们使用了三个特征（即卫星仰角、右旋圆极化信号的 C/N_0 ，以及右旋圆极化信号与左旋圆极化信号之间的 C/N_0 差值），并采用决策树算法进行分类。在 [40]评估的四种算法中，决策树算法在我们的测试中表现最佳，因此我们选择了该算法。

II. GPS 信号采集与标注

A. GPS 信号采集

与单极化天线不同，我们用于 GPS 信号采集的双极化天线在同一个天线中同时具有右旋圆极化和左旋圆极化输出。如图 1 所示，通过在每个天线输出端连接接收机，即可采集右旋圆极化信号和左旋圆极化信号。

图 2 所示的硬件设置用于采集 GPS L1 信号，以生成机器学习训练和测试数据集。硬件设置由一副 Antcom 3G1215RL-AA-XT-1 双极化天线、两台 NovAtel PwrPak7 接收机和一台笔记本电脑组成。为与使用双单极化天线的现有方法进行比较， [40],GPS 信号采集在相同的五个地点进行，这些地点标记为 A、B、C、D 和 E，且均具有相似的多径环境。对于双极化天线和双单极化天线，训练和测试数据集均以相同方式创建。训练和测试数据集中的数据样本数量见表 I。

B. 真实信号接收状态的生成

为所收集的信号生成真实信号接收状态（即仅非视距、仅视距或 LOS+NLOS），对于准备用于监督学习的标注数据集是必要的。我们使用来自*3dbuildings* [41]的商业三维建筑地图，以及名为*Wireless InSite* [42], 的商业射线追踪软件，为每个卫星信号生成了真实信号接收状态。在射线追踪仿真中，我们假设所有三维建筑表面均由混凝土构成。这与 [40]采用的方法相同。最后，通过将生成的真实信号接收状态与从信号中提取的特征向量进行拼接，构建了用于监督学习的标注训练和测试数据集。所采用的特征将在下一节中进行说明。

III. 特征选择与决策树算法

A. 从双极化天线信号中选取的特征

本研究采用的特征包括卫星仰角、右旋圆极化信号的 C/N_0，以及右旋圆极化信号与左旋圆极化信号之间的 C/N_0 差值。Guermah 等人 [43] 提出了一种利用卫星仰角以及右旋圆极化信号与左旋圆极化信号之间的 C/N_0 差值进行分类的信号接收状态分类方法。为提高分类准确率，我们增加了一个特征（即右旋圆极化信号的 C/N_0 ）。这些特征具有以下特点。

- 卫星高度角：与高高度角卫星相比，低高度角卫星更容易受到多径影响，因为它更可能被建筑物遮挡或反射。

- C/N_0 右旋圆极化信号的： C/N_0 是接收载波（即信号）功率与噪声密度之比，其单位为 dB-Hz。通常，当信号发生反射时，由于存在额外的传播路径 [35]，会产生传播损耗；因此，仅视距条件下的 C/N_0值高于其他条件。如图 3 所示，仅视距条件下的 C/N_0范围为 36–51 dB-Hz，而仅非视距条件和 LOS+NLOS 条件下的该值则广泛分布于 25–51 dB-Hz 范围内。

- C/N_0 右旋圆极化和左旋圆极化信号之间的差异：右旋圆极化和左旋圆极化天线分量对

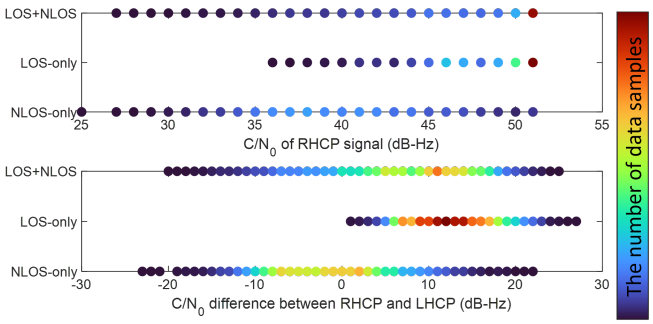


图 3. 从采集的数据样本中提取的右旋圆极化信号的 C/N_0 分布，以及右旋圆极化信号与左旋圆极化信号之间的 C/N_0 差值。

表 II 使用双极化天线的所提方法与使用双单极化天线的现有方法所采用的特征

双单极化天线 [40]	双极化天线
卫星仰角	卫星仰角
C/N_0	C/N_0 右旋圆极化信号的
时间差 C/N_0	C/N_0 右旋圆极化 以及左旋圆极化信号
增量之间的差异 伪距与伪距率	
双差伪距 残差	

分别为直达信号和多径信号。因此， C/N_0 RHCP 信号与 LHCP 信号之间的差值是检测多径条件的有效参数。图 3 显示，在仅视距条件下，所有 C/N_0 差值（即 RHCP C/N_0 – LHCP C/N_0）均为正值，而在仅非视距条件和 LOS+NLOS 条件下则存在负值。

表 II 总结了本研究所用的特征，并与 [40] 中的现有方法进行了比较。我们没有使用从双单极化天线信号中提取的五个特征，而只使用了从双极化天线信号中提取的三个特征。

B. 用于分类的决策树算法

我们使用决策树算法对信号接收条件进行分类。决策树算法是一种非常稳健且广为人知的分类与预测方法，因为它能够简化复杂的决策过程，使决策者能够更好地理解问题的解决方案[44]。目前有多种类型的决策树算法，例如迭代二分器3（ID3）、ID3 的扩展算法（C4.5）以及分类与回归树（CART）。本研究采用了 *scikit-learn* [45] 提供的基于 CART 的决策树算法的优化版本。此外，我们还使用了 *GridSearchCV* （包含在 *scikit-learn* 中）函数来进行超参数调优。

IV. 结果与讨论

使用表 I 中的 T0 训练数据集，我们基于决策树算法训练了一个机器学习模型。随后，我们使用 T1、T2 和 T3 测试数据集测试了该模型的性能。图 4 展示了每个测试数据集的总体分类准确率结果，其中 T1、T2 和 T3 的准确率分别为 86%、76% 和 58%。T1 的分类准确率最高，因为它由与训练数据集位于同一地点所采集的数据样本组成。另一方面，T2 和 T3 数据集采集自不同地点，因此与 T1 相比准确率较低。此外，T3 的准确率相对低于 T2，这似乎是由于两个数据集中的数据样本数量不同所致。如表 I 所示，T3 的数据样本数量多于 T0（即训练数据集），约为 T2 的两倍。通常，当使用远大于训练数据集的测试数据集时，可能会包含许多未经训练的特征域数据样本，从而导致性能下降 [46]。

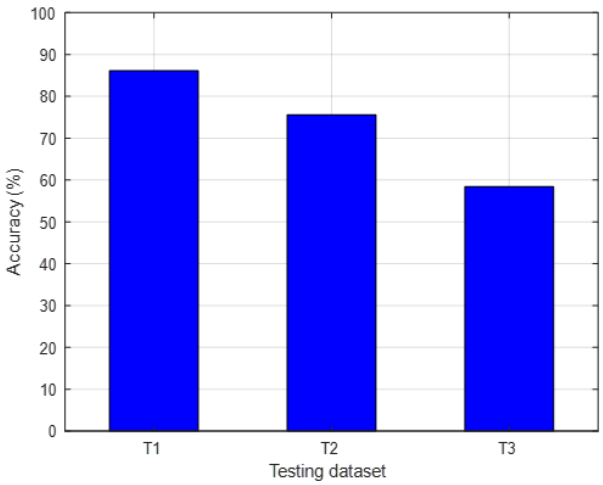


图 4. 三个测试数据集的总体分类准确率。

图 5 展示了仅非视距、仅视距以及视距+非视距条件下的分类准确率结果。我们训练的机器学习模型对仅视距条件的分类最为准确；然而，视距+非视距条件下的分类准确率相对较低。出现这些结果是可以预期的，因为视距+非视距环境比仅非视距或仅视距环境更加多样。因此，为提高视距+非视距条件下的分类准确率，应使用额外的视距+非视距数据样本进行训练。

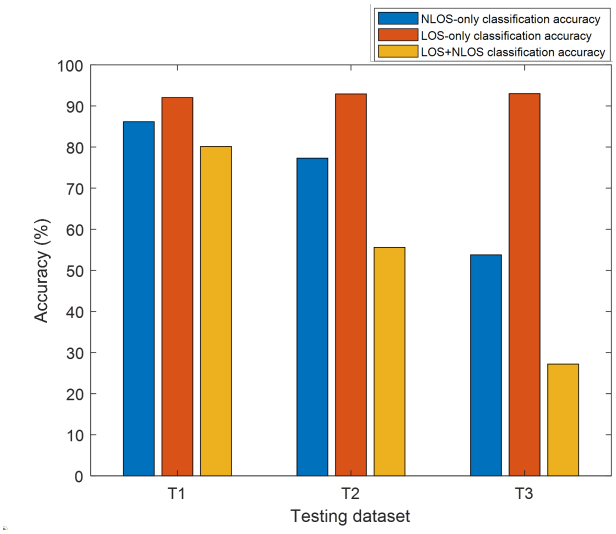


图 5. 三个测试数据集在不同信号接收条件下的分类准确率。

最后，我们将基于双极化天线的机器学习模型的性能与使用双单极化天线的现有方法进行了比较。我们比较了两项性能结果：分类准确率（T1）和分类准确率（T2+T3），如表 III 所示。前者是使用从与训练数据集相同位置采集的测试数据集 T1 得到的测试结果，后者是使用从不同位置采集的测试数据集 T2 和 T3 得到的测试结果。需要注意的是，相较于分类准确率（T1）的结果，分类准确率（T2+T3）的结果与实际应用中的预期性能更加相关。

如表 III 所示，分类准确率（T1）的结果相近，两种情况下均约为 86%。使用双极化天线的方法，其分类准确率（T2+T3）为 64.47%，表明其性能略优于使用两根单极化天线的方法。考虑到双极化天线方案的外形尺寸更小，我们可以得出结论：所提出的方法比现有方法更有优势。

表 III 双单极化天线与双极化天线的分类准确率比较

	分类准确率 (T1)	分类准确率 (T2+T3)
双单极化天线	86.81%	63.10%
双极化天线	86.12%	**64.47%**

V. 结论

在本研究中，我们开发了一种基于决策树的方法，利用双极化天线对GPS信号接收条件进行分类，并将其性能与采用双单极化天线的现有方法进行了比较。

使用在不同于训练数据集的地点采集的数据集进行测试，分类准确率为 64.47%，略高于现有方法的准确率（63.10%）。如果进一步采集分类准确率较低的 LOS+NLOS 条件数据并用于训练，预计性能还会有所提升。总之，由于还具有外形尺寸紧凑这一优势，双极化天线方案比双单极化天线方案更有利。

参考文献

[1] F. Causa 和 G. Fasano，“改善 GNSS 挑战环境中的导航：多无人机协作与广义精度衰减因子”，*IEEE* 航空航天与电子系统汇刊，第 57 卷，1479，2021 年第 3 期，页码 1462– 。[2] S. Kim、J. Park、J.-K. Yun 和 J. Seo，“通过强化学习在具有静态障碍物的虚拟开放空间中为无人飞行器进行运动规划”，载于 *ICCAS* 会议录，2020 年，第 784–787 页。[3] H. Yoon、H. Seok、C. Lim 和 B. Park，“在高仰角遮蔽区域提升无人机导航性能的在线 SBAS 服务”，传感器，第 20 卷，第 11 期，第 1–20 页，2020 年。[4] A. K. Sun、H. Chang、S. Pullen、H. Kil、J. Seo、Y. J. Morton 和 J. Lee，“深度信号衰落的基于马尔可夫链的随机建模：电离层闪烁条件下双频 GNSS 航空应用的可用性评估”，空间天气，第 19 卷，第 9 期，第 1–19 页，2021 年 9 月。[5] H. Lee、S. Pullen、J. Lee、B. Park、M. Yoon 和 J. Seo，“通过最优参数膨胀提升智能航空运输中单频 GBAS 的可用性”，*IEEE* 智能交通系统汇刊，第 23 卷，第 10 期，第 17 801–17 808 页，2022 年 10 月。[6] K. Sun、H. Chang、J. Lee、J. Seo、Y. Jade Morton 和 S. Pullen，“电离层闪烁条件下双频 GNSS 航空应用的性能优势：一种新的衰落过程建模方法”，载于 *ION ITM* 会议录，2020 年，第 889–899 页。[7] M. Jia、H. Lee、J. Khalife、Z. M. Kassas 和 J. Seo，“融合机会蜂窝信号的多星座 GNSS 地面车辆导航完整性监测”，载于 *IEEE ITSC* 会议录，2021 年，第 3978–3983 页。[8] H. Lee、J. Seo 和 Z. Kassas，“使用 GPS 和蜂窝信号进行城市自动驾驶车辆导航的基于完整性的路径规划策略”，载于 *ION GNSS* 会议录+，2020 年，第 2347–2357 页。[9] H. Lee、T. Kang 和 J. Seo，“用于城市区域基于 LTE 的无人机定位的安全距离可视化工具”，先进导航技术期刊，第 23 卷，第 5 期，第 408–414 页，2019 年。[10] H. Lee、J. Seo 和 Z. Kassas，“城市道路安全预测：卫星导航视角”，*IEEE* 智能交通系统杂志，第 14 卷，第 6 期，第 94–106 页，2022 年 11–12 月。[11] H. Lee 和 J. Seo，“多路径环境中基于机器学习、利用 LTE 信道冲激响应进行测距的初步研究”，载于 *IEEE ICCE-Asia* 会议录，2020 年。[12] W. Kim、P.-W. Son、S. G. Park、S. H. Park 和 J. Seo，“利用改进的 ASF 地图首次演示韩国 eLoran 在狭窄水道中的精度”，*IEEE* 航空航天与电子系统汇刊，第 58 卷，第 2 期，第 1492–1496 页，2022 年 4 月。[13] S. Lee、E. Kim 和 J. Seo，“通过数字预失真对 SFOL DME 脉冲整形以实现高精度 DME”，*IEEE* 航空航天与电子系统汇刊，第 58 卷，第 3 期，第 2616–2620 页，2022 年 6 月。[14] J. H. Rhee、S. Kim、P.-W. Son 和 J. Seo，“面向未来韩国全国性 eLoran 系统的增强精度模拟器”，*IEEE Access*，第 9 卷，第 115 042–115 052 页，2021 年 8 月。[15] K. Park、W. Kim 和 J. Seo，“初始姿态估计误差对松耦合智能手机 GPS/IMU 集成系统的影响”，载于 *ICCAS* 会议录，2020 年，第 800–803 页。[16] G. MacGougan、G. Lachapelle、R. Klukas、K. Siu、L. Garin、J. Shewfelt 和 G. Cox，“独立式高灵敏度接收机的性能分析”，*GPS Solutions*，第 6 卷，第 3 期，第 179–195 页，2002 年。[17] N. Shen、L. Chen、L. Wang、X. Lu、T. Tao、J. Yan 和 R. Chen，“基于坐标时间序列窗口匹配的特定地点实时 GPS 多路径抑制”，*GPS Solutions*，第 24 卷，14，2020 年第 3 期，页码 1– 。[18] Y. Lee 和 B. Park，“利用 DGNSS 和基于 CMC 的多路径抑制技术实现深度城市区域的无缝精确定位”，载于 ION GNSS+ 会议论文集，2020 年 9 月，第 1690–1719 页。[19] H. Lee 和 J. Seo，“用于目标移动设备定位的贪心算法与预测式无人机轨迹优化方法的性能评估及混合应用”，载于 ION ITM 会议论文集，2023 年 1 月，第 161–171 页。[20] N. Zhu、J. Marais、D. B´etaille 和 M. Berbineau，“城市环境中的 GNSS 定位完整性：文献综述”，《IEEE 智能交通系统汇刊》，第 19 卷，第 9 期，第 2762–2778 页，2018 年。[21] P. Misra 和 P. K. Enge，《全球定位系统：信号、测量值与性能》。Ganga-Jamuna Press，1994 年。[22] N. Kubo、K. Kobayashi 和 R. Furukawa，“利用连续时间序列 C/N0 检测 GNSS 多路径”，《Sensors》，第 25 卷，第 20 期，第 14 页，2020 年。[23] T. Suzuki 和 Y. Amano，“利用机器学习对 GNSS 信号相关输出进行 NLOS 多路径分类”，《Sensors》，第 21 卷，第 7 期，第 1–19 页，2021 年。[24] P. Closas 和 C. Fern´andez-Prades，“一种用于基于天线阵列的 GNSS 接收机的统计多路径检测器”，《IEEE 无线通信汇刊》，第 10 卷，第 3 期，第 916–929 页，2011 年。[25] L. Massarweh、M. Fortunato 和 C. Gioia，“利用 Xiaomi Mi 8 智能手机评估基于 Android 原始 GNSS 测量值的实时多路径检测”，载于 IEEE/ION PLANS 会议论文集，2020 年 4 月，第 1111–1122 页。[26] H. Lee 和 J. Seo，“用于基于接收信号强度的 LTE 信号室内 LOS/NLOS 分类的机器学习算法性能比较”，《定位、导航与授时期刊》，第 11 卷，第 4 期，第 361–368 页，2022 年。[27] X. Bai、W. Wen 和 L.-T. Hsu，“利用朝向天空的鱼眼相机和 LiDAR 辅助城市峡谷中的 GNSS 单点定位”，《IET 智能交通系统》，第 14 卷，第 8 期，第 908–918 页，2020 年。[28] L. Wang、P. D. Groves 和 M. Ziebart，“利用大型虚拟现实城市模型评估城市峡谷中的多星座 GNSS 性能”，《导航学报》，第 65 卷，第 459–476 页，2012 年。[29] S. Miura、S. Hisaka 和 S. Kamijo，“利用三维地图进行 GPS 多路径检测与校正”，载于 IEEE ITSC 会议论文集，2013 年 10 月，第 1–7 页。[30] P. R. R. Strode 和 P. D. Groves，“利用三频信噪比测量值检测 GNSS 多路径”，《GPS Solutions》，第 20 卷，第 399–412 页，2016 年。[31] P. D. Groves、Z. Jiang、B. Skelton、P. Cross、L. Lau、Y. Adane 和 I. Kale，“利用双极化天线实现的新型多路径抑制方法”，载于 ION GNSS 会议论文集，2010 年 9 月，第 140–151 页。[32] S. Kim、H. Lee 和 K. Park，“基于双极化天线载噪密度比测量值的 GPS 多路径检测”，载于 ICCAS 会议论文集，2021 年，第 1099–1103 页。[33] M. Caamano、O. G. Crespillo、D. Gerbeth 和 A. Grosch，“用于陆基应用的基于时间差分码减载波的 GNSS 多路径检测”，载于 ENC 会议论文集，2020 年 11 月，第 1–12 页。[34] L. Xu 和 J. Rife，“利用多普勒频移测量值检测 NLOS 和多路径”，载于 ION GNSS+ 会议论文集，2019 年 9 月，第 4064–4075 页。[35] L.-T. Hsu，“一种基于机器学习方法的 GNSS 多路径检测技术”，载于 IEEE ITSC 会议论文集，2017 年 10 月，第 1–6 页。[36] T. Suzuki、K. Kusama 和 Y. A mano，“利用卷积神经网络进行 NLOS 多路径检测”，载于 ION GNSS+ 会议论文集，2020 年 9 月，第 2989–3000 页。[37] R. Sun、L. Fu、G. Wang、Q. Cheng 和 L.-T. Hsu，“利用经过优化的自适应神经模糊推理系统与双极化 GPS 天线提高城市峡谷中的单点定位精度”，《导航学会导航期刊》，第 68 卷，第 1 期，第 41–60 页，2021 年。[38] T. Suzuki、Y. Nakano 和 Y. Amano，“利用机器学习检测城市环境中的 NLOS 多路径”，载于 ION GNSS+ 会议论文集，2017 年 9 月，第 3958–3967 页。[39] R. Sun、G. Wanga、W. Zhang、L.-T. Hsu 和 W. Y. Ochieng，“一种基于梯度提升决策树的 GPS 信号接收分类算法”，《应用软计算期刊》，第 86 卷，第 1–12 页，2020 年。[40] S. Kim、J. Byun 和 K. Park，“一种利用双天线的基于机器学习的 GPS 多路径检测方法”，载于 ASCC 会议论文集，2022 年 5 月，第 691–695 页。[41] 3Dbuildings。https://3dbuildings.com/。[42] Remcom。Wireless InSite。[43] B. Guermah、H. E. Ghazi、T. Sadiki 和 H. Guermah，“ 一种基于信息融合的鲁棒 GNSS LOS/ 多路径信号分类器，以及……”

智能交通系统的机器学习”，载于*IEEE ICTMOD* 会议论文集，2018年11月，第94–100页。[44] I. S. Damanik、A. P. Windarto、A. Wanto、Poningsih、S. R. Andani 和 W.Saputra，“使用遗传算法优化 C4.5 算法中的决策树”，载于*ICCSAM* 会议论文集，2018年，第1–6页。[45] F. Pedregosa、G. Varoquaux、A. Gramfort、V. Michel、B. Thirion、O.Grisel、M.Blondel、P. Prettenhofer、R. Weiss、V.Dubourg、J.Vander- plas、A. Passos、D. Cournapeau、M. Brucher、M. Perrot 和 E. Duches- nay，“Scikit-learn：Python 中的机器学习”，机器学习研究期刊，第12卷，第2825–2830页，2011年。[46] A. R´acz、D. Bajusz 和 K. H´eberger，“数据集大小和训练/测试集划分比例对定量构效关系/定量构性关系多类别分类的影响”，*Molecules*，第26卷，第4期，第1–16页，2021年。