



计算机应用研究
Application Research of Computers
ISSN 1001-3695, CN 51-1196/TP

《计算机应用研究》网络首发论文

题目: 基于随机森林算法的谷物粉分类及硬件加速
作者: 徐雷钧, 项厚友, 倪利华
DOI: 10.19734/j.issn.1001-3695.2021.09.0418
收稿日期: 2021-09-26
网络首发日期: 2021-12-08
引用格式: 徐雷钧, 项厚友, 倪利华. 基于随机森林算法的谷物粉分类及硬件加速 [J/OL]. 计算机应用研究. <https://doi.org/10.19734/j.issn.1001-3695.2021.09.0418>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

基于随机森林算法的谷物粉分类及硬件加速^{*}徐雷钧[†], 项厚友, 倪利华

(江苏大学 电气信息工程学院, 镇江 212013)

摘要: 针对传统谷物粉种类检测速度较慢的问题, 本研究基于 Zynq 平台实现随机森林算法辅助微波无损检测技术对谷物粉种类进行高效准确识别。通过对随机森林模型硬件实现的分析研究, 提出了一种改进模型参数结构, 有效节省了硬件存储资源的消耗。为了缩短算法预测时间并降低系统功耗, 在硬件实现时引入提前终止识别机制, 在保证准确率不变的前提下避免不必要的决策树预测过程。针对 Zedboard 开发板, 设计一种模型参数存储方案, 充分利用片上资源保证系统正常工作。实验结果表明, 与传统 CPU 实现随机森林算法相比, 该实现方案在 Zynq 上运行的实测时间缩短约 54.2%, 同时没有引起识别精度的损失。

关键词: 随机森林; Zynq; FPGA; 机器学习; 可重构计算

中图分类号: TP18 **doi:** 10.19734/j.issn.1001-3695.2021.09.0418

Grain flour classification and hardware acceleration based on random forest algorithm

Xu Leijun[†], Xiang Houyou, Ni Lihua

(School of Electrical Information Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract: In view of the slow detection speed of traditional grain flour species, this study implements the random forest algorithm-assisted microwave nondestructive testing technology to identify grain flour species efficiently and accurately based on Zynq platform. By analyzing the hardware implementation of the random forest model, an improved model parameter structure is proposed in this paper, which can effectively save the consumption of hardware storage resources. In order to shorten the prediction time of the algorithm and reduce the power consumption of the system, the early termination recognition mechanism is introduced in the hardware implementation to avoid the unnecessary decision tree prediction process on the premise of keeping the accuracy unchanged. In view of the Zedboard development board, a storage scheme for model parameters is designed to make full use of on-board resources and ensure the normal operation of the system. The experimental results show that the implementation time of the scheme with Zynq is about 54.2% shorter than that of the random forest algorithm with traditional CPU, and no loss of recognition accuracy is caused.

Key words: random forest; zynq; FPGA; machine learning; reconfigurable computing

0 引言

为了降低生产成本, 提高市场竞争力, 不法商家弄虚作假, “面粉中掺杂滑石粉”等侵犯消费者权益甚至危害生命健康的公共卫生事件屡见不鲜, 因此对谷物进行准确、高效分类识别具有重大的现实意义^[1-4]。微波检测以其无损、非接触、非污染、灵敏度高等优点, 已广泛应用于含水量检测、复合材料检测、医疗人体检测等领域。可以利用微波探测产生的幅值、相位信息等作为谷物粉分类的数据源^[5-8]。

分类算法是谷物粉识别的重要部分, 随机森林是一种集成学习方法, 对于异常值与缺失值有较强的容忍度, 通过平均决策树降低了过拟合的风险, 能够处理高维度数据且不需要做特征选择。由于优异的性能表现, 随机森林被广泛应用于计算机视觉领域、企业风险评估领域、金融危机预警等领域^[9-15]。算法在 CPU 上实现速度较慢, 无法满足实际应用中对于装置低时延的要求。随机森林算法内部决策树相互独立, 可以实现高度并行处理。相较于其他算法包含大量算数运算, 随机森林算法仅进行比较运算, 其在硬件实现时将消耗更少的资源与运算时间, 因此随机森林算法极其适合硬件实现。现场可编程门阵列(Field Programmable Gate Arrays, FPGA)具

有高度并行性、可重构性, 同时功耗低、体积小, 是机器学习算法硬件加速的一个热门研究领域^[16-22]。

Xiang Lin^[23]等人提出三种随机森林算法的硬件实现架构, 分别为存储器核心架构、比较器核心架构、综合核心架构。三种方案均存在不足, 存储核心架构数据量太大消耗过多存储资源, 比较器核心架构以及综合核心架构无法快速更新模型参数, 灵活性较差。郑小敏^[24]等完成了随机森林手势识别算法的嵌入式软件实现, 但是在硬件实现过程中为了节省存储资源的消耗, 其采用动态加载模型参数的方案, 这将消耗大量的时间用于模型参数的读取, 无法达到硬件加速的效果。

本文基于 Zynq 平台实现了随机森林的预测模型, 结合微波检测技术, 实现了对谷物粉类别的预测。该设计提出了一种精简模型参数结构, 降低了硬件实现时存储资源的消耗。同时, 该设计引入预测提前终止机制, 降低了预测的平均时间。整个系统能够脱离 PC 独立运行, 提高了装置的便携性与可移动性。

1 谷物粉分类系统介绍

本文采用的数据产生设备为所在课题组设计的微波透射

收稿日期: 2021-09-26; 修回日期: 2021-12-01 基金项目: 国家自然科学基金资助项目(61874050); 江苏省自然科学基金项目(SJ CX20_1403); 江苏省农业科技自主创新项目(CX(17)3001)

作者简介: 徐雷钧(1976-), 男(通信作者), 江苏海门人, 教授, 博导, 博士, 主要研究方向为太赫兹集成电路设计与谷物无损探测技术(xlking@ujs.edu.cn); 项厚友(1996-), 男, 安徽安庆人, 硕士, 主要研究方向为机器学习算法硬件加速实现; 倪利华(1995-), 男, 四川内江人, 硕士, 主要研究方向为微波无损检测。

扫频装置, 在该系统原型的基础上将识别算法功能移植到 FPGA 平台上, 实现算法硬件加速同时增加系统的灵活性。图 1 所示为本文谷物粉识别系统采用的实验装置。其中, 1 为微控制器, 2 为微波源, 3 为微带天线、载物盒与金属腔体, 4 为混频器, 5 为滤波器, 6 为幅相检测模块, 7 为 Zedboard, 8 为串口屏。

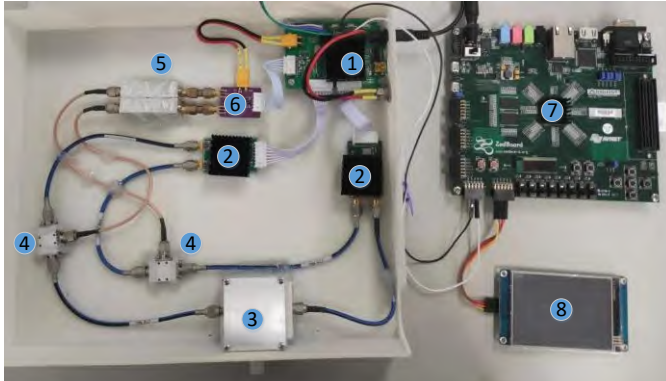


图 1 谷物粉分类系统

Fig. 1 Grain flour classification system

发射天线主动发射电磁波, 信号经过待测样本被接收天线采集。不同介电常数的待测样品对微波的吸收能力不一样, 从而接收天线采集到的信号会有幅值和相位的不同。本文以待测样本在不同频率下对应的幅值电压值作为数据特征, 特征数据经过分类器后完成谷物粉的预测。

实验装置以扫频模式进行特征数据采集, 起始频率设置为 2.2GHz, 终止频率为 4.4GHz, 步进 100MHz, 每组数据累计 150 个频率点幅值信息, 也就是 150 个特征值。本文选择面粉(类标签 1), 奶粉(类标签 2), 小苏打粉(类标签 3), 粘米粉(类标签 4), 滑石粉(类标签 5)作为待测样本。

2 随机森林算法

2.1 随机森林简介

随机森林(Random Forest, RF)算法于 2001 年由 Leo Breiman 和 Adele Cutler 首次提出, 其实质为采用集成学习(Ensemble Learning)的理念组合一系列彼此独立决策树, 是一种适用于回归与分类的算法。在随机森林的构建过程中, 利用 Bootstrap 方法重采样, 随机生成互异的训练集, 基于每个训练集构造一个决策树。内部节点进行分割时不再考察所有属性的基尼值(Gini), 而是随机选择若干属性进行对比。由于在决策树的生成过程中引入了两个随机性, 使得随机森林具有优异的抗噪声能力, 降低了过拟合的风险^[25-26]。随机森林算法的框架如图 2 所示。

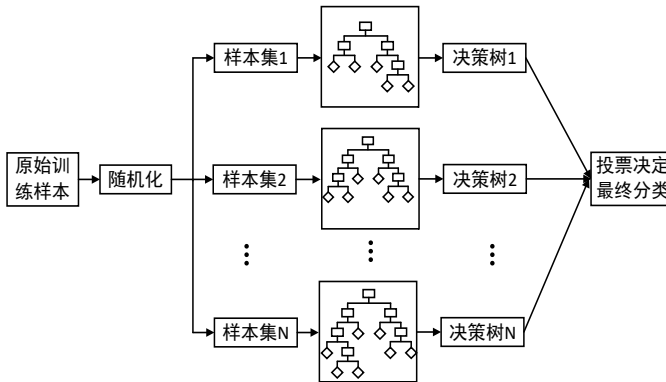


图 2 随机森林算法构架

Fig. 2 Framework of random forest algorithm

随机森林算法的预测基本流程如下:

a) 基于 Bootstrap 方法重采样, 随机生成 N 个训练集, 产生的训练集与原始训练集 P 样本总数相等。由于是有放回抽取, 产生的训练集中大概率包含重复样本, 保证了训练集之

间的差异性。

b) 利用训练集构建相对应的决策树 $K_1, K_2 \dots K_N$ 。在每一个内部节点选择分裂特征前, 从训练集中的 M 个特征中随机选取 $m(m \leq M)$ 个作为当前内部节点的分裂特征集, 选择对应基尼值(Gini)最小的特征作为分裂特征。每棵树都生长到底, 不进行剪枝操作。特征 a 对应的基尼值计算公式如下:

$$\text{Gini}(D') = 1 - \sum_{i=1}^{|D'|} p_i^2 \quad (1)$$

$$\text{Gini_index}(D, a) = \sum_{i=1}^V \frac{|D'_i|}{|D|} \text{Gini}(D'_i) \quad (2)$$

其中, D 指的是一个数据集, V 指的是在该数据集中特征 a 可能取值数目。

c) 对于测试集数据 X , 每个决策树给出独立的预测结果 $K_1(X), K_2(X) \dots K_N(X)$, 即投出自己的一票。

d) 统计每一个决策树的预测结果, 将所得票数最多的预测值作为最终的预测结果。

2.2 连续属性问题

连续属性问题, 是指属性的可取值数目为无限, 内部节点不能直接根据连续属性进行分裂^[27]。本文中的数据为不同频率下对应的幅值信息, 均为连续属性。解决连续属性问题的策略就是采用二分法对连续属性进行处理, 基本流程如下:

a) 连续属性 a 的取值从小到大排序, 设存在 n 个不同的取值, 记为 $\{a^1, a^2, a^3, \dots, a^n\}$ 。

b) 构建划分点集 $T_a = \left\{ \frac{a^i + a^{i+1}}{2} \mid 1 \leq i \leq n-1 \right\}$, 基于划分点 $t(t \in T_a)$ 将样本集 D 分为 D_t 和 D_t^+ 两个子集, D_t^+ 包含那些在属性 a 上取值大于 t 的样本, D_t 则为在属性值 a 上小于等于 t 的样本集合。

c) 计算每一个划分点对应的基尼值, 选择对应基尼值最小的划分点对样本集合进行划分。

2.3 随机森林模型超参数设定

用户自定义函数 T (决策树棵树)与 m (随机选择特征变量的个数)影响随机森林模型的性能, 因此, 需要选择合适的随机森林模型超参数保证预测的精度。通过比较不同 T 与 m 组合对应随机森林算法识别的准确度, 来确定最佳的 T 和 m 值。为了降低随机性的影响, 当 T 和 m 的值确定之后, 构建 100 个随机森林模型, 然后取其准确率的平均值, 作为当前确定参数下的准确率。图 3 展现了不同随机森林超参数组合对应的预测精度, 综合考虑模型精确度与运行时间, 设定决策树的数量为 150, 特征数量为 11。

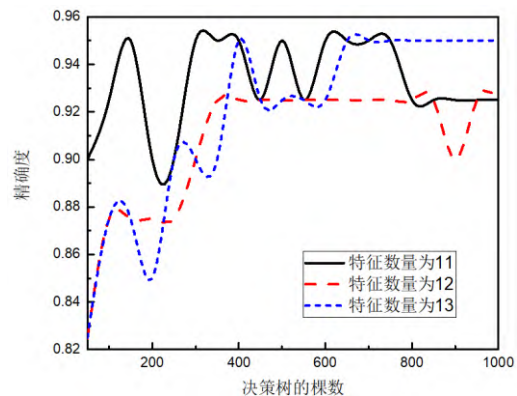


图 3 随机森林参数组合与预测精度

Fig. 3 Combination of random forest parameters and prediction accuracy

3 基于 Zynq 平台的谷物粉分类

随机森林单元总体设计如图 4 所示, Zynq 的 PS(ARM)

端首先将存储在 SD 卡中的模型参数写入 DDR, 接着通过 AXI_LITE 接口配置并启动 AXI_DMA IP 核, 自定义 IP 核 Randomforest 通过 AXI-STREAM 接口与 AXI_DMA IP 核连接, 接收模型参数用于构建随机森林模型。随机森林模型部署完毕后, 将待测样品置入微波探测平台, 启动装置, 生成的预测数据通过 Uart 接口与自定义 IP 核 Sample_Rx 连接, 该模块将解析得到的特征数据通过自定义接口传输到 Randomforest 核。经过随机森林模型运算得到预测类型, 该值通过自定义接口传输到 Screen_Drive 模块, 用于驱动串口屏显示预测类型。

随机森林算法的硬件实现需要消耗大量的存储资源 Bram, 而 Bram 资源分布在 FPGA 中的特定区域, 在系统分配时, 跟附近的 SLICE 距离较远, 导致路径延迟过大。此时, 需要降低时钟运行频率, 否则会导致时序违例, 引起系统异常工作。因此在设计随机森林算法硬件实现的时候, 需要充分考虑资源的使用, 平衡空间与时间。

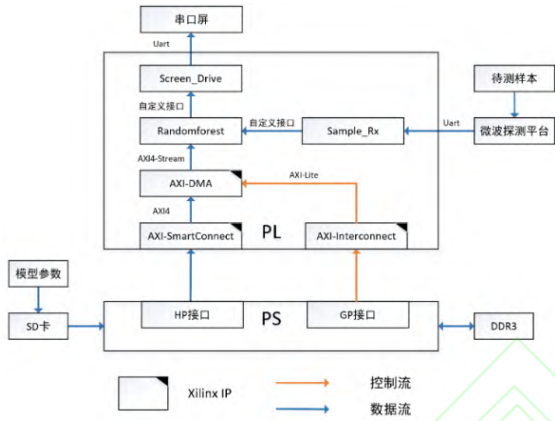


图 4 系统结构

Fig. 4 System structure

4 随机森林的硬件实现

4.1 随机森林模型参数结构

采用文献[23]提出的存储核心架构在 FPGA 上实现随机森林模型时, 需要将所有决策树转换成二叉树, 这种方法逻辑控制简单, 但是模型参数文件较大, 达到 2.89MB。硬件实现时需要消耗大量存储资源, 并且随着决策树层数的增加, 节点数将呈指数级递增, 相应的模型参数大小也将大幅增长。

为了减少 FPGA 存储资源的消耗, 降低 FPGA 布局布线难度, 提高系统运行频率, 需要最大限度的简化模型, 仅组合存储随机森林的最简有效信息。综合考虑硬件访问, 本文提出一种数据格式, 每个节点数据包含三个域, 如图 5 所示。

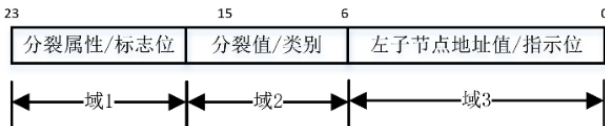


图 5 节点的数据结构

Fig. 5 Data structure of node

通过层次遍历决策树, 给每一个节点标记一个序号。对于决策树中的最后一个叶子节点, 域 1 取值为 0, 域 2 为对应的类标签, 域 3 取值为 1。对于其他叶子节点, 域 1 取值为 0, 域 2 为对应的类标签, 域 3 取值大于 1。对于内部节点, 域 1 取值大于 0, 域 2 为对应的分裂值, 域 3 取值大于 1, 为该内部节点左子节点的序号。第一个域是分裂属性与叶子节点标志位, 它的取值范围是(0~150), ($2^7 < 151 < 2^8$), 所以用 8 位二进制数存储该域值; 第二个域是分裂值或类标签(当域 1 取值为 0 时, 该节点为叶子节点, 该域为类标签; 否则该节点为内部节点, 该域为分裂值), 经校验最大分裂值为

433($2^8 < 151 < 2^9$), 需要 9 位二进制数来表示。第三个域为左子节点序号值与决策树最后一个叶子节点指示位, 经计算单棵决策树最大节点数为 69($2^6 < 69 < 2^7$), 该域值至少需要 7 位二进制数存储。故一个节点数据结构占 3 个字节, 节点数据的含义以及相应域特征如表 1 所示。

表 1 节点数据含义及相应域特征

Tab. 1 Node data meaning and corresponding domain characteristics

含义	特征
叶子节点(决策树中最后一个)	域 1 取值为 0, 域 3 取值为 1
叶子节点(非决策树中最后一个)	域 1 取值为 0, 域 3 取值大于 1
内部节点	域 1 取值大于 0, 域 3 取值大于 1

本文保留决策树的结构, 采用特定格式存储节点信息, 最终生成的模型参数文件大小为 84KB。对本文中的精简模型数据与文献[23]中的模型进行比较, 经验证, 在减少模型参数的同时, 并没有损失识别精度。但是会增加硬件逻辑控制的复杂性, 少量增加逻辑资源的消耗。

4.2 模型参数片上存储

Zedboard 开发板有 140 个 36K 的 Bram, 一个 36K 的 Bram 可以配置为两个独立的 18K Bram, 所以 Zedboard 含有 280 个 18K 的 Bram。本文训练好的随机森林模型拥有 150 棵决策树。单棵决策树的参数文件大小约为 2.16K, 理论上可以将每棵决策树模型参数存入一个 18K Bram, 150 棵决策树同时开始预测过程, 这种方案的预测速度最快。但是, 这种实现方案消耗大量 Bram, 会导致系统布局布线复杂, 引起 Bram 输出数据到下一级寄存器的路径延时巨大, 造成时序违例, 系统不能正常工作。同时, 这种实现方案需要消耗 77213 个 LUT, 数量已经超过 Zedboard 开发板上的 LUT 资源。

查阅 Xilinx 官方文档[28], Bram 有 16K x 1、8K x 2、4K x 4、2K x 9、1K x 18、512 x 36 六种原型方案, 所有的 Bram 配置方案均在这些原型的基础上叠加拼接得到。为了平衡时间与空间, 本设计在一个 18K 的 Bram 中存储 7 个决策树模型参数, 共计消耗 22 个 18K Bram。

决策树在 Bram 中的存储结构如图 6 所示, 单棵决策树的节点数据按照节点序号顺序存储, 每棵决策树的基地址以 69(单棵决策树的最多节点数)为单位递增。这种存储方式, 通过决策树的基地址结合节点序号, 硬件电路可以轻易访问任意决策树以及任意节点。



图 6 单个 18K Bram 存储结构

Fig. 6 Single 18K Bram storage structure

4.3 随机森林模型预测过程

考虑到测试数据量较小并且 Dram(分布式 RAM)有更好的时序表现, 将测试数据存储在 Dram 中。当测试数据采集完毕, 系统进入预测过程。22 棵决策树为一组, 并行开始预测过程。随机森林的预测过程流程图如图 7 所示。

首先读取 Bram 中基地址为 0, 偏移地址也为 0 的节点数据, 判断该节点是内部节点还是叶子节点, 如果是叶子节点, 该棵决策树预测结束, 预测结果为节点数据对应的类标签值; 如果该节点是内部节点, 在节点数据中分离出划分属性、划分值、左子节点序号值 M。在待测数据中取出对应划分属性的特征值, 将特征值与划分值进行比较, 如果特征值大于划分值, 那么下一个比较节点为左子节点(基地址为 0, 偏移地址为 M-1), 否则下一个比较节点为右子节点(基地址

为 0, 偏移地址为 M)。直到读取到 22 棵决策树的叶子节点, 一轮预测结束。

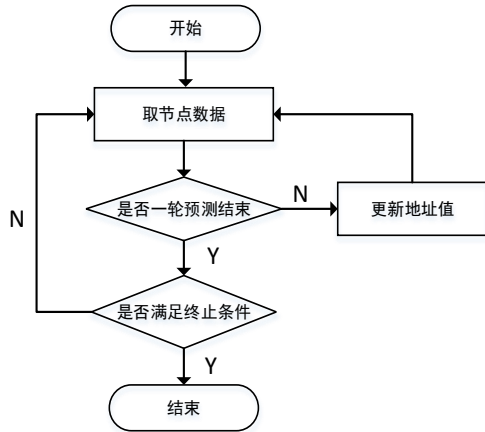


图 7 随机森林的预测过程流程

Fig. 7 Flow chart of prediction process of random forest

接着统计各棵决策树的预测值, 并且判断是否存在某一类别的被预测次数大于随机森林中决策树棵树的一半, 如果存在, 随机森林预测过程结束, 预测结果就是该类标签。例如有 78 棵决策树预测该待测数据属于标签 1 的类, 那么随

机森林预测过程提前终止, 预测结果为标签 1 的类。如果不存在, 则开启第二轮 22 棵决策树的预测, 总体流程与第一轮基本一致, 区别在于, 首先读取 Bram 中基地址为 69, 偏移地址为 0 的节点数据。依此类推, 第三轮的基地址为 138, 第四轮为 207, 第五轮为 276, 第六轮为 345, 第七轮为 414。如果提前终止条件没有被触发, 那么待测数据需要通过随机森林中每一棵决策树的判断, 也就是经历 7 轮预测过程, 才能给出最后的预测结果。

4.4 基于 Zynq 芯片开发流程

Zynq 内部集成了 PS(ARM)和 PL(FPGA)两部分, 与传统的 FPGA 开发不同, Zynq 开发流程除了 FPGA 部分的硬件逻辑设计, 还需要 ARM 部分的设备驱动程序开发。

本文 ZYNQ 芯片主要开发流程如下: 创建 VIVADO 工程, 添加 block design, 通过部署 IP 核构建硬件平台。在 block design 中添加 ZYNQ7 Processing System、AXI_DMA 等官方 IP 以及实现随机森林预测功能、测试数据采集功能、串口屏驱动功能三个自定义 IP, 并完成各个 IP 的配置, 硬件平台结构如图 8 所示。接着在 Vivado 里生成顶层 HDL 文件, 添加约束文件, 再编译生成比特流文件。最后导出硬件信息到 SDK 软件开发环境, 在 SDK 环境中编写驱动软件代码, 结合比特流文件调试 ZYNQ 系统。

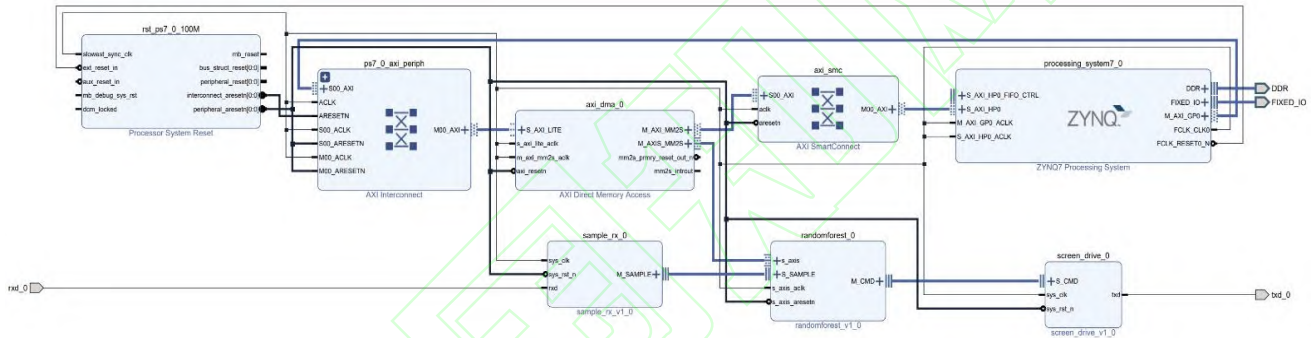


图 8 硬件平台结构

Fig. 8 Hardware platform structure

5 结果与总结

5.1 行为级描述仿真结果

构建随机森林算法模型, 用于谷物粉种类的识别。利用 Ultraedit 软件编写 Verilog 代码, 调用 Vivado 软件进行编译综合, 仿真采用的是 Modelsim 软件。输入面粉(类标签为 1)对应的测试数据作为仿真激励, 仿真结果如图 9 所示, predict_state 表示预测过程的状态机, m_predict_result 表示 150 棵决策树预测结果的统计值, result 表示随机森林最终的预测结果, result_vld 表示预测结果有效。如图所示, 随机森林模型正确预测出该谷物种类为面粉。

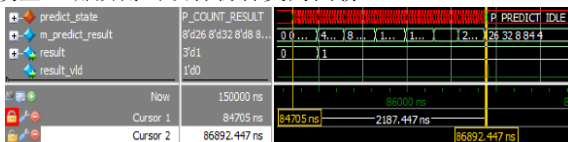


图 9 Modelsim 仿真图

Fig. 9 Modelsim simulation diagram

5.2 性能评估

5.2.1 识别精度

本文测试集为 50 组待测样本对应的特征数据, 每种待测样本对应十组数据。基于相同的训练集和测试集, 将本文采用的随机森林算法与其他主流分类算法比较, 同时比较随机森林算法软件实现与硬件实现对应的准确率, 结果如表 2 所示。

实验结果表明: 本文使用的随机森林算法识别的准确率

达到 98%, 明显高于其他分类算法。同时, 由于本文采用的训练集与预测集数据均经过定点化处理, 硬件实现没有引入误差, 所以随机森林算法硬件实现的准确率与软件实现一致, 没有造成精度损失。

表 2 基于不同算法与实现方式识别结果

Tab. 2 Recognition results based on different algorithms and implementation methods

所用算法	准确率/%
BP 神经网络	72.00
支持向量机	86.00
决策树	78.00
随机森林(软件实现)	98.00
随机森林(硬件实现)	98.00

5.2.2 片上资源利用率

本文基于 Xilinx Zynq-7000 平台的 Zedboard 开发板, 该平台 FPGA 芯片为 XC7Z020CLG484, FPGA 的时钟频率设定为 100MHZ。表 3 给出了算法模型实现后的 FPGA 硬件资源使用状况。

表 3 FPGA 硬件资源使用表

Tab. 3 FPGA hardware resource usage table

资源	使用资源	可利用资源	占比
LUT	13256	53200	24.92%
Filp-Flops	4318	106400	4.06%
BRAM(36K)	12	140	8.57%
LUTRAM	204	17400	1.17%

从表 3 中可以看出, 本设计在 Zynq 中实现的随机森林模型片内资源利用率处于较低的水平。

5.2.3 识别速度

表 4 是在 Zynq 与 CPU 两种平台上, 实现相同架构, 相同待测数据的预测时间的比较。Zynq 平台运行时间频率为 100MHZ 时钟频率。CPU 平台是在 11th Gen Intel(R) Core(TM) i7-11800H @ 2.30GHz, 16.0 GB 内存配置的 PC 上运行, 利用 MATLAB 2020b 软件完成随机森林预测过程时间测量。比较数据可以得出, ZYNQ 的计算效率相比与 CPU 实现提高了近 2.2 倍, 加速效果明显。

表 4 识别时间对比

Tab. 4 Identification time comparison

平台	计算时间
ZYNQ	2.2ms
CPU	4.8ms

5.3 板级验证

在 Vivado 20108.3 软件上对工程进行综合、实现、生成 bit 流文件以及硬件信息。该设计基于 AMD 公司的 ZYNQ 系列 FPGA 芯片 XC7Z020CLG484 进行验证, 并在 Zedboard 开发板上进行了板级验证, 输入为面粉(类标签为 1)对应的待测数据, 开发板连接串口屏显示随机森林模型的预测结果, 预测结果正确, 如图 10 所示。

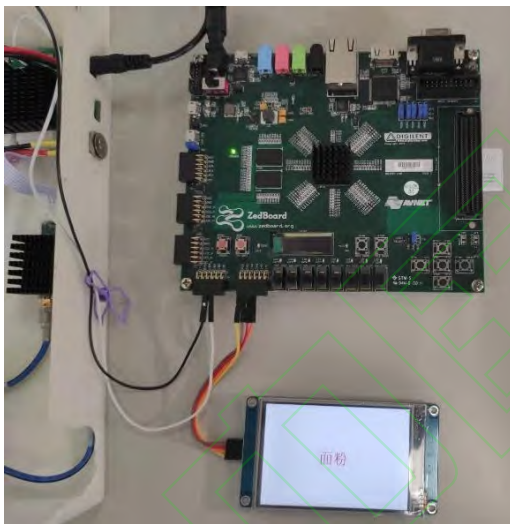


图 10 预测结果

Fig. 10 Prediction results

6 结束语

本文设计了一种基于 Zynq 平台的随机森林算法模型, 并联合微波探测平台应用于谷物分类领域, 增强设备的移动性, 减少预测时间。本文提出更精简的节点数据结构, 针对 Zedboard 开发板, 综合考虑性能、功耗、面积, 设计一种折中的存储空间分配方案。经过实验测试, 相较于 CPU, 基于 Zynq 实现随机森林预测模型, 预测时间减少了 54.2%, 同时预测精度完全一致。本设计可以推广应用于其他分类场景。

参考文献:

[1] 潘立刚, 张缙, 陆安祥, 等. 农产品质量无损检测技术研究进展与应用 [J]. 农业工程学报, 2008, 24 (S2): 325-330. (Pan Ligang, Zhang Jin, Lu Anxiang, *et al.* Research progress and application of nondestructive testing technology for agricultural product quality [J]. Journal of Agricultural Engineering, 2008, 24 (S2): 325-330.)

[2] Yu Xiaokang, Wang Zhiwen, Wang Yuhang, *et al.* Edge Detection of Agricultural Products Based on Morphologically Improved Canny Algorithm [J]. Mathematical Problems in Engineering, 2021, 2021:

[3] Zhu Min, Huang Dan, Hu Xinjun, *et al.* Application of hyperspectral technology in detection of agricultural products and food: A Review. [J]. Food science & nutrition, 2020, 8 (10):

[4] Xin Pingmo, Zhi Guopan. Study and Application of Machine Vision Technique for Quality Detection of Agricultural Products [J]. Advanced Materials Research, 2015, 3683:

[5] 杨玉娥, 何存富, 吴斌. 微波无损检测热障涂层下金属表面裂缝的参数优化 [J]. 复合材料学报, 2013, 30 (03): 149-153. (Yang Yue, He Cunfu, Wu Bin. Parameter optimization of microwave nondestructive testing of metal surface cracks under thermal barrier coatings [J]. Journal of composites, 2013, 30 (03): 149-153.)

[6] Li, Haigh, Soutis, *et al.* A review of microwave testing of glass fibre-reinforced polymer composites [J]. Nondestructive Testing and Evaluation, 2019, 34 (4):

[7] Jan B, Nadine S, Jyrgen E, *et al.* Minority Carrier Lifetime Measurements on 4H-SiC Epiwafers by Time-Resolved Photoluminescence and Microwave Detected Photoconductivity [J]. Materials Science Forum, 2019, 4682:

[8] A S, Kh S, J R, *et al.* Low power microwave tests on RF gun prototype of the Iranian Light Source Facility [J]. Iranian Journal of Physics Research, 2017, 17 (2):

[9] 李兵, 韩睿, 何怡刚, 等. 改进随机森林算法在电机轴承故障诊断中的应用 [J]. 中国电机工程学报, 2020, 40 (04): 1310-1319+1422. (Li Bing, Han Rui, He Yigang, *et al.* Application of improved random forest algorithm in motor bearing fault diagnosis [J]. Chinese Journal of Electrical Engineering, 2020, 40 (04): 1310-1319+1422.)

[10] 胡梦璐, 王佳, 张亚云, 等. 基于随机森林评价的兰州市主城区校园地表灰尘重金属污染 [J]. 环境科学, 2020, 41 (04): 1838-1846. (Hu Mengjun, Wang Jia, Zhang Yayun, *et al.* Heavy metal pollution of campus surface dust in the main urban area of Lanzhou Based on random forest assessment [J]. Environmental Science, 2020, 41 (04): 1838-1846.)

[11] 郭澎涛, 李茂芬, 罗微, 等. 基于多源环境变量和随机森林的橡胶园土壤全氮含量预测 [J]. 农业工程学报, 2015, 31 (05): 194-200+202+201. (Guo Pengtao, Li Maofen, Luo Wei, *et al.* Prediction of soil total nitrogen content in rubber plantation based on multi-source environmental variables and random forest [J]. Journal of Agricultural Engineering, 2015, 31 (05): 194-200+202+201.)

[12] Wall W A, Hohmann M G, Just M G, *et al.* Characterizing past fire occurrence in longleaf pine ecosystems with the Mid-Infrared Burn Index and a Random Forest classifier [J]. Forest Ecology and Management, 2021, 500:

[13] Kato D, Yoshitsugu K, Maeda N, *et al.* Positioning Error Calibration of Industrial Robots Based on Random Forest [J]. ijat, 2021, 15 (5):

[14] Piao Y, Jeong S, Park S, *et al.* Analysis of Land Use and Land Cover Change Using Time-Series Data and Random Forest in North Korea [J]. Remote Sensing, 2021, 13 (17):

[15] V R K, M R. B, J A M. Prediction of maximum surface settlement caused by earth pressure balance shield tunneling using random forest [J]. Journal of Artificial Intelligence and Data Mining, 2017, 5 (1):

[16] 王超, 王腾, 马翔, 等. 基于 FPGA 的机器学习硬件加速研究进展 [J]. 计算机学报, 2020, 43 (06): 1161-1182. (Wang Chao, Wang Teng, Ma Xiang, *et al.* Research progress of hardware acceleration of machine learning based on FPGA [J]. Journal of Computer Science, 2020, 43 (06): 1161-1182.)

[17] 孙孝辉, 宋庆增, 金光浩, 等. 基于 ARM+FPGA 平台的二值神经网络加速方法研究 [J]. 计算机应用研究, 2020, 37 (03): 779-783. (Sun Xiaohui, Song Qingzeng, Jin Guanghao, *et al.* Research on acceleration method of binary neural network based on ARM+FPGA platform [J].

- Computer Application Research, 2020, 37 (03): 779-783.)
- [18] 王晨, 王晓凯, 徐鹤丁. 基于 FPGA 的图像自动白平衡研究和实现 [J]. 计算机应用研究, 2020, 37 (S1): 346-347+350. (Wang Chen, Wang Xiaokai, Xu Heding. Research and implementation of image automatic white balance based on FPGA [J]. Computer Application Research, 2020, 37 (S1): 346-347+350.)
- [19] 闫小盼, 夏玉立, 赵军锁, 等. 遥感视频实时云检测的 FPGA 高效实现方法 [J]. 计算机应用研究, 2021, 38 (06): 1794-1799. (Yan Xiaopan, Xia Yuli, Zhao Junsuo, *et al.* FPGA efficient implementation method of remote sensing video real-time cloud detection [J]. Computer Application Research, 2021, 38 (06): 1794-1799. 2)
- [20] Choi Y, Jeong D, Lee M, *et al.* FPGA Implementation of the Range-Doppler Algorithm for Real-Time Synthetic Aperture Radar Imaging [J]. Electronics, 2021, 10 (17):
- [21] Chen Rui, Qiu Chengfeng, Liu Zhaojun, *et al.* P - 5. 4: Design and Implementation of Mini-LED Multi-Screen Splicing Display System Based on FPGA [J]. SID Symposium Digest of Technical Papers, 2021, 52:
- [22] Zhang Jinglong, Qin Xue, Zhao Mingyang, *et al.* FPGA implementation of a real-time digital pulse processing analysis for radiation detectors. [J]. Applied radiation and isotopes: including data, instrumentation and methods for use in agriculture, industry and medicine, 2021, 176:
- [23] Xiang L, R D S B, Donald E T. Random Forest Architectures on FPGA for Multiple Applications [P]. Great Lakes Symposium on VLSI 2017, 2017.
- [24] 郑小敏, 李翔宇. 随机森林手势识别算法的高效嵌入式软件实现 [J]. 计算机工程, 2021, 47 (07): 218-225. (Zheng Xiaomin, Li Xiangyu, Efficient embedded software implementation of random forest gesture recognition algorithm [J]. Computer Engineering, 2021, 47 (07): 218-225.)
- [25] 王鹏新, 齐璇, 李俐, 等. 基于随机森林回归的玉米单产估测 [J]. 农业机械学报, 2019, 50 (07): 237-245. (Wang Pengxin, Qi Xuan, Li Li, *et al.* Estimation of maize yield per unit area based on random forest regression [J]. Journal of Agricultural Machinery, 2019, 50 (07): 237-245.)
- [26] 陈元鹏, 罗明, 彭军还, 等. 基于网格搜索随机森林算法的工矿复垦区土地利用分类 [J]. 农业工程学报, 2017, 33 (14): 250-257+315. (Chen Yuanpeng, Luoming, Peng Junhuan, *et al.* Land use classification of industrial and Mining Reclamation Area Based on grid search random forest algorithm [J]. Journal of Agricultural Engineering, 2017, 33 (14): 250-257+315.)
- [27] 周志华. 机器学习 [M]. 北京: 清华大学出版社, 2016: 171-196. (Zhou Zhihua. Machine Learning [M]. Beijing: Tsinghua University Press, 2016: 171-196.)