融入专家决策过程的视频监控系统：检测牛分娩征兆的案例研究

Ryosuke Hyodo*、Susumu Saito*†、Teppei Nakano*†、Makoto Akabane*、Ryoichi Kasuga‡、Tetsuji Ogawa*
* 早稻田大学，日本东京
† Intelligent Framework Lab，日本东京
‡ Farmers Support，日本鹿儿岛

arXiv:2301.03926v1 [cs.HC] 10 Jan 2023

摘要——通过在畜牧业领域开展用户研究，我们验证了可解释人工智能框架对于视频监控系统的有效性。通过融入专家的决策过程，可以使系统具备可解释性。人工智能系统在现实应用中日益普及，尤其是在与人类决策相关的领域，而其可解释性至关重要。然而，目前针对现实应用中基于机器学习的系统，评估并解决其可解释性问题的标准方法仍然相对较少。本研究考察了一种视频监控人工智能系统框架，该系统融入了具备丰富领域知识的专家针对通知目标的决策过程，从而呈现预测背后的推理依据。一般的黑箱 人工智能系统只能呈现最终概率值，而所提框架能够提供与专家决策相关的信息，预计更有助于他们进行决策。在本案例研究中，我们基于所提框架设计了一个检测牛分娩征兆的系统，并通过对畜牧业相关人员开展用户研究（N=6）对系统进行了评估。与黑箱 人工智能系统进行比较后发现，许多参与者会参考所呈现的预测结果依据，且六名参与者中有五名选择了他们未来愿意使用的系统，即所提系统。研究表明，我们需要设计能够考虑预测结果依据的用户界面。

索引词-可解释人工智能、机器学习、用户研究、精准畜牧业

I. 引言

近年来，人工智能（AI）和机器学习（ML）已应用于与人类决策相关的领域，包括医学领域 [1]。在开发人工智能系统时，由于其简洁性，端到端 方法通常被视为一项不可或缺的技术。然而，有人指出，这种方法的内部行为是黑箱式的，不足以支持人类决策 [2]。与此同时，人们对人工智能系统的可解释性日益关注，并将其称为可解释人工智能（XAI） [3], [4]。可解释人工智能的代表性技术包括 CAM [5] 和 Grad-CAM [6]，等可视化技术，以及局部可解释模型无关解释（LIME） [7] 和 Shapley 加法解释（SHAP） [8] 等事后解释方法。然而，许多人工智能和机器学习方面的开发往往缺乏可用性，也缺少面向实际决策者的实用可解释性 [3], [4], [9]。

近年来，人机交互（HCI）领域已认识到以人为中心的评估的重要性，这种评估将用户评估纳入人工智能系统的可解释性研究中。利用实验数据集评估这些解释技术的用户研究日益增多 [10], [11], [12]。然而，在工业领域为高度专业化的任务实现人工智能系统时，前述通用解释技术以及基于实验数据集开展的一般用户评估所得结论，并不能提升系统对专家的可解释性。因此，高度专业化任务中的系统设计和实证实验至关重要；然而，针对工业应用的可解释人工智能框架及其通过用户研究进行验证的研究仍然不足 [13]。

本研究探讨了一种视频监控系统的框架，该系统将专家的决策过程融入神经网络，从而能够为专家的判断提供相关解释。专家观看视频，并基于其领域知识考虑各种属性后，判断是否存在异常。我们设计了一种神经网络，将这一过程融入异常检测中。具体而言，我们基于领域知识提取与异常相关的不同特征，并通过以提取的特征作为输入的各个流来判断这些特征是否异常。在此，明确提取专家用于作出判断的特征，使这些特征能够作为可解释信息呈现给专家

我们通过对畜牧农户开展用户研究，将上述框架应用于基于视频的牛产犊征兆检测（图1）。由于牛在产犊期间死亡会给农民造成重大损失， [14]，准确检测分娩征兆对畜牧管理至关重要。尽管直接安装在牛身上的接触式传感器已得到广泛使用，但从管理效率和动物福利的角度来看，采用摄像头这一非接触式传感器的检测系统具有重要价值。首先，为应用上述框架，我们回顾了动物研究相关文献，并通过对农民进行访谈，收集有关产犊征兆的领域知识。根据研究结果，我们设计了一个系统，明确提取与产犊相关的牛的姿势、旋转和运动的统计信息，并据此识别产犊征兆。我们利用特征提取器提供的信息设计了通知屏幕的用户界面。在用户界面中，

向农民展示牛的姿势频率、旋转量、运动量以及运动模式（俯视图）。与仅显示产犊征兆概率值的端到端系统用户界面相比，呈现系统内部状态有望让农民更容易理解系统。我们邀请参与畜牧业的人士（N=6）对端到端系统和所提出系统的用户界面进行了评估。我们希望研究结果能有助于开发以人为中心的可解释人工智能系统，使其有效融入专家知识。

本文其余部分安排如下。第二节简要介绍该框架，该框架使我们能够理解预测背后的推理过程。第三节介绍基于所提框架的畜牧业分娩检测系统案例研究。第四节讨论案例研究的发现以及所提框架的未来工作。第五节简要总结全文。

第二节 框架

所提框架将用户的决策过程融入预测网络，并支持依据其领域知识进行解释。典型的基于机器学习的系统采用端到端方法，利用专家标注对通知目标建模。相比之下，所提框架采用以人为中心的方法，通过用户访谈对通知目标建模。本节介绍所提框架设计流程的以下四个阶段：

- 与专家访谈
- 设计信息流，以提取用于表征监测对象的信息
- 整合多个信息流的信息，设计检测器
- 设计通知界面

第一阶段，我们就通知目标访谈专家。在访谈中，我们请他们描述检测目标时的决策过程，并确定他们在日常工作中关注的特征。随后，我们将这些特征细分为更具体的特征，使非专家也能够判断。

在下一阶段，我们从专家检测通知目标时关注的特征中提取更具体的特征，然后基于提取的特征开发流网络。将特征细分为更具体、即使非专家也能够判断的特征，便可加入众包标注。这对于确保监控系统的可持续性十分重要，因为这样我们无需依赖该领域的专家，就能增加训练数据的数量。

在第三阶段，我们基于不同的属性特征，开发一个整合多个流网络的网络。系统根据该网络的后验概率向用户发送通知。在此，即使对每个流网络的检测未必有效，只要各流网络的预测结果互为补充，就能提高检测效果。

最后，我们设计呈现给用户的用户界面。除了通知目标的最终后验概率外，还可以将从各个流网络中提取的多个特征作为预测结果的依据呈现出来。这些特征的设计考虑了用户访谈中的反馈，因此能够为用户的决策过程提供必要的信息。

III. 案例研究

本节将通过与畜牧农户开展用户研究，验证所提框架在检测牛产犊迹象方面的有效性。首先，我们介绍基于第二节所述框架的系统设计，然后介绍用户实验的流程、结果和发现。

A. 系统设计

本节介绍基于所提框架设计的产犊检测系统。我们将介绍有关分娩征兆的领域知识，以及能够明确提取与产犊相关特征的网络结构。最后，我们将说明用户界面如何向农民呈现预测结果的依据。

1) 视频中可观察到的分娩征兆：本节概述有关分娩征兆的领域知识。我们采访了农民，并查阅了动物科学文献，以识别与产犊相关的牛的行为。动物科学领域已对与产犊相关的姿势和行为变化进行了广泛研究。这部分对应于所提框架第一阶段的实施。

以下是可从图像中观察到的典型姿势类分娩征兆：

- **站立与卧姿之间的转换**：产犊前约两至六小时，姿势变化的次数（例如，在站立与卧姿之间转换）会变得更加频繁 [15], [16], [17] ，而产犊前两小时卧躺时间会增加 [17]。
- **抬尾**：产犊前约四至六小时，抬尾行为变得更加频繁 [17]，且产犊前尾部的位置会抬高 [18], [19]。

在视频中观察到以下基于行为的产犊征兆：

- **旋转和转向次数增加**：产犊前四小时可观察到具有特征性的行走模式（例如，旋转和转向），产犊前三小时变得更加频繁 [20]。
- **无目的行走时间增加**：产犊当天的行走时长增加 [16], [17]。无目的行走时间显然会在产犊前约140分钟增加 [19]。

2）系统架构：我们设计了一个多流网络 [21] ，根据先前研究中识别出的产犊征兆，提取有关姿势、旋转和运动的统计信息，并根据具体情况整合这三个流。本部分对应于所提框架第二和第三阶段的实现

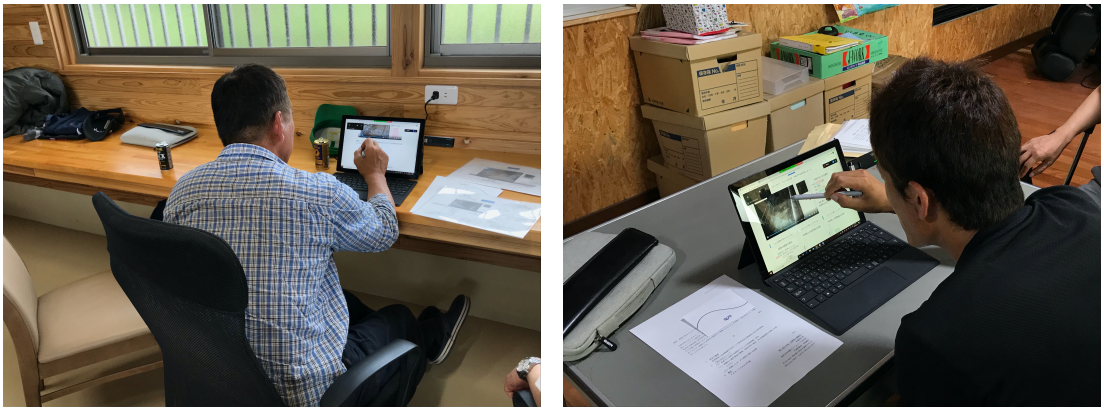

图1. 畜牧农户的用户研究。

。具体而言，每个流都根据以下特征，针对每段 30 分钟的输入视频识别产犊征兆：

- **基于姿势的特征**：使用 ResNet-50 [22] 对每个视频帧中牛站立、躺下和抬尾的状态进行捕捉，然后通过时间池化技术将其累积为相应的频率。

- **基于旋转的特征**：使用 ResNet-50 从每个视频帧中提取身体朝向信息，并通过测量身体朝向的变化，使用 M-measure [23], [24] 将其累积为反映牛旋转情况的统计数据。
- **基于运动的特征**：使用 YOLOv3 [25] 检测每个视频帧中的牛体区域，并将各帧之间位置的差异累积为反映牛运动情况的统计数据。

与产犊相关的特征旨在从非专家能够判断的信息中提取，例如姿势、颈部和尾部位置以及位置坐标。这样既可以通过众包收集数据，也使重新训练特征提取机制变得更加容易。

3）通知屏幕的用户界面：本节将介绍所提系统用户界面的设计。这部分对应所提框架第四阶段的实现。除了最终的产犊征兆后验概率外，所提框架还会为每一帧提供以下与产犊征兆相关的信息：

- 牛的姿势后验概率：1）尾巴抬起的站立牛，2）尾巴未抬起的站立牛，3）卧倒的牛，以及 4）无法判断

- 牛身体朝向的热图
- 牛的位置坐标

以通俗易懂的形式展示这些数据，将有助于农民估算牛何时开始产犊。

图2展示了根据上述信息设计的用户界面。除了监控视频和系统在此场景中的后验概率值外，界面还显示了姿势频率、旋转量、运动量以及从正上方观察到的牛的轨迹。右上方的条形图显示了系统根据30分钟的视频帧得出的产犊征兆后验概率。图中呈现了四个图表，分别根据姿势、旋转和运动的统计数据，显示后验概率及其均值。用户可以查看系统将哪些信息视为产犊征兆。姿势、旋转和运动的统计数据显示在屏幕底部。左侧的圆形图展示了这30分钟内各种姿势的频率，该频率根据特征提取器得到的姿势分类后验概率计算得出。此图可视为基于姿势的流程预测结果的依据。中间显示的旋转量与正常状态的比较，是根据特征提取器估算的牛身体朝向热图计算得出的。通过这一指标，可以量化牛在这30分钟内相较于其默认状态转动的幅度，并将其解读为预测结果的依据。在原型用户界面中，我们将身体朝向角度的时间序列变化可视化。采访一位农民后，我们采用条形图简化呈现方式。右侧显示的运动量与正常状态的比较，则根据牛的位置坐标变化量计算得出。与正常状态相比，牛在这30分钟内的运动量可以量化，并可解读为基于运动的流程预测结果的依据。最后，监控视频右侧显示牛的位置俯视图，呈现从房间正上方观察到的牛的位置轨迹。无需持续观看视频，用户也能了解牛整体的运动模式。

B. 用户研究

为评估所提框架，我们通过实验将其与端到端系统进行了比较。实验总时长约为60至90分钟，包括指导、练习、实验第1部分、第2部分以及实验后调查（图3）。首先，参与者观看了一段



图2. 所提系统（系统B）的用户界面。向参与者展示的用户界面为日语。

介绍实验的5分钟教学视频。随后，我们在实验开始前取得了参与者签署的书面知情同意，并向参与者发放了价值5,000日元（约合50美元）的礼品卡。作为练习，参与者分别回应了所提系统和端到端系统发出的两条通知。在本环节中，指导参与者出声思维 [26]，并通过主动评论实验过程中看到的内容和想到的想法，练习回答问题。实验分为两个部分，中间设有休息时间，参与者总共回应了36条通知。参与者被分为两个实验组，每组以伪随机顺序查看用户界面，以控制界面顺序产生的顺序效应。实验结束后，我们使用 Google 表单开展了定性实验后调查。

实验中，根据实验组以伪随机顺序向参与者展示了两个用户界面。用于比较的端到端系统的用户界面如图4所示。与所提系统的用户界面相比，端到端系统仅展示预测结果的后验概率图。对于相同的30分钟序列，端到端系统用户界面中展示的后验概率值，与所提系统用户界面中各数据流的平均后验概率值相同。为避免系统名称造成偏差，在向参与者展示用户界面时，我们将端到端系统称为系统A，将所提系统称为系统B。

通知单元为30分钟视频序列，共准备了18个序列。正例定义为产犊前3至0小时的时间段，负例定义为产犊前24至27小时的时间段。准备的序列中，真正例和假正例各有6个序列，假负例和真负例各有3个序列。换言之，我们假设通知阈值已显著降低，并包含一定数量的假正例。

参与者需要回答界面底部的以下两个问题（最多四个）：

Q1-1 你是否在这段 30 分钟的场景中识别出了分娩征兆？（-3：完全没有，.....3：完全识别出）Q1-2 你给出这个答案的原因是什么？（口头回答）

参与者需要针对**Q1-2**中的以下提示作答：1）你在屏幕上看到了什么？2）你在想什么？3）你是如何做出判断的？这种提问形式旨在尽可能充分地揭示他们的思维过程。回答上述问题后，如果**Q1-1** 的答案是“既不是”（0）或“有预测性”（1-3），则还需要回答以下问题：

Q2-1 看到这个用户界面后，你会采取什么行动？（1：立即开始协助，2：开始做准备，3：什么也不做）Q2-2 你给出这个答案的原因是什么？（口头回答）

Q2-1 的候选答案是指畜牧农户常见决策行为的术语。这些术语用于 Gyuonkei[1]，，这是一种在日本广泛使用的接触式分娩通知传感器。在这里，“立即开始协助”指立即开始协助产犊的决定，而“开始做准备”指在产犊前约 24 小时进行的准备工作。这些问题旨在鼓励参与者使用用户界面进行决策。每位参与者都需要

[1]Gyuonkei，http://www.gyuonkei.jp

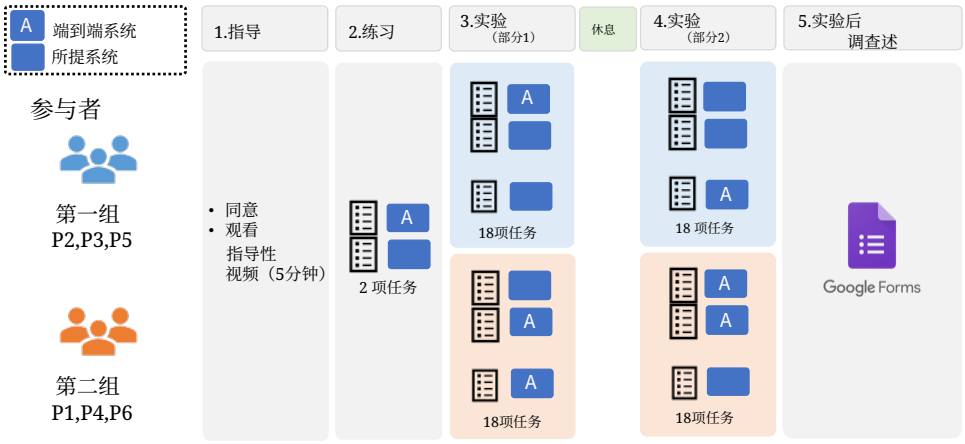

图3. 实验流程示意图。每个实验组的任务顺序各不相同。

针对上述问题作答，共涉及 36 条通知（2 个用户界面 × 18 个序列）。实验后调查的问题如下：

Q’1 您未来想使用哪个人工智能系统？（1. 系统A，2. 系统B）Q’2 您为什么选择该系统？（如愿意，可口头回答。）Q’3 系统A呈现的信息有多大用处？（-2. 毫无用处，……，2. 非常有用）a 分娩征兆概率图表。Q’4 系统B呈现的信息有多大用处？（-2. 毫无用处，……，2. 非常有用）a 分娩征兆概率图表。b 姿势频率图表。c 旋转量图表。d 运动量图表。e 俯视图（从正上方观察到的牛的位置）。Q’5 哪个系统提供了更准确的预测？（1. 系统A，2. 系统B，3. 两者都不是）Q’6 与另一个系统相比，您在**Q’1** 中选择的系统有哪些优点？（如愿意，可口头回答。）Q’7 您会对系统B做出哪些改进（如果有）？（如愿意，可口头回答。）

共有六位从事畜牧业的人士参与了用户实验。参与者的年龄、性别以及在畜牧业的工作年限如表I所示。参与者分为两组：农民（P1-4）和农业学院人员（P5-6）。其中，P5是教授，P6是动物科学专业的学生。

实验在计算机上进行，参与者访问了事先提供的实验页面网址。我们使用网络会议应用程序（Zoom）记录参与者的回答。为记录参与者在出声思维过程中的发言，我们通过Zoom在线录音，并在本地备份。

C. 结果与发现

参与者根据监控视频判断某种行为是否为产犊征兆，但他们的回答方式

表I 实验参与者详情。四位是农民（P1-P4），另外两位是在农业学院学习的人员（P5-P6）。

ID	组别	Age	Sex	工作年限
P1	II	30’s	M	10-19
P2	I	50’s	M	20-29
P3	I	20’s	F	5-9
P4	II	40’s	M	20-29
P5	I	40’s	M	20-29
P6	II	10’s	F	0-1

会因所呈现的用户界面类型而异。呈现端到端系统的界面时，大多数口头评论都与视频有关。然而，呈现所提系统的界面时，部分评论则与统计数据有关。一位参与者在谈及一张比较运动量与正常状况的图表时分享了自己的判断。P5 表示“我认为这头牛快要产犊了，因为数据显示它运动很多，而且旋转增加了。”。根据回答趋势，我们认为，呈现所提系统的内部状态有助于判断某种行为是否为产犊征兆。

实验后调查的回答结果如图5所示。六名参与者中有五名选择了所提系统，认为这是他们未来希望使用的系统（图5，Q’1）。他们认为，系统的内部状态有助于识别分娩征兆。P1表示“将通常凭感觉判断的运动量以图表形式呈现，更容易理解。”P2表示“能够看到有关牛的详细信息很有帮助。”P3表示“所呈现的图表有助于判断某种行为是否属于分娩征兆。”，P5表示“由于数据呈现得很详细，我认为经验较少的农民也能更有把握地做出决定。”

此外，部分评论符合这样的假设：端到端系统的用户界面是一个黑箱，不足以支持决策。P1 表示：“ *[*端到端系统的*]* 预测概率本身并不能告诉我们



图4. 端到端系统（系统A）的用户界面。向参与者展示的用户界面为日语。

该怎么做。我不知道该如何利用它。”。相比之下，P4 是本题中唯一选择端到端系统的人，他说：“我认为所提系统中的指标并不具有判断性，因为很多情况下，系统呈现的信息与我的认知不同。”P4 是一名农民，受邀协助访谈，以收集有关产犊征兆的领域知识。他表示，系统呈现的信息不如他预期的准确。其中一个不足之处是，本应在框架设计的第二阶段开展用户评估。由于我们在这一阶段没有收到专家的充分反馈，提取的特征并不总是足够有用的指标。此外，有人指出，主动呈现系统的内部状态（包括错误）可能会干扰判断，并导致参与者不信任系统。

接下来，我们讨论用户界面上的哪些信息有用（图5、Q’3、Q’4）。我们发现，参与者对分娩征兆概率趋势的看法各不相同。只有P4认为端到端系统提供的概率信息更有用，而其他参与者则认为所提系统更有用或两者差不多。现在，我们将注意力转向所提系统中各项统计数据的用户界面结果，这些统计数据涉及姿势、旋转和运动（图5、Q’4）。来自农业学院的两位参与者（P5、P6）认为姿势频率图表很有用，而其余四位农民则认为总体而言用处不大。这可能是因为农民可以查看原始数据，以便更准确地了解姿势频率，因此不太需要参考系统呈现的信息。另一方面，具有学术背景的农业学院参与者认为，能够以定量方式直观呈现姿势频率很有用。

更多参与者认为，运动量图表比姿势频率图表更有用，因为无法从原始数据中获取的元信息有助于判断。事实上，许多参与者表示，与正常状况进行比较对他们作答很有帮助。P1 表示“以前我都是凭直觉判断旋转量和运动量，所以用图表显示后更容易看出来。”P6 表示 “*[* 运动量图表 *]* 让人更容易注意到与正常状况相比变化的程度。” 旋转图表往往不如运动图表有用，因为参与者会关注牛的运动强度，并将其作为判断是否出现产犊征兆的标准之一，而旋转量与他们的判断并无直接关系。因此，以用户易于理解的方式呈现特征十分重要，而“运动强度”这类较为抽象的表达可能更合适。

在回答最后一个问题（Q’7）时，参与者指出了所呈现信息准确性方面的问题，以及显示信息时用户界面会缩小视频的问题。对于前一个问题，P4表示“我觉得图表与分娩征兆没有直接关联。”，P6则表示“人工智能在很多情况下都将姿势判定为无法确定，这让我感到担忧。在这种情况下，我更愿意自己判断。”对于后一个问题，P5表示，“可读性。我不希望视频太小。”

IV. 讨论

黑箱人工智能系统被认为不足以支持人类决策。一位农民指出，端到端系统没有指导他们在看到分娩征兆概率值后应该采取什么行动。几位农民指出了我们所提系统的优势，并表示，详细信息有助于新手农民自信地做出正确决策，而且理由可视化也很有帮助。尤其是，包含无法从原始数据中获取的元信息在内的统计信息往往特别有用。相反，我们发现，当呈现的信息有误时，系统的可靠性会受到不利影响。多项研究将这种现象称为算法厌恶 [27], [28], ，即用户在发现算法出错后便不再信任算法。为了向用户呈现更多信息，

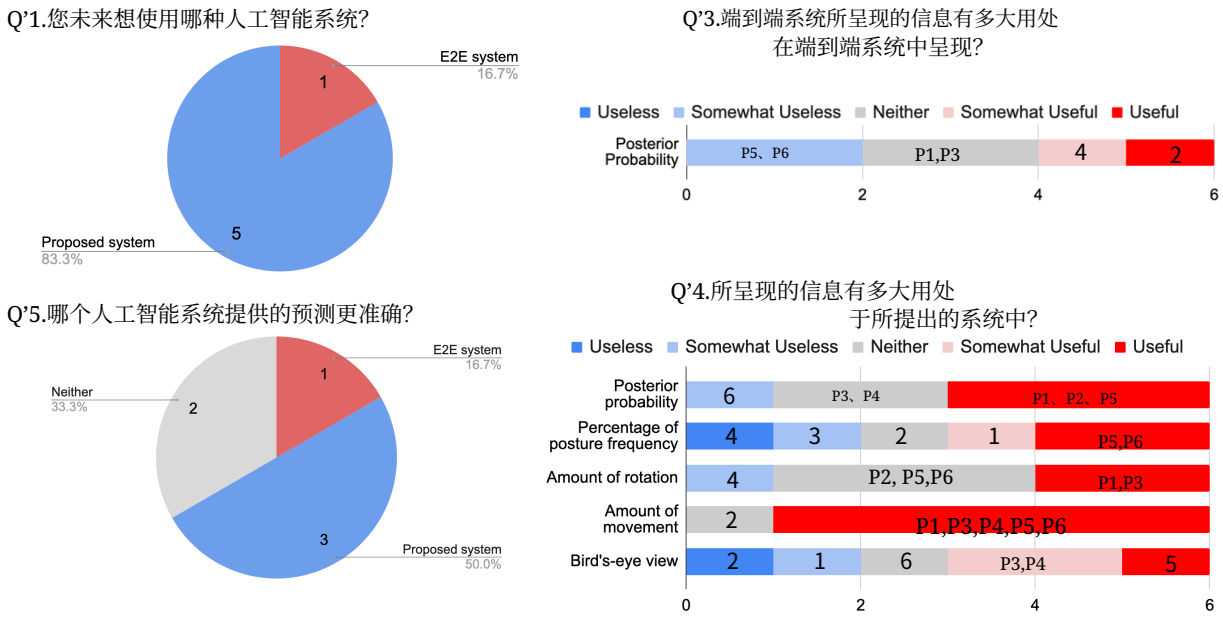

图5. 实验后调查中部分问题的回答。

所提框架需要将信息准确性纳入考量。此外，在框架设计的第二阶段，我们本应进行用户评估，因为提取的特征对用户而言可能并非足够有用的指标。参与者在反馈中指出，随着呈现的信息增多，用户界面也变得更加复杂。在考虑实际应用开发时，除了智能手机可能会让屏幕更难查看之外，还需要考虑老年用户。我们需要更精细的用户界面设计，例如将摘要界面与详细分析界面分开。我们预计，通过用户研究评估可解释人工智能框架，将有助于推动可解释人工智能在各类工业领域中的应用。

五、结论

本研究提出了一种视频监控系统框架，将专家决策过程融入系统架构，使预测结果的原因能够得到解释。我们通过一项针对畜牧业从业者的用户研究（N=6），评估了基于所提框架的产犊征兆检测系统。我们将所提框架与端到端系统进行了比较，六位参与者中有五位选择所提框架作为未来愿意使用的系统。此外，所提框架还用于呈现系统的内部状态，以帮助用户做出决策并识别系统错误。然而，我们发现，呈现不准确的系统内部状态可能会干扰用户判断，并导致他们不再信任该系统。未来，我们计划研究向用户呈现的信息的准确性。

致谢

我们感谢参与实验的农民以及鹿儿岛县农业大学。我们还要感谢鹿儿岛脑中心进行了富有成效的讨论。

参考文献

[1] E. Beede, E. Baylor, F. Hersch, A. Iurchenko, L. Wilcox, P. Ruamvi- boonsuk, and L. M. Vardoulakis, “在诊所部署的用于检测糖尿病视网膜病变的深度学习系统：以人为中心的评估”，载于*2020*年计算机系统人因学*CHI*会议（*CHI ’20*）论文集，2020年，第1—12页。[2] A. Abdul, J. Vermeulen, D. Wang, B. Y. Lim, and M. Kankanhalli, “可解释、可问责且易理解的系统的发展趋势与轨迹：人机交互研究议程”，载于*2018*年计算机系统人因学*CHI*会议（*CHI’18*）论文集，2018年，第1–18页。[3] F. Doshi-Velez and B. Kim, “迈向可解释机器学习的严谨科学”，2017年。[4] Z. C. Lipton, “模型可解释性的神话”，2017年。[5] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, “用于判别性定位的深度特征学习”，载于计算机视觉与模式识别（*CVPR*），2016年，第2921–2929页。[6] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM：通过基于梯度的定位对深度网络进行可视化解释”，国际计算机视觉杂志，第128卷，第2期，第336–359页，2019年10月。[7] M. T. Ribeiro, S. Singh, and C. Guestrin, “我为什么应该信任你？”：解释任意分类器的预测，载于*2016*年计算语言学协会北美分会会议论文集：演示（*KDD ’16*），2016年6月，第1135—1144页。[8] S. M. Lundberg and S.-I. Lee, “解释模型预测的统一方法”，载于神经信息处理系统进展第*30*卷，I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett 编。Curran Associates, Inc.，2017年，第4765–4774页。[9] T. Miller, “人工智能中的解释：来自社会科学的见解”，2018年。

[10] H. Kaur, H. Nori, S. Jenkins, R. Caruana, H. Wallach, and J. Wort- man Vaughan, “解读可解释性：了解数据科学家如何使用机器学习可解释性工具”，载于*2020*年计算机系统人因学*CHI*会议（*CHI ’20*）论文集，2020年，第1–14页。[11] A. Abdul, C. von der Weth, M. Kankanhalli, and B. Y. Lim, “Cogam：衡量并调节机器学习模型解释中的认知负荷”，载于*2020*年计算机系统人因学*CHI*会议（*CHI ’20*）论文集，2020年，第1–14页。[12] M. Yin, J. Wortman Vaughan, and H. Wallach, “准确率对机器学习模型信任度的影响”，载于*2019*年计算机系统人因学*CHI*会议（*CHI ’19*）论文集，2019年，第1–12页。[13] D. Wang, Q. Yang, A. Abdul, and B. Y. Lim, “设计理论驱动的以用户为中心的可解释人工智能”，载于*2019*年计算机系统人因学*CHI*会议（*CHI ’19*）论文集，2019年，第1–15页。[14] B. J. McGuirk, R. Forsyth, and H. Dobson, “英国奶牛群难产的经济成本”，《兽医记录》，第161卷，第20期，第685–687页，2007年。[15] M. Speroni, M. Malacarne, R. Federico, P. Franceschi, and A. Summer, “姿势变化增加是奶牛即将分娩的指标”，《农业》，第8卷，第11期，第1–11页，2018年。[16] M. Saint-Dizier and S. Chastant-Maillart, “预测牛分娩时间的方法与农场设备”，《兽医杂志》，第205卷，第3期，第349–356页，2015年。[17] M. Jensen, “奶牛在分娩前后的行为”，《应用动物行为科学》，第139卷，第3期，第195–202页，2012年。[18] L. P. Bueno, D. J. Tainturier, and Y. Ruckebusch, “利用实用预警系统检测母牛和母马的分娩”，《兽医学》，第16卷，第6期，第599–605页，1981年。[19] J. L. Owens, T. N. Edey, B. M. Bindon, and L. R. Piper, “为双胎产犊而选育的牛群中的分娩行为与犊牛存活率”，《应用动物行为科学》，第13卷，第4期，第321–333页，1985年。[20] K. Sugawara, S. Saito, T. Nakano, M. Akabane, T. Kobayashi, and T. Ogawa, “基于视频预测分娩：利用与日本黑毛肉牛分娩征兆相关的行为信息”，载于欧洲精准畜牧会议（*ECPLF 2019*），2019年，第663–669页。[21] R. Hyodo, T. Nakano, and T. Ogawa, “利用视频帧进行母牛分娩检测的特征表示学习”，载于第*25*届国际模式识别会议（*ICPR 2020*），2021年1月，第4131–4136页。[22] K. He, X. Zhang, S. Ren, and J. Sun, “用于图像识别的深度残差学习”，载于计算机视觉与模式识别（*CVPR 2016*），2016年，第770–778页。[23] H. Hermansky, E. Variani, and V. Peddinti, “平均时间距离：根据语音信号的时间属性预测自动语音识别错误”，载于国际声学、语音与信号处理会议（*ICASSP 2013*），2013年，第7423–7426页。[24] T. Ogawa, S. Mallidi, E. Dupoux, J. Cohen, N. Feldman, and H. Her- mansky, “一种新的高效准确率预测度量及其在基于多流的无监督自适应中的应用”，载于国际模式识别会议（*ICPR 2017*），2017年，第2222–2227页。[25] J. Redmon and A. Farhadi, “YOLOv3：渐进式改进”，*CoRR*，第abs/1804.02767卷，2018年。 [在线]。网址：http://arxiv.org/ abs/1804.02767[26] M. Jaspers, T. Steen, C. Bos, and M. Geenen, “出声思维法：用户界面设计指南”，《国际医学信息学杂志》，第73卷，第781–95页，2004年12月。[27] B. Dietvorst, J. Simmons, and C. Massey, “算法厌恶：人们看到算法出错后会错误地避开算法”，《实验心理学杂志：通论》，第144卷，2014年11月。[28] M. Dzindolet, L. Pierce, H. Beck, and L. Dawe, “视觉检测任务中人工与自动化辅助工具的感知效用”，《人因》，第44卷，第79–94页，2002年2月。