患者如何描述其疾病症状？采用人工参与整理的深度多标签文本分类，实现患者问题自我报告的自动标注

Lakshmi Arbatti、Abhishek Hosamath、VikramRamanarayanan 和 Ira Shoulson*Modality.AI, Inc.*<lakshmi.arbatti, abhishek.hosamath, v, ira.shoulson>@modality.ai

arXiv:2305.04905v1 [cs.CL] 8 May 2023

摘要——**美国食品药品监督管理局日益重视临床和研究环境中的患者报告问题。本文探讨了规模最大的在线数据集之一，其中包含 170,141 条帕金森病患者（PwP）针对“帕金森病给他们带来哪些困扰，以及如何影响其日常功能”这一问题作出的开放式自我报告回答（称为“原话记录”），也称为帕金森病患者问题报告©。将此类原话记录分类到多个临床相关症状类别是一项重要任务，需要经过多个步骤——专家整理、多标签文本分类（MLTC）方法，以及大量带标签的训练数据。此外，对如此庞大的数据集进行人工标注既繁琐又昂贵。我们提出一种新颖的解决方案：利用 2,341 条（共 170,141 条）由九位标注人员（包括临床专家和帕金森病患者）标注的原话记录，构建基线数据集。我们运用自然语言处理技术开发基于规则的语言词典，并构建基于图数据库的专家短语查询系统，将标注扩展至其余队列，从而生成机器标注数据集；最后，针对两个数据集构建基于 Keras-TensorFlow 的 MLTC 模型。在留出测试集上，机器标注模型在 65 个症状类别上的 F1分数达到 95%，显著优于基线模型。**

索引词—**多标签、文本分类、患者报告、短语查询提取**

I. 引言

医疗保健和临床研究是一个信息密集型行业 [1]。电子健康记录（EHR）的出现以及大量临床笔记的可用性，引起了众多研究人员的关注，并推动了自然语言处理（NLP）领域的发展。为了实现现有信息的自动化处理、准确分析和有效利用，需要从这些大型语料库中提取数据，并将其整理成计算机能够理解的结构化形式。将此类信息分类为结构化报告和标签，是医学文本分析中最常见的方法之一。研究人员已经构建了多种预训练语言模型，并利用这些临床笔记以及数百万条其他数据进行训练。从历史上看，这些自然语言处理算法一直被用于执行此类分类任务。其中一个例子是将主诉的自由文本分诊内容分类到预先确定的综合征类别中。类别 [2]。然而，具有临床相关性的分类需要专业知识，以便从自由文本中提取特定领域的信息和描述符。文本分析不存在适用于所有情况的通用解决方案或现成方案。因此，有必要采用集成方法，以针对具体用例探索和分类数据。采用人在环（human-in-the-loop）方法，可以确保在模型训练的整个过程中，对系统结果的准确率和临床相关性进行验证，其中包括生成金标准测试集、数据扩展和模型验证。

A. 背景和 *PD-PROP*© 数据收集

患者的声音正受到越来越多的研究界和监管机构的关注，这在很大程度上得益于专注于特定疾病的倡导组织、美国《21世纪治愈法案》的颁布[1], 以及美国食品药品监督管理局（FDA）的以患者为中心的药物开发（PFDD）倡议[2]。患者对自身疾病的描述至关重要，但传统上，这些信息通常通过分类量表收集，并由研究环境中的临床医生进行评分。由于患者之间存在较大的个体差异，且缺乏量化方法，直接获取患者的原话报告一直被认为难以实现。然而，当患者被询问，尤其是在保密的在线环境中，被问及疾病最令其困扰的方面时，与面对面接受临床专科医生问诊相比，他们所提供的回答往往更加细致入微，也更具洞察力。面对面交流通常平均只有约27秒，而且可能受到临床医生预期的影响。在线研究平台的出现以及医学信息学的发展，使系统性地收集和分析患者的实际体验或感受成为可能。FoxInsight（FI）[3] 是由 Michael J Fox 帕金森病研究基金会（MJFF） [3] 资助的规模最大的持续运营在线研究平台之一，该平台从自愿同意参与研究的、确诊患有帕金森病（PD）的成年人中收集数据，其中包括

[1]专员办公室。21世纪治愈法案。美国食品药品监督管理局：https://www.fda.gov/regulatory-information/selected-amendments-fdc-act/21st-century-cures-act[2]药物评价与研究中心以患者为中心的药物开发。美国食品药品监督管理局：https: //https://www.fda.gov/drugs/development-approval-process-drugs/cder-patient-focused-drug-development[3]Fox insight：https://foxinsight.michaeljfox.org

其他问卷，即 PD-PROP [4], [5]。目前，针对亨廷顿病、多发性硬化、抑郁症等疾病应用 PROP© 的类似研究也正在开展。因此，亟须建立一套流程方法学（包括预训练模型），以便在不同疾病和研究场景中应用时，能够对这类问题报告进行分类。

B. 相关工作

多标签文本分类（MLTC）有多种应用场景，例如流派检测[6],主题建模 [7] [8],以及电子健康记录（EHR）中的普通医学文本挖掘 [9]。Doc2Vec [8], 、通用句子编码器 [10]等深度学习向量嵌入算法是强大的工具，能够在大型向量空间中检测文档相似性。然而，这些方法仍存在局限，因为需要对生成的文档相似性进行人工评估，才能从类别聚类中提取更多见解。FasTag 方法通过自动标注临床记录，将其与用于计费的 ICD-9 和 ICD-10 编码进行匹配，在兽医数据中取得了较为理想的准确率（91%），但在人类医疗记录中的准确率较低（71%） [11]。使用 BERT 等预训练模型 [12]得到的标签分类结果更加通用，因为这些模型适用于多种输入数据类型，例如电子健康记录数据或临床笔记 [13]。此外，将原话记录归类到特定的临床症状类别具有挑战性，因为这些记录可能包含非常细微的差异，以至于不同的人实际上可能会以不同的方式报告相同的症状。此外，使用通用预训练模型需要在训练阶段投入大量数据资源和计算能力 [14]。过去，在领域密集型分类问题中，传统的基于规则的技术往往能够取得最佳性能。基于规则的词典结构简单易用且易于实现，但在遇到未知实体时表现不佳，并且往往导致较低的召回率，因为这些规则仅适用于非常特定的数据集 [15]。据我们所知，目前尚无工具能够以具有临床意义的方式，按照不同的症状类别收集患者问题报告并自动为其添加标签。

II. 方法学

我们的方法学由图1所示的三步法组成。第一步包括对原话记录进行初步分析，为整理团队提供所需的知识，以定义疾病特异性症状和领域。这也有助于定义基于规则的流程，以构建包含同义词以及相似术语和短语的语言词典。第二步，我们应用这些规则和生成的语言词典，将数据扩展至整个队列。最后，在第三步，我们训练一个深度学习模型来执行多标签文本分类。

A. 逐字分析

为简洁起见，PROP 问题“帕金森病给您带来的最困扰的问题是什么？”称为问题，而“这在什么方面...“这个问题是否困扰您（影响您的日常功能或完成待办事项的能力）？”作为后果。我们分析由问题和后果拼接而成的报告，以下称之为“原话记录”。我们发现，与完整的原话记录相比，单独分析问题或后果所提供的信息较少。为了初步了解原话记录的语言句法结构，我们通过依存句法分析检查了其中几条，以评估句子结构中是否存在共同模式（图 2）。回答主要由动词、名词和形容词组合构成。

1) 人机协同标注：接下来，标注人员相互讨论了帕金森病患者（PwP）常见的典型问题和症状，并整理出 14 个领域（即 PD 症状的广泛类别），每个领域包含若干具体症状，共计 65 种。例如，在领域和症状的定义中，姿势不稳被定义为一个领域，而平衡、下降和跌倒恐惧则被定义为姿势不稳领域内的症状。在定义的 14 个领域中，8 个与运动相关，6 个与非运动相关，包括认知、精神、睡眠、自主神经功能障碍、疲劳和疼痛。这一区分非常重要，因为所开发的模型能够独立分类非运动症状，而这些症状在不同疾病中普遍存在。有关更详细的症状列表，请参阅表 VII。

本次工作的结果形成了一份症状定义表，该表不仅定义了需要分类的症状，还明确列出了 65 种症状各自的纳入和排除标准。症状定义表为深入理解已定义症状提供了经过临床验证的可靠参考，从而弥合了研究人员与广大用户群体之间的差距，使双方能够更好地解读深度学习模型预测的结果标签，其示例如（表 I）所示。

在170,141条原话记录中，约有2,341条（每个症状类别最多抽取50条原话记录），即1.4%，由专家进行了整理标注。专家对症状进行了标记，并提供了原话记录中能够表明该分类的术语或短语。

2）语言词典的生成：我们在这一流程中的下一步是利用句子结构知识和人工整理的输出结果，构建一个语言词典[4] 。该词典采用三步法构建：

1) 词性提取 - 我们使用词性，特别是名词、形容词和动词，对症状报告的各个方面进行探索性可视化，以帮助标注人员定义领域和症状。2) *word2vec* 模型基于临床试验和 PubMed 数据训练，用于检测同义词 - 这一阶段是为每个.....构建语言词典的重要组成部分。

[4]https://plato.stanford.edu/archives/win2022/entries/ natural-language-ontology/

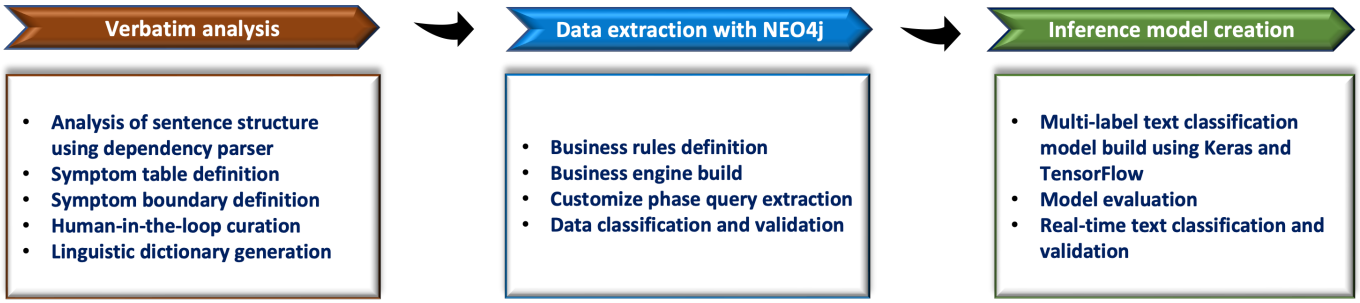

图1. 描述该流程各个阶段的流程图

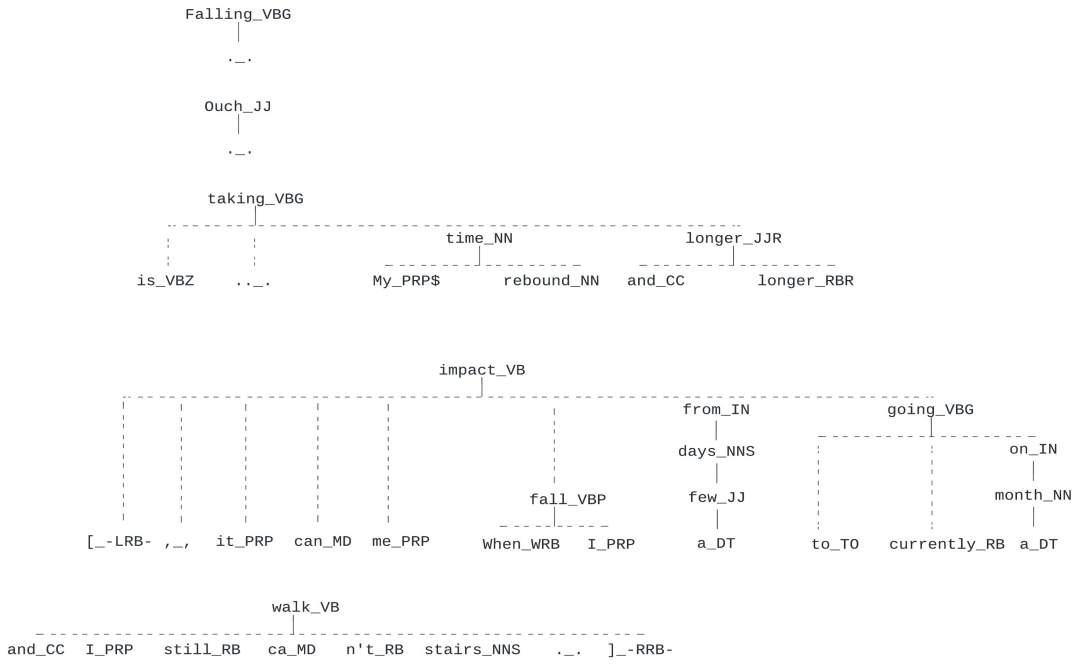

图2. 展示样本文本词性的依存树

特定术语。例如，当我们查询 word2vec 模型，要求其提供与“dystonia”相似概率最高的 4 个术语时——dystonia 是患者提及的一种病症——模型能够正确识别出 cramping、calf、ankle 等同义词，这些术语常用于与患者报告肌张力障碍是其困扰问题时相似的语境中（表 II）。3）提取统一医学语言系统 (UMLS)[5] 受控唯一标识符 (CUI)，以建立一份全面的表格，列出患者通常会用来描述特定症状的所有可能术语 [16] 和短语，从而生成语言词典，随后可将其输入专家系统，以提取适当的原话记录进行量化。UMLS CUI 提取示例见表 III。以这条原话记录为例：”抑郁症，对动作感到非常不愉快

问题、对未来的恐惧。 *[*我无法再像以前那样轻松地做那些事情了*]*”。在上面的例子中，我们看到，相关术语如“不快乐（C0476477）”被用于构建抑郁症状（C0086132）的语言词典。

B. 使用 *Neo4j* 进行数据提取

NEO4j 的[6] 全文搜索引擎由 Apache Lucene 提供支持，结合自定义搜索规则 [17]–[19]，将标注从经过筛选的 2,341 条原话记录扩展到包含超过 170,141 条原话记录的整个数据集 [11], [20], [21]。由原话记录以及所有其他参与者和就诊相关信息组成的整个数据集，均按照 Neo4j 的图模式进行组织，该模式使用节点和关系来存储数据。系统针对 Partic- 等唯一标识信息设置了索引，以实现高效的数据库运行性能

[5]https://www.nlm.nih.gov/research/umls/index.html

[6]https://neo4j.com/

表 I RBD 类症状定义表及示例短语。

领域	症状	概念边界		示例短语
		包括	不包括	
睡眠	REM 睡眠行为障碍（快速眼动睡眠行为障碍） like 症状	表现梦境、动作或发声- 睡眠期间出现的抽动；可能只有在.....时才会意识到 向伴侣报告；明确报告 REM 睡眠问题；夜惊	睡眠中无意识地移动 进一步说明	梦境行为化，快速眼动睡眠行为障碍， 快速眼动睡眠行为障-碍，睡觉时大喊

表 II
通过查询 word2vec 模型，获取与“肌张力障碍”（患者常提及的一种病症）最相似的 4 个术语及其概率。

word2vec 同义词	相似度概率	由所提出的算法分类的示例原文
’小腿’	0.776	左侧小腿收缩，脚趾蜷曲。 [偶尔难以开车和行走。]
’神经病变’	0.776	由于肌肉神经刺激不良导致肌肉量减少、脚趾蜷曲和神经病变
’踝部’	0.743	足部因腿部和臀部的肌肉紧张而感到不适。 [这主要是一种烦恼。脚趾蜷曲会导致锤状趾，难以找到穿着舒适的支撑性鞋具。] 疼痛、由肌张力障碍引起的踝部疼痛、肩痛以及腿抽筋导致的间歇性疼痛。
’运动障碍’	0.729	[踝部疼痛使行走困难。肩痛限制上半身活动。腿抽筋 打断了我本就断断续续的睡眠。] 运动障碍。 [导致颈部和肩部肌张力障碍，以及脚趾痉挛]

表 III CUI 关系示例

症状 CUI	相关概念唯一标识符				
抑郁症状 (C0086132)	快感缺失 (C0178417)	烦躁不安的情绪 (C0233477)	忧郁 (C0025193)	不快乐 (C0476477)	情绪低落 （C0344315）

表 IV 快速眼动睡眠行为障碍样症状术语表。

领域	症状	序号	术语1	术语2	术语3	术语4	术语5	术语6
睡眠	RBD like 症状	1	演绎 my 梦境	生动梦境	打斗 in 我的睡眠	在.....附近表现异常 the term 噩梦	尖叫 AND 睡眠	剧烈翻滚 near the term 梦

ipantID[7]。人口统计学等研究层面信息存储在患者节点中，并在其下创建多个相互链接的就诊节点，用于存储就诊层面的信息，例如就诊编号以及每次就诊的原话报告。系统还为每个就诊节点创建了扩展节点并建立索引，用于存储拼接后的原话问题和后果报告，以实现高速查询。所有节点均通过关系相互连接（图 3）。研究人员利用全文搜索功能制定了系统的症状标签标注规则，包括单词查询、短语查询、通配符查询、范围查询、正则表达式查询、模糊查询等[8]。这些规则针对每种症状分别创建，并根据以下内容存储在术语表中：

- 标注人员（帕金森病临床医生和经验丰富的专家）提供的症状纳入与排除标准
- 标注人员提供的标注和术语/短语
- 与上一节中定义的算法密切相关的术语，该算法是使用 word2vec 分类、UMLS 和概念唯一标识符
- ICD-10 编码。

表 IV 列出了一个快速眼动睡眠行为障碍（RBD）术语表示例。

使用 Python 脚本连接 Neo4j 数据库，以进行标注并提取原始文本数据。结合规则使用查询，遍历所有原话记录并对其进行标注。随后，使用 Pand- das[9] 进一步重塑所得数据，移除未分类和空白行，从而获得易于理解的分析数据集，其中包含原话记录及其相关症状。针对每种症状，从分析数据集中随机抽取约 1%，供标注人员验证。为进一步检验规则和机器分类的能力，验证数据集中的症状还加入了“阴性项”，即并非该症状、但与待验证症状密切相关的内容（例如，睡眠症状样本中加入了预测为报告疲劳的原话记录）。以下是密切相关症状的原话记录示例：

[7]https://neo4j.com/docs/cypher-manual/current/indexes-for-search-performance/ [8]https://lucene.apache.org/core/

[9]https://pandas.pydata.org

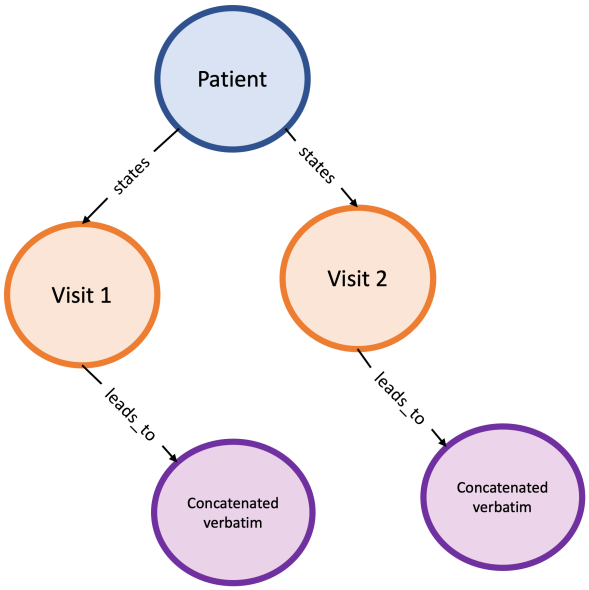

图 3. 原始文本数据库的图表示

- 内在震颤：“做事缓慢，内在颤抖。这是我在患帕金森病（*PD*）之前从未需要面对过的问题 *[”]*
- 焦虑：“焦虑——广泛性焦虑。 *[*导致难以在公众面前讲话，而且声音会颤抖*]”*

随后将验证结果与机器分类结果进行比较，并针对准确率较低的症状进一步优化和微调所构建的规则。经标注人员手动验证（表 VII），优化后的模型能够对 1555 条原话记录中的各种症状进行标注，准确率为 96%–100%。随后使用该数据集训练了一个可扩展的机器学习模型。

C. 推理模型创建 / *Keras* 多标签文本分类模型

在总计 170,141 条可用原话记录样本中，有 2,341 条由人工标注。随后，从这组样本中抽取约 445 条均匀分布的样本，作为留出测试集用于模型评估。剩余的 1,896 条样本构成“基线”训练集。使用该训练集训练了“基线模型”。第二个评估集则由剩余的 167,800 条样本构成，这些样本通过我们所述的流程进行了标注。在该数据集上训练的模型构成了“机器标注模型”。两个模型均采用相同的 Tensorflow Keras[10] 深度学习模型架构进行训练。在构建机器标注模型时，为消除类别不平衡，从 10,519 种独特的多标签组合中剔除了罕见的标签分类，即出现频率为 1 的标签组合，最终留下 159,115 条原话记录可用于模型训练。

[10]https://keras.io

采用了 90:10 的训练集-测试集划分比例。随后，将测试集进一步按照 50:50 的比例划分为测试集和验证集。使用 Sklearn 的 [11]train-test-split 函数完成这一任务，最终得到包含 143,203 条报告的训练集、7956 条测试样本和 7956 条验证样本。数据通过 Keras 的*StringLookup* 函数进行多热编码预处理，并使用 Keras 的*TextVectorization* 函数将数据转换为向量。该函数首先将数据转换为二元词组（bi-grams），然后使用 TD-IDF（词频-逆文档频率）进行表示。随后构建了深度学习模型架构，其中包含两个全连接隐藏层，分别具有 512 个和 256 个神经元，以及一个输出层。在尝试的多种激活函数组合中，隐藏层采用 ReLU （修正线性单元） [22] 函数、输出层采用 sigmoid 单元的组合取得了最佳结果。鉴于训练数据中标签组合的复杂性，ReLU 凭借其简单性以及避免 sigmoid 或 tanh 函数所面临的梯度消失问题的能力，最适合用于模型的隐藏层。输出层选择 sigmoid 函数，是因为在多标签分类模型中，每个标签的预测输出均为二元值。随后，使用二元交叉熵损失函数对模型进行编译，因为每条原话记录在被分类时，对应的每个类别的预测值均为 0 或 1。通过实证实验发现，运行 50 个训练轮次能够获得准确率、F1 分数、精确率和召回率的最佳组合。准确率采用 Keras 的 CategoricalAccuracy[12] 指标进行计算，其数值表示预测标签与实际标签相匹配的百分比。表 V 中的示例说明了各项衍生指标的计算方法。类别总数为 66 个，其中包括 65 种症状，以及一个表示无症状或“未知”的类别。需要注意的是，较低的精确率或较低的召回率并不一定意味着模型表现不佳。模型实际上可能检测到了整理团队尚未定义的模式，因此，这类原话记录应当被标记出来，以便进一步调查。

III. 结果

基线模型，即在包含 1896 个训练样本的人工标注数据集上训练的模型，在留出测试集上的表现较差，准确率、召回率和 F1 分数均较低，这表明训练数据不足。由于这是一个多标签分类任务，同时也受到过拟合的影响，该模型的精确率仍然相对较高。如表 VI 所示，我们提出的机器标注模型在留出测试集的所有指标上均优于基线模型。准确率提高了 42%，召回率提高了 79%，表明机器标注模型在识别特定多标签组合方面具有较高的效率。为了进一步排除参与者偏差的可能性，即模型通过学习同一参与者的症状而产生过拟合，

[11]https://scikit-learn.org [12]https://www.tensorflow.org/api_docs/python/tf/keras/metrics/ 分类准确率

表 V 展示我们如何计算多标签分类实验的准确率和其他指标的示例。（TP：真正例，FP：假正例，FN：假负例，TN：真负例）

示例真实/预测标签	指标值
Y_{true}：['平衡'] Y_{pred}：['下降']	分类准确率：0 真正例：0，真负例：65，假正例：0，假负例：1 精确率：0 ，召回率：0 ，F1 分数：0
Y_{true}：['平衡', '下降'] Y_{pred}：['平衡']	分类准确率：1 真正例：1，真负例：64，假正例：0，假负例：1 精确率：1 ，召回率：0.5 ，F1 分数：0.677
Y_{true}：['平衡'] Y_{pred}：['平衡'、'下降']	分类准确率：0 真正例：1，真负例：64，假正例：1，假负例：0 精确率：0.5 ，召回率：1 ，F1 分数：0.677
Y_{true}：['平衡', '下降'] Y_{pred}：['平衡', '下降']	分类准确率：1 真正例：2，真负例：64，假正例：0，假负例：0 精确率：1 ，召回率：1 ，F1 分数：1

为了进一步研究报告模式，我们对测试集中的一个子集进行了深入分析，该子集仅包含其原话记录未纳入训练集的参与者（125 名独立参与者，共 241 条原话记录）。与原始机器标注模型相比，我们仅观察到召回率和 F1 分数略有下降，这表明该模型在未见过的数据上也能表现良好。

总体而言，随着可用数据增多，改善出现频率较低的标签组合中的类别不平衡问题后，精确率还有进一步提升的空间。

表 VI 基线模型和机器标注机器学习模型在留出测试集上的指标

测量	基线模型	机器已标注模型	机器已标注模型（no 参与者偏差）
准确率	0.5217	0.9468	0.9499
F1分数	0.2407	0.9513	0.9499
精确率	0.8125	0.9736	0.9877
召回率	0.1413	0.9303	0.9160

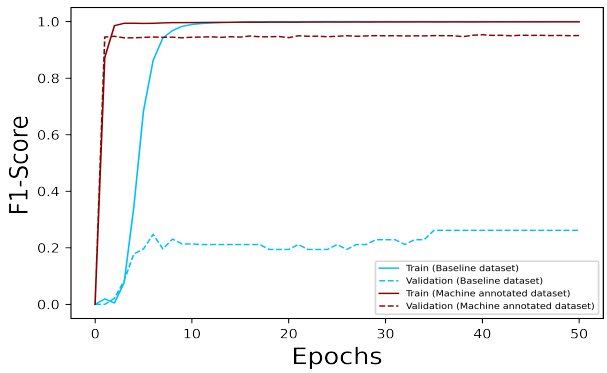

图 4. 在基线数据集和机器标注数据集上训练的模型的 F1 分数

该模型与预处理层结合使用，构建了一个推理模型，可直接根据原始输入文本逐字预测帕金森病（PD）症状。为展示该推理模型的应用和能力，我们结合推理模型、Python、Flask[13],和 Chart.js[14] 开发了一个 Web 应用程序，创建了一个 Web 界面，可将用户输入的 PD 病情描述实时加权分类，并以图表形式显示。图 5 展示了该试点界面对一段逐字输入文本进行分类的截图。

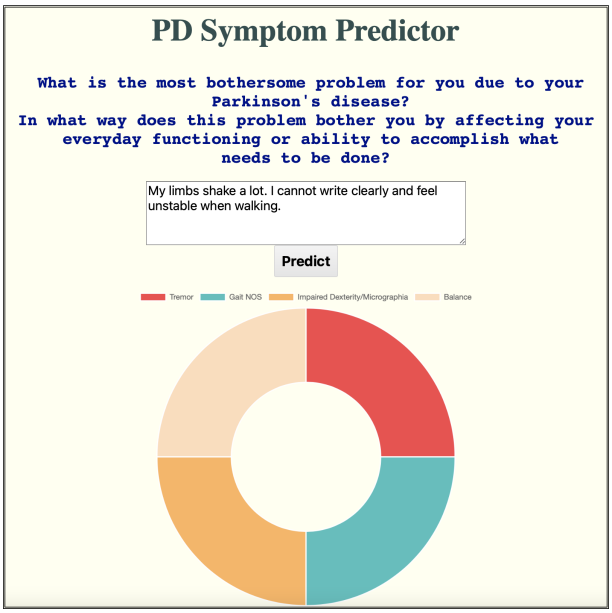

图 5. PD 症状预测器试点界面

IV. 讨论与后续步骤

本文提出了一种新颖的方法，用于将患者的原话记录分类为多个具有临床相关性的症状类别。首先，我们构建了一个基线子集，其中包含由九名标注人员（包括临床专家和患者）标注的原话记录。随后，我们利用基于规则的语言词典和基于 Neo4j 的专家短语查询系统，对其余原话记录进行机器标注。最后，我们训练了一个基于 Keras-TensorFlow 的多标签文本分类模型。我们进一步证明，在留出测试集上，机器标注模型在 65 个症状类别上的 F1分数达到 95%，显著优于基线模型。

我们提出的建模方法具有多项优势。首先，尽管语言词典及其生成的模型是基于 PD 数据集构建的，但该模型的突出特点在于其通用性，即能够应用于其他疾病的症状。这是因为多种运动症状和非运动症状并不局限于特定疾病。其次，通过制定纳入和排除标准，为每种症状明确界定范围，从而构建症状表，这一做法提供了临床

[13]https://flask.palletsprojects.com [14] https://www.chartjs.org

与派生输出的相关性。本项目旨在尽可能避免使用医学术语，以体现患者的声音，并方便非医疗背景的整理团队成员和未来的数据使用者使用症状表。我们将拟定的边界发送给标注人员，并不断完善，直至达成共识。

未来工作也有多种方向。虽然我们证明了模型即使面对来自未曾见过的参与者的原话记录也表现良好，但这些记录仅占留出测试集的一小部分。未来，我们计划在规模更大的未曾见过的参与者数据队列上评估模型的性能。此类评估还可能帮助发现新的、迄今尚未定义的症状，随后可将其纳入症状表，以便在主动机器学习框架下开展进一步研究和分类。目前，研究人员也在着手扩充语言词典，并为亨廷顿病和多发性硬化症等其他疾病建立症状表。随着语言词典和症状表不断完善，该模型也可协助评估前临床期和临床期疾病状态，进一步帮助研究人员和临床医生为患者提供适当且及时的照护。

致谢

本研究部分经费由迈克尔·J·福克斯基金会提供。作者还感谢FI参与者填写PD-PROP问卷并提供回答。

参考文献

[1] A. B. Wilcox 和 G. Hripcsak，“领域知识在自动化医学文本报告分类中的作用”，美国医学信息学会杂志，第 10 卷，第 4 期，第 330–338 页，2003年。

[2] W. W. Chapman、L. M. Christensen、M. M. Wagner、P. J. Haug、O. Ivanov、J. N. Dowling 和 R. T. Olszewski，“利用自然语言处理将自由文本分诊主诉归入综合征类别”，医学人工智能，第 33 卷，40，2005年 1，31期，第 – 页。[3] L. Smolensky 、 N. Amondikar 、 K. Crawford 、 S. Neu 、 C. M. Kopil、M. Daeschler、L. Riley、23andMe研究团队、E. Brown、A. W. Toga 和 C. Tanner，“Fox insight 收集帕金森病患者在线纵向报告的结局和遗传数据”，第 7 卷，第 1 期，第 67 页，2020年。[4] M. Javidnia、L. Arbatti、A. Hosamath、S. W. Eberly、D. Oakes 和 I. Shoulson，“帕金森病患者对姿势不稳和下降的原话报告症状的预测价值”，第 11 卷，第 4 期，第 1957–1964 页，2021年，出版商：IOS出版社。 [在线]。可访问：https: // content.iospress.com/articles/journal-of-parkinsons-disease/jpd 212036

[5] I. Shoulson、L. Arbatti、A. Hosamath、S. W. Eberly 和 D. Oakes，“将原话报告的姿势不稳症状作为在线帕金森病试验结局的纵向队列研究”，第 12 卷，第 6 期，第 1969–1978 页，2022年，出版商：IOS出版社。 [在线]。可访问：https://content.iospress.com/articles/journal-of-parkinsons-disease/jpd 223274

[8] P. Karvelis、D. Gavrilis、G. Georgoulas 和 C. Stylios，“使用 Doc2Vec 进行主题推荐”，载于*2018*年国际神经网络联合会议（*IJCNN*），2018年，第 1–6 页。

[6] M. M. Hasan、S. T. Dip、T. Rahman、M. S. Akter 和 I. Salehin，“基于电影字幕的多标签电影类型分类：参数优化的混合分类器”，载于*2021*年第*4*届先进电气与通信技术国际研讨会（*ISAECT*），2021年，第1–6页。[7] K. Nawab、G. Ramsey 和 R. Schreiber，“利用自然语言处理从患者体验反馈中提取有意义的信息”，第11卷，第2期，第242–252页，2020年，期号：2。 [在线]。可访问：https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7113005/ [9] Y. Zhang、R. Henao、Z. Gan、Y. Li 和 L. Carin，“利用卷积残差模型从医学纯文本中进行多标签学习”，载于医疗保健机器学习大会。PMLR，2018年，第280–294页。[10] D. Cer、Y. Yang、S.-y. Kong、N. Hua、N. Limtiaco、R. S. John、N. Constant、M. Guajardo-Cespedes、S. Yuan、C. Tar、Y.-H. Sung、B. Strope 和 R. Kurzweil，“通用句子编码器”，2018年。 [ ] 1803 11175 在线。可访问：https://arxiv.org/abs/ [11] 。G. R. Venkataraman、A. L. Pineda、O. J. B. D. W. Iv、A. M. Zehnder、S. Ayyar、R. L. Page、C. D. Bustamante 和 M. A. Rivas，“FasTag：非结构化医学叙述的自动文本分类”，第15卷，第6期，第e0234647页，2020年，期号：6，出版方：公共科学图书馆。 [在线]。可访问：https://journals.plos.org/ plosone/article?id =10 1371 0234647 。/journal.pone。J. Devlin、M.-W. Chang、K. Lee 和 K. Toutanova，“BERT：用于语言理解的深度双向变换器预训练”，2019年。 [ ] 1810 04805 在线。可访问：https://arxiv.org/abs/ [12] 。R. J. Turner、F. Coenen、F. Roelofs、K. Hagoort、A. H¨arm¨a、P. D. Gr¨unwald、F. P. Velders 和 F. E. Scheepers，“从自由文本中提取信息以辅助跨诊断精神病学：构建针对临床医生需求定制的自然语言处理流程”，第22卷，第1期，第407页，2022年，期号：1。 [在线]。可访问：https: // doi.org/10.1186/s12888-022-04058-z [13] [14] M. Pranji´c、M. Robnik-˘Sikonja 和 S. Pollak，“BERT 和 Doc2Vec 模型在 IPTC 主题代码预测数据集上的评估”，2020年。 [在线]。可访问：https://ailab.ijs.si/dunja/SiKDD2021/Papers/ Pranjicetal.pdf [15] E. H. Houssein、R. E. Mohamed 和 A. A. Ali，“生物医学自然语言处理的机器学习技术：综合综述”，*IEEE Access*，第9卷，第140 628–140 653页，2021年。[16] A. Sabbir、A. Jimeno-Yepes 和 R. Kavuluru，“利用神经概念嵌入进行基于知识的生物医学词义消歧”，第2017卷，第163–170页，2017年。 [在线]。可访问：https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5792196/ [17] J. Fries、S. Wu、A. Ratner 和 C. R´e，“Swellshark：一种无需标注数据的生物医学命名实体识别生成模型”，2017年。 [ ] 1704 06360 在线。可访问：https://arxiv.org/abs/ [18] 。Y. Wang、S. Sohn、S. Liu、F. Shen、L. Wang、E. J. Atkinson、S. Amin 和 H. Liu，“采用弱监督和深度表征的临床文本分类范式”，第19卷，第1期，第1页，2019年，期号：1。 [ ] 10 1186 12911 018 0723 6 在线。可访问：https://doi.org/ [19] 。/s - - - M. Cusick、P. Adekkanattu、T. R. Campion、E. T. Sholle、A. Myers、S. Banerjee、G. Alexopoulos、Y. Wang 和 J. Pathak，“利用弱监督和深度学习对临床笔记进行分类，以识别当前的自杀意念”，第136卷，第95–102页，2021年。 [在线]。可访问：https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8009838/ [20] Z. Zhang、R. Takanobu、Q. Zhu、M. Huang 和 X. Zhu，“面向任务的对话系统的最新进展与挑战”，2020年，文献类型：arXiv: 2003 07490 [ ] 2003 07490 。在线。可访问：https://arxiv.org/abs/ [21] 。H. A. Mohammad、S. Sivarajkumar、S. Viggiano、D. Oniani、S. Visweswaran 和 Y. Wang，“使用自然语言处理从阿尔茨海默病患者的临床笔记中提取睡眠信息”，2022年。 [在线]。可访问：http://medrxiv.org/lookup/doi/ 10.1101/2022.03.29.22273078 [22] A. F. Agarap，“使用修正线性单元（ReLU）进行深度学习”，2019年。[在线]。可访问：https://arxiv.org/abs/1803.08375

表 VII Neo4j流程提取的症状验证指标

领域	症状	F1分数
震颤	震颤	1
	内在震颤	1
僵直	僵硬	1
运动迟缓	缓慢	0.9714
	面部表情	1
姿势不稳	平衡	1
	跌倒	1
	跌倒恐惧	1
Gait	未特指的步态	0.9859
	步态冻结	1
其他运动症状	灵活性受损	1
	肌张力障碍	1
	姿势	1
	言语	1
波动	关期——药物相关	1
	Off 时段——未提及药物	1
	随机关期/不可预测的关期	1
	药物不起作用，鼻子	1
运动障碍搭配形式	运动障碍	1
睡眠	日间过度嗜睡（ES）	1
	入睡困难	1
	睡眠维持障碍	1
	清晨早醒	1
	睡眠质量差（未另行说明）	1
	RLS/坐立不安	1
	类似快速眼动睡眠行为障碍的症状	1
	未特指的异态睡眠	0.9859
	梦境	1
疲劳	身体疲劳	1
	精神疲劳	1
认知	记忆	0.9565
	注意力/专注力	1
	认知迟缓/精神疲劳	0.9565
	语言/词汇检索	1
	精神警觉性/意识状态	1
	视觉空间能力	0.9565
	执行功能/工作记忆	1
	未特指的认知障碍	1
情感/动机/思维- 知觉/其他精神症状	抑郁症状	0.9524
	死亡与自杀意念	1
	冷漠	1
	焦虑/担忧	1
	妄想/精神病性症状	1
	孤独/隔离	1
	冲动控制	1
	幻觉/错觉/存在感/经过感	1
	假性延髓情感	1
	未特指的负面情绪	1
Pain	疼痛/不适	1
	痉挛或抽搐	1
	头痛	1
自主神经功能障碍	大便失禁	1
	便急	1
	腹胀/饱胀感	1
	排便频率改变	1
	未特指的腹部不适	1
	尿失禁	1
	多汗	1
	尿频	1
	体位变化时出现头晕	1
	吞咽困难	1
	体温调节障碍	1
	性功能障碍	1
	腹泻	1
	恶心	1