

项目计划书

——基于复杂脑网络的轻度认知障碍

计算机辅助诊断

项目概述：近年来，静息态核磁功能成像技术(resting state functional Magnetic Resonance Imaging, rs-fMRI)和复杂网络理论为脑科学研究开辟了新方法。基于此背景，本项目针对轻度认知障碍疾病(Mild Cognitive Impairment, MCI)设计并研发了辅助诊断软件 MCICAD(Mild Cognitive Impairment Computer Aided Diagnosis)，目标用户主要为医生和相关研究者两个群体。诊断流程分为四步：(1) 首先对 MCI 组和正常对照组(CN 组)的原始 rs-fMRI 数据进行预处理；(2) 再通过皮尔逊偏相关系数和稀疏度阈值法构建出每个被试的脑功能网络，运用网络分析理论从脑网络中提取特征；(3) 采用 LASSO(Least Absolute Shrinkage and Selection Operator)方法进行特征选择，得出最优特征子集；(4) 利用支持向量机方法进行分类，使用留一法交叉验证进行评估。基于上述流程，本实验比较了阈值特征和 AUC 特征构建出的两种分类模型，利用 100 个样本训练出了留一法交叉验证准确率为 90% 分类模型。最后，为方便目标用户的使用，我们将本项目研究的 MCI 诊断流程整合为计算机辅助诊断软件 MCICAD(Mild Cognitive Impairment Computer Aided Diagnosis)，该软件将流程分成 4 个模块进行：预处理、网络分析、构建分类模型、诊断待测数据，在本项目中，每个步骤每项功能都通过了测试验证。通过此软件，医生可以精准快速的对患者进行 MCI 疾病的辅助诊断，研究者可以方便快捷地进行诊断优化研究以加快研究进程。

1 项目目标简述与需求分析说明

1.1 目标简述

轻度认知障碍(Mild Cognitive Impairment, MCI)是介于正常老化和阿尔茨海默病(Alzheimer Disease, AD)之间的过渡阶段，患者存在记忆或其它认知障碍，但达不到 AD 的诊断标准。研究表明，MCI 患者转化为 AD 的风险远高于正常老年人^[1]。由于 AD 的病因和发病机制尚不明确，目前还没有有效的治疗方法，而在人口老龄化形势日趋严峻的今天，这俨然成为一个严重的社会问题。但对于 MCI 患者，临床上已有方法可通过药物干预等方法有效改善症状、延缓病情发展。因此，对 MCI 患者的临床前预警和早期干预治疗尤为重要，而如何在 MCI 阶段就准确地诊断出病情，成为治疗 MCI 并有效降低 AD 发病率的关键。

目前虽然有医学和神经心理学方法的 MCI 诊断方法，然而由于 MCI 疾病临床表征轻微、致病因素复杂、病情反复，以及不同方法所得数据具有差异性等特点给医生的诊断造成了很大困难。神经影像学的快速发展，提供了一个有效的，安全系数高的方式来研究与疾病相关的大脑结构和功能上的差异，而运用复杂网络分析方法构建脑网络的方法已被证明比使用原始影像数据有更好的诊断效果。因此，本项目旨在基于 rs-fMRI 数据和复杂网络理论，对 MCI 疾病的计算机辅助诊断算法进行研究，并在此基础上为医生和相关研究者开发一个基于复杂脑网络的轻度认知障碍的计算机辅助诊断软件，帮助医生精确快速诊断，帮助 MCI 疾病相关研究者便捷开展研究并加快研究进程。

1.2 市场需求分析

MCI 计算机辅助诊断软件面向两类目标用户，医生和 MCI 疾病相关研究者。主要需要实现以下功能：

对于医生，提供辅助诊断功能，利用输入的 fMRI 数据，和默认模型及参数设置可完

成预处理、脑网络构建、特征提取、特征选择、分类等功能，最终须输出软件通过机器学习诊断出的结果（患病或不患病）；

对于 MCI 疾病相关研究者，除了可以使用默认模型及参数设置实现辅助诊断外，还提供自定义模型构建功能，若用户有足够的样本（已知是否患病）作为输入，可以通过自定义每一个步骤的参数，构建新的分类器模型，可以通过留一法交叉验证准确率比较各模型的优劣，并可用自定义的模型对未知样本进行分类诊断。

2 项目系统设计方案

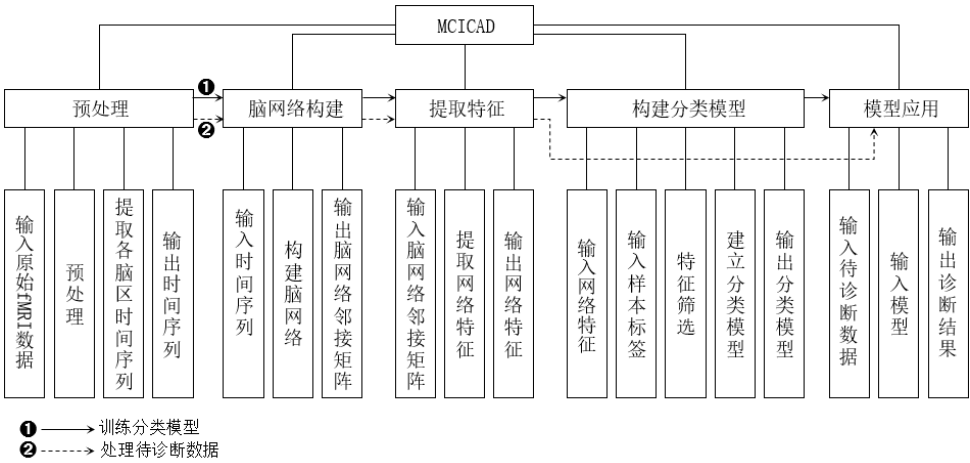


图 1 系统架构图

3 项目软件设计

3.1 算法设计

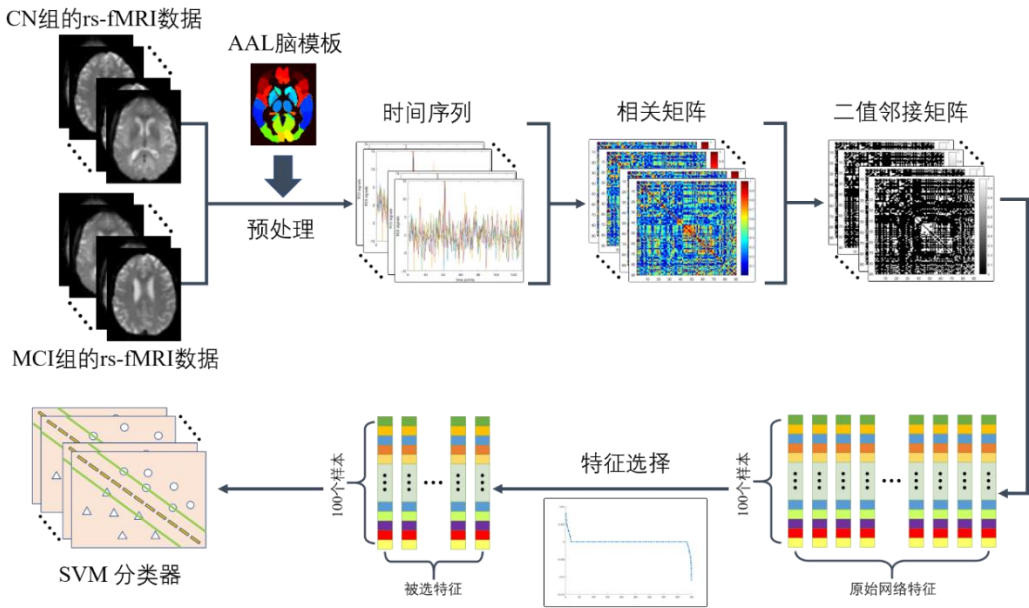


图 2 MCI 分类原理流程图

3.1.1 预处理

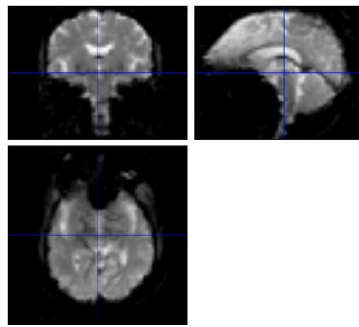


图3 MCI 组被试的 rs-fMRI 图像

对于获取到的 rs-fMRI 原始数据（如图 3 所示）需要先进行预处理，本实验的预处理步骤通过严超赣^[2]等人开发的 DPABI 软件实现，使用的预处理方法包括时间层校正、头动校正、去噪，空间标准化和平滑等，从而起到校正和标准化数据的作用。

3.1.2 网络分析

3.1.2.1 脑功能网络的构建

构建脑功能网络需要从经过预处理的 rs-fMRI 数据中可以提取出大脑各脑区的平均时间序列，以脑区时间序列之间的同步性表示脑区间的功能关系。本实验中的脑区是基于 AAL^[3]（Automated Anatomical Labeling atlas）标准脑模板进行划分的，AAL 标准脑模板将人脑划分为 116 个脑区，其中大脑区为 90 个左右对称区域，其余 26 个区域为小脑区。由于认知功能是由大脑主导的，对 MCI 进行分类研究时只针对 90 个大脑区进行网络构建。

通过复杂网络分析理论，脑功能网络中的节点可定义为脑模板划分出的脑区，网络的边可由大脑中任意两个脑区时间序列之间的相关系数定义，此处采用皮尔逊偏相关系数进行度量，由此可以得到相关矩阵。然对得到的相关矩阵进行去对角线元素、绝对值化处理，再采用稀疏度阈值法对网络进行阈值化，可得到二值邻接矩阵，即脑功能网络（如图 4 所示）。

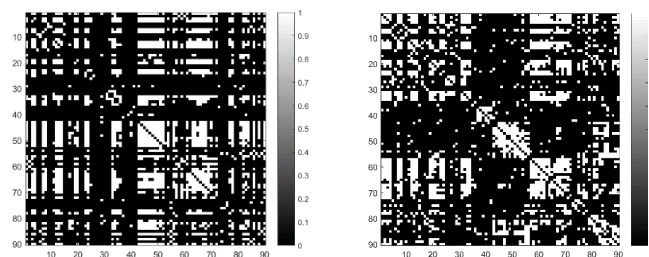


图4 脑网络的邻接矩阵可视化结果（左为 MCI 组样本，右为 NC 组样本）

3.1.2.3 特征提取

本实验计算了六组网络度量作为特征：节点度中心性节点全局效率，节点局部效率，介数中心性，网页排名（PageRank）以及节点度的匹配系数。其中除了匹配系数是网络全局特征，其他特征均为网络节点特征。节点特征就是对每一个节点进行度量，对应于大脑就是针对每一个脑区，所以每一组节点特征都有 90 个；而全局特征则是对网络整体进行度量。由于 MCI 仅是大脑中部分脑区发生病变，反映到整个大脑时变化并不明显，通过对一些全局特征计算相关性进行对比后，本实验最终选择了匹配系数作为脑功能网络的全局度量。

由于不同的阈值下构建出的脑网络不同，从中提取的网络特征的取值范围也不同，那么就会对分类模型的效果产生影响。因此，实验将选定阈值范围中所有阈值下计算出的特征进行积分，求得曲线下面积（Area Under The Curve, AUC），作为不同阈值下的网络特征的综合度量。这样一来，本实验中的所有网络特征都有阈值和 AUC 两种表现形式。

3.1.3 构建分类模型

3.1.3.1 特征选择

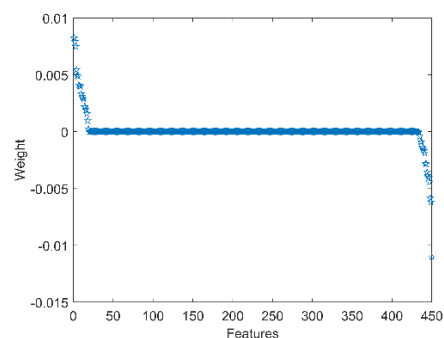


图5 特征选择

本步骤位于构建分类器之前，对接下来的步骤至关重要。有效的特征选择能够降低数据处理量，节省时间，避免数据冗余，减轻噪声影响，有助于快速建立分类模型并提高分类效果。

本项目采用的是LASSO(least absolute shrinkage and selection operator)算子以选择具有高区分性的特征子集。LASSO是一种基于稀疏线性回归方法的算子，通过向线性模型中加入L1惩罚项，将值很小的系数压缩至0，并将这部分系数所对应的特征变量舍弃，通过改变正则化参数Lambda的大小来决定舍弃的范围，从而实现对本样本数据的特征选择。不同特征对应的LASSO权重值排序后如图5所示。

3.1.3.2 SVM 分类

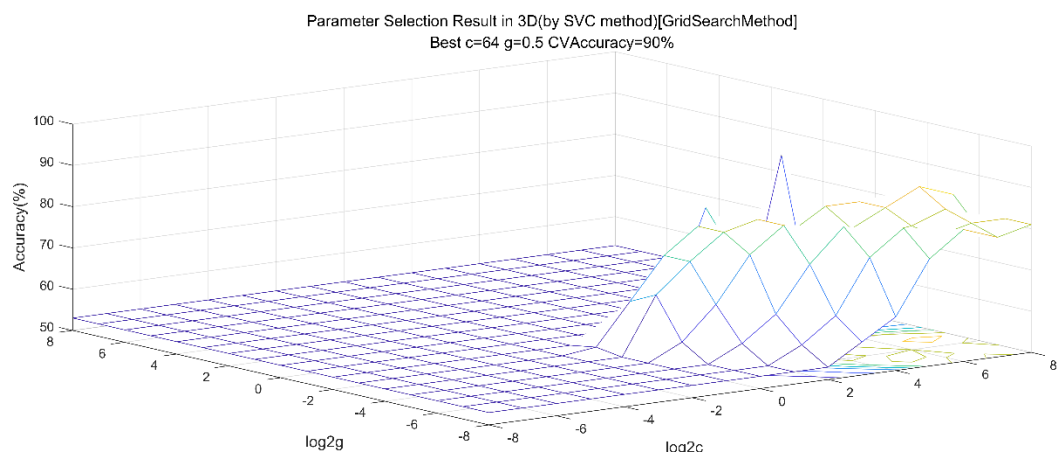


图6 网格寻参（脑网络阈值为0.3时）

SVM是一种发展成熟、应用广泛的机器学习算法，且对于小样本数据有较强的防止过拟合的能力。因此本实验选择使用支持向量机模型进行分类研究，具体步骤利用Chang.^[4]等人开发的Libsvm工具包实现。我们首先将特征选择后的数据进行归一化计算，将数据缩放至 $[-1, 1]$ 区间内，以加快训练速度并提高准确率。再结合网格寻参法和留一法交叉验证将归一化后的数据通过在一定范围内遍历参数进行训练和测试（如图6所示），并选出留一法交叉验证准确率最高的参数下训练出的分类模型，由此可以使用此模型对未知类别的数据进行分类，以达到MCI辅助诊断的目的。

3.2 软件设计

为了实现辅助诊断的功能，构建出可以直接使用的分类模型，本项目基于上述方法训练出了提供给医生用户直接进行诊断的默认模型。本项目使用的所有数据均从ADNI(Alzheimer's Disease Neuroimaging Initiative)公开数据库(<http://adni.loni.ucla.edu/>)中获得。通过条件控制和筛选，最终选取了100个样本用以

构建默认模型，样本情况见下表1。

表 1 样本的基本信息

| | 正常组 (CN) | 病患 (MCI) |
|-----------|------------|------------|
| 数量 | 47 | 53 |
| 男/女 | 19/28 | 24/29 |
| 年龄 | 75.74±4.62 | 72.11±4.22 |
| MMSE 量表评分 | 29.18±1.19 | 26.98±1.86 |
| CDR 量表评分 | 0.00±0.14 | 0.50±0.27 |

MMSE: Mini-Mental State Examination; CDR: Clinical Dementia Rating

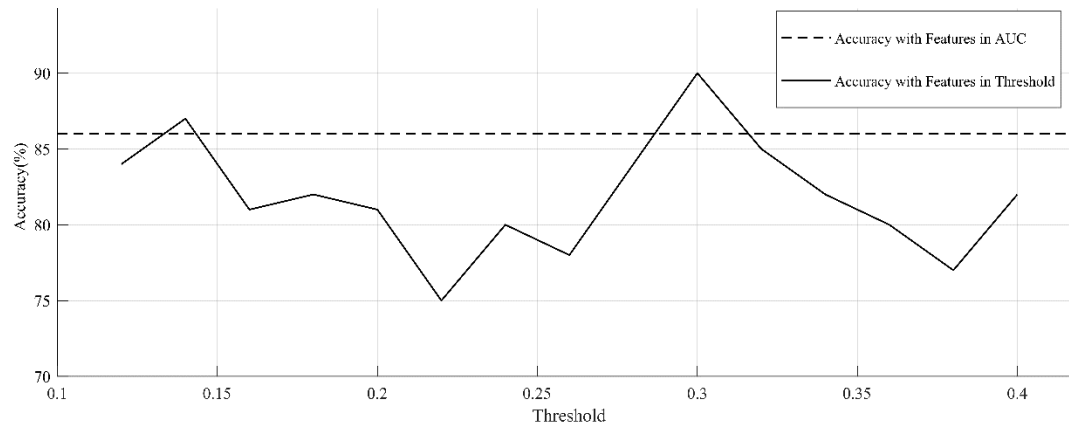


图 7 AUC 特征与阈值特征分类准确率对比

在构建网络时，本实验选取的阈值范围为[0.12, 0.4]，步长为0.02，总共15个阈值。从图7中可以看出，阈值选定为0.3时准确率达到峰值，为90%，而AUC特征的准确率可以达到86%，高于绝大多数阈值。本项目开发的软件将这两个模型设定为用户和直接利用的默认模型，用户可以在软件中导入默认模型以及处理好的待诊断数据样本直接进行分类诊断。

此外，软件还支持自定义参数设定，研究者也可以使用自己的训练数据通过设定不同参数来得到更适合自己的研究课题的分类模型。

4 项目系统集成与验证

本项目将需要实现的功能整合成一个基于MATLAB开发的软件MCICAD(Mild Cognitive Impairment Computer Aided Diagnosis)。

4.1 功能集成

4.1.1 MCICAD 主界面

最终完成的 MCICAD 一共分为四个模块（如图 8 所示）：预处理、网络分析、构建分类模型、诊断待测数据。用户可以通过 ReadMe 了结更多信息。

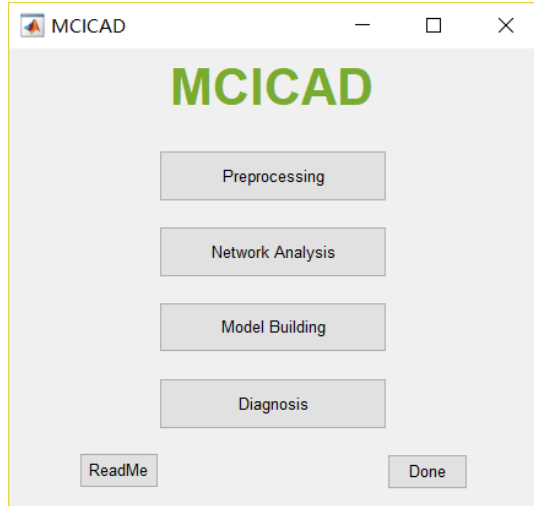


图 8 MCICAD 主界面

4.1.2 预处理

本模块采用了严超赣等人开发的预处理软件 DIBABI[1]中的 rs-fMRI 预处理模块 DPARSFA，用户导入数据后可以直接载入我们提供的默认设置文件进行预处理，得到输入样本的相关矩阵，也可通过调整各选项自定义预处理方案（如图 9 所示）。

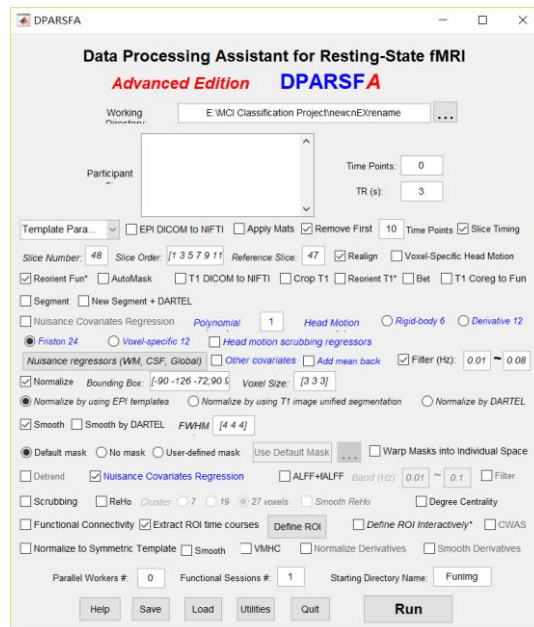


图 9 DIPARSFA 界面

4.1.2 网络分析

该模块整合了构建脑网络和提取特征两个步骤，输入文件夹中只能含有上一步骤的到的相关矩阵，用户还需要指定输出文件夹。对于构建脑网络，用户可以自定义阈值范围与间隔，构建好的脑网络最终都将被存入输出文件夹中的“\ThreshNet”文件夹中。之后用户可以选择想要提取的网络特征，这些特征都将存入输出文件夹中的“\NetworkFeature”中，这个文件夹里每个特征的阈值特征和 AUC 特征都将单独存储在各自的“.mat”文件中，此外还会有一个“orifea.mat”文件用来存储格式化了的特征，方便下一步操作。对于医生用户，此步骤目的为对待诊断的数据进行网络分析，产生诊断步骤所需的输入数据，软件界面显示的参数值为默认设置（如图 10 所示），指定输入输出文件夹后即可直接运行。

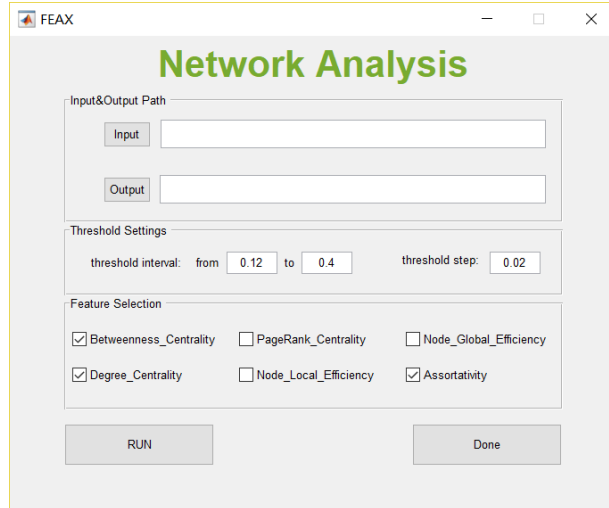


图 10 网络分析界面

4.1.3 构建分类模型

这一步骤专门为研究者设计，用以通过此软件使用不同的数据训练模型，医生可以跳过这一步骤直接使用软件提供的默认模型在下一步骤中对样本进行诊断。用户需要输入上一步骤输出的“orifea.mat”文件，此外还要输入样本标签的 Excel 文件。这一模块一共分三个步骤（如图 11 所示）：

（1）特征选择

在这里用户需要选择整个模型训练是基于阈值特征还是 AUC 特征，如果是基于阈值特征，软件将在最后比较不同阈值分类准确率得出最佳阈值，而如果是基于 AUC 将不会有这一步骤。此外用户还可以指定正则化参数 Lambda，以及选择的特征数目。在特征选择时，Lambda 越小，被选中的特征数目就越多，而对于最终特征选择数目，如果用户输入“0”，那么软件将保留特征选择中所有权重值非零的特征作为最优特征子集，如果用户输入的是非零值 n，那么软件就会根据权重的绝对值对特征进行排序并保留前 n 个特征作为最终的最优特征子集。

（2）训练模型

用户需先设定网格寻参所需的正则化参数 c、高斯核函数的标准差 g 的范围和步长，可以训练出交叉验证网格寻参下准确率最高的模型，此处使用的交叉验证方法默认为留一法，即折数为样本数，用户也可以根据需求自定义折数，如三折、十折等。

（3）保存模型

这一模块保存的文件除了训练好可以直接使用的 SVM 模型，同时也会保留最优阈值的序列以及该阈值下选择的参数标签以供使用者参考。

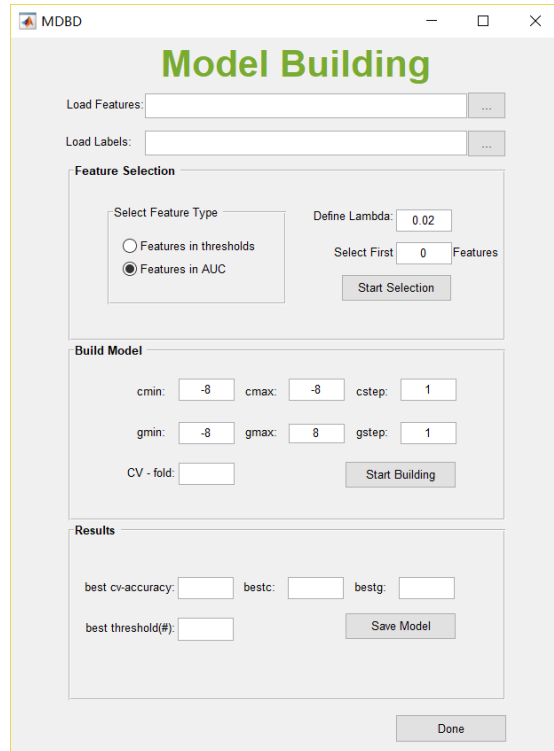


图 11 构建分类模型界面

4.1.4 诊断待测样本

这一步骤可以应用构建好的分类模型来诊断未知患病与否的样本（如图 12 所示），研究者可以在这一模块中测试上一步骤构建出的模型在新样本中的泛化能力，而医生则可以直接使用软件提供的默认模型对待诊断样本进行诊断。这一模块的输入文件必须是网络分析模块中得到的“orifea.mat”文件。

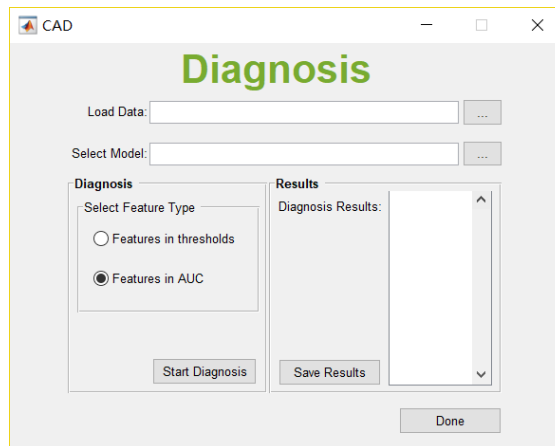


图 12 诊断待测样本界面

4.2 测试与验证

本项目针对功能和性能两方面以医生及相关研究者两类用户的需求进行了技术标准与指标的制定与测试。测试分为功能测试和性能测试，其中功能测试又分为主界面、预处理、网络分析、模型构建与诊断五个步骤进行测试，最终软件成功实现了每个模块的数据导入、参数设置，数据处理及数据存储等功能。而在性能测试方面，本项目用留一法测试了构建模型的交叉验证准确率（CV-accuracy），并测试了预测结果的准确率（accuracy）、精确率（precision）和召回率（recall）。结果显示，软件 MCICAD 在功能测试和性能测试中均表

现优异，能够帮助医生实现基于复杂脑网络的 MCI 计算机辅助诊断，同时也能协助此领域的研究人员更方便且高效地进行实验。

4.2.1 项目标准与技术指标

- (1) 基本功能正常：能够分别面向医生和研究人员成功完成预处理、脑网络构建、特征提取、特征选择、分类等功能。能够对各参数进行自定义，并且读取保存文件功能正常。
- (2) 模型构建交叉验证准确率 $\geq 80\%$ 。
- (3) 预测结果准确率 $\geq 75\%$ ，精确率 $\geq 75\%$ ，召回率 $\geq 75\%$ 。

4.2.2 测试计划

- (1) 测试构建分类模型（含预处理、网络分析、模型构建）过程是否完善，模型构建用留一法求出交叉验证准确率（CV-accuracy）并检验是否达标。
- (2) 测试模型应用（诊断）流程是否完善，包括默认模型测试和新建模型测试。
计算并检验诊断结果准确率（accuracy）、精确率（precision）和召回率（recall）是否达标

4.2.3 测试样本数据

| 测试内容 样本（个） | 构建模型功能测试 (Model Building) | 诊断功能测试 (Diagnosis) |
|---------------|------------------------------|-----------------------|
| 认知正常组（CN） | 35 | 15 |
| 轻度认知障碍组（MCI） | 35 | 15 |
| 总计 | 70 | 30 |

4.2.4 测试结果小结

根据功能测试结果可知，构建分类模型（含预处理、网络分析、模型构建）过程完善，模型应用（诊断）流程完善，默认模型测试和基于 30 个样本的新建模型均能正常使用，参数设置，数据存取等功能正常，用于研究时提供了较高的便捷性与自由度。

由性能测试结果可知，基于 AUC 特征的默认模型通过留一法验证得到交叉验证准确率 CV-accuracy = 83%，基于阈值特征的默认模型的交叉验证准确率 CV-accuracy = 90%，两个结果均高于技术指标 80%。同时，对测试样本的诊断结果各项数据皆达到标准：使用基于 AUC 特征的默认模型（以下简称 aucmodel）进行测试时，accuracy（准确率）= 83.3%，precision（精确率）= 77.8%，recall（召回率）= 93.3%。使用基于阈值特征的默认模型（以下简称 thdmodel）进行测试时，accuracy（准确率）= 90.0%，precision（精确率）= 92.9%，recall（召回率）= 86.7%。

两个默认模型对测试样本的分类准确率、精确率和召回率都超出了标准，值得注意的是，由于 aucmodel 的训练数据为一个阈值区间内特征值的积分，它相较于 thdmodel 具有更高的泛化能力，利于应对更多不同的数据。同时，其召回率更是达到了 93.3% 的高标准，这意味着 aucmodel 能够精准地判别出大部分的 MCI 患者。

对于 thdmodel，其训练数据为具体阈值下的特征值，故在最佳阈值下能表现出更高的分类准确率（90.0%）和精确率（92.9%），这意味着 thdmodel 能在测试数据的阈值被限定的情况下得到准确率极高的分类结果。

5 项目总结

本项目通过系统整体架构的设计、软件设计与开发、功能与性能的测试几个步骤，基于 rs-fMRI 数据和复杂网络分析法，结合稀疏线性回归和支持向量机等机器学习方法，对轻度认知障碍疾病进行了辅助诊断研究，并在此基础上开发了 MCI 疾病的辅助诊断软件 MCICAD。项目意义在于帮助医生分担工作量并提高诊断准确率，降低患者向不可逆的阿尔兹海默症转化的风险，以及为相关研究者提供便利，以加快 MCI 疾病的研究进程。本项目所开发的软件（MCICAD）结合最新的研究成果，将图论知识和机器学习应用于 MCI 辅助诊断，完善了从输

入 fMRI 数据到得出分类结果的全过程。MCICAD 提供了 MCI 辅助诊断的整体流程，并且细化为不同功能模块，每个模块都会输出过程结果文件，方便用户根据自己的需求使用。在实际应用中，根据使用者的不同，MCICAD 可以发挥不同的功能，医生可以直接输入 fMRI 数据得到机器分类结果作为诊断的参考，以此大大提高 MCI 诊断的效率和准确率；而对于相关研究者则可以根据自己的研究课题和样本进行进一步的研究，通过自定义各个步骤的参数得到不同的自定义分类模型。

6 参考文献

- [1] Petersen R C, Negash S. Mild cognitive impairment: An overview [J]. CNS Spectrums, 2008, 13(1): 45-53.
- [2] Yan C G , Wang X D , Zuo X N , et al. DPABI: Data Processing & Analysis for (Resting-State) Brain Imaging[J]. NEUROINFORMATICS, 2016, 14(3):339-351.
- [3] Tzourio-Mazoyer N , Landeau B , Papathanassiou D , et al. Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain[J]. Neuroimage, 2002, 15(1):273-289.
- [4] Chang C C , Lin C J . A Library for Support Vector Machines[J]. 2011.