

## 附录：强对流天气人工智能应用数据集说明文档

短时强降水、雷暴大风、冰雹人工智能应用数据集，均分为训练数据和测试数据，其中测试数据按照天气过程、考虑季节和区域均匀分布等因素，设置 A 榜测试数据(TestSetA)和 B 榜测试数据 (TestSetB)，A 榜测试数据中特征数据在挑战活动初期提供，标签数据保留用于评估测试，作为调优模型的参考基准；B 榜测试数据中特征数据在挑战活动后期提供，标签数据保留用于评估模型的真实性能。训练数据、A 榜测试数据、B 榜测试数据的比例为 80%：10%：10%。

### （一）短时强降水人工智能应用数据集

#### 1. 数据集组织结构和数据内容

##### （1）数据集组织结构

短时强降水人工智能应用数据集包括训练数据集、测试数据集（A 榜测试数据-TestA, B 榜测试数据-TestB）和数据说明文档三个部分的内容。其中训练数据集和测试数据集均为数据集实体，数据说明文档部分主要包括用于数据读取的每个样本的区域范围（起始经纬度）、量纲单位的说明信息及数据读取程序示例等，数据集组织结构见附图 1。



附图 1 短时强降水数据集组织结构

短时强降水人工智能应用数据集以短时强降水天气过程为单元，每 6 分钟一个样本，数据集共包含 1.38 万个过程，每个天气过程数据由该过程对应时间窗口范围内的若干

样本组成，共计 98.4 万个样本。

数据集实体的组织结构见附图 2，每个样本包含用于模型输入的降水量强度标签以及和降水生消演变相关的 10 个雷达特征和 27 个水汽条件、抬升条件和不稳定条件特征（基于国产再分析产品计算），数据均为二维空间数据，详细内容见附表 1、2。

## 2. 数据集格式说明

数据内容	数据格式	排列顺序	区域范围	备注
降水强度标签	numpy 的 npy 格式，二维网格空间数据	自西向东、自北向南排列，第一个数据点位于西北角，最后一个数据点位于东南角	标签范围见 Label-Format-File 目录下的 Label-Format-Zi ii.txt； 雷达特征范围见 RADA-Format-File 下的 RADA-Format-Zi iiii.txt； 环境条件特征范围见 NWP-Format-File 下的 NWP-Format-Zi iiii.txt。	Python 读取程序： import numpy as np data = np.load(fpa th+file)
雷达特征参量				
环境条件特征参量				

## 3. 数据集时间属性

（1）时间范围：5 年时序长度，本次活动不提供年份信息，只提供月份、日期和时刻信息。

（2）时间分辨率：①降水强度标签数据：时间分辨率为降水过程对应雷达站的体扫时间分辨率，约 6 分钟；②雷达特征参量数据：10 种雷达特征参量数据时间分辨率为对应

雷达站的体扫时间分辨率，约 6 分钟；③环境大气特征参量：27 种环境大气特征参量时间分辨率为逐小时。

#### 4. 数据集空间属性

(1) 空间范围：覆盖中东部地区 18 个省份，不同过程选取的雷达站不同，数据的覆盖范围不尽相同。①每个强降水过程空间范围为以对应雷达站为中心的周边  $3^{\circ} \times 3^{\circ}$  区域，约  $300\text{km} \times 300\text{km}$ ，覆盖  $301 \times 301$  个网格。②环境大气特征参量：空间范围为以降水过程对应的雷达站位置最邻近的环境大气特征参量网格为中心的  $4.98^{\circ} \times 4.98^{\circ}$  区域，覆盖  $167 \times 167$  个网格点。

(2) 空间分辨率：①降水标签数据： $0.01^{\circ} \times 0.01^{\circ}$ ；②雷达特征参量：每部雷达对应产品的空间分辨率为  $0.01^{\circ} \times 0.01^{\circ}$ ；③环境大气特征参量：空间分辨率为  $0.03^{\circ} \times 0.03^{\circ}$ 。

#### 5. 特征值属性

数据类型		特征值	含义描述
降水标签	降水强度标签	-9	缺失值
雷达特征参量	所有产品	-32768	缺失值
		-1280	雷达无有效回波
环境大气特征	所有产品	-9999	缺失值

#### 6. 训练数据和测试数据划分

按 80%: 10%: 10% 的比例，考虑年份、季节和区域均匀分布因素，选取 1.1 万个过程，78.8 万个样本作为训练数据，选取 1385 个过程，9.8 万个样本作为 A 榜测试数据，选取 1382 个过程，9.8 万个样本作为 B 榜测试数据，其中 A、B

榜测试数据只提供雷达特征和环境条件特征，不提供标签数据。

短时强降水					
训练集（80%）		A 榜测试集（10%）		B 榜测试集（10%）	
过程数	样本数	过程数	样本数	过程数	样本数
1.1 万	78.8 万	1385	9.8 万	1382	9.8 万

- **Data**
  - **CP (降水代码)**
    - Trainset (训练数据集)
      - REG1 (区域1代码)
        - CaseID1 (过程1 ID号)
          - LABEL (标签数据)
            - CP\_Label\_RA\*time1.npy(time1时刻降水强度标签)
            - CP\_Label\_RA\*time2.npy(time2时刻降水强度标签)
            - ..... (其他时刻标签)
        - RADAR (雷达特征)
          - CR (组合反射率)
            - CP\*time1\_CR.npy (time1时刻组合反射率)
            - CP\*time2\_CR.npy (time2时刻组合反射率)
            - ..... (其他时刻组合反射率)
          - HBR (混合扫描反射率)
            - CP\*time1\_HBR.npy(time1时刻混合扫描反射率)
            - CP\*time2\_HBR.npy(time2时刻混合扫描反射率)
            - ..... (其他时刻混合扫描反射率)
          - ..... (其他8个雷达特征)
        - NWP (基于自主再分析产品计算的环境条件特征)
          - CAPE (对流有效位能)
            - CP\*time1\_CAPE.npy (time1对流有效位能)
            - CP\*time2\_CAPE.npy (time2对流有效位能)
            - ..... (其他时刻对流有效位能)
          - KI (KI指数)
            - CP\*time1\_KI.npy (time1时刻KI指数)
            - CP\*time2\_KI.npy (time2时刻KI指数)
            - ..... (其他时刻KI指数)
      - CaseIDn..... (其他过程ID号)
    - REGx..... (其他区域代码)

附图 2 短时强降水数据集实体组织结构



附表 1 短时强降水过程对应的雷达特征参量列表

产品名称	产品标识	数据单位
组合反射率	CR	0.1dBZ
混合扫描反射率	HBR	0.1dBZ
2km 等高面反射率	CAP02	0.1dBZ
3km 等高面反射率	CAP03	0.1dBZ
4km 等高面反射率	CAP04	0.1dBZ
5km 等高面反射率	CAP05	0.1dBZ
6km 等高面反射率	CAP06	0.1dBZ
7km 等高面反射率	CAP07	0.1dBZ
回波顶高	ET	0.1km
垂直累积液态水含量	VIL	0.1Kg m <sup>-2</sup>

附表 2 短时强降水过程对应的环境大气条件特征参量列表

名称	简写	含义描述
动力抬升条件特征量	925hPa 散度	DVG925
	850hPa 散度	DVG850
	200hPa 散度	DVG200
	925hPa 风速	WS925
		反映低层速度场辐合的强弱程度，由同一水平面或等压面的风场计算求得。短时强降水事件中，低层辐合演变一般具有单波谷或双波谷特征，且中尺度特征明显。在强降水前期低层散度为辐合或由辐散转为辐合，表现为负值；强降水集中期低层辐合发展至最强；之后逐渐减弱，强降水后期低层转为辐散，表现为正值。
		与 925hPa 散度相似。
		反映对流层高层速度场辐散的强弱程度。短时强降水事件中，该值基本都为正值，即表现为辐散特征。
		低层风速能较好地反映出动力和水汽条件，其大小和方向与短时强降水或暴雨的发生密切相关：短时强降水与低层偏南风及风速呈正相关关系，其值越大，对发生短时强降水越有利；低空偏南气流的稳定维持利于降水过

名称		简写	含义描述
			程持续，低空急流的建立有利于短时强降水的形成；急流轴上中小尺度系统的时间尺度在 0.5 ~ 4h，其对强降水出现有极强的预示作用，其中低空急流中尺度风速脉动出现比强降水提前出现。如果低空的偏南气流遇地形抬升，则更易形成短时强降水。低空急流的加强和向近地面扩展伴随着动量下传，会促使低层辐合扰动增强，有利于触发不稳定能量强烈释放，从而有利于发生短时强降水，且降水强度与超低空急流风速脉动（变化）有一定的对应关系，有很好的短时临近预报预警指示作用。
	700hPa 风速	WS700	反映出对流层低层到中层过渡中的动力抬升和水汽条件。与低层风速的作用相似，该层的偏南风速越大，就越有利于促进形成短时强降水。
	500hPa 风速	WS500	衡量对流层中层的动力和水汽条件因子之一。 该层风速反映出的短波波动会造成降水强度的强弱波动变化：当高空气流由偏西风转为西南风，中低空西南气流和中高空西南气流同向叠加时，对流层中层偏南风急流建立并且向超低空传递，高空冷空气的加入和低层偏南气流加强促使中小尺度系统生成，就会导致促进形成短时强降水或其强度增强。
水汽条件特征量	1000hPa 比湿	Q1000	比湿是在一团湿空气中水汽的质量与该团空气总质量（水汽质量加上干空气质量）的比值，其单位是 g/kg。它是直接反映水汽条件的特征量之一，与短时强降水密切相关，即比湿越大就越有利于发生短时强降水。
	850hPa 比湿	Q850	反映对流层低层的水汽状况。该值越大越有利于形成短时强降水，且雨强



名称	简写	含义描述
		可能越大。
700hPa 比湿	Q700	反映出对流层低层向中层转换过程中的水汽状况，其值的增大对短时强降水也具有促进作用。
1000hPa 相对湿度	RH1000	表示空气中的绝对湿度与同温度和气压下的饱和绝对湿度的比值，反映边界层内的水汽状况，该值越大预示着水汽越饱和。短时强降水发生前或过程中，该值基本都接近 100%。
700hPa 相对湿度	RH700	与低层相对湿度类似，其值越大越有利于发生短时强降水或雨强增大。在地势较高处的我国中部偏西地区，该层高度相当于东部地区的低层，它反映的是当地的低层水汽条件，是判别对流或强降水天气的重要参考特征量之一。
500hPa 相对湿度	RH500	衡量对流层中层水汽状况的特征量。该值越大就越有利于形成短时强降水。
整层可降水量	PWAT	衡量空中水汽含量的参量，反映了水汽的输送和分布特征，其值的大小与地面降水呈正相关关系。可由地面湿度参量（露点温度等）、探空资料计算或地基 GPS 探测资料反演得到，三种计算方法的结果接近。
降水效率	PE	单位时间内降水量与云内凝结液态水含量之比。短时强降水的形成不仅需要充沛的空中水汽，还要有水汽降落至地面的有利条件，二者缺一不可，因此高降水效率是短时降水形成的关键因素。
地面至 850hPa 平均露点温度	TdSfc850	露点温度是在空气中水汽含量不变，保持气压一定的情况下，使空气冷却达到饱和时的温度，水蒸气与水达到平衡状态的温度，气象上通过饱和水汽压公式(马格拉斯经验公式)的近似求解得到。该值越高，就预示着温度

名称		简写	含义描述
			较高且湿度较大。地面至 850hPa 的平均露点温度反映的是对流层低层及以下的热力和水汽条件，其值越高就越有利于形成短时强降水。
	700-400hPa 平均温度露点差	TTdMean74	温度露点差是相对湿度的另一种度量，表示空气干湿程度或水汽饱和程度。700-400hPa 平均温度露点差反映的是对流层中层的水汽状况，该值越大越有利于短时强降水的发生及其雨强的增大。
	700-400hPa 最大温度露点差	TTdMax74	与上述平均值类似，但其反映出的水汽状况能较好地指示短时强降水的强度潜势。
	湿球 0℃ 高度	HTw0	将不同气压层的湿球温度点连接起来，则得到湿球温度垂直廓线，而该垂直廓线与 0℃ 等温线的交点对应高度即为湿球 0℃ 层高度（Wet Bulb Zero, WBZ）。当对流层整层的相对湿度较大时，WBZ 高度只是略低于阈值（常为 4.0 km，不同季节或地区有所差异），仍属于相对较高的值，这时伴随发生短时强降水的几率较大。
	100mb 以下 平均气块抬升凝结高度 (LCL)	LCL	抬升凝结高度 (LCL) 是未饱和湿空气块干绝热上升，刚开始凝结的高度，常用于代表云底的高度，可通过干绝热过程中的温度递减率和露点温度递减率估计。该值取决于近地空气层的湿度大小：湿度大时 LCL 低，反之则高。100mb 以下平均气块抬升凝结高度反映出了对流层内的湿度状况，尤其对流层低层的水汽条件；该值越低，则低层水汽条件越好，发生短时强降水的可能性就越大。
	100mb 以下 平均不稳定抬升凝结高	muLCL	可反映出由动力条件造成对流抬升引起水汽凝结的高度，其结果与对流凝结高度 (CCL) 接近。该值越低，则发

名称		简写	含义描述
	度		生对流性降水的可能性就越大。
热力/能量条件特征量	K 指数	KI	利用温度和露点反映出大气热力状况的特征量，根据其计算公式： $K = (T_{850} - T_{500}) + T_{d850} - (T - T_d) \times 700$ ，第一项为温度垂直递减率，第二项为低层水汽条件，第三项是中层饱和程度，因此 K 指数可以反映出大气的层结稳定情况，越大时层结越不稳定，业务经验表明该值越大越有利于发生短时强降水。
	500hPa 抬升指数	LI500	气块从低层 900m 高度沿干绝热线上升，到达凝结高度后再沿湿绝热线上升至 500hPa 时所具有的温度 $T_s$ 与 500hPa 等压面上的环境温度 $T_{500}$ 的差值。反映的是对流层中层的稳定程度，当该值 $<0$ 时，大气层结不稳定，且负值越大，不稳定程度越大，就越有利于形成短时强降水。
	300hPa 抬升指数	LI300	与 500hPa 抬升指数类似，但其反映出了对流层高层的不稳定程度。
	0℃层高度	HT0	0℃温度所在高度反映的是环境温度在垂直方向上的分布，对流云的伸展高度受其影响很大，因此与大风、冰雹等的出现密切相关，即该值高度越低时越有利于形成风雹（满足其他对流条件后），相反时则更有利于产生短时强降水。
	-10℃层高度	HT10	与 0℃层高度类似，该值越大也越有利于形成短时强降水。另外，它也与大气中闪电能否形成有较好关联，对判识短时强降水伴随对流活动的强弱程度有一定参考指示作用。
	自由对流高度	LFC	气块温度与环境温度之差由负值转为正值的转折点的高度，常用埃玛图上的状态曲线与层结曲线相交点所在的

名称	简写	含义描述
		高度来表征。该值越低就越有利于形成短时强降水。
对流有效位能	CAPE	在自由对流高度之上，气块可从正浮力做功而获得的能量，常用埃玛图上的正面积大小来表示。反映的是大气中不稳定能量的大小，与对流活动的强弱程度直接相关。该值越大，就越有利于形成对流性降水天气。

## （二）雷暴大风人工智能应用数据集

### 1. 数据集组织结构和数据内容

雷暴大风人工智能应用数据集包括训练数据集、测试数据集（A 榜测试数据-TestA, B 榜测试数据-TestB）和数据说明文档三个部分的内容。其中训练数据集和测试数据集均为数据集实体，数据说明文档部分主要包括用于数据读取的每个样本的区域范围（起始经纬度）、量纲单位的说明信息及数据读取程序示例等，数据集组织结构见附图 3。

雷暴大风人工智能应用数据集以雷暴大风天气过程为单元，每 6 分钟一个样本，数据集共包含 5237 个过程，每个天气过程数据由该过程对应时间窗口范围内的若干样本组成，共计 29.8 万个样本。

数据集实体的组织结构见附图 4，每个样本包含用于模型输入的大风强度标签以及和雷暴大风生消演变相关的 7 个雷达特征和 25 个水汽条件、抬升条件和不稳定条件特征（基于国产再分析产品计算），数据均为二维空间数据，详细内容见附表 3、4。

### 2. 数据集格式说明

数据内容	数据格式	排列顺序	区域范围	备注
大风强度 标签	numpy 的 npy 格式， 二维网格空间 数据	自西向东、自北向南排列，第一个数据点位于西北角，最后一个数据点位于东南角	标签范围见 Label-Format-File 目录下的 Label-Format-Ziii.txt; 雷达特征范围见 RADA-Format-File 下的 RADA-Format-Ziiii.txt; 环境条件特征范围见 NWP-Format-File 下的 NWP-Format-Ziiii.txt。	Python 读取程序： <pre>import numpy as np data = np.load(fpath+file)</pre>
雷达特征 参量				
环境条件 特征 参量				

### 3. 数据集时间属性

(1) 时间范围：5 年时序长度，本次活动不提供年份信息，只提供月份、日期和时刻信息。

(2) 时间分辨率：①大风强度标签数据：时间分辨率为雷暴大风过程对应雷达站的体扫时间分辨率，约 6 分钟；②雷达特征参量数据：7 种雷达特征参量数据时间分辨率为对应雷达站的体扫时间分辨率，约 6 分钟；③环境大气特征参量：25 种环境大气特征参量时间分辨率为逐小时。

### 4. 数据集空间属性

(1) 空间范围：覆盖中东部地区 18 个省份，不同过程选取的雷达站不同，数据的覆盖范围不尽相同。①每个雷暴大风过程空间范围为以对应雷达站为中心的周边  $3^{\circ} \times 3^{\circ}$  区域，约  $300\text{km} \times 300\text{km}$ ，覆盖  $301 \times 301$  个网格。②环境大气



特征参量：空间范围为以雷暴大风过程对应的雷达站位置最邻近的环境大气特征参量网格为中心的  $4.98^{\circ} \times 4.98^{\circ}$  区域，覆盖  $167 \times 167$  个网格点。

(2) 空间分辨率：①大风标签数据： $0.01^{\circ} \times 0.01^{\circ}$ ；②雷达特征参量：每部雷达对应产品的空间分辨率为  $0.01^{\circ} \times 0.01^{\circ}$ ；③环境大气特征参量：空间分辨率为  $0.03^{\circ} \times 0.03^{\circ}$ 。

## 5. 特征值属性

数据类型		特征值	含义描述
大风标签数据	大风强度标签	-9	缺失值
雷达特征参量	所有产品	-32768	缺失值
		-1280	雷达无有效回波
环境大气特征参量	所有产品	-9999	缺失值

## 6. 训练数据和测试数据划分

按 80%: 10%: 10% 的比例，考虑年份、季节和区域均匀分布因素，选取 4173 个过程，24 万个样本作为训练数据，选取 526 个过程，2.9 万个样本作为 A 榜测试数据，选取 538 个过程，2.9 万个样本作为 B 榜测试数据，其中 A、B 榜测试数据只提供雷达特征和环境条件特征，不提供标签数据。

雷暴大风					
训练集 (80%)		A 榜测试集 (10%)		B 榜测试集 (10%)	
过程数	样本数	过程数	样本数	过程数	样本数
4173	24 万	526	2.9 万	538	2.9 万



- **Data**
  - **TSW ( 雷暴大风代码 )**
    - Trainset ( 训练数据集 )
    - TestSet ( 测试数据集 )
      - TestA ( A榜测试集 )
      - TestB ( B榜测试集 )
  - **DOC ( 数据集说明文档 )**
    - Label\_Format\_DOC ( 标签格式文档 )
      - Label\_Format\_Ziiii1.txt ( 雷达站1对应样本格式说明 )
      - Label\_Format\_Ziiii2.txt ( 雷达站2对应样本格式说明 )
      - ..... ( 其他范围的样本格式说明 )
    - Radar\_Format\_DOC ( 雷达特征格式文档 )
      - Radar\_Format\_Ziiii1.txt ( 雷达站1对应样本格式说明 )
      - Radar\_Format\_Ziiii2.txt ( 雷达站2对应样本格式说明 )
      - ..... ( 其他范围的样本格式说明 )
    - NWP\_Format\_DOC ( 环境条件特征格式说明 )
      - NWP\_Format\_Ziiii1.txt ( 雷达站1对应样本格式说明 )
      - NWP\_Format\_Ziiii2.txt ( 雷达站2对应样本格式说明 )
      - ..... ( 其他范围的样本格式说明 )
    - 读取程序示例
      - Label\_import.py ( 标签读取程序示例 )
      - Radar\_import.py ( 雷达特征读取程序示例 )
      - NWP\_import.py ( 环境条件特征读取程序示例 )

附图 3 雷暴大风数据集组织结构

- **Data**

- **TSW ( 雷暴大风代码 )**

- **Trainset ( 训练数据集 )**

- REG1 ( 区域1代码 )

- CaseID1 ( 过程1 ID号 )

- LABEL ( 标签数据 )

- TSW\_Label\_WA\*time1.npy(time1时刻大风强度标签)

- TSW\_Label\_WA\*time2.npy(time2时刻大风强度标签)

- ..... ( 其他时刻标签 )

- RADAR ( 雷达特征 )

- CR ( 组合反射率 )

- TSW\*time1\_CR.npy ( time1时刻组合反射率 )

- TSW\*time2\_CR.npy ( time2时刻组合反射率 )

- ..... ( 其他时刻组合反射率 )

- R05 ( 0.5仰角反射率 )

- TSW\*time1\_VIL.npy(time1时刻0.5仰角反射率)

- TSW\*time2\_VIL.npy(time2时刻0.5仰角反射率)

- .....(其他时刻0.5仰角反射率)

- NWP ( 基于自主再分析产品计算的环境条件特征 )

- CAPE ( 对流有效位能 )

- TSW\*time1\_CAPE.npy(time1时刻对流有效位能)

- TSW\*time2\_CAPE.npy(time2时刻对流有效位能)

- ..... ( 其他时刻对流有效位能 )

- LCL ( 抬升凝结高度 )

- TSW\*time1\_LCL.npy(time1时刻抬升凝结高度)

- TSW\*time2\_LCL.npy(time2时刻抬升凝结高度)

- .....(其他时刻抬升凝结高度)

- CaseIDn..... ( 其他过程ID号 )

- REGx..... ( 其他区域代码 )

附图 4 雷暴大风数据集实体组织结构

附表 3 雷暴大风过程对应的雷达特征参量列表

产品名称	产品标识	数据单位
0.5° 基本反射率	R05	dBZ
0.5° 基本径向速度	V05	m/s
1.5° 基本径向速度	V15	m/s
组合反射率	CR	dBZ
垂直累积液态水含量	VIL	kg/m <sup>2</sup>
0.5° 基本反射率东西向梯度	RG1	dbz/km
0.5° 基本反射率南北向梯度	RG2	dbz/km

附表 4 雷暴大风过程对应的环境大气条件特征参量列表

名称		单位	简称	含义描述
动力抬升条件特征量	925hPa 风速	m/s	V925	对流层低层的风速能较好地反映出动力和水汽条件，其辐合特征的形成和维持是雷暴活动的判识指标之一。当 925hPa 风速达到急流标准时，这种低空的高速气流具有很强的超地转风特征和明显的日变化，有利于对流或中尺度天气的发展，也导致对流活动呈现出明显的日变化特征。
	0 ~ 1km 垂直风切变	m/s	VP01	风速在 1km 高度以内的变化。雷暴云体中的强烈下降气流区和积雨云的前缘阵风锋区常与低空风切变环境紧密联系，是形成雷暴大风的重要环境因素。强的低空风切变也易使得雷暴演变成龙卷。
	0 ~ 3km 垂直风切变	m/s	VP03	与 0 ~ 1km 垂直风切变相似，也能很好地反映出对流层低层的风在垂直方向的变化程度，是衡量雷暴强弱程度的最常用判识参量之一，也与雷暴大风的潜势密切相关，在中尺度天气预报业务中有重要地位。
	0 ~ 6km	m/s	VP0	作用同 0 ~ 3km 垂直风切变，也是衡量

名称		单位	简称	含义描述
	垂直风切变		6	中尺度天气发展潜势和强弱程度的重要判识指标之一，被业务和科研广泛使用。弱的垂直风切变能增强降水效率，而强的垂直风切变更容易导致中尺度涡旋等系统的形成，反映出具有更显著的上升气流垂直螺旋度，形成向上的扰动气压梯度力，导致气流上升更高，因而增大形成雷暴大风的可能性。
水汽条件特征量	1000hPa比湿	Kg/kg	Q1000	比湿是直接反映水汽条件的特征量之一，因此与短时强降水密切相关。该值常与对流层中层的水汽条件结合分析，若其他对流条件具备、且发现层结具有“上干下湿”或呈“喇叭口”分布特征时，往往就会形成雷暴大风。
	850hPa比湿	Kg/kg	Q850	与1000hPa比湿类似，能反映出对流层低层的水汽状况。
	1000hPa相对湿度	%	RH1000	反映出边界层内的水汽状况，该值越大预示着水汽越饱和。当该值较小而对流层中层的水汽条件较好时，往往更有利于形成雷暴大风。
	850hPa相对湿度	%	RH850	与1000hPa相对湿度相似，是反映对流层低层水汽条件的特征量之一。
	700hPa相对湿度	%	RH700	与850hPa相对湿度相似，是反映对流层中低层水汽条件的特征量之一。
	500hPa相对湿度	%	RH500	衡量对流层中层水汽状况的特征量。该值越大就越有利于形成短时强降水，反之就有利于雷暴大风。
	抬升凝结高度(LCL)	m	LCL	是未饱和湿空气块干绝热上升，刚开始凝结的高度，常用于代表云底的高度，可通过干绝热过程中的温度递减率和露点温度递减率估计。该值取决于近地空气层的湿度大小：湿度大时LCL低，反之则高。抬升凝结高度反映出了对流层内的湿度状况，尤其对流层低层的水



名称		单位	简称	含义描述
				汽条件；该值越低，则低层水汽条件较差时，较有利于发生雷暴大风。
热力/ 能量 条件 特征 量	700hPa 温度	K	T70 0	对流层中低层温度反映的是热力条件。如果 700hPa 上存在着高温区或有暖脊影响，则对流发展的可能性会提高。
	850-500 hPa 垂直 温差	K	DT8 55	表征对流层低层到中层的气温随高度降低程度，随区域、经纬度、季节等的不同而变化，总体而言夏季大而冬季小。从各区域分布来看，华北、西北、西南地区月温度递减率绝对值最大，其他地区则相对较小。它与风雹类强对流天气关系密切，该值越大越有利于形成雷暴大风或冰雹，其高于 25℃ 的指标（不同地区略有差别，也跟季节有关）多被用来作为判识雷暴大风或冰雹可能发生的潜势条件。
	700-500 hPa 垂直 温差	K	DT7 05	与 850-500hPa 垂直温差类似，但能克服可能存在的边界层或低层逆温的不利影响，也对高架雷暴的潜势有更好的参考价值。其高于 16℃ 的指标（不同地区略有差别）多被用来作为判识雷暴大风或冰雹可能发生的潜势条件。
	地面至 850hPa 平均露 点温度	K	TdS fc8 50	地面至 850hPa 的平均露点温度反映的是对流层低层即以下的热力和水汽条件。当该值较高时，就意味着有相对高温高湿的环境，对流层低层及其以下的水汽就容易被抬升而形成降水；此时如果对流层中层较干，则往往容易形成雷暴大风，反之则更易形成短时强降水，发生雷暴大风的可能性略小些。而当该值较小时，如果其他对流条件较好（如 CAPE 较大等），则不容易产生短时强降水而更易形成雷暴大风。
	500hPa 抬升指 数	K	LI5 00	气块从低层 900m 高度沿干绝热线上升，到达凝结高度后再沿湿绝热线上升至 500hPa 时所具有的温度 Ts 与 500hPa

名称		单位	简称	含义描述
				等压面上的环境温度 T500 的差值。反映的是对流层中层的稳定程度, 当该值 <0 时, 大气层结不稳定, 且负值越大, 不稳定程度越大。
	300hPa 抬升指数	K	LI300	与 500hPa 抬升指数类似, 但其反映出了对流层高层的不稳定程度。
	下沉对流有效位能	J/kg	DCAPE	对流发展到一定时期后, 下沉运动将产生并逐渐取代上升运动成为主要特征, 尤其在对流风暴进入成熟阶段后。在风暴体内, 当不饱和空气中有降水蒸发或凝结层的降水融化, 就会有下沉气流产生, 用气块的密度温度随环境气压变化的积分来表征。该特征量是描述雷暴大风可能性及其强度的重要参量之一, 其值的大小与雷暴大风可能性及潜在强度呈很较好的正相关关系, 被广泛应用于风暴特性的预报和研究。
	对流有效位能	J/kg	CAPE	对流有效位能 (Convective available potential energy, CAPE) 是一种与大气中对流现象密切关联的指标, 是强对流天气分析预报的重要参数。CAPE 被定义为抬升凝结高度 (Lifting Condensation Level, LCL) 之上大气正浮力做功所获得的能量。当气块受外力抬升, 气块的重力小于浮力时, 一部分位能可以释放, 由于这部分位能对大气对流有积极的作用, 并可以转化为大气动能, 故称其为对流有效位能。对流有效位能和雷暴大风呈正相关关系。
	对流抑制能量		CIN	对流抑制能量是气块到达自由对流高度的状态曲线与层结曲线围成的面积, 反映的是阻止气块抬升的能量大小。
				CIN 的存在不利于对流的发展, 当该值较大时, 大气处于相对稳定的状态, 不利雷暴产生; 但如果当天的 CAPE 值较



名称		单位	简称	含义描述
				大, 则较大的 CIN 反而易被加热或加湿过程消耗掉, 则会倾向于发展出来对流, 因此发生雷暴大风的可能性会有很大程度的增加。
	地面开始的 CIN	J/kg	SBCIN	反映对流层下层到地面间的对流抑制程度, 地面气块需要突破负浮力区所造成的对流抑制能, 才能继续向上发展深对流。当对流抑制能存在时, 负浮力区通常会分布在地表到自由对流高度之间, 因此它能较好地考虑到虚温订正情况, 从而能较好地获得对流层低层的对流抑制状况, 因为低层大气的干空气平流和地表空气的冷却都会增强对流抑制能, 两者均会减少近地面空气的虚温, 使得垂直虚温分布在低层附近出现逆温结构, 上方虚温较高的空气阻碍下方空气块的上升运动。
	100 mb 以下平均 CIN	J/kg	MLCIN	100 mb 以下平均 CIN 是对近似整个对流层的温度随高度/气压的变化进行了积分计算, 与 SBCIN 类似, 其反映了近似整个对流层的对流抑制程度。
	低层温度平流	°C/hr	ITA	<p>温度平流是指温度从较暖空气向较冷空气方向或较冷空气向较暖空气方向输送, 除能直接引起某地大气热力结构的变化之外, 还可引起其他大气物理属性的变化。根据天气图上风场和温度场的配置, 可以判断温度平流的性质和强度。</p> <p>低层出现暖平流而高空出现冷平流, 就会形成上干冷、下暖湿的不稳定层结特征, 在其他有利条件的配合下, 可产生包括雷暴大风在内的对流性天气。</p> <p>当低层温度梯度越大时, 平流越强; 风速越大, 平流也越强; 或风速与温度等</p>

名称		单位	简称	含义描述
				值线夹角越趋于正交，平流越强，这些都是判断低层温度平流的判据，也是判断雷暴大风潜势的重要参考。
	0℃层高度	m	H0	0℃温度所在高度反映的是环境温度在垂直方向上的分布，对流云的伸展高度受其影响很大，因此与大风、冰雹等的出现密切相关，即该值高度越低时越有利于形成雷暴大风或冰雹（满足其他对流条件后）。
	自由对流高度（LFC）	m	LFC	气块温度与环境温度之差由负值转为正值的转折点的高度，常用埃玛图上的状态曲线与层结曲线相交点所在的高度来表征。 LFC 较高时，如果 0～3 或 0～6 km 垂直风切变很小，则有利于发生雷暴大风。

### （三）冰雹人工智能应用数据集

#### 1. 数据集组织结构和数据内容

冰雹人工智能应用数据集包括训练数据集、测试数据集（A 榜测试数据-TestA, B 榜测试数据-TestB）和数据说明文档三个部分的内容。其中训练数据集和测试数据集均为数据集实体，数据说明文档部分主要包括用于数据读取的每个样本的区域范围（起始经纬度）、量纲单位的说明信息及数据读取程序示例等，数据集组织结构见附图 5。



附图 5 冰雹数据集组织结构

冰雹人工智能应用数据集以冰雹天气过程为单元，每 6 分钟一个样本，数据集共包含 1669 个过程，每个天气过程数据由该过程对应时间窗口范围内的若干样本组成，共计 10 万个样本。

数据集实体的组织结构见附图 6，每个样本包含用于模型输入的冰雹是否发生标签以及和冰雹生消演变相关的 8 个

雷达特征和 4 个水汽条件、抬升条件和不稳定条件特征（基于国产再分析产品计算），数据均为二维空间数据，详细内容见附表 5、6。

### 2. 数据集格式说明

数据内容	数据格式	排列顺序	区域范围	备注
冰雹标签	Excel	—	中东部观测站	
雷达特征参量	numpy 的 npy 格式，二维网格空间数据	自西向东、自北向南排列，第一个数据点位于西北角，最后一个数据点位于东南角	雷达特征范围见 RADA_Format_File 下的 RADA_Format_Ziiii.txt；环境条件特征范围见 NWP_Format_File 下的 NWP_Format_Ziiii.txt。	Python 读取程序： import numpy as np data = np.load(fpath+file)
环境条件特征参量				

### 3. 数据集时间属性

（1）时间范围：5 年时序长度，本次活动不提供年份信息，只提供月份、日期和时刻信息。

（2）时间分辨率：①冰雹是否发生标签数据：时间分辨率为冰雹过程对应雷达站的体扫时间分辨率，约 6 分钟；②雷达特征参量数据：8 种雷达特征参量数据时间分辨率为对应雷达站的体扫时间分辨率，约 6 分钟；③环境大气特征参量：4 种环境大气特征参量时间分辨率为逐小时。

### 4. 数据集空间属性

（1）空间范围：覆盖中东部地区 18 个省份，不同过程选取的雷达站不同，数据的覆盖范围不尽相同。①每个冰雹

过程空间范围为以对应雷达站为中心的周边  $3^{\circ} \times 3^{\circ}$  区域，约  $300\text{km} \times 300\text{km}$ ，覆盖  $301 \times 301$  个网格。②环境大气特征参量：空间范围为以冰雹过程对应的雷达站位置最邻近的环境大气特征参量网格为中心的  $4.98^{\circ} \times 4.98^{\circ}$  区域，覆盖  $167 \times 167$  个网格点。

(2) 空间分辨率：①冰雹标签数据：单点的观测站；②雷达特征参量：每部雷达对应产品的空间分辨率为  $0.01^{\circ} \times 0.01^{\circ}$ ；③环境大气特征参量：空间分辨率为  $0.03^{\circ} \times 0.03^{\circ}$ 。

### 5. 特征值属性

数据类型		特征值	含义描述
冰雹标签数据	冰雹是否发生标签	无	无
雷达特征参量	所有产品	-32768	缺失值
		-1280	雷达无有效回波
环境大气特征参量	所有产品	-9999	缺失值

### 6. 训练数据和测试数据划分

按 80%: 10%: 10% 的比例，考虑年份、季节和区域均匀分布因素，选取 1336 个过程，8.2 万个样本作为训练数据，166 个过程，9238 个样本作为 A 榜测试数据，选取 167 个过程，9172 个样本作为 B 榜测试数据，其中 A、B 榜测试数据只提供雷达特征和环境条件特征，不提供标签数据。

冰雹					
训练集 (80%)		A 榜测试集 (10%)		B 榜测试集 (10%)	
过程数	样本数	过程数	样本数	过程数	样本数
1336	8.2 万	166	9238	167	9172



- **Data**
  - **HA (冰雹代码)**
    - TrainSet (训练数据集)
      - REG1 (区域1代码)
        - CaseID1 (过程1 ID号)
          - LABEL (标签数据)
            - Label-HA.xlsx(样本范围内观测站是否发生冰雹)
          - RADAR (雷达特征)
            - VIL (垂直累积液态水含量)
              - HA\*time1\_VIL.npy(time1时刻垂直液态水含量)
              - HA\*time2\_VIL.npy(time2时刻垂直液态水含量)
              - .....(其他时刻垂直液态水含量)
            - ET45 (45dBZ回波顶高)
              - HA\*time1\_ET45.npy(time1时刻45dBZ回波顶高)
              - HA\*time2\_ET45.npy(time2时刻45dBZ回波顶高)
              - .....(其他时刻45dBZ回波顶高)
          - NWP (基于自主再分析产品计算的环境条件特征)
            - H0 (0度层高度)
              - HA\*time1\_H0.npy(time1时刻0度层高度)
              - HA\*time2\_H0.npy(time2时刻0度层高度)
              - .....
            - H10 (-10度层高度)
              - HA\*time1\_H10.npy(time1时刻-10度层高度)
              - HA\*time2\_H10.npy(time2时刻-10度层高度)
              - .....(其他时刻-10度层高度)
        - CaseIDn..... (其他过程ID号)
      - REGx..... (其他区域代码)

附图 6 冰雹数据集实体的组织结构



表 5 冰雹过程对应的雷达特征

产品名称	产品标识	数据单位
组合反射率	CR	dBZ
垂直累积液态水含量	VIL	kg/m2
20dBZ 回波顶高	ET20	km
20dBZ 回波底高	EB20	km
30dBZ 回波顶高	ET30	km
30dBZ 回波底高	EB30	km
45dBZ 回波顶高	ET45	km
45dBZ 回波底高	EB45	km

表 6 冰雹过程对应的环境条件特征

名称	单位	简称
温度 0℃ 层高度	m	H0
温度-10℃ 层高度	m	H10
温度-20℃ 层高度	m	H20
湿球温度 0℃ 层高度	m	WBZ