

中鋼人工智慧挑戰賽-字元辨識

競賽說明

中鋼 50 年來致力於鋼鐵產業，有舉足輕重的地位，為世界知名鋼鐵廠。隨著人工智慧科技在許多領域已超越人類智能，中鋼放眼下一個 50 年的發展，已定調「精緻鋼鐵」、「綠能產業」是我們經營發展的兩大主軸，透過「智能化」讓中鋼再精進升級，希冀保持領先的競爭力。為慶祝中鋼 50 周年，特別舉辦「人工智慧挑戰賽」，挑戰賽的主題與智能化鋼胚的管理有密切的相關。

本次參賽者要以人工智慧演算法，實現鋼胚上印刷或手寫序號的自動判別，將有助於問題源的返溯及品質的控管。參賽者除了要能辨識正常印刷之序號（含數字及英文字母），還得針對手寫序號、序號上下顛倒、重複印刷序號、或序號模糊等情況做有效的處理，以發展出適合實際產線運作的序號辨識系統。

歡迎各路投入人工智慧好手，參加本次競賽，我們等著您來挑戰！

報名規範

1. 參賽者須為年滿 18 歲之本國籍國民，學生或業界人士皆可報名參加（中鋼公司及集團公司現職同仁、接觸題目之中山大學人員及趨勢科技公司現職員工除外）。
2. 參賽隊伍人數 1 至 3 人。
3. 參賽者只能加入一隊，一旦報名後，不得更換隊伍。
4. 報名截止後，不可再變更團隊成員名單及人數。
5. 比賽期間，隊伍不可合併與分割。
6. 參賽者報名參加本競賽即同意主辦單位於頒獎典禮當天拍攝（包含攝影、錄影等），並得將相關影片或照片等運用在本次競賽之新聞稿及競賽宣傳文宣中，參賽者不得異議。

領獎資格與方式

1. 比賽排名前五名之隊伍，每隊至少推派一員參加競賽發表會，報告實做方法、模型與運作流程，並接受提問。經確認無違反競賽規定，才能實際獲獎。如有遇違反規定或放棄領獎情事，則依序遞補。
2. 以匯款方式發放獎金，得獎隊伍應推派有本國新台幣帳戶之隊員代表受領獎金，並以該隊員為中華民國稅務申報人，依規定簽署包含稅務、個資使用等相關之文件。

獎項說明

- 冠軍：新台幣 15 萬元、獎狀乙紙。
 - 亞軍：新台幣 8 萬元、獎狀乙紙。
 - 季軍：新台幣 5 萬元、獎狀乙紙。
 - 佳作：新台幣 1 萬元。
 - 佳作：新台幣 1 萬元。
 - 中山大學評審團獎：依參加競賽發表會五組隊伍所採用解題方法之創意、及發表會報告內容，擇一隊加發新台幣 2 萬元。
-

評分標準

1. 於 2021/10/4(一)競賽開始時會公佈 Public Dataset，其中包含訓練集與測試集。
 2. 參賽隊伍每隊對於 Public Dataset 與 Private Dataset 預測結果，每日可提交結果之次數合計至多 2 次。
 3. 提交檔案請使用.csv 檔，內容格式需符合比賽格式規定。
 4. 比賽過程中，Leaderboard 系統會對所提交之測試集結果依據評分方式進行評測。
 5. 依各隊伍所有上傳結果中最優之分數，提供各隊伍之排名。
 6. 主辦單位將於 2021/11/8(一)提供 Private Dataset 的加密壓縮檔下載，並於 2021/11/9(二)14:00 於官網公佈解壓縮密碼。
 7. 參賽隊伍可於 2021/11/9(二)當天 14:00 到 22:00 期間，針對 Private Dataset 的預測至多提交 2 次結果，逾期不再接受上傳。
 8. 最終比賽排名會依據上傳最優之 Private Dataset 的結果分數排名，數值越小者排名越高。當分數一樣時，則依上傳時間先後來排名。先上傳者排名在前。
-

評分方式

$$\text{score} = \sum_{i=1}^N \text{err}(\text{pred_str}_i, \text{label_str}_i) + \frac{1}{N} \sum_{i=1}^N \text{lev}(\text{pred_str}_i, \text{label_str}_i) / \max(|\text{pred_str}_i|, |\text{label_str}_i|)$$

- N ：總測試資料筆數，包含列印或手寫序號的扁鋼胚圖片
- label_str_i ：第 i 筆圖片資料標記之序號，不含空白
- pred_str_i ：第 i 筆圖片資料預測結果之序號（之後與標記序號的比較時一律不考慮英文字母大小寫的差異）
- $\text{err}(\text{pred_str}_i, \text{label_str}_i)$ ：當 pred_str_i 與 label_str_i 完全一樣時為 0；否則為 1
- $|\text{pred_str}_i|$ 、 $|\text{label_str}_i|$ ：分別為第 i 筆預測結果之序號與標記之序號的長度

- $lev(pred_str_i, label_str_i)$ ：預測序號 $pred_str_i$ 與標記序號 $label_str_i$ 之 Levenshtein 距離(https://en.wikipedia.org/wiki/Levenshtein_distance/)，Levenshtein 距離函數之 reference implementation 可參考 (<https://gist.github.com/marty1885/4db962bbdf5657756a6a4cf9b94ab4#file-score-py-L70>)
- 上述採用的計分方式，著重於圖片的序號是否能夠完全偵測正確。當隊伍提交的結果中序號偵測完全正確的圖片張數一樣時，評量部分序號正確比例的 Levenshtein 距離評分才會對排名有實質影響。

競賽規則

1. 參賽隊伍可以使用自行製作之資料或額外開源資源來增進模型訓練結果，惟務必使用 Machine Learning 來進行辨識，禁止使用任何人工標記產生上傳結果。若有使用額外資料，需於比賽結束後提供來源資訊。如有違反資料之使用規範，參賽者將喪失領獎與參賽資格。如有爭議，主辦單位保有最終決定權。
2. 禁止使用非開源 Auto Machine Learning 相關之自動建模服務。
3. 禁止使用現成之 OCR 模型/系統進行辨識。
4. 禁止使用非開源之 classification/recognizing 服務 (例如：IBM Watson/Azure Cognitive Services)。
5. 競賽結束後前五名之隊伍須繳交實作程式，包含(但不限於)：前處理程式碼、訓練程式碼、辨識程式碼、各項參數之設定(包括訓練權重)、執行環境說明。主辦單位將依需求驗證程式之可執行性與可重現性，必要時得請求參賽隊伍協助重現實作程式的結果。上述驗證結果將影響最後能否獲獎之資格。
6. 資料集僅限於此次比賽使用，參賽者不得為自己或他人利益而洩漏或交付資料集予非參賽者。
7. 不可私下共享程式及特徵值。
8. 如有需要，主辦單位有權在比賽途中調整資料集。
9. 如有下列情事，主辦單位得無需告知參賽者，逕行取消參賽者資格或領獎資格：
 - 已有具體事證，所屬隊伍有任何抄襲、作弊、或詐欺等行為。
 - 已有具體事證，所屬隊伍有侵害他人智慧財產權之情事。
 - 已有具體事證，所屬隊伍有對 Leaderboard 系統進行攻擊。
 - 已有具體事證，所屬隊伍影響其他參賽隊伍導致不公平事例發生。
 - 已有具體事證，所屬隊伍違反本比賽活動辦法、或「T-Brain AI 實戰吧平台服務」使用條款、或「中鋼人工智慧挑戰賽-字元辨識」參賽者使用條款。
10. 主辦單位保有對活動與競賽規則解釋及裁決的權利。
11. 比賽排名前五名之隊伍需於比賽結束後，2021/11/13 之前將實作程式及設計文件釋出至公

開程式碼託管平台，完成程式碼開源分享。程式碼必須以 [AGPL](<https://www.gnu.org/licenses/agpl-3.0.en.html>) 或者 [SSPL](<https://www.mongodb.com/licensing/server-side-public-license>) 釋出至公開程式碼託管平台 (例如 [GitHub](<https://github.com/>) [SourceForge](<https://sourceforge.net/>))。未於期限前釋出並繳程式碼連結者將失去領獎資格。

12. 主辦單位將選取排名前五名之隊伍，參加 2021/11/27(六)於中鋼公司舉辦之競賽發表會，上台分享實作內容與方法。報告內容應包括 (但不限於) 實做方法、模型與運作流程。
13. 參賽隊伍應依前述規定繳交作品通過初審，以及競賽發表會複審，確認無違反競賽規定之情事後方可獲獎。

競賽時程

項目	時程	說明
報名時程	2021/9/13 - 2021/10/3	開放報名。
比賽時程	2021/10/4 - 2021/11/9	競賽隊伍可上傳答案，評分系統將開始針對每個隊伍的答案進行評分。
競賽 Public Dataset 訓練集與測試集下載	2021/10/4 - 2021/11/9	競賽隊伍可於比賽正式開始同日下午下載 Public Dataset 訓練集及測試集，並上傳答案，每一日曆日可上傳之次數上限為 2 次。
競賽測試集 Private Dataset 下載	2021/11/8	<ol style="list-style-type: none"> 1. 2021/11/8 凌晨 00:00 起競賽隊伍可下載競賽測試集 Private Dataset 的加密壓縮檔。 2. 解壓縮密碼於 2021/11/9 下午 14:00 於官網公告。
競賽測試集 Private Dataset 預測	2021/11/9	測試集預測結果總上傳次數上限為 2 次。上傳答案需依照規定之格式，以避免上傳失敗的問題。答案上傳期限為 2021/11/9 下午 14:00~22:00，逾期不再接受上傳。
公布 Private Leaderboard	2021/11/10	官網公佈 Private Dataset 的成績排名及分數。
作品繳交	2021/11/10 - 2021/11/13	<ol style="list-style-type: none"> 1. 排名前五名之隊伍依規定繳交作品及完成程式碼開源分享。 2. 進行初審，若有不符規定者則進行相關遞補程序。
公布參加競賽發表會組別名單	2021/11/19	於 T-brain 平台公布入圍名單。
競賽發表會暨	2021/11/27 (六)	1. 入圍隊伍進行作品報告，評審團進行作品複審。

頒獎典禮		2. 進行中山大學評審團獎之評審。 3. 頒獎典禮程序及內容，將另行公布。
------	--	--

主辦單位：中國鋼鐵股份有限公司

中國鋼鐵股份有限公司(中鋼)為國內鋼鐵產業的龍頭，是一家勇於創新、執行力強的公司，以「追求成長，持續節能環保及價值創新，成為值得信賴的全球卓越鋼鐵企業」為願景，積極落實「團隊、企業、踏實、求新」中鋼精神，以及「增進社會福祉、落實實際績效、發揮群體力量、講求人性管理」的經營理念。除持續深耕鋼鐵本業外，中鋼發揮經營綜效，進行多角化經營，目前轉投資事業共有 28 家，業務範圍涵蓋鋼鐵核心、工程事業、工業材料、物流貿易，以及服務投資等五大事業群，形成中鋼集團，並定位為「總部設於臺灣，以亞洲為核心，集鋼鐵與材料、工程及服務、礦業與資源的重環保與節能的集團」，朝向資源節約與環境友好的全球卓越鋼鐵企業努力。

協辦單位：國立中山大學人工智慧研究暨產業推廣中心

國立中山大學屹立南台灣，為一所人文與科技並重的精緻研究型大學，具備人才培育與學術研究的能量與競爭優勢，蔚然成為南臺灣學術重鎮，不僅躋身全國重點研究型大學，也立足國際知名一流大學行列。「人工智慧研究暨產業推廣中心」，整合校內的人工智慧相關研究、教學及產學經驗，協助校內跨領域之研究聯繫與合作；並將校內的人工智慧研究量能與產業結合，獲得更大更廣之研究成果及業界的產能提升。

贊助單位：趨勢科技

趨勢科技為資訊安全解決方案全球領導廠商，致力建立一個安全的資訊交換世界。我們專為消費者、企業及政府機構設計的創新解決方案，能為資料中心、雲端工作負載、網路、端點裝置提供多層式安全防護。我們的產品皆彼此整合、共享威脅情報，提供環環相扣的威脅防禦與集中式的掌握及調查能力，實現更好、更快的防護。趨勢科技全球共超過 6,000 名員工，遍及 50 個國家，並擁有全世界最先進的全球威脅研究及情報，是企業保護連網環境的最佳夥伴。

Q&A

若對於比賽有任何問題，歡迎來信至 CSCAlcup@gmail.com。

客服時段：週一至週五 09:00-12:00，14:00-17:00