

**Supplementary Appendix S2. Inclusion and exclusion criteria, as provided to researchers conducting abstract and full-text screening.**

Inclusion	Exclusion
<ol style="list-style-type: none"> <li>1. Is published in the English language</li> <li>2. Is a peer-reviewed primary research article</li> <li>3. Is not a retracted article</li> <li>4. Utilises automated machine learning               <ol style="list-style-type: none"> <li>a. Google, Microsoft, Apple, Sony, Amazon.</li> <li>b. JADBio, H2O.ai.</li> <li>c. Auto-WEKA, auto-Sklearn, autoKeras, TPOT, TuPaQ, ATM, Automatic Frankenstein, ML-Plan, Autostacker, AlphaD3M, Collaborative filtering <i>List is not exhaustive! Use of any platform that automates any element of ML engineering is eligible.</i></li> </ol> </li> <li>5. Algorithm is applied in a clinical context               <ol style="list-style-type: none"> <li>a. Human patients</li> <li>b. Related to disease diagnosis, management, or prognosis</li> </ol> </li> </ol>	<ol style="list-style-type: none"> <li>1. Is not published in the English language</li> <li>2. Is not a peer-reviewed primary research article               <ol style="list-style-type: none"> <li>a. Review (systematic, scoping...)</li> <li>b. Grey literature (e.g. preprint)</li> <li>c. Conference paper</li> <li>d. Commentary</li> <li>e. Case report</li> </ol> </li> <li>3. Is a retracted article               <ol style="list-style-type: none"> <li>a. Misleading</li> <li>b. Fraud</li> <li>c. Unethical</li> </ol> </li> <li>4. Does not utilise automated machine learning <i>No mention of autoML platform.</i></li> <li>5. Algorithm is not applied in a clinical context               <ol style="list-style-type: none"> <li>a. No aspect relevant to patients</li> </ol> </li> </ol>

**Exclusion criteria are explicitly ranked.** Papers should be evaluated first by criterion 1, then 2, then 3 etc. When full-text screening, the first number ‘failed’ will need to be recorded. In Rayyan, this enables rapid exclusion by searching for e.g. ‘Italian’, ‘Japanese’ (non-English language), then ‘conference’, ‘comment’, ‘case report’ (non-peer reviewed research article).

**Examples are not exhaustive!** All published, English, peer-reviewed research articles testing any form of automated machine learning in a clinical context should be included. It may be useful to enter exclusion examples in the Rayyan search tool.