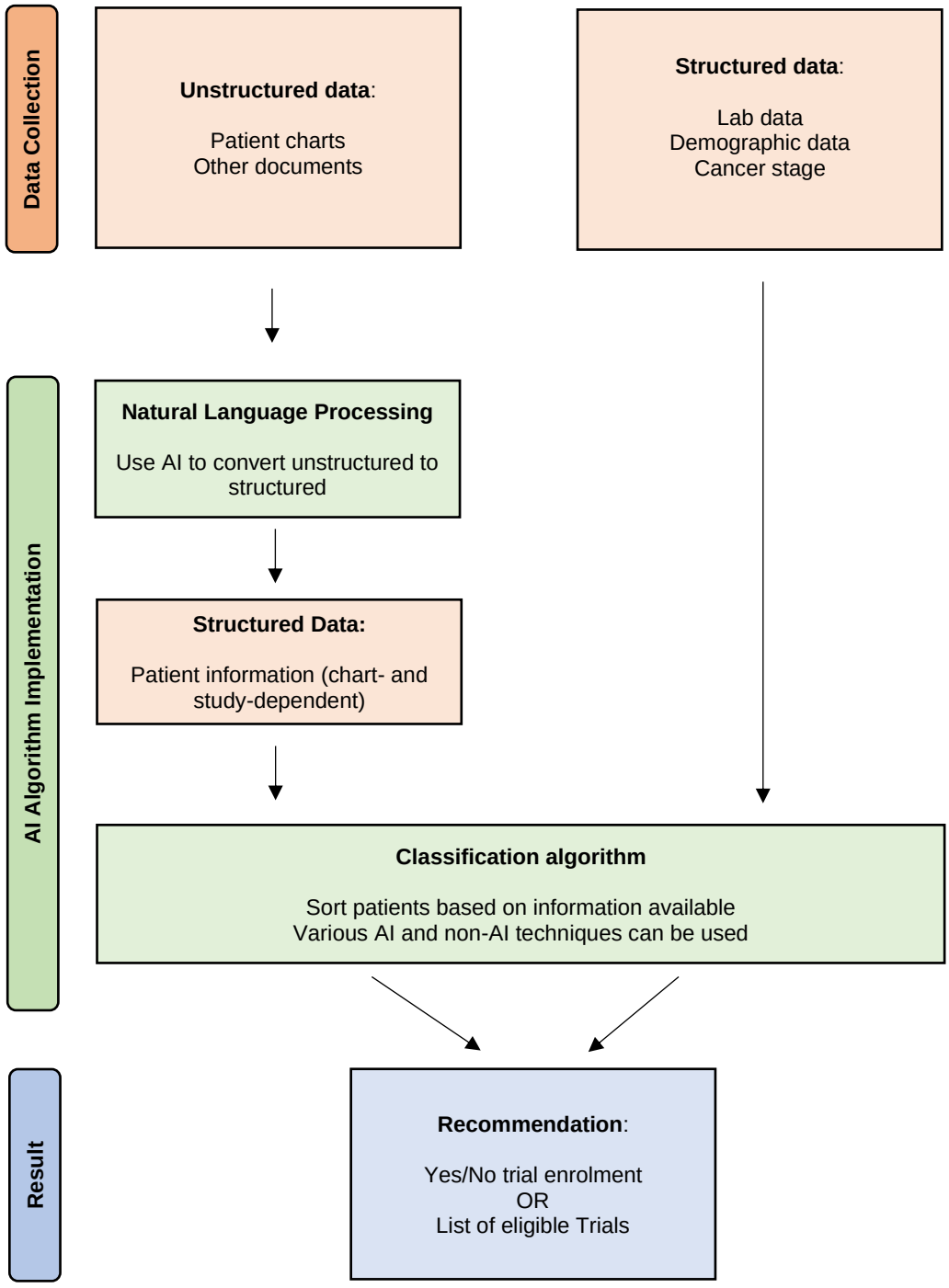
Supplementary Figure 1. Example Clinical Trial Enrolment Algorithm Pipeline



Supplementary Box 1. Search Strategies

PubMed (524)
(AI[tw] OR ar\ficial intelligence[tw] OR ar\ficial intelligence[mh] OR machine learning[tw] OR deep
learning[tw] OR transfer learning[tw] OR data mining[tw] OR natural language processing[tw] OR knowledge
acquisi\on[tw] OR machine intelligence[tw] OR computa\onal intelligence[tw])
AND
(clinical trial[pt] OR Controlled Clinical Trial[pt] OR clinical trial*[tw] OR clinical stud*[tw] OR cancer trial*[tw]
OR clinical trials as topic[mh])
AND
(cancer*[tw] OR oncology[tw] OR neoplasms[mh])
AND
(matching[tw] OR enrollment[tw] OR enrolment[tw] OR recruitment[tw] OR eligibility[tw] OR eligible[tw] OR
par\cipa\on[tw] OR selec\on[tw] OR admission[tw] OR enlistment[tw] OR acceptance[tw] OR registra\on[tw] OR registered[tw])

EMBASE (639) & Cochrane (10)
(exp ar\ficial intelligence/ or AI.mp. or ar\ficial intelligence.mp. or machine learning.mp. or deep learning.mp. or transfer learning.mp. or data mining.mp. or natural language processing.mp. or knowledge
acquisi\on.mp. or machine intelligence.mp. or computa\onal intelligence.mp.)
and
(exp clinical trial/ or exp controlled clinical trial/ or controlled study/ or randomized controlled trial/ or
clinical trial*.mp. or “clinical trial (topic)”/ or cancer trial*.mp.)
and
(cancer*.mp. or exp oncology/ or oncology.mp. or exp neoplasm/)
and
(matching.mp. or enrollment.mp. or enrolment.mp. or recruitment.mp. or eligibility.mp. or eligible.mp. or
par\cipa\on.mp. or selec\on.mp. or admission.mp. or enlistment.mp. or acceptance.mp. or registra\on.mp.
or registered.mp.)

Supplementary Table 1. Group of Patients, Within Included Trials

Study	Group	Description
Beck et al, 2020[15]	Group 1	Phase III study of alpelisib plus fulvestrant in men and postmenopausal women with advanced breast cancer
	Group 2	Phase II study of letrozole with or without alpelisib or buparlisib, for neoadjuvant treatment of postmenopausal women
	Group 3	Phase III study of buparlisib with fulvestrant in postmenopausal women
	Group 4	Phase III study of ribociclib in combination with fulvestrant for treatment of men and postmenopausal women
Calaprice-Whitty et al, 2020[16]	Group 1	Breast cancer trial, with good enrolment
	Group 2	Non-small cell lung cancer, with moderate enrolment
	Group 3	Non-small cell lung cancer, with no enrolment
Cesario et al, 2021[17]	Group 1	Breast cancer
	Group 2	Lung cancer
Haddad et al, 2021[20]	Group 1	Patients with attributes manually verified by humans
	Group 2	Patients without any processing by humans
Ni et al, 2015[22]	Group 1	Retrospective workload evaluation
	Group 2	Physician chart review
Zeng et al, 2014[23]	Group 1	Adapted gene mention disambiguation component
	Group 2	Adapted gene mention disambiguation component, to identify genes as selection criteria

Supplementary Box 2. Description of AI Algorithms, for Included Studies

The methodologies of the AI algorithms used are described in detail below, when available. It is important to note that for Alexander et al.[14], Beck et al.[15], and Haddad et al.[20], no further details can be given as the algorithm used (WCTM) is proprietary and thus exact details of model structure and methods cannot be obtained. Similarly, no further details can be provided about Mendel.ai, used by Calaprice-Whitty et al.[16]

Cesario et al.[17]
While the DRA is developed in-house and in collaboration with an Italian initiative, and the authors clearly state that their matching algorithm uses machine learning, no further details were provided, either in the paper or online.

Cuggia et al.[18]
The model used by Cuggia et al. first employs a series of boolean operations to determine if pre-coded patient data meets given criteria. The paper is not clear on whether or not this data is coded entirely manually, or if there is a level of automation involved. In cases where some information about the patient is missing, additional criteria can be implemented to deal with these edge cases.[31] By building a complex list of criteria, the algorithm is able to emulate the decision-making process to sort patients into included and excluded groups.

Delorme et al.[19]
Word2Vec, a neural network-based NLP model, was used to build the word embedding. This was followed by UMAP clustering and dimension reduction of the dictionary to form semantic clusters. For a given set of patient data, relevant dictionary clusters were determined and then a random forest model was used on the clusters to predict trial eligibility.

The GitHub link for this project can be found at
https://github.com/DITEP/NLP_for_ScreenFail_prediction

Meystre et al.[21]
An Apache UIMA-based NLP algorithm was used to extract relevant clinical information from patient records. A support vector machine (SVM) was then used to cluster the extracted clinical information for each patient into eligibility criteria, which could be assessed to determine the eligibility of that patient for a given particular study.

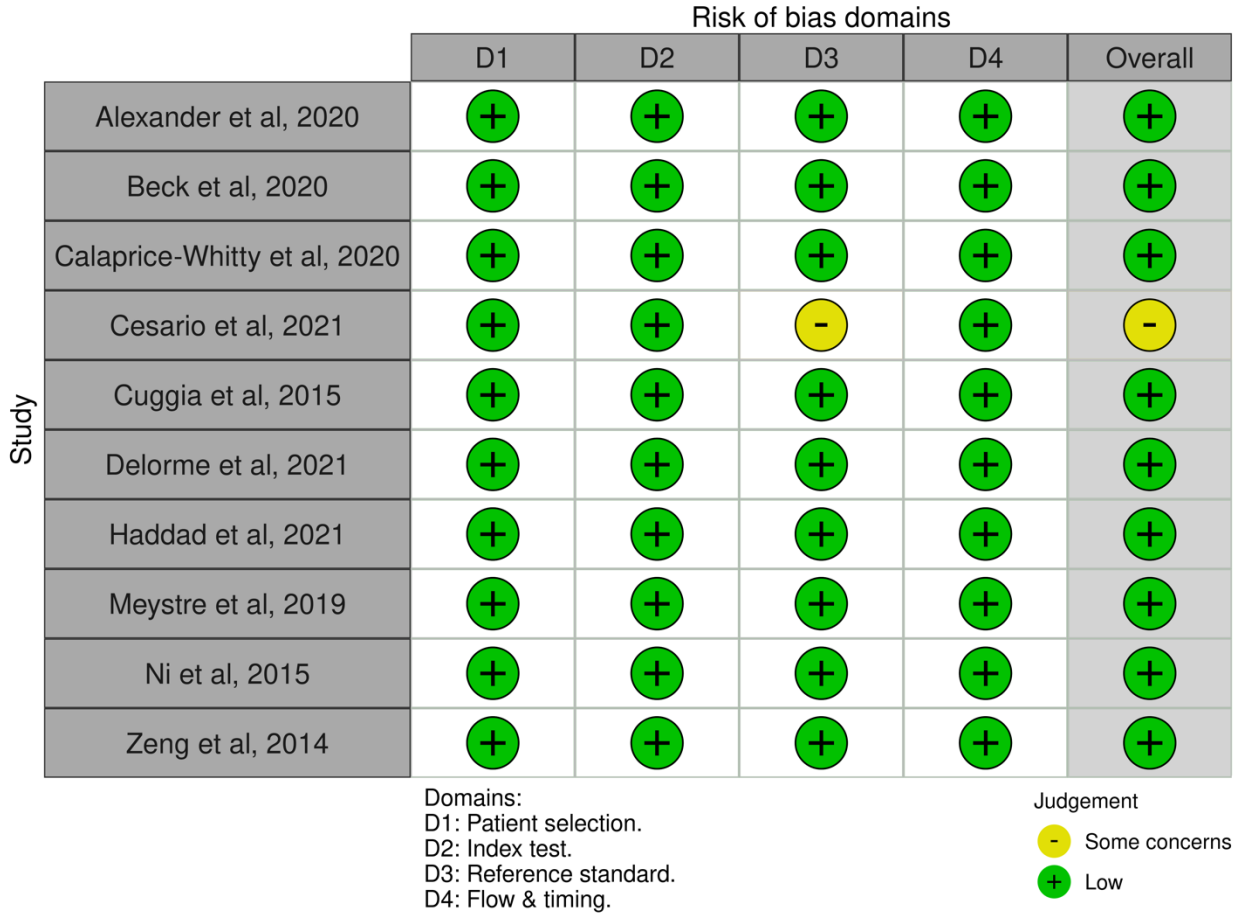
Ni et al.[22]
An in-house NLP algorithm was designed, composed of several steps. First, EHR information was filtered based on demographic criteria for the particular trial. Studies that passed this step were processed using well-established medical term dictionaries such as SNOMED, to extract relevant clinical information and associate it with particular identifiers. This process included negation detection using a method based on the NegEx algorithm as well as Apache cTAKES for clinical information extraction. Identifiers for each patient were stored and compared with vectors similarly extracted from the trial description itself. Comparisons between the two groups

of identifiers were used to generate a similarity score, reflecting how likely a patient is to be eligible.

Zeng et al.[23]

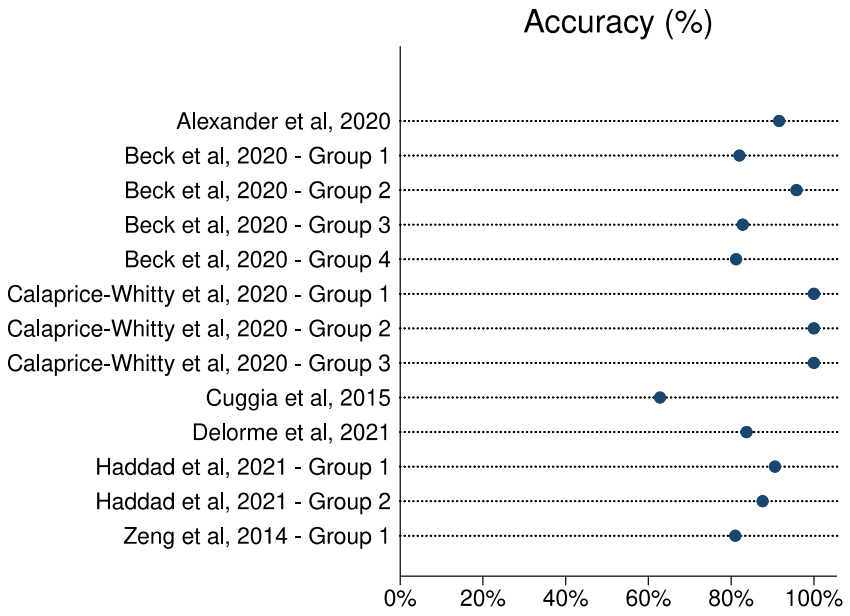
This paper used a modified version of an algorithm from Wu et al.[32], which extracts gene information from clinical trial descriptions and documentation. This classifier was composed of a series of CVMs which extracted genetic information from documents and clustered it into various categories of information. Zeng et al. re-trained this model on a new, nonoverlapping set of clinical trials to determine if the algorithm is generalizable to the task of automatically determining enrolment criteria for clinical trials.

Supplementary Figure 2. Study Quality of Included Studies

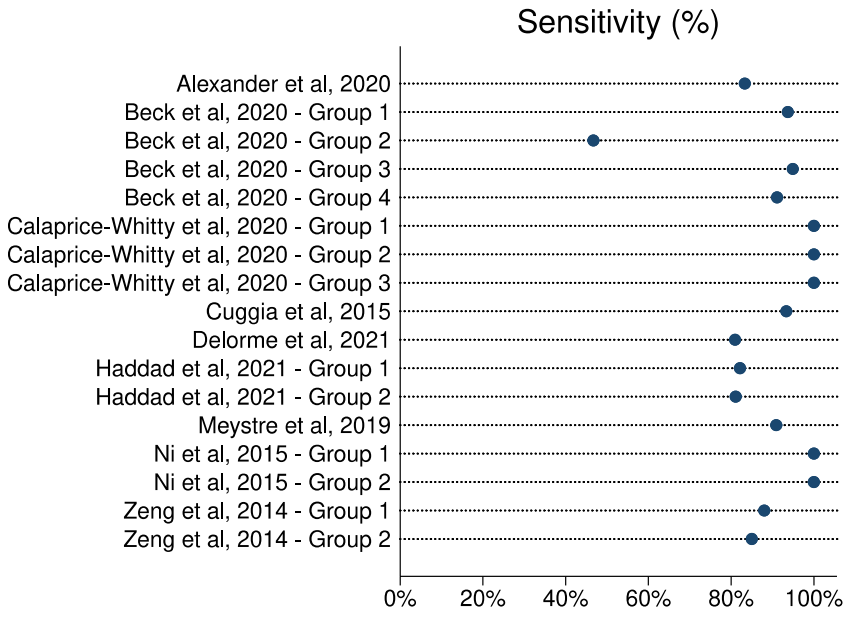


Supplementary Figure 3. Predictive Ability of Artificial Intelligence. **A** Accuracy **B** Sensitivity **C** Specificity

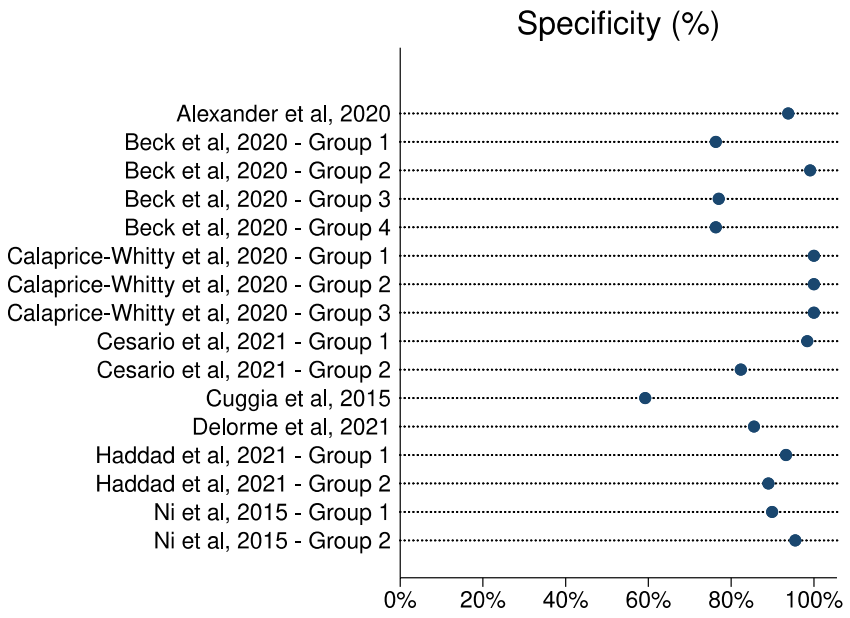
A



B

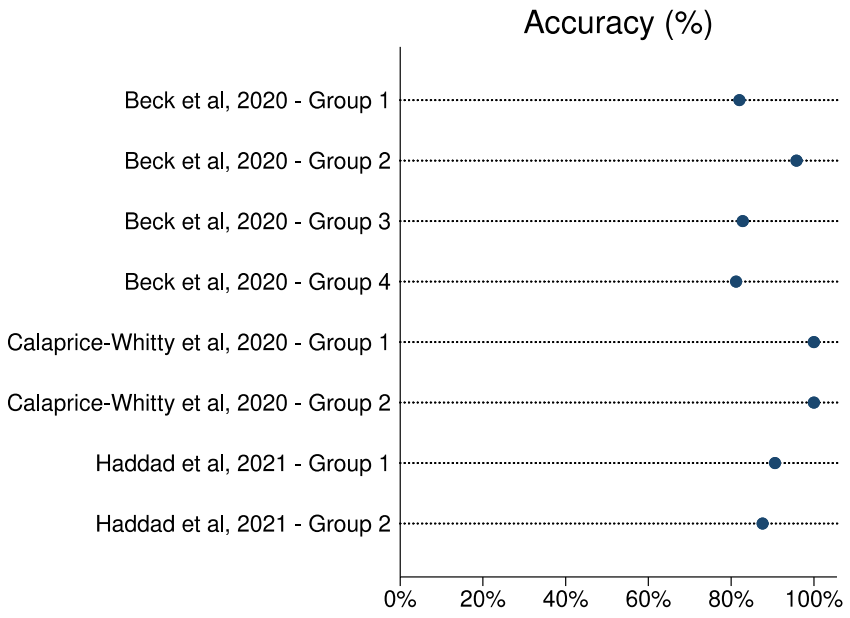


C

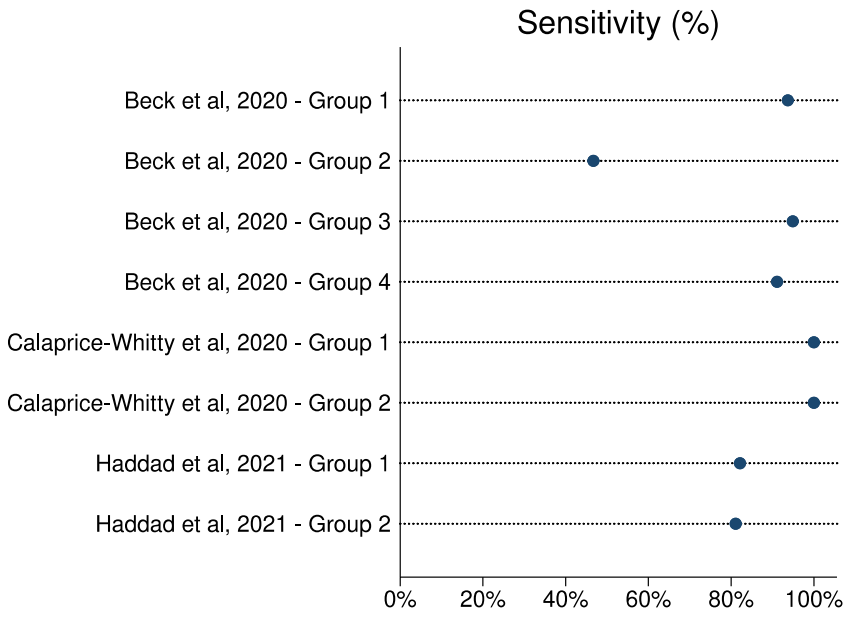

Specificity (%)
Alexander et al, 2020
Beck et al, 2020 - Group 1
Beck et al, 2020 - Group 2
Beck et al, 2020 - Group 3
Beck et al, 2020 - Group 4
Calaprice-Whitty et al, 2020 - Group 1
Calaprice-Whitty et al, 2020 - Group 2
Calaprice-Whitty et al, 2020 - Group 3
Cesario et al, 2021 - Group 1
Cesario et al, 2021 - Group 2
Cuggia et al, 2015
Delorme et al, 2021
Haddad et al, 2021 - Group 1
Haddad et al, 2021 - Group 2
Ni et al, 2015 - Group 1
Ni et al, 2015 - Group 2
0%
20%
40%
60%
80%
100%

Supplementary Figure 4. Predictive Ability of Artificial Intelligence, of Studies Included in Meta-Analysis **A** Accuracy **B** Sensitivity **C** Specificity **D** Positive Predictive Value **E** Negative Predictive Value

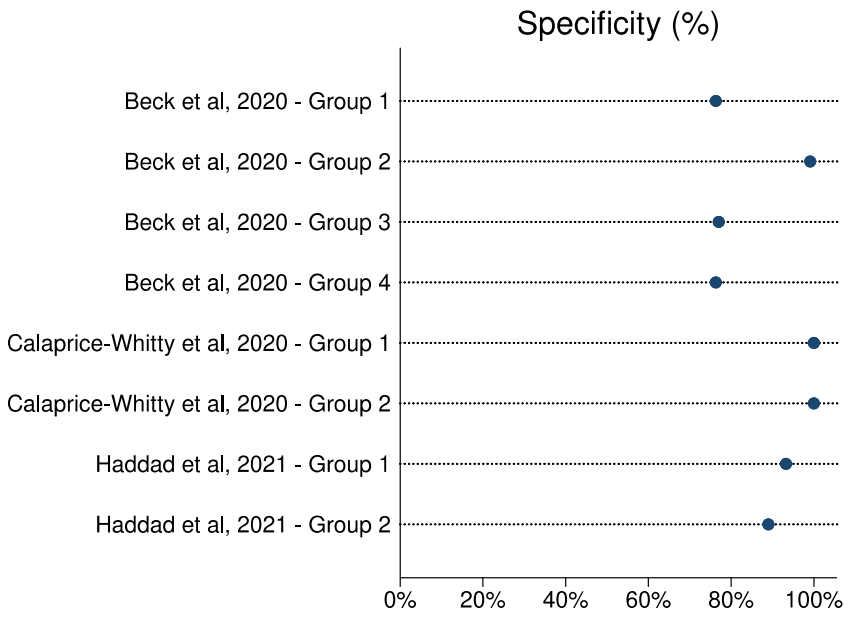
A



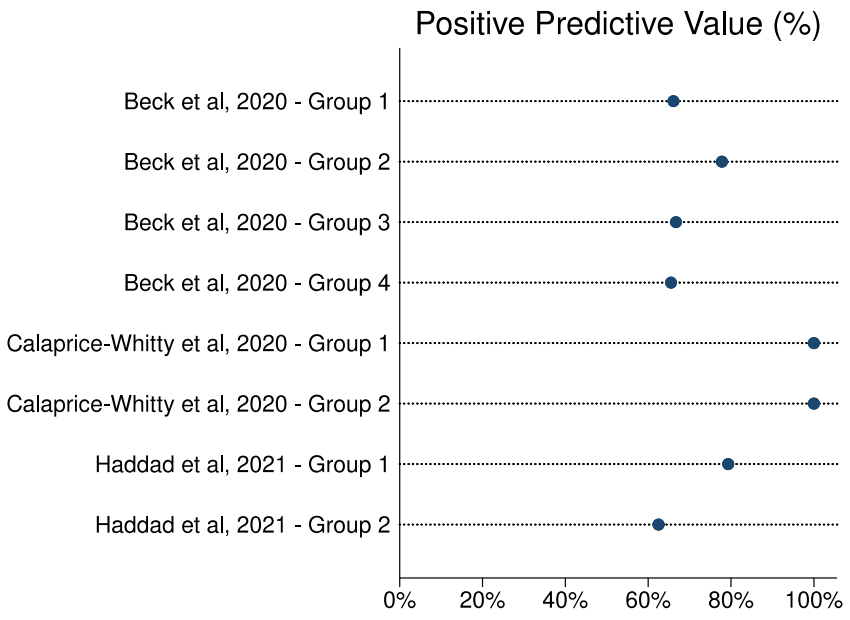
B



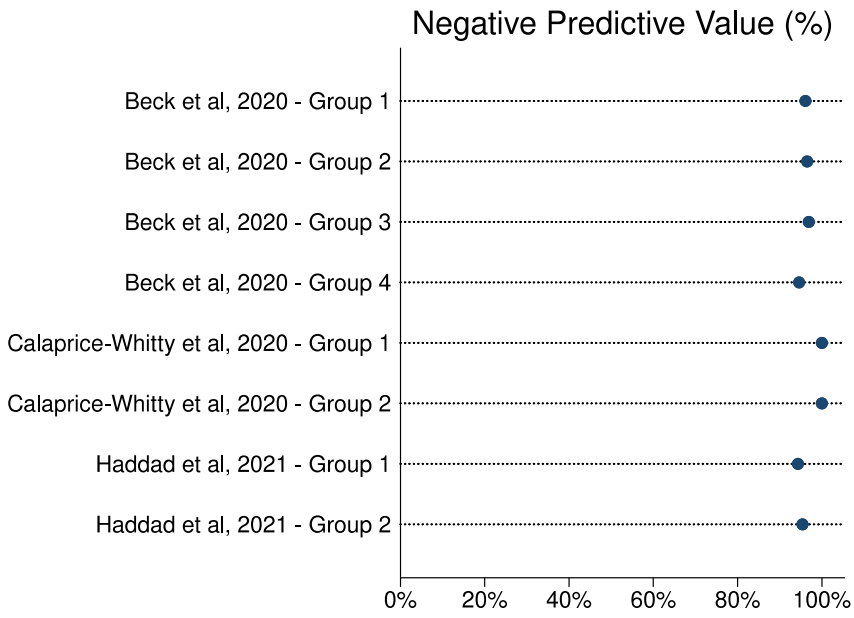
C


Specificity (%)
Beck et al, 2020 - Group 1
Beck et al, 2020 - Group 2
Beck et al, 2020 - Group 3
Beck et al, 2020 - Group 4
Calaprice-Whitty et al, 2020 - Group 1
Calaprice-Whitty et al, 2020 - Group 2
Haddad et al, 2021 - Group 1
Haddad et al, 2021 - Group 2
0%
20%
40%
60%
80%
100%

D


Positive Predictive Value (%)
Beck et al, 2020 - Group 1
Beck et al, 2020 - Group 2
Beck et al, 2020 - Group 3
Beck et al, 2020 - Group 4
Calaprice-Whitty et al, 2020 - Group 1
Calaprice-Whitty et al, 2020 - Group 2
Haddad et al, 2021 - Group 1
Haddad et al, 2021 - Group 2
0%
20%
40%
60%
80%
100%

E


Negative Predictive Value (%)
Beck et al, 2020 - Group 1
Beck et al, 2020 - Group 2
Beck et al, 2020 - Group 3
Beck et al, 2020 - Group 4
Calaprice-Whitty et al, 2020 - Group 1
Calaprice-Whitty et al, 2020 - Group 2
Haddad et al, 2021 - Group 1
Haddad et al, 2021 - Group 2
0%
20%
40%
60%
80%
100%

Supplementary Table 2. Comparison of Positive Predictive Value, of Industry-Developed and In-House Algorithms

Industry-Developed		**In-House Algorithms**	
Alexander et al, 2020[14]	76.5%	Cuggia et al, 2015[18]	21.2%
Beck et al, 2020 – Group 1[15]	66.1%	Delorme et al, 2021[19]	78.7%
Beck et al, 2020 – Group 2[15]	77.8%	Meystre et al, 2019[21]	89.7%
Beck et al, 2020 – Group 3[15]	66.7%	Ni et al, 2015 – Group 1[22]	12.6%
Beck et al, 2020 – Group 4[15]	65.5%	Ni et al, 2015 – Group 2[22]	35.7%
Calaprice-Whitty et al, 2020 – Group 1[16]	100.0%	Zeng et al, 2014 – Group 1[23]	55.0%
Calaprice-Whitty et al, 2020 – Group 2[16]	100.0%	Zeng et al, 2014 – Group 2[23]	69.0%
Calaprice-Whitty et al, 2020 – Group 3[16]	100.0%		
Haddad et al, 2021 – Group 1[20]	79.3%		
Haddad et al, 2021 – Group 2[20]	62.5%		