Language-Based Detection of Depression with Machine Learning: Systematic Review and Meta-Analysis

Supplementary Tables and Figures

Table S1. PRISMA checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Title p.1
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Abstract p. 2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Intro p. 3
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Intro p. 3
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Methods pp. 4-5
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Methods p. 4
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Methods p. 4
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Methods p. 4
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Methods pp. 5-6
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Methods pp. 6-7
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Methods pp. 6-7
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Methods pp. 5-6
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Methods pp. 6-7
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Methods pp. 7-8
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Methods pp. 7-8
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Methods pp. 7-8
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Methods pp. 7-8
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Methods pp. 7-8

Section and Topic	Item #	Checklist item	Location where item is reported
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Methods pp. 7-8
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Methods pp. 7-8
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Methods pp. 7-8
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Results p. 8
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Figure 1
Study characteristics	17	Cite each included study and present its characteristics.	Results pp. 8-11 and Tables 1 and 2
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Table S2
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Tables 1 and 2, all figures
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Results pp. 8-12
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Results pp. 10-12
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Results pp. 10-12
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Results pp. 10-11
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Results pp. 10-12
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Results pp. 10-12
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Discussion pp. 12-13
	23b	Discuss any limitations of the evidence included in the review.	Discussion pp. 13-14
	23c	Discuss any limitations of the review processes used.	Discussion pp. 13-14
	23d	Discuss implications of the results for practice, policy, and future research.	Discussion p. 14-15
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Methods p. 4

Section and Topic	Item #	Checklist item	Location where item is reported
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Methods p. 4
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Methods p. 4
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Acknowledgements p. 16
Competing interests	26	Declare any competing interests of review authors.	Competing Interests p. 16
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	p. 8 and p. 16

Table S2. Quality Assessment of Studies Included in the Systematic Review and Meta-Analyses

Study	Sample Size	Balanced Sample	Ground Truth
Abilkaiyrkyzy 2024[49]	-	+	+
Aloshban 2021[50]	-	+	+
Antoniou 2022[51]	+	+	-
Banerjee 2021[52]	+	+	+
Boian 2025[53]	+	+	+
Burkhardt 2022[54]	+	?/-	+
Cao 2025[55]	-	+	+
Chen 2024[33] (CMDC) (Same as Zou 2023)	-	+	+
Chen 2024[33] (EATD) (Same as Shen 2022)*	+	+	+
Cohen 2023[56]	-	+	+
Cook 2016[57]	+	+	+
deHond 2024[58]	+	-	+
Demiroglu 2020[27]	-	+	+
Demiroglu 2020[27]	-	+	+
Gao 2024[59]	+	+	+
Guo 2024[60]	+	+	+
Hayati 2022[38]	-	+	+
He 2022 (same data as Yuan 2021)[61]	+	+	+
Howes 2014[62]	+	?	+
Iyortsuun 2024[28] (Same as Shen 2022)	+	+	+
Joharee 2023[63]	+	+	+
Krishnamurti 2022[64]	+	-	+
Li 2023[65]	+	+	+
Liu 2022[66]	+	+	+
Munthuli 2023[67]	-	N/A	+
Nobles 2018[68]	-	+	+
Howes 2014[62]	-	+	-
Oh 2024[69]	-	+	+
Ohse 2024[70]	-	+	+
Orhan 2019[71]	-	+	+

Ozkanca 2018[81]	-	N/A	+
Parkeaw 2025[72]	+	+	+
Pérez-Toro 2022[73]	-	+	-
Podina 2025[74]	+	+	+
Qin 2025[9]	-	+	+
Ren 2024[75]	-	+	+
Resnik 2013[76]	+	+	+
Rutowski 2020[77]	+	+	+
Shen 2022[29]	+	+	+
Shin 2022[78]	+	+	+
Shin 2024[79]	-	+	+
Smirnova 2018[80]	+	+	+
Smirnova 2019[81]	+	+	+
Sood 2023[82]	+	+	+
Tao 2023[83]	+	+	+
Tlachac 2020[84]	+	+	+
Tlachac 2022a[85]	+	+	+
Tlachac 2022b[30]	?	?	+
Tlachac 2022c[86]	-	+	+
Weber 2025[87]	-	N/A	+
Wright-Berryman 2023[88]	+	+	+
Xue 2024[31] (Same as Shen 2022)	+	+	+
Ye 2021[89]	+	+	+
Yuan 2021[90]	+	+	+
Zhang 2024[91]	+	+	+
Zou 2023[92]	-	+	+

Note. N/A indicates studies using continuous predictions or item-level classifications rather than binary group outcomes, for which a balanced sample is not applicable. Each quality assessment criterion was rated as either positive (+) when fulfilled or negative (-) when not fulfilled.

Table S3. Summary of Gender-Based Reporting and Percentage of Female Participants Across Studies

Study	Gender reported	Separate analysis reported	% Female participants	Comment
Abilkaiyrkyzy 2024	X	X		
Aloshban 2021	V	X	70	
Antoniou 2022	V	X	64.7	
Banerjee 2021	X	X	-	
Boian 2025	V	X	67.8	
Burkhardt 2022	V	X	79	
Cao 2025	V	X	52	
Chen 2024a (CMDC) (Same as Zou 2023)	X	X	-	
Chen 2024a (EATD) (Same as Shen 2022)*	X	X	-	
Chen 2024b (EATD) (Same as Shen 2022)*	V	X	76.5	
Cohen 2023	V	X	65	
Cook 2016	V	X	53	
deHond 2024	V	X	80	
Demiroglu 2020 (Turkish)	X	X	-	
Demiroglu 2020 (German)	X	X	-	
Gao 2024	X	X	-	
Guo 2024	V	X	49	
Hayati 2022	V	V	50	Models performance were better for female
He 2022 (same data as Yuan 2021)	V	X	75	
Howes 2014	X	X	-	
Iyortsuun 2024 (Same as Shen 2022)	X	X	-	
Joharee 2023	Only pregnant women		100	
Krishnamurti 2022	V	X	66	
Li 2023	V	X	77.2	
Liu 2022	V	X	50	
Munthuli 2023	V	X	85	
Nobles 2018	V	X	69	
Oh 2024	V	X	78.3	
Ohse 2024	V	X	56.6	
Orhan 2019	V	X	82.9	
Parkeaw 2025	V	X	51.8	
Pérez-Toro 2022	V	X	74.1	
Podina 2025	X	X	-	
Qin 2025	V	X	83	

Ren 2024	X	X	-	
Resnik 2013	X	X	-	
Rutowski 2020	X	X	-	
Shen 2022	V	X	80	
Shin 2022	V	X	81	
Shin 2024	V	X	77	
Smirnova 2018	V	X	77	
Smirnova 2019	X	X	-	
Sood 2023	X	X	-	
Tao 2023	X	X	-	
Tlachac 2020	V	X	59	
Tlachac 2022a	X	X	-	
Tlachac 2022b	X	X	-	
Tlachac 2022c	X	X	-	
Weber 2025	V	X	79	
Wright-Berryman 2023	X	X	-	
Xezonaki 2020	V	X	51	
Xue 2024 (Same as Shen 2022)	V	V	50	Model performance was better for female
Ye 2021	V	X	60 (from total sample)	
Yuan 2021	V	X	68	

Table S4. Overview of Gender-Informed Modeling Across DAIC Studies

Study	Included gender in the model
Agarwal 2022	X
Agarwal 2024	X
Al-Hanai 2018	X
Ansari 2023	X
Burdisso 2023*	X
Cao 2022	X
Chen 2024	X
Correia 2016	X
Dang 2017	X
Danner 2023	X
Fang 2023	X
Firoz 2023a	X
Firoz 2023b	Adding gender to the model improved the model
Flores 2023	X
Guo 2024	X
Hadzic 2024	X
Hong 2022	X
Iyortsuun 2024	X
Jo 2022*	X
Kokkera 2023	X
Lam 2019	X
Lau, 2021	X
Lau 2023	X
Li 2022a	X
Li 2022b	X
Li 2023	X
Lin 2020	X
Lopez-Otero 2017	X
Lorenc 2022	X
Lu 2023	X
Rodrigues Makiuchi 2019*	X
Mallol-Ragolta 2019	X
Mao 2023	X
Milintsevich 2023	X
Niu 2021	X
Pampouchidou 2016	Training separate gender-specific models improved performance, though results were reported for the combined gender-based strategy
Prabhu 2022	X
Qureshi 2019	X
Qureshi 2020	X
Qureshi 2021	Including gender improved the MSE but worsened the MAE compared to the gender-unaware model
Rasipuram 2022	X
Ray 2019*	X
Rinaldi 2020	X
Rohanian 2019	X
Sadeghi 2023*	X
Sadeghi 2024*	X
Samareh 2018	Including gender improved the RMSE but worsened the MAE compared to the gender-unaware model

Senn 2022	X
Shen 2022	X
Stasak 2017	X
Stepanov 2018	X
Sun 2017	X
Tlachac 2022	X
Toto 2021	X
Marriwala 2023	X
Van Steijn 2022*	X
Villatoro-Tello 2021*	X
Williamson 2016	X
Xezonaki 2020	X
Xia 2024	X
Xiao 2021	X
Xu 2023*	X
Xue 2024	X
Yadav 20231/14/26 3:49:00 PM	X
Yang 2017a	Model performance was better for female
Yang 2017b	MAE was slightly better for male and RMSE slightly better for female.
Yang 2018	X
Yang 2019	X
Zhang 2020a*	X
Zhang 2020b	X
Zhang 2024b	X
Zhang 2024c	X
Zhao 2022	X

Table S5. Leave-One-Out Sensitivity Analyses for Pooled Classification Accuracy

Study		Accuracy	95% CI low	95% CI high
Abilkaiyrkyzy	2024	0.802	0.765	0.835
Aloshban	2021	0.802	0.764	0.834
Antoniou	2022	0.801	0.763	0.834
Boian	2025	0.799	0.761	0.832
Cook	2016	0.805	0.770	0.836
deHond	2024	0.795	0.758	0.828
Demiroglu	2020a	0.803	0.766	0.835
Demiroglu	2020b	0.799	0.761	0.832
Gao	2023	0.795	0.759	0.827
Guo	2024	0.799	0.761	0.832
Hayati	2022	0.802	0.765	0.834
Iyortsuun	2024	0.804	0.768	0.836
Joharee	2023	0.802	0.765	0.835
Li	2023	0.799	0.761	0.832
Munthuli	2023	0.797	0.760	0.830
Nobles	2018	0.803	0.765	0.835
Oh	2024	0.800	0.763	0.833
Ohse	2024	0.797	0.760	0.830
Orhan	2019	0.798	0.761	0.831
Parkeaw	2025	0.801	0.763	0.834
Pérez-Toro	2022	0.803	0.766	0.835
Podina	2025	0.799	0.761	0.832
Qin	2025	0.800	0.763	0.833
Ren	2024	0.804	0.768	0.836
Resnik	2013	0.800	0.763	0.833
Rutowski	2020	0.801	0.764	0.834
Shin	2022	0.800	0.762	0.833
Shin	2024	0.797	0.760	0.830
Smirnova	2019	0.793	0.758	0.825
Sood	2023	0.797	0.760	0.830
Tao	2023	0.803	0.766	0.835
Tlachac	2020	0.800	0.763	0.833
Tlachac	2022a	0.805	0.769	0.836
Tlachac	2022b	0.802	0.765	0.835
Tlachac	2023	0.802	0.765	0.835
Weber	2025	0.799	0.762	0.832
Wright-Berryman	2023	0.803	0.766	0.835
Ye	2021	0.798	0.760	0.831

Yuan	2021	0.803	0.766	0.835
Zhang	2020a	0.799	0.762	0.832
Zhang	2024a	0.803	0.767	0.836
Zhang	2024c	0.797	0.760	0.830
Zou	2022	0.797	0.760	0.829

Table S6. Leave-One-Out Sensitivity Analyses for Pooled Classification Precision

Study		Accuracy	95% CI low	95% CI high
Cao	2025	0.782	0.720	0.834
Gao	2023	0.768	0.706	0.821
Guo	2024	0.787	0.731	0.835
Iyortsuun	2024	0.777	0.712	0.830
Li	2023	0.768	0.706	0.821
Tao	2023	0.780	0.716	0.833
Yuan	2021	0.780	0.716	0.833
Zhang	2024a	0.781	0.718	0.834
Zou	2022	0.773	0.709	0.826
Burkhardt	2022	0.782	0.720	0.834
deHond	2024	0.767	0.705	0.820
Nobles	2018	0.780	0.716	0.833
Resnik	2013	0.785	0.724	0.835
Sood	2023	0.775	0.710	0.829
Tlachac	2020	0.779	0.715	0.832
Tlachac	2022a	0.785	0.724	0.835
Wright-Berryman	2023	0.780	0.717	0.833
Zhang	2024c	0.776	0.712	0.830
Aloshban	2021	0.763	0.705	0.812
Boian	2025	0.776	0.711	0.830
Cook	2016	0.782	0.719	0.834
Demiroglu	2020a	0.777	0.713	0.831
Demiroglu	2020b	0.772	0.708	0.825
Munthuli	2023	0.769	0.706	0.822
Ohse	2024	0.774	0.710	0.828
Parkeaw	2025	0.777	0.712	0.831
Podina	2025	0.778	0.714	0.831
Shin	2024	0.778	0.714	0.832

Table S7. Leave-One-Out Sensitivity Analyses for Pooled Classification Recall

Study		Accuracy	95% CI low	95% CI high
Cao	2025	0.767	0.682	0.834
Gao	2023	0.751	0.668	0.820
Guo	2024	0.766	0.681	0.833
Iyortsuun	2024	0.768	0.684	0.835
Li	2023	0.757	0.671	0.826
Tao	2023	0.764	0.679	0.833
Yuan	2021	0.768	0.684	0.835
Zhang	2024a	0.766	0.682	0.834
Zou	2022	0.755	0.670	0.823
Burkhardt	2022	0.771	0.689	0.836
deHond	2024	0.779	0.708	0.838
Nobles	2018	0.761	0.676	0.830
Qin	2025	0.761	0.676	0.830
Resnik	2013	0.770	0.687	0.836
Rutowski	2020	0.763	0.678	0.832
Sood	2023	0.761	0.675	0.830
Tlachac	2020	0.755	0.670	0.824
Tlachac	2022a	0.749	0.667	0.817
Tlachac	2023	0.755	0.670	0.824
Wright-Berryman	2023	0.769	0.685	0.835
Zhang	2024c	0.759	0.674	0.829
Aloshban	2021	0.771	0.689	0.836
Boian	2025	0.762	0.676	0.831
Cook	2016	0.774	0.695	0.837
Demiroglu	2020a	0.765	0.681	0.833
Demiroglu	2020b	0.763	0.678	0.832
Munthuli	2023	0.759	0.673	0.828
Oh	2024	0.751	0.668	0.819
Ohse	2024	0.760	0.674	0.829

Pérez-Toro	2022	0.768	0.684	0.835
Podina	2025	0.755	0.670	0.824
Shin	2022	0.765	0.680	0.833
Shin	2024	0.766	0.682	0.834

Table S8. Leave-One-Out Sensitivity Analyses for Pooled Classification AUC

Study		Accuracy	95% CI low	95% CI high
Antoniou	2022	0.789	0.698	0.859
Burkhardt	2022	0.795	0.707	0.862
Cohen	2023	0.800	0.720	0.861
deHond	2024	0.801	0.722	0.861
Krishnamurti	2022	0.778	0.687	0.849
Liu	2022	0.792	0.701	0.860
Oh	2024	0.781	0.689	0.852
Pérez-Toro	2022	0.792	0.703	0.860
Qin	2025	0.777	0.689	0.846
Ren	2024	0.797	0.710	0.862
Rutowski	2020	0.784	0.691	0.855
Shin	2022	0.773	0.686	0.842
Wright-Berryman	2023	0.789	0.697	0.858
Zou	2022	0.759	0.689	0.817

Table S9. Leave-One-Out Sensitivity Analyses for Pooled Classification Balanced Accuracy

Study		Accuracy	95% CI low	95% CI high
Shin	2022.000	0.705	0.620	0.778
Oh	2024.000	0.721	0.637	0.792
Rutowski	2020.000	0.712	0.625	0.785
Wright-Berryman	2023.000	0.738	0.675	0.792
Pérez-Toro	2022.000	0.716	0.631	0.788
deHond	2024.000	0.725	0.643	0.794
Shin	2024.000	0.708	0.622	0.782
Munthuli	2023.000	0.699	0.618	0.769
Boian	2025.000	0.708	0.622	0.781
Cook	2016.000	0.724	0.642	0.794
Podina	2025.000	0.702	0.619	0.774
Tlachac	2020.000	0.705	0.620	0.777
Li	2023.000	0.708	0.622	0.781
Nobles	2018.000	0.716	0.630	0.789
Tlachac	2023.000	0.717	0.632	0.789
Tlachac	2022a	0.723	0.640	0.793

Figure S1. Funnel Plot of Study-Level Classification Accuracy

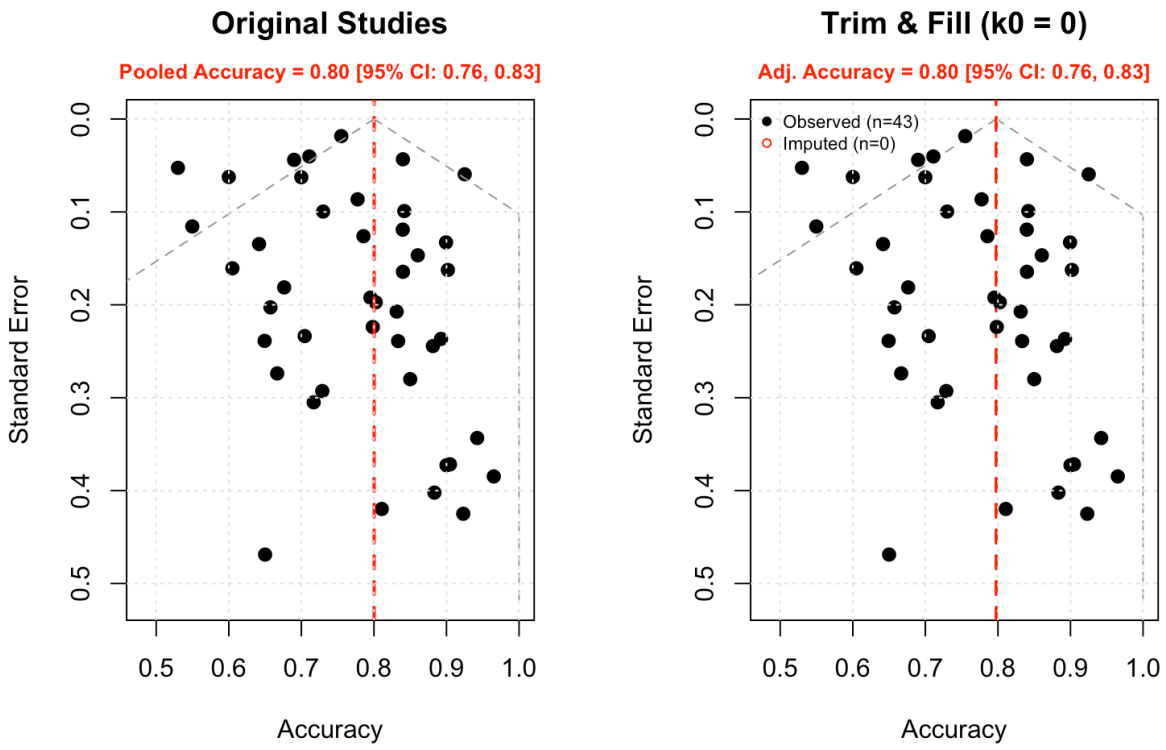

Figure S2. Forest Plot of Study-Level Classification Precision with Pooled Random-Effects Estimate

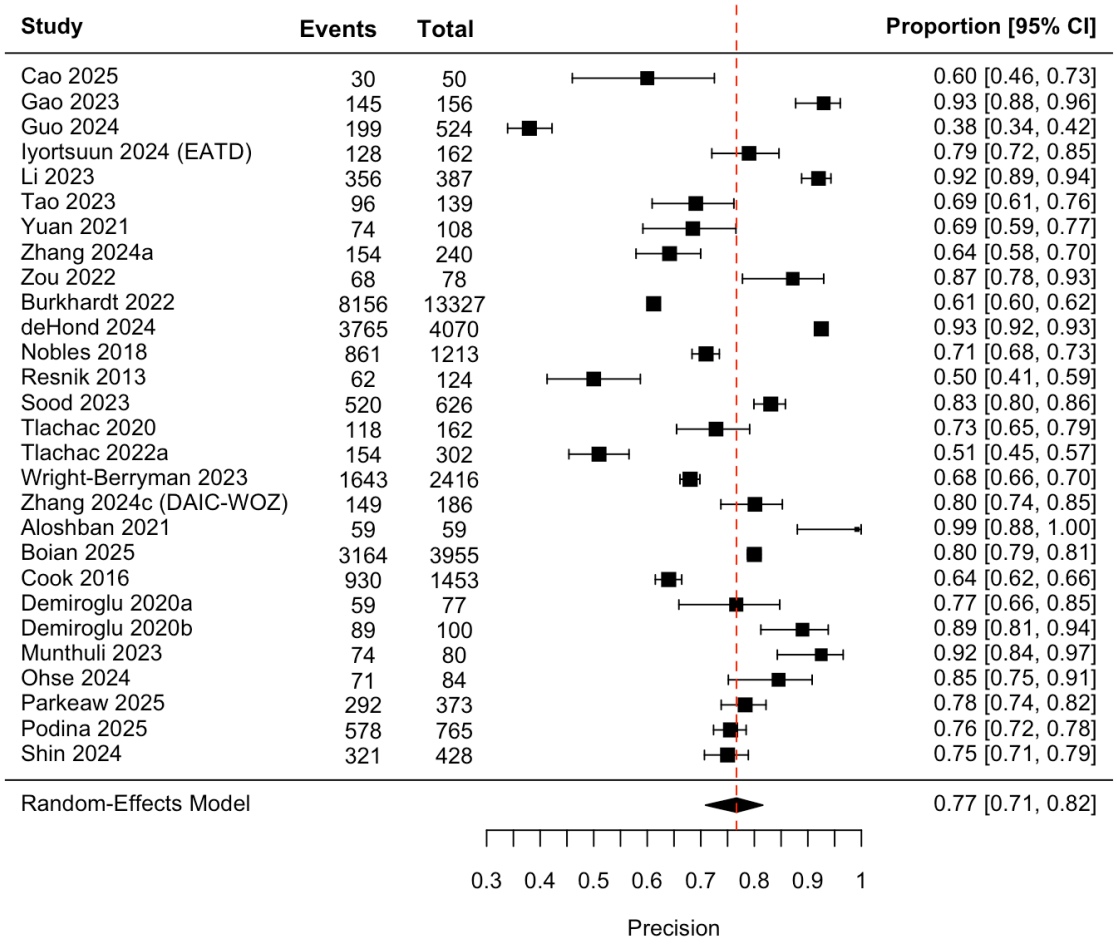

Figure S3. Funnel Plot of Study-Level Classification Precision

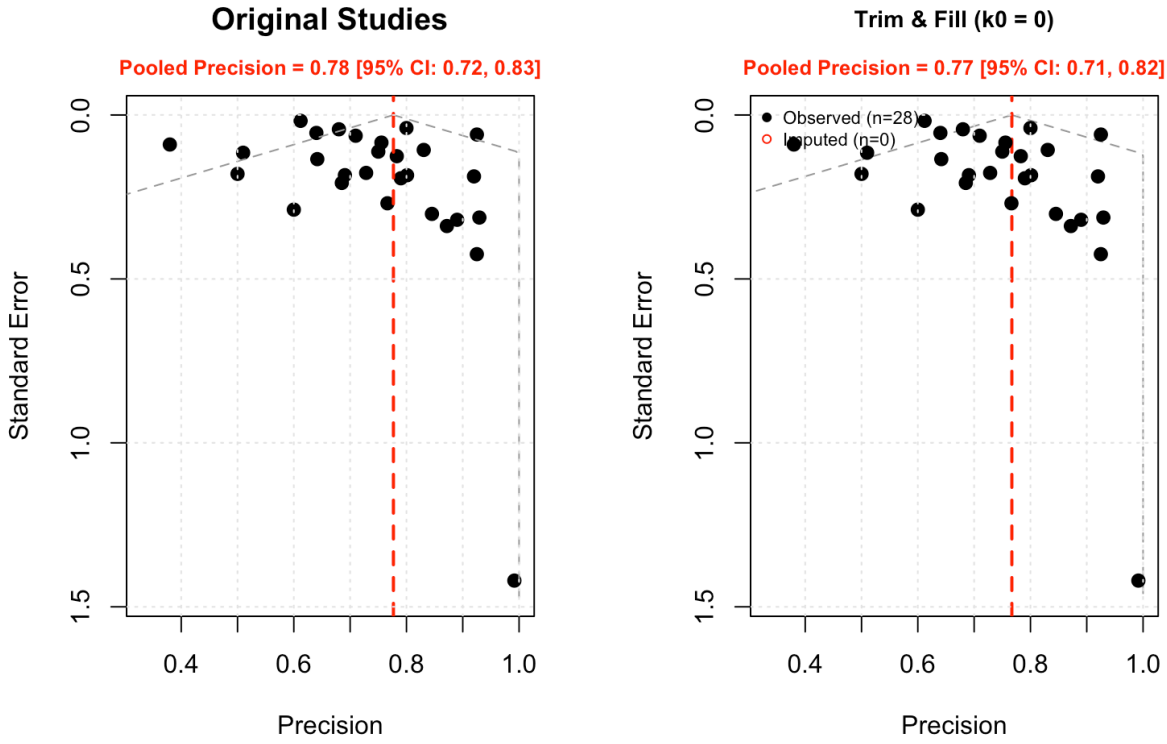

Figure S4. Forest Plot of Study-Level Classification Recall with Pooled Random-Effects Estimate

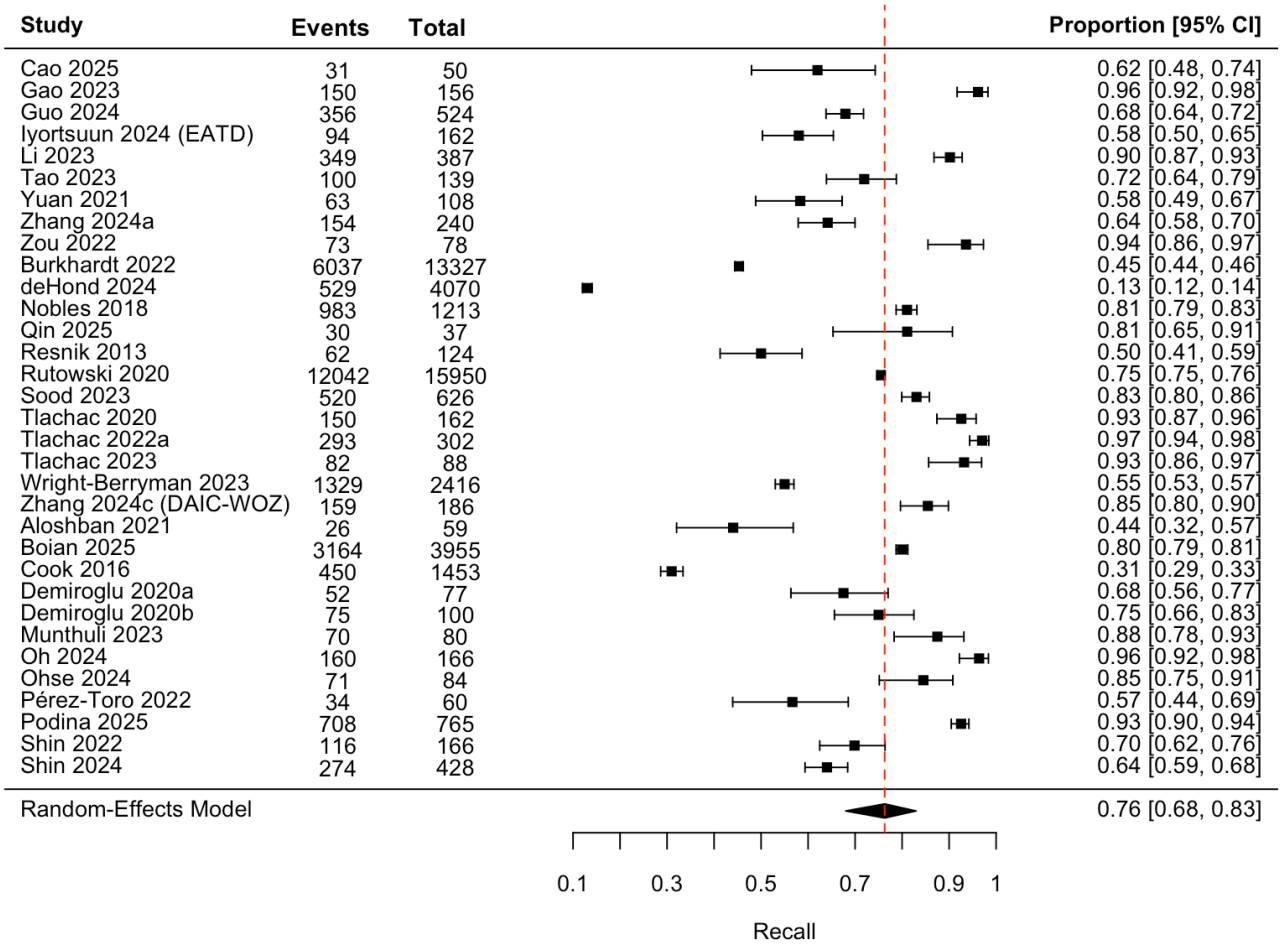

Figure S5. Funnel Plot of Study-Level Classification Recall

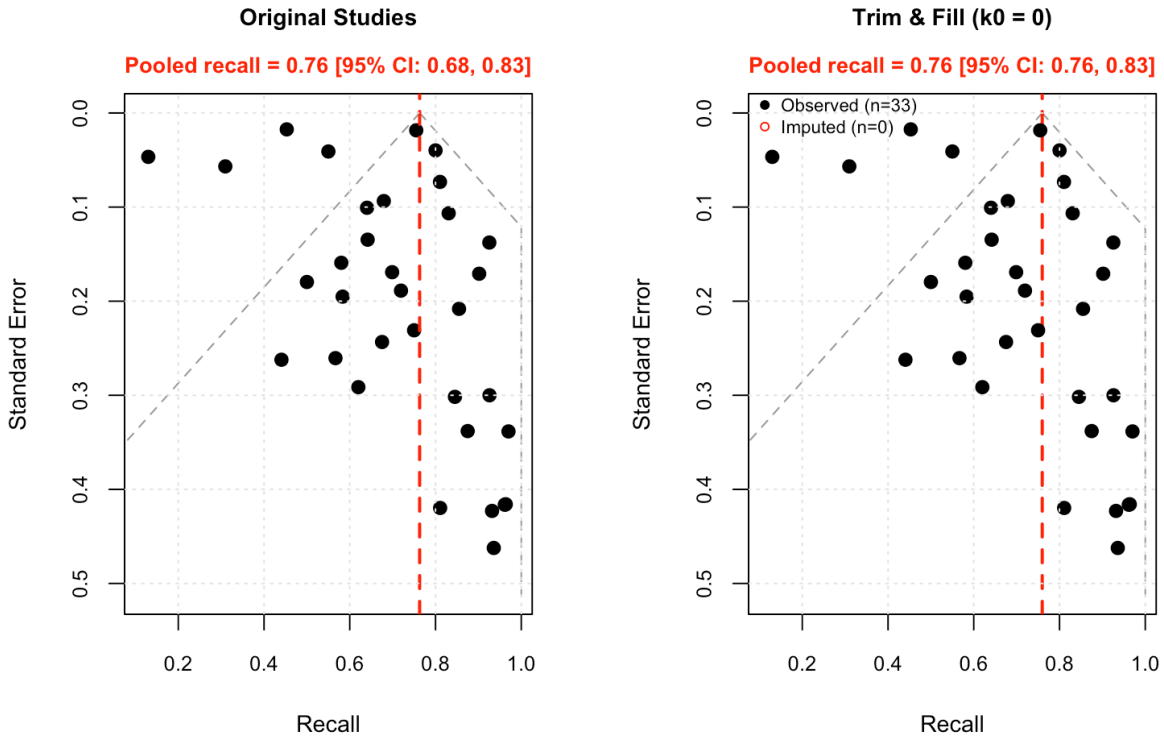

Figure S6. Forest Plot of Study-Level Classification AUC With Pooled Random-Effects Estimate

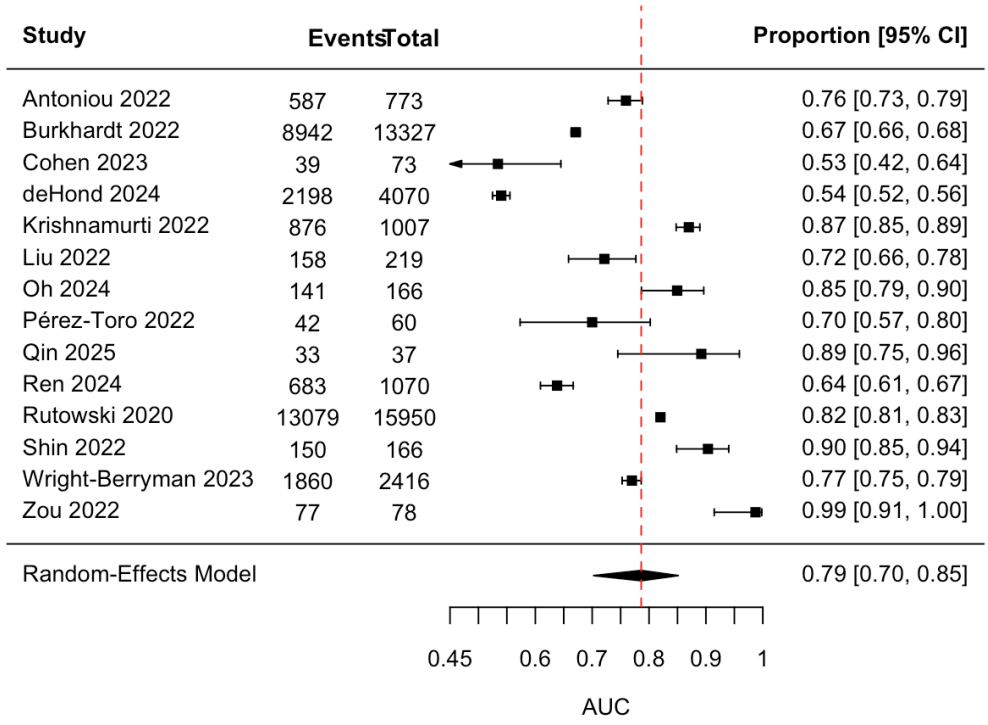

Figure S7. Funnel Plot of Study-Level Classification AUC

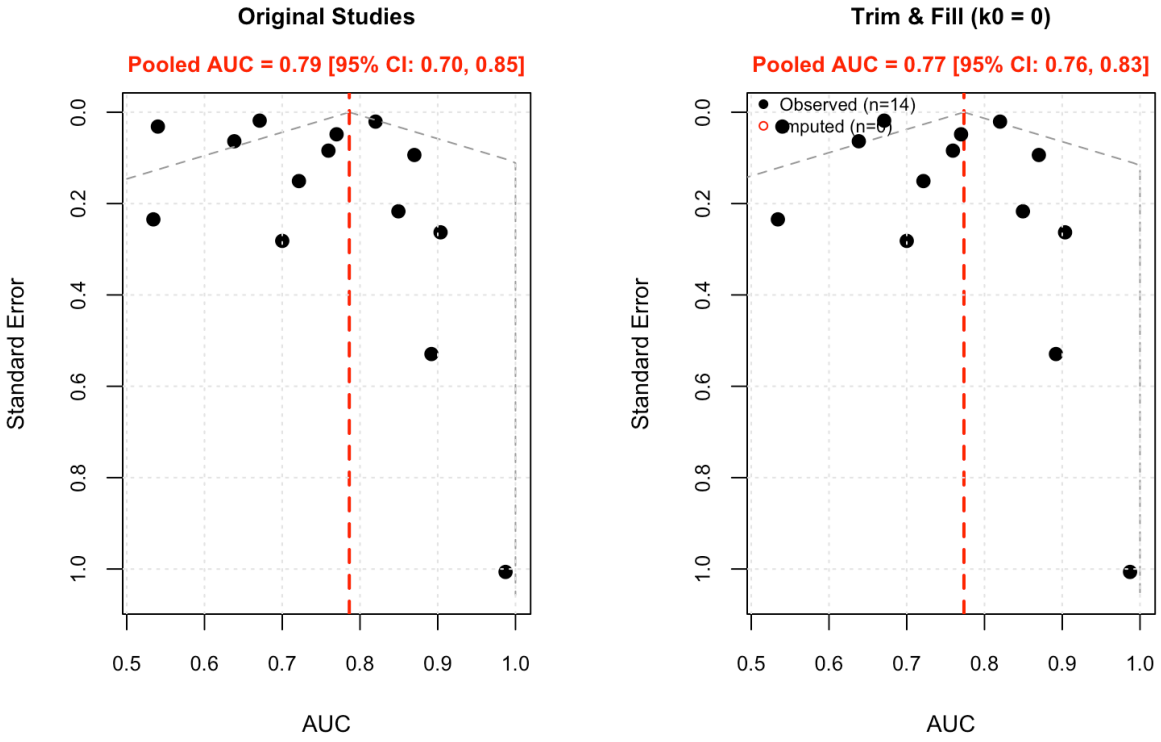

Figure S8. Forest Plot of Study-Level Classification Balanced Accuracy With Pooled Random-Effects Estimate

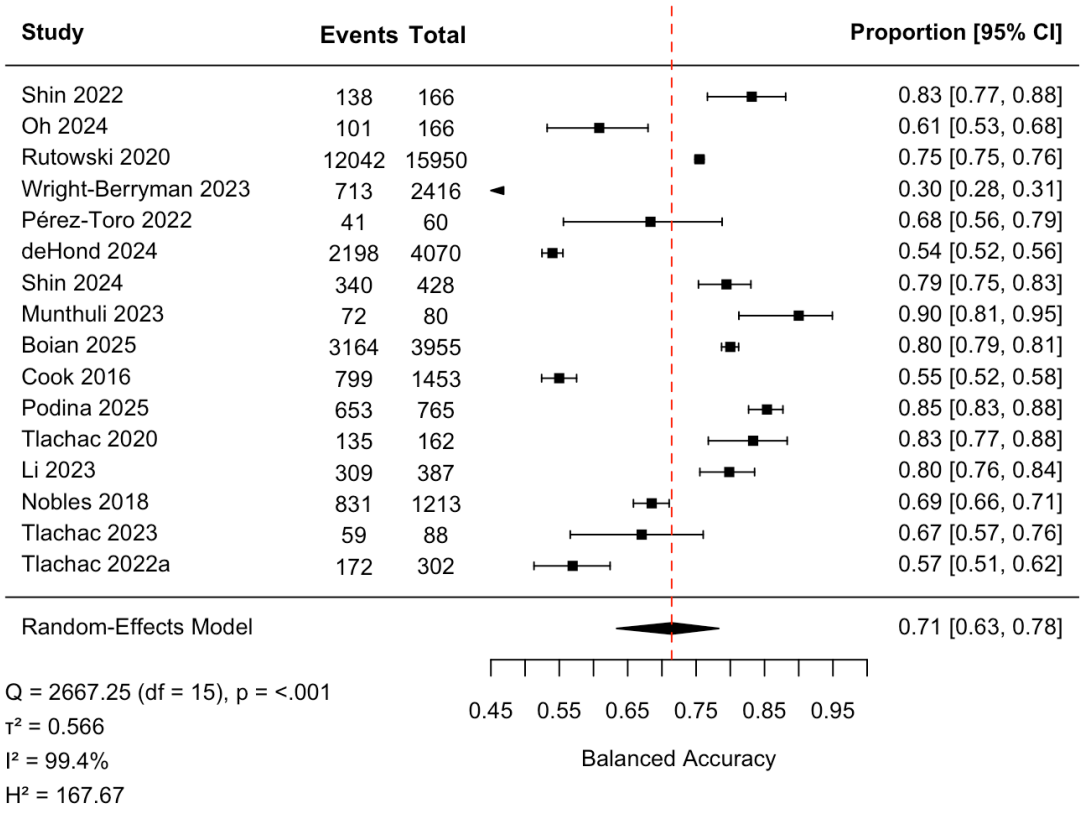

Figure S9. Funnel Plot of Study-Level Classification Balanced Accuracy