

## Model Performance - Balanced Accuracy

## Phenotype

Model

Spore Form.

Motility

Biosafety

Animal Path.

Plant Path.

Host Assoc.

Extreme Env.

Cell Shape

Gram Stain

Biofilm

Balanced Accuracy

gemini-2.5-pro	0.97 (3612)	0.88 (2949)	0.87 (3711)	0.79 (2050)	0.72 (3729)	0.77 (2665)	0.72 (3350)	0.71 (3295)	0.69 (3840)	0.55 (253)
deepseek-r1	0.94 (3611)	0.88 (2950)	0.91 (3711)	0.78 (2051)	0.74 (3729)	0.72 (2665)	0.73 (3350)	0.72 (3295)	0.65 (3840)	0.58 (253)
claude-3.5-sonnet	0.94 (3609)	0.90 (2947)	0.91 (3708)	0.80 (2049)	0.76 (3726)	0.73 (2663)	0.72 (3348)	0.71 (3293)	0.65 (3837)	0.51 (253)
grok-3-mini	0.95 (3612)	0.88 (2950)	0.90 (3711)	0.82 (2051)	0.78 (3729)	0.71 (2665)	0.72 (3350)	0.71 (3295)	0.65 (3840)	0.51 (253)
gpt-5	0.95 (3611)	0.91 (2949)	0.87 (3704)	0.78 (2050)	0.71 (3727)	0.79 (2664)	0.71 (3349)	0.72 (3294)	0.65 (3839)	0.51 (253)
gpt-4o	0.93 (3612)	0.86 (2950)	0.87 (3711)	0.82 (2051)	0.73 (3729)	0.81 (2665)	0.66 (3350)	0.71 (3295)	0.64 (3840)	0.55 (253)
gemini-pro-1.5	0.97 (3607)	0.76 (2945)	0.89 (3705)	0.82 (2046)	0.79 (3723)	0.76 (2659)	0.63 (3344)	0.71 (3289)	0.64 (3833)	0.57 (252)
gpt-oss-120b	0.96 (3603)	0.86 (2947)	0.77 (3707)	0.74 (2048)	0.71 (3718)	0.80 (2659)	0.71 (3346)	0.71 (3289)	0.65 (3832)	0.59 (252)
o1-mini	0.94 (3578)	0.86 (2921)	0.84 (3681)	0.77 (2035)	0.73 (3695)	0.76 (2649)	0.69 (3320)	0.71 (3269)	0.67 (3807)	0.52 (251)
claude-sonnet-4	0.95 (3612)	0.87 (2950)	0.85 (3711)	0.82 (2051)	0.75 (3729)	0.68 (2665)	0.70 (3350)	0.71 (3295)	0.65 (3840)	0.50 (253)
gemini-2.0-flash-lite-001	0.96 (3612)	0.79 (2950)	0.82 (3711)	0.79 (2051)	0.75 (3729)	0.77 (2665)	0.67 (3350)	0.71 (3294)	0.65 (3840)	0.54 (253)
gpt-4.1-nano	0.80 (3610)	0.83 (2950)	0.87 (3709)	0.75 (2048)	0.69 (3726)	0.77 (2663)	0.68 (3347)	0.92 (3290)	0.58 (3837)	0.55 (252)
gpt-4	0.92 (3612)	0.84 (2950)	0.78 (3711)	0.75 (2051)	0.74 (3729)	0.79 (2665)	0.72 (3350)	0.63 (3295)	0.65 (3840)	0.62 (253)
gpt-5-nano	0.91 (3611)	0.86 (2949)	0.84 (3710)	0.72 (2050)	0.73 (3728)	0.78 (2664)	0.69 (3349)	0.71 (3294)	0.68 (3839)	0.50 (253)
deepseek-chat-v3-0324	0.91 (3606)	0.87 (2943)	0.85 (3709)	0.76 (2051)	0.75 (3727)	0.70 (2663)	0.69 (3350)	0.69 (3294)	0.65 (3779)	0.53 (253)
gpt-oss-20b	0.91 (3607)	0.87 (2947)	0.83 (3705)	0.73 (2048)	0.72 (3720)	0.78 (2662)	0.66 (3347)	0.67 (3287)	0.66 (3837)	0.55 (253)
llama-3-70b-instruct:nitro	0.92 (3610)	0.80 (2949)	0.83 (3709)	0.83 (2050)	0.75 (3727)	0.66 (2664)	0.63 (3349)	0.71 (3288)	0.65 (3838)	0.50 (253)
gemini-flash-1.5	0.95 (3611)	0.82 (2950)	0.82 (3710)	0.85 (2050)	0.78 (3729)	0.63 (2664)	0.63 (3349)	0.62 (3294)	0.63 (3839)	0.50 (253)
wizardlm-2-8x22b	0.89 (3601)	0.84 (2943)	0.76 (3700)	0.80 (2045)	0.75 (3718)	0.69 (2658)	0.69 (3340)	0.70 (3282)	0.64 (3829)	0.50 (253)
claude-3-haiku:beta	0.92 (3604)	0.82 (2942)	0.82 (3704)	0.84 (2046)	0.74 (3722)	0.78 (2659)	0.53 (3344)	0.66 (3286)	0.64 (3832)	0.50 (253)
gemini-pro	0.89 (3581)	0.81 (2923)	0.88 (3678)	0.83 (2028)	0.67 (3694)	0.73 (2644)	0.60 (3319)	0.65 (3263)	0.62 (3805)	0.54 (251)
gemini-2.5-flash-lite	0.91 (3612)	0.86 (2950)	0.74 (3711)	0.71 (2051)	0.75 (3729)	0.54 (2665)	0.68 (3350)	0.70 (3290)	0.64 (3840)	0.54 (253)
gpt-3.5-turbo-0125	0.91 (3612)	0.79 (2950)	0.66 (3711)	0.74 (2051)	0.73 (3729)	0.70 (2665)	0.63 (3350)	0.68 (3284)	0.64 (3840)	0.50 (253)
palm-2-chat-bison-32k	0.86 (3610)	0.76 (2948)	0.77 (3709)	0.79 (2051)	0.70 (3727)	0.76 (2665)	0.61 (3348)	0.58 (3293)	0.61 (3838)	0.53 (253)
mixtral-8x7b-instruct:nitro	0.92 (3612)	0.78 (2950)	0.58 (3711)	0.70 (2051)	0.71 (3729)	0.59 (2665)	0.62 (3350)	0.57 (3264)	0.61 (3840)	0.50 (253)
mistral-7b-instruct	0.86 (3609)	0.69 (2948)	0.71 (3691)	0.74 (2050)	0.69 (3726)	0.65 (2663)	0.57 (3349)	0.49 (3287)	0.49 (3805)	0.61 (253)
openchat-7b	0.76 (3609)	0.70 (2947)	0.52 (3708)	0.74 (2051)	0.68 (3726)	0.66 (2663)	0.57 (3347)	0.50 (3292)	0.68 (3837)	0.59 (253)
llama-3-8b-instruct:nitro	0.86 (3612)	0.77 (2950)	0.54 (3711)	0.78 (2051)	0.69 (3729)	0.60 (2665)	0.57 (3350)	0.47 (3290)	0.54 (3757)	0.50 (253)
llama-3-sonar-small-32k-chat	0.88 (3610)	0.64 (2948)	0.61 (3708)	0.76 (2050)	0.66 (3726)	0.50 (2663)	0.63 (3348)	0.40 (3292)	0.58 (3837)	0.50 (253)
phi-3-mini-128k-instruct	0.72 (3603)	0.70 (2945)	0.71 (3699)	0.59 (2046)	0.62 (3722)	0.58 (2658)	0.57 (3343)	0.50 (3275)	0.51 (3831)	0.48 (253)
gemma-7b-it	0.67 (3584)	0.52 (2928)	0.54 (3683)	0.69 (2035)	0.66 (3701)	0.58 (2643)	0.50 (3321)	0.28 (3268)	0.43 (3810)	0.50 (252)