

Supplementary material for the paper titled **TB Screening from
Cough Audio: Baseline Models, Clinical Variables, and
Uncertainty Quantification**

George P. Kafentzis, Efstratios Selisios

January 11, 2026

1 Per-fold Acoustic Only models: LR

Table 1: **Acoustic-Only.** Logistic Regression trained on acoustic features only (dataset: $N=9772$, $D=261$, speakers = 1082). Per-fold results and mean \pm std over 10 outer folds are reported for waveform- and cougher-level evaluation, plus conformal prediction outputs.

| Waveform-level classification (per fold) | | | | | | | | | | |
|--|-------------------|--------|---------|--------|--------|--------|--------|--------|--------|--------|
| Fold | n_{test} | τ | ROC-AUC | PR-AUC | ACC | UAR | Sens | Spec | PPV | NPV |
| 1 | 960 | 0.2500 | 0.7527 | 0.5682 | 0.6479 | 0.6784 | 0.7645 | 0.5923 | 0.4721 | 0.8406 |
| 2 | 945 | 0.3333 | 0.7549 | 0.5624 | 0.7333 | 0.6861 | 0.5870 | 0.7851 | 0.4915 | 0.8431 |
| 3 | 1043 | 0.3505 | 0.7235 | 0.4383 | 0.6663 | 0.6518 | 0.6172 | 0.6865 | 0.4463 | 0.8141 |
| 4 | 988 | 0.3974 | 0.6444 | 0.4282 | 0.6306 | 0.5894 | 0.4776 | 0.7012 | 0.4245 | 0.7441 |
| 5 | 976 | 0.3256 | 0.7121 | 0.4587 | 0.6342 | 0.6564 | 0.7031 | 0.6097 | 0.3905 | 0.8524 |
| 6 | 1021 | 0.2857 | 0.6375 | 0.3814 | 0.6484 | 0.6157 | 0.5465 | 0.6848 | 0.3828 | 0.8085 |
| 7 | 979 | 0.3214 | 0.6818 | 0.4859 | 0.6639 | 0.6486 | 0.6078 | 0.6895 | 0.4709 | 0.7945 |
| 8 | 919 | 0.2928 | 0.6651 | 0.5229 | 0.6007 | 0.6122 | 0.6438 | 0.5805 | 0.4169 | 0.7778 |
| 9 | 911 | 0.2924 | 0.6441 | 0.3627 | 0.6037 | 0.6158 | 0.6458 | 0.5859 | 0.3977 | 0.7962 |
| 10 | 1030 | 0.3836 | 0.5967 | 0.4390 | 0.6311 | 0.5789 | 0.4011 | 0.7568 | 0.4740 | 0.6981 |

| Cougher-level classification (per fold) | | | | | | | | | | |
|---|-------------------|--------|---------|--------|--------|--------|--------|--------|--------|--------|
| Fold | n_{test} | τ | ROC-AUC | PR-AUC | ACC | UAR | Sens | Spec | PPV | NPV |
| 1 | 108 | 0.204 | 0.7712 | 0.5698 | 0.4907 | 0.6167 | 0.9000 | 0.3333 | 0.3418 | 0.8966 |
| 2 | 108 | 0.304 | 0.7772 | 0.5944 | 0.7222 | 0.7251 | 0.7308 | 0.7195 | 0.4524 | 0.8939 |
| 3 | 108 | 0.250 | 0.7589 | 0.4333 | 0.6111 | 0.7027 | 0.8929 | 0.5125 | 0.3906 | 0.9318 |
| 4 | 108 | 0.334 | 0.6430 | 0.3984 | 0.6111 | 0.5705 | 0.4828 | 0.6582 | 0.3415 | 0.7761 |
| 5 | 108 | 0.308 | 0.6790 | 0.4435 | 0.6574 | 0.6875 | 0.7500 | 0.6250 | 0.4118 | 0.8772 |
| 6 | 108 | 0.290 | 0.5483 | 0.3283 | 0.5648 | 0.5346 | 0.4667 | 0.6026 | 0.3111 | 0.7460 |
| 7 | 107 | 0.332 | 0.6801 | 0.4865 | 0.6636 | 0.6238 | 0.5333 | 0.7143 | 0.4211 | 0.7971 |
| 8 | 108 | 0.282 | 0.6929 | 0.5661 | 0.6019 | 0.5955 | 0.5806 | 0.6104 | 0.3750 | 0.7833 |
| 9 | 109 | 0.308 | 0.6224 | 0.3049 | 0.6239 | 0.6134 | 0.5926 | 0.6341 | 0.3478 | 0.8254 |
| 10 | 110 | 0.362 | 0.6190 | 0.3745 | 0.6727 | 0.5757 | 0.3438 | 0.8077 | 0.4231 | 0.7500 |

| Classification summary (mean \pm std over folds) | | |
|--|-------------------------------|------------------------------|
| Metric | Waveform ($\mu \pm \sigma$) | Cougher ($\mu \pm \sigma$) |
| Threshold (τ) | 0.3233 \pm 0.0455 | 0.2974 \pm 0.0450 |
| ROC AUC | 0.6813 \pm 0.0530 | 0.6792 \pm 0.0745 |
| PR AUC | 0.4648 \pm 0.0701 | 0.4500 \pm 0.1025 |
| ACC | 0.6460 \pm 0.0377 | 0.6219 \pm 0.0639 |
| UAR | 0.6333 \pm 0.0363 | 0.6246 \pm 0.0620 |
| Sensitivity | 0.5994 \pm 0.1052 | 0.6273 \pm 0.1856 |
| Specificity | 0.6672 \pm 0.0725 | 0.6218 \pm 0.1290 |
| PPV | 0.4367 \pm 0.0395 | 0.3816 \pm 0.0455 |
| NPV | 0.7969 \pm 0.0477 | 0.8277 \pm 0.0672 |

| Conformal prediction summary (mean \pm std over folds) | | | |
|--|----------|-------------------------------|---|
| Level | α | Coverage ($\mu \pm \sigma$) | Set size ($\mu \pm \sigma$) [Singleton] |
| Waveform | 0.10 | 0.8981 \pm 0.0389 | 1.479 \pm 0.064 [0.521 \pm 0.064] |
| Waveform | 0.05 | 0.9436 \pm 0.0366 | 1.668 \pm 0.095 [0.332 \pm 0.095] |
| Cougher | 0.10 | 0.9039 \pm 0.0419 | 1.442 \pm 0.078 [0.558 \pm 0.078] |
| Cougher | 0.05 | 0.9492 \pm 0.0318 | 1.642 \pm 0.083 [0.358 \pm 0.083] |

2 Per-fold Acoustic Only models: CatBoost

Table 2: **Acoustic-Only.** CatBoost trained on acoustic features only (dataset: $N=9772$, $D=261$, speakers = 1082). Per-fold results and mean \pm std over 10 outer folds are reported for waveform- and cougher-level evaluation, plus conformal prediction outputs.

| Waveform-level classification (per fold) | | | | | | | | | | |
|--|-------------------|--------|---------|--------|--------|--------|--------|--------|--------|--------|
| Fold | n_{test} | τ | ROC-AUC | PR-AUC | ACC | UAR | Sens | Spec | PPV | NPV |
| 1 | 960 | 0.3338 | 0.7268 | 0.5446 | 0.6500 | 0.6555 | 0.6710 | 0.6400 | 0.4706 | 0.8031 |
| 2 | 945 | 0.3513 | 0.7784 | 0.5765 | 0.6889 | 0.7122 | 0.7611 | 0.6633 | 0.4444 | 0.8870 |
| 3 | 1043 | 0.3608 | 0.7212 | 0.4506 | 0.6731 | 0.6585 | 0.6238 | 0.6932 | 0.4543 | 0.8182 |
| 4 | 988 | 0.3369 | 0.6771 | 0.4406 | 0.5931 | 0.6414 | 0.7724 | 0.5104 | 0.4213 | 0.8293 |
| 5 | 976 | 0.3494 | 0.7096 | 0.4467 | 0.6506 | 0.6562 | 0.6680 | 0.6444 | 0.4005 | 0.8452 |
| 6 | 1021 | 0.3249 | 0.6834 | 0.4194 | 0.6543 | 0.6459 | 0.6283 | 0.6636 | 0.4005 | 0.8331 |
| 7 | 979 | 0.3652 | 0.6954 | 0.4868 | 0.6711 | 0.6271 | 0.5098 | 0.7444 | 0.4756 | 0.7696 |
| 8 | 919 | 0.2290 | 0.6917 | 0.4915 | 0.5571 | 0.6297 | 0.8288 | 0.4306 | 0.4040 | 0.8438 |
| 9 | 911 | 0.2456 | 0.6424 | 0.3822 | 0.5532 | 0.6214 | 0.7897 | 0.4531 | 0.3794 | 0.8357 |
| 10 | 1030 | 0.2889 | 0.6428 | 0.4842 | 0.6417 | 0.6190 | 0.5412 | 0.6967 | 0.4937 | 0.7353 |

| Cougher-level classification (per fold) | | | | | | | | | | |
|---|-------------------|--------|---------|--------|--------|--------|--------|--------|--------|--------|
| Fold | n_{test} | τ | ROC-AUC | PR-AUC | ACC | UAR | Sens | Spec | PPV | NPV |
| 1 | 108 | 0.302 | 0.7496 | 0.5449 | 0.6204 | 0.6449 | 0.7000 | 0.5897 | 0.3962 | 0.8364 |
| 2 | 108 | 0.292 | 0.8086 | 0.6490 | 0.6019 | 0.6984 | 0.8846 | 0.5122 | 0.3651 | 0.9333 |
| 3 | 108 | 0.362 | 0.7705 | 0.4453 | 0.6852 | 0.6946 | 0.7143 | 0.6750 | 0.4348 | 0.8710 |
| 4 | 108 | 0.392 | 0.6626 | 0.4226 | 0.6481 | 0.5849 | 0.4483 | 0.7215 | 0.3714 | 0.7808 |
| 5 | 108 | 0.302 | 0.6527 | 0.3793 | 0.5833 | 0.6143 | 0.6786 | 0.5500 | 0.3455 | 0.8302 |
| 6 | 108 | 0.240 | 0.5776 | 0.3728 | 0.4444 | 0.5333 | 0.7333 | 0.3333 | 0.2973 | 0.7647 |
| 7 | 107 | 0.320 | 0.7169 | 0.5331 | 0.6822 | 0.6673 | 0.6333 | 0.7013 | 0.4524 | 0.8308 |
| 8 | 108 | 0.240 | 0.7453 | 0.5175 | 0.6019 | 0.6726 | 0.8387 | 0.5065 | 0.4062 | 0.8864 |
| 9 | 109 | 0.314 | 0.6292 | 0.3251 | 0.6055 | 0.6136 | 0.6296 | 0.5976 | 0.3400 | 0.8305 |
| 10 | 110 | 0.284 | 0.6611 | 0.4077 | 0.6545 | 0.6550 | 0.6562 | 0.6538 | 0.4375 | 0.8226 |

| Classification summary (mean \pm std over folds) | | |
|--|-------------------------------|------------------------------|
| Metric | Waveform ($\mu \pm \sigma$) | Cougher ($\mu \pm \sigma$) |
| Threshold (τ) | 0.3186 \pm 0.0481 | 0.3048 \pm 0.0474 |
| ROC AUC | 0.6969 \pm 0.0405 | 0.6974 \pm 0.0720 |
| PR AUC | 0.4723 \pm 0.0576 | 0.4597 \pm 0.0990 |
| ACC | 0.6333 \pm 0.0483 | 0.6127 \pm 0.0688 |
| UAR | 0.6467 \pm 0.0273 | 0.6379 \pm 0.0518 |
| Sensitivity | 0.6794 \pm 0.1073 | 0.6917 \pm 0.1198 |
| Specificity | 0.6140 \pm 0.1090 | 0.5841 \pm 0.1160 |
| PPV | 0.4344 \pm 0.0387 | 0.3846 \pm 0.0497 |
| NPV | 0.8200 \pm 0.0424 | 0.8387 \pm 0.0489 |

| Conformal prediction summary (mean \pm std over folds) | | | |
|--|----------|-------------------------------|---|
| Level | α | Coverage ($\mu \pm \sigma$) | Set size ($\mu \pm \sigma$) [Singleton] |
| Waveform | 0.10 | 0.8957 \pm 0.0424 | 1.451 \pm 0.082 [0.549 \pm 0.082] |
| Waveform | 0.05 | 0.9353 \pm 0.0383 | 1.590 \pm 0.116 [0.410 \pm 0.116] |
| Cougher | 0.10 | 0.9040 \pm 0.0386 | 1.427 \pm 0.096 [0.573 \pm 0.096] |
| Cougher | 0.05 | 0.9502 \pm 0.0358 | 1.601 \pm 0.094 [0.399 \pm 0.094] |

3 Per-fold Fused-features models: LR

Table 3: **Fused Features.** Logistic Regression trained on fused features (dataset: $N=9772$, $D=277$, speakers = 1082). Per-fold results and mean \pm std over 10 outer folds are reported for waveform- and cougher-level evaluation, plus conformal prediction outputs.

| Waveform-level classification (per fold) | | | | | | | | | | |
|--|-------------------|--------|---------|--------|--------|--------|--------|--------|--------|--------|
| Fold | n_{test} | τ | ROC-AUC | PR-AUC | ACC | UAR | Sens | Spec | PPV | NPV |
| 1 | 960 | 0.1967 | 0.8276 | 0.6871 | 0.6604 | 0.7324 | 0.9355 | 0.5292 | 0.4866 | 0.9451 |
| 2 | 945 | 0.2785 | 0.8612 | 0.7671 | 0.6952 | 0.7453 | 0.8502 | 0.6404 | 0.4555 | 0.9236 |
| 3 | 1043 | 0.3589 | 0.8169 | 0.5900 | 0.7287 | 0.7493 | 0.7987 | 0.7000 | 0.5216 | 0.8946 |
| 4 | 988 | 0.3388 | 0.7684 | 0.6165 | 0.7055 | 0.6976 | 0.6763 | 0.7189 | 0.5262 | 0.8279 |
| 5 | 976 | 0.3659 | 0.8363 | 0.6374 | 0.7828 | 0.7722 | 0.7500 | 0.7944 | 0.5647 | 0.8994 |
| 6 | 1021 | 0.2483 | 0.7173 | 0.5078 | 0.6543 | 0.6578 | 0.6654 | 0.6503 | 0.4050 | 0.8446 |
| 7 | 979 | 0.3077 | 0.7811 | 0.6481 | 0.6987 | 0.6917 | 0.6732 | 0.7103 | 0.5137 | 0.8270 |
| 8 | 919 | 0.3529 | 0.8480 | 0.7168 | 0.7911 | 0.7902 | 0.7877 | 0.7927 | 0.6389 | 0.8891 |
| 9 | 911 | 0.3400 | 0.7991 | 0.5924 | 0.7475 | 0.7118 | 0.6236 | 0.8000 | 0.5690 | 0.8339 |
| 10 | 1030 | 0.2919 | 0.7854 | 0.6126 | 0.7214 | 0.6911 | 0.5879 | 0.7943 | 0.6097 | 0.7791 |

| Cougher-level classification (per fold) | | | | | | | | | | |
|---|-------------------|--------|---------|--------|--------|--------|--------|--------|--------|--------|
| Fold | n_{test} | τ | ROC-AUC | PR-AUC | ACC | UAR | Sens | Spec | PPV | NPV |
| 1 | 108 | 0.260 | 0.8372 | 0.6898 | 0.6759 | 0.7551 | 0.9333 | 0.5769 | 0.4590 | 0.9574 |
| 2 | 108 | 0.326 | 0.8598 | 0.7812 | 0.7315 | 0.7444 | 0.7692 | 0.7195 | 0.4651 | 0.9077 |
| 3 | 108 | 0.334 | 0.8228 | 0.5572 | 0.7315 | 0.7491 | 0.7857 | 0.7125 | 0.4889 | 0.9048 |
| 4 | 108 | 0.360 | 0.7660 | 0.6206 | 0.7130 | 0.6838 | 0.6207 | 0.7468 | 0.4737 | 0.8429 |
| 5 | 108 | 0.400 | 0.8020 | 0.6087 | 0.7963 | 0.7464 | 0.6429 | 0.8500 | 0.6000 | 0.8718 |
| 6 | 108 | 0.148 | 0.6868 | 0.4694 | 0.5648 | 0.6269 | 0.7667 | 0.4872 | 0.3651 | 0.8444 |
| 7 | 107 | 0.356 | 0.7805 | 0.6277 | 0.6916 | 0.6433 | 0.5333 | 0.7532 | 0.4571 | 0.8056 |
| 8 | 108 | 0.412 | 0.8580 | 0.7098 | 0.8056 | 0.7673 | 0.6774 | 0.8571 | 0.6562 | 0.8684 |
| 9 | 109 | 0.386 | 0.7823 | 0.5462 | 0.7798 | 0.7170 | 0.5926 | 0.8415 | 0.5517 | 0.8625 |
| 10 | 110 | 0.288 | 0.7933 | 0.5903 | 0.7364 | 0.7127 | 0.6562 | 0.7692 | 0.5385 | 0.8451 |

| Classification summary (mean \pm std over folds) | | |
|--|-------------------------------|------------------------------|
| Metric | Waveform ($\mu \pm \sigma$) | Cougher ($\mu \pm \sigma$) |
| Threshold (τ) | 0.3080 \pm 0.0548 | 0.3270 \pm 0.0789 |
| ROC AUC | 0.8041 \pm 0.0430 | 0.7989 \pm 0.0511 |
| PR AUC | 0.6376 \pm 0.0729 | 0.6201 \pm 0.0895 |
| ACC | 0.7185 \pm 0.0459 | 0.7226 \pm 0.0700 |
| UAR | 0.7239 \pm 0.0411 | 0.7146 \pm 0.0486 |
| Sensitivity | 0.7348 \pm 0.1089 | 0.6978 \pm 0.1166 |
| Specificity | 0.7130 \pm 0.0884 | 0.7314 \pm 0.1193 |
| PPV | 0.5291 \pm 0.0702 | 0.5055 \pm 0.0830 |
| NPV | 0.8664 \pm 0.0517 | 0.8711 \pm 0.0428 |

| Conformal prediction summary (mean \pm std over folds) | | | | |
|--|----------|-------------------------------|-------------------------------|---------------------|
| Level | α | Coverage ($\mu \pm \sigma$) | Set size ($\mu \pm \sigma$) | [Singleton] |
| Waveform | 0.10 | 0.8982 \pm 0.0427 | 1.339 \pm 0.093 | [0.661 \pm 0.093] |
| Waveform | 0.05 | 0.9576 \pm 0.0192 | 1.533 \pm 0.067 | [0.467 \pm 0.067] |
| Cougher | 0.10 | 0.9104 \pm 0.0355 | 1.322 \pm 0.084 | [0.678 \pm 0.084] |
| Cougher | 0.05 | 0.9621 \pm 0.0188 | 1.510 \pm 0.068 | [0.490 \pm 0.068] |

4 Per-fold Fused-features models: CatBoost

Table 4: **Fused Features.** CatBoost trained on fused features (dataset: $N=9772$, $D=277$, speakers = 1082). Per-fold results and mean \pm std over 10 outer folds are reported for waveform- and cougher-level evaluation, plus conformal prediction outputs.

| Waveform-level classification (per fold) | | | | | | | | | | |
|--|-------------------|--------|---------|--------|--------|--------|--------|--------|--------|--------|
| Fold | n_{test} | τ | ROC-AUC | PR-AUC | ACC | UAR | Sens | Spec | PPV | NPV |
| 1 | 960 | 0.3924 | 0.8140 | 0.6998 | 0.6865 | 0.7027 | 0.7484 | 0.6569 | 0.5099 | 0.8455 |
| 2 | 945 | 0.3065 | 0.8706 | 0.7029 | 0.7386 | 0.7891 | 0.8947 | 0.6834 | 0.5000 | 0.9483 |
| 3 | 1043 | 0.4619 | 0.8444 | 0.6775 | 0.7728 | 0.7600 | 0.7294 | 0.7905 | 0.5878 | 0.8771 |
| 4 | 988 | 0.4402 | 0.7517 | 0.5757 | 0.6994 | 0.6259 | 0.4263 | 0.8254 | 0.5299 | 0.7571 |
| 5 | 976 | 0.4017 | 0.8298 | 0.5877 | 0.7254 | 0.7661 | 0.8516 | 0.6806 | 0.4866 | 0.9280 |
| 6 | 1021 | 0.3229 | 0.7076 | 0.4762 | 0.6729 | 0.6275 | 0.5316 | 0.7234 | 0.4074 | 0.8119 |
| 7 | 979 | 0.3663 | 0.8123 | 0.6770 | 0.7569 | 0.7439 | 0.7092 | 0.7786 | 0.5929 | 0.8548 |
| 8 | 919 | 0.3636 | 0.8059 | 0.6802 | 0.7595 | 0.7341 | 0.6644 | 0.8038 | 0.6120 | 0.8372 |
| 9 | 911 | 0.3226 | 0.8139 | 0.6486 | 0.7420 | 0.7419 | 0.7417 | 0.7422 | 0.5492 | 0.8716 |
| 10 | 1030 | 0.2154 | 0.8336 | 0.6816 | 0.5913 | 0.6715 | 0.9451 | 0.3979 | 0.4617 | 0.9298 |

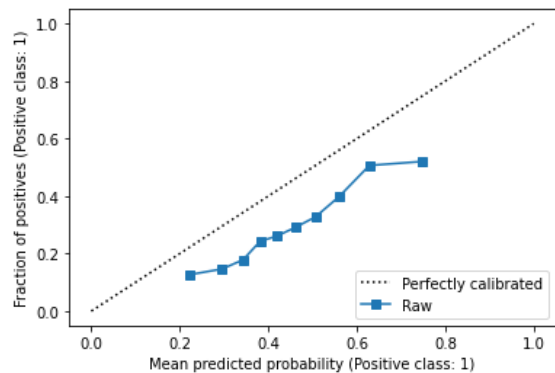
| Cougher-level classification (per fold) | | | | | | | | | | |
|---|-------------------|--------|---------|--------|--------|--------|--------|--------|--------|--------|
| Fold | n_{test} | τ | ROC-AUC | PR-AUC | ACC | UAR | Sens | Spec | PPV | NPV |
| 1 | 108 | 0.342 | 0.8415 | 0.6978 | 0.6574 | 0.7218 | 0.8667 | 0.5769 | 0.4407 | 0.9184 |
| 2 | 108 | 0.184 | 0.8605 | 0.7060 | 0.5833 | 0.7125 | 0.9615 | 0.4634 | 0.3623 | 0.9744 |
| 3 | 108 | 0.394 | 0.8730 | 0.7273 | 0.8056 | 0.8107 | 0.8214 | 0.8000 | 0.5897 | 0.9275 |
| 4 | 108 | 0.410 | 0.7560 | 0.5928 | 0.7037 | 0.6556 | 0.5517 | 0.7595 | 0.4571 | 0.8219 |
| 5 | 108 | 0.342 | 0.7980 | 0.5730 | 0.7222 | 0.7545 | 0.8214 | 0.6875 | 0.4792 | 0.9167 |
| 6 | 108 | 0.126 | 0.6955 | 0.4821 | 0.5556 | 0.6513 | 0.8667 | 0.4359 | 0.3714 | 0.8947 |
| 7 | 107 | 0.308 | 0.7810 | 0.6097 | 0.7383 | 0.7368 | 0.7333 | 0.7403 | 0.5238 | 0.8769 |
| 8 | 108 | 0.182 | 0.8270 | 0.6870 | 0.7130 | 0.7602 | 0.8710 | 0.6494 | 0.5000 | 0.9259 |
| 9 | 109 | 0.298 | 0.8123 | 0.6255 | 0.7431 | 0.7299 | 0.7037 | 0.7561 | 0.4872 | 0.8857 |
| 10 | 110 | 0.338 | 0.8598 | 0.6999 | 0.8091 | 0.7640 | 0.6562 | 0.8718 | 0.6774 | 0.8608 |

| Classification summary (mean \pm std over folds) | | |
|--|-------------------------------|------------------------------|
| Metric | Waveform ($\mu \pm \sigma$) | Cougher ($\mu \pm \sigma$) |
| Threshold (τ) | 0.3594 \pm 0.0716 | 0.2924 \pm 0.0961 |
| ROC AUC | 0.8084 \pm 0.0468 | 0.8104 \pm 0.0551 |
| PR AUC | 0.6407 \pm 0.0726 | 0.6401 \pm 0.0774 |
| ACC | 0.7145 \pm 0.0544 | 0.7031 \pm 0.0838 |
| UAR | 0.7162 \pm 0.0574 | 0.7297 \pm 0.0487 |
| Sensitivity | 0.7242 \pm 0.1580 | 0.7854 \pm 0.1224 |
| Specificity | 0.7083 \pm 0.1231 | 0.6741 \pm 0.1433 |
| PPV | 0.5237 \pm 0.0640 | 0.4889 \pm 0.0945 |
| NPV | 0.8661 \pm 0.0587 | 0.9003 \pm 0.0421 |

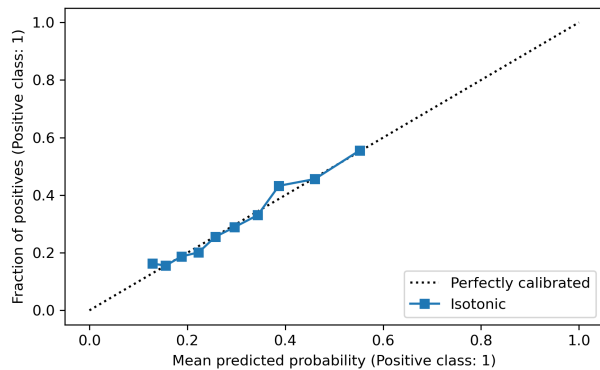
| Conformal prediction summary (mean \pm std over folds) | | | | |
|--|----------|-------------------------------|-------------------------------|---------------------|
| Level | α | Coverage ($\mu \pm \sigma$) | Set size ($\mu \pm \sigma$) | [Singleton] |
| Waveform | 0.10 | 0.9018 \pm 0.0448 | 1.340 \pm 0.122 | [0.660 \pm 0.122] |
| Waveform | 0.05 | 0.9508 \pm 0.0271 | 1.490 \pm 0.095 | [0.510 \pm 0.095] |
| Cougher | 0.10 | 0.9066 \pm 0.0380 | 1.306 \pm 0.066 | [0.694 \pm 0.066] |
| Cougher | 0.05 | 0.9602 \pm 0.0249 | 1.522 \pm 0.089 | [0.478 \pm 0.089] |

5 Calibration

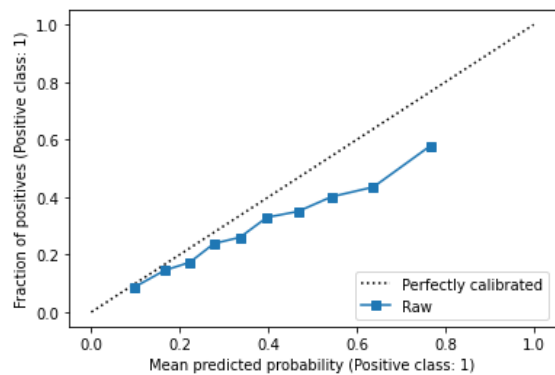
5.1 Waveform-level - Acoustic-only models



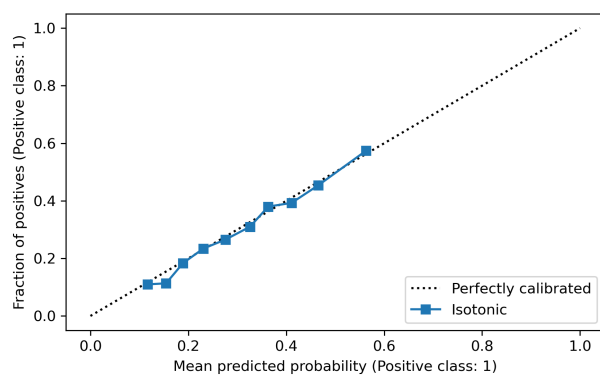
(a) Reliability plot: LR (before calibration)



(b) Reliability plot: LR (after calibration)



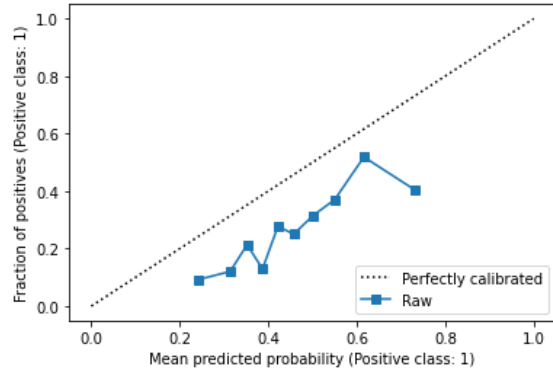
(c) Reliability plot: CatBoost (before calibration)



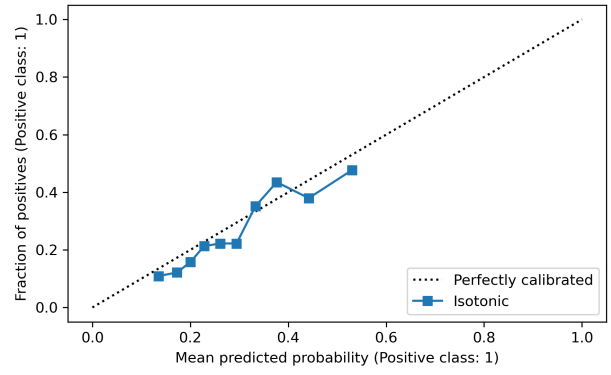
(d) Reliability plot: CatBoost (after calibration)

Figure 1: Waveform-level reliability plots for acoustic-only trained models.

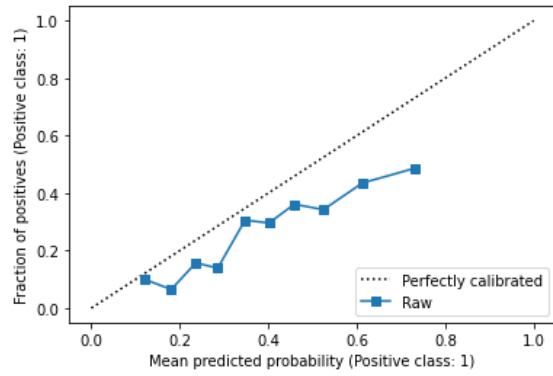
5.2 Cougher-level - Acoustic-only models



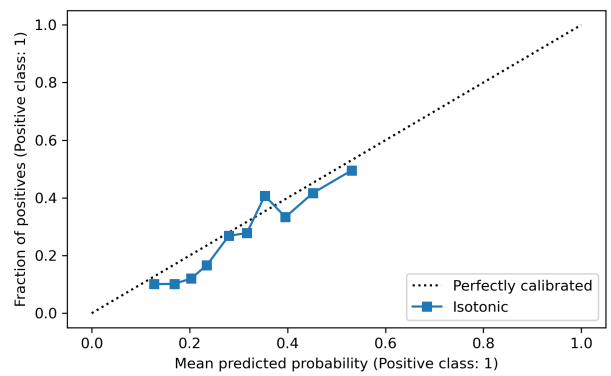
(a) Reliability plot: LR (before calibration)



(b) Reliability plot: LR (after calibration)



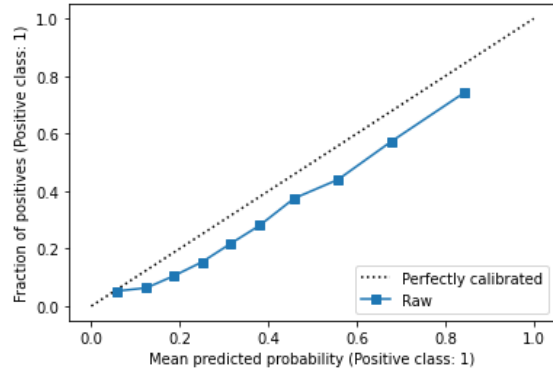
(c) Reliability plot: CatBoost (before calibration)



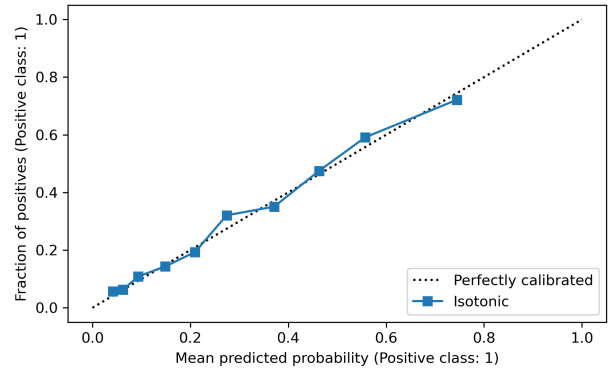
(d) Reliability plot: CatBoost (after calibration)

Figure 2: Cougher-level reliability plots for acoustic-only trained models.

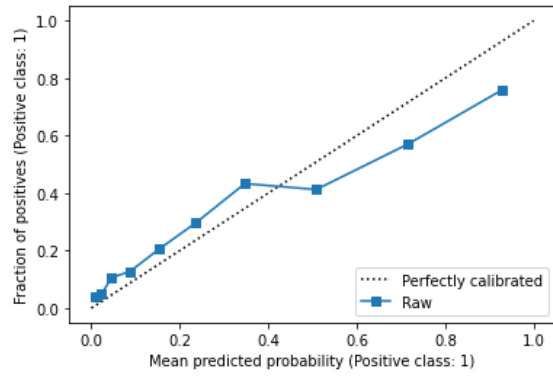
5.3 Waveform-level - Fused-features models



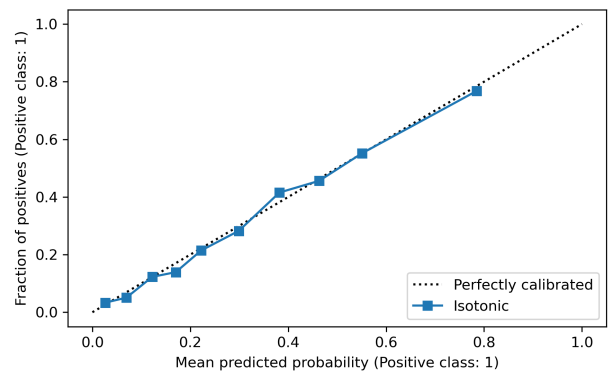
(a) Reliability plot: LR (before calibration)



(b) Reliability plot: LR (after calibration)



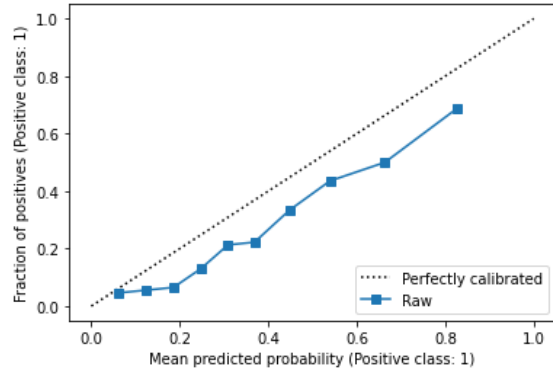
(c) Reliability plot: CatBoost (before calibration)



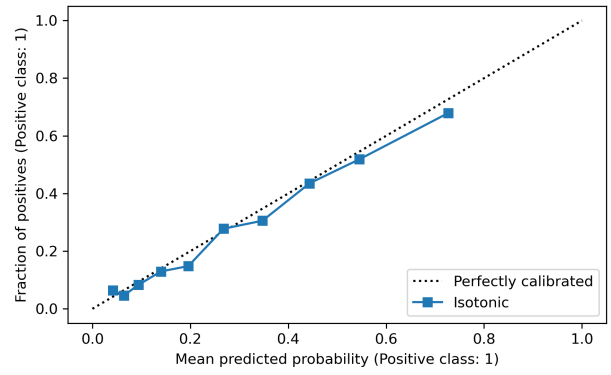
(d) Reliability plot: CatBoost (after calibration)

Figure 3: Waveform-level reliability plots for feature-fused trained models.

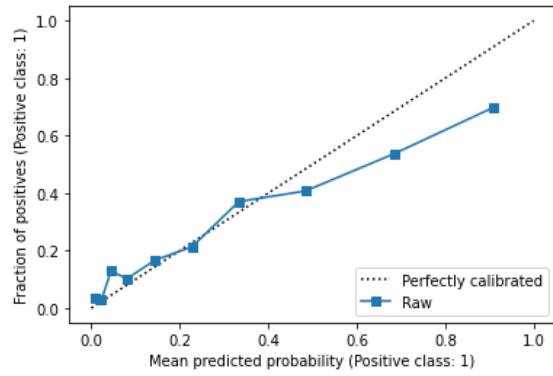
5.4 Cougher-level - Fused-features models



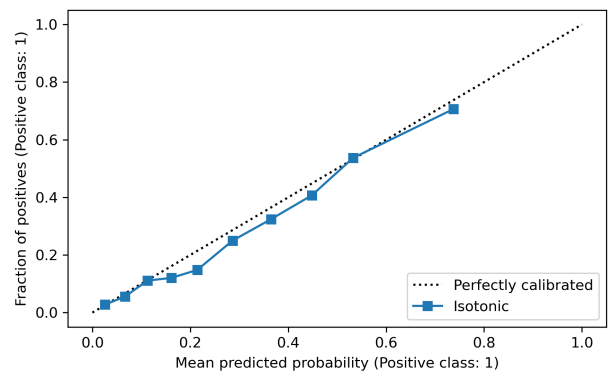
(a) *Reliability plot: LR (before calibration)*



(b) *Reliability plot: LR (after calibration)*



(c) *Reliability plot: CatBoost (before calibration)*



(d) *Reliability plot: CatBoost (after calibration)*

Figure 4: *Cougher-level reliability plots for feature-fused trained models.*

6 Selective correctness metrics

Let (x_i, y_i) denote a held-out (test) speaker, with true label $y_i \in \{0, 1\}$, calibrated positive-class probability $p_i = \hat{p}(y = 1 \mid x_i)$, and a point prediction obtained via a fixed threshold (here, Youden’s J threshold) as

$$\hat{y}_i = \mathbb{I}\{p_i \geq \tau_J\}.$$

Let $C_i = C(x_i) \subseteq \{0, 1\}$ be the conformal prediction set at miscoverage level α (reported at the speaker level). We define:

$$\text{Acc} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{\hat{y}_i = y_i\}, \quad \text{Acc} \mid \text{singleton} = \frac{1}{n_s} \sum_{i: |C_i|=1} \mathbb{I}\{\hat{y}_i = y_i\},$$

$$\text{Acc} \mid \text{ambiguous} = \frac{1}{n_a} \sum_{i: |C_i|>1} \mathbb{I}\{\hat{y}_i = y_i\}, \quad P(\text{singleton} \mid \text{correct}) = \frac{\sum_{i=1}^n \mathbb{I}\{|C_i| = 1 \wedge \hat{y}_i = y_i\}}{\sum_{i=1}^n \mathbb{I}\{\hat{y}_i = y_i\}}.$$

Macro values are computed per outer fold and summarized as mean \pm standard deviation across folds. **Pooled** values are computed by aggregating all held-out coughers across folds (micro-average).

6.1 LR-audio only

| α | Overall point accuracy | | Acc singleton | | Acc ambiguous | | P(singleton correct) | |
|----------|------------------------|--------|-------------------|--------|-------------------|--------|----------------------|--------|
| | Macro | Pooled | Macro | Pooled | Macro | Pooled | Macro | Pooled |
| 0.10 | 0.622 \pm 0.064 | 0.622 | 0.779 \pm 0.090 | 0.781 | 0.417 \pm 0.051 | 0.421 | 0.697 \pm 0.090 | 0.701 |
| 0.05 | 0.622 \pm 0.064 | 0.622 | 0.836 \pm 0.085 | 0.829 | 0.503 \pm 0.083 | 0.506 | 0.479 \pm 0.102 | 0.477 |

Table 5: Selective correctness metrics for LR trained on audio-only features (cougher-level). Macro values are mean \pm standard deviation across outer folds; pooled values aggregate all held-out coughers across folds.

6.2 CB-audio only

| α | Overall point accuracy | | Acc singleton | | Acc ambiguous | | P(singleton correct) | |
|----------|------------------------|--------|-------------------|--------|-------------------|--------|----------------------|--------|
| | Macro | Pooled | Macro | Pooled | Macro | Pooled | Macro | Pooled |
| 0.10 | 0.613 \pm 0.069 | 0.613 | 0.752 \pm 0.112 | 0.744 | 0.433 \pm 0.064 | 0.437 | 0.695 \pm 0.092 | 0.695 |
| 0.05 | 0.613 \pm 0.069 | 0.613 | 0.858 \pm 0.119 | 0.843 | 0.454 \pm 0.079 | 0.460 | 0.552 \pm 0.104 | 0.549 |

Table 6: Selective correctness metrics for CatBoost trained on audio-only features (cougher-level). Macro values are mean \pm standard deviation across outer folds; pooled values aggregate all held-out coughers across folds.

6.3 LR-fusion

| α | Overall point accuracy | | Acc singleton | | Acc ambiguous | | P(singleton correct) | |
|----------|------------------------|--------|-------------------|--------|-------------------|--------|----------------------|--------|
| | Macro | Pooled | Macro | Pooled | Macro | Pooled | Macro | Pooled |
| 0.10 | 0.723 ± 0.070 | 0.723 | 0.842 ± 0.074 | 0.842 | 0.472 ± 0.063 | 0.471 | 0.788 ± 0.062 | 0.790 |
| 0.05 | 0.723 ± 0.070 | 0.723 | 0.920 ± 0.045 | 0.919 | 0.537 ± 0.093 | 0.534 | 0.623 ± 0.060 | 0.623 |

Table 7: Selective correctness metrics for LR trained on fused (acoustic+clinical) features (cougher-level). Macro values are mean \pm standard deviation across outer folds; pooled values aggregate all held-out coughers across folds.

6.4 CB-fusion

| α | Overall point accuracy | | Acc singleton | | Acc ambiguous | | P(singleton correct) | |
|----------|------------------------|--------|-------------------|--------|-------------------|--------|----------------------|--------|
| | Macro | Pooled | Macro | Pooled | Macro | Pooled | Macro | Pooled |
| 0.10 | 0.703 ± 0.084 | 0.703 | 0.802 ± 0.108 | 0.799 | 0.492 ± 0.076 | 0.486 | 0.787 ± 0.039 | 0.788 |
| 0.05 | 0.703 ± 0.084 | 0.703 | 0.902 ± 0.053 | 0.899 | 0.519 ± 0.119 | 0.524 | 0.615 ± 0.099 | 0.611 |

Table 8: Selective correctness metrics for CatBoost trained on fused (acoustic+clinical) features (cougher-level). Macro values are mean \pm standard deviation across outer folds; pooled values aggregate all held-out coughers across folds.