



Understanding MammoScreen Score™

MammoScreen is an AI-driven concurrent reading assistive tool designed to aid radiologists in interpreting mammograms. This document outlines how users can understand and interpret the outputs generated by MammoScreen.

The MammoScreen Score definition

MammoScreen detects and characterizes lesions within mammography images. For each detected lesion, MammoScreen assigns a score from 2 to 10, the MammoScreen Score 1 being assigned for breasts where no lesions are detected. A low score (1-4) indicates low suspicion of malignancy, while a high score (7-10) indicates high suspicion of malignancy, an intermediate score (5 or 6) is provided for indeterminate suspicion. Those three main categories are translated and displayed as color-coded categories: green (low suspicion), yellow (indeterminate suspicion), and red (high suspicion). The Table 1 below illustrates the different score categories.

Detected lesions are displayed only if assigned with a score between 5 and 10, because the Positive Predictive Value (PPV) of those categories (i.e., yellow and red) exceeds 2%.

Further stratification within the three main categories is intended to add granularity: the more extreme the score (e.g., 1 or 10), the greater the software's confidence in its assessment.

MammoScreen assigns a score to each breast based on the highest score assigned to a lesion, and the overall score for a study is the higher of the two breast scores.

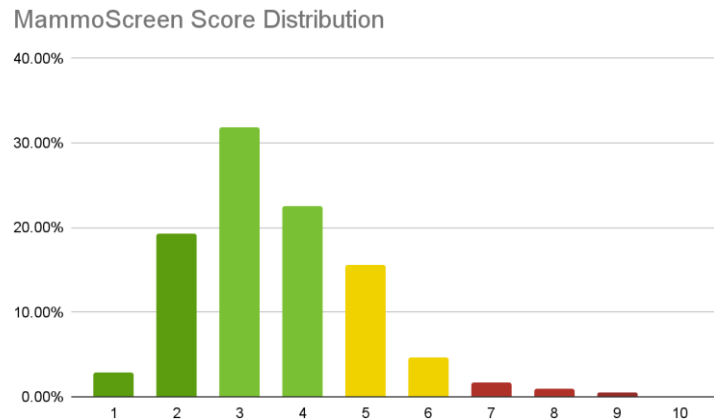
Table 1 MammoScreen Score categories

1	2	3	4	5	6	7	8	9	10
Lowest suspicion		Low suspicion		Indeterminate		High suspicion		Highest suspicion	

MammoScreen Score distribution

The distribution of MammoScreen Score in a typical screening population is shown on Figure 1 below, based on real-world data from 99,190 exams.

Figure 1 MammoScreen Score distribution, in a typical screening setting



When grouped by categories, the distribution (Table 2) shows a clear clustering: a high number of low-suspicion cases, a low number of high-suspicion cases, and a reasonable number of indeterminate cases. This pattern suggests that the MammoScreen Score could serve as a useful classifier of cases requiring attention.

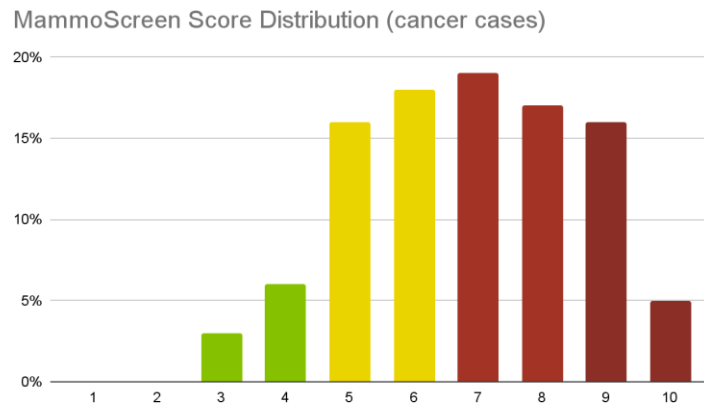
Table 2 MammoScreen Score distribution per scores categories

1	2	3	4	5	6	7	8	9	10
76.6%				20.2%		3.1%			

Cancer distribution

The MammoScreen Score on cancer cases is shown below. The data set is the same as above, where the overall cancer prevalence was 6 per 1,000 mammograms.

Figure 2 Percentage of cancer in the MammoScreen Score, in a typical screening setting



Grouping with the same categories as above, we observe the following cancer distribution (table 3), which also shows clear clustering and suggests a useful classifier of cases requiring attention.

Table 3 Percentage of cancer in the different MammoScreen Score categories, in a typical screening setting

1	2	3	4	5	6	7	8	9	10
8.6%				34.2%		57.2%			

MammoScreen Score as a cancer prevalence indicator

The per-score prevalence rates are provided in the Table 4 below.

Table 4 Cancer prevalence in the different MammoScreen Score categories, in a typical screening setting

1	2	3	4	5	6	7	8	9	10
1/1,520				1/100		1/9			

This grouping again suggests a clear clustering: the green category includes cases with low suspicion of malignancy Negative Predictive Value (NPV) of 99.93%, the yellow category is slightly above the natural



prevalence, and the red category consists of cases highly suggestive of malignancy (PPV of 11.11%), supporting its potential as a classifier for cases requiring attention.

Validation Results

Rigorous internal benchmark tests are also conducted to evaluate model performances, and notably to evaluate improvements of every new version over the previous ones.

For these tests, a dataset from three U.S. sites was used, including 1,053 cancer cases. Table 5 below shows the distribution of MammoScreen Score categories and corresponding cancer prevalence from these benchmark tests:

Table 5 Percentage of cases and cancer prevalence in the different MammoScreen Score categories, based on in silico tests with a typical screening distribution

	1	2	3	4	5	6	7	8	9	10
Percentage of cases	16%		56%			24%		3%		1%
Cancer prevalence	1/32,000		1/1,300			1/164		1/20		1/3

With this dataset, NPV for green was 99.94%, prevalence for yellow was 6 per 1000, and PPV for red was 9%.

The MammoScreen Score distribution and cancer prevalence observed in the real-world dataset align well with the results from these internal benchmark tests.

This consistency suggests that the MammoScreen Score remains stable across different settings, confirming the robustness of the algorithm and its generalizability.

Consequently, one can expect an improvement of radiologists' performance when using MammoScreen. During a retrospective Multi-Reader Multi-Case study¹, all 25 participating radiologists—both breast imaging fellowship-trained specialists and general radiologists—showed improved performance, regardless of the experience level. For a typical screening distribution, cancer detection increased by 12%, and recall rate decreased by 11%². MammoScreen has also shown to reduce inter-reader variability. The Intra-Class Correlation Coefficient (ICC) increased from 0.60 in the unassisted condition to 0.74 with

¹ S. Pacile & Al. Evaluation of a multi-instant multi-modal AI system supporting interpretive and noninterpretive functions. Journal of Breast Imaging, 2024; wbae062.

² Based on Reference 2, the increase in cancer detection rate was derived from the 8-percentage point improvement in sensitivity. The reduction in recall rate was estimated by applying the natural disease prevalence to the study results, which demonstrated a 2-percentage point gain in specificity.



MammoScreen assistance, suggesting a more standardized level of care for individuals undergoing screening mammography.

Overall, these results demonstrate the score's utility as a classifier, its role in improving radiologists' performance, how it generalizes to all categories of reader and thus also improves inter-reader variability.

Conclusion

The MammoScreen Score is a robust, understandable and meaningful tool for screening mammography.

- The robustness of the score has been verified by its stable distribution in multiple settings and geographies.
- The Score is understandable as the same metric is used across lesion, breasts and case level.
- The Score is meaningful as its distribution and cancer prevalence can be used as a robust classifier of the cases requiring attention, and reliable support for recall decision making.
- The Score leads to better performance for all radiologists individually, regardless of their experience, and collectively as the inter-reader variability is reduced.