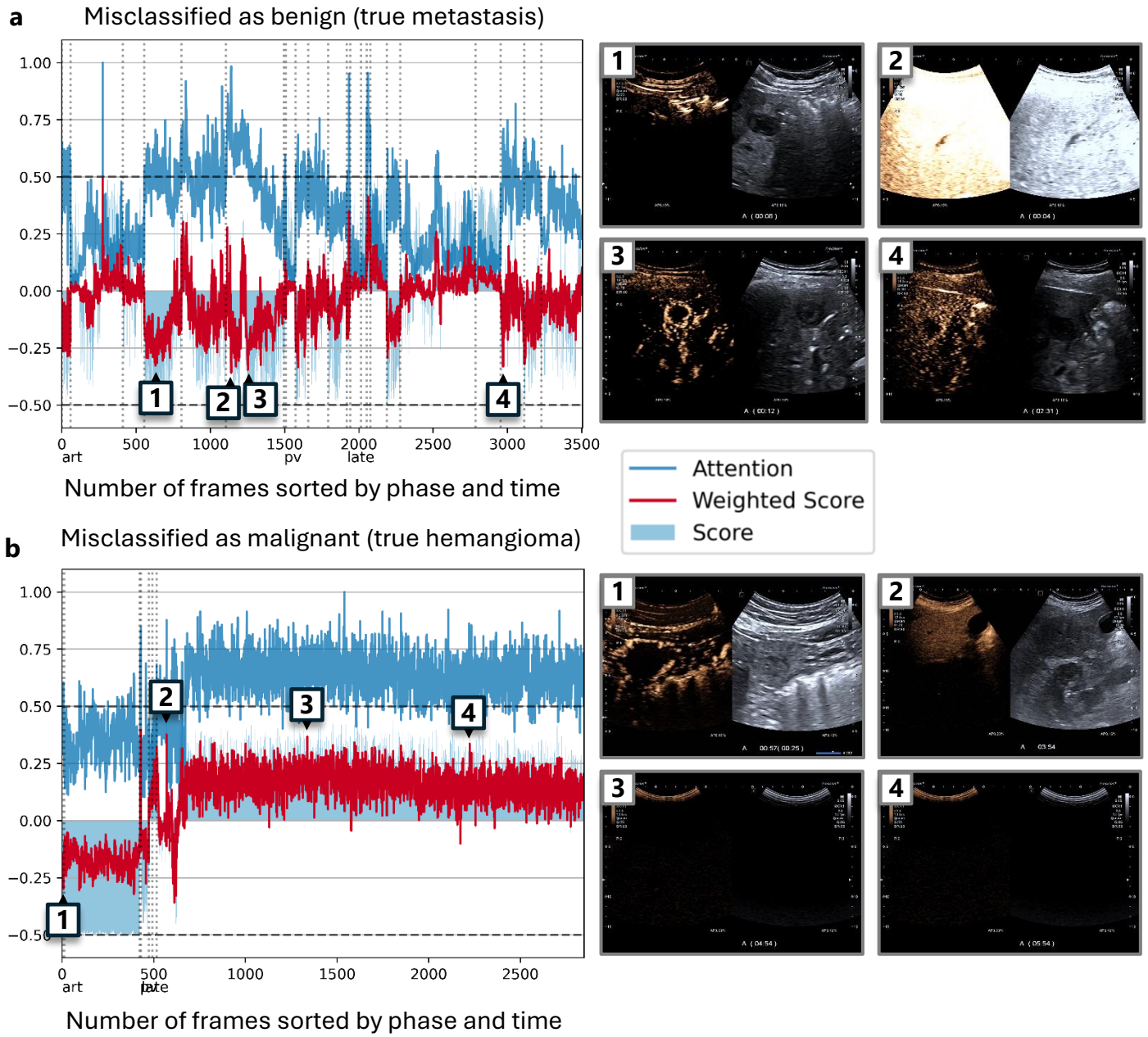

Supplemental Figure 1. Attempting to uncover the combined model's basis of decision-making for misclassified cases. The explainability method is shown for two exemplified lesions of patients misclassified as **a) benign** (true metastasis) and **b) malignant** (true hemangioma). The weighted score (red line), the multiplication of attention (blue line) and raw prediction score (light blue vertical bar), is plotted for each image in the patient's examination as calculated by the model. Within the plots attention varies between 0 and 1 for low and high attention, respectively; the (weighted) score varies between -0.5 and 0.5 for "definitely benign" and "definitely malignant" diagnoses, with 0 representing "unsure". The calculated results are sorted on the x-axis in the following order: arterial phase (art), portal venous (pv), late phase; standalone images are placed before those extracted from videos. Gray dashed lines indicate changes from one video to another. Adjacent: the four most important images as decided by the model, numbers indicate chronologically labeled location on the plot axis. In **a)** no distinct pattern of the model's attention can be seen, with the highest scoring pictures depicting a so-called flash and a frame showing the lesion in arterial phase depicting features similar to a hemangioma. In **b)** the model leaned correctly toward benign during the arterial phase. However, in the late phase, the weighted scores inverted toward malignant, likely due to confusion caused by an artifact generated when the ultrasound probe was removed, but the video continued recording. Not all frames deemed relevant by the model displayed distinctive features recognizable as relevant by human experts.