1 Supplementary

1.1 Preliminary analysis

We perform a preliminary analysis where we examine the link between the crevasse binary map (CNN and NeRD) and the ice rate factor field A using two different ice rate factor thresholds: an A value for temperate ice and an A value 10 times large. We evaluate the confusion matrix charts (Fig. S1) and F1 scores (Table T1 and T2) to assess the performance of the model and assess whether the dataset is unbalanced. The confusion matrix chart summarises the four possible outcomes of a binary classification analysis between predicted and true classes: True positive, True Negative, False Positive and False Negative. These values are used to estimate different performance metrics such as accuracy, precision, recall and F1 score, which allow us to quantify how often the model makes correct and incorrect predictions for the presence of crevasses at a given location. The precision metric measures the proportion of true positive predictions among all positive predictions, while the recall metric is the proportion of true positive prediction among all true positive cases. The F1 score is the "harmonic mean" of these two metrics, and ranges from 0 to 1 (Rizk et al., 2019). It is useful when both false positives and false negatives are equally undesirable, and the dataset is unbalanced.

Since the crevasse datasets are highly skewed towards the non-crevassed class (99 %), with such an imbalanced distribution of classes, a "naïve" classifier will likely show a biased high accuracy, as it would predict most non-crevassed regions correctly, while performing poorly on the minority class. This is clearly reflected in the confusion matrix charts and the F1 scores reported in Fig. S1 and Tables T1 and T2, for Filchner-Ronne and Pine Island ice shelves, respectively. The confusion charts summarise the results of a binary classification analysis, between predictions and observations. In both cases, for all crevasses and heavily damaged crevasses, the non-crevassed region is correctly detected, yet the model fails to detect the crevasse class, which is a minority in the dataset. This is further reflected in the F1 score analysis. A high weighted F1 score is measured for both *A-threshold*s, which would suggest that the model identifies crevasses accurately while minimizing incorrect hits and misses. However, these scores are unreliable as the majority no-crevasses class, which dominates the sample, provides very high performances compared to the minority crevasse class (0.0074 and 0.0024, in Table T1), overall dominating the final score. This is visible on both Filchner-Ronne and Pine Island ice shelves, as most samples belongs to the non-crevassed class. As a result, the F1 scores for all crevasse maps and *A-threshold*s are considerably influenced by the majority class, making it challenging to evaluate accurately the model's performance on the minority (crevasse) class.

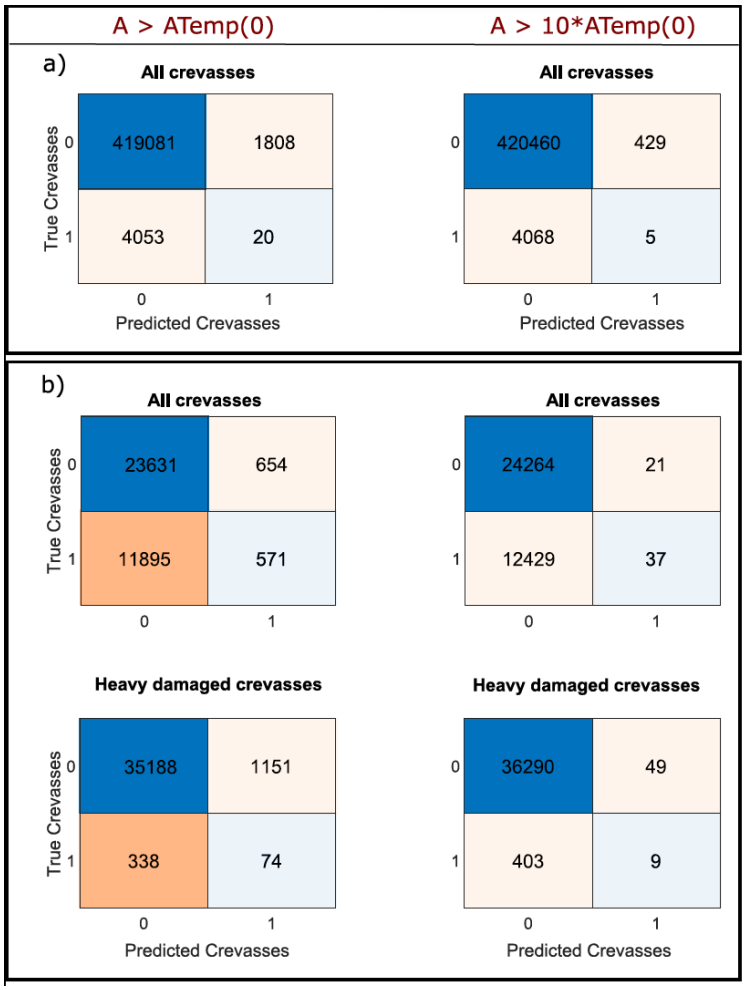

Figure S1 Confusion matrices for two classification cases, when *A* is greater than *A* for temperate ice measured by experiments and when *A* is ten times greater than *A* for temperate ice, for Filchner-Ronne Ice Shelf for all crevasses (panel a), and for Pine Island Ice Shelf for all and heavily damaged crevasses (panel b). All matrices summarize the predicted and true class labels. Each table is divided into four quadrants, depicting the four outcomes of a binary classification analysis: true negative, true positive, false positive and false negative. In both cases (panels a and b), the non-crevassed region is correctly detected, thus providing a high "accuracy", which is biased and incorrect since the non-crevassed class is dominating the dataset.

Table T 1 Display of F1 scores for Filchner-Ronne Ice Shelf when relating the inverted ice rate factor field *A* and all crevasses as mapped by the CNN of Lai et al., (2020). The high weighted F1 score (0.9841 and 0.9854), for the two *"A-threshold"* cases (A(0°C) and 10*A(0 °C)) suggests that the model identifies crevasses accurately while minimizing incorrect hits and misses. These scores are unreliable because the no-crevasses class still dominates the sample (Weight_NoCrev, 99 %), resulting in high F1 scores (0.9936 and 0.9950) compared to the F1 scores of the minority crevasse class (0.0074 and 0.0024).

Filchner Ronne Ice Shelf		Weight_Crev	Weight_NoCrev	F1_Crev	F1_NoCrev	F1_Weighted
All Crevasses	A(0 °C)	0.0097	0.9903	0.0074	0.9936	0.9841
	10*A(0 °C)	0.0097	0.9903	0.0024	0.9950	0.9854

Table T 2 Display of F1 scores for Pine Island Ice Shelf for all crevasses and heavily damaged crevasses; Overall, a weighted F1 score of 0.53-0.55 and 0.96 -0.98 indicates that the model's overall performance across all classes is moderate to high. However, the performance of the model seems to be imbalanced towards the No-Crevasse class (which define 66 % and 99 % of the distribution respectively). As the F1 score for crevasses does not exceed 0.11, it suggests that the performance of the current crevasse detection model is weak when classifying crevasses and that certain techniques are necessary to balance the dataset.

Pine Island Ice Shelf		Weight_Crev	Weight_NoCrev	F1_Crev	F1_NoCrev	F1_Weighted
All Crevasses	A(0 °C)	0.34	0.66	0.10	0.78	0.55
	10* A(0 °C)	0.34	0.66	0.02	0.80	0.53
Highly Damaged Crev	A(0°C)	0.01	0.99	0.08	0.97	0.96
	10*A(0 °C)	0.01	0.99	0.11	0.99	0.98

1.2 Ice rate factor field *A* and satellite image for Filchner-Ronne and Pine Island ice shelves.

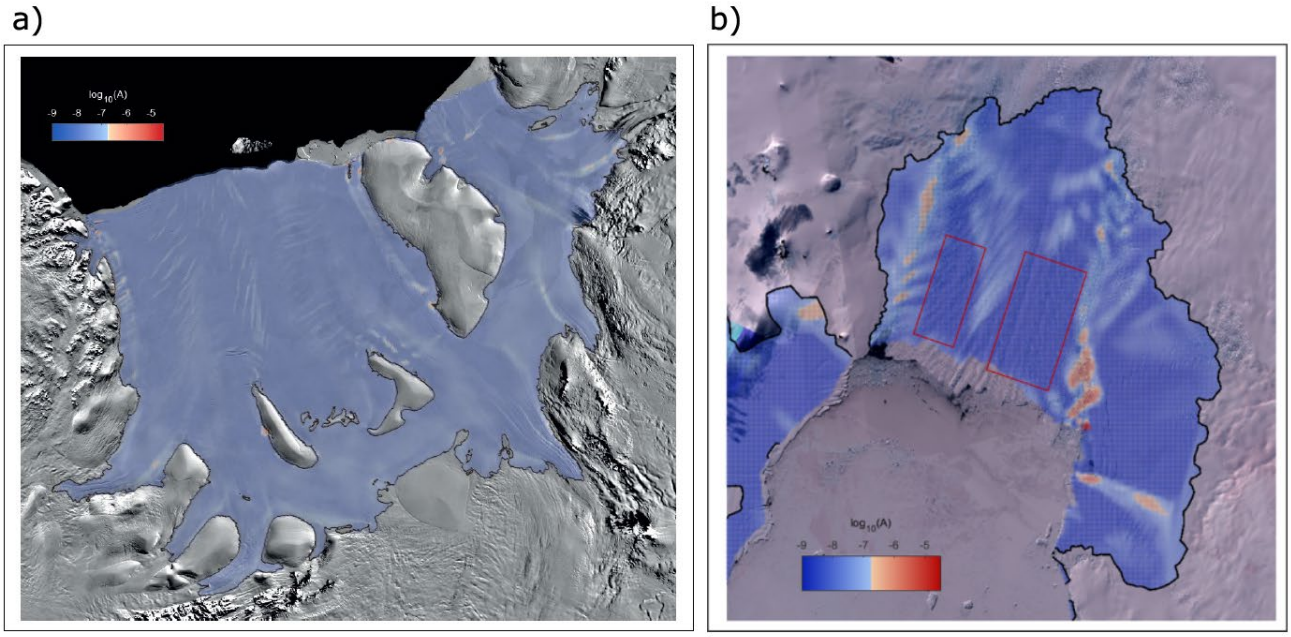

Figure S2 Detailed display of the inverted ice rate factor field *A* for Filchner-Ronne (a) and Pine Island Ice Shelf (b), using colour scale described in Fig. 1. Red boxes in b) display a discrepancy between areas displaying an ice rate factor well below the reference *A* for temperate ice and the abundance of remotely sensed crevasses depicted from the composite Sentinel (S2) image of austral summer 2019-2020.

1.3 Calculation of the Optimal Operating Point

To further identify an ideal *"A-threshold"* which best compromises the FPR and TPR in this analysis, we calculate the OPTtimal operating PoinT (OPTPT) on the ROC curve. The Optimal Operating point on the ROC curve is calculated by finding the *"A-threshold"* which best comprises the pair of False Positive Rate (FPR) and True Positive Rate (TPR). This is performed by finding the slope S of a line which satisfies the relationship:

$$S = \frac{\mathrm{Cost(P/N)} - \mathrm{Cost(N/N)}}{\mathrm{Cost(N/P)} - \mathrm{Cost(P/P)}} * \frac{N}{P}$$

where the terms "Cost(N/P)" and "Cost(P|N)" refer to the costs incurred from misclassifying a positive class as a negative class and a negative class as a positive class, respectively. P is the total number of positive instances, which is the sum of true positives (TP) and false negatives (FN). N is the total number of negative instances, which is the sum of true negatives (TN) and false positives (FP). The optimal operating point is found by moving the straight line with slope S from the upper left corner of the ROC plot (FPR = 0, TPR = 1) down and to the right, until it intersects the ROC curve (tangent to the curve).

1.4 Sensitivity analysis to changes of the regularisation parameter (γ_s)

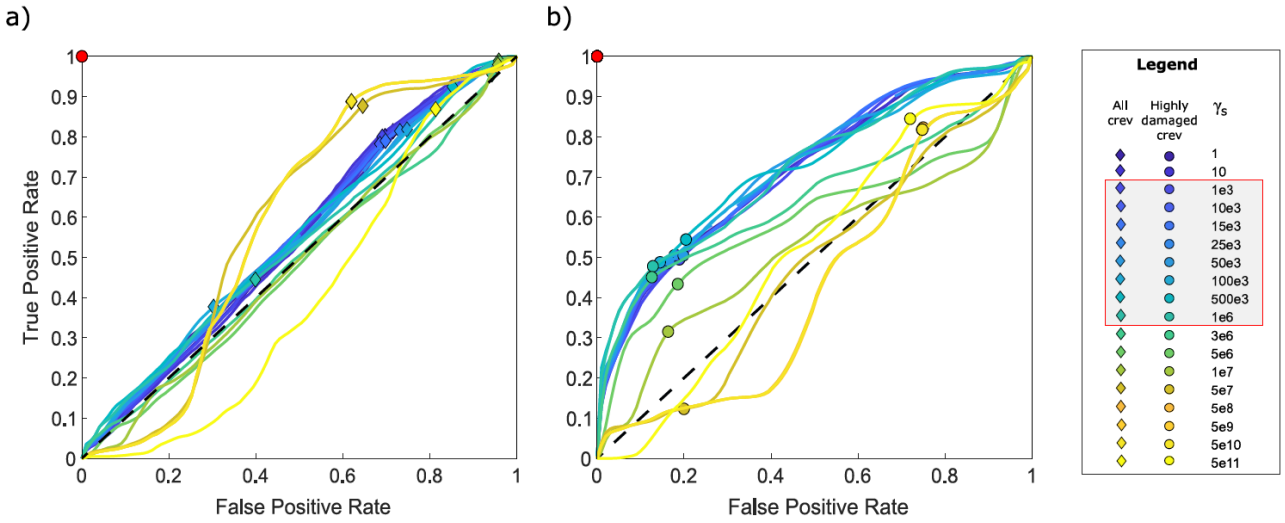

Figure S3: **ROC curve analysis for 2000 classification tests applied to Pine Island Ice Shelf, for all crevasses as mapped via NeRD (panel a) and for only heavily damaged crevasses (panel b), when testing for different regularisation values of γ_s and maintaining a constant γ_a=1. Notably, we observe a relatively consistent distribution of curves, up until γ_s reaches values > 10^6, at which point the modest observed relationship visible for heavily damaged crevasses no longer holds, and the ROC curves become close to that of a random classifier.**

1.5 Simulations fitting velocities for the month of November 2019 and February 2020 for Pine Island Ice Shelf

We conduct a comparative analysis evaluating the L-curve and ROC plot results for simulations whose ice rate factor fitted either the November 2019 (grey line) or the February 2020 (black line) velocities observed on Pine Island Ice Shelf. Both sets of simulations display a similar L-curve, with ROC plots results being insensitive to the value of the regularisation parameter γ_s when within the L-corner range.

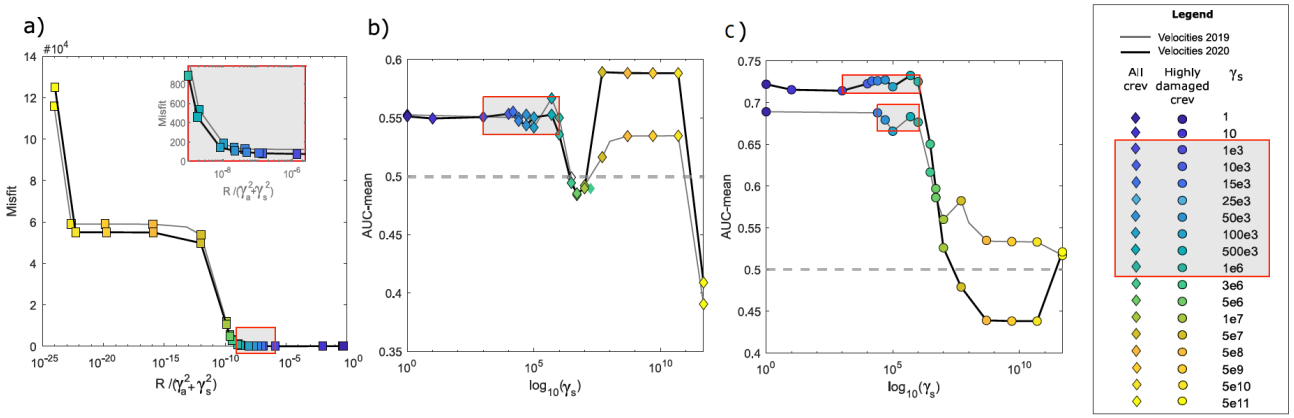

Figure S4 a) L-curve analysis comparing simulations when fitting velocities from November 2019 (grey line) and February 2020 (black line). A zoomed-in plot of L-corner range is displayed as a shaded grey box, with red contouring. b) AUC-mean across varying regularisation values (γ_s) for all crevasses as depicted by the NeRD method for Pine Island Ice Shelf, for November 2019 and February 2020; Values within the L-corner range are highlighted by the grey-shaded box. c) AUC-mean across varying regularisation values (γ_s) for all heavily damaged crevasses as depicted by the NeRD method for Pine Island Ice Shelf, for November 2019 and February 2020; Values within the L-corner range are highlighted by the grey-shaded box.

1.6 True Positive Rate and False positive rate for all crevasses and heavily damaged crevasses

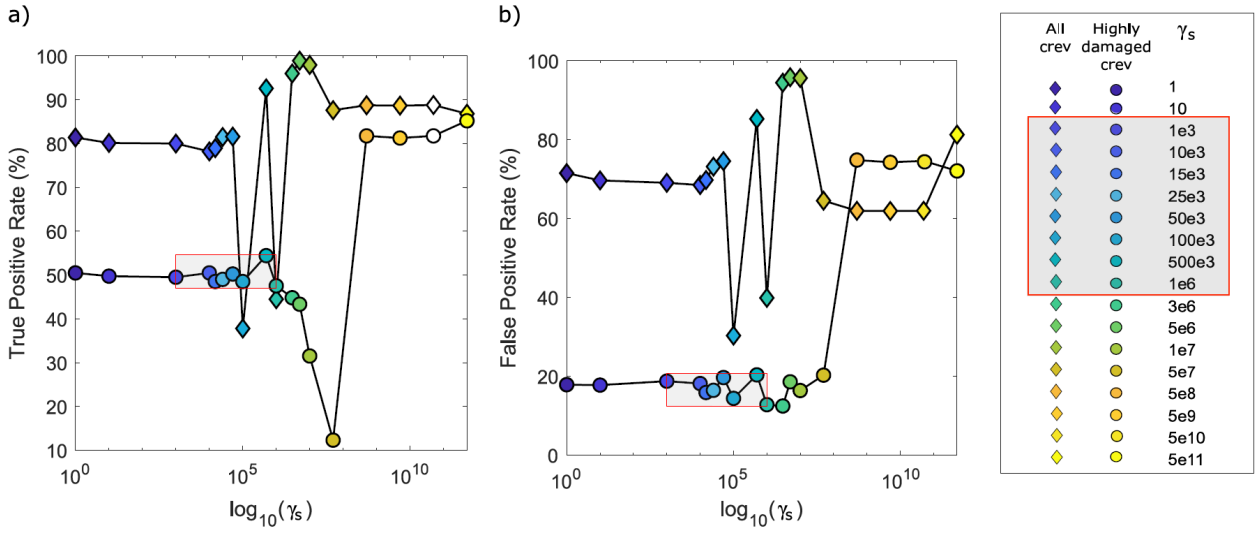

Figure S 5 True Positive Rate (a) and False Positive Rate (b) for Pine Island Ice Shelf across varying regularisation values (γ_s), for all crevasses (diamonds) and heavily damaged crevasses (circles, as in Figure 5) for simulations that fitted velocities for February 2020. Within each panel, a red-outlined box filled with grey shading highlights the acceptable values for γ_s —within the L-corner range.

1.7 Map of True Positive, False Positive and False Negative when applying the mean optimal *A-Threshold* estimated from the ROC curve.

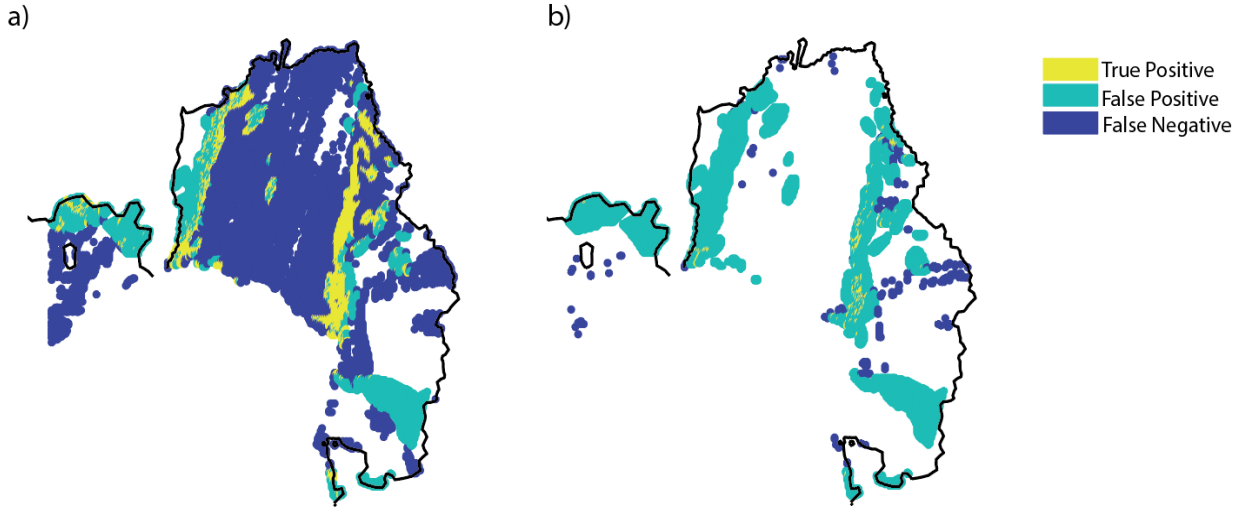

Figure S 6 True Positives (crevasses that are correctly detected as crevasses), False Positive (crevasses that are incorrectly predicted as crevasses) and False Negatives (crevasses that were missed) for Pine Island Ice Shelf, when applying the mean optimal *A-Threshold* estimated from the ROC curve analysis, for all crevasses (a), and for heavily damaged crevasses (b). Simulations inverted for A by fitting velocities for the month of February 2020, with a regularisation parameter of y_a = 1 and γ_s=25000.

References

Lai, C. Y., Kingslake, J., Wearing, M. G., Chen, P. H. C., Gentine, P., Li, H., Spergel, J. J., and van Wessem, J. M.: Vulnerability of Antarctica's ice shelves to meltwater-driven fracture, Nature, 584, 574–578, https://doi.org/10.1038/s41586-020-2627-8, 2020.

Rizk, Y., Hajj, N., Mitri, N., and Awad, M.: Deep belief networks and cortical algorithms: A comparative study for supervised classification, Appl. Comput. Informatics, 15, 81–93, https://doi.org/10.1016/J.ACI.2018.01.004, 2019.