Supplementary Material

1 Fine-tuning on MIMIC data

Table (1) Accuracy, F1-score and training time on GPU of fine-tuning on MIMIC data with the different optimiser using NN1, and at different learning and drop out rates.

	Accuracy (Learning rate, Drop out rate)					
Optimiser	(0.0001, 0.15)	(0.0001, 0.2)	(0.0001, 0.25)	(0.0005, 0.15)	(0.0005, 0.2)	(0.0005, 0.25)
Adam	0.902 ± 0.011	0.903 ± 0.012	0.903 ± 0.012	0.903 ± 0.012	0.903 ± 0.011	0.903 ± 0.012
AdamW	0.902 ± 0.013	0.903 ± 0.013	0.904 ± 0.012	0.908 ± 0.013	0.908 ± 0.011	0.906 ± 0.010
SGD	0.808 ± 0.025	0.816 ± 0.023	0.802 ± 0.028	0.872 ± 0.013	0.872 ± 0.013	0.872 ± 0.010
	F1-score (Learning rate, Drop out rate)					
Optimiser	(0.0001, 0.15)	(0.0001, 0.2)	(0.0001, 0.25)	(0.0005, 0.15)	(0.0005, 0.2)	(0.0005, 0.25)
Adam	0.910 ± 0.010	0.911 ± 0.011	0.911 ± 0.010	0.912 ± 0.011	0.911 ± 0.009	0.911 ± 0.011
AdamW	0.911 ± 0.011	0.912 ± 0.011	0.912 ± 0.011	0.915 ± 0.011	0.916 ± 0.010	0.914 ± 0.009
SGD	0.838 ± 0.020	0.845 ± 0.018	0.835 ± 0.022	0.885 ± 0.012	0.884 ± 0.011	0.885 ± 0.009
	Time (Min) (Learning rate, Drop out rate)					
Adam	52.012 ± 31.871	54.622 ± 36.095	53.690 ± 33.143	24.953 ± 11.361	28.247 ± 13.267	29.620 ± 12.057
AdamW	60.175 ± 36.935	57.892 ± 34.549	58.152 ± 35.111	28.470 ± 17.422	30.002 ± 21.738	26.483 ± 17.515
SGD	104.448 ± 43.670	104.433 ± 43.664	104.443 ± 43.626	94.610 ± 40.999	104.432 ± 43.617	103.918 ± 43.134

Table (2) Accuracy, F1-score and training time on GPU of fine-tuning on MIMIC data with the different optimiser using NN1, and at different learning and drop out rates.

	Accuracy (Learning rate, Drop out rate)					
Optimiser	(0.0001, 0.15)	(0.0001, 0.2)	(0.0001, 0.25)	(0.0005, 0.15)	(0.0005, 0.2)	(0.0005, 0.25)
Adam	0.943 ± 0.011	0.944 ± 0.010	0.942 ± 0.014	0.943 ± 0.015	0.945 ± 0.014	0.942 ± 0.012
AdamW	0.942 ± 0.007	0.945 ± 0.009	0.945 ± 0.012	0.945 ± 0.010	0.944 ± 0.012	0.944 ± 0.006
SGD	0.852 ± 0.021	0.850 ± 0.012	0.846 ± 0.019	0.925 ± 0.012	0.920 ± 0.013	0.922 ± 0.013
	F1-score (Learning rate, Drop out rate)					
Optimiser	(0.0001, 0.15)	(0.0001, 0.2)	(0.0001, 0.25)	(0.0005, 0.15)	(0.0005, 0.2)	(0.0005, 0.25)
Adam	0.947 ± 0.010	0.948 ± 0.010	0.946 ± 0.013	0.947 ± 0.014	0.949 ± 0.013	0.946 ± 0.011
AdamW	0.946 ± 0.007	0.949 ± 0.008	0.949 ± 0.011	0.949 ± 0.009	0.948 ± 0.012	0.948 ± 0.006
SGD	0.867 ± 0.019	0.865 ± 0.012	0.861 ± 0.017	0.932 ± 0.011	0.926 ± 0.012	0.929 ± 0.011
	Time (Min) (Learning rate, Drop out rate)					
Adam	10.450 ± 8.321	8.648 ± 2.085	10.335 ± 4.967	7.272 ± 2.570	9.092 ± 3.768	11.217 ± 3.999
AdamW	5.588 ± 0.549	6.293 ± 2.313	6.208 ± 2.604	6.108 ± 2.612	4.797 ± 0.670	5.293 ± 1.869
SGD	106.710 ± 44.473	115.078 ± 43.474	106.702 ± 44.418	96.905 ± 39.984	85.037 ± 39.787	88.497 ± 44.997

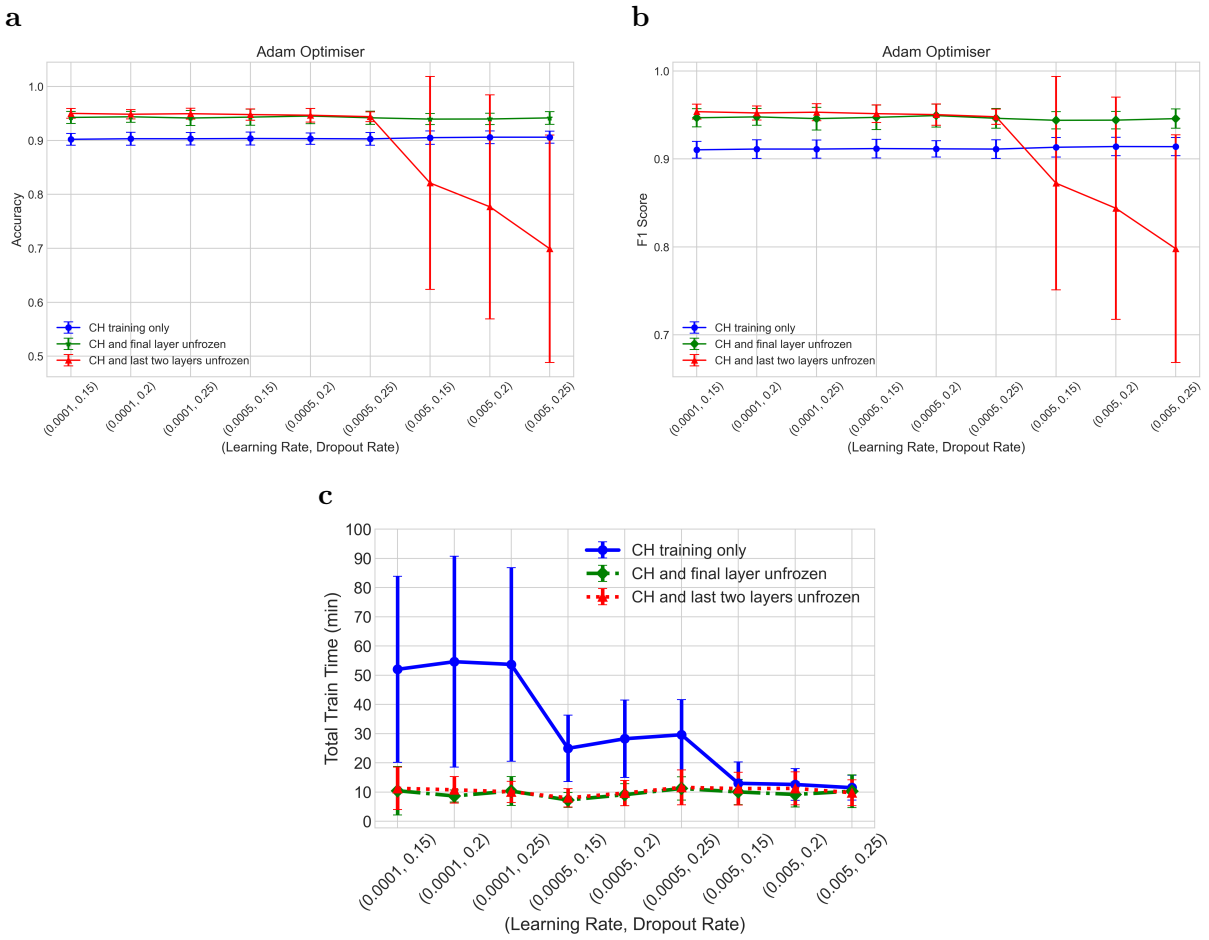

Figure (1) Results of fine-tuning using the Adam optimiser: (a) Accuracy of the validation data; (b) F-1 score on the validation data; (c) training time.

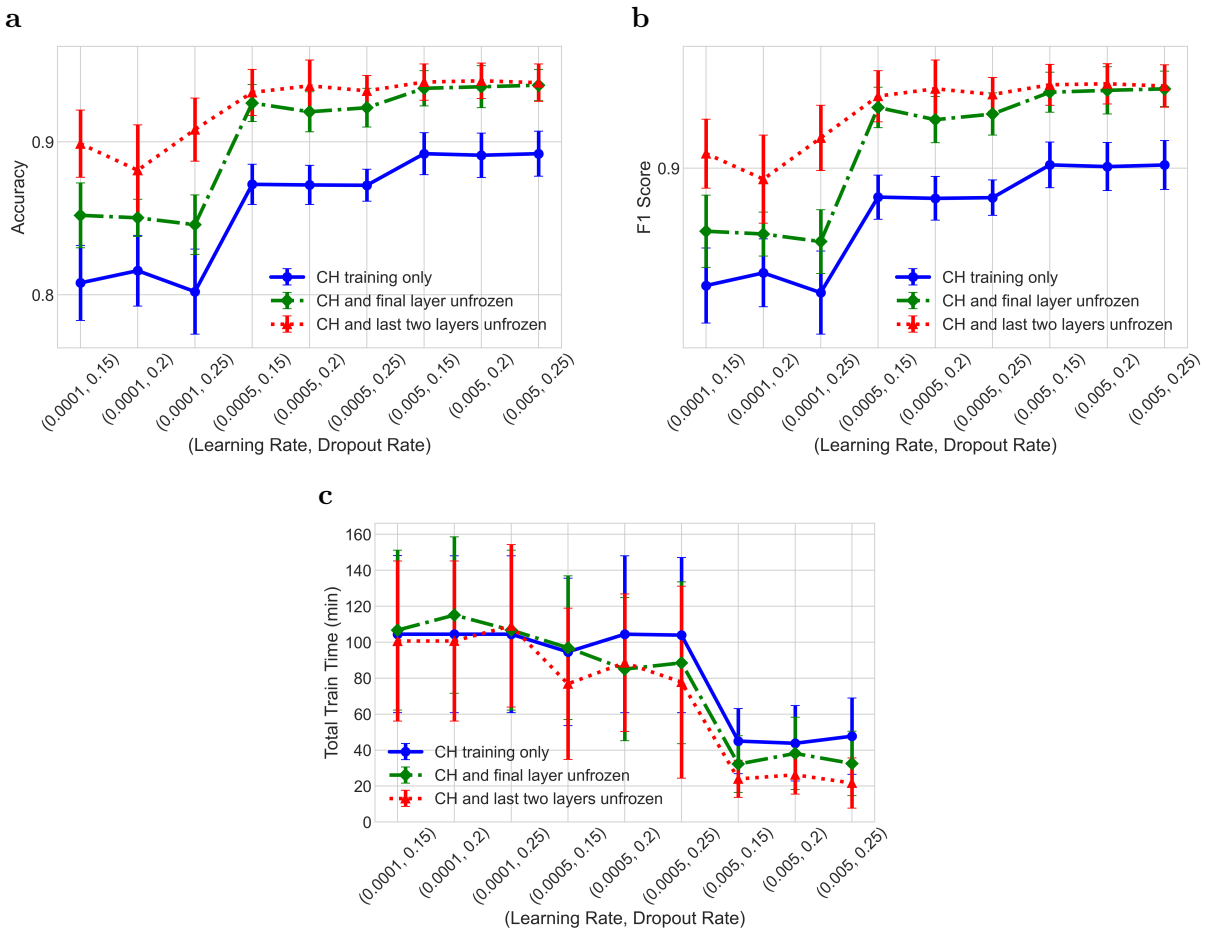

Figure (2) Results of fine-tuning using the SGD optimiser: (a) Accuracy of the validation data; (b) F-1 score on the validation data; (c) training time.

2 Prediction with fine-tuned models

Table (3) Prediction results on In-house Two using fine-tuned NN1 with the different optimisers and at different learning and drop out rates.

	Accuracy (Learning rate, Drop out rate)					
Optimiser	(0.0001, 0.15)	(0.0001, 0.2)	(0.0001, 0.25)	(0.0005, 0.15)	(0.0005, 0.2)	(0.0005, 0.25)
Adam	0.672 ± 0.008	0.672 ± 0.009	0.673 ± 0.007	0.683 ± 0.009	0.682 ± 0.004	0.686 ± 0.009
AdamW	0.675 ± 0.015	0.682 ± 0.011	0.682 ± 0.010	0.695 ± 0.012	0.696 ± 0.010	0.693 ± 0.008
SGD	0.521 ± 0.034	0.516 ± 0.024	0.519 ± 0.024	0.598 ± 0.012	0.596 ± 0.012	0.599 ± 0.016
	F1-score (Learning rate, Drop out rate)					
Optimiser	(0.0001, 0.15)	(0.0001, 0.2)	(0.0001, 0.25)	(0.0005, 0.15)	(0.0005, 0.2)	(0.0005, 0.25)
Adam	0.612 ± 0.013	0.612 ± 0.015	0.611 ± 0.015	0.628 ± 0.010	0.629 ± 0.008	0.630 ± 0.011
AdamW	0.616 ± 0.019	0.622 ± 0.012	0.621 ± 0.012	0.641 ± 0.010	0.640 ± 0.009	0.636 ± 0.009
SGD	0.561 ± 0.032	0.566 ± 0.018	0.558 ± 0.045	0.553 ± 0.015	0.543 ± 0.020	0.558 ± 0.019

Table (4) Prediction results on In-house Two using fine-tuned NN2 using the different optimisers and at different learning and drop out rates.

	Accuracy (Learning rate, Drop out rate)					
Optimiser	(0.0001, 0.15)	(0.0001, 0.2)	(0.0001, 0.25)	(0.0005, 0.15)	(0.0005, 0.2)	(0.0005, 0.25)
Adam	0.814 ± 0.020	0.818 ± 0.010	0.815 ± 0.014	0.818 ± 0.016	0.828 ± 0.013	0.824 ± 0.014
AdamW	0.812 ± 0.015	0.820 ± 0.013	0.808 ± 0.017	0.828 ± 0.010	0.825 ± 0.016	0.818 ± 0.014
SGD	0.596 ± 0.028	0.573 ± 0.047	0.593 ± 0.031	0.727 ± 0.019	0.722 ± 0.026	0.726 ± 0.020
	F1-score (Learning rate, Drop out rate)					
Optimiser	(0.0001, 0.15)	(0.0001, 0.2)	(0.0001, 0.25)	(0.0005, 0.15)	(0.0005, 0.2)	(0.0005, 0.25)
Adam	0.756 ± 0.019	0.756 ± 0.017	0.756 ± 0.019	0.763 ± 0.016	0.775 ± 0.021	0.769 ± 0.015
AdamW	0.762 ± 0.014	0.769 ± 0.013	0.761 ± 0.011	0.785 ± 0.012	0.781 ± 0.020	0.774 ± 0.016
SGD	0.521 ± 0.053	0.547 ± 0.049	0.519 ± 0.040	0.686 ± 0.020	0.665 ± 0.027	0.677 ± 0.013

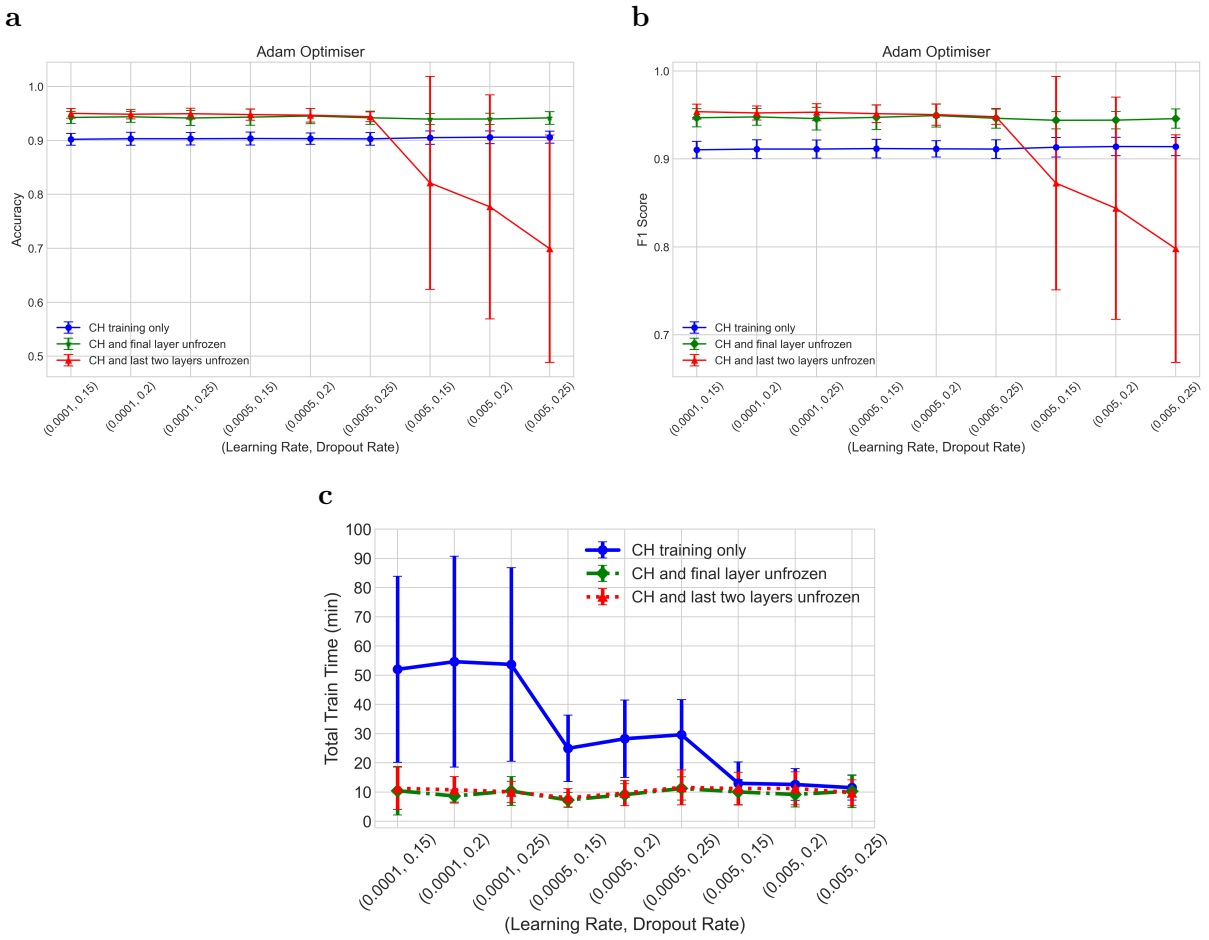

Figure (3) Results of prediction on In-house Two using models fine-tuned with the Adam optimiser: (a) Prediction accuracy; (b) F-1 score; (c) Prediction on CPU time.

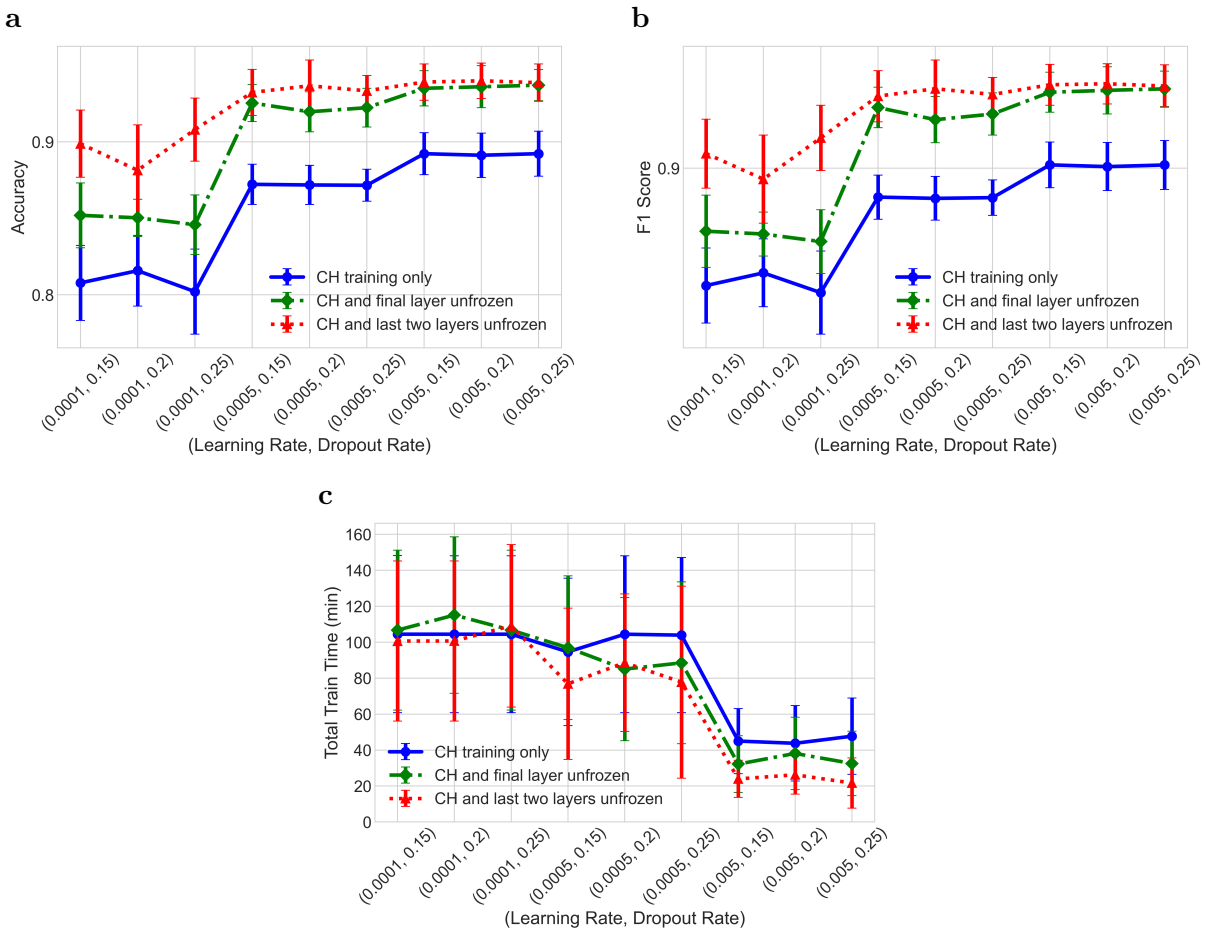

Figure (4) Results of prediction on In-house Two using models fine-tuned with the SGD optimiser: (a) Prediction accuracy; (b) F-1 score; (c) Prediction on CPU time.

3 Further fine-tuning on In-House One Data

Table (5) Accuracy, F1-score and training time on CPU of further fine-tuning (domain adaptation) on In-house One data with the different optimiser using NN2, and at different learning and drop out rates.

	Accuracy (Learning rate, Drop out rate)					
Optimiser	(0.0001, 0.15)	(0.0001, 0.2)	(0.0001, 0.25)	(0.0005, 0.15)	(0.0005, 0.2)	(0.0005, 0.25)
Adam	0.936 ± 0.008	0.929 ± 0.012	0.928 ± 0.011	0.938 ± 0.009	0.943 ± 0.011	0.940 ± 0.009
AdamW	0.923 ± 0.011	0.932 ± 0.015	0.922 ± 0.013	0.934 ± 0.013	0.941 ± 0.013	0.934 ± 0.006

	F1 Score (Learning rate, Drop out rate)					
Optimiser	(0.0001, 0.15)	(0.0001, 0.2)	(0.0001, 0.25)	(0.0005, 0.15)	(0.0005, 0.2)	(0.0005, 0.25)
Adam	0.930 ± 0.009	0.923 ± 0.013	0.922 ± 0.011	0.932 ± 0.009	0.937 ± 0.013	0.934 ± 0.009
AdamW	0.916 ± 0.011	0.927 ± 0.015	0.916 ± 0.014	0.928 ± 0.013	0.936 ± 0.014	0.928 ± 0.007

	Time (min) (Learning rate, Drop out rate)					
Optimiser	(0.0001, 0.15)	(0.0001, 0.2)	(0.0001, 0.25)	(0.0005, 0.15)	(0.0005, 0.2)	(0.0005, 0.25)
Adam	198.092 ± 59.439	184.190 ± 48.655	185.488 ± 45.923	131.315 ± 40.023	127.185 ± 63.776	98.440 ± 21.057
AdamW	111.487 ± 32.506	104.337 ± 24.244	178.830 ± 14.122	166.618 ± 32.676	148.200 ± 20.756	93.328 ± 21.506

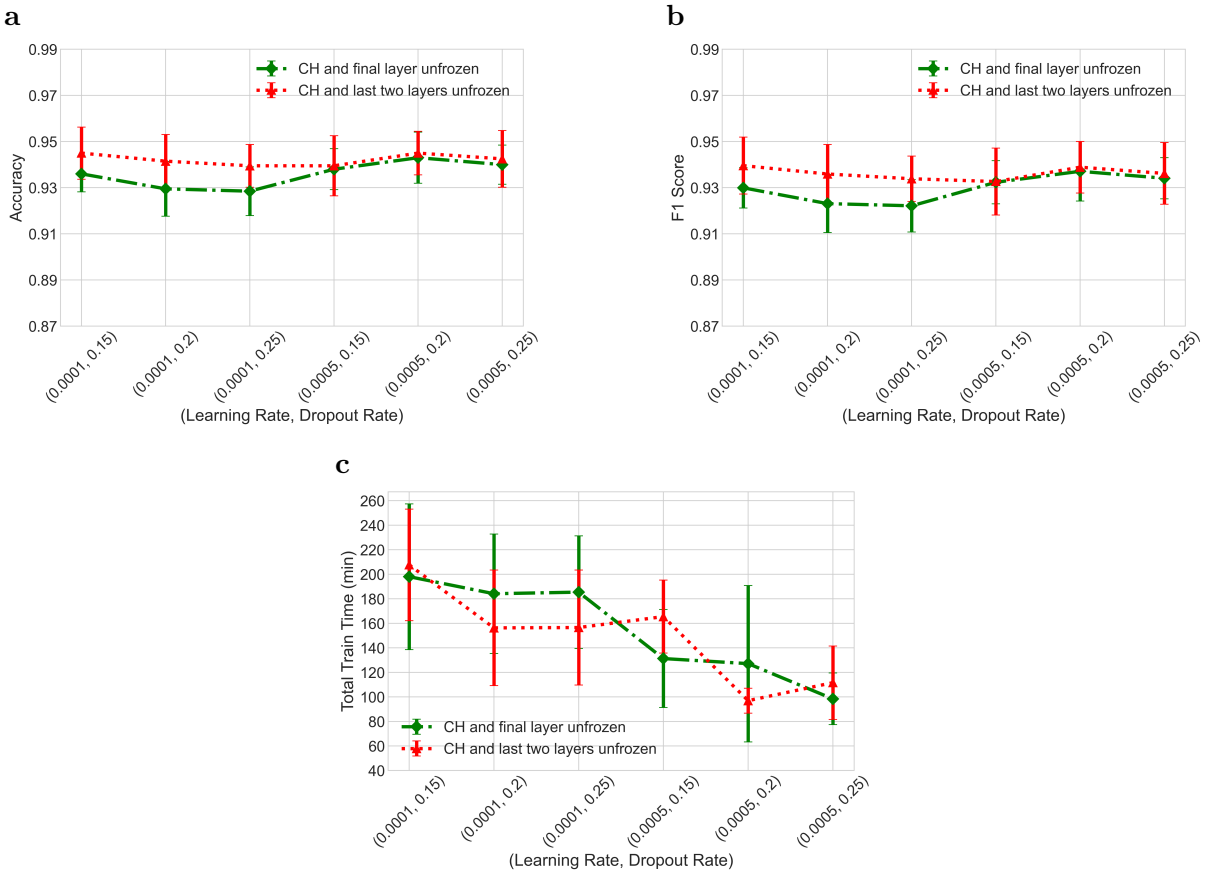

Figure (5) Results of further fine-tuning on In-house One data with the Adam optimiser: (a) Prediction accuracy; (b) F-1 score; (c) Training time on CPU (min).

4 Prediction with domain adapted models

Table (6) Accuracy, F1-score and training time on CPU of further fine-tuning (domain adaptation) on In-house One data with the different optimiser using NN2, and at different learning and drop out rates.

	Accuracy (Learning rate, Drop out rate)					
Optimiser	(0.0001, 0.15)	(0.0001, 0.2)	(0.0001, 0.25)	(0.0005, 0.15)	(0.0005, 0.2)	(0.0005, 0.25)
Adam	0.933 ± 0.004	0.931 ± 0.005	0.930 ± 0.004	0.933 ± 0.005	0.935 ± 0.004	0.932 ± 0.006
AdamW	0.925 ± 0.006	0.923 ± 0.005	0.923 ± 0.007	0.930 ± 0.003	0.931 ± 0.005	0.928 ± 0.005

	F1 Score (Learning rate, Drop out rate)					
Optimiser	(0.0001, 0.15)	(0.0001, 0.2)	(0.0001, 0.25)	(0.0005, 0.15)	(0.0005, 0.2)	(0.0005, 0.25)
Adam	0.922 ± 0.005	0.920 ± 0.006	0.919 ± 0.005	0.922 ± 0.005	0.924 ± 0.004	0.920 ± 0.006
AdamW	0.913 ± 0.007	0.911 ± 0.005	0.911 ± 0.007	0.918 ± 0.004	0.919 ± 0.006	0.916 ± 0.006

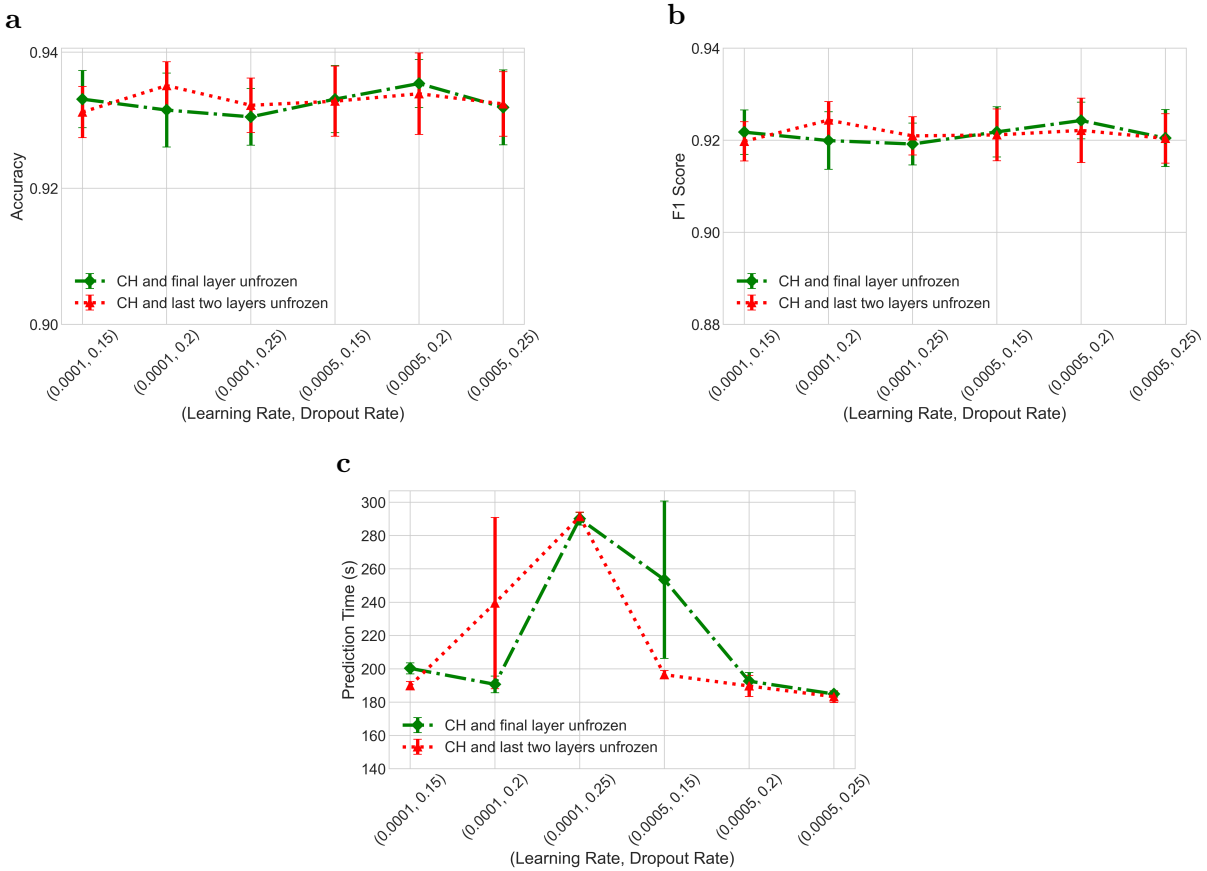

Figure (6) Results of prediction on In-house Two data with the domain adapted models optimised with Adam optimiser: (a) Prediction accuracy; (b) F-1 score; (c) Time taken on CPU (s)