



Weekly Meeting with Dr. Hannah

Hasan Shaikh

Quantitative Imaging Research and
Artificial Intelligence Lab (QIRAIL)

Table 1. Comparative AUC performance for three clinical outcomes in head and neck cancer across the original published study (“Paper Results”), our reproduced models on the same dataset (“Our Results”), and our proposed CNN trained on our in-house dataset (“Our Results with Our Dataset”)

Events	Cohort	Paper Results (CNN)	Our Results (CNN)	Our Dataset Results (CNN)
		Cohort split (CI 95%) / 5-fold CV	Cohort split (CI 95%) / 5-fold CV	Cohort split (CI 95%) / 5-fold CV
Distant Metastasis (2y)	Train	0.91 [0.84, 0.96] / 0.87 (0.84–0.92)	0.81 / 0.87	-
	Val	0.89 [0.81, 0.96] / 0.86 (0.77–0.96)	0.84 / 0.85	-
	Test	0.89 [0.79, 0.98] / 0.83 (0.76–0.90)	0.81 / 0.74	-
Locoregional failure (2y)	Train	0.76 [0.64, 0.88] / 0.77 (0.72–0.86)	0.71 / 0.81	-
	Val	0.77 [0.58, 0.92] / 0.76 (0.72–0.84)	0.72 / 0.80	-
	Test	0.45 [0.32, 0.57] / 0.53 (0.48–0.59)	0.49 / 0.57	-
Overall survival (4y)	Train	0.84 [0.75, 0.92] / 0.82 (0.68–0.94)	0.75 / 0.78	-
	Val	0.80 [0.66, 0.91] / 0.77 (0.62–0.96)	0.77 / 0.79	-
	Test	0.67 [0.57, 0.77] / 0.63 (0.57–0.72)	0.67 / 0.75	-

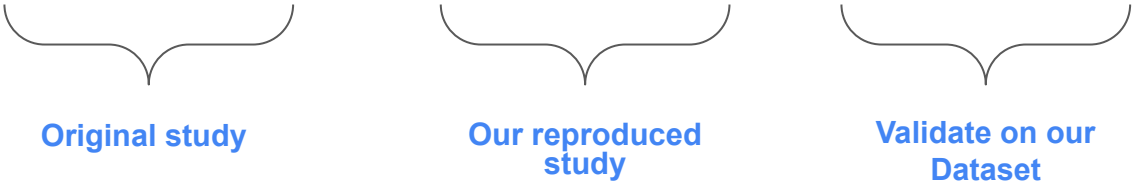


Table 2. Comparative performance (AUCs) including clinical data for predicting multiple clinical outcomes in head and neck cancer across the original study results (*Paper Results*), our reproduced CNN and ANN models (*Our Results*), and our proposed CNN trained on our in-house dataset (*Our Dataset Results*)

Events	Cohort	Paper Results (CNN)	Our Results (CNN)	Paper Results (ANN)	Our Results (ANN)	Our Dataset Results (CNN)
		Cohort split (CI 95%) / 5-fold CV	Cohort split (CI 95%) / 5-fold CV	Cohort split (CI 95%) / 5-fold CV	Cohort split (CI 95%) / 5-fold CV	Cohort split (CI 95%) / 5-fold CV
Distant Metastasis (2y)	Train	0.91 [0.86, 0.95] / 0.88 (0.81–0.93)	0.87 / 0.86	0.87 [0.78, 0.93] / 0.87 (0.81–0.92)	0.84 / 0.76	- / -
	Val	0.89 [0.79, 0.98] / 0.88 (0.81–0.93)	0.86 / 0.89	0.79 [0.65, 0.93] / 0.83 (0.79–0.88)	0.68 / 0.80	- / -
	Test	0.93 [0.86, 0.99] / 0.88 (0.86–0.90)	0.87 / 0.88	0.87 [0.78, 0.95] / 0.86 (0.81–0.89)	0.89 / 0.82	- / -
Locoregional failure (2y)	Train	0.84 [0.76, 0.93] / 0.77 (0.62–0.87)	0.75 / 0.92	0.71 [0.61, 0.80] / 0.74 (0.70–0.84)	0.78 / 0.78	- / -
	Val	0.70 [0.54, 0.84] / 0.72 (0.60–0.84)	0.70 / 0.87	0.66 [0.48, 0.82] / 0.71 (0.60, 0.81)	0.54 / 0.69	- / -
	Test	0.59 [0.47, 0.70] / 0.57 (0.53–0.60)	0.57 / 0.54	0.41 [0.29, 0.54] / 0.53 (0.50, 0.54)	0.33 / 0.40	- / -
Overall survival (4y)	Train	0.74 [0.64, 0.84] / 0.83 (0.74–0.94)	0.74 / 0.88	0.83 [0.74, 0.90] / 0.83 (0.77–0.85)	0.78 / 0.84	- / -
	Val	0.74 [0.58, 0.86] / 0.81 (0.73–0.93)	0.72 / 0.92	0.75 [0.62, 0.87] / 0.76 (0.71–0.78)	0.74 / 0.79	- / -
	Test	0.69 [0.59, 0.79] / 0.68 (0.63–0.71)	0.69 / 0.68	0.63 [0.52, 0.73] / 0.63 (0.61, 0.64)	0.65 / 0.56	- / -

