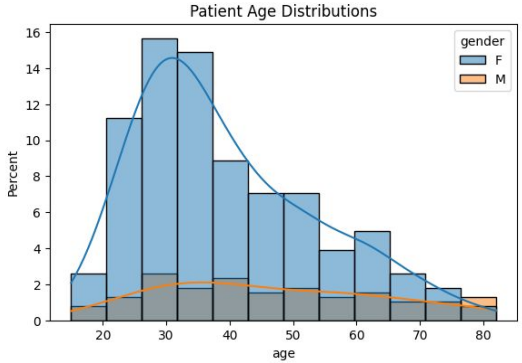
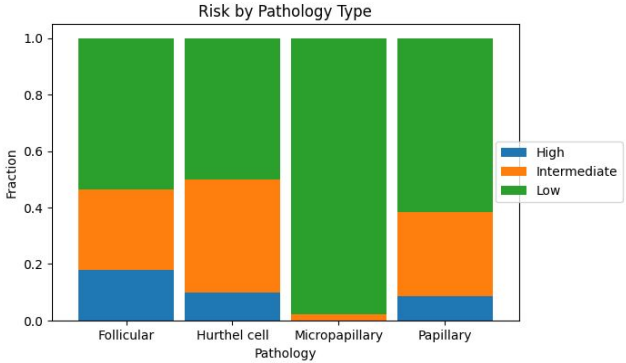
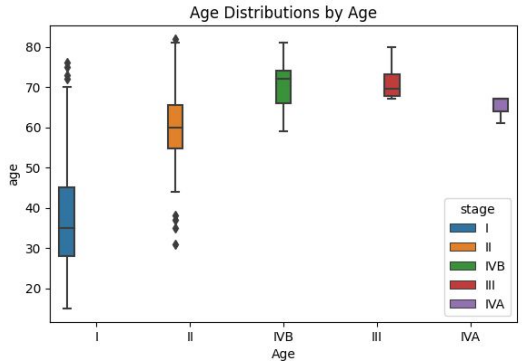
Thyroid Cancer Recurrence Risk Prediction

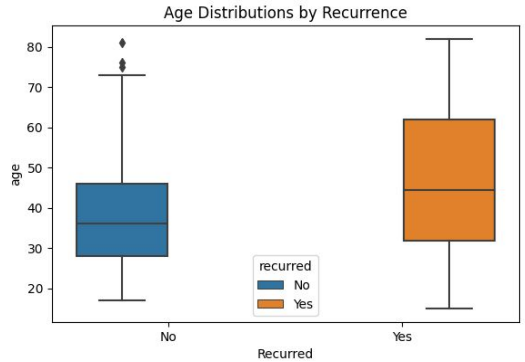
Cole Khamnei and Shree Rout

Thyroid Cancer Dataset Feature Distributions









Thyroid Cancer Dataset

Features		Train	Validation	Total
Gender	F	224 (79%)	88 (88%)	312 (81%)
	M	59 (21%)	12 (12%)	71 (19%)
Smoking	No	245 (87%)	89 (89%)	334 (87%)
	Yes	38 (13%)	11 (11%)	49 (13%)
History Of Smoking	No	259 (92%)	96 (96%)	355 (93%)
	Yes	24 (8%)	4 (4%)	28 (7%)
History Of Radiotherapy	No	277 (98%)	99 (99%)	376 (98%)
	Yes	6 (2%)	1 (1%)	7 (2%)
Thyroid Function	Euthyroid	245 (87%)	87 (87%)	332 (87%)
	Clinical Hyperthyroidism	17 (6%)	3 (3%)	20 (5%)
	Clinical Hypothyroidism	6 (2%)	6 (6%)	12 (3%)
	Subclinical Hyperthyroidism	4 (1%)	1 (1%)	5 (1%)
	Subclinical Hypothyroidism	11 (4%)	3 (3%)	14 (4%)
Physical Examination	Single nodular goiter-left	64 (23%)	25 (25%)	89 (23%)
	Multinodular goiter	105 (37%)	35 (35%)	140 (37%)
	Single nodular goiter-right	106 (37%)	34 (34%)	140 (37%)
	Normal	4 (1%)	3 (3%)	7 (2%)
	Diffuse goiter	4 (1%)	3 (3%)	7 (2%)
Adenopathy	No	206 (73%)	71 (71%)	277 (72%)
	Right	36 (13%)	12 (12%)	48 (13%)
	Extensive	6 (2%)	1 (1%)	7 (2%)
	Left	10 (4%)	7 (7%)	17 (4%)
	Bilateral	23 (8%)	9 (9%)	32 (8%)
	Posterior	2 (1%)	0 (0%)	2 (1%)
Pathology	Micropapillary	30 (11%)	18 (18%)	48 (13%)
	Papillary	217 (77%)	70 (70%)	287 (75%)
	Follicular	23 (8%)	5 (5%)	28 (7%)
	Hurthel cell	13 (5%)	7 (7%)	20 (5%)

Features		Train	Validation	Total
Focality	Uni-Focal	178 (63%)	69 (69%)	247 (64%)
	Multi-Focal	105 (37%)	31 (31%)	136 (36%)
Risk	Low	182 (64%)	67 (67%)	249 (65%)
	Intermediate	78 (28%)	24 (24%)	102 (27%)
	High	23 (8%)	9 (9%)	32 (8%)
T	T1a	31 (11%)	18 (18%)	49 (13%)
	T1b	29 (10%)	14 (14%)	43 (11%)
	T2	115 (41%)	36 (36%)	151 (39%)
	T3a	75 (27%)	21 (21%)	96 (25%)
	T3b	12 (4%)	4 (4%)	16 (4%)
	T4a	14 (5%)	6 (6%)	20 (5%)
	T4b	7 (2%)	1 (1%)	8 (2%)
N	N0	197 (70%)	71 (71%)	268 (70%)
	N1b	70 (25%)	23 (23%)	93 (24%)
	N1a	16 (6%)	6 (6%)	22 (6%)
M	M0	272 (96%)	93 (93%)	365 (95%)
	M1	11 (4%)	7 (7%)	18 (5%)
Stage	I	244 (86%)	89 (89%)	333 (87%)
	II	26 (9%)	6 (6%)	32 (8%)
	IVB	7 (2%)	4 (4%)	11 (3%)
	III	3 (1%)	1 (1%)	4 (1%)
	IVA	3 (1%)	0 (0%)	3 (1%)
Response	Indeterminate	46 (16%)	15 (15%)	61 (16%)
	Excellent	152 (54%)	56 (56%)	208 (54%)
	Structural Incomplete	65 (23%)	26 (26%)	91 (24%)
	Biochemical Incomplete	20 (7%)	3 (3%)	23 (6%)
Recurred	No	204 (72%)	71 (71%)	275 (72%)
	Yes	79 (28%)	29 (29%)	108 (28%)

Paper Methods: Model Selection

- Preprocessing: Standard scaling and one hot encoding
- CV Grid Search: All models were test using 2-fold cross validation for a variety of parameters
- Models with the highest AUC on CV were selected per paper
- Models:
  - K-Nearest Neighbors
  - Decision Trees
  - Random Forests
  - Artificial Neural Network
  - SVM
- New Models:
  - AdaBoost
  - Logistic Regression

Methods: Model Parameter Search

Paper Models:

- K-Nearest Neighbors: K: [1, 10],
- Decision Trees: max depth [2, 10], criterion (gini or entropy)
- Random Forests: N trees [20, 100], max depth [1, 10], criterion (gini or entropy)
- SVM: C [0.001, 1], Kernel (linear or RBF)
- Artificial Neural Network:
    - # hidden layers [1, 5]
    - # hidden layer sizes [20, 50]
    - Activation: relu or tanh
    - Dropout: [0, 1]
    - Learning rate: [0.001, 1]

New Models:

- AdaBoost: N trees [20, 100], max depth [1, 10], criterion (gini or entropy)
- Logistic Regression

Model Results: Metrics

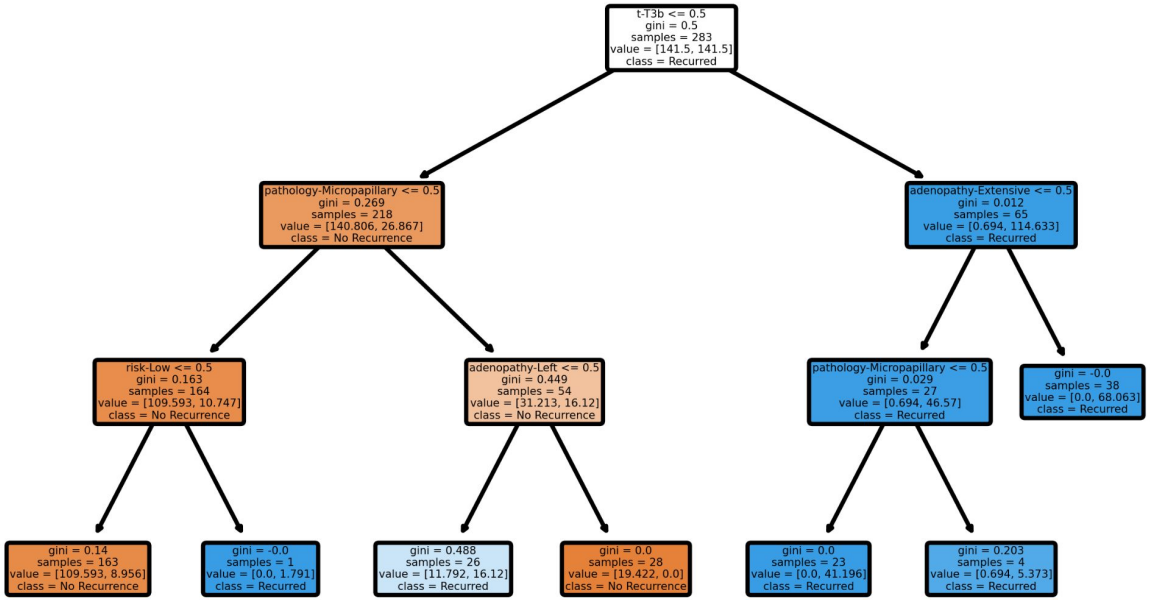
Paper Models:

	Feature Set	Sensitivity	Specificity	PPV	NPV	AUC	Accuracy
ANN	ATA Risk	86.20%	88.73%	75.75%	94.02%	89.21%	88.0%
	ATA Risk Excluded	93.10%	97.18%	93.10%	97.18%	97.96%	96.0%
	Full	89.65%	97.18%	92.85%	95.83%	98.49%	95.0%
DecisionTree	ATA Risk	86.20%	88.73%	75.75%	94.02%	89.21%	88.0%
	ATA Risk Excluded	89.65%	95.77%	89.65%	95.77%	93.54%	94.0%
	Full	82.75%	97.18%	92.30%	93.24%	91.64%	93.0%
KNN	ATA Risk	86.20%	88.73%	75.75%	94.02%	89.21%	88.0%
	ATA Risk Excluded	82.75%	98.59%	96.0%	93.33%	95.04%	94.0%
	Full	86.20%	97.18%	92.59%	94.52%	95.07%	94.0%
RandomForest	ATA Risk	86.20%	88.73%	75.75%	94.02%	89.21%	88.0%
	ATA Risk Excluded	93.10%	98.59%	96.42%	97.22%	99.17%	97.0%
	Full	89.65%	97.18%	92.85%	95.83%	98.73%	95.0%
SVM	ATA Risk	86.20%	88.73%	75.75%	94.02%	89.21%	88.0%
	ATA Risk Excluded	89.65%	98.59%	96.29%	95.89%	98.73%	96.0%
	Full	86.20%	95.77%	89.28%	94.44%	98.00%	93.0%

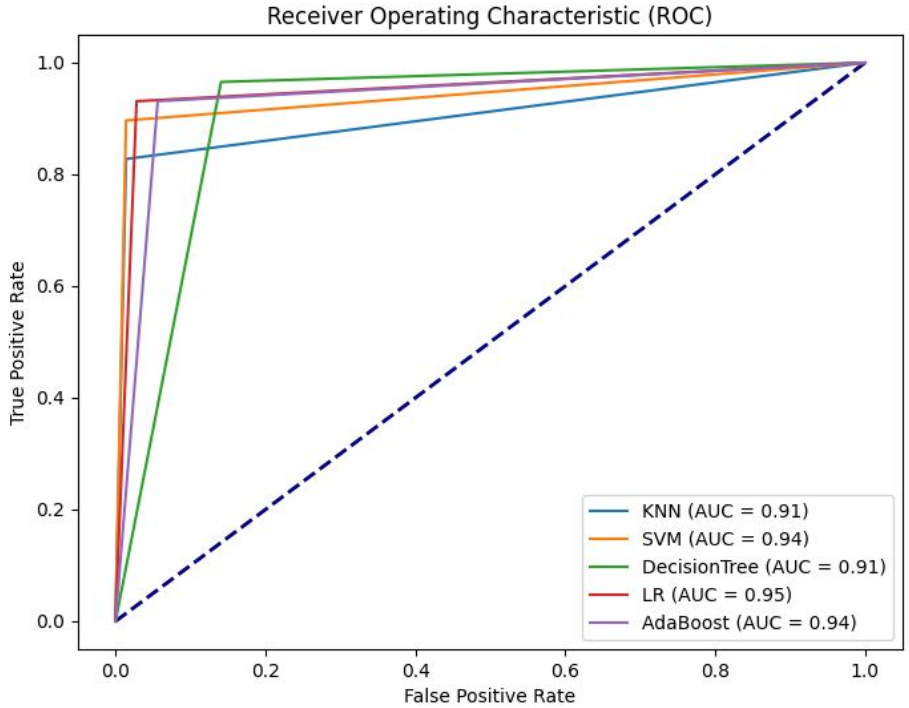
New Models:

	Feature Set	Sensitivity	Specificity	PPV	NPV	AUC	Accuracy
AdaBoost	ATA Risk	86.20%	88.73%	75.75%	94.02%	89.21%	88.0%
	ATA Risk Excluded	93.10%	94.36%	87.09%	97.10%	97.52%	94.0%
	Full	93.10%	94.36%	87.09%	97.10%	97.28%	94.0%
LR	ATA Risk	86.20%	88.73%	75.75%	94.02%	89.21%	88.0%
	ATA Risk Excluded	93.10%	97.18%	93.10%	97.18%	98.30%	96.0%
	Full	93.10%	95.77%	90.0%	97.14%	98.34%	95.0%

Model Results: Example Feature Splits



Model Results: ROC



Limitations and Future Directions

- More complex models saw marginal improvements
- Larger dataset may improve model training abilities
- Additional features would likely provide more decisional information
- Dataset consisted most of stage I tumors
    - Significantly lower recurrence rate and risk
    - Clinical value is limited