

Table with columns: UB_Accuracy, UB_FScore, UB_ROC_AUC, UB_Precision, UB_Recall, UB_F_Accuracy, UB_F_PScore, UB_F_ROC_AUC, UB_F_Precision, UB_F_Recall, UB_F_TNR, UB_F_FPR, UB_F_FNR, UB_Accuracy, UB_M_FScore, UB_M_ROC_AUC, UB_M_Precision, UB_M_Recall, UB_M_FNR, UB_M_TNR, UB_M_FPR. Rows 0-812. Includes Mean, Standard Deviation, T test, and P value at the bottom.

Experiment 3.1.1 (Unbalanced training data, no feature selection) - Gaussian Naive Bayes Classifier

Table with 28 columns representing different models and 99 rows of results. Each row shows a number (likely a test case or metric) and 28 numerical values corresponding to the models. The table is dense with data, including a 'Mean' row at the bottom.

Summary row with 28 columns, each containing a mean value for the corresponding model from the main table above.

Summary row with 28 columns, each containing a standard deviation value for the corresponding model from the main table above.

Summary row with 28 columns, each containing a test accuracy value for the corresponding model from the main table above.

Summary row with 28 columns, each containing a test accuracy value for the corresponding model from the main table above.

Summary of Results for Experiment 3.1.2

	Random Forest Classifier			Logistic Regression Classifier			Support Vector Machine			Gaussian Naive Bayes		
	Mean	Standard Deviation	T Test Significance (Female to Male)	Mean	Standard Deviation	T Test Significance (Female to Male)	Mean	Standard Deviation	T Test Significance (Female to Male)	Mean	Standard Deviation	T Test Significance (Female to Male)
All Accuracy	81.66	2.33		74.53	1.96		83.30	1.75		74.75	1.97	
All FScore	76.52	3.09		65.63	2.88		80.07	2.25		62.80	3.26	
All ROC_AUC	80.22%	2.41%		72.40%	1.97%		82.70%	1.85%		71.67%	1.97%	
All Precision	84.80%	3.38%		78.52%	4.30%		82.71%	3.84%		85.28%	3.80%	
All Recall	69.87%	4.42%		56.57%	3.85%		77.91%	4.71%		49.83%	3.77%	
Females Accuracy	84.75	3.03	0.00	77.71	2.42	0.00	89.03	2.23	0.00	78.47	2.35	0.00
Females FScore	71.41	5.74	0.00	51.69	5.87	0.00	82.18	3.72	0.00	51.55	5.38	0.00
Females ROC_AUC	78.47%	3.76%	0.13	67.10%	2.86%	0.00	86.87%	3.01%	0.00	67.12%	2.59%	0.00
Females Precision	85.44%	6.07%	0.88	82.22%	8.04%	0.00	84.22%	5.64%	0.14	85.13%	7.00%	0.37
Females Recall	61.70%	7.03%	0.00	38.12%	6.09%	0.00	80.88%	6.90%	0.00	37.19%	5.02%	0.00
Females FNR	38.30	7.03	0.00	61.88	6.09	0.00	19.12	6.90	0.00	62.81	5.02	0.00
Females TNR	95.24	2.03	0.00	96.09	2.04	0.00	92.86	3.00	0.00	97.05	1.48	0.00
Females FPR	4.76	2.03	0.00	3.91	2.04	0.00	7.14	3.00	0.00	2.95	1.48	0.00
Females TPR	61.70	7.03	0.00	38.12	6.09	0.00	80.88	6.90	0.00	37.19	5.02	0.00
Males Accuracy	78.58	3.05		71.35	3.22		77.56	2.93		71.04	3.22	
Males FScore	79.10	3.21		71.86	3.41		78.78	3.10		68.20	3.96	
Males ROC_AUC	79.07%	2.93%		71.89%	3.07%		77.80%	2.95%		72.57%	2.82%	
Males Precision	84.49%	3.92%		77.47%	4.44%		81.90%	4.62%		85.37%	4.05%	
Males Recall	74.59%	4.93%		67.34%	5.33%		76.24%	5.26%		57.01%	5.09%	
Males FNR	25.41	4.93		32.66	5.33		23.76	5.26		42.99	5.09	
Males TNR	83.55	4.16		76.44	4.82		79.36	5.69		88.12	3.48	
Males FPR	16.45	4.16		23.56	4.82		20.64	5.69		11.88	3.48	
Males TPR	74.59	4.93		67.34	5.33		76.24	5.26		57.01	5.09	
	Random Forest Results			Logistic Regression Results			Support Vector Machine Results			Gaussian Naive Bayes Results		
Averaged over 100 experiments	Sex Disparity (Male Mean - Female Mean)		T Test P Value	Sex Disparity (Male Mean - Female Mean)		T Test P Value	Sex Disparity (Male Mean - Female Mean)		T Test P Value	Sex Disparity (Male Mean - Female Mean)		T Test P Value
Accuracy	-6.17		0.00	-6.36		0.00	-11.47		0.00	-7.43		0.00
FScore	7.69		0.00	20.17		0.00	-3.40		0.00	16.65		0.00
ROC_AUC	0.60%		0.13	4.79%		0.00	-9.06%		0.00	5.45%		0.00
Precision	-0.94%		0.88	-4.75%		0.00	-2.32%		0.14	0.24%		0.37
Recall	12.88%		0.00	29.22%		0.00	-4.64%		0.00	19.82%		0.00
False Negative Rate	-12.88		0.00	-29.22		0.00	4.64		0.00	-19.82		0.00
True Negative Rate	-11.69		0.00	-19.65		0.00	-13.49		0.00	-8.93		0.00
False Positive Rate	11.69		0.00	19.65		0.00	13.49		0.00	8.93		0.00
True Positive Rate	12.88		0.00	29.22		0.00	-4.64		0.00	19.82		0.00

S.1.2 Balanced training data, no feature selection - Random Forest Classifier

Table with columns: ID, A Accuracy, B FScore, B ROC AUC, B Precision, B Recall, B F Accuracy, B F ROC AUC, B F Precision, B F Recall, B F F1 Score, B F F2 Score, B F F3 Score, B M Accuracy, B M ROC AUC, B M Precision, B M Recall, B M F1 Score, B M F2 Score, B M F3 Score. Rows 0-9999, plus Mean and Standard Deviation rows.

Table with 31 columns: B Accuracy, B FScore, B ROC AUC, B Precision, B Recall, B F Accuracy, B F Score, B F ROC AUC, B F Precision, B F Recall, B F FPR, B F PPV, B F NPV, B Accuracy, B FScore, B ROC AUC, B Precision, B Recall, B F Accuracy, B F Score, B F ROC AUC, B F Precision, B F Recall, B F FPR, B F PPV, B F NPV, B Accuracy, B FScore, B ROC AUC, B Precision, B Recall, B F Accuracy, B F Score, B F ROC AUC, B F Precision, B F Recall, B F FPR, B F PPV, B F NPV. Rows 0-999 represent individual data points, and the bottom section contains Mean, Standard Deviation, and Test Definition statistics.

Table with 31 columns: B Accuracy, B FScore, B ROC AUC, B Precision, B Recall, B F Accuracy, B F Score, B F ROC AUC, B F Precision, B F Recall, B F FPR, B F TNR, B F FPP, B F PPV, B F NPV, B M Accuracy, B M FScore, B M Precision, B M Recall, B M FPR, B M TNR, B M FPP, B M NPV. Rows 0-999 represent individual patients.

Summary statistics table with 31 columns: Mean, Standard Deviation, T Test (Pearson's M), B Accuracy, B FScore, B ROC AUC, B Precision, B Recall, B F Accuracy, B F Score, B F ROC AUC, B F Precision, B F Recall, B F FPR, B F TNR, B F FPP, B F PPV, B F NPV, B M Accuracy, B M FScore, B M Precision, B M Recall, B M FPR, B M TNR, B M FPP, B M NPV.

3.1.2 Balanced training data, no feature selection - Gaussian Naïve Bayes Classifier

Table with 25 columns: B.Accuracy, B.FScore, B.ROC.AUC, B.Precision, B.Recall, B.F.Accuracy, B.F.SFscore, B.F.ROC.AUC, B.F.Precision, B.F.Recall, B.F.NTR, B.F.TNR, B.F.FPR, B.M.Accuracy, B.M.FScore, B.M.ROC.AUC, B.M.Precision, B.M.Recall, B.M.FNR, B.M.TNR, B.M.FPR, B.M.FRR, B.M.TPR. Rows 1-99 contain individual data points, and rows 100-101 contain Mean and Std Deviation for each metric.

Summary of Results for Experiment 3.1.3 - All Classifiers

	Random Forest Classifier			Logistic Regression Classifier			Support Vector Machine			Gaussian Naive Bayes		
	Mean	Standard Deviation	T Test Significance (Female to Male)	Mean	Standard Deviation	T Test Significance (Female to Male)	Mean	Standard Deviation	T Test Significance (Female to Male)	Mean	Standard Deviation	T Test Significance (Female to Male)
All Accuracy	76.88	2.68		70.34	2.31		70.24	2.36		71.56	2.60	
All FScore	75.65	2.94		65.88	2.86		64.00	3.24		64.57	3.62	
All ROC_AUC	76.92%	2.65%		70.35%	2.21%		70.38%	2.25%		71.67%	2.24%	
All Precision	80.32%	4.24%		77.38%	4.23%		81.75%	4.61%		86.07%	3.68%	
All Recall	71.72%	4.10%		57.56%	3.84%		52.77%	3.98%		51.86%	4.50%	
Females Accuracy	74.45	4.79	0.00	72.40	4.26	0.01	72.22	4.03	0.01	73.93	4.45	0.00
Females FScore	63.50	7.71	0.00	53.13	7.06	0.00	50.66	7.56	0.00	52.19	8.13	0.00
Females ROC_AUC	71.49%	5.08%	0.00	66.65%	3.92%	0.00	65.92%	3.96%	0.00	67.24%	4.06%	0.00
Females Precision	72.51%	9.58%	0.00	77.21%	8.66%	0.44	82.53%	9.80%	0.90	88.92%	7.98%	0.03
Females Recall	57.62%	10.12%	0.00	41.05%	7.62%	0.00	36.98%	7.22%	0.00	37.43%	7.92%	0.00
Females FNR	42.38	10.12	0.00	58.95	7.62	0.00	63.02	7.22	0.00	62.57	7.92	0.00
Females TNR	85.37	6.39	0.00	92.26	3.47	0.00	94.86	3.28	0.00	97.06	2.18	0.00
Females FPR	14.63	6.39	0.00	7.74	3.47	0.00	5.14	3.28	0.00	2.94	2.18	0.00
Females TPR	57.62	10.12	0.00	41.05	7.62	0.00	36.98	7.22	0.00	37.43	7.92	0.00
Males Accuracy	77.87	3.12		69.51	3.19		69.47	2.85		70.62	3.04	
Males FScore	78.87	3.14		68.92	3.34		67.16	3.36		67.48	3.98	
Males ROC_AUC	78.10%	3.19%		70.25%	3.15%		70.82%	2.67%		72.23%	2.73%	
Males Precision	82.37%	4.43%		77.45%	4.66%		81.66%	4.77%		85.51%	4.30%	
Males Recall	75.83%	3.85%		62.29%	4.25%		57.27%	4.47%		55.97%	5.15%	
Males FNR	24.17	3.85		37.71	4.25		42.73	4.47		44.03	5.15	
Males TNR	80.38	5.24		78.22	5.08		84.36	4.50		88.49	3.54	
Males FPR	19.62	5.24		21.78	5.08		15.64	4.50		11.51	3.54	
Males TPR	75.83	3.85		62.29	4.25		57.27	4.47		55.97	5.15	
	Random Forest Results			Logistic Regression Results			Support Vector Machine Results			Gaussian Naive Bayes Results		
Sex Performance Disparities (Male Mean - Female Mean, n=100)	Sex Performance Disparities (Male Mean - Female Mean, n=100)		T Test P Value	Sex Performance Disparities (Male Mean - Female Mean, n=100)		T Test P Value	Sex Performance Disparities (Male Mean - Female Mean, n=100)		T Test P Value	Sex Performance Disparities (Male Mean - Female Mean, n=100)		T Test P Value
Accuracy	3.42		0.00	-2.90		0.01	-2.75		0.01	-3.31		0.00
FScore	15.36		0.00	15.79		0.00	16.50		0.00	15.29		0.00
ROC_AUC	6.61%		0.00	3.60%		0.00	4.90%		0.00	4.99%		0.00
Precision	9.85%		0.00	0.24%		0.44	-0.87%		0.90	-3.41%		0.03
Recall	18.21%		0.00	21.24%		0.00	20.30%		0.00	18.54%		0.00
False Negative Rate	-18.21		0.00	-21.24		0.00	-20.30		0.00	-18.54		0.00
True Negative Rate	-4.99		0.00	-14.04		0.00	-10.50		0.00	-8.57		0.00
False Positive Rate	4.99		0.00	14.04		0.00	10.50		0.00	8.57		0.00
True Positive Rate	18.21		0.00	21.24		0.00	20.30		0.00	18.54		0.00

Experiment 3.1.3 (Unbalanced training data, with feature selection) - Random Forest Classifier

Table with columns: LB Accuracy, LB F1 Score, LB Precision, LB Recall, UB Accuracy, UB F1 Score, UB Precision, UB Recall, UB MTHN, UB MTPR. The table contains 27 rows of performance metrics for various models, with 'Mean' and 'Std Deviance (F6 to M4)' rows at the bottom.

Table with 28 columns: UB Accuracy, UB F1 Score, UB ROC AUC, UB Precision, UB Recall, UB F1 Accuracy, UB F1 Precision, UB F1 Recall, UB F1 FNR, UB F1 FPR, UB F1 TPR, UB AUC Accuracy, UB AUC F1 Score, UB AUC Precision, UB AUC Recall, UB AUC FNR, UB AUC TNR, UB AUC FPR, UB AUC TPR. Rows 1-99 contain numerical data for various models. Rows 100-105 contain summary statistics for Mean, Standard Deviation, T Test, and P Value (Female to Male).

Table with columns: UB Accuracy, UB FScore, UB ROC_AUC, UB Precision, UB Recall, UB F_Accuracy, UB F_Precision, UB F_Recall, UB F_FNR, UB F_FPR, UB F_FNR, UB F_FPR, UB F_TPR, UB_M_Accuracy, UB_M_PScore, UB_M_ROC_AUC, UB_M_Precision, UB_M_Recall, UB_M_FNR, UB_M_FPR, UB_M_TPR. Rows include individual data points and summary statistics like Mean, Standard Deviation, and UB Significance P-value to Mean.

Table with 13 columns: LB, Accuracy, UB, F_Score, UB, ROC_AUC, UB, Precision, UB, Recall, UB, F_Accuracy, UB, F_Precision, UB, F_Recall, UB, F_FNR, UB, F_FPR, UB, Accuracy, UB, ROC_AUC, UB, Precision, UB, Recall, UB, F_PNR, UB, F_FPR. Rows contain numerical data for various metrics across different categories.

Mean, Standard Deviation, T Test Significance (Female to Male)

