Automated assay of four-protein biomarker panel for improved early detection of ovarian cancer (Supplementary Materials)

January 9, 2021

Table S1: Confusion matrices when applying KNN to identify the samples of patients in early stage cancer and healthy patients in data set 1 (E1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 1 (E1 vs H1) were used to train the model. Note that in this experiment the cohort of early stage cancer samples and healthy samples in data set 1 was divided into 60% for training and 40% for testing. This experiment is corresponding to experiment ID 1 in Table 3.



Using CA-125 alone



Table S2: Confusion matrices when applying logistic regression to identify the samples of patients in early stage and healthy patients in data set 1 (E1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 1 (E1 vs H1) were used to train the model. Note that in this experiment the cohort of early stage cancer samples and healthy samples in data set 1 was divided into 60% for training and 40% for testing. This experiment is corresponding to experiment ID 1 in Table 3.

	Actual label					Actual label			
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence
			v	0.18					0.18
	D'	C	0	PPV		D'		0	PPV
cteo	Disease	6	0	1	ctec	Disease	4	0	1
di	TT 14 h	1	20	NPV	di	TT 14 h	2	20	NPV
Pr_{f}	Healthy	1	32	0.97	\Pr	Healthy	3	32	0.91
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.86	1	0.97			0.57	1	0.92

Using CA-125 alone

Using four proteins

Table S3: Confusion matrices when applying random forest to identify the samples of patients in early stage and healthy patients in data set 1 (E1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 1 (E1 vs H1) were used to train the model. Note that in this experiment the cohort of early stage cancer samples and healthy samples in data set 1 was divided into 60% for training and 40% for testing. This experiment is corresponding to experiment ID 1 in Table 3.

	Actual label					Actual label				
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence	
ī				0.18					0.18	
ted	Disease	5	7	0.42	ted	Disease	7	0	PPV 1	
Predic	Healthy	2	25	NPV 0.93	Predic	Healthy	0	32	NPV 1	
-		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy	
		0.71	0.78	0.77			1	1	1	

Using CA-125 alone

Table S4: Confusion matrices when applying SVM to identify the samples of patients in early stage and healthy patients in data set 1 (E1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 1 (E1 vs H1) were used to train the model. Note that in this experiment the cohort of early stage cancer samples and healthy samples in data set 1 was divided into 60% for training and 40% for testing. This experiment is corresponding to experiment ID 1 in Table 3.



Using CA-125 alone



Table S5: Confusion matrices when applying KNN to identify the samples of patients in early stage and healthy patients in data set 2 (E2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 1 (E1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 2 in Table 3.

	Actual label					Actual label			
		Disease	Healthy	Prevalence 0.09			Disease	Healthy	Prevalence 0.09
cted	Disease	11	23	PPV 0.32	cted	Disease	13	2	PPV 0.87
Predi	Healthy	8	176	NPV 0.96	Predi	Healthy	6	196	NPV 0.97
		Sensitivity 0.58	Specificity 0.88	Accuracy 0.86			Sensitivity 0.68	Specificity 0.99	Accuracy 0.97

Using CA-125 alone

Using four proteins

Table S6: Confusion matrices when applying logistic regression to identify the samples of patients in early stage and healthy patients in data set 2 (E2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 1 (E1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 2 in Table 3.

	Actual label					Actual label				
		Dicease	Hoolthy	Prevalence			Dicesce	Hoolthy	Prevalence	
		Disease	meaniny	0.09			Disease	meaniny	0.09	
-	Dianaga	10	96	PPV	- P	Diagona	11	1	PPV	
cte	Disease	12	20	0.32	cte	Disease	11	1	0.92	
ipe	Hoolthy	7	172	NPV	ibe	Hoolthy	8	108	NPV	
Pre	meaniny	1	175	0.96	P_{re}	meaniny	8	190	0.96	
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy	
		0.63	0.87	0.85			0.58	0.99	0.96	

Using CA-125 alone

Table S7: Confusion matrices when applying random forest to identify the samples of patients in early stage and healthy patients in data set 2 (E2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 1 (E1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 2 in Table 3.

	Actual label					Actual label			
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence
				0.09					0.09
Ч	Diagona	11	25	PPV	5	Diagona	14	C	PPV
cte	Disease		30	0.24	cte	Disease	14	0	0.7
di	Haalthar	0	164	NPV	ipi	Haalthar	F	102	NPV
\Pr	пеаниу	0	104	0.95	$\mathbf{Pr}_{\mathbf{f}}$	пеаниу	5	195	0.97
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.58	0.82	0.80			0.74	0.97	0.95

Using CA-125 alone



Table S8: Confusion matrices when applying SVM to identify the samples of patients in early stage and healthy patients in data set 2 (E2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 1 (E1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 2 in Table 3.

		Actua	l label			Actual label				
		Disease	Healthy	Prevalence 0.09			Disease	Healthy	Prevalence 0.09	
cted	Disease	9	12	PPV 0.43	cted	Disease	14	9	PPV 0.61	
Predi	Healthy	10	187	NPV 0.95	Predi	Healthy	5	190	NPV 0.97	
		Sensitivity 0.47	Specificity 0.94	Accuracy 0.90			Sensitivity 0.74	Specificity 0.95	Accuracy 0.94	



Using four proteins

Table S9: Confusion matrices when applying KNN to identify the samples of ovarian cancer patients and healthy patients in data set 2 (C2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 1 (E1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 3 in Table 3.

Actual label					Actual label				
	Disease	Healthy	Prevalence 0.29			Disease	Healthy	Prevalence 0.29	
Disease	44	6	PPV 0.88	cted	Disease	40	0	PPV 1	
Healthy	36	193	NPV 0.84	Predi	Healthy	40	199	NPV 0.83	
	Sensitivity 0.55	Specificity 0.97	Accuracy 0.85			Sensitivity 0.5	Specificity 1	Accuracy 0.86	
]	Disease Healthy	Disease 44 Healthy 36 Sensitivity 0.55	DiseaseHealthyDisease446Healthy36193SensitivitySpecificity0.550.97	DiseaseHealthyPrevalence 0.29Disease446PPV 0.88Healthy36193NPV 0.84SensitivitySpecificity 0.55Accuracy 0.970.85	DiseaseHealthyPrevalence 0.29Disease446PPV 0.88Healthy36193NPV 0.84SensitivitySpecificity 0.55Accuracy 0.85	DiseaseHealthyPrevalence 0.29Disease446PPV 0.88Healthy36193NPV 0.84SensitivitySpecificity 0.55Accuracy 0.85Healthy	DiseaseHealthyPrevalence 0.29DiseaseDiseaseDisease446PPV 0.88Disease40Healthy36193NPV 0.84Disease40SensitivitySpecificity 0.55Accuracy 0.85Sensitivity 0.85Sensitivity 0.55Sensitivity 0.55	DiseaseHealthyPrevalence 0.29DiseaseHealthyDiseaseHealthyDisease446PPV 0.88Disease400Healthy36193NPV 0.84Healthy40199SensitivitySpecificityAccuracy 0.550.970.85SensitivitySpecificity 0.85Sensitivity	

Using CA-125 alone

Table S10: Confusion matrices when applying logistic regression to identify the samples of ovarian cancer patients and healthy patients in data set 2 (C2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 1 (E1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 3 in Table 3.

	Actual label					Actual label			
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence
			5	0.29				J	0.29
Ч	Diagona	45	C	PPV	ъ	Diagona	24	1	PPV
cte	Disease	40	0	0.88	cte	Disease	- 34	1	0.97
di	Hoolthy	25	102	NPV	di	Hoolthy	16	108	NPV
Pre	neariny	- 35	195	0.85	$\mathbf{Pr}_{\mathbf{f}}$	meaniny	40	190	0.81
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.58	0.97	0.85			0.42	1	0.83
								·	

Using CA-125 alone



Table S11: Confusion matrices when applying random forest to identify the samples of ovarian cancer patients and healthy patients in data set 2 (C2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 1 (E1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 3 in Table 3.

	Actual label					Actual label			
		Disease	Healthy	Prevalence 0.29			Disease	Healthy	Prevalence 0.29
cted	Disease	47	15	PPV 0.76	cted	Disease	50	3	PPV 0.94
Predi	Healthy	33	184	NPV 0.85	Predi	Healthy	30	196	NPV 0.87
		Sensitivity 0.59	Specificity 0.92	Accuracy 0.83			Sensitivity 0.63	Specificity 0.98	Accuracy 0.88

Using CA-125 alone

Using four proteins

Table S12: Confusion matrices when applying SVM to identify the samples of ovarian cancer patients and healthy patients in data set 2 (C2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 1 (E1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 3 in Table 3.

	Actual label					Actual label				
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence	
				0.29				, J	0.29	
γ	Dicesce	30	9	PPV	q	Diconco	52	5	PPV	
cte	Disease	- 59	2	0.95	cte	Disease	- 55	5	0.91	
jbe	Hoolthr	41	107	NPV	jpe	Hoolthr	27	104	NPV	
Pre	meaniny	41	197	0.83	\Pr	meaniny	21	194	0.88	
-		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy	
		0.49	0.99	0.85			0.66	0.97	0.89	

Using CA-125 alone

Table S13: Confusion matrices when applying KNN to identify the samples of patients in late stage and healthy patients in data set 1 (L1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 1 (L1 vs H1) were used to train the model. Note that in this experiment the cohort of late stage cancer samples and healthy samples in data set 1 was divided into 60% for training and 40% for testing. This experiment is corresponding to experiment ID 4 in Table 3.

		Actua	l label		Actual label				
		Disease	Healthy	Prevalence 0.41			Disease	Healthy	Prevalence 0.41
cted	Disease	17	0	PPV 1	cted	Disease	22	0	PPV 1
Predi	Healthy	5	32	NPV 0.86	Predie	Healthy	0	32	NPV 1
		Sensitivity	Specificity 1	Accuracy 0.91			Sensitivity	Specificity 1	Accuracy
		0.11	1	0.01			I	1	1

Using CA-125 alone



Table S14: Confusion matrices when applying logistic regression to identify the samples of patients in late stage and healthy patients in data set 1 (L1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 1 (L1 vs H1) were used to train the model. Note that in this experiment the cohort of late stage cancer samples and healthy samples in data set 1 was divided into 60% for training and 40% for testing. This experiment is corresponding to experiment ID 4 in Table 3.

		Actua	l label				Actua	l label	
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence
_		Discuse	meaning	0.41			Dibease	incarting	0.41
7	Digongo	17	1	PPV	لم ا	Dicesce	22	0	PPV
cte	Disease	17	1	0.92	cte	Disease	22	0	1
jbe	Hoolthy	F	91	NPV	di	Hoolthy	0	20	NPV
Pre	Heattiny	5	31	0.94	Pr_{f}	meaniny	0	32	1
-		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.92	0.94	0.89			1	1	1

Using CA-125 alone

Using four proteins

Table S15: Confusion matrices when applying random forest to identify the samples of patients in late stage and healthy patients in data set 1 (L1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 1 (L1 vs H1) were used to train the model. Note that in this experiment the cohort of late stage cancer samples and healthy samples in data set 1 was divided into 60% for training and 40% for testing. This experiment is corresponding to experiment ID 4 in Table 3.

		Actua	l label			Actual label			
		Disease	Healthy	Prevalence 0.41			Disease	Healthy	Prevalence 0.41
cted	Disease	20	2	PPV 0.91	cted	Disease	22	0	PPV 1
Predi	Healthy	2	30	NPV 0.94	Predi	Healthy	0	32	NPV 1
-		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.91	0.94	0.93			1	1	1

Using CA-125 alone

Table S16: Confusion matrices when applying SVM to identify the samples of patients in late stage and healthy patients in data set 1 (L1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 1 (L1 vs H1) were used to train the model. Note that in this experiment the cohort of late stage cancer samples and healthy samples in data set 1 was divided into 60% for training and 40% for testing. This experiment is corresponding to experiment ID 4 in Table 3.



Using CA-125 alone



Table S17: Confusion matrices when applying KNN to identify the samples of patients in late stage and healthy patients in data set 2 (L2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 1 (L1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 5 in Table 3.

		Actua	l label			Actual label			
	·	Disease	Healthy	Prevalence 0.23			Disease	Healthy	Prevalence 0.23
cted	Disease	40	24	PPV 0.63	cted	Disease	42	5	PPV 0.89
Predi	Healthy	21	175	NPV 0.89	Predi	Healthy	19	194	NPV 0.91
		Sensitivity 0.66	Specificity 0.88	Accuracy 0.83			Sensitivity 0.69	Specificity 0.97	Accuracy 0.91

Using CA-125 alone

Using four proteins

Table S18: Confusion matrices when applying logistic regression to identify the samples of patients in late stage and healthy patients in data set 2 (L2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 1 (L1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 5 in Table 3.

		Actua	l label		Actual label				
		Dicease	Hoolthy	Prevalence			Disease	Hoolthy	Prevalence
		Disease	meaniny	0.23			Disease	meaniny	0.23
-	Disease	40	24	PPV	q	Disease	45	F	PPV
cte	Disease	40	24	0.63	cte	Disease	40	5	0.9
ipe	Hoolthy	21	175	NPV	ibe	Hoolthy	16	104	NPV
Pre	Healthy	21	175	0.89	Pr_{r}	Heattiny	10	194	0.92
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.66	0.88	0.83			0.74	0.97	0.92

Using CA-125 alone

Table S19: Confusion matrices when applying random forest to identify the samples of patients in late stage and healthy patients in data set 2 (L2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 1 (L1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 5 in Table 3.

		Actual	l label				Actua	l label	
	·	Disease	Healthy	Prevalence 0.23			Disease	Healthy	Prevalence 0.23
cted	Disease	40	24	PPV 0.63	cted	Disease	46	10	PPV 0.82
Predic	Healthy	21	175	NPV 0.89	Predic	Healthy	15	189	NPV 0.93
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.66	0.88	0.83			0.75	0.95	0.9

Using CA-125 alone



Table S20: Confusion matrices when applying SVM to identify the samples of patients in late stage and healthy patients in data set 2 (L2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 1 (L1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 5 in Table 3.

		Actua	l label			Actual label				
		Disease	Healthy	Prevalence 0.23			Disease	Healthy	Prevalence 0.23	
cted	Disease	38	6	PPV 0.86	cted	Disease	45	7	PPV 0.87	
Predi	Healthy	23	193	NPV 0.89	Predi	Healthy	16	192	NPV 0.92	
		Sensitivity 0.62	Specificity 0.97	Accuracy 0.89			Sensitivity 0.74	Specificity 0.96	Accuracy 0.91	



Using four proteins

Table S21: Confusion matrices when applying KNN to identify the samples of ovarian cancer patients and healthy patients in data set 2 (C2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 1 (L1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 6 in Table 3.

	Actual label					Actual label				
		Digongo	Hoolthy	Prevalence			Digongo	Hoolthy	Prevalence	
		Disease	Heattiny	0.29			Disease	meaniny	0.29	
7	Disease	52	22	PPV	q	Disease	56	4	PPV	
cte	Disease	- 55	- 55	0.62	cte	Disease	50		0.93	
ibe	Hoolthy	27	166	NPV	edi	Hoolthy	24	105	NPV	
Pro	meaniny	21	100	0.86	\Pr	meaniny	24	195	0.89	
-		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy	
		0.66	0.83	0.78			0.7	0.98	0.9	



Table S22: Confusion matrices when applying logistic regression to identify the samples of ovarian cancer patients and healthy patients in data set 2 (C2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 1 (L1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 6 in Table 3.

		Actua	l label				Actua	l label	
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence
_		Discuse	mountify	0.29			Dibease	incartiny	0.29
7	Disease	50	22	PPV	5	Digongo	61	7	PPV
cte	Disease	52	22	0.7	cte	Disease	01	1	0.9
jdi	Haalthar	00	177	NPV	did	Haalthar	10	109	NPV
Pre	пеанну	20	177	0.86	$\mathbf{Pr}_{\mathbf{f}}$	пеанну	19	192	0.91
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.65	0.89	0.82			0.76	0.96	0.91

Using CA-125 alone



Table S23: Confusion matrices when applying random forest to identify the samples of ovarian cancer patients and healthy patients in data set 2 (C2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 1 (L1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 6 in Table 3.

	Actual label					Actual label				
		Disease	Healthy	Prevalence 0.29			Disease	Healthy	Prevalence 0.29	
cted	Disease	52	32	PPV 0.62	cted	Disease	62	11	PPV 0.85	
Predi	Healthy	28	167	NPV 0.86	Predi	Healthy	18	188	NPV 0.91	
		Sensitivity 0.65	Specificity 0.84	Accuracy 0.78			Sensitivity 0.78	Specificity 0.94	Accuracy 0.9	

Using four proteins

Table S24: Confusion matrices when applying SVM to identify the samples of ovarian cancer patients and healthy patients in data set 2 (C2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 1 (L1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 6 in Table 3.

	Actual label					Actual label				
		Disease	Hoolthy	Prevalence			Disease	Hoolthy	Prevalence	
		Disease	Heattiny	0.29			Disease	meaniny	0.29	
γ	Disoaso	47	8	PPV	q	Disease	61	8	PPV	
cte	Disease	- 11	0	0.85	cte	Disease	01	0	0.88	
edi	Hoalthy	22	101	NPV	edi	Hoalthy	10	101	NPV	
Pr	meaniny		131	0.85	$\mathbf{P}_{\mathbf{r}}$	Heartiny	13	131	0.91	
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy	
		0.59	0.96	0.85			0.76	0.96	0.9	



Table S25: Confusion matrices when applying KNN to identify the samples of patients in early stage and healthy patients in data set 2 (E2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 1 (C1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 7 in Table 3.

		Actua	l label				Actua	l label	
		Disease	Healthy	Prevalence 0.09			Disease	Healthy	Prevalence 0.09
ted	Disease	15	135	PPV 0.1	ted	Disease	16	46	PPV 0.26
Predic	Healthy	4	64	NPV 0.94	Predic	Healthy	3	153	NPV 0.98
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.79	0.32	0.36			0.84	0.77	0.76



Table S26: Confusion matrices when applying logistic regression to identify the samples of patients in early stage and healthy patients in data set 2 (E2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 1 (C1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 7 in Table 3.

	Actual label					Actual label				
		Disease	Healthy	Prevalence 0.09			Disease	Healthy	Prevalence 0.09	
cted	Disease	14	105	PPV 0.12	cted	Disease	17	53	PPV 0.24	
Predi	Healthy	5	94	NPV 0.95	Predi	Healthy	2	146	NPV 0.99	
		Sensitivity 0.74	Specificity 0.47	Accuracy 0.5			Sensitivity 0.89	Specificity 0.73	Accuracy 0.75	

Using CA-125 alone

Using four proteins

Table S27: Confusion matrices when applying random forest to identify the samples of patients in early stage and healthy patients in data set 2 (E2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 1 (C1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 7 in Table 3.

	Actual label					Actual label				
		Digongo	Hoolthy	Prevalence			Digongo	Hoolthy	Prevalence	
		Disease	meaniny	0.09			Disease	meaniny	0.09	
γ	Disease	19	80	PPV	q	Disease	17	119	PPV	
cte	Disease	12	89	0.12	cte	Disease	11	115	0.13	
jbe	Hoolthy	7	110	NPV	jpe	Hoolthr	2	96	NPV	
Pre	meaniny	1	110	0.94	\Pr	meaniny	2	80	0.98	
-		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy	
		0.63	0.55	0.56			0.89	0.43	0.47	



Table S28: Confusion matrices when applying SVM to identify the samples of patients in early stage and healthy patients in data set 2 (E2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 1 (C1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 7 in Table 3.

		Actua	l label				Actua	l label	
		Disease	Healthy	Prevalence 0.09			Disease	Healthy	Prevalence 0.09
	D:	19	75	PPV	-	D:	17	100	PPV
cte	Disease	13	(5)	0.15	cte	Disease	17	108	0.14
edi	Healthy	6	194	NPV	edi	Healthy	2	91	NPV
Pr	incareiry	Ū	121	0.95	$\mathbf{P}_{\mathbf{r}}$	incarony	2	51	0.98
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.68	0.62	0.63			0.89	0.46	0.5

Using CA-125 alone



Table S29: Confusion matrices when applying KNN to identify the samples of patients in late stage and healthy patients in data set 2 (L2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 1 (C1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 8 in Table 3.

		Actua	l label			Actual label				
		Disease	Healthy	Prevalence 0.23			Disease	Healthy	Prevalence 0.23	
cted	Disease	44	27	PPV 0.61	cted	Disease	42	6	PPV 0.88	
Predi	Healthy	17	172	NPV 0.91	Predi	Healthy	19	193	NPV 0.91	
		Sensitivity 0.72	Specificity 0.86	Accuracy 0.83			Sensitivity 0.69	Specificity 0.97	Accuracy 0.9	

Using CA-125 alone

Using four proteins

Table S30: Confusion matrices when applying logistic regression to identify the samples of patients in late stage and healthy patients in data set 2 (L2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 1 (C1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 8 in Table 3.

		Actua	l label			Actual label				
		Dicease	Hoolthy	Prevalence			Disease	Hoolthy	Prevalence	
		Disease	meaniny	0.23			Disease	meaniny	0.23	
7	Disoaso	41	18	PPV	q	Disease	45	19	PPV	
cte	Disease	41	10	0.69	cte	Disease	40	12	0.79	
edi	Hoalthy	20	181	NPV	edi	Hoalthy	16	187	NPV	
Pr	meaniny	20	101	0.9	$\mathbf{P}_{\mathbf{r}}$	meaniny	10	107	0.92	
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy	
		0.67	0.91	0.85			0.74	0.94	0.89	



Table S31: Confusion matrices when applying random forest to identify the samples of patients in late stage and healthy patients in data set 2 (L2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 1 (C1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 8 in Table 3.

		Actua	l label				Actua	l label	
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence
				0.23					0.23
-	Diagona	41	20	PPV	ק	Digongo	46	11	PPV
cte	Disease	41	29	0.59	cte	Disease	40	11	0.81
ģi	Hoolthy	20	170	NPV	ibe	Hoolthy	15	199	NPV
Pre	meaniny	20	170	0.89	$\mathbf{P}_{\mathbf{r}\epsilon}$	meaniny	15	100	0.93
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.67	0.85	0.81			0.75	0.94	0.9
							•	·	

Using CA-125 alone



Table S32: Confusion matrices when applying SVM to identify the samples of patients in late stage and healthy patients in data set 2 (L2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 1 (C1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 8 in Table 3.

		Actua	l label			Actual label			
		Disease	Healthy	Prevalence 0.23			Disease	Healthy	Prevalence 0.23
cted	Disease	39	12	PPV 0.76	cted	Disease	45	11	PPV 0.8
Predi	Healthy	22	187	NPV 0.89	Predi	Healthy	16	188	NPV 0.92
		Sensitivity 0.64	Specificity 0.94	Accuracy 0.87			Sensitivity 0.74	Specificity 0.94	Accuracy 0.89

Using CA-125 alone

Using four proteins

Table S33: Confusion matrices when applying KNN to identify the samples of ovarian cancer patients and healthy patients in data set 2 (C2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 1 (C1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 9 in Table 3.

		Actua	l label			Actual label				
		Digongo	Hoolthy	Prevalence			Digongo	Hoolthy	Prevalence	
		Disease	meaniny	0.29			Disease	meaniny	0.29	
7	Disoaso	50	35	PPV	q	Disease	58	19	PPV	
cte	Disease		55	0.63	cte	Disease		15	0.82	
ibe	Hoolthy	21	164	NPV	edi	Hoolthy	22	186	NPV	
Pro	meaniny	21	104	0.89	$\mathbf{Pr}_{\mathbf{r}}$	meaniny	22	180	0.89	
-		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy	
		0.74	0.82	0.8			0.73	0.93	0.87	



Table S34: Confusion matrices when applying logistic regression to identify the samples of ovarian cancer patients and healthy patients in data set 2 (C2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 1 (C1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 9 in Table 3.

		Actua	l label				Actua	l label	
		Disease	Healthy	Prevalence 0.29			Disease	Healthy	Prevalence 0.29
cted	Disease	56	27	PPV 0.67	cted	Disease	61	12	PPV 0.84
Predie	Healthy	24	172	NPV 0.88	Predic	Healthy	19	187	NPV 0.91
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.7	0.86	0.82			0.76	0.94	0.89

Using CA-125 alone



Table S35: Confusion matrices when applying random forest to identify the samples of ovarian cancer patients and healthy patients in data set 2 (C2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 1 (C1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 9 in Table 3.

		Actua	l label			Actual label				
		Disease	Healthy	Prevalence 0.29			Disease	Healthy	Prevalence 0.29	
cted	Disease	55	41	PPV 0.57	cted	Disease	62	13	PPV 0.83	
Predi	Healthy	25	158	NPV 0.86	Predi	Healthy	18	186	NPV 0.91	
		Sensitivity 0.69	Specificity 0.79	Accuracy 0.76			Sensitivity 0.78	Specificity 0.93	Accuracy 0.89	

Using CA-125 alone

Using four proteins

Table S36: Confusion matrices when applying SVM to identify the samples of ovarian cancer patients and healthy patients in data set 2 (C2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 1 (C1 vs H1) were used to train the model. This experiment is corresponding to experiment ID 9 in Table 3.

		Actua	l label			Actual label			
		Dicease	Hoalthy	Prevalence			Disease	Hoalthy	Prevalence
		Disease	Heattiny	0.29			Disease	meaniny	0.29
g	Disease	51	14	PPV	g	Disease	62	11	PPV
cte	Disease	51	14	0.78	cte	Disease	02	11	0.85
edi	Hoalthy	20	185	NPV	edi	Hoalthy	18	188	NPV
Pr	meaniny	23	105	0.86	\mathbf{Pr}	meanity	10	100	0.91
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.64	0.93	0.85			0.76	0.94	0.9



Table S37: Confusion matrices when applying KNN to identify the samples of patients in early stage and healthy patients in data set 2 (E2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 2 (E2 vs E2) were used to train the model. Note that in this experiment the cohort of early stage cancer samples and healthy samples in data set 2 was divided into 60% for training and 40% for testing. This experiment is corresponding to experiment ID 9 in Table 3. This experiment is corresponding to experiment ID 10 in Table 3.

		Actua	l label				Actua	l label	
		Disease	Healthy	Prevalence			Disease	Hoalthy	Prevalence
_		Discuse	incartiny	0.08			Discuse	incarcity	0.08
7	Disease	1	ŋ	PPV	q	Diconco	0	1	PPV
cte	Disease	1	2	0.33	cte	Disease	2	1	0.67
ibe	Hoolthy	6	77	NPV	ibe	Hoolthy	5	78	NPV
Pre	meaniny	0	11	0.93	P_{re}	Heattiny	5	18	0.94
-		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.14	0.97	0.91			0.29	0.99	0.93

Using CA-125 alone



Table S38: Confusion matrices when applying logistic regression to identify the samples of patients in early stage and healthy patients in data set 2 (E2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 2 (E2 vs E2) were used to train the model. Note that in this experiment the cohort of early stage cancer samples and healthy samples in data set 2 was divided into 60% for training and 40% for testing. This experiment is corresponding to experiment ID 10 in Table 3.

		Actua	l label				Actua	l label	
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence
		21000000	incarony	0.08			Dibotabo	licarting	0.08
Ч	Diagona	1	0	PPV	-	Diagona	4	1	PPV
cte	Disease	1	2	0.33	cte	Disease	4	1	0.8
ibe	Hoolthy	6	77	NPV	di	Hoolthy	9	70	NPV
\Pr	Heattiny	0	11	0.93	\Pr	meaniny	0	10	0.96
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.14	0.97	0.91			0.57	0.99	0.95

Using CA-125 alone

Using four proteins

Table S39: Confusion matrices when applying random forest to identify the samples of patients in early stage and healthy patients in data set 2 (E2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 2 (E2 vs E2) were used to train the model. Note that in this experiment the cohort of early stage cancer samples and healthy samples in data set 2 was divided into 60% for training and 40% for testing. This experiment is corresponding to experiment ID 10 in Table 3.

		Actua	l label			Actual label				
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence	
_ [PPV					PPV	
cted	Disease	1	5	0.17	cted	Disease	3	2	0.6	
edic	Healthy	6	74	NPV	edic	Healthy	4	77	NPV	
$\mathbf{P}_{\mathbf{r}}$	liounij			0.93	Pr	licarony	-		0.95	
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy	
		0.14	0.94	0.87			0.43	0.97	0.93	

Using CA-125 alone

Table S40: Confusion matrices when applying SVM to identify the samples of patients in early stage and healthy patients in data set 2 (E2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 2 (E2 vs E2) were used to train the model. Note that in this experiment the cohort of early stage cancer samples and healthy samples in data set 2 was divided into 60% for training and 40% for testing. This experiment is corresponding to experiment ID 10 in Table 3.

		Actua	l label				Actua	l label	
		Disease	Hoalthy	Prevalence			Disease	Hoalthy	Prevalence
_		Disease	Heartiny	0.08			Disease	ITEATUIY	0.08
g	Disease	1	9	PPV	g	Disease	2	1	PPV
cte	Discuse	1	2	0.33	cte	Discuse	2	I	0.67
edi	Hoalthy	6	77	NPV	edi	Hoalthy	5	78	NPV
Pr	meaniny	0	11	0.93	\mathbf{Pr}	meaniny	0	10	0.94
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.14	0.97	0.91			0.29	0.99	0.93

Using CA-125 alone

Using four proteins

Table S41: Confusion matrices when applying KNN to identify the samples of patients in early stage and healthy patients in data set 1 (E1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 2 (E2 vs E2) were used to train the model. This experiment is corresponding to experiment ID 11 in Table 3.

		Actua	l label			Actual label				
		Disease	Healthy	Prevalence 0.18			Disease	Healthy	Prevalence 0.18	
cted	Disease	5	1	PPV 0.83	cted	Disease	12	0	PPV 1	
Predi	Healthy	13	79	NPV 0.86	Predi	Healthy	6	80	NPV 0.93	
		Sensitivity 0.28	Specificity 0.99	Accuracy 0.86			Sensitivity 0.67	Specificity 1	Accuracy 0.94	

Using CA-125 alone

Using four proteins

Table S42: Confusion matrices when applying logistic regression to identify the samples of patients in early stage and healthy patients in data set 1 (E1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 2 (E2 vs E2) were used to train the model. This experiment is corresponding to experiment ID 11 in Table 3.

		Actual	l label		Actual label				
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence
		Discuse filoartily	0.18			Discuse	incartiny	0.18	
7	Disease	4	0	PPV	g	Disease	15	0	PPV
cte	Discuse	т	0	1	cte	Discuse	10	0	1
edi	Healthy	14	80	NPV	edi	Healthy	3	80	NPV
Pr	incantity	14	00	0.85	Pr	incartiny	5	00	0.96
-		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.22	1	0.86			0.83	1	0.97

Using CA-125 alone

Table S43: Confusion matrices when applying random forest to identify the samples of patients in early stage and healthy patients in data set 1 (E1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 2 (E2 vs E2) were used to train the model. This experiment is corresponding to experiment ID 11 in Table 3.

	Actual label					Actual label			
	·	Disease	Healthy	Prevalence			Disease	Healthy	Prevalence
				0.18					0.18
ا ت	Disease	6	4	PPV	ק	Dicesce	16	0	PPV
cte	Disease	0	4	0.6	cte	Disease	10	0	1
jdi	Haalthar	10	76	NPV	did	Haalthar	0	20	NPV
Pre	пеаниу	12	70	0.86	P_{re}	пеанну	2	00	0.98
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.33	0.95	0.84			0.89	1	0.98
								·	



Table S44: Confusion matrices when applying SVM to identify the samples of patients in early stage and healthy patients in data set 1 (E1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 2 (E2 vs E2) were used to train the model. This experiment is corresponding to experiment ID 11 in Table 3.

	Actual label					Actual label			
		Disease	Healthy	Prevalence 0.18			Disease	Healthy	Prevalence 0.18
cted	Disease	4	0	PPV 1	cted	Disease	15	0	PPV 1
Predi	Healthy	14	80	NPV 0.85	Predi	Healthy	3	80	NPV 0.96
		Sensitivity 0.22	Specificity 1	Accuracy 0.86			Sensitivity 0.83	Specificity 1	Accuracy 0.97

Using CA-125 alone

Using four proteins

Table S45: Confusion matrices when applying KNN to identify the samples of ovarian cancer patients and healthy patients in data set 1 (C1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 2 (E2 vs E2) were used to train the model. This experiment is corresponding to experiment ID 12 in Table 3.

	Actual label					Actual label				
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence	
		Discuse	incartiny	0.48			Discuse	incartiny	0.48	
7	Disease	າຊ	1	PPV	q	Disease	24	0	PPV	
cte	Disease	28	1	0.97	cte	Disease	24	0	1	
jģi	Haalthar	45	70	NPV	di	Haaltha	40	80	NPV	
Pre	пеанну	40	19	0.64	\Pr	пеаниу	49	80	0.62	
-		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy	
		0.38	0.99	0.7			0.33	1	0.68	



Table S46: Confusion matrices when applying logistic regression to identify the samples of ovarian cancer patients and healthy patients in data set 1 (C1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 2 (E2 vs E2) were used to train the model. This experiment is corresponding to experiment ID 12 in Table 3.

		Actua	l label				Actua	l label	
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence
		21000000	incarony	0.48			21000000	inconting	0.48
7	Digongo	26	0	PPV	ъ	Dicesce	20	0	PPV
ste	Disease	20	0	1	cte	Disease	30	0	1
jģi	Haalthar	47	80	NPV	die	Haalthar	25	80	NPV
Pré	пеанну	47	80	0.63	\Pr	пеаниу		80	0.7
1		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.36	1	0.69			0.52	1	0.77
								a	
	Using CA-125 alone			Using four proteins					

Table S47: Confusion matrices when applying random forest to identify the samples of ovarian cancer patients and healthy patients in data set 1 (C1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 2 (E2 vs E2) were used to train the model. This experiment is corresponding to experiment ID 12 in Table 3.

		Actua	l label			Actual label			
		Disease	Healthy	Prevalence 0.48			Disease	Healthy	Prevalence 0.48
cted	Disease	31	3	PPV 0.91	cted	Disease	42	0	PPV 1
Predi	Healthy	42	77	NPV 0.65	Predi	Healthy	31	80	NPV 0.72
		Sensitivity 0.42	Specificity 0.96	Accuracy 0.71			Sensitivity 0.58	Specificity 1	Accuracy 0.8

Using four proteins

Table S48: Confusion matrices when applying SVM to identify the samples of ovarian cancer patients and healthy patients in data set 1 (C1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in early stage cancer and healthy patients in data set 2 (E2 vs E2) were used to train the model. This experiment is corresponding to experiment ID 12 in Table 3.

	Actual label					Actual label				
		Disease	Healthy	Prevalence 0.48			Disease	Healthy	Prevalence 0.48	
ted	Disease	27	0	0.48 PPV 1	ted	Disease	50	0	0.48 PPV 1	
Predic	Healthy	46	80	NPV 0.63	Predic	Healthy	23	80	NPV 0.78	
		Sensitivity 0.37	Specificity 1	Accuracy 0.7			Sensitivity 0.68	Specificity 1	Accuracy 0.85	



Table S49: Confusion matrices when applying KNN to identify the samples of patients in late stage and healthy patients in data set 2 (L2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 2 (L2 vs L2) were used to train the model. Note that in this experiment the cohort of late stage cancer samples and healthy samples in data set 2 was divided into 60% for training and 40% for testing. This experiment is corresponding to experiment ID 13 in Table 3.

		Actual	l label			Actual label				
		Disease	Hoalthy	Prevalence			Disonso	Hoalthy	Prevalence	
		Disease	Healthy	0.23			Disease	meaniny	0.23	
7	Disease	19	ŋ	PPV	q l	Disease	16	1	PPV	
cte	Disease	15	2	0.87	cte	Disease	10	1	0.94	
ipe	Hoolthy	11	77	NPV	ipe	Hoolthy	Q	78	NPV	
Pre	meaniny	11	11	0.88	\mathbf{P}_{rc}	meaniny	0	18	0.91	
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy	
		0.54	0.97	0.87			0.67	0.98	0.91	

Using CA-125 alone

Using four proteins

Table S50: Confusion matrices when applying logistic regression to identify the samples of patients in late stage and healthy patients in data set 2 (L2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 2 (L2 vs L2) were used to train the model. Note that in this experiment the cohort of late stage cancer samples and healthy samples in data set 2 was divided into 60% for training and 40% for testing. This experiment is corresponding to experiment ID 13 in Table 3.

	Actual label				Actual label				
	·	Disease	Healthy	Prevalence			Disease	Healthy	Prevalence
		Dibease	incarony	0.23			Disease	ileaniy	0.23
q	Disease	11	0	PPV	q	Disease	16	n	PPV
cte	Disease	11	2	0.85	cte	Disease	10	2	0.89
ibe	Hoolthy	19	77	NPV	ipe	Hoolthy	8	77	NPV
P_{re}	Healthy	15		0.86	\Pr	meaniny	8	11	0.91
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.46	0.97	0.85			0.67	0.97	0.9

Using CA-125 alone

Using four proteins

Table S51: Confusion matrices when applying random forest to identify the samples of patients in late stage and healthy patients in data set 2 (L2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 2 (L2 vs L2) were used to train the model. Note that in this experiment the cohort of late stage cancer samples and healthy samples in data set 2 was divided into 60% for training and 40% for testing. This experiment is corresponding to experiment ID 13 in Table 3.

		Actua	l label			Actual label				
		Disease	Healthy	Prevalence 0.23			Disease	Healthy	Prevalence 0.23	
cted	Disease	14	10	PPV 0.58	cted	Disease	17	3	PPV 0.85	
Predi	Healthy	10	69	NPV 0.87	Predi	Healthy	7	76	NPV 0.92	
		Sensitivity 0.58	Specificity 0.87	Accuracy 0.81			Sensitivity 0.71	Specificity 0.96	Accuracy 0.9	

Using CA-125 alone

Table S52: Confusion matrices when applying SVM to identify the samples of patients in late stage and healthy patients in data set 2 (L2 vs H2), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 2 (L2 vs L2) were used to train the model. Note that in this experiment the cohort of late stage cancer samples and healthy samples in data set 2 was divided into 60% for training and 40% for testing. This experiment is corresponding to experiment ID 13 in Table 3.

		Actua	l label				Actua	l label	
		Disease	Hoalthy	Prevalence			Disease	Hoalthy	Prevalence
_		Disease	meaniny	0.23			Disease	ITEATURY	0.23
g	Disease	11	2	PPV	g	Disease	16	1	PPV
cte	Disease	11	2	0.85	cte	Disease	10	I	0.94
edi	Hoolthy	19	77	NPV	edi	Hoolthy	8	78	NPV
Pr	Healthy	15	11	0.86	$\mathbf{P}_{\mathbf{r}}$	meaniny	8	10	0.91
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.46	0.97	0.85			0.67	0.99	0.91

Using CA-125 alone



Table S53: Confusion matrices when applying KNN to identify the samples of patients in late stage and healthy patients in data set 1 (L1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 2 (L2 vs L2) were used to train the model. This experiment is corresponding to experiment ID 14 in Table 3.

	Actual label					Actual label			
		Disease	Healthy	Prevalence 0.41			Disease	Healthy	Prevalence 0.41
cted	Disease	43	0	PPV 1	cted	Disease	49	1	PPV 0.98
Predi	Healthy	12	80	NPV 0.87	Predi	Healthy	6	79	NPV 0.93
		Sensitivity 0.78	Specificity 1	Accuracy 0.91			Sensitivity 0.89	Specificity 0.99	Accuracy 0.95

Using CA-125 alone

Using four proteins

Table S54: Confusion matrices when applying logistic regression to identify the samples of patients in late stage and healthy patients in data set 1 (L1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 2 (L2 vs L2) were used to train the model. This experiment is corresponding to experiment ID 14 in Table 3.

	Actual label					Actual label				
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence	
I				0.41					0.41	
cted	Disease	42	0	1	cted	Disease	48	0	1	
Predi	Healthy	13	80	NPV 0.86	Predi	Healthy	7	80	NPV 0.92	
-		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy	
		0.76	1	0.9			0.87	1	0.95	

Using CA-125 alone

Table S55: Confusion matrices when applying random forest to identify the samples of patients in late stage and healthy patients in data set 1 (L1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 2 (L2 vs L2) were used to train the model. This experiment is corresponding to experiment ID 14 in Table 3.

		Actua	l label				Actua	l label	
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence
		Discuse	incartiny	0.41			Disease	incarony	0.41
q	Disease	49	11	PPV	لم ا	Disease	50	0	PPV
cte	Disease	42	11	0.79	cte	Disease	52	9	0.85
di	Hoolthy	19	60	NPV	di	Hoolthr	9	71	NPV
\Pr	neariny	10	09	0.84	P_{re}	Heattiny	0	11	0.96
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.76	0.86	0.82			0.95	0.89	0.91

Using CA-125 alone



Table S56: Confusion matrices when applying SVM to identify the samples of patients in late stage and healthy patients in data set 1 (L1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 2 (L2 vs L2) were used to train the model. This experiment is corresponding to experiment ID 14 in Table 3.

		Actua	l label				Actua	l label	
		Disease	Healthy	Prevalence 0.41			Disease	Healthy	Prevalence 0.41
cted	Disease	42	0	PPV 1	cted	Disease	48	0	PPV 1
Predi	Healthy	13	80	NPV 0.86	Predi	Healthy	7	80	NPV 0.92
		Sensitivity 0.76	Specificity 1	Accuracy 0.9			Sensitivity 0.87	Specificity 1	Accuracy 0.95

Using CA-125 alone

Using four proteins

Table S57: Confusion matrices when applying KNN to identify the samples of ovarian cancer patients and healthy patients in data set 1 (C1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 2 (L2 vs L2) were used to train the model. This experiment is corresponding to experiment ID 15 in Table 3.

	Actual label					Actual label				
		Diconco	Hoolthy	Prevalence			Disease	Hoolthy	Prevalence	
		Disease	meaniny	0.48			Disease	Healthy	0.48	
g	Disease	53	0	PPV	g	Disease	65	4	PPV	
cte	Discuse		0	1	cte	Discuse	00	т	0.94	
edi	Healthy	20	80	NPV	edi	Healthy	8	76	NPV	
Pr	incareiry	20	00	0.8	$\mathbf{P}_{\mathbf{r}}$	incarony	0	10	0.9	
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy	
		0.73	1	0.87			0.89	0.95	0.92	



Table S58: Confusion matrices when applying logistic regression to identify the samples of ovarian cancer patients and healthy patients in data set 1 (C1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 2 (L2 vs L2) were used to train the model. This experiment is corresponding to experiment ID 15 in Table 3.

		Actua	l label				Actua	l label	
	·	Disease	Healthy	Prevalence 0.48			Disease	Healthy	Prevalence 0.48
cted	Disease	51	0	PPV 1	cted	Disease	59	0	PPV 1
Predic	Healthy	22	80	NPV 0.78	Predic	Healthy	14	80	NPV 0.85
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.7	1	0.86			0.81	1	0.91

Using CA-125 alone



Table S59: Confusion matrices when applying random forest to identify the samples of ovarian cancer patients and healthy patients in data set 1 (C1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 2 (L2 vs L2) were used to train the model. This experiment is corresponding to experiment ID 15 in Table 3.

	Actual label					Actual label			
		Disease	Healthy	Prevalence 0.48			Disease	Healthy	Prevalence 0.48
ted	Disease	52	11	PPV 0.83	ted	Disease	68	10	PPV 0.87
Predic	Healthy	21	69	NPV 0.77	Predic	Healthy	5	70	NPV 0.93
		Sensitivity 0.71	Specificity 0.86	Accuracy 0.79		·	Sensitivity 0.93	Specificity 0.88	Accuracy 0.9

Using CA-125 alone

Using four proteins

Table S60: Confusion matrices when applying SVM to identify the samples of ovarian cancer patients and healthy patients in data set 1 (C1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from patients in late stage cancer and healthy patients in data set 2 (L2 vs L2) were used to train the model. This experiment is corresponding to experiment ID 15 in Table 3.

	Actual label					Actual label			
		Disease	Healthy	Prevalence 0.48			Disease	Healthy	Prevalence 0.48
cted	Disease	50	0	PPV 1	cted	Disease	63	0	PPV 1
Predi	Healthy	23	80	$\begin{array}{c} \mathrm{NPV} \\ 0.78 \end{array}$	Predi	Healthy	10	80	NPV 0.89
-		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.68	1	0.85			0.86	1	0.93



Table S61: Confusion matrices when applying KNN to identify the samples of patients in early stage and healthy patients in data set 1 (E1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 2 (C2 vs H2) were used to train the model. This experiment is corresponding to experiment ID 16 in Table 3.

		Actual	l label				Actua	l label	
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence
	D:	10	2	PPV		D:	17	10	PPV
cteo	Disease	12	2	0.86	cted	Disease	17	12	0.59
Predi	Healthy	6	78	NPV 0.93	Predi	Healthy	1	68	NPV 0.99
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.67	0.85	0.93			0.94	0.85	0.87



Table S62: Confusion matrices when applying logistic regression to identify the samples of patients in early stage and healthy patients in data set 1 (E1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 2 (C2 vs H2) were used to train the model. This experiment is corresponding to experiment ID 16 in Table 3.

	Actual label					Actual label			
		Disease	Healthy	Prevalence 0.18			Disease	Healthy	Prevalence 0.18
cted	Disease	11	1	PPV 0.92	cted	Disease	17	7	PPV 0.71
Predi	Healthy	7	79	NPV 0.92	Predi	Healthy	1	73	NPV 0.99
		Sensitivity 0.61	Specificity 0.99	Accuracy 0.92			Sensitivity 0.94	Specificity 0.91	Accuracy 0.92

Using CA-125 alone

Using four proteins

Table S63: Confusion matrices when applying random forest to identify the samples of patients in early stage and healthy patients in data set 1 (E1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 2 (C2 vs H2) were used to train the model. This experiment is corresponding to experiment ID 16 in Table 3.

	Actual label					Actual label				
		Digongo	Hoolthy	Prevalence			Digongo	Hoolthy	Prevalence	
		Disease	meaniny	0.18			Disease	meaniny	0.18	
7	Dicease	19	10	PPV	q	Disease	17	25	PPV	
cte	Disease	15	19	0.41	cte	Disease	11	- 55	0.33	
jdi	Hoolthy	F	61	NPV	jpe	Hoolthr	1	45	NPV	
Pre	meaniny	5	01	0.92	P_{r_0}	meaniny	1	40	0.98	
-		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy	
		0.72	0.76	0.76			0.94	0.56	0.63	



Table S64: Confusion matrices when applying SVM to identify the samples of patients in early stage and healthy patients in data set 1 (E1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 2 (C2 vs H2) were used to train the model. This experiment is corresponding to experiment ID 16 in Table 3.

		Actua	l label				Actua	l label	
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence
ا ب			v	0.18					0.18
	Disease	10	0	PPV	q	Disease	17	9	PPV
cte	Disease	10	0	1	cte	Disease	17		0.65
jbi	Haalthar	8	80	NPV	Predi	Hoolthy	1	71	NPV
Pre	Healthy			0.91		meaniny	1	(1	0.99
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.56	1	0.92			0.94	0.89	0.9
			24.105.1				** .	a	

Using CA-125 alone



Table S65: Confusion matrices when applying KNN to identify the samples of patients in late stage and healthy patients in data set 1 (L1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 2 (C2 vs H2) were used to train the model. This experiment is corresponding to experiment ID 17 in Table 3.

		Actua	l label			Actual label				
		Disease	Healthy	Prevalence 0.41			Disease	Healthy	Prevalence 0.41	
Predicted	Disease	42	0	PPV 1	Predicted	Disease	48	2	PPV 0.96	
	Healthy	13	80	NPV 0.86		Healthy	7	78	NPV 0.92	
		Sensitivity 0.76	Specificity 1	Accuracy 0.9			Sensitivity 0.87	Specificity 0.98	Accuracy 0.93	

Using CA-125 alone

Using four proteins

Table S66: Confusion matrices when applying logistic regression to identify the samples of patients in late stage and healthy patients in data set 1 (L1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 2 (C2 vs H2) were used to train the model. This experiment is corresponding to experiment ID 17 in Table 3.

		Actua	l label			Actual label					
		Digoago	Hoolthy	Prevalence			Digongo	Hoolthy	Prevalence		
p		Disease	meaniny	0.41			Disease	meaniny	0.41		
	Disease	49	0	PPV	q	Disease	18	0	PPV		
cte	Disease	42	0	1	cte	Disease	10		1		
ibe	Healthy	13	80	NPV	edi	Hoolthr	7	80	NPV		
Pre				0.86	$\mathbf{P}_{\mathbf{r}\epsilon}$	meaniny			0.92		
-		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy		
		0.76	1	0.9			0.87	1	0.95		



Table S67: Confusion matrices when applying random forest to identify the samples of patients in late stage and healthy patients in data set 1 (L1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 2 (C2 vs H2) were used to train the model. This experiment is corresponding to experiment ID 17 in Table 3.

		Actua	l label		Actual label					
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence	
	D:	43	10	PPV		Disease	51	7	PPV	
cted	Disease			0.81	cteo				0.88	
Predic	Healthy	12	70	NPV 0.85	Predi	Healthy	4	73	NPV 0.95	
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy	
		0.78	0.88	0.83			0.93	0.91	0.92	
								_		



Table S68: Confusion matrices when applying SVM to identify the samples of patients in late stage and healthy patients in data set 1 (L1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 2 (C2 vs H2) were used to train the model. This experiment is corresponding to experiment ID 17 in Table 3.

	Actual label					Actual label					
		Disease	Healthy	Prevalence 0.41			Disease	Healthy	Prevalence 0.41		
Predicted	Disease	41	0	PPV 1	Predicted	Disease	48	0	PPV 1		
	Healthy	14	80	NPV 0.85		Healthy	7	80	NPV 0.92		
		Sensitivity 0.75	Specificity 1	Accuracy 0.9			Sensitivity 0.87	Specificity 1	Accuracy 0.95		

Using CA-125 alone

Using four proteins

Table S69: Confusion matrices when applying KNN to identify the samples of ovarian cancer patients and healthy patients in data set 1 (C1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 2 (C2 vs H2) were used to train the model. This experiment is corresponding to experiment ID 18 in Table 3.

	Actual label					Actual label					
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence		
			, i i i i i i i i i i i i i i i i i i i	0.48					0.48		
Ч	Dianaga	50	0	PPV	ъ	Diagona	GE	0	PPV		
cte	Disease	52	0	1	cte	Disease	00		1		
jģi	Haalthar	21	80	NPV	ibi	G II hhe	8	80	NPV		
Pre	Healthy			0.78	\Pr	Healthy			0.91		
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy		
		0.71	1	0.86			0.89	1	0.95		



Table S70: Confusion matrices when applying logistic regression to identify the samples of ovarian cancer patients and healthy patients in data set 1 (C1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 2 (C2 vs H2) were used to train the model. This experiment is corresponding to experiment ID 18 in Table 3.

		Actua	l label				Actua	l label	
		Disease	Healthy	Prevalence			Disease	Healthy	Prevalence
		Discuse	mountify	0.48			Dibease	incartiny	0.48
7	Digongo	51	0	PPV	لم ا	Dicesce	66	0	PPV
cte	Disease	51	0	1	cte	Disease	00		1
j	TT 14 h	22	80	NPV	bib	Healthy	7	80	NPV
Pre	meaniny			0.78	$\mathbf{P}_{\mathbf{r}\epsilon}$		1		0.92
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy
		0.7	1	0.86			0.9	1	0.95
	Using CA-125 alone						Using	four proteins	

Table S71: Confusion matrices when applying random forest to identify the samples of ovarian cancer patients and healthy patients in data set 1 (C1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 2 (C2 vs H2) were used to train the model. This experiment is corresponding to experiment ID 18 in Table 3.

		Actua	l label			Actual label					
		Disease	Healthy	Prevalence 0.48			Disease	Healthy	Prevalence 0.48		
Predicted	Disease	56	10	PPV 0.85	Predicted	Disease	67	10	PPV 0.87		
	Healthy	17	70	NPV 0.8		Healthy	6	70	NPV 0.92		
		Sensitivity 0.77	Specificity 0.88	Accuracy 0.82			Sensitivity 0.92	Specificity 0.88	Accuracy 0.9		

Using four proteins

Table S72: Confusion matrices when applying SVM to identify the samples of ovarian cancer patients and healthy patients in data set 1 (C1 vs H1), using only CA-125 (left panel) and all four proteins (right panel). The samples from ovarian cancer patients and healthy patients in data set 2 (C2 vs H2) were used to train the model. This experiment is corresponding to experiment ID 18 in Table 3.

		Actua	l label			Actual label				
		Disease	Healthy	Prevalence 0.48			Disease	Healthy	Prevalence 0.48	
Predicted	Disease	49	0	PPV 1	Predicted	Disease	64	0	PPV 1	
	Healthy	24	80	NPV 0.77		Healthy	9	80	$\begin{array}{c} \mathrm{NPV} \\ 0.9 \end{array}$	
		Sensitivity	Specificity	Accuracy			Sensitivity	Specificity	Accuracy	
		0.67	1	0.84			0.88	1	0.94	





Figure S1: The AUC and Youden index when applying KNN, logistic regression, and random forest to identify control and healthy samples in experiment 1 in Table 8, using only CA-125 and all four proteins.



Figure S2: The AUC and Youden index when applying KNN, logistic regression, and random forest to identify control and healthy samples in experiment 2 in Table 8, using only CA-125 and all four proteins.



Figure S3: The AUC and Youden index when applying KNN, logistic regression, and random forest to identify control and healthy samples in experiment 3 in Table 8, using only CA-125 and all four proteins.



Figure S4: The AUC and Youden index when applying KNN, logistic regression, and random forest to identify control and healthy samples in experiment 4 in Table 8, using only CA-125 and all four proteins.



Figure S5: The AUC and Youden index when applying KNN, logistic regression, and random forest to identify control and healthy samples in experiment 5 in Table 8, using only CA-125 and all four proteins.



Figure S6: The AUC and Youden index when applying KNN, logistic regression, and random forest to identify control and healthy samples in experiment 6 in Table 8, using only CA-125 and all four proteins.



Figure S7: The AUC and Youden index when applying KNN, logistic regression, and random forest to identify control and healthy samples in experiment 7 in Table 8, using only CA-125 and all four proteins.



Figure S8: The AUC and Youden index when applying KNN, logistic regression, and random forest to identify control and healthy samples in experiment 8 in Table 8, using only CA-125 and all four proteins.



Figure S9: The AUC and Youden index when applying KNN, logistic regression, and random forest to identify control and healthy samples in experiment 9 in Table 8, using only CA-125 and all four proteins.



Figure S10: The AUC and Youden index when applying KNN, logistic regression, and random forest to identify control and healthy samples in experiment 10 in Table 8, using only CA-125 and all four proteins.



Figure S11: The AUC and Youden index when applying KNN, logistic regression, and random forest to identify control and healthy samples in experiment 11 in Table 8, using only CA-125 and all four proteins.



Figure S12: The AUC and Youden index when applying KNN, logistic regression, and random forest to identify control and healthy samples in experiment 12 in Table 8, using only CA-125 and all four proteins.



Figure S13: The AUC and Youden index when applying KNN, logistic regression, and random forest to identify control and healthy samples in experiment 13 in Table 8, using only CA-125 and all four proteins.



Figure S14: The AUC and Youden index when applying KNN, logistic regression, and random forest to identify control and healthy samples in experiment 14 in Table 8, using only CA-125 and all four proteins.



Figure S15: The AUC and Youden index when applying KNN, logistic regression, and random forest to identify control and healthy samples in experiment 15 in Table 8, using only CA-125 and all four proteins.



Figure S16: The AUC and Youden index when applying KNN, logistic regression, and random forest to identify control and healthy samples in experiment 16 in Table 8, using only CA-125 and all four proteins.



Figure S17: The AUC and Youden index when applying KNN, logistic regression, and random forest to identify control and healthy samples in experiment 17 in Table 8, using only CA-125 and all four proteins.



Figure S18: The AUC and Youden index when applying KNN, logistic regression, and random forest to identify control and healthy samples in experiment 17 in Table 8, using only CA-125 and all four proteins.