

ВСЕРОССИЙСКАЯ НАУЧНО-ПРАКТИЧЕСКАЯ КОНФЕРЕНЦИЯ
«ЗДРАВООХРАНЕНИЕ БУДУЩЕГО: ОРИЕНТИР НА КАЧЕСТВО»

8-9 декабря 2023 г.

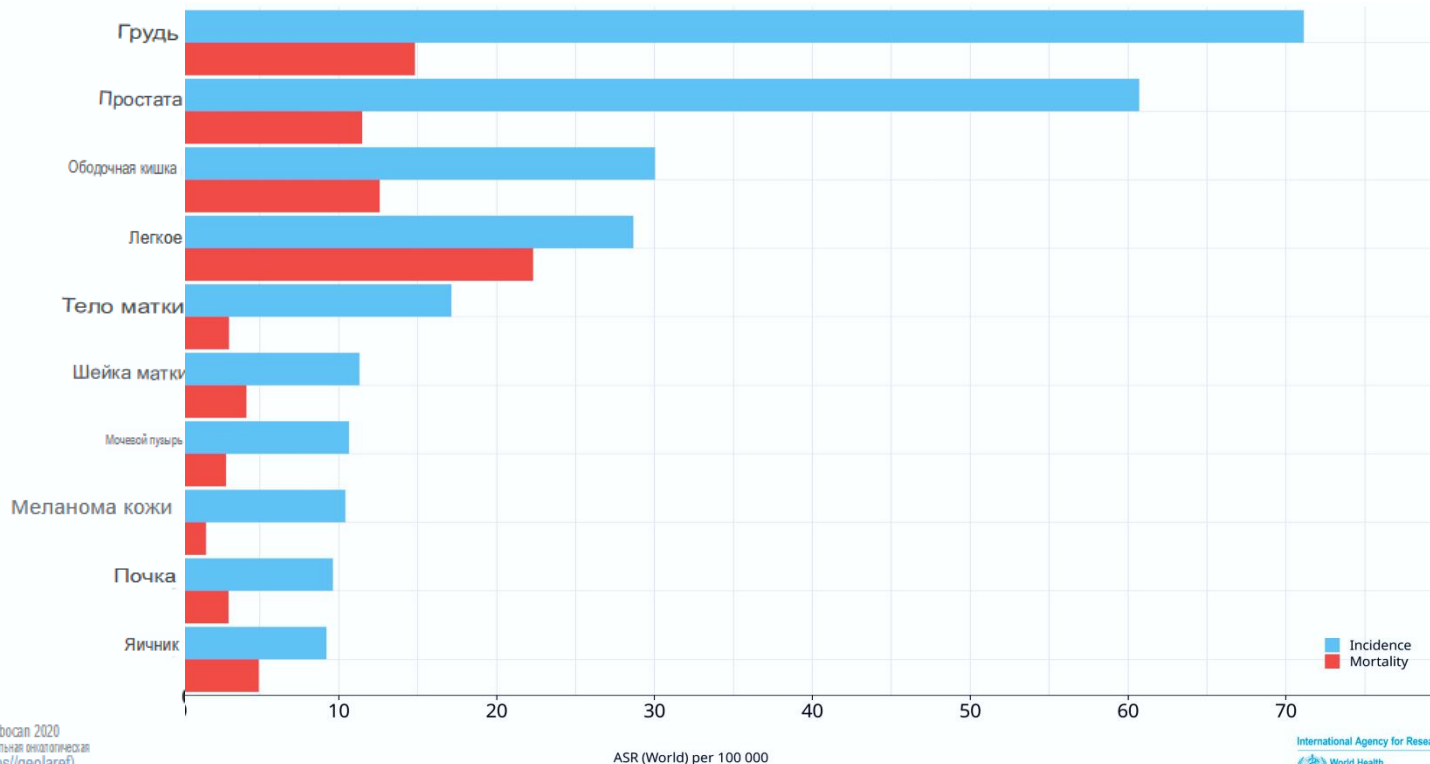
**ДИАГНОСТИКА РАКА МОЛОЧНОЙ
ЖЕЛЕЗЫ С ИСПОЛЬЗОВАНИЕМ
МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ**

Студенческая наука «Здравоохранение будущего»

Кухто Яна Сергеевна
Научный руководитель
к.ф.-м.н., доцент Е. Е. Голденко

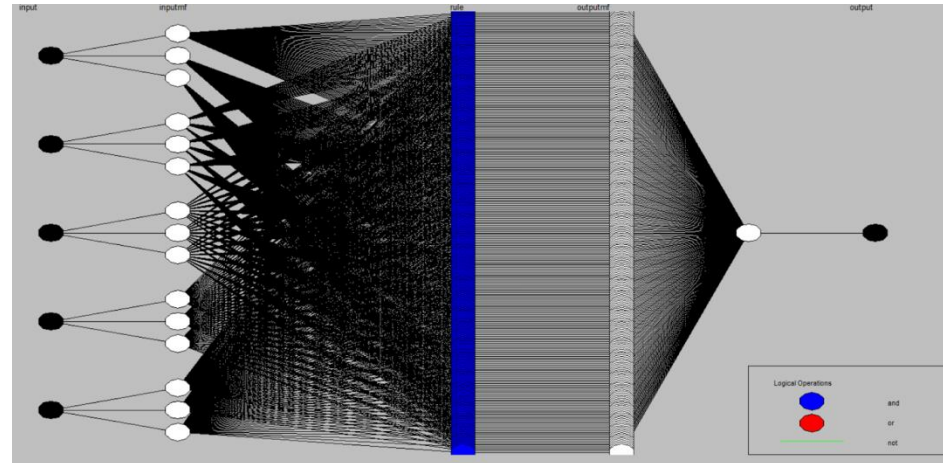
АКТУАЛЬНОСТЬ И ЦЕЛЬ ИССЛЕДОВАНИЯ

Цель: Диагностика рака молочной железы по профилю экспрессии генов с использованием адаптивной нейронной сети и методов машинного обучения.



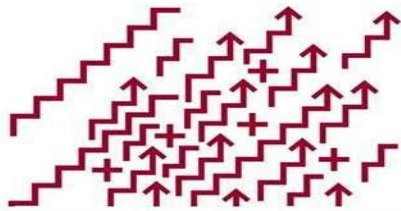
Материалом для исследования является база данных результатов ДНК-микрочипирования 590 образцов ткани молочной железы. Для решения задачи классификации были использованы адаптивная нейронная сеть ANFIS и методы машинного обучения: NB, RF, SVM, KNN

	A	B	C	D	E	F	G	H	I	J
1	Hybridization REF	TCGA-AO-A	TCGA-A8-/	TCGA-A8-/	TCGA-A8-/	TCGA-A8-/	TCGA-A8-/	TCGA-A8-/	TCGA-A8-/	TCGA-A8-/
2	ELMO2	1.21941666	0.43033333	0.43983333	1.242667	-0.22833	-0.34308	0.442167	0.752333	0.287583
3	CREB3L1	-0.15225	0.4445	1.88225	1.781	0.5335	0.79425	1.43475	1.12325	1.1055
4	RPS11	2.193375	-0.100625	0.055625	0.2695	0.722375	0.7395	0.525125	0.5345	0.178125
5	PNMA1	-0.21875	-0.1675	-0.026	0.14975	0.4275	0.284	0.482	-0.921	-0.34875
6	MMP2	-0.1486666	-0.206	0.9701666	-1.07317	-0.28367	-0.43583	0.702833	-0.378	0.102833
7	C10orf90	-1.955	-2.0185	-3.2575	-2.74075	-1.848	-2.1535	-2.17525	-2.2335	-1.806
8	ZHX3	0.21083333	0.1425	-0.194333	0.379667	-0.554	-0.027	0.078	-0.2895	-0.63717
9	ERCC5	0.42975	-0.90075	-0.73925	-0.23675	0.09875	0.312	0.29425	-0.26275	-0.08175
10	GPR98	-0.970875	-2.210625	-0.200375	-1.7785	-1.851	-0.472	-1.30225	1.61125	-0.27575
11	RXFP3	0.523	0.497	0.2415	0.761	0.7265	0.309	-0.1245	0.6925	0.5605
12	APBB2	1.458	2.04375	1.30125	0.025	0.7	1.29975	2.03775	0.25225	0.2335
13	PRO0478	-1.234	-0.67125	-0.717125	-1.05638	-0.76275	-0.88188	-1.04563	-0.65675	-1.46863
14	KLHL13	-3.43475	-4.157	-0.76075	-4.4365	-3.881	-1.42625	-1.97925	-2.32125	0.13675
15	PRSSL1	-0.7603333	0.5711666	-0.947	-0.468	0.7015	-0.96333	-0.78233	-0.79483	-0.47767
16	PDCL3	0.53733333	0.18533333	0.1308333	0.501333	0.072833	0.4805	0.538667	0.7565	0.409333



ОЦЕНКА КАЧЕСТВА КЛАССИФИКАЦИИ

Модель	Accuracy (%)	Точность(%)	Полнота(%)	F1
NB	97,6	97,5	97,9	97,6
RF	98,6	98,5	98,7	98,6
SVM	98,6	98,5	98,7	98,6
KNN	98,1	98,0	98,2	98,1
ANFIS	95,0	95,0	94,9	94,9



- В результате предобработки данных и устранения дисбаланса классов объем обучающей выборки был увеличен с 681 наблюдений до 1058.
- При снижении размерности данных выявлены 5 наиболее значимых для классификации генов (HOOK3, SYP, C6orf57, RAB1F, HRASLS).
- Наибольшая точность классификации РМЖ получена при использовании моделей RF, SVM.

