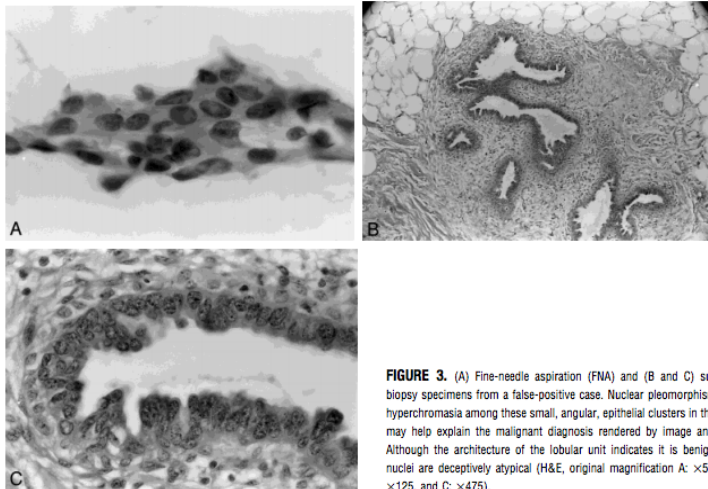# Breast Cancer(Diagonostic) Data Set

Classification of Breast Tumors as Malignant or Benign

Amir Ghaderi & Mezbah Uddin

13$^{\text{th}}$ February 2017

FNA Digitized Images: Source: UW

FIGURE 3. (A) Fine-needle aspiration (FNA) and (B and C) surgical biopsy specimens from a false-positive case. Nuclear pleomorphism and hyperchromasia among these small, angular, epithelial clusters in the FNA may help explain the malignant diagnosis rendered by image analysis. Although the architecture of the lobular unit indicates it is benign, the nuclei are deceptively atypical (H&E, original magnification A: ×500, B: ×125, and C: ×475).

Figure 1: OLTAS

Overview of the dataset: Source : UCI machine learning repo

id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	symmetry_mean
842302	M	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001	0.1471	0.2419
842517	M	20.57	17.77	132.9	1326	0.08474	0.07864	0.0869	0.07017	0.1812
84300903	M	19.69	21.25	130	1203	0.1096	0.1599	0.1974	0.1279	0.2069
84348301	M	11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	0.1052	0.2597
84358402	M	20.29	14.34	135.1	1297	0.1003	0.1328	0.198	0.1043	0.1809
843786	M	12.45	15.7	82.57	477.1	0.1278	0.17	0.1578	0.08089	0.2087
844359	M	18.25	19.98	119.6	1040	0.09463	0.109	0.1127	0.074	0.1794
84458202	M	13.71	20.83	90.2	577.9	0.1189	0.1645	0.09366	0.05985	0.2196
844981	M	13	21.82	87.5	519.8	0.1273	0.1932	0.1859	0.09353	0.235
84501001	M	12.46	24.04	83.97	475.9	0.1186	0.2396	0.2273	0.08543	0.203
845636	M	16.02	23.24	102.7	797.8	0.08206	0.06669	0.03299	0.03323	0.1528
84610002	M	15.78	17.89	103.6	781	0.0971	0.1292	0.09954	0.06606	0.1842
846226	M	19.17	24.8	132.4	1123	0.0974	0.2458	0.2065	0.1118	0.2397
846381	M	15.85	23.95	103.7	782.7	0.08401	0.1002	0.09938	0.05364	0.1847
84667401	M	13.73	22.61	93.6	578.3	0.1131	0.2293	0.2128	0.08025	0.2069
84799002	M	14.54	27.54	96.73	658.8	0.1139	0.1595	0.1639	0.07364	0.2303
848406	M	14.68	20.13	94.74	684.5	0.09867	0.072	0.07395	0.05259	0.1586
84862001	M	16.13	20.68	108.1	798.8	0.117	0.2022	0.1722	0.1028	0.2164
849014	M	19.81	22.15	130	1260	0.09831	0.1027	0.1479	0.09498	0.1582
8510426	B	13.54	14.36	87.46	566.3	0.09779	0.08129	0.06664	0.04781	0.1885
8510653	B	13.08	15.71	85.63	520	0.1075	0.127	0.04568	0.0311	0.1967
8510824	B	9.504	12.44	60.34	273.9	0.1024	0.06492	0.02956	0.02076	0.1815

Data Set Features

- Dataset Source: UCI Machine learning Repository
- 569 digitized images
- Class Distribution: 2 Classes, Benign(357) & Malignant(212).
- 10 Features : radius,texture,perimeter,area,smoothness,compactness
- cocavity,concave points,symmetry,fractal dimensions
- No missing attributes

Classification Exploration

- Classifiers to be chosen based on accuracy .
- Dimensionality reduction: PCA,SVD
- Parallelized k-fold validation

- Thank You.