

Feature Selection

Subject	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6
Normal	3	5	5.1			
Normal	8.7	9	5			
Normal	6	8	4.9			
Normal	6.5	7	5.2			
AD	8	16	5			
AD	8.5	15	5			
AD	9.2	20	4.8			
AD	7.9	20	4.9			

Regularization

$$J(w) = -\frac{1}{m} \left[\sum_{i=1}^m y^{(i)} \log h_w(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_w(x^{(i)})) \right] + \lambda \sum_{j=1}^n w_j^2$$

Repeat {

$$w_0 := w_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)}) x_0^{(i)}$$

$$w_j := w_j (1 - \alpha \frac{\lambda}{m}) - \alpha \frac{1}{m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

}

($j = 0, 1, 2, 3, \dots, n$)

Regularization

$$J(w) = -\frac{1}{m} \left[\sum_{i=1}^m y^{(i)} \log h_w(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_w(x^{(i)})) \right] + \lambda \sum_{j=1}^n |w_j|$$

Regularization

$$J(w) = -\frac{1}{m} \left[\sum_{i=1}^m y^{(i)} \log h_w(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_w(x^{(i)})) \right] + \lambda \sum_{j=1}^n |w_j|$$

Repeat {

$$w_0 := w_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)}) x_0^{(i)}$$

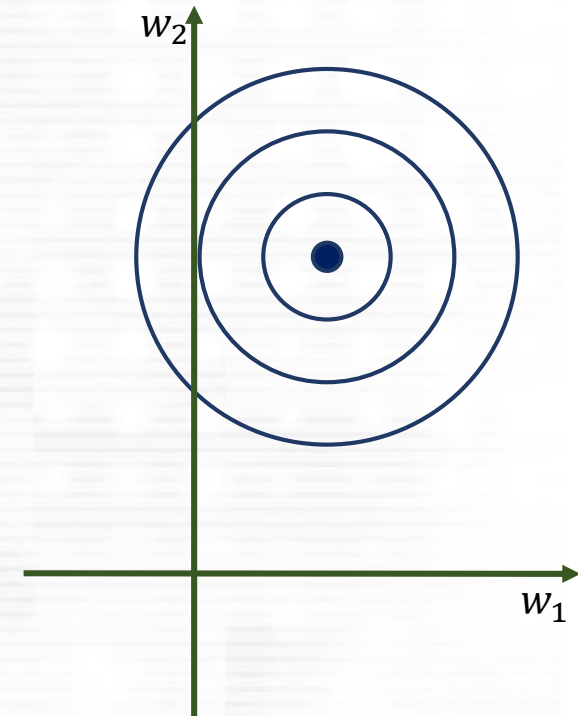
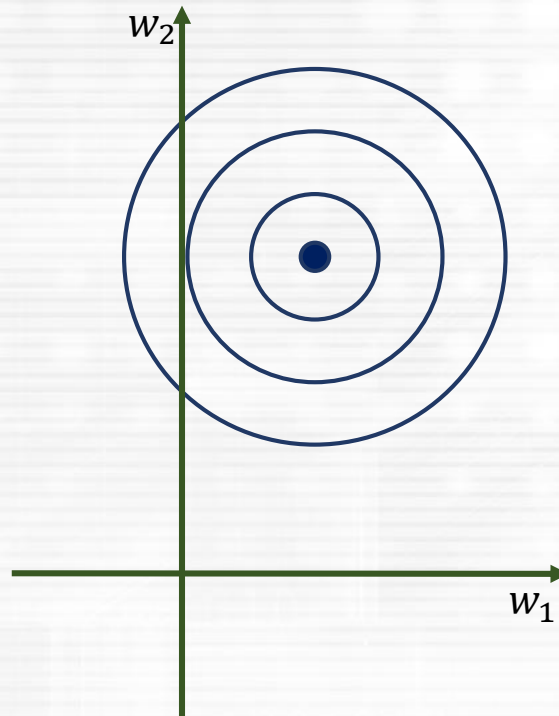
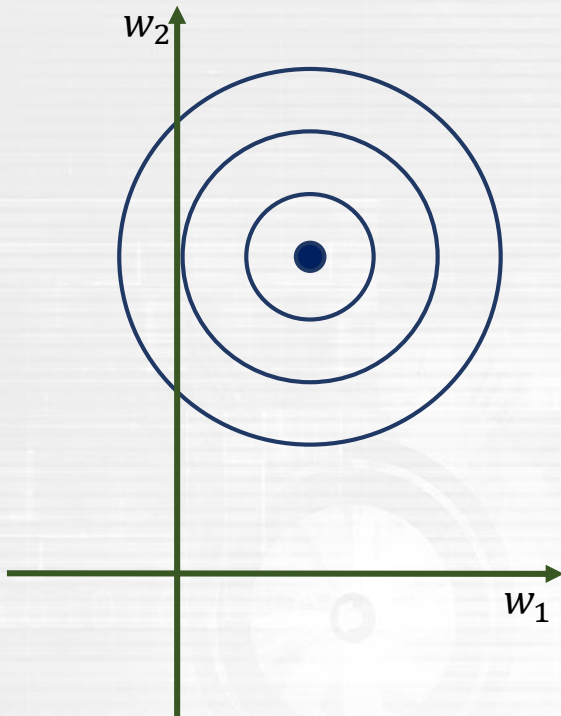
$$w_j := w_j - \alpha \frac{\lambda}{m} \text{sgn}(w_j) - \alpha \frac{1}{m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

}

$$(j = 0, 1, 2, 3, \dots, n)$$

Regularization

$$J(w) = -\frac{1}{m} \left[\sum_{i=1}^m y^{(i)} \log h_w(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_w(x^{(i)})) \right]$$



Feature Selection

Subject	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6
Normal	3	5	5.1			
Normal	8.7	9	5			
Normal	6	8	4.9			
Normal	6.5	7	5.2			
AD	8	16	5			
AD	8.5	15	5			
AD	9.2	20	4.8			
AD	7.9	20	4.9			

Entropy

- Amount of information: $\log\{p(x)\}$

- Entropy: $H(x) = -\sum_x p(x) \log\{p(x)\}$

- Joint entropy:
$$H(X, Y) = -\sum_{i=1}^N \sum_{j=1}^M p(x_i, y_j) \log\{p(x_i, y_j)\}$$

- Mutual information:
$$I(X; Y) = \sum_{i=1}^N \sum_{j=1}^M p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}$$

$$I(X; Y) = H(X) + H(Y) - H(X, Y)$$

Decision Tree

Subject	Feature 1	Feature 2	Feature 3
Normal	3	5	5.1
Normal	8.7	9	5
Normal	6	8	4.9
Normal	6.5	7	5.2
AD	8	16	5
AD	8.5	15	5
AD	9.2	20	4.8
AD	7.9	20	4.9

mRMR Feature Selection

Minimum-redundancy-maximum-relevance(mRMR) feature selection

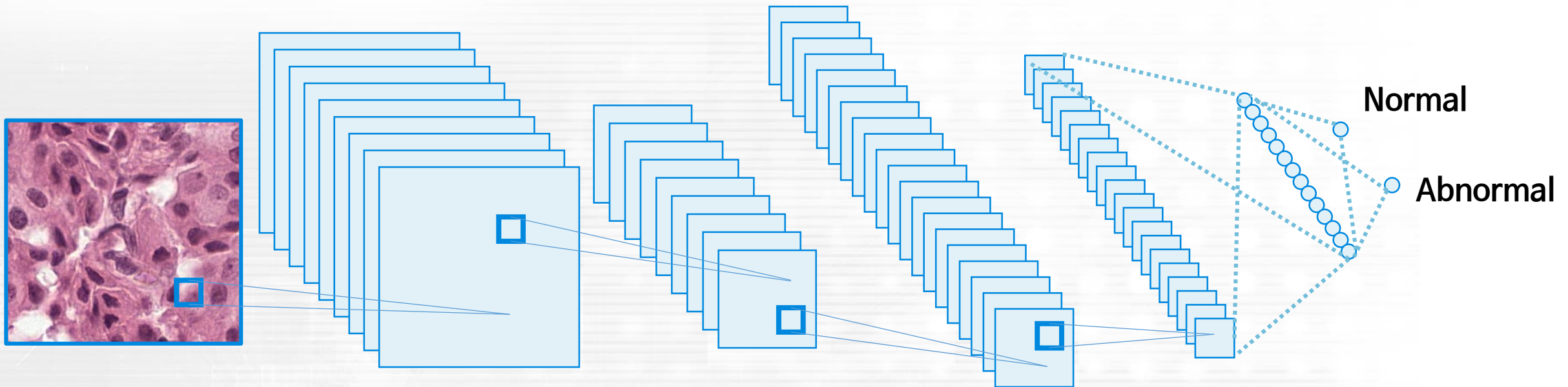
Subject	Feature 1	Feature 2	Feature 3
Normal	3	5	5.1
Normal	8.7	9	5
Normal	6	8	4.9
Normal	6.5	7	5.2
AD	8	16	5
AD	8.5	15	5
AD	9.2	20	4.8
AD	7.9	20	4.9

$$D(S, c) = \frac{1}{|S|} \sum_{f_i \in S} I(f_i; c)$$

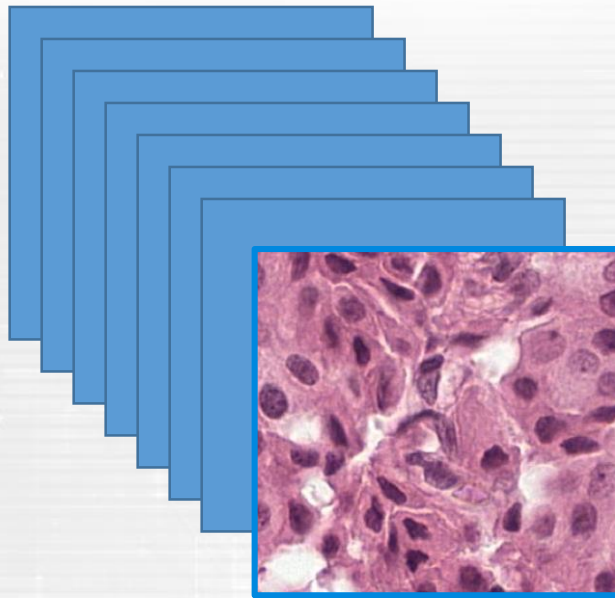
$$R(S) = \frac{1}{|S|^2} \sum_{f_i, f_j \in S} I(f_i; f_j)$$

$$mRMR = \max_s \left[\frac{1}{|S|} \sum_{f_i \in S} I(f_i; c) - \frac{1}{|S|^2} \sum_{f_i, f_j \in S} I(f_i; f_j) \right]$$

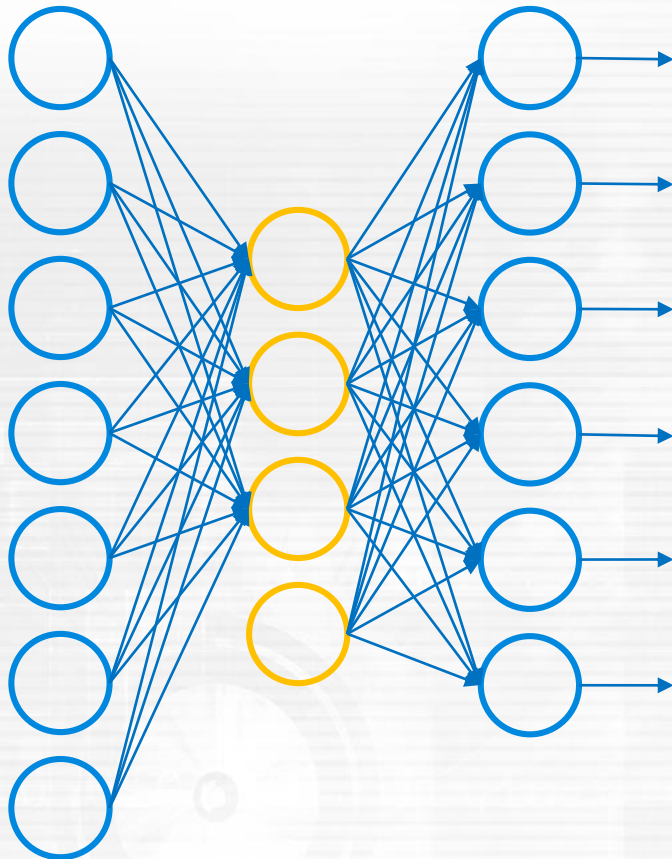
Feature Extraction



Feature Extraction



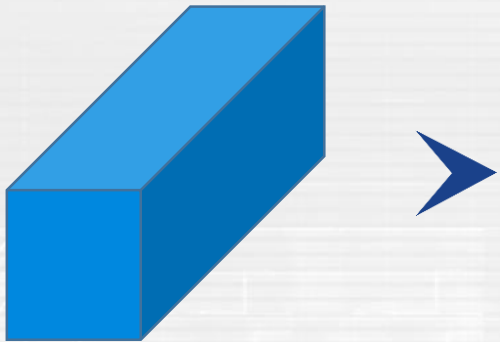
Auto-Encoder



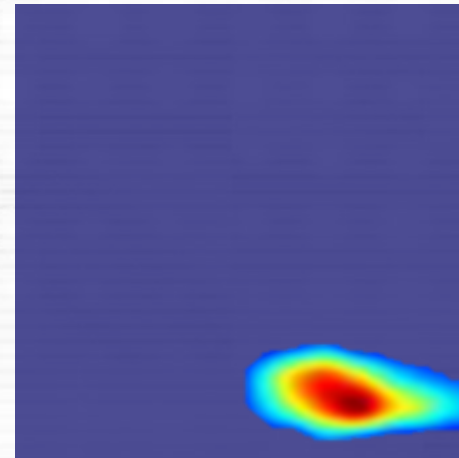
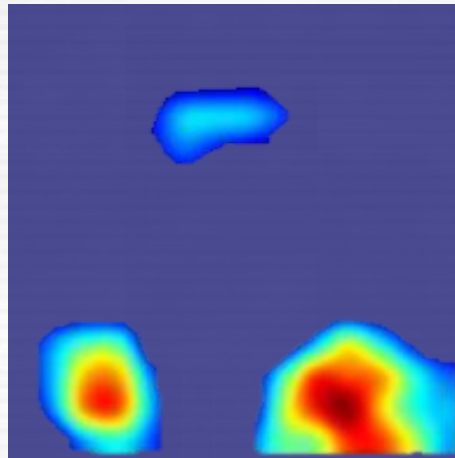
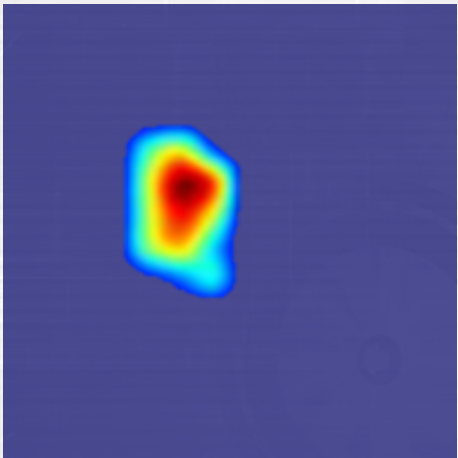
Class Activation Map



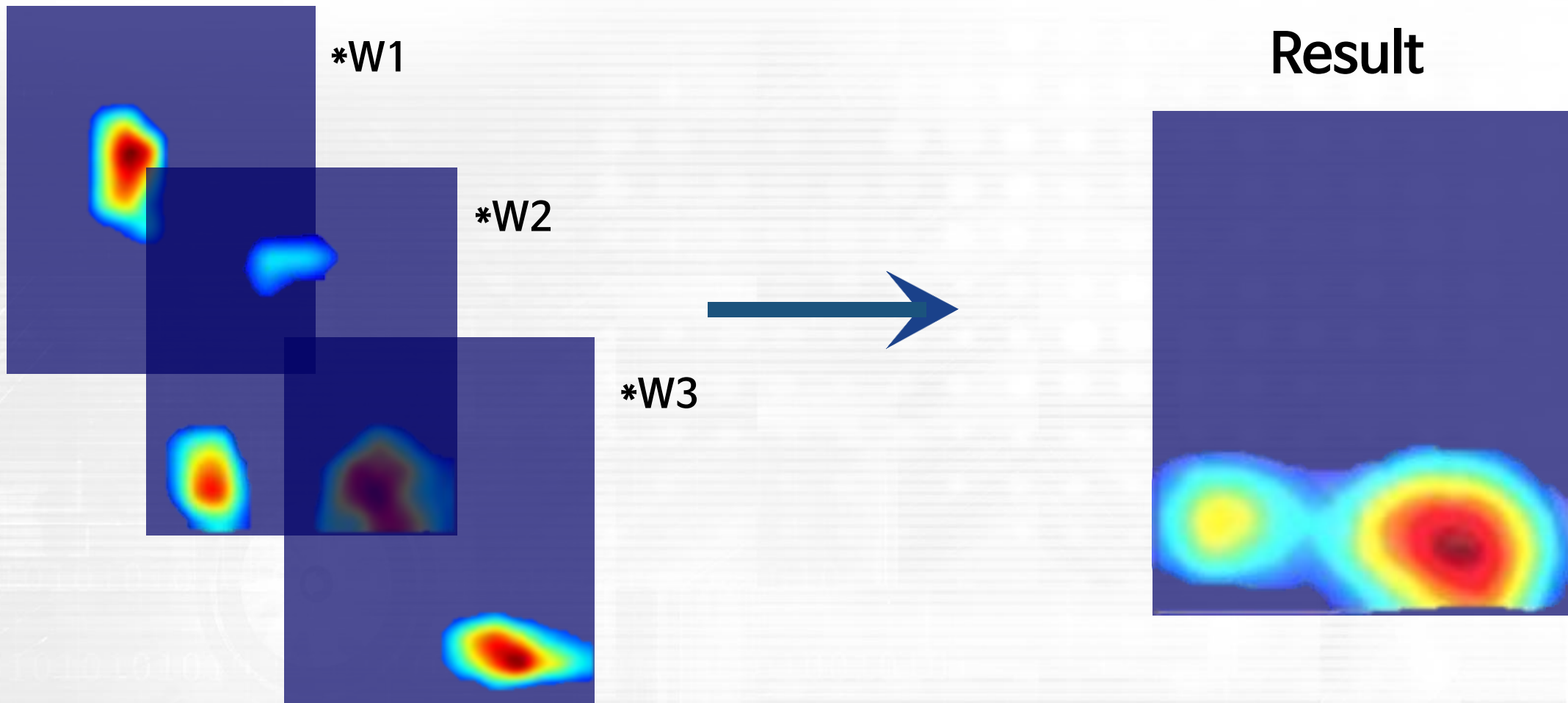
Global Average Pooling



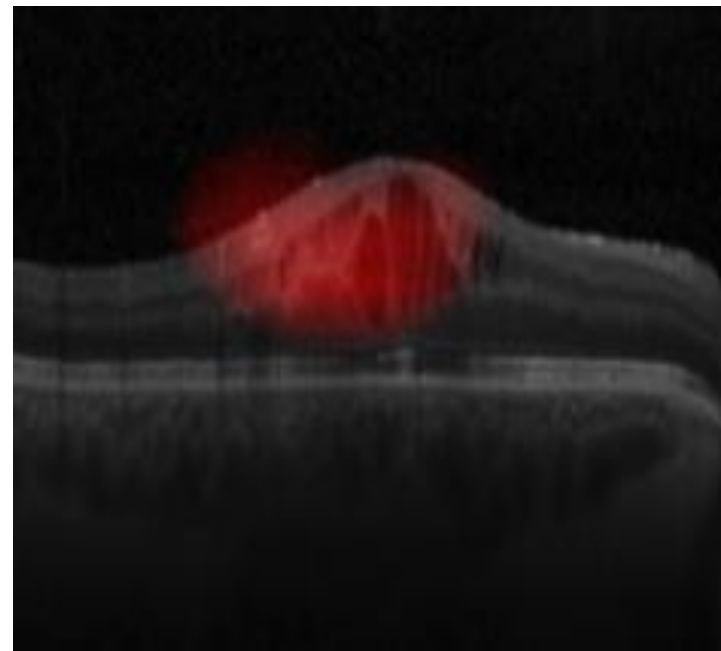
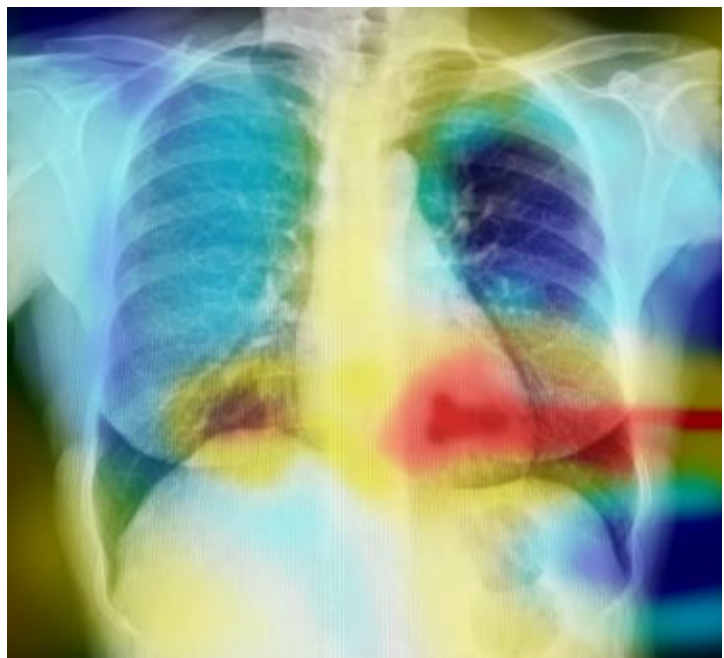
Class Activation Map



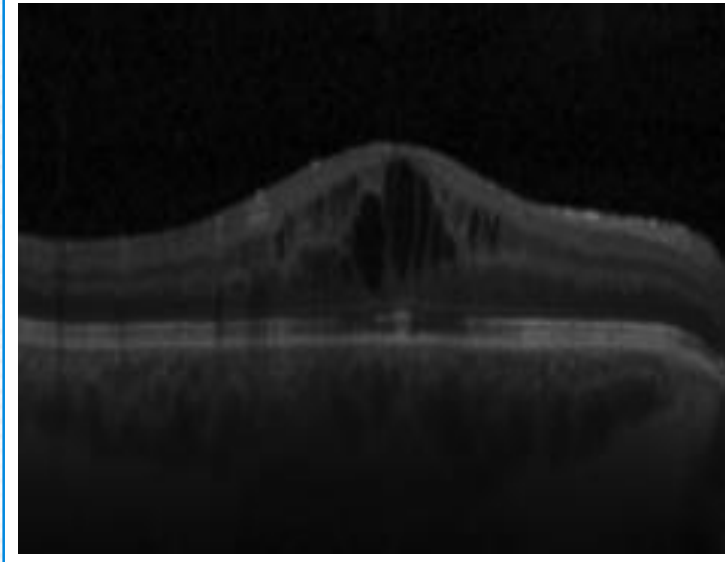
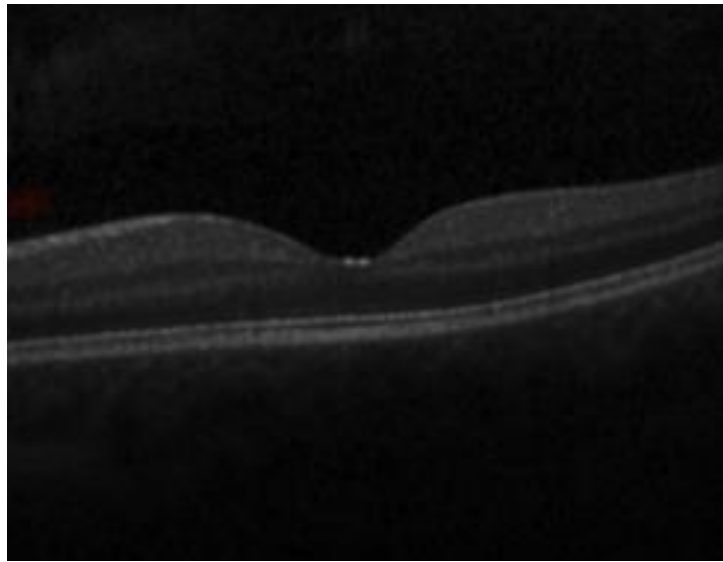
Class Activation Map



Applications



Weakly Supervised Learning



Multiple Instance Learning



Multiple Instance Learning

