Image Classification for Age-related Macular Degeneration Screening using Hierarchical Image Decomposition and Graph Mining

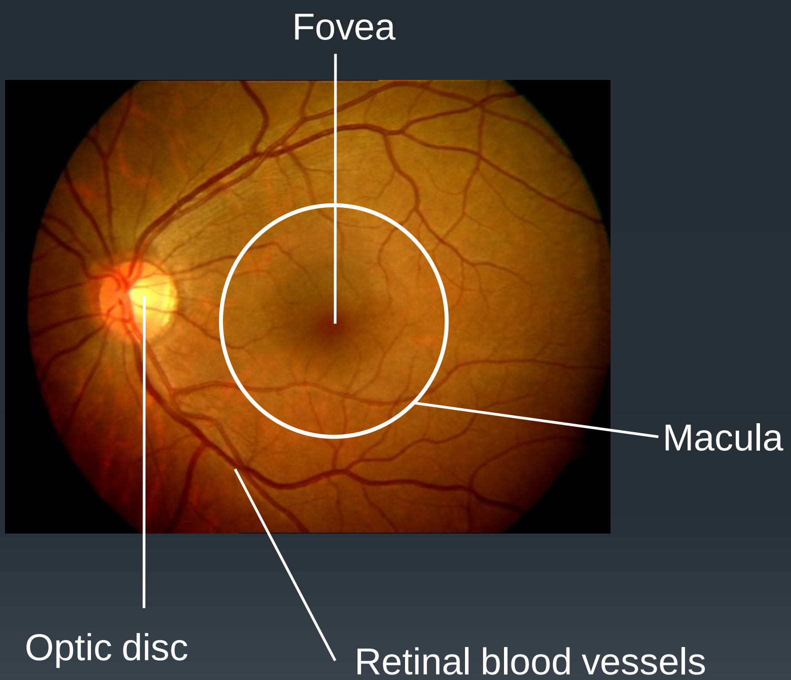
Mohd Hanafi Ahmad Hijazi, Chuntao Jiang, Frans Coenen and Yalin Zheng

ECML PKDD 2011
Athens, Greece
26 September 2011

Presentation outline

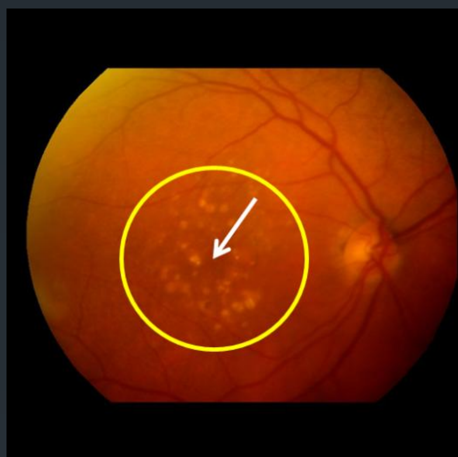
- Age-related Macular Degeneration (AMD)
- AMD screening challenges
- Overview of the proposed AMD screening approach
- Image pre-processing
- Image decomposition
- Graph mining
- Feature selection
- Classifier generation
- Experimental setup
- Results and discussion
- Conclusion and future work

Age-related Macular Degeneration

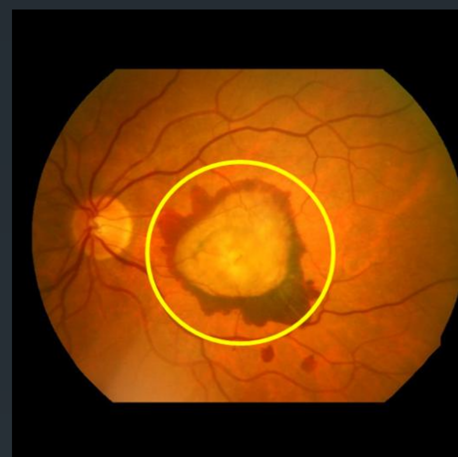


- AMD
  - Cells of the macula become damaged and stop functioning at a later stage of life
  - Loss of vision/ blindness
- Early diagnosis achieved through the detection of drusen (hard or soft)

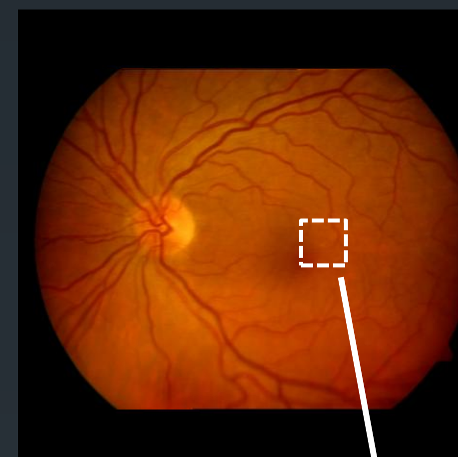
More example of AMD images



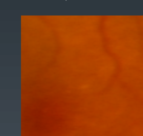
[a]



[b]



[c]



- Challenges:
  - Difficulties to localise the optic disc and macula
  - Difficulties to identify drusen

Overview of the proposed AMD screening approach

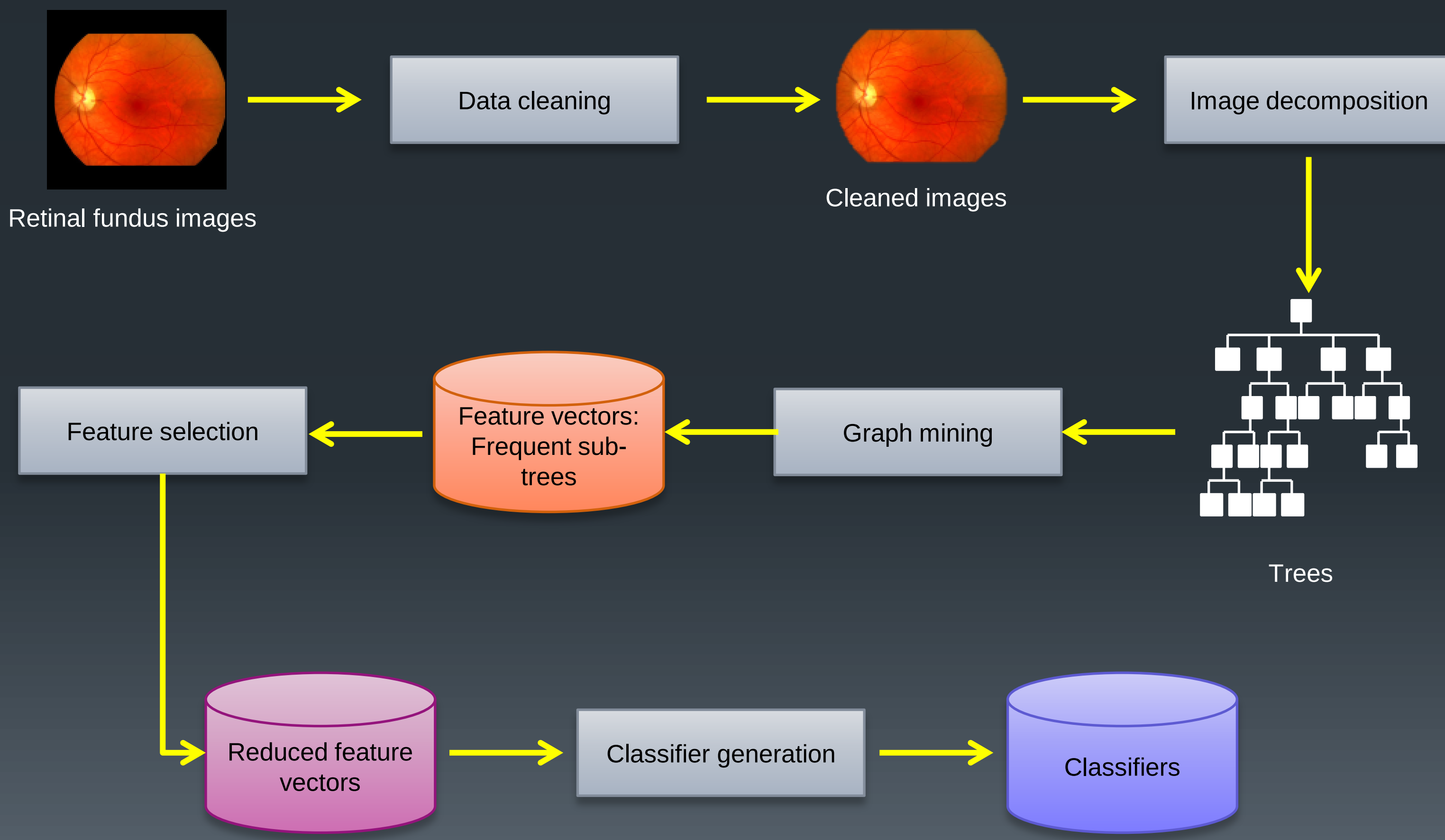


Image pre-processing

- Image enhancement
  - Colour normalisation using histogram specification approach.
  - Illumination normalisation using luminosity estimates (Foracchia et al., 2005) and adaptive histogram equalisation.
- Noise removal
  - Retinal blood vessel removal
  - Blood vessel segmentation using wavelet and supervised pixels classification approach (Soares et al., 2006).
  - ‘Vessel’ pixels value is replaced by a null value, and will not be counted in tree generation.
- Output: retinal images with the dark background and blood vessels pixels labelled with ‘null’ value.

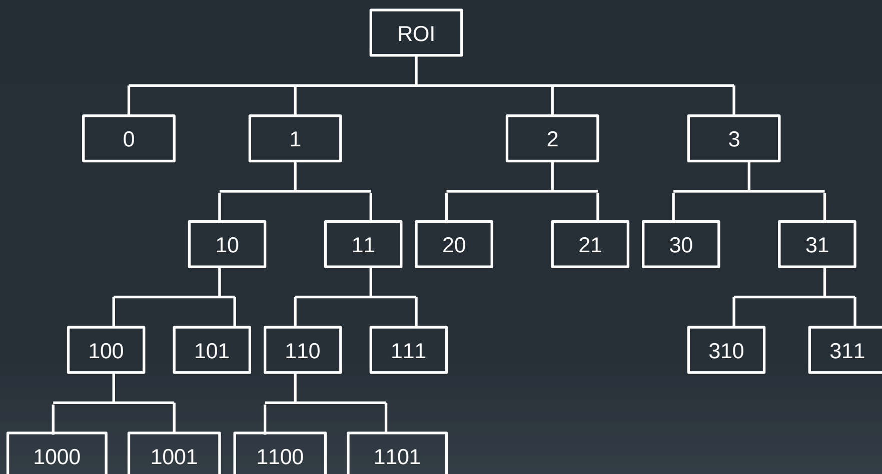
Image decomposition

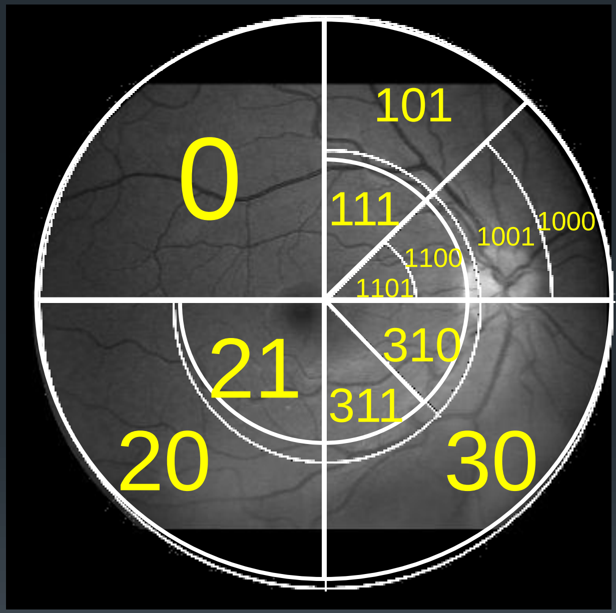
- Circular and angular partitioning.
- Each region (node) is split into two child nodes, except for depth 1 (four child nodes).
- Splitting criterion – how well the parent node represents its child nodes intensity values, using:

$$\omega = \frac{1}{s}\sum_{i=1}^{s}\sqrt{(\mu_P - \mu_i)^2}$$

- Parent node is homogeneous if $\omega < \tau$ (a predefined threshold) → no splitting is required.
- Output: Trees that represent retinal images (one tree per image).

Image decomposition





- Each node weight is defined by the average intensity value of that particular node.
- Each edge weight is defined by the intensity distance between parent and its child nodes.

Graph mining

- Aim: to identify weighted Frequent Sub-Trees (wFSTs) using a weighted frequent sub-graph mining algorithm (weighted gSpan).
- A sub-tree, *g*, is ‘frequent’ if it satisfies two conditions:
  - **C1:** $\boldsymbol{N_{wr} \times sup(g) \geq \sigma}$
    - N_{wr} is the accumulated weights of nodes in *g*, *sup*(*g*) is the occurrences of *g* in the dataset and σ is the *minimum support* threshold.
  - **C2:** $\boldsymbol{E_{wr} \geq \lambda}$
    - E_{wr} is the accumulated weights of edges in *g* and λ is the *minimum weight* threshold.
- Output: wFSTs (sub-trees that satisfies both conditions).

Feature selection

- Identify and rank wFSTs that displayed strong discriminatory power.
- Get each wFST weight using L2-regularized L2-loss SVM model.
- Sort wFSTs according to their weights absolute value.
- Select only top *K* wFSTs for classification.
- Output: Reduced number of wFSTs.

Classifier generation

- Two establish classification technique was employed:
  - Naïve Bayes
    - Works well on data with independent attributes,
    - Does not require user defined parameters
  - SVM
    - Acknowledge as one of the most effective classification method.

Experimental setup

- Data acquired from 2 databases
  - 258 retinal images: 160 AMD, 98 Normal.
- Evaluation metrics:
  - Specificity = TN/(FP+TN).
  - Sensitivity = TP/(TP+FN).
  - Accuracy = (TP+TN)/(TP+FN+TN+FP).
- Experiments aims:
  - The effect of tree depth (D_{max}) on classification,
  - How feature selection improved the classification result,
  - Comparison with other approaches.
- Experiments were conducted using Tenfold Cross Validation (TCV).

Performances using various D_{max} values – Naïve Bayes

σ (%)	5					6					7				
	λ	*F*	*Sens*	*Spec*	*Acc*	*λ*	*F*	*Sens*	*Spec*	*Acc*	*λ*	*F*	*Sens*	*Spec*	*Acc*
10	40	764	65	60	**63**	20	7125	64	53	60	80	3433	66	42	57
20	60	291	68	50	62	80	498	70	42	59	80	3433	66	42	57
30	60	291	68	50	62	20	1746	66	49	60	80	3433	66	42	57
40	20	248	66	48	59	80	498	70	42	59	80	3433	66	42	57
50	20	181	69	45	60	80	498	70	42	59	80	3433	66	42	57
60	20	130	67	46	59	20	559	69	45	60	20	3893	66	42	57
70	20	99	69	33	55	20	404	68	34	55	20	2623	69	41	58
80	20	71	80	13	55	20	283	72	27	55	20	1706	70	36	57
90	20	55	86	16	60	20	180	79	22	58	20	955	72	28	55

Performances using various D_{max} values – LibSVM

σ (%)	5					6					7				
	λ	***F***	***Sens***	***Spec***	***Acc***	***λ***	***F***	***Sens***	***Spec***	***Acc***	***λ***	***F***	***Sens***	***Spec***	***Acc***
10	40	764	86	43	**70**	20	7125	87	39	69	60	11461	96	15	66
20	20	594	89	37	69	20	3103	91	30	68	60	11461	96	15	66
30	20	365	95	16	65	60	1358	89	33	68	60	11461	96	15	66
40	80	118	100	0	62	20	1135	92	30	68	20	9043	97	8	63
50	80	118	100	0	62	20	779	94	11	62	80	3433	96	8	63
60	20	130	83	30	63	20	559	99	1	62	20	3893	96	11	64
70	20	99	99	4	63	20	404	98	0	60	20	2623	99	4	63
80	20	71	100	0	62	20	283	100	1	62	20	1706	97	8	63
90	20	55	99	3	63	20	180	100	0	62	20	955	96	10	63

Performances using various D_{max} values - Discussion

- Most results generated by various σ and λ values produced best accuracy when $D_{max} = 5$ (using both Naïve Bayes and LibSVM).
- The best accuracy achieved by Naïve Bayes was 63%.
- The best accuracy achieved by LibSVM was 70%.
- Both of the best results were generated when σ and λ were 10% and 40%, and $D_{max} = 5$.

Performances using different feature size – Naïve Bayes, $D_{max} = 7$

σ (%)	$P_{0.05}$					$P_{0.1}$					$P_{0.4}$				
	λ	*K*	*Sens*	*Spec*	*Acc*	*λ*	*K*	*Sens*	*Spec*	*Acc*	*λ*	*K*	*Sens*	*Spec*	*Acc*
10	20	3671	94	96	**95**	20	7342	92	92	92	40	16278	73	67	71
20	20	1407	91	93	91	20	2814	88	82	85	20	11257	71	65	68
30	20	748	88	82	85	20	1496	86	78	83	20	5983	71	61	67
40	20	452	85	80	83	20	904	84	72	79	20	3618	71	56	65
50	20	291	85	72	80	80	343	86	64	78	20	2330	73	52	65
60	20	195	84	68	78	20	389	86	67	78	20	1558	76	51	66
70	20	131	83	63	75	20	262	83	56	72	20	1050	74	47	64
80	20	85	82	45	68	20	171	81	50	69	20	683	74	45	63
90	20	48	83	37	65	20	96	84	41	67	20	382	78	37	62

Performances using different feature size – LibSVM, D_{max} = 7

σ (%)	$P_{0.05}$					$P_{0.1}$					$P_{0.4}$				
	λ	K	Sens	Spec	Acc	λ	K	Sens	Spec	Acc	λ	K	Sens	Spec	Acc
10	20	3671	100	100	**100**	20	7342	100	100	100	40	16278	99	81	92
20	80	172	100	0	62	80	343	100	0	62	60	4585	99	8	65
30	20	748	99	80	92	20	1496	99	94	97	20	5983	98	70	87
40	80	172	100	0	62	80	172	100	0	62	80	1374	99	5	64
50	20	291	97	54	81	20	583	96	84	91	20	2330	95	56	80
60	80	172	100	0	62	80	172	100	0	62	80	1374	99	5	64
70	20	131	100	0	62	20	262	100	0	62	20	1050	100	0	63
80	20	85	100	0	62	20	171	100	1	62	20	683	99	6	64
90	20	48	100	0	62	20	96	100	1	62	20	382	98	10	65

Performances using different feature size – Discussion

- D_{max} = 7 performed better than other D_{max} values.
- The best accuracy was 95% (Naïve Bayes) and 100% (LibSVM).
- Both results were generated using σ = 10%, λ = 20% and feature size, K, is 3671.
- Feature selection has improved the classification results, in particular when D_{max} = 7.

Comparison with other AMD classification approaches

- Results were generated using leave-one-out approach.

Approach	Dataset size	Sensitivity	Specificity	Accuracy
Barriga et al. [1]	100	75	50	-
Brandon and Hoover [2]	97	-	-	87
Chaum et al. [3]	395	-	-	88
Proposed approach	258	89	99	**93**

UNIVERSITY OF LIVERPOOL

Conclusion and future work

- An approach of retinal image classification for AMD screening has been described.
- Images were represented by trees, where angular and circular image partitioning was used to decompose images.
- Best results were achieved when deeper tree level *(D_{max} = 7)* and feature selection were applied.
- Current/ future work:
  - To apply this approach to multiclass problems:
    - AMD grading, and
    - Screening other retinal diseases (e.g diabetic retinopathy).

UNIVERSITY OF LIVERPOOL

Thank you.