Biomedical imaging informatics in ocular disease diagnosis

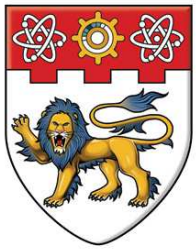
Zhuo, Zhang

2015

Zhuo, Z. (2015). Biomedical imaging informatics in ocular disease diagnosis. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/62523

https://doi.org/10.32657/10356/62523



NANYANG
TECHNOLOGICAL
UNIVERSITY

Biomedical Imaging Informatics in Ocular Disease Diagnosis

Zhuo Zhang
School of Computer Engineering

A thesis submitted to the Nanyang Technological University in partially fulfillment of the requirement for the degree of

Doctor of Philosophy

2015

Acknowledgements

I would like to give my grateful thanks to my adviser, Dr. Kwoh Chee Keong, for his great guidance, patience and support during my study.

I would also like to thank my colleagues, Jimmy, Yanwu, Lixin, Damon, Meng, Cheng Jun, Xingtin, Ruchir, Huijuan and many others for their encouragement, friendships and insightful discussions.

I thank my dear friends Pan Hong, Xiaoqun, Ge Yu, Suisheng, who gave me a lot of mental support during this journey, our lunch gatherings and discussions motivated me all the time.

This thesis was supported by the scholarship of Scientific Staff Development Award (SSDA) from A*STAR (Agency of Science, Technology and Research) Singapore, for which I am very thankful.

Lastly, I sincerely thank my family for everything, it's their support made this journey possible.

Author's publications related to this thesis

Journal publication

1. Zhang Z, Xu Y, Liu J, Wong DWK, Kwoh CK, Saw SM, Wong TY. Automatic Diagnosis of Pathological Myopia from Heterogeneous Biomedical Data, *PlosOne*, 2013, 8(6): e65736.

2. Liu J, Zhang Z, Wong DWK, Xu YW, Yin FS, Cheng J, Tan NM, Kwoh CK, Xu D, Tham YC, Aung T, Wong TY. Automatic Glaucoma Diagnosis through Medical Imaging Informatics, *J Am Med Inform Assoc (JAMIA)*, 2013, 20(6):1021-1027.

3. Zhang Z, Srivastava R, Liu H, Chen X, Duan L, Wong DWK, Kwoh CK, Wong TY and Liu J. A survey on computer aided diagnosis for ocular diseases, *BMC Medical Informatics and Decision Making*, 2014, 14:80.

4. Zhang Z, Kwoh CK, Aung T, Wong TY, Liu J. Automatic Glaucoma Diagnosis with mRMR-based Feature Selection, *Journal of Biometrics & Biostatistics* 2012, S7:008.

5. Duan L, Xu Y, Zhang Z, Wong DWK, Wong TY, Liu J. A Unified Multiple Kernel Learning Framework for the Detection of Ocular Diseases Using Multiple Informatics Domains, *Biomedical Engingeering, IEEE Transctions on*, under review.

Conference publication

1. Chen X, Xu Y, Duan L, Zhang Z, Wong DWK and Liu J. Multiple Ocular Diseases Detection by Graph Regularized Multi-label Learning, *in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014.*

2. Zhang Z, Xu Y, Liu J, Kwoh CK. Identify Predictive SNP groups in Genome Wide Association Study: A Sparse Learning Approach, *International Conference on Computational Systems Biology and Bioinformatics, 2012, 11:107–114.*

3. Xu Y, Liu J, Zhang Z, Tan NM, Wong DWK, Saw WM, Wong TY. Learn to Recognize Pathological Myopia in Fundus Images Using Bag-of-feature and Sparse Learning Approach, *in Proceedings of IEEE International Symposium on Biomedical Imaging, 2013, pp. 888–891.*

4. Zhang Z, Cheng J, Liu J, Yeo CM, Saw WM, Chui CK. Pathological Myopia Detection from Selective Fundus Image Features, *in Proceedings of IEEE. Conf. on Industrial Electronics and Applications (ICIEA), 2012, pp. 1742–1745..*

5. Zhang Z, Kwoh CK, Liu J, Yin F, Wirawan A, Cheung C, Baskaran M, Aung T and Wong TY. MRMR optimized classification for automatic glaucoma diagnosis, *Conf Proc IEEE Eng Med Biol Soc, 2011, pp. 6228–31.*

6. Tan Z, Zhang Z, Liu J, Kwoh CK, Ong SH, Teo YY, Khor CC, Tai ES, Aung T, Vithana E and Wong TY. A hybrid framework for genome wide epistasis discovery, *Conf Proc IEEE Eng Med Biol Soc, 2011, pp. 6479–82.*

7. Zhang Z, Liu J, Kwoh CK, Sim X, Tay WT, Tan Y, Yin F and Wong TY. Learning in glaucoma genetic risk assessment, *Conf Proc IEEE Eng Med Biol Soc, 2010, pp. 6182–5.*

8. Zhang Z, Yin F, Liu J, Wong DWK, Tan NM, Lee BH, Cheng J and Wong TY, $ORIGA^{light}$: an online retinal fundus image database for glaucoma analysis and research, *Conf Proc IEEE Eng Med Biol Soc, 2010, pp. 3065–8.*

Table of contents

List of figures

List of tables

Nomenclature

AM-FM Amplitude Modulation-Frequency Modulation

AMD Age-related Macular Degeneration

ANN Artificial Neural Network

AODI Automatic Ocular Disease Diagnosis through Biomedical Imaging Informatics

BIF Biologically Inspired Feature

BMII Biomeidcal Imaging Informatics

CAD Compter Aided Diagnosis

CASNET Causal association network

CDR Cup-to-Disc Ratio

CDSS Clinical Decision Support Systems

CNV Choroidal Neovascularization

CSLO Confocal Scanning Laser Ophthalmoscopy

DFP Digital Fundus Photograph

DR Diabetic Retinopathy

FFA Fundus Fluorescein angiography

GRI Glaucoma Risk Index

GWAS Genome-Wide Association Studies

HRT Heidelberg Retina Tomography

ICA Independent Component Analysis

IOP Intra-Ocular Pressure

LASSO Least Absolute Shrinkage Selector Operation

LDA Linear Discriminant Analysis

MGG Mixture of Generalized Gaussian

MKL Multiple Kernel Learning

MRA Moorfields Regression Analysis

MRI Magnetic Resonance Imaging

NCO Neural Canal Opening

NRR Neuro-Retinal Rim

OCT Optical Coherence Tomography

OMIM Online Mendelian Inheritance in Man

PCA Principal Component Analysis

PM Pathological Myopia

PPA Parapapillary atrophy

QCQP Quadratically Constrained Quadratic Programming

QDA Quadratic Discriminant Analysis

RNFL Retina Nerve Fibre Layer

ROI Region Of Interest

SAP Standard Automated Perimetry

SD-OCT Spectral Domain-Optical Coherence Tomography

SEAD Symptomatic Exudate-Associated Derangements

SIFT Scale-invariant feature transform

SILP Semi-Infinite Linear Program

SiMES Singapore Malay Eye Study

SLIC Simple Linear Iterative Clustering

SLP Scanning Laser Polarimetry

SMO Sequential Minimal Optimization

SNP Single-Nucleotide Polymorphism

SNR Signal to Noise Ratio

SS-OCT Swept Source-Optical Coherence Tomography

SVM Support Vector Machines

Abstract

Computer aided diagnosis (CAD) system allows cost effective and prompt disease diagnosis, which has both clinical and social significance. Current ocular CAD systems typically account for only one type of data, e.g. medical image which may yield sub-optimal accuracy as the training data itself lack the complete aspects for decision making. A new challenge in CAD research is to integrate the distinct attributes of clinical research that are provided by different types of biomedical data. By combining heterogeneous data sources, a CAD system would integrate the complementary pieces of information and provide a more holistic appreciation of the multiple risk factors, thus improves disease detection accuracy.

This PhD study aims to fill in the blank by proposing an innovative system AODI (*Automatic Ocular Disease Diagnosis through Biomedical Imaging Informatics*), which focuses on CAD for ocular diseases, aiming to boost the diagnosis accuracy through intelligently combining image, SNP (Single-Nucleotide Polymorphism) and clinical data. AODI enables a data-driven approach that takes advantage of ever-growing heterogeneous data sources and improves the performance when more data or additional information becomes available. We investigate the recent advancements in

kernel learning and deploy multiple kernel learning (MKL) algorithms for AODI. We conduct experiments to predict major ocular diseases including glaucoma, age-related macula disease (AMD), and pathological myopia (PM), using heterogeneous data sets covering image, SNP and clinical data which are obtained from a holistic population study conducted in Singapore. We also perform comprehensive statistical analysis to validate the improvement in the accuracy of predictions and prove the effectiveness of the proposed framework.

To our best knowledge, AODI is the first published work using MKL to integrate multiple kinds of information including image, SNP and clinical data for ocular disease screening/diagnosis. Using MKL, the resulting classifier optimizes the contribution from each sub-kernel through learning an adapted kernel function from each of the heterogeneous feature sets. Such a framework paves a holistic way for automatic and objective disease diagnosis and screening. Moreover, our work on feature selection for SNP data tackles the challenge of SNP selection by innovatively grouping SNPs into functional groups (genes, interacting proteins and biological pathways), and thus explores the biomedical knowledge by sparse learning. Finally, we innovatively incorporates classemes (pre-learned classifiers trained from individual informatics domains) into MKL, and further improves the performance of ocular disease detection.

Chapter 1

Introduction

Computer aided diagnosis (CAD) system allows cost effective and prompt disease diagnosis, which has both clinical and social significance. Current ocular CAD systems typically account for only one type of data, e.g. medical image which may yield sub-optimal performance as the training data itself lack the complete aspects for decision making. In the era of information explosion, data from multiple sources are becoming progressively available. For examples, retinal fundus cameras, an important retinal image acquisition device, can be found in community polyclinics and optical shops; the health screening outreach programs have allowed individuals access the clinical data which was hard-to-access previously; the recent massive reduction in genotyping costs has made it possible to acquire SNP (Single-Nucleotide Polymorphism) data with low cost. A CAD system that combines these heterogeneous data sources would integrate the complementary pieces of information and provide a more holistic appreciation of the multiple risk factors to improves disease detection. Nevertheless, there is no previous work attempting to combine image, SNP and clinical data for ocular disease detection. A possible reason could be that only until recently such data has become available on a large scale. Moreover, researchers working on these heterogeneous data sets usually have different technological background with different foci, e.g. image processing researchers mostly focused on image analysis, and bioinformaticians are

interested in discovering disease associated genetic markers, etc. In such a situation, a new challenge in the advanced CAD systems is to develop an innovative framework to efficiently integrate the distinct attributes of clinical research that are provided by different types of biomedical data.

This PhD study aims to fill in the blank by establishing a computational framework, which integrates and conducts inference from a collection of biomedical data. We propose an innovative system AODI (*Automatic Ocular Disease Diagnosis through Biomedical Imaging Informatics*) which focuses on CAD of ocular diseases and aims to to boost the diagnosis accuracy through intelligently combining various biomedical and imaging data. AODI enables a data-driven approach that takes advantage of ever-growing heterogeneous data sources and improves the performance when more data or additional information becomes available. We investigate the recent advancements in kernel learning and develop multiple kernel learning (MKL) algorithms for AODI. We conduct experiments to validate AODI through predicting major ocular diseases including glaucoma, age-related macula disease (AMD), and pathological myopia (PM), using heterogeneous data sets covering image, SNP and clinical data, obtained from a holistic population study conducted in Singapore. We also performed comprehensive statistical analysis to validate the improvement in the accuracy of predictions to prove the effectiveness of the proposed framework.

The contributions of the thesis are summarized as follows: 1) AODI system paves the holistic way for automatic and objective disease diagnosis and screening. Using MKL, the resulting classifier optimizes the contribution from each sub-kernel through learning an adapted kernel function from each of the heterogeneous feature sets. To our best knowledge, AODI is the first published work using MKL to integrate information of image, SNP and clinical data for ocular disease screening/diagnosis; 2) Our work on feature selection for SNP data tackles the challenge of SNP selection by innova-

tively grouping SNPs into functional groups (genes, interacting proteins and biological pathways), and thus explores biomedical knowledge by sparse learning; 3) To further improve the performance, we develop AODI-MKL^{clm} which incorporates classemes (pre-learned classifiers trained from individual informatics domains) with MKL. and 4) The experiment conducted on a large-scale dataset (2,258 subjects) from a population study demonstrates that AODI-MKL^{clm} is significantly better than CAD using image alone (up to 18.5% improvement of accuracy), or standard SVMs using data from individual domains (5.5% improvement of accuracy) as well as the traditional MKL method of (3% improvement of accuracy).

The rest of this thesis is organized as follows. In chapter 2 we conduct a comprehensive literature review on current CAD system for ocular disease diagnosis. Chapter 3 describes our AODI-PM framework, with experimental result and analysis. Chapter 4 solves a sub problem in the AODI framework on SNP selection. Chapter 5 describes a unified AODI-MKL^{clm} framework incorporates pre-learned classemes trained from individual informatics domains, which further improves the disease detection performance. Finally, chapter 6 concludes the study.

Chapter 2

A survey on computer aided diagnosis for ocular diseases

This chapter reviews ocular CAD systems for various types of data. For each data type we investigate the algorithms to detect different ocular diseases. Their advantages and shortcomings are summarized and discussed. Through the survey we conclude that, while CAD for ocular diseases has shown considerable progress over the past years, the development of fully automatic CAD systems that are able to embed clinical knowledge and integrate heterogeneous data sources still show great potential for future breakthrough.

2.1 Background

Patients with ocular diseases are often unaware of the asymptomatic progression of the said disease [1] until at a later stage when treatment is less effective in preventing vision impairment [2]. Though regular eye screenings enable early detection and timely intervention of such diseases, it would put a significant strain on limited clinical resources. CAD systems, which automate the process of ocular disease detection, are urgently required to alleviate the burden on the clinicians.

Owing to the fast pace of technological advancements in both hardware and software, many CAD systems dedicated to ocular diseases diagnosis have been proposed recently, though most of them are still undergoing evaluation or clinical validation. For example, Fujita *et al.*[3] discussed an emerging CAD system using fundus images for the detection of glaucoma, hypertensive retinopathy as well as diabetic retinopathy (DR). Their project has entered the final stage of development, and commercialized CAD systems ought to appear by its completion.

Though such fully automated systems are not yet on the market, semi-automated and manual computer systems incorporating these CAD systems are relatively widely used, with several clinical publications already reporting on their usage. Examples of the development of such systems include IVAN [4] from University of Wisconsin and more recently SIVA [5] from National University of Singapore for semi-automated vascular analysis. Software packages allowing for processing of data garnered from these systems also exist: ADRES 3.0 [6] is used for the grading of DR and has been commercialized and deployed for use in diabetic centres and general physician clinics in India; the Singapore Eye Research Institute has also been running clinical trials for the diagnosis of several ocular diseases (e.g., pathological myopia (PM), DR and age related macular degeneration (AMD)) using a same set of protocols for ophthalmic image analysis. [1].

This survey covers three types of data for CAD systems: clinical data, image based data and genetic data. Clinical data refers to a patient's demographic information (e.g., age, race etc.) and data acquired from clinical laboratory tests or exams (e.g., intra-ocular pressure) but excludes data acquired from digital imaging or genomic tests, detailed to Section 2.2. Image based data refers to images captured using an imaging device for observing the pathology in the affected part of the eye, details are in Section 2.3.1. Genetic information refers to any data obtained from an individual's

[1]http://www.seri.com.sg/Research%20Professionals/Page.aspx?id=142

DNA, genes or proteins, detailed in Section 2.4. These definitions are specific to this document and may vary depending on context. Of the three, CAD systems using clinical data has been widely studied [7–9]. As far as CAD using genetic information is concerned, recent advancements in genotyping technology have made individual genetic information more commonly available, it is still unfeasible to utilize genetic information for CAD systems on a large scale presently. Perhaps with time, genetic information will gain an essential role in medicine by complementing clinical data with additional genetic interpretations [10]. We cover genetic data as a possible input to future CAD systems. A considerable amount of the survey is focused on the usage of image based data in CAD systems as they are by far the most important type of data in ocular disease diagnosis.

2.2 CAD of Ocular Diseases Based on Clinical Data

One of the pioneer research works on Clinical Decision Support Systems (CDSS), CASNET [11] (causal-associational network), was developed in late 1970s to assist the diagnosis of glaucoma. Clinical data used in CASNET covered symptoms reported by the patient, e.g., 'ocular pain' , 'decreased visual acuity' and various eye examination results, e.g. visual acuity, Intra-Ocular pressure (IOP), anterior chamber depth, angle closure, pupil abnormality and corneal edema. [12]. CASNET used a descriptive disease model to interpreter clinical findings for glaucoma. It represented early medical expert systems, providing a framework that describes the knowledge of experts and simulates various aspects of the clinical cognitive process.

In 2002, Chan *et al.*[13] reported the first implementation of Support Vector Machines (SVM) in glaucoma diagnosis. Clinical data used in the research was the output from a visual field test, Standard Automated Perimetry (SAP). The learning algorithms included SVM, Discriminant Analysis, multilayer perceptron, mixture of

Gaussian etc. The author observed that machine learning methods out performed the best indexes from SAP.

In 2011, Bizios *et al.*[14] conducted a study investigating the use of Artificial Neural Networks (ANNs) on data fusion of SAP parameters and the Retina Nerve Fibre Layer Thickness for diagnosis of glaucoma. The results showed that a fusion of SAP and OCT data yields a accuracy higher than using either of the two alone. This was the first reported study using fused data for glaucoma diagnosis. A recent study [15] defines the risk factors in glaucoma development by study the association of Heidelberg Retina Tomography II (HRTII) structural measurements, central corneal thickness and IOP.

There are a number of large scale or population-based eye studies conducted in various countries. For examples, Blue Mountains Eye Study (Australia) [16], Singapore Malay Eye Study (SiMES) [17], Singapore Indian Eye Study (SINDI) [18], Singapore Chinese Eye Study (SCES) [19] etc. Many research works conducted on various ocular diseases have been published based on the data collected in these eye studies.

2.3 CAD of Ocular Diseases Based on Imaging

Large-scale systematic research and development of CAD from radiology and medical images began in the early 1980s. The first report on retinal image analysis was published in 1973, focusing on vessel segmentation [20]. In 1984, [21] reported an image analysis method for lesion detection related to DR. Over the past 30 years, ophthalmological imaging has paved a way for the development of CAD systems for many ocular diseases (DR [22], AMD [23], glaucoma [24] and cataract [25]). Such diagnostic systems, if used in large-scale screening programs, will offer great cost saving and remove observer bias.

Table 2.1 Imaging Modalities and Diseases to observe

Imaging Modalities	Technology	Targets	Diseases observed
Retina Fundus	2D, larger areas of the fundus compared to what can be seen with handheld ophthalmoscopes	(macular; posterior pole; retina; optic disc) interior surface of eye	DR, glaucoma, AMD
OCT	3D; high resolution cross-sectional imaging	Cornea thickness, retinal nerve fibre layer tissue, macular thickness	Glaucoma, macular degeneration and edema
Heidelberg Retina Tomography (HRT)	2D; confocal scanning laser ophthalmoscope	Retina	Glaucoma
Slit Lamp	2D high-intensity light source stereoscopic magnified view of the eye structures	eyelid, scelra, conjunctiva, iris, lens, cornea	Cataract
RetCam	2D wide angle imaging	anterior segment, anterior chamber	Anterior segment lesions, Retinopathy of Prematurity
Scanning laser polarimetry (SLP)	cross-sectional high resolution	thickness of RNFL	Glaucoma
FFA	2D, angiogram	retinal lesions assessment	AMD, DR

2.3.1 Imaging Modalities

Figure 2.1 shows the anatomy of an eye. Various medical imaging devices have been developed to capture the different parts of the eye, observing various pathological signs. Table 2.1 lists the anatomical structure(s) and the associated disease(s) each imaging modality is able to observe.

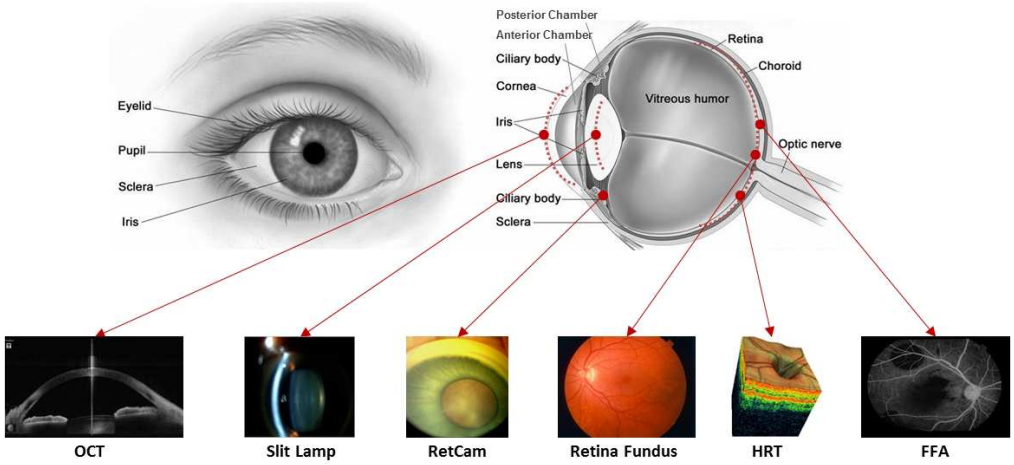


Fig. 2.1 Ocular Anatomy and various image modalities

We obtain the statistics of research works done on major ocular image modalities by searching the *IEEEXplore* publication database. Figure 2.2(a) shows the number of publications related to various ocular imaging modalities, while Figure 2.2(b) shows the number of publications on CAD for ocular diseases using retinal images. The observation is further substantiated by a distribution of the works surveyed in this chapter, in Table 2.2, wherein the works are arranged according to the disease and the associated imaging modality.

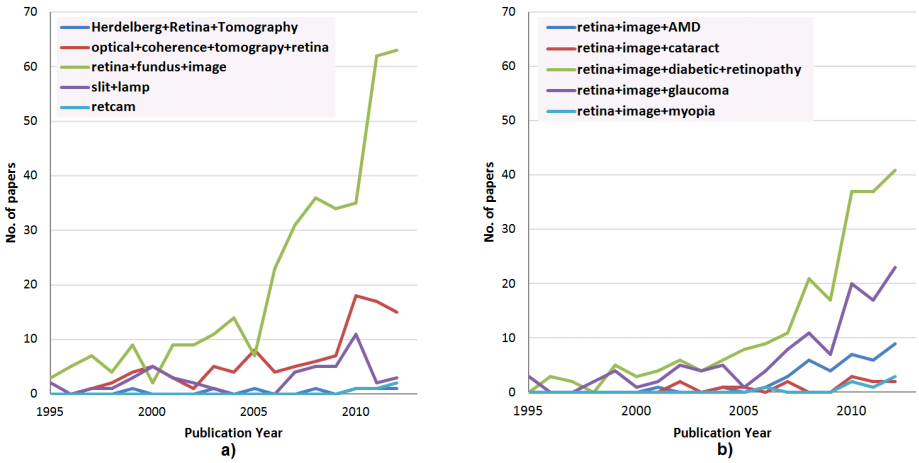


Fig. 2.2 Publication trends for ocular disease detection. (a) Number of publications each year for different ocular imaging modality (b)Number of publications each year for different ocular disease detection using retinal image (queries to IEEEXplore are as on May 2013)

To date, the most studied disease is DR, followed by glaucoma and AMD, as shown in Figure 2.2(a). Among all the imaging modalities, Digital Fundus Photograph (DFP) has been attracting intensive interest, as information extracted from fundus are useful in detecting a number of diseases including stroke, hypertension, heart disorders, peripheral vascular disease and DR [155]. Moreover, the availability of inexpensive fundus imaging cameras makes eye examination simple and cost effective. Another modality which is gaining interest in the research community is OCT. First proposed in 1991 [156], OCT has been widely applied to medical imaging especially for imaging the eye. The most important advantage of OCT compared with DFP is that it provides

Table 2.2 A distribution of works on CAD of major ocular diseases based on imaging.

Modality	AMD	Cataract	DR	Glaucoma	PM
OCT	[23] [26]		[27]	[28][29][30] [34][35][36] [39]	[31] [32] [33] [37] [38]
Slit Lamp		[40] [41] [42] [25] [43] [44] [45] [46] [47] [48]			
SLP				[49][50][51][52] [52]	
Retina Fundus	[53] [54] [55] [64] [65] [66] [75] [76] [77] [84] [85] [86] [93] [94] [95] [102] [103] [110] [111] [112] [118] [119] [126] [127] [133] [134] [135] [138] [95]		[56] [57] [58] [67] [68] [69] [78] [79] [80] [87] [88] [89] [96] [97] [98] [104] [105] [106] [113] [114] [115] [120] [121] [122] [128] [129] [130] [87] [136] [137] [139] [140] [141] [142] [143] [144] [145] [146] [147] [148] [149] [150]	[24] [59] [60] [70] [71] [72] [81] [82] [83] [90][91][92] [99][100][101] [107][108][109] [116][117][70] [123][124][125] [131][132]	[61] [62] [63] [73] [61] [74]
HRT				[151][152][153] [154]	

quantifiable depth information enabling a 3D scan of the target part. Therefore it is possible to detect pathologies with topological changes in-vivo. Although a powerful tool [157], in early years, the progress of OCT-based ocular disease detection has been constrained by the speed of OCT imaging. Early version of OCT required lengthy amounts of time to capture an image. In recent years, with the advancement of SD-OCT(spectral domain OCT), which needs only 6 seconds to take a high resolution image, OCT-based ocular disease detection methods are increasing in popularity [158].

The images associated with the above mentioned modalities often need preprocessing to remove noise and improve contrast before they can be analyzed further using CAD methods.

2.3.1.1 Image Preprocessing

Some of the common preprocessing methods are histogram equalization [69, 78], shade correction [79, 80, 87], convolution with a Gaussian mask [88], median filtering [89] and blood vessel removal [96] [97].

Most of the contrast enhancement techniques use histogram equalization [69, 78]. Shade correction is often used to normalize illumination [79, 80] [87]. For noise re-

duction, the commonly used techniques are convoluting with a Gaussian mask [88] or using a median filter [89]. Some of the methods also use blood vessel removal as a preprocessing step since they can be detected as false positives while detecting red lesions, especially MAs [96, 97].

The choice of a suitable preprocessing method depends on the desired effect. [98] experimentally showed that contrast limited adaptive histogram equalization [104] effectively improves local contrast but also introduces noise. Similarly, vessel removal is used to reduce false positives which can be found during red lesion detection. Considering this subjective nature of the preprocessing methods, [98] proposed to choose the best pair of preprocessing and segmentation methods through a fusion algorithm.

The remaining part of this section surveys the works on detecting the major ocular diseases, e.g. DR, PM, AMD and glaucoma. Though DFP is still the main stream modality, OCT is rapidly gaining widespread adoption. Therefore we focus on these two modalities.

2.3.2 Algorithms for image based ocular disease detection

2.3.2.1 Diabetic Retinopathy

As a side effect of diabetes, DR occurs when the blood vessels in the eye get blocked due to high sugar content in the blood [159]. Lesions appearing on the retinal surface are visible in a DFP. Figure 2.3 (a) and 2.3 (b) show the DFPs of a normal eye and a DR affected eye, respectively. DR-related lesions can be categorized into **red lesions**, including Haemorrhages and Microaneurysms (MA), and **bright lesions** and cotton-wool spots, shown in Figure 2.3 (c). There are several works [137] which detect other symptoms as well.

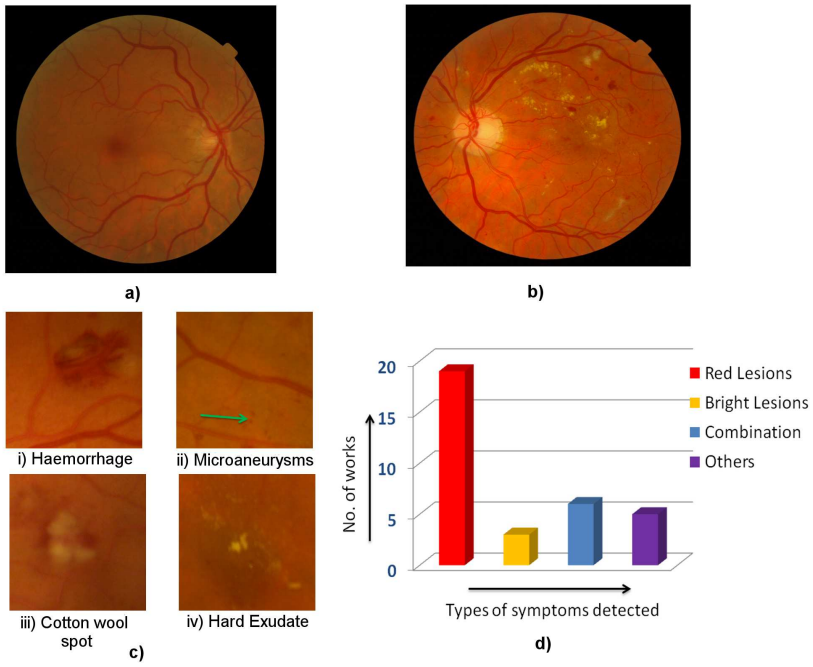


Fig. 2.3 DR symptoms observed by DFP (a) DFP of a normal eye (b) DFP of an eye affected with DR (c) Common lesions associated with DR (d) A distribution showing number of works detecting each type of symptom.

DFP for DR Detection Detection of DR using DFP typically involves four steps 1) Preprocessing to enhance lesions, 2) Segmentation of candidate lesions, 3) Feature extraction from candidate lesions 4) Classifying candidate lesions into normal and lesions. The green channel of the DFP is preferred for analysis since since in which the retina has a strong contrast [89].

Segmentation is usually based on morphological operations [105, 160]. Lay and Baudoin *et al.*[105] were among the first to propose automatic segmentation of MAs. They performed morphological opening of images using structuring elements of different orientations and subtracted the resultant image from the original one, though it is hard to choose an optimal size of the structuring element [88].

Apart from morphological approaches, researchers have used Gabor filters [121], Gaussian correlation filters [122], curvelet transforms [96], wavelet transforms [161],

local image properties [162, 163], or just the intensity values in the green channel [88, 128] for extracting candidate lesions.

Some of the works detected both bright and red lesions [97, 128, 140, 144, 145] while [146] and [137] attempted to detect neovascularization in addition to the lesions. Individual detections were then fused in these works to predict the severity of DR.

OCT Imaging for Detecting DR Other than DFPs, OCT images can also be used for DR detection. An OCT image can analyze different retina layers and is able to detect cystoid fluids. Wilkins *et al.*[27] proposed to detect Cystoid Macular Edema which is one of the symptoms of DR. They presented a method for segmenting retinal cyst without going further for DR detection. A drawback with the OCT images is that they are prone to noise during capture and a poor Signal to Noise Ratio (SNR) can affect the segmentation accuracy [27].

Most of the DR related CAD research has focused on detecting lesions associated with DR. Few works [147] have gone further to convert lesion detection to DR detection. Even for DR detection, most of the works presented their results as a binary detection, i.e whether DR is present or not in an eye. It might be useful to provide a grade to the severity of DR. Recent works [148] have attempted to bypass the lesion detection and used non-clinical features, as described in Section 2.3.3 for DR detection. Future research can focus on filling these gaps.

2.3.2.2 Glaucoma

Glaucoma is defined by the structural changes of the optic nerve head, causing by progressive degeneration of optic nerve fibres. Due to its asymptomatic nature in early stage, glaucoma has the name of 'the silent thief of sight'. Although glaucoma cannot be cured, timely diagnosis is critical for disease control [164, 165]. Glaucoma diagnosis usually takes into consideration of a patient's IOP, medical history, visual field loss

tests, and ophthalmoscopy assessment on OD. In 2D images, the OD contains two separate zones: the bright one in the center is the optic cup, and the peripheral region is the neuroretinal rim [81]. Glaucoma causes the cup enlargement with respect to OD, which is one of the essential clinical indicators. Various parameters related to cupping have been used in glaucoma detection, including CDR (vertical cup to disc ratio) [166], disc diameter [167, 168], ISNT rule [169], peripapillary atrophy (PPA) [170] and notching [171]. The most popular measurement is CDR, which is computed as the ratio of the vertical cup diameter to vertical disc diameter clinically, shown in Figure 2.4. There are four imaging modalities measure parameters of the ONH for glaucoma

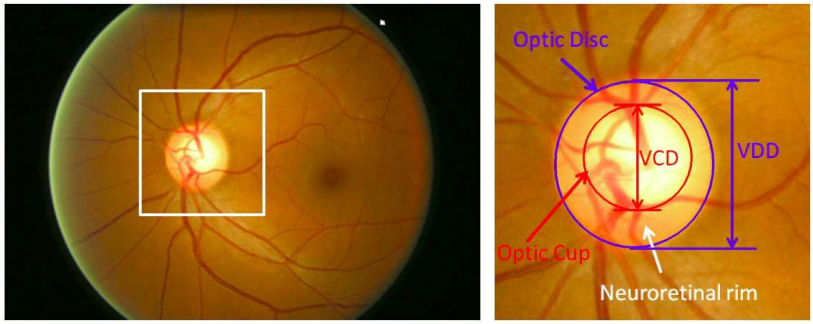


Fig. 2.4 Major structures of the optic disc in DFP. The region enclosed by the blue line is the optic disc; the central bright zone enclosed by the red line is the optic cup; and the region between the red and blue lines is the neuroretinal rim.

assessment, namely, DFP; OCT; CSLO (Confocal Scanning Laser Ophthalmoscopy) and SLP (Scanning Laser Polarimetry).

DFP for Glaucoma Detection DFP is one of the main and popular modalities to diagnose glaucoma. Acquired noninvasively, DFP is a suitable modality for large-scale glaucoma screening where a CAD system detects signs of suspicious of glaucoma present in an image. Then, only those suspicious images will be passed to ophthalmologists for further examination.

Glaucoma detection based on DFP falls into three strategies: 1) detection with no disc parametrization, 2) detection with disc parametrization in stereo DFP, and 3) detection with disc parametrization with monocular DFP.

For detecting glaucoma without the disc parametrization, a set of features are computed at the image-level without cup and OD segmentation. Then, a classifier determines an image as glaucomatous or not. Bock et al.[81] reported an glaucoma detection system using appearance-based dimension reduction technique to compress different generic feature types , then a probabilistic classifier combined these features types to get the Glaucoma Risk Index(GRI). Several other papers [71, 72, 90–92, 99] have also adopted this strategy for glaucoma detection.

For the other two strategies, cup and OD are segmented to estimate the relevant disc parameters. The strategy based on monocular DFP utilizes the 2-D projection of retinal structures to compute the areas of OD and cup. As shown in Figure 2.4, in a monocular DFP, OD is a elliptic region partially obstructed by blood vessels. After segmenting the OD and cup [83, 100, 107], vertical CDR is estimated to detect glaucoma [70, 108, 109, 116, 117]. In a recent work [108], Cheng *et al.*developed a glaucoma screening system based on OD and cup segmentation using superpixel classification. The authors used histograms and centre surround statistics to classify each superpixel for OD segmentation. In addition, the location information was added to the feature space for the cup segmentation.

Different from monocular DFP, a stereo set of DFP contains partial depth information, which can be used to characterize the inner region of OD, e.g., the cup and rim. A number of works based on stereo DFP have been carried out to detect glaucoma[101, 123–125, 131, 132]. For example, [124] reported an method on stereo color photographs segmentation, to detect the OD, cup and rim using pixel feature

classification. In their system, the feature space is formed by a depth map, combined with the outputs from a Gaussian steerable filter bank.

OCT Imaging for Detecting Glaucoma OCT is relatively new compared to fundus photography, so is the history of OCT-based image analysis techniques, which, is growing fast as OCT is becoming an important modality for glaucoma detection. While stereo fundus photography is able to extract limited information regarding the 3D shape of the OD nerve head, OCT provides true 3D image. Figure 2.5 gives three spectral-domain OCT images in glaucoma [34]. There are mainly two categories of disc/cup segmentation from OCT images for glaucoma detection[172]: 1) rely on pixel classification which utilize the depth-columns of OCT voxels, with the reference defined by manual planimetry from stereo fundus photographs and 2) a graph theoretic approach to segment neural canal opening and cup structures directly from 3D OCT images

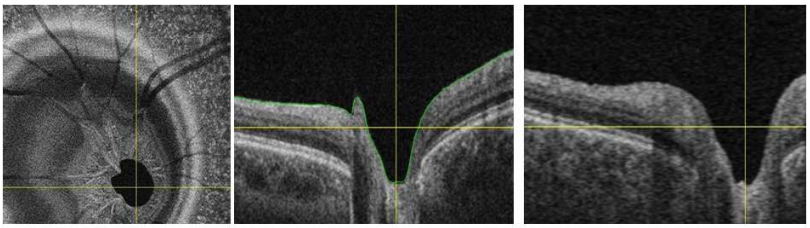


Fig. 2.5 SD-OCT: Cross-sectional images showing OCT volume. a) volume in image plane; b) image showing the horizontal line of a); c) image showing the vertical line of a)

For the first strategy of segmenting ONH, a series of studies [34–36] have been performed. Lee et al.[35] proposed an algorithm segmenting the OD, cup and rim in SD–OCT scans centered on the optic nerve head. Their system first segmented intra–retinal surfaces with a fast multiscale 3D graph search method; then the retina volume was flattened for features derived from OCT voxel intensities and intra-retinal

surfaces. As a further study, [34] presented a fully automatic method for OCT-based OD, cup and rim segmentation. The method performed automated planimetry from close-to-isotropic SD-OCT scans directly.

For the second strategy of segmenting ONH, a variety of studies[28–30, 39] directly segment the neural canal opening (NCO) as well as the cup from 3D OCT images. Hu et al.[28] introduced a scheme for segmenting the optic disc margin of ONH in SD-OCT images using a graph-theoretic approach. They created planar 2D projection images using a small number of slices surrounding the Bruch's Membrane Opening plane. In addition, since there are large vessels in images, the information from the vessels segmentation was used to suppress the vasculature influence to reduce the segmentation difficulty. In order to study the association between the NCO-based metrics and the clinical disc margin, Hu et al.[30] proposed an approach for NCO and cup segmentation at the level of the Retinal Pigment Epithelium and Bruch's Membrane complex in SD-OCT volumes.

CSLO Imaging for Detecting Glaucoma CSLO produces quantitative measurements of the ONH and posterior segment by using a diode-laser light source. A commercially available CSLO device is the Heidelberg Retina Tomograph (HRT) that is able to detect the structural alterations in glaucoma. An example of an HRT image is shown in Figure 2.6 [81]. Numerous studies [151–154] have reported that HRT measurements are highly reproducible. In [153, 154], after outlining the optic disc border manually, the system generated geometric parameters such as the volume, depth and shape of the cup, and retinal height variations along the rim contour. The authors applied discriminant analysis to combine these parameters into a quantitative parameters. Due to manual outlining of the OD border, the gained parameters are not fully objective. Burgansky-Eliash et al.[151] utilized a non–linear shape model on the topographic ONH shape for glaucoma classification, which overcame the subjectivity

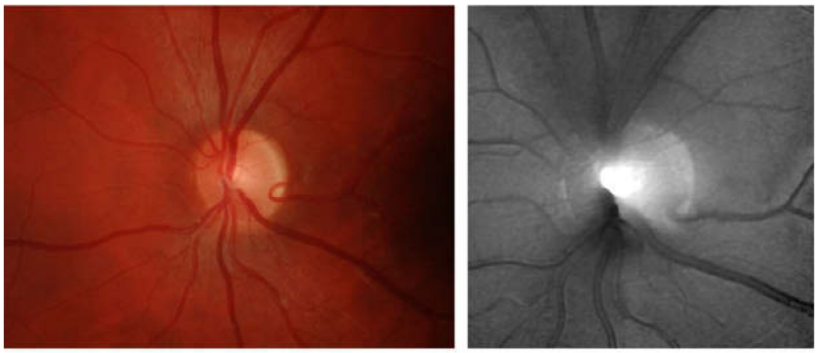


Fig. 2.6 Central retina image. ONH-centered DFP is used for glaucoma detection by glaucoma risk index; and HRT 2.5-dimensional topography images are used for glaucoma probability score detection

of contour based methods. The work reported in [152], quantified the progression of glaucomatous degeneration over years by using the HRT Topographic Change Analysis to automatically locate and quantify the temporal glaucomatous structural ONH changes.

SLP Imaging for Detecting Glaucoma SLP is another available imaging modality for glaucoma detection. The structural change of the ONH in the course of glaucoma leads to nerve fibres degeneration, yielding a thinning of the Retina Nerve Fibre Layer (RNFL). SLP is able to measure the thickness of the RNFL for glaucoma detection. In SLP image, the retina is illuminated by polarized light, and thus by detecting the polarization change of the reflected light, the thickness of RNFL can be obtained [49].

SLP is commercialized by Carl Zeiss Meditec (Dublin, CA) as the GDxVCC, which comprises a scanner with a software program assisting the acquisition procedure to analyze the scan and derive various parameters to generate the Nerve Fiber Indicator. The software can be considered as a soft classification of glaucoma likelihood images generated by the GDx VCC are shown in Figure 2.7 [173]. From SLP data many

glaucomatous progression detection strategies can be formulated. Based on repeated GDxVCC SLP measurements, Vermeer et al.[51] tested several strategies to determine the optimal one for clinical use. Medeiros et al.[52] presented a scheme for differentiating between glaucomatous and control cases, which extracted global and sectoral geometric parameters related to RNFL thickness.

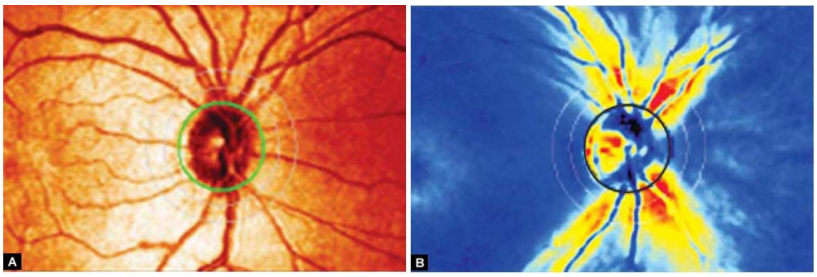


Fig. 2.7 GDx VCC images. (A) The reflectance image displayed as a colored intensity map (greater reflectance = lighter color). (B) The retardation map converted to RNFL thickness

Utilizing DFP and OCT to detect glaucoma are two popular and active directions with OCT having a shorter history. Till now, time-domain OCT and SD-OCT have been widely utilized to perform glaucoma detection [28–30, 34–36, 39]. However, swept-source OCT (SS-OCT) has not been further exploited for the research of glaucoma. For DFP, the combined analysis of stereo DFP and OCT for extracting disc parameters may boost current performance of state-of-the-art algorithms.

2.3.2.3 Age-related Macular Degeneration

AMD is a deterioration of the eye's macula. It causes vision loss at the central region, and blur or distortion at the peripheral region, as shown in Figure 2.8. Depending on the presence of exudates, AMD is classified into dry AMD (non-exudative AMD) and wet AMD (exudative AMD), with drusen and exudation as major symptoms respectively [174]. AMD can be detected from DFP, OCT, X-ray, and Magnetic Resonance Imaging (MRI). Among them, DFP is perhaps the most widely used one for AMD de-

(a) The view from a healthy eye (b) The view from an eye suffering AMD

Fig. 2.8 Vision damage caused by AMD. Image showing views from a normal eye and an eye affected with AMD.

tection, while OCT is rapidly growing in use. Most of the approaches detecting AMD from DFPs focus on detecting drusen using local thresholding[53, 55], wavelets [53], background modeling[85] and saliency [93] etc. Some of the works have also attempted to bypass drusen detection and directly predict AMD [102, 103, 110, 111, 118, 119, 175]. On the other hand, it is easier to observe exudates and edema in OCT images. OCT can segment out retinal layers, and texture and thickness of these layers can help in separating normal region from region corresponding to exudates[23, 26].

DFP for Detecting AMD The existing automatic AMD detection methods focus mainly on detecting drusen, the symptom of early AMD. Several other methods walk a step further to grade AMD. In DFPs, drusen appear as small bright spot with particular size and orientation, as shown in Figure 2.9 (b). Because the intensity and color of the image may vary with different imaging condition, finding local maxima is a more effective method than global thresholding is. Local maxima are found through geodesic method [53], Histogram based Adaptive Local Thresholding [55], and Otsu method based adaptive threshold [64]. After maxima detection, the candidates are further classified according to contrast, size and shape. Apart from spatial domain, frequency domain has also been used for drusen detection. For example, multi-scale

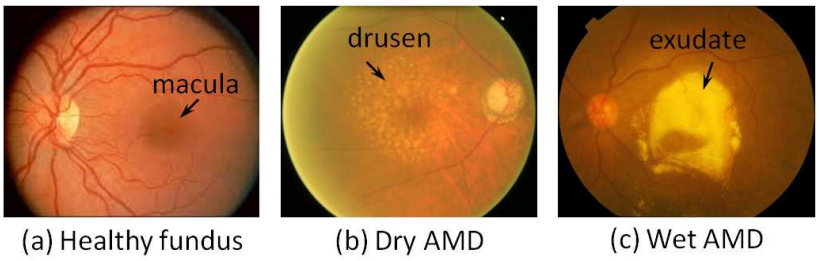


Fig. 2.9 The symptoms of AMD seen in DFP. DFP of a healthy eye and an eye affected with dry and wet AMD. Presence of drusen and exudates can be seen.

and multi-orientation wavelet is used to detect drusen in a hierarchical framework [53] or through Support Vector Data Description [176], which is derived from support vector machine [66]. Furthermore, AM-FM (amplitude-modulation frequency modulation), an mathematical technique, has been employed to extract multi-scale features for pathological structure (e.g. drusen) classification on a retinal image [75].

In recent years, with the rapid progress of computer vision and machine learning, many advanced techniques have been introduced for drusen detection, e.g., novel feature descriptor such as ICA [76] and biologically inspired features [66], feature selection schemes such as AdaBoost [77], and parameter choosing approaches [54]. A latest work, Thalia [177] is a system for drusen lesion image detection and AMD assessment by using a hierarchical word transform (HWI).

There are other methods using background modeling [85] and saliency [93]. The background modeling method [85] first segments the healthy structure of eye and blood vessels and the inverse of the healthy parts provide the drusen detection result. The saliency based method [93] first detects the salient regions and then classifies them as blood vessel, hard exudates or drusen. In [86], a general framework was proposed to detect and characterize target lesions concurrently. In the framework, a feature space, including the confounders of both true positive (e.g., drusen near to other drusen) and false positive samples (e.g., blood vessels), is automatically derived from a set of reference image samples. Subsequently a Haar filter was used to build the

transformation space and Principal Component Analysis (PCA) was used to generate the optimal filter.

Since drusen is one of the main early symptoms of AMD, most of the existing work on AMD detection take drusen detection and segmentation as basis. The overlap of drusen with macular is used to measure the severity of AMD [94, 95]. The performance of such methods is restricted by the accuracy of drusen detection. To bypass drusen detection and segmentation, in recent years, researchers have started to seek for methods detecting AMD directly from DFPs. An early work along this direction was a histogram based representation followed by Case-Based Reasoning [102]. Though the results were satisfactory, the observations indicated that relying on the retinal image color distribution alone was insufficient. Thus the authors upgraded the method by using a spatial histogram technique that included color and spatial information [103]. The latest work from the same team comprises hierarchical image decomposition and applied a weighted frequent sub-tree mining algorithm to identify sub-graphs, which were incorporated into a feature vector representation (one vector per image) for classification [110, 111]. These methods detect AMD from the scope of a single image. Another strategy is to use content-based image retrieval. Region based and lesion based features were tested and obtained satisfactory performance in [118] and [119].

The above mentioned works detect dry (non-exudate) AMD. Till now, there are few works on wet AMD detection except the one proposed in [112] where the basic idea is that the vessels in the DFP seem different under dry and wet AMD. Thus the method first detected the vessels, using a wavelet based method. Subsequently the area, standard deviation, and other features describing the distribution of the vessels were used as features for classification.

OCT imaging for Detecting AMD As mentioned in Section 2.3.1, it is easier to observe edema and exudates in OCT. [23] reports the method on characterizing

normal macular appearance in SD-OCT volumes to detect local retinal abnormality. The approach characterizes texture and thickness properties across the macular using features extracted from each of the 10 intraretinal layers segmented automatically. In [26], the authors improved the approach by employing a probabilistically constrained combined graph search-graph cut method by integrating the candidate volumes into the graph cut cost function as probability constraints.

The aforementioned works show that although OCT imaging is increasingly prevalent, DFP is the major image modality for AMD detection and screening. However with the progress of SD-OCT, OCT based AMD detection and screening is emerging as a new area of focus.

2.3.2.4 Pathological Myopia

Caused by posterior staphyloma and deficient corrected acuity, PM is a type of severe and progressive nearsightedness characterized by changes in the fundus of the eye. PM is primarily a genetic condition [178] though both genetic and environmental factors have shown associations with its onset and progression [179]. Different from myopia which is caused by the lengthening of the eyeball, PM is coupled with the degeneration of retina, which can lead to blindness without treatment. The accurate detection of PM will facilitate timely intervention, better disease control to slow down the progression of the disease.

PM has been detected mostly from DFPs where retinal degeneration is observed in the form of PPA [180, 181], the thinning of retinal layers around the optic nerve and is characterized by a pigmented ring like structure around the optic disc. Apart from DFPs, there have been studies to detect PM from OCT images [182] however CAD systems for detecting PM from OCT images have not emerged yet.

DFP for Detecting PM An observable sign for PM detection is PPA, an atrophy of pre -existing retina tissue. The PAMELA system developed by Liu *et al.*[61] was the first CAD system for PM detection. In PAMELA, features were extracted from a sectional texture map by entropy analysis in the optic disc ROI, and SVM learning achieved a 85% specificity and 90% sensitivity. Later on, Tan *et al.*[62] reported a PPA detection method using a variational level set approach. The method used a disc difference approach to locate PPA by obtain a difference in the two areas, e.g., optic disc with PPA and the fundamental optic disc. It reported a 95% accuracy. The above two methods were based on a rather small data set of only 40 images. A recent advance in PPA detection was reported in [63], which was tested on a much larger dataset containing 1584 images. The authors presented a method using biologically inspired feature (BIF) which mimics the cortical processes for visual perception, to detect PPA. The approach first segments a focal region (ROI) from DFP, then extracts the BIF by selective pair-wise discriminant analysis for sparse learning. The authors reported that negative sparse transfer learning outperforms the positive one and achieves an accuracy of more than 90% on PPA detection.

Different features have been extracted from DFP for PM detection. APAMEA extracted a texture feature obtained through entropy analysis. In [63] BIF was used for sparse learning. The study conducted by Zhang *et al.*[73] developed a combined approach integrating SIFT features extracted from DFP with genetic information as well as other clinical data. The study demonstrated that, by learning from multiple data sources, the classifier can achieve a more accurate prediction result. It is the first reported study to combine heterogeneous data including image, genetic and text data for PM detection.

SS-OCT Imaging for Detecting PM SS-OCT uses a frequency swept laser as a light source [31] and has less roll-off of sensitivity with tissue depth compared with

conventional SD-OCT instruments. SS-OCT instruments use a longer wavelength (usually in the $1\mu m$ range) that has improved their ability to penetrate deeper into tissues [183] than SD-OCT. Though CAD systems based on SS-OCT have not emerged, some clinical studies have discovered that SS-OCT could be a powerful machine for PM analysis. A recent study conducted in Japan [38] reported that SS-OCT can detect optic nerve pits or pit-like changes in PM eyes. [37] analyzed the association between the sclera shape and the myopic retinochoroidal lesions, and concluded that SS-OCT offers important information on deformations of the sclera which are related to myopic fundus lesions. Such clinical discoveries provide strong evidences for the use of SS-OCT as a good candidate for future PM-CAD development.

2.3.3 Summary on image based ocular CAD

Feature extraction plays an essential role in ocular image based CAD systems. From the survey, we observe two broad classes of features used in the ocular CAD systems. Approaches using either of them are described below:

2.3.3.1 Approaches Using Clinical Features

Many of the retinal image based CAD systems employ clinical domain knowledge during the feature selection and decision making processes. Such systems focus on identifying disease associated landmarks from images. A number of clinically relevant features can be extracted from the identified landmarks. For example, the following image cues are highly related to glaucoma: large optic CDR [184]; appearance of optic Disc haemorrhage [185]; thinning of the neuroretinal rim (NRR) or notching of the NRR [171] and presence of PPA [170]. These features based on clinical knowledge can be described as clinical features.

The early efforts in retinal image analysis focused on optic disc localization. [186] used specialized template matching to locate optic disc, followed by a global elliptical and local deformable contour model for disc segmentation. [123] presented a deformable-model-based algorithm for the OD boundary detection in DFP. Later efforts spent on optic cup detection. [124] analyzed stereo-based DFPs for rim and cup segmentation via pixel feature classification. [187] detected the optic cup using vessel kinking analysis. [188] proposed method based on depth discontinuity to estimate the cup boundary. After cup and disc detection CDR can be obtained, based on which CAD systems for automatic glaucoma detection were developed [24, 59, 60, 70]. Cheng *et al.*[63, 189] developed PPA detection algorithms for Pathological Myopia (PM) detection. Liang *et al.*[95] focused on detecting drusen presented in retina for automatic AMD detection. Other researchers worked on CAD systems for DR based on various vasculature segmentation algorithms, e.g., matched filters [56, 57], vessel tracking [58] or morphological processing [67, 68].

The advantages of using clinical features in CAD systems are: the CAD results can be interpreted and presented with clinical meaning, furthermore, the prior knowledge allows modeling the disease detection with a small data set, which is critical when the training data is insufficient.

However, the detection models built using clinical features have a number of limitations:

- The modeling process is either localization or segmentation dependent. For example, [24, 59] detect glaucoma based on optic cup and disc segmentation, a small error in disc localization may propagate downstream and finally yield an error in detection.

- A model built upon prior knowledge may not evolve with the growing available data.

- As different diseases may possess different landmark features, the system developed for one disease may not be adaptable for other diseases. For example, CDR–based glaucoma diagnosis may not be suitable for detection of other ocular diseases.

- Such systems usually need to learn from manually curated ground truth images, which is not only time consuming but also prone to human error.

- Finally and most importantly, detection of one particular disease associated landmark may be neither necessary nor sufficient condition for disease detection. For example, [61, 74] proposed to recognize PM based on PPA detection, however, having PPA does not imply having PM.

Detecting all the retinal changes in DFPs is much more difficult compared to detecting a particular landmark. Statistical learning based on image feature extraction can be a possible solution to address these challenges. The following section casts light on this possibility.

2.3.3.2 Approaches Using Image Features not Related to Clinical Diagnosis

With the increasing availability of image databases and advances in statistical learning, new CAD systems are shifting to approaches using image features not used in or related to clinical diagnosis, e.g., Common image features relate to the *content* of the image such as color, texture and gradient.

Many image feature extraction techniques can be applied to retinal image based CAD systems. Bock *et al.*[71] used an appearance based approach to quantitatively generate a glaucomatic risk index from retina images. Cheng *et al.*[82] used Focal Biologically Inspired Feature for glaucoma type classification. Wang *et al.*[190] presented

a DFP mosaic algorithm based on Scale-Invariant Feature Transform (SIFT) feature [191] to overcome geometric distortion between different fields of view of DFPs as well as low contrast issue. Extracted SIFT features were described using vectors to detect the matching feature point pairs between two images. The approach then computes the transformation matrix according to the purified matching point and generate a panoramic picture which improves their CAD systems. Xu et al [180] presented a CAD system for PM detection based on SIFT features extracted from a DFP. The system achieved a high AUC value (98.4%) compared with the earlier approaches to detect PM using particular image cues [74].

Another example is the use of superpixels [192, 193]. A superpixel is a perceptually consistent unit with all pixels in a group being similar in color and texture. It reduces the complexity of images from thousands of pixels to only a few hundred superpixels. Algorithms such as Simple Linear Iterative Clustering (SLIC) [194] have been developed to aggregate nearby pixels into superpixels whose boundaries closely match true image boundaries. Many features can be computed from superpixels such as shape, color, location and texture, and they are applicable for subsequent classification [83]. presented a superpixel based learning framework based on retinal structure priors for glaucoma detection. The use of superpixels results a more effective and descriptive representation as compare to pixel-based techniques, while requires shorter computing time than what needed in methods based on sliding windows.

Non-clinical features can be considered to be associated with a data driven approach, which has shown many advantages over the approach using clinical features. Extracting non-clinical features is followed by learning from the labeled examples, therefore fewer manual ground truth labeling is needed as compared to the approaches using clinical features. As these systems do not rely on particular image landmarks, they avoid the error cascading due to initial segmentation or localization. Non-clinical

features are generalized features which make it possible for the system to transfer knowledge learned from one disease to other diseases. Such feature extraction can facilitate learning algorithms such as multi-task learning [195, 196] and transfer learning [197]. Furthermore, since the techniques apply statistical evaluation, the performance of the systems is expected to improve when more data is available. The result of such systems can be a quantifiable score other than Yes or No, which is particularly useful in clinical assessment. The use of non-clinical features for CAD is a promising area for future CAD systems.

2.4 Predicting Ocular Diseases Based on Genetic Information

Genetic information can be used to predict heritable disease related genotypes, mutations or phenotypes for clinical purposes [198]. Ocular diseases are highly inheritable and thus the genetic information can provide valuable insights for disease risk and disease prognosis.

2.4.1 Heritability of Ocular Diseases

Heritability is the proportion of phenotypic variation in a population that is attributable to genetic variation among individuals [199]. The observable characteristics of an organism, or 'Phenotype' [200], can be represented as a linear function of genetic and environmental factors: $Phenotype(P) = Genotype(G) + Environment(E)$. The heritability can be represented as $H^2 = G/P$ where H^2 represents the heritability due to all genetic effects. Since the beginning of the 20th century, heritability studies have been conducted on numerous diverse biological and psychological human traits. Among these, attempts have been made to estimate the genetic contribution to human

Table 2.3 Heritability for ocular diseases or disease related traits

Disease / Traits	Heritability Value	Source
AMD	0.7	[206]
AMD	0.75	[207]
AMD	0.71	[208]
AMD	0.46-0.71	[209]
AMD	0.45	[210]
AMD (small hard drusen)	0.63	[211]
CCT	0.95	[212]
CCT	0.72	[213]
CDR	0.48	[214]
CDR	0.66	[213]
Corneal astigmatism	0.6	[215]
Corneal curvature	0.71	[215]
Cortical cataract	0.24	[216]
Cortical cataract	0.58	[217]
Glaucoma	0.63	[218]
Glaucoma	0.7	[219]
Glaucoma (shallow anterior chamber)	0.92	[220]
Hyperopia	0.75	[221]
Hyperopia	0.86-0.89	[217]
IOP	0.47-0.51	[222]
IOP	0.3	[223]
IOP	0.36	[214]
IOP	0.56-0.64	[224]
Noncongenital cataract	0.15-0.32	[225]
Nuclear cataract	0.356	[216]
Nuclear cataract	0.48	[226]
Ocular refraction	0.89-0.94	[227]
Pathological Myopia	0.306	[228]
Pathological Myopia	0.8	[229]

longevity and lifespan [201, 202], and a person's susceptibility to becoming a smoker [203, 204].

In 1992, the first ophthalmic twin study was conducted to investigate the heredity of refractive error [205]. Since then, over 100 articles have been published in the scientific literature examining the genetic contribution to variation in ophthalmic traits. Table 2.3 summarizes the heritability of various ocular diseases or ocular related phenotypes as reported in the literature. It is observed that the heritability values reported in different studies vary from each other, because the value is population related.

The range of heritability values are shown in Figure 2.10, from which it is observed that Central Corneal Thickness is the most heritable trait while PM spans a wider range due to its population dependence, and cataract seems a less heritable disease.

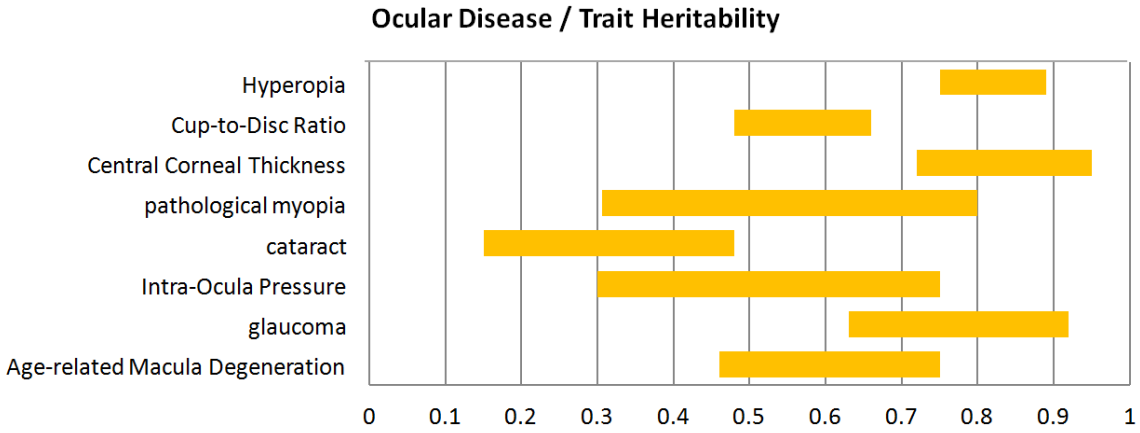


Fig. 2.10 Heritability for various ocular traits. (The range of heritability values for different ocular traits. A higher heritability value means a higher change of inheriting the trait)

2.4.2 Knowledgebases of Genetic Markers for Ocular Diseases

Over the past 20 years, biomedical research community has spent huge efforts in identifying genetic markers for heritable diseases, through classical linkage studies [230] or recent Genome-wide association studies [231]. The disease related biomarkers include genes, genetic mutations and SNPs have been continuously accumulated in various biomedical databases, which are usually called *knowledgebases*. This section introduces the knowledgebases highly relevant to this study.

- OMIM - Online Mendelian Inheritance in Man

  OMIM is a catalog of human genes and genetic disorders and traits. It particularly focus on the molecular relationship between genetic variation and phenotypic expression [232]. It is thus considered to be a phenotypic companion to the

Human Genome Project [233]. On 8 May 2013, it has more than 14,000 disease related gene entries in stock.

- GWAS Catalogue - Catalogue of Published Genome-Wide Association Studies (GWAS)

  GWAS is an approach to rapidly scan markers across the complete sets of genome (DNA) of many people to find genetic variations associated with a particular disease [234]. The first GWAS published in 2005 [235] was associated with an ocular disease. It investigated AMD and found two SNPs that are significantly associated with AMD . Since then, similar successes have been reported using GWAS to identify genetic variations that contribute to risk of type 1 diabetes [236], Parkinson's disease [237], heart disorders [238], obesity [239] etc. The GWAS Catalogue http://www.genome.gov/gwastudies/ is a collection of GWAS discovered SNPs, hosted by NHGRI (National Human Genome Research Institute) . SNP-trait associations listed in the GWAS Catalogue are limited to those with $p - values < 1.0 \times 10^{-5}$. As on 8 May 2013, the catalog includes 1594 humane GWA studies which examined over 200 diseases and identified more than 10,000 disease associated SNPs.

2.4.2.1 Ocular disease related SNPs

Figure 2.11 shows the ocular disease related SNPs found from the OMIM and GWAS Catalogue knowledgebases. The SNPs identified in the databases have many potential applications, for example, new discoveries on underlying biology and risk prediction, personalized medicine, or better interpretation of disease etiology. For risk prediction, it was suggested that, the combination effect of a larger number of SNPs may result in a signal that is stronger than those from the unrelated individual SNPs. [240].

Fig. 2.11 Ocular disease related SNPs found in OMIM and GWAS Catalogue. (query made on May 8th, 2013)

2.4.3 Discovering Novel Disease Related SNPs from Large-Scale Genome Wide Association Study

Computational methods investigating for SNP-trait association study [241, 242] have been developed. Such methods treat SNPs as individual players in one's genetic profile. Following these methods, efforts [243–245] have been expanded to investigate those SNPs which have little effects on disease risk individually but seriously affect the disease risk jointly, the phenomenon being known as epistatic interaction, where the effects of one gene are believed to be modified by one or several other genes. The single-locus and epistasis SNP detection based algorithms test individual SNPs or pair of SNPs without taking into consideration, the underlying biological intertwining mechanism. Whereas, the real gene-gene interaction participating in biological pathway are often composed of a group of arbitrary number of SNPs. To date, exhaustively detecting significant SNP groups of arbitrary size is still computationally infeasible [244].

Recently, machine learning especially sparse learning algorithms have been introduced to GWAS data analysis. This is intended to tackle the challenge of identifying a group of N potent but interwinely correlated SNPs, some of which may not pass the stringent threshold by themselves. Penalized regression based on Least Absolute Shrinkage Selector Operation (LASSO) [246] has recently been explored for GWAS analysis. Some researchers [247, 248] have proposed 2-step approaches for Genome-wide association analysis via shortlisting a group of marginal predictors using penalized likelihood maximization for further higher order interaction detection. Hoggart *et al.*[249] have proposed a method to simultaneously analyze all SNPs in genome-wide and re-sequencing association studies. D'Angelo [250] have combined LASSO and principal- components analysis for detection of gene-gene interactions in genome-wide association studies. These approaches are not global due to the 2-stage process and none of them have considered incorporating prior knowledge into the model building. Prior knowledge can be combined with GWAS to improve the power of the association study [251]. It can also model dependencies and moderate the curse of dimensionality.

2.5 Discussion

Two major observations can be drawn from the above survey. First, there is a trend of transition of acquiring knowledge about CAD from learning rules from human experts to machine learning from largely available data. In the 80s, research was dedicated to constructing knowledge-bases from inputs of physicians [252, 253] for CAD tools. Building such systems required a lot human intervention, e.g. experts' inputs, and can be considered as a 'semi-automatic' way for knowledge acquisition. Over the years, the alternative approach of automatic knowledge acquisition without inputs from clinicians or experts, has become more popular [254, 255]. One such way of

knowledge acquisition is to capture patterns in data using non-clinical features, as described in Section 2.3.3.2. This approach offers a number of advantages:

- Knowledge learned from datasets are more precise in comparison with what constructed from the inputs from human experts, who could be less objective or consistent. An increased precision of CAD systems will make them more reliable for a mass screening application.

- Medical datasets embed local epidemiological patterns. Hence the learned knowledge can result in more accurate CAD tools, as disease and symptom patterns vary from one region to another [256]. A system learnt using data obtained from a particular region can be expected to be more precise in performing mass screening in that same region. The human experts, however, may not have sufficient experience in a particular region.

The second trend is the integration of heterogeneous data sources [257]. CAD tools rely on single data source may yield sub-optimal accuracy as the training data itself lack the complete aspects for decision making [258]. In the era of information explosion, data from multiple sources are becoming progressively available. For examples, retinal fundus cameras can be found in community polyclinics and optical shops; the health screening outreach programs have allowed individuals access the clinical data which was hard-to-access previously; the recent massive reduction in genotyping costs has made it possible to acquire SNP data with low cost.

Each of these heterogeneous data sources (image features, personal profile data, SNP data) is likely to contain a different perspective on the disease risk of an individual, based on the pathological, environmental and genetic mechanisms of the disease. Furthermore, integration of different data sources in CAD systems can also help in early detection since some of the early symptoms of the disease may appear in one data source but not the other.

There is no previous work attempting to combine these three types of data for automatic ocular disease detection. Possible reasons could be that only until recently such data has become available on a large scale. Also, researchers working on these heterogeneous data sets usually come from different domains with different foci, e.g. computer vision and image understanding researchers focused on DFP analysis, bioinformaticians are interested in discovering disease associated SNP or SNP groups. Effectively combining these data can maximize the information gain and pave the way for a holistic approach for automatic and objective disease detection and screening. In this thesis, Chapter 3 and 5 describe our pioneer works on combining the three types of data for ocular diseases detection.

Converse to the integration of multiple data sources, there is a possibility of using the same image to detect multiple diseases since many ocular diseases may have common symptoms. Along this line, there are already machine learning algorithm such as multi-task learning which look to solve similar problems. Chapter 5 describes our work on predicting multiple ocular diseases from single retinal image.

Chapter 3

Multiple Kernel Learning for Heterogeneous Data Fusion – Automatic Detection of Pathological Myopia

In this chapter, we describe an AODI-PM (*Automatic Ocular Disease Diagnosis through Biomedical Imaging Informatics*) framework which uses multiple kernel learning (MKL) to combine heterogeneous data sets for automatic ocular disease detection. We first briefly review the development of Kernel method, the Support Vector Machines (SVM), followed by the formulations of MKL. We then present AODI-PM, a CAD framework for PM diagnosis based a heterogeneous data sources. Through the use of MKL, AODI-PM intelligently fuses image, SNP and clinical data to enhance the disease detection accuracy. We conduct experiment on SiMES [17] dataset to validate the proposed system. In SiMES, data of retinal fundus imaging, genotyping and clinical were collected from 2,258 subjects in a population-based study. The results show that AUC of PM detection in AODI-PM achieves 0.888, better than results obtained from any individual data source or fusion of two data sources.

3.1 Introduction on Multiple Kernel Learning

3.1.1 Support Vector Machines and Kernel Method

Support vector machine (SVM) was first proposed by Boser, Guyon and Vapnik [259] in 1992, though the very first mention of the Generalized Portrait algorithm (implemented by SVMs) was much earlier in 1963 [260]. The basic idea of SVMs is to find a hyperplane which separates the d-dimensional data by maximizing the margin between two classes. Suppose we have n training examples $\{\mathbf{x}_i, y_i|_{i=1}^n\}$, each example has d inputs ($\mathbf{x}_i \in \mathbb{R}^d$), and a class label ($y_i \in \{-1, 1\}$). A hyperplanes in $\mathbb{R}^d$ can be parameterized by a vector ($\mathbf{w}$) and a constant (b), expressed in the equation

$$\mathbf{w}' \cdot \mathbf{x} + b = 0 \tag{3.1}$$

where $\mathbf{w}$ is in fact the vector orthogonal to the hyperplane. SVMs look for the hyperplane that maximizes the geometric distance to the closest data points. Which is equivalent to maximizes the margin $\frac{2}{\|\mathbf{w}\|^2}$ and thus requires optimization problem:

$$\min \tau(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 \tag{3.2}$$

$$\text{s.t. } y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1, \forall i \tag{3.3}$$

Where τ in Equation (3.2) is the objective function, which is subject to the inequality constraints in Equation (3.3). This form a constrained optimization problem, which can be solved by *Lagrangian multipliers* [261], a strategy for finding the local maxima and minima of a function subject to equality constraints [262]. A Lagrangian function to derive necessary conditions for conditional extrema of functions of several variables,

is defined for Lagrangian multipliers $\alpha_i \geq 0$,

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i (y_i(\mathbf{w}'\mathbf{x}_i) + b) - 1). \tag{3.4}$$

The *Lagrangian* L has to be minimized with respect to the primal variables $\mathbf{w}$ and b and maximized with respect to the dual variables α_i. Then, the dual problem can be formulated as

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j), \tag{3.5}$$

$$\text{s.t. } \alpha_i \geq 0(\forall i), \text{ and } \sum_{i=1}^{n} \alpha_i y_i = 0 \tag{3.6}$$

This minimization problem is what is known as a Quadratic Programming Problem. Fortunately, many techniques have been developed to solve them. From the derivation of these equations, the optimal hyperplane can be obtained:

$$\mathbf{w} = \sum_{i}^{n} \alpha_i y_i \mathbf{x}_i \tag{3.7}$$

In real problems, noise would cause a large overlap among the classes, the ideal separating hyperplane may not exist. To allow for the possible violation of examples, soft margin SVM was proposed [263, 264] by introducing the slack variables ξ_i for each of the training data point. By considering the different loss function for the slack variables, the primal SVM objective functions are given as in the following respectively for the loss function:

$$\min(\tau\mathbf{w}, \xi) = \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{i=1}^{n} \xi_i \tag{3.8}$$

$$\text{s.t. } y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1 - \xi_i, \text{ and } \xi_i \geq 0, \forall i \tag{3.9}$$

where the C is a constant larger than 0, which determines the trade-off between margin maximization and training error minimization. The dual problem can then be derives as follows:

$$\max_{\alpha} \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j), \tag{3.10}$$

$$\text{s.t. } 0 \leq \alpha_i \leq C(\forall i), \text{ and } \sum_{i=1}^{n} \alpha_i y_i = 0 \tag{3.11}$$

The only difference between Equations (3.5 3.6) and Equation (3.10 3.11) is the upper bound C on the Lagrange multipliers α_i. In the soft margin problem, the influence of possible outlier patterns is suppressed by C.

When data is in a linear space, the construction of Equation (3.10) requires the dot products $(\mathbf{x}_i \cdot \mathbf{x}_j)$. Since example data is often not linearly separable, SVMs cleverly cast the data into a higher dimensional space where the data is separable. We define a nonlinear mapping $\mathbf{z} = \phi(\mathbf{x})$ that maps $\mathbf{x}$ from the original space $\mathbb{R}^n$ to the high dimensional space (RKHS) $\mathcal{H}$, $\mathbf{x} \mapsto \mathbf{x} := \phi(\mathbf{x})$, where $\mathbf{x}$ denotes the vectorial representation of $\mathbf{x}$ in $\mathcal{H}$. Usually, the linear classifier $f(\mathbf{x}) = w'\mathbf{x}+b$ can be constructed in the original low dimensional space $\mathbb{R}^n$, and the linear classifier $f(\mathbf{x}) = w'\phi(\mathbf{x})+b$ in $\mathcal{H}$ corresponds to a particular the nonlinear classifier in the original low dimensional space.

When dimension of $\mathcal{H}$ is exponentially larger than $\mathbb{R}^n$, the computation of $\phi(\mathbf{x})$ becomes infeasible. SVMs introduce a *kernel trick* [265] to avoid the problem. For mapping $\mathbf{z} = \phi(\mathbf{x})$, if we replace all occurrences of $\mathbf{x}$ with $\phi(\mathbf{x})$. Our QP problem in

Equation (3.10,3.11) would become:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j (\phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)), \tag{3.12}$$

$$\text{s.t. } \alpha_i \geq 0 (\forall i), \text{ and } \sum_{i=1}^{n} \alpha_i y_i = 0 \tag{3.13}$$

The key insight of *kernel trick* is that the higher-dimensional space does not need to be dealt with directly (calculate $\phi(\mathbf{x})$), but only the formula for the dot-product $(\phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j))$ in that space is needed, which largely eliminates computation complexity.

Such approach belongs to kernel-based learning method, which embeds the data into a Euclidean space, and then conduct search for linear relations among the embedding data points. For any two vectors $\mathbf{x}, \mathbf{x}'$ belong to sample data $\{\mathbf{x}_i, y_i|_{i=1}^n\}$, a kernel is defined in [266] as a function K, that for all $\mathbf{x}, \mathbf{x}'$ satisfies $K(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle$ where ϕ is a mapping from $\boldsymbol{X} = \{\mathbf{x}_i|_{i=1}^n\}$ to the feature space $\mathcal{H}$.

3.1.2 General Kernel Learning Method

It is widely acknowledged that a key factor in an SVM's performance is the choice of the kernel. However, in practice, very few different types of kernels have been used due to the difficulty of appropriately tuning the parameters. For examples, there are four commonly used basic forms for kernel function $K(\mathbf{x}, \mathbf{x}')$:

- linear: $K(\mathbf{x}, \mathbf{x}') := \mathbf{x}^T \mathbf{x}'$
- polynomial: $K(\mathbf{x}, \mathbf{x}') := (\gamma \mathbf{x}^T \mathbf{x}' + r)^d, \gamma > 0$
- radial basis function (RBF): $K(\mathbf{x}, \mathbf{x}') := exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2), \gamma > 0.$
- sigmoid: $K(\mathbf{x}, \mathbf{x}') := tanh(\gamma \mathbf{x}^T \mathbf{x}' + r).$

Here, γ, r and d are kernel parameters. The choice of kernel and the corresponding feature space are central choices that must generally be made by a human uses. While

this provides opportunities for prior knowledge to be brought to bear, it can be difficult in practice to find prior justification for the use of one kernel instead of another [267]. Research have been conducted to explore model selection methods that allow kernels to be chosen in a more automatic way based on data. Chapelle et. el. [268] proposed to automatically tune the multiple parameters for the kernel while training the SVM model. The method is based on minimizing some estimates of the generalization error of the SVMs. They can learn the multiple hyper-parameters such as the bandwidth scaling parameters for each dimension of the features in the RBF kernel. The empirical results show that the Radius-margin bound is good for the square hinge loss SVM [269]. However, the objective function proposed is non-convex, thus the global solution can not be guaranteed. Also, the method can be quite complex when there are more training samples. The Hyper-kernel regularization was proposed by Ong and Smola et. el. [270–272], and it learns the kernel by defining a reproducing kernel Hilbert space on the space of the kernels. The proposed problem has l^2 new parameters as the kernel coefficients, and is formulated as the Semi-definite Programming (SDP) problem. [273] further proposed to use the Second-Order Cone Programming (SOCP) to improve the efficiency for solving the hyper-kernel regularization problem.

The pioneer work for learning the linear combination of the a set of pre-defined multiple base kernels was first proposed in [267, 274] by minimizing the structural risk functional similarly with the SVM, which is the major research direction in the recent ten years. Besides the linear combination there are also works that learns the non-linear combination of the multiple base kernels such as [275, 276].

3.1.3 Multiple kernel learning

In recent years, Multiple kernel learning methods are proposed to use multiple kernels instead of selecting one specific kernel function and its corresponding parameters. The

original problem of MKL is formulated as follows. Given a set of basic kernel functions, $\{K_m(\mathbf{x}, \mathbf{x}')|m = 1, ...M\}$ and a set of training examples $S = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_n, y_n)\}$, the goal of MKL is to optimize a cost function $Q(K_\mu(\mathbf{x}, \mathbf{x}'), S)$ where $K_\mu(\mathbf{x}, \mathbf{x}')$ is a linear combination of basic kernels,

$$K_\mu(\mathbf{x}, \mathbf{x}') = \sum_{m=1}^{M} \mu_m K_m(\mathbf{x}, \mathbf{x}'), \mu_m \geq 0, \sum_{m}^{M} \mu_m = 1 \tag{3.14}$$

and the feature space $\mathcal{H}$ is mapped by:

$$\mathbf{x} \mapsto \phi_\mu(\mathbf{x}) = (\sqrt{\mu_1}\phi_1(\mathbf{x}), ...\sqrt{\mu_M}\phi_M(\mathbf{x}))^T \in \mathcal{H} \tag{3.15}$$

where $\phi_i(\mathbf{x})$ is the mapping to the $\mathcal{H}_i$ feature space associated with the K_m. Two criteria, the margin based quality as well as the alignment based quality are used as the objective function for the optimization, and will be detailed respectively.

Lanckriet *et al.*[267, 274] proposed to use the maximum margin criterion similarly to SVM. Using $\nu - SVC$ [264] as the classification model, the MKL problem is defined as:

$$\min_{\mu} \max_{\alpha} -\frac{1}{2} \sum_{m=1}^{M} \mu_m (\alpha \odot y)' K_m (\alpha \odot y) \tag{3.16}$$

$$\text{s.t. } \alpha' y = 0, \alpha' 1 = 1, 0 \leq \alpha \leq C, \tag{3.17}$$

$$K \in \mathcal{K}, \tag{3.18}$$

where $\mathcal{K}$ is the feasible set constructed from the give M base kernels.

If only restricting the combined kernel matrix to be semi-positive definite $K = \sum_{m=1}^{M} \mu_m K_m \succcurlyeq 0$, the previous problem is formulated as a Semi-definite Programming (SDP) problem [267]. If the combined kernel matrix is restricted to be the convex combination where $\sum_{m=1}^{M} \mu_m = 1, \mu_i \geq 0$, the problem can be formulated as the

quadratically constrained quadratic programming (QCQP) problem. Based on the formulation Equation (3.16), the Semi-Infinite Linear Program (SILP) formulation was proposed in [277] by solving the following problem:

$$\max_{\alpha,\lambda} \lambda \tag{3.19}$$

$$\text{s.t.} \sum_{m=1}^{M} \mu_m = 1, \mu_m \geq 0, \tag{3.20}$$

$$\frac{1}{2} \sum_{m=1}^{M} \mu_m (\alpha \odot y)' K_m (\alpha \odot y) \geq \lambda$$

$$\alpha' y = 0, \alpha' 1 = 1, 0 \leq \alpha \leq C,$$

SILP formulation makes the large-scale MKL problem applicable by recycling the single kernel SVM and thus the complexity of the problem is equivalent to that of the optimization of SVM. The above problem is non-smooth due to the mixed-norm structure for the complexity. Actually, as denoted in [278] that $(\sum_{i=1}^{M} \|f_i\|\mathcal{H})^2 = \min_\mu \sum_{i=1}^{M} \frac{\|f_i\|^2 \mathcal{H}_i}{\mu_i}$ with the simplex constraint for $\mu \in \{\mu | 0 \leq \mu, \sum_{i=1}^{M} \mu_i = 1\}$. Then the above formulation was further formulated as:

$$\min_{\mu,\{f_m\},b,\rho} \frac{1}{2} \sum_{m=1}^{M} \frac{\|f_m\|^2 \mathcal{H}_i}{\mu_m} + C \sum_{i=1}^{n} \epsilon_i - \rho \tag{3.21}$$

$$\text{s.t. } y_i (\sum_{m=1}^{M} f_m(x_i) + b) \leq \rho - \epsilon_i, \epsilon_i \geq 0,$$

$$0 \leq \mu, \sum_{m=1}^{M} \mu_m = 1$$

The solver of the above SimpleMKL [278] approach alternates between the optimization of kernel weights and the optimization of SVM classifiers. In each step, given the current solution of kernel weights, it solves a classical SVM with the combined kernel; then a specific procedure is used to update the kernel weights. The advantage of the alternating scheme is that SVM solvers can be very efficient due to recent advances

in large scale optimization [279]. However, such approaches for updating the kernel weights are still time consuming.

In summary, there are two active research directions in MKL. One is to improve the efficiency of MKL algorithms. Early work on SDP algorithm [267, 280] and the block-norm regularization method based on SOCP [281] require intensive computation. Following work on alternating approaches [277, 278, 282] were more efficient by alternating between the optimization of kernel weights and the optimization of SVM classifiers. The advantage of the alternating scheme is that SVM solvers can be very efficient due to recent advances in large scale optimization [279]. Based on the observation of the equivalence between MKL and group lasso regularizer [283], Xu *et al.*[284] formulated a close-form solution for optimizing the kernel weights, leading to an more efficient algorithm for solving MKL.

The second direction is to improve the accuracy of MKL by exploring the possible combination ways of base kernels. L1-norm of the kernel weights, also known as the simplex constraint, is mostly used in MKL methods. The advantage of the simplex constraint is that it leads to a sparse solution, i.e., only a few base kernels among mynas carry significant weights. However, as argued in [285], the simplex constraint may discard complementary information when base kernels encodes orthogonal information, leading to suboptimal performance. To improve the accuracy in this scenario, an L2-norm of the kernel weights, known as a ball constraint, is introduced in their work. Another extension is to explore the grouping property or the mixed-norm combination, [286, 287] or the possibility of non-linear combination of kernels [275, 288]. Although the solution space has been enlarged, the non-linear combination usually results in non-convex optimization problem, leading to even higher computational cost. Moreover, the solution of nonlinear combination is difficult to interpret.

3.2 MKL for heterogeneous data fusion

3.2.1 Applications of MKL

Due to its interpretability and good performance, MKL has been considered as a favorable technique for fusing multiple data sources or identifying feature subsets. The application of MKL have been seen in a number of domains.

The early applications of MKL was in the bioinformatics domain. In 2004, Lanckriet *et al.*[280] proposed an MKL-based computational framework for combining heterogeneous genome-wide datasets and reported an application of the approach to tackle the classification problem of membrane proteins and yeast ribosomal. The proposed SDP/SVM algorithm outperformed the SVM trained on any single dataset or trained upon a naive combination of kernels. Following this work, Sonnenburg *et al.*[289] used a more efficient algorithms, SILP, to transform the complexity of the MKL problem to the optimization of the SVM. The authors applied their proposed methods in the prediction task on identifying acceptor splice site as well as alternatively spliced exons. The authors demonstrated that the approach could handle thousands of examples while combine hundreds of kernels within reasonable time, and identifies a few statistically significant positions successfully.

In computer vision, Harchaoui et al. [290] defined a family of kernels representing their respective image segmentation graphs. Leveraging the natural structure of images, the kernels are based on soft matching of subtree–patterns of the respective graphs thus are robust to the uncertainty of the associated segmentation process. The proposed kernels enable efficient supervised learning to classify natural images with a SVM. In 2009, Vedaldi et al. [291] proposed an MKL approach to learn an optimal combination of exponential x^2 kernels, each of which captures a different feature channel.

Their images features cover the feature descriptors at different levels of spatial organization, and the distribution of edges, dense and sparse visual words. The authors proposed a novel three-stage classifier to combines linear, quasi-linear, and non-linear kernel SVMs. The result confirmed that increase of kernel non-linearity will raise the discriminative power, at the cost of higher computational complexity.

Due to the fact that, only until recently abundant image data and SNP data has become available in large scale, there is no report work combining image features and SNP data for disease prediction. These heterogeneous data sets were used to be analyzed in different domains by image understanding researchers, data mining experts and bioinformaticians, the field of the information integration is still very young.

3.2.2 AODI-PM framework for PM detection

As introduced in Section 2.3.2.4, PM is one of the leading causes of blindness and the accurate detection of PM will facilitate timely intervention and enable better disease control to slow down the progression of the retina degeneration.

Current clinical practice in detecting PM relies heavily on manual efforts. A complete eye exam conducted for patients which usually takes up to 60 minutes. These exams include check on the patient's medical history, a set of physical eye evaluation tests including visual acuity, refraction and visual field test. Complicated ophthalmological devices are involved, for examples, slit lamp evaluates the anterior sections and lens, tonometry measures the intra-ocular pressure ophthalmoscopy observes the retina. Research on retinal imaging algorithms and the development of CAD systems to automatically detect PM will alleviates the burden on the clinicians by eliminating the need for manual examinations. Recently a number of systems were proposed for automatic PM screening through retinal fundus image processing. For example, Liu

et al.[74] described the PAMELA system to diagnose PM by detecting parapapillary atrophy (PPA) around the optic disc in fundus images.

Other than image based only approaches, Zhang *et al.*[292] investigated a method to identify an optimal set of essential features from a combination of clinical data and information extracted from fundus image, in order to improve the accuracy of PM detection. With genotyping technologies out-pacing Moore's Law since 2008 [293], it has become much more affordable to obtain genetic information, which, could be an important data source for decision making due to the fact that PM is highly heritable. The current genotyping technics produce SNP data which offers partial view of a person's genetic profile. The available knowledgebase of disease associated SNPs can be used as a form of genetic prior knowledge in estimating the likelihood of disease occurrence.

Different data source, including fundus image, demographic/clinical and SNP data, contains a different perspective on the disease risk of an individual, based on the pathological, environmental and genetic mechanisms of the disease. These perspectives may potentially be complementary, such that a combination of the data from these independent sources are able to provide a more comprehensive and holistic assessment of the disease[280].

We propose a computer-aided diagnosis framework for the detection of PM called AODI-PM. The AODI-PM framework automatically detects PM based on a combination of heterogeneous sources, i.e. imaging data, demographic/clinical data, and genotyping data. We use an MKL-based approach to optimize modeling, learning and classification. Figure 3.1 illustrates the architecture of the proposed AODI-PM framework.

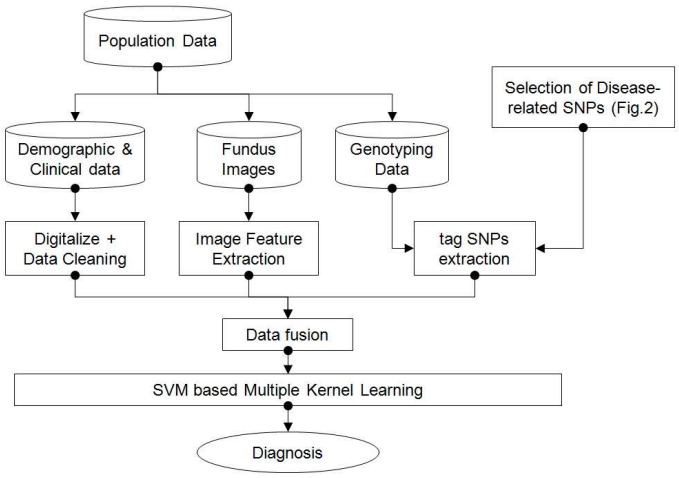


Fig. 3.1 Architecture of AODI-PM framework

3.3 Data Description and Feature Extraction

We evaluate the proposed AODI-PM framework on the Singapore Malay Eye Study (SiMES) database [17]. SiMES examined a population-based, cross-sectional, age stratified, random sample of 3280 Malays (78.7% participation rate) aged 40 to 80 years living in Singapore. In this population study, a subject's image data, blood sample for genotyping as well as clinical diagnosis were acquired simultaneously during the clinic visit. We use the clinical diagnosis of PM (made by multiple clinicians) as the gold standard to evaluate our approach.

In current clinical settings, fundus images are easily available at polyclinics and even optical shops. Furthermore, the cost of genotyping chips has decreased dramatically in recent years, a trend of which would greatly increase the accessibility of a person's genotyping data in the near future. The objective of our study is to develop a computational tool facilitating automatic prediction for applications such as health screening when clinicians are not present but abundant data is available.

The following data is used to evaluate the proposed AODI-PM framework:

- Fundus Image Data: The images were acquired using a 45° FOV Canon CRDGi retinal fundus camera with a 10D SLR backing, at an image resolution of 3072×2048 pixels

- Demographic/clinical Data: The eye screening record in SiMES contains demographic/clinical data such as age and gender, medical histories (e.g. diabetes etc.) and ocular examination data. The clinical diagnosis of pathological myopia is used as the gold standard label in this study.

- Genotyping (SNP) Data: subjects were genotyped on Illumina 610quad arrays, followed by a stringent quality control (QC) procedure [294]. The QC process excludes the subjects with a missing call rate $> 5\%$, filters out monomorphic SNPs, non-autosomal SNP and SNPs with minor allele frequency (MAF) $< 5\%$. A Hardy-Weinberg equilibrium (HWE) test was also conducted to detect genotyping artifacts [295]. The final SNP data set contains $2,542$ individuals with $557,824$ SNPs on 22 autosomal chromosomes.

3.3.1 Knowledge-based Feature Selection in SNP Data

It has been shown that there is an interplay between genetic factors and environmental influences [179] in myopia, with an estimated heritability of myopia at 0.306 [296]. A set of myopia related genes were discovered in linkage studies [297–299], and recent genome wide association studies (GWAS) further identified several loci highly associated with PM [300–304]. This valuable knowledge forms a *smart prior* in our framework, and using such a *smart prior* for feature selection enables us to overcome the *curse of dimensionality* raised by the overwhelming number of SNPs as compared to samples. We propose a holistic approach to identify myopia-related SNPs using the following steps:

- Identify susceptibility loci from a group of myopia-related genes

  We use the OMIM (Online Mendelian Inheritance in Man) database [232] to obtain disease related SNPs. OMIM contains information on known genetic disorders and over 12,000 genes, with carefully examined reference literature. We searched OMIM with query item *myopia [TI]* and found 40 entries, from which a list of myopia-related SNPs were extracted as shown in Table 3.1

- Obtain susceptibility loci from recent published genome wide association study

  We used the NHGRI GWAS catalog [305] to search for PM-associated SNPs discovered by recent Genome Wide Association Studies [300–304]. The SNPs and their references are listed in Table 3.2

- Match tag SNPs genotyped in SiMES data

  The SNPs identified in the above steps may not appear as markers genotyped in SiMES data. Based on the fact that Illumina 610quad arrays are derived from the International HapMap Project [306] with one tag SNP every 5–6 kb across the genome in the CEU,CHB+JPT and YRI populations, we use GVS (Genome Variation Server) [307] to find corresponding tag SNPs. The GVS database contains 11.8 million SNPs with corresponding genotyping data and provides a set of tools for the analysis of SNP data. For each SNP identified in Steps 1 and 2, we set a range of 3kb both up- and down-stream with a LD-score $r^2 > 0.8$ as the search criteria to catch the corresponding tag SNPs.

Figure 3.2 illustrates the steps described above. A detailed list of the extracted SNPs are listed in appendix tables (in a separate file). In total 87 SNPs are matched in SiMES genotyping data and these SNPs are used to form a sub-feature space for learning.

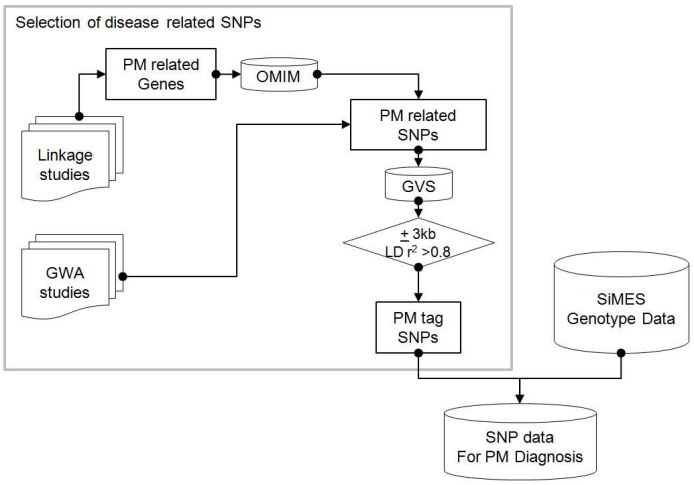


Fig. 3.2 Knowledge-based SNP selection in genotyping data

Table 3.1 PM (PM) related SNPs found from Genetic Linkage Studies

Genes	Location	OMIM ID	PM SNP	Source
MYP2	18p11.31	160700	rs1034762, rs1635529, rs1793933, rs3803183, rs17122571	Young, Ronan, Drahozal et al. (1998), Mutti et al. (2007), Metlapally et al. (2009)
MYP3	12q21-q23	603221	rs3832846, rs17853500, rs3759223, rs10860860, rs2946834, rs6214	Young, Ronan, Alvear et al. (1998), Lin et al. (2010), Metlapally et al. (2010)
MYP7	11p13	609256	rs1506, rs592859, rs608293, rs628224, rs662702, rs667773, rs694617, rs1540320, rs1806155, rs1806158, rs1806159, rs1806180, rs1894620, rs2071754, rs2239789, rs3026389, rs3026401	Hammond et al.(2004)
MYP11	4q22-q27	609994	rs113432966, rs112669274, rs112391551, rs112356377, rs111691784, rs111322719	Zhang, Guo et al. (2005)
MYP12	2q37.1	609995	rs111706042	Paluru et al. 2005
MYP13	Xq23-q25	300613	rs113695792, rs111774596	Zhang, Guo et al. 2006
MYP14	1p36	610320	rs113328794	Stambolian et al. (2004)
TGIF	18p11.31	602630	rs121909066, rs121909067, rs121909068, rs121909069, rs121909070, rs28939693	Gripp et al. (2000)

Table 3.2 PM associated SNPs found in Genome-wide Association Studies (GWAS)

Genes	Location	PM SNP	Source
GJD2	15q14	rs634990	Solouki et al. 2010, Nature Genet.
RASGRF1	15q25	rs939661	Hysi et al. 2010, Nature Genet.
CTNND2	5q15	rs6885224, rs12716080	Li et al. 2011, Ophthalmology
MIPEP	13q12.12	rs9318086	Shi et al. 2011b, AJHG
ZC3H11B LAMA2 CD55	1q41 6q22.33 1q32.2	rs4373767 rs12193446 rs1572275	Fan et al. 2012, PloS Genetics
ZNF644	1p22.2	rs6680123	Shi et al. 2011a, Plos Genetics
MYP11	4q25	rs10034228, rs1585471	Li et al. 2011, Hum Mol Genet.
BLID GLULP3	11q24.1	rs577948 rs12275397	Nakanishi et al. 2009, Plos Genetics

3.3.2 Demographic and clinical data preprocessing

Both environmental and genetic factors have been associated with the onset and progression of myopia. Some of the known environmental risk factors of myopia include *close up work, educational level, IQ, outdoor activity, academic achievement* and *a introvert personality* [179]. These risk factors are partially represented in the demographic and clinical variables obtained from the population study protocol. The data is cleaned by removing subjects or variables with more than 5% missing values. We digitized the categorical parameters and scaled all variables to range of $[0, 1]$. The clean set contains 44 parameters as listed in Table 3.3, with 2,258 subjects data matched with image and SNP data.

Table 3.3 List of Demographic & clinical variables used in AODI-PM

Age	Blood LDL Cholesterol	Can read
Age Group	Blood HDL Cholesterol	Can write
Gender	Triglycerides	Alcoholic drink categories
Height	Hypertension	Ever Smoke
Weight	Hypertension treatment & control	Current smoker
Diastolic Blood Pressure	Albumin-Creatinine ratio	Angina
Systolic Blood Pressure	Diabetes I	Heart Attack
Pulse Pressure	Diabetes II	Stroke
Mean arterial pressure	Job Categories	Hypercholessterolemia
BMI	Race	Thyroid Condition
Blood Creatinine	Marital Categories	Chronic Kidney Disease indicator
Blood Glucose	Income Categories	hyperlipidemia
Blood HbA1c Categories	Type of place living in	Metabolic syndrome
Blood Glycosylated Haemoglobin	Place of birth	Microalbuminuria
Blood Total Cholesterol	Education categories	

We conducted a univariant analysis for all parameters. P-values are obtained by conducting the Student's T-test for numerical variables and the Chi-square test for categorical variables. The following parameters were found to have an associated with PM with P-value < 0.05 : *Age* (p=0.019), *Job Category* (p=0.007), *Income* (p=0.003), *Type of place living in* ($p < 0.0005$), *Education* ($p < 0.0005$), *Ever Smoke* and *Current Smoke* (both $p < 0.005$).

3.3.2.1 Semantic Image Feature Analysis for Fundus Image

Semantic image features, also known as high-level features, differ from low-level local features as they are global features which are location-independent. In this work, the bag-of-words (BOW) model [308] is introduced to semantic image feature extraction.

BOW is a simplified representation used in image retrieval by treating local image features as words. In natural language processing, a bag-of-words is a sparse vector of occurrence counts of words; that is, a sparse histogram over the vocabulary. Correspondingly, in computer vision, a bag-of-words is a sparse vector of occurrence counts of a vocabulary of local image features (codebook), which is a location-independent

global feature. The properties of local features, such as intensity, rotation, scale and affine invariants can also be preserved. Figure 3.3 illustrates the described method.

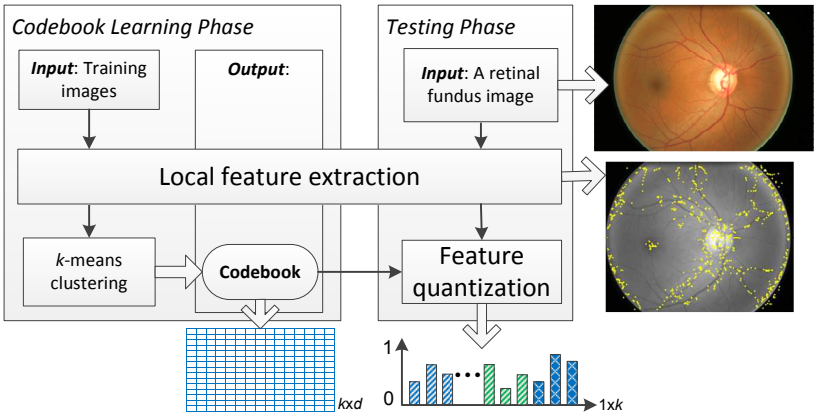


Fig. 3.3 Semantic image feature extraction

Many visual features can be extracted from grids or superpixels [83] to form local features, such as HOG [309], BIF [82] and color histograms [109] which are related to edges, textures and intensity, respectively. In this work, SIFT (Scale-invariant feature transform) [191] features are used as local features. SIFT has been widely used in object detection and classification, due to its intensity, rotation, scale and affine invariant properties. In this implementation, the Harris-Laplacian (HAR) and Hessian-Laplacian (HES) detectors [310] were used to generate SIFT features from each retinal fundus image. This is mainly because both detectors produce complementary features: HAR locates corner features, while HES extract blob features. Each SIFT feature was represented as a 128-dimensional histogram and each dimension was quantized into an integer between 0 and 255.

To reduce computational costs and avoid feature noise from the retinal image field of view limits, the images were resized to a height of 256 pixels by keeping the original aspect ratio, and only feature points within 0.95 radius to the center were collected for further processing. In addition, the SIFT feature extraction was performed on the

green channel only, since the retinal images are less well differentiated in the red and blue channels.

After obtaining all the SIFT features from training images, $k-$means clustering was used to generate the codebook by randomly selecting half of the training images, with each cluster centroid representing a visual word. After which the BOW global features (*i.e.*, occurrence counts of the visual words in a retinal image) of each training and testing image were obtained in the quantization procedure. To balance the dimensions of different features, we empirically set k=100. L_1-normalization is performed to standardize features before training and testing.

3.3.3 Data Fusion

The features extracted from each of the three heterogeneous data sets were merged via subject matching. The final dataset contains $2,258$ subjects with demographic/clinical data, fundus image and SNP data. Among the $2,258$ individuals, 58 had been diagnosed with PM while the rest were normal. The distribution of pathological myopia subjects in the dataset is representative of the prevalence of PM in the population. The range of each feature dimension was normalized to the range of $[0, 1]$ in order to avoid magnitude differences among the dimensions.

3.4 Experiments and Results

3.4.1 Learning Algorithms

In our experiments, when only one type of feature set (e.g., SNP) is used, a linear kernel $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$ based basic SVM classifier is utilized, where the corresponding label of $\mathbf{x}_i$ is determined by $\mathbf{w}^T \phi(\mathbf{x}_i) + \mu$. When incorporating feature sets from multiple data sources, multiple kernel learning (MKL) is applied to learn the adapted kernel

function for each feature set, and to optimize the contribution of each sub-kernel for the resulting classifier. In such cases, a convenient approach is to consider that $K(\mathbf{x}_i, \mathbf{x}_j)$ is actually a convex combination of the basis kernels:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \sum_{m=1}^{M} d_m K_m(\mathbf{x}_i, \mathbf{x}_j), d_m \geq 0 \text{ and } \sum_{m=1}^{M} d_m = 1, \tag{3.22}$$

where M is the total number of kernels. Each basis kernel K_m may either use the full set of features describing samples or subsets of features stemming from different data sources [281]. Within this MKL framework, the problem of data representation through the kernel is then transferred to the selection of weights d_m. In AODI-PM, we use basis kernels based on each single data source, and demonstrate that models based on a combination of multiple sources are better than those using a single data source. For efficiency, one linear kernel is initialized for each feature type. There are many MKL solver toolboxes which are publicly available, such as SimpleMKL [278] and Group Lasso [284]. The LIBLINEAR toolbox [311] is used to train linear SVM models for each individual data source, and the Group Lasso [284] toolbox is used to train MKL models.

3.4.2 Experimental Methods for AODI-PM

To demonstrate that the combination of multiple data sources can enhance detection accuracy in our PM-BII framework, we report and compare the diagnosis performance of 7 methods using the following different features and their combinations:

1. Demographic/clinical data only (referred to as **D**)

2. SNP data, genetic information only (referred to as **G**)

3. low-level direct image features only (referred to as **I**)

4. combined demographic/clinical data and SNP data ($\mathbf{D} + \mathbf{G}$)

5. combined demographic/clinical data and image features ($\mathbf{D} + \mathbf{I}$)

6. combined SNP data and image features ($\mathbf{G} + \mathbf{I}$)

7. combined all three data source, $\mathbf{D} + \mathbf{G} + \mathbf{I}$ (AODI-PM)

For fair comparison, we performed 10 independent tests, with two rounds of stratified cross-validation conducted per test. This was carried out in the following way. In each test, all subjects were randomly divided into non-overlapping sets of equal size, A and B. In the first round, we used all the positive subjects and the same number of randomly selected negative subjects from set A as training set, due to the imbalanced in the number of positive (PM) and negative (normal) subjects. The trained model is then used for testing set B. The second round was conducted in the same approach but with subjects from set B used for training and those from set A used for testing. In total, 20 groups of evaluation results were collected for each of the 7 methods for analysis.

We assess the classification performance using the area under the ROC (receiver operating characteristic) curve (AUC) *which evaluates the overall performance* and a balanced accuracy with a fixed 85% specificity. The balanced accuracy ($\bar{P}$), sensitivity (P_+) and specificity (P_-) are defined as

$$\bar{P} = \frac{P_+ + P_-}{2}, P_+ = \frac{TP}{TP + FN}, P_- = \frac{TN}{TN + FP}, \tag{3.23}$$

where TP and TN denote the number of true positives and negatives, respectively, and FP and FN denote the number of false positives and negatives, respectively.

3.4.3 Results

Table 3.4 shows the results for the different input data, both single and combined, on their ability to detect PM, measured using the specificity, sensitivity and area under the ROC curve (AUC). The mean and standard variation (SD) values of AUC of each method were calculated based on the results obtained from the 20 sets of cross validation testing as described in the above Methods section. At the screening-based specificity setpoint of 0.85, in comparing only the models from single sources, the results show that the use of imaging data provided the best prediction of PM (Sensitivity $P_+ = 0.71$), compared to that of SNP data (Sensitivity $P_+ = 0.52$) and showed a large improvement over detection using only demographic data (Sensitivity $P_+ = 0.27$). Comparatively, detection using only demographic/clinical data was the least accurate compared to the other single sources.

Table 3.4 Sensitivity and AUC results for the various sources combinations. Results show AODI-PM is better able to detect PM compared to the other individual or combined sources.

source	combination type	sensitivity (specificity=0.85)	AUC	
			mean	SD
SNP(G)	Single	0.52	0.774	0.038
retinal image(I)		0.71	0.852	0.044
demographic/clinical(D)		0.27	0.607	0.044
G+I	Two	0.73	0.875	0.032
D+G		0.56	0.792	0.037
D+I		0.71	0.863	0.033
D+G+I	Multiple	**0.77**	**0.888**	0.032

Notes:

D Demographic/clinical data; **G** SNP data, genetic information; **I** low-level direct image features

$\mathbf{D + G}$ combined demographic/clinical data and SNP data

$\mathbf{D + I}$ combined demographic/clinical data and image features

$\mathbf{G + I}$ combined SNP data and image features

$\mathbf{D + G + I}$ combined all three data source -(AODI-PM)

When multiple (2 or more) data sources are combined, the general trend shows that the sensitivities from combining sources outperforms their component sources at specificity of 0.85, with the best performing model based on the combination of SNP data, demographic/clinical data and imaging data in the AODI-PM framework (Sensitivity $P_+ = 0.77$).

This trend can also be observed using the calculated AUC metrics, with the corresponding ROC plots presented in Figure 3.4 and box plot of AUC distribution based on the 20 rounds of cross validation tests shown in Figure 3.5. The AODI-PM prediction

model combining demographic/clinical data, SNP data and imaging data generated the best AUC metrics ($AUC = 0.888$), outperforms all other source combinations or single sources. Compared to the single sources, the use of AODI-PM resulted in significant improvements over demographic/clinical data $\mathbf{D}$ (46.3%, $p < 0.005$) and genetic information $\mathbf{G}$ (14.7%, $p < 0.005$), as well as a modest improvement over imaging data $\mathbf{I}$ (4.2%, $p = 0.19$). Furthermore, the results also show AODI-PM performing moderately better than the combined models obtained from the combinations of any two sources, resulting in improvements of 12.1%, 2.9% and 1.5% in AUC over $\mathbf{D} + \mathbf{G}$, $\mathbf{D} + \mathbf{I}$ and $\mathbf{G} + \mathbf{I}$ respectively.

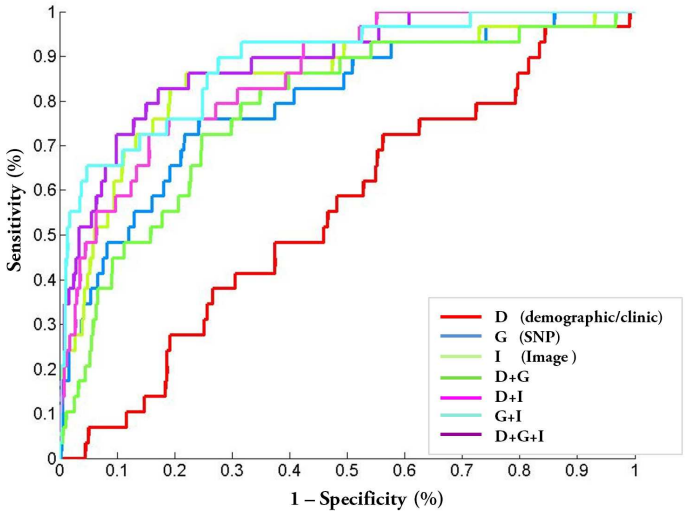


Fig. 3.4 ROC (receiver operating characteristic) curve of various methods

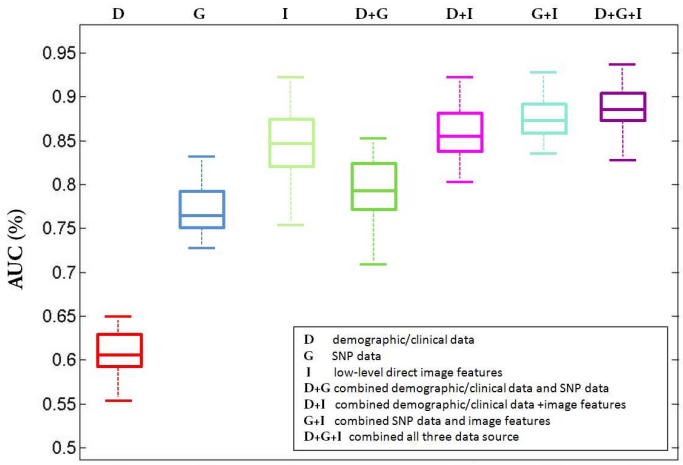


Fig. 3.5 Boxplot of AUC to compare various methods

3.5 Discussion

Our experimental results also seem to suggest an advantage in combining any two sources over the use of their component sources. For example, the use of SNP and retinal image information $\mathbf{D} + \mathbf{G}$ produced an *AUC* of 0.792, which is better than the individual *AUC*s from **D** (30.4%) and **G** (2.3%) respectively. This trend can also be observed for the other two combinations $\mathbf{G} + \mathbf{I}$ and $\mathbf{D} + \mathbf{I}$ over their components.

In this work, we have tested the use of different combinations of data for the detection of PM. These data sources can be described as imaging data, SNP data and demographic/clinical data. Based on the results of the experiments, the following observations can be made:

1. AODI-PM approach of combining imaging, SNP and demographic/clinical data outperformed single data sources and two-source combinations

   In our experiments, we have shown that the combination of imaging, genomic and demographic data in the AODI-PM framework was able to achieve an *AUC*

of up to 0.888. The AODI-PM prediction results outperform the models based on other data combinations, as well as the individual component sources.

2. Advantages in combining different data types

   Furthermore, the experiments also support combining different data types for pathological myopia prediction. In the experiments based on the combinations of any two different types, we observed that the results were better than the models which only use the individual data components. This was most obvious in the use of demographic/clinical data $\mathbf{D}$, which when used in conjunction with any other data type registered an improvement of at least 30.4% ($p < 0.005$) in PM detection. Although the use of individual data can possibly be used for detection, our results show that it is advantageous to include at least one other data type in the model. This suggests that the data types are indeed complementary.

3. Usefulness of demographic/clinical data

   The results show that the performance of AODI-PM ($\mathbf{D} + \mathbf{G} + \mathbf{I}$) is comparable to that of using SNP and imaging information $\mathbf{G} + \mathbf{I}$. However, the addition of demographic/clinical data $\mathbf{D}$ to genetic information $\mathbf{G}$ or $\mathbf{D}$ to $\mathbf{I}$ does show an improvement in detection accuracy. This suggests that in the overall AODI-PM framework the inclusion of demographic/clinical data $\mathbf{D}$ may not be strictly necessary, particularly when both genetic information $\mathbf{G}$ and imaging information $\mathbf{I}$ are included, and further suggests some possible redundancy in the use of demographic/clinical data $\mathbf{D}$ with genetic and imaging information $\mathbf{G} + \mathbf{I}$. Nonetheless, a model that is built using imaging information I or SNP data $\mathbf{G}$ alone would benefit from the inclusion of demographic/clinical data $\mathbf{D}$.

4. We observe the limited significance of adding SNP and demographics data into the prediction model, with a modest 4.2%($p = 0.19$) improvement of AUC. This may be due to the limited number of subjects in our study. Increasing data available in future studies could allow us to draw more significant conclusions.

Demographic/clinical data, imaging data and SNP data can provide different perspectives towards disease detection. With the large quantity of potential data that can be obtained, the challenge is to combine these data in a holistic fashion to make the best use of their individual advantages. Computer-based informatics methodologies offer such an opportunity to intelligently fuse these data sources. We have proposed AODI-PM, a framework powered by MKL, for PM diagnosis by combining heterogeneous biomedical data, including demographic data, imaging data and SNP data. Our experiments show that the AODI-PM framework is able to detect PM with high accuracy, and supports the use of data fusion over any single or two-source combination. These promising results encourage further exploration of the AODI-PM framework for the detection of other ocular diseases.

In this work, SNP feature selection is achieved by acquiring knowledge regarding disease associated SNPs, as illustrated in Figure 3.2. However, not every disease are well studied in terms of SNPs. In the case of no much information about disease related SNPs available, another way for feature selection in SNP data could be achieved by incorporating the grouping information of SNPs with available information of gene and biological pathways, in the next chapter a sparse-learning approach is developed for such approach.

Chapter 4

Identify Predictive SNP groups in Genome Wide Association Study: A Sparse Learning Approach

Genome-Wide Association Study (GWAS) aims to identify genetic variants that are significantly associated with genetic traits. To analyze GWAS data that often contains 0.5 to 1 million Single Nucleotide Polymorphisms (SNPs) genotyped from thousands of individuals, stringent statistical significant thresholds are pre-defined for multiple testing adjustment, e.g., with p-value $< 10^{-8}$ for single SNP detection and at least $< 10^{-12}$ for SNP-SNP interaction detection. Such stringent thresholds were used for efficiency computation but it hinders the discovery of many true genetic variants and more practical approaches are needed to conduct GWAS.

We propose a machine learning approach to identify groups of predictive SNPs in GWAS analysis. Our method differs from other methods by first translating the genomics knowledge into SNP grouping as priors, then selecting a list of most predictive SNP groups using linear regression regularized by group sparse constraints, solved by Group-lasso (Least Absolute Shrinkage and Selection Operator). The selected SNPs groups compose a sparse feature space which yields a higher predictive power for continuous trait prediction.

We conduct experiment on SiMES (Singapore Malay Eye Study) data set, with 3280 Malay individuals genotyped on Illumina 610 quad arrays. We investigate one discrete trait (Glaucoma) and two glaucoma-related quantitative traits, optic Disc-Cup-Ratio (CDR) and Intra-ocular Pressure (IOP). The hypothesis is that, with more biological knowledge embedded, a learning mechanism yields higher predictive power. Our preliminary results support the above hypothesis. Further analysis reveals that our approach can identify groups of SNPs highly associated with a particular genetic trait, in spite of the small sample size and the incomplete biological knowledge.

4.1 Genome Wide Association Study

Genome Wide Association Study (GWAS) uses high-throughput genotyping technologies to assay hundreds of thousands of single nucleotide polymorphisms (SNPs) and associate them with the phenotype of interest; the identified SNPs are often used as genetic markers to identify the causes and access risks of disease. In a typical GWAS data set, the number of SNPs (usually $> 500k$) far exceeds the number of sampled individuals ($< 10k$) by at least 50 fold. Computational methods have been investigated for SNP-trait association study due to its large number. Pioneer works [241, 242] focus on detecting statistically significant (e.g., with marginal effect) SNPs associated with a trait. Recently, efforts [243–245] have been expanded to investigate those SNPs with little effects on disease risk individually but influence the disease risk jointly, which is known as epistatic interaction in genetic analysis where the effects of one gene are believed to be modified by one or several other genes.

The single-locus and epistasis SNP detection based algorithms test individual SNPs or pair of SNPs without taking into consideration of the underline biological intertwining mechanism, whereas, the real gene-gene interaction participating in biological pathway are often composed by a group of SNPs with arbitrary numbers. However,

to date, exhaustively detecting significant SNP groups of arbitrary size is still computational infeasible.

In conventional GWAS, a significant statistical threshold is often set and only candidates passing the threshold are follow up. However, since SNPs are often correlated via linkage disequilibrium, the M most significant individual SNPs identified by simple linear regression may not constitute an optimal set for following up. It's more important to find a small group of N potent but interwinely correlated SNPs (some of them may not pass the stringent threshold by themselves) for following up study. In machine learning, such problem is classified as feature selection issue and regression methods are often used to tackle the challenges. However, a normal forward and backward stepwise regression cant address the sparse and correlated nature of genetic analysis. We explore penalized regression for its powerful engine and ability to perform continuous model selection when compared to the conventional regression approaches. It is also computationally suitable for large data analysis and adapts readily to the interactions of group members.

4.2 Sparse Learning in Genome Wide Association Study

Penalized regression based on the Least Absolute Shrinkage Selector Operation (lasso) [246] was recently explored for GWAS analysis. Several lasso based approaches for GWAS analysis have been proposed. Some researchers [247, 248] proposed 2-step approaches for Genome-wide association analysis via shortlisting a group of marginal predictors using penalized likelihood maximization for further higher order interaction detection. Hoggart and others [249] proposed a method to simultaneously analyze all SNPs in genome-wide and re-sequencing association studies. D'Angelo [250] com-

bined lasso and principal-components analysis for detection of gene-gene interactions in genome-wide association studies. These approaches are not global due to the 2-stage process and none of them have considered incorporating prior knowledge into the model building.

Prior knowledge can be combined into GWAS to improve the power of association study [251], it can also model dependencies and moderate the curse of dimensionality. In this study, we propose a holistic approach to identify groups of predictive SNPs in preliminary GWAS analysis. Our method translates prior knowledge of proteomics and biological pathways into SNP groups; then applies linear regression regularized by group sparse constraint to select a small number of most predictive SNP groups, and deploys the group-lasso as the solver for the regularized linear regression.

4.3 Methodology

4.3.1 Dataset

The presented work is based on data collected in a population based study, Singapore Malay Eye Study (SiMES) [17]. SiMES is a large-scale population based study to assess the causes and risk factors of blindness and visual impairment in Singapore Malay community, conducted over a 3 year period from 2004 to 2007 by Singapore Eye Research Institute and funded by the National Medical Research Council. A total of 3,280 individuals comprising Malay adults aged between 40 to 80 are genotyped on the Illumina 610 quad arrays. We only analyze the autosomal SNPs, and conduct a stringent quality control procedure and the final set contains 2,542 individuals with 557,824 SNPs on 22 autosomal chromosomes. Clinical data collected for SiMES covers diagnosis information and various measurement of optic parameters. We choose three ocular traits because their significance in ophthalmological study and their heritability

presented in previous research [312–315]. The discrete trait includes 121 glaucoma subjects and 2421 normal subjects. Two quantitative traits are optic Cup-to-Disc ratio (CDR), ranged from $0.08 \sim 1.0$ and Intra-ocular pressure (IOP), ranged from $6.0 \sim 73$ mmHg. The optic disc is the anatomical location of the eye's "blind spot", the area where the optic nerve and blood vessels enter the retina. CDR is a measurement used in ophthalmology and optometry to assess the progression of glaucoma, which produces additional pathological cupping of the optic disc due to an increase in Intra-ocular Pressure (IOP). Single-SNP based GWAS has been reported previously for CDR [313, 314] and IOP [315].

4.3.2 Knowledge-based SNP grouping

We group SNPs into cascading layers of functional units, as illustrated in Figure 4.1. The first layer is individual SNPs; the second layer contains groups of SNPs located in the same genetic region. We use dbSNP [316] to annotate gene related SNPs. SNPs fall in extron, intron and flanking area within 10K distance from a gene are composed into one group. The approach makes it possible to detect well-annotated functional genes that are important to access the risk of interest genetic trait. The top-related SNPs in layer one are ranked by p-values obtained in basic association test as illustrated in Section 4.4. The *functional SNP groups* from layer two are selected by linear regression regularized by group sparse constraint which is illustrated in the Section 4.3.4. In this preliminary study, we compare the predictive power of top-related individual SNPs selected from the first layer and SNP groups selected from the second layer. In future work, we will further introduce the third layer, where SNPs involved in protein-protein interaction genes are grouped together; and the fourth layer where SNPs that occur in the genes participating in one particular biological pathway form a group.

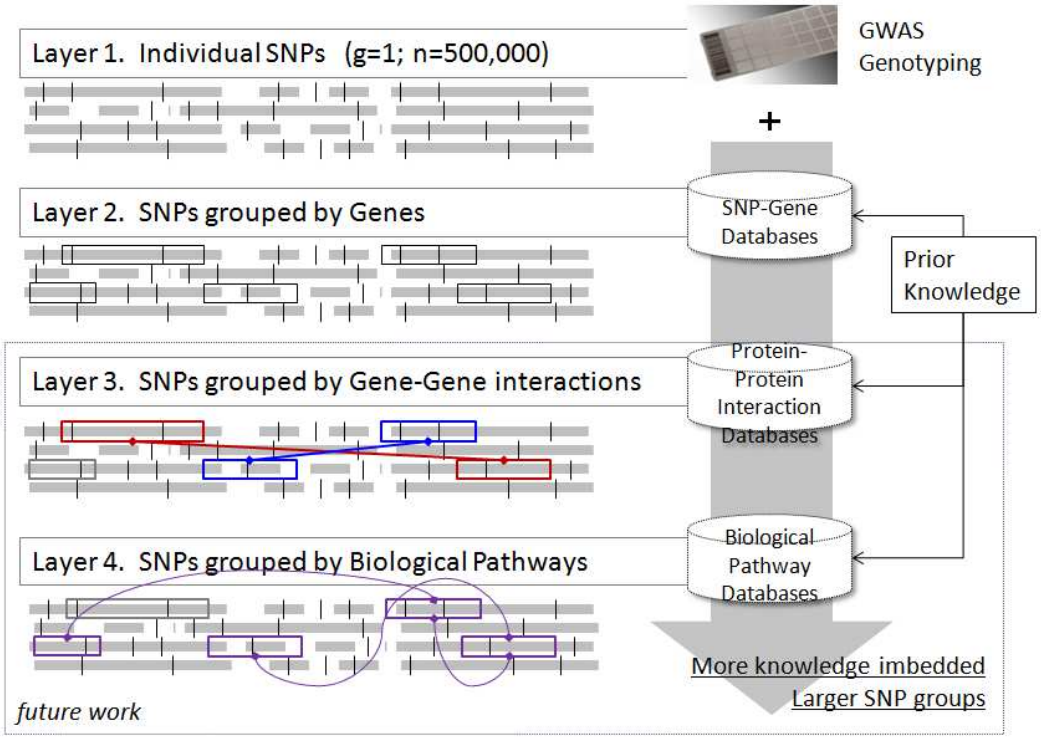


Fig. 4.1 A knowledge-based multi-layer SNP grouping mechanism

4.3.3 Linear SVR based continuous trait prediction

To predict IOP and CDR, which are continuous real values, from the very high dimensional SNP data, linear support vector regression(SVR) [311] is utilized for its efficiency; at the mean time, the accuracy can not be guaranteed since the feature dimension is significantly higher than the number of training samples [317]. To improve from its initial implementation for feature selection and dimensional reduction, $L2$-regularized linear SVR [311] is introduced. For a sample with SNP feature $\mathbf{x}_i$, corresponding to a regression value y_i (*i.e.*, CDR or IOP value), a weighting vector $\mathbf{w}$ is learned to predict the regression value using $\mathbf{w}^T\mathbf{x}_i + b$, by minimizing the following objective function:

$$\min_{\mathbf{w},b} \sum_{i=1}^{l} \|y_i - \mathbf{w}^T\mathbf{x}_i - b\|^2 + C\|\mathbf{w}|_2 \tag{4.1}$$

where the penalty factor coefficient C affects the trade-off between model complexity and non–separable samples.

4.3.4 SNP group selection Based on Group Sparsity Constraint

SNPs may not affect a particular trait individually, rather in a cooperative way which usually act in pairs or groups [318]. Thus, taking the relationship of SNPs in a *functional group* may lead to higher prediction accuracy, and more importantly useful biological insights. By inspecting the contribution of the various layers, we can also infer the risk is caused by genes, protein complexes or regulatory pathways. At the same time, identifying and using only the effective elements of the original features can bring about improvement in speed, and reduce computational cost.

For a sample represented by an original feature $\mathbf{x}_i$ consisting of g feature groups (we treat each gene as a group), we denote its regression value (*i.e.*, CDR or IOP value) as y_i. We adopt the linear regression model $\mathbf{w}^T\mathbf{x}_i + b$ to obtain the estimated value, where $\mathbf{w}$ is the weighting vector and b is the bias, and minimize the following objective function:

$$\min_{\mathbf{w},b} \sum_{i=1}^{l} \|y_i - \mathbf{w}^T\mathbf{x}_i - b\|^2 + \lambda \sum_{j=1}^{g} \|\mathbf{w}_j\|_2 \tag{4.2}$$

where $\mathbf{w}_j$ is the corresponding weight of the j^{th} feature group, g is the number of groups, l is the number of training samples and λ is used to control the sparsity of $\mathbf{w}$. In Equation (4.2), the first term represents the regression error and the second term is a $L_{1,2}$-norm based regularizer to enforce the group sparsity. Considering the features are intrinsically organized in groups, we use an $L_{1,2}$-norm based regularizer to select features from only a sparse set of groups. In the experiments, we use the group-lasso method provided in SLEP toolbox [319] to solve Equation (4.2).

After $\mathbf{w}$ is obtained, it can be used as a feature selection mask to generate the final features, *i.e.*, the j^{th} group of features is selected when $\|\mathbf{w}_j\|_2 > 0$. Usually, the selected feature has much lower dimension than the original feature, and thus the subsequent prediction can be greatly speed up and the memory storage can also be reduced significantly.

Comparing Equation (4.1) with (4.2), one can observe that Equation (4.1) is a special case of Equation (4.2), in which all features are considered in a unique group; while in real cases, such high dimensional features are naturally grouped into many groups according to the functionality. In our study, the SNPs fall in the same genes, proteins or biological pathways will perform biological function in different groups, the group sparse model we deployed using Equation (4.2) allows each functional group play as an individual unit, the SNPs may present in or absent from the model as a group.

The convergence rate of $L2$–regularized linear SVR is at least linear [311]. With the efficient projection algorithm implemented by [319], the $L_{1,2}$–norm algorithm achieves the convergence rate of $O(1/k^2)$, where k denotes the iteration number.

4.4 Experiment and Result

Single SNP based analysis

We first perform basic association analysis on the three traits for all individual SNPs, using software package PLINK [241]. The test basically calculates chi-squared statistic for each SNP against the respective traits. The resulted p-values are illustrated in the Manhattan plot as shown in Figure 4.2. Subplot a, b, c are Manhattan plots for glaucoma trait, CDR trait and IOP trait respectively. We observe that, with default genome wide significant setting ($P < 10^{-8}$), three SNPs are identified as significant

SNPs for glaucoma. There is no significant SNP for CDR trait, and more than 20 significant SNPs found for IOP. We rank the SNPs by their individual p-value and selected the top ranked SNPs as related features to construct prediction model. Top 400 SNPs for CDR and IOP trait are selected respectively.

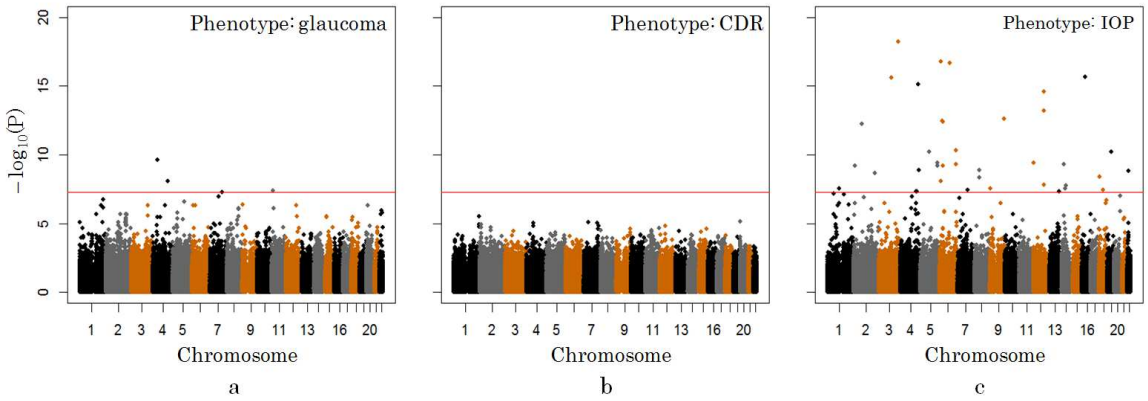


Fig. 4.2 Manhattan plot for basic association analysis. a.Glaucoma; b.CDR; C.IOP

4.4.1 Prediction based on selected SNPs

From the post QC GWAS data, we exclude samples with missing CDR or IOP values and focus on SNPs fall in genes or within 10K flanking area of genes (as a preliminary study). It results in 2531 valid samples with 246,123 SNPs. In the 2531 samples, 1265 are randomly selected for training, which cover the whole range of regression values (CDR and IOP), the rest 1266 sample are used for testing.

We select following three settings to build prediction model:

- Setting 1. The full SNPs feature set, contains about 246K SNPs.
- Setting 2. The high related feature set, 50-400 dimension, using top SNPs filtered from association test as mentioned in last section.
- Setting 3. Sparse group feature set, < 500 dimension, SNP grouping is composed using prior knowledge and selected by group sparsity constraint. The SNPs

grouping method is illustrated in Figure 4.1, in layer 2 the grouping unit is gene, layer 3 the group composes SNPs from a pair of interacting genes, and in layer 4 each group contains genes involved in a particular biological pathway. In this preliminary study, we focus on layer 2 grouping.

To compare the prediction power of different SNP sets, we build the linear regression model for CDR prediction and IOP prediction based on the three setting. The regression learning model of Setting 1 and 2 use Equation (4.1) and Setting 3 uses Equation (4.2). For fair comparison, optimal parameters are obtained with cross-validation on the training samples for each method. The parameter tuning are conducted as following:

- The regularizer coefficient C for linear SVR is set as
  $C \in \{0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000\}$
- The related feature set of Setting 2 is composed by top SNPs selected based on association test (as described in last section). 50, 100, 200 and 400 dimension top related features (SNPs) are tested
- For Setting 3, the group sparsity regularizer coefficient is set as
  $\lambda \in \{2000, 4000, 8000, 16000\}$

4.4.2 Comparison of prediction power based on three feature sets

The regression results are evaluated by lowest average error obtained in each experiment, as listed in Table 1. In Setting 1, regression model is built on a feature space composed by all SNPs. We use the result of this setting as baseline to evaluate other Settings. Using all SNPs for learning introduces several issues. Firstly the sample size is too small compare to feature dimension and the learning can be easily resulted

in overfitting during the training. We observe from Table 1 that training error is rather small (1.46% for CDR and 0.27% for IOP) but the testing error is much larger (12.62% for CDR and 3.94% for IOP). Moreover, the learning process consumes substantial memory and computational time. As most of the features are actually noise for the learning, in Setting 2, the feature space is composed by only the statistically most relevant SNPs, e.g., the SNPs with the lowest p-value. The regression result of Setting 2 leads to a poorer performance as compare to baseline, the reason can be the information loss in the whole genome context. In Setting 3, we use SNP groups to compose feature space, each group implies a functional unit in biological context which allows SNPs in the same group jointly affect the trait. In both CDR / IOP cases the best testing performance are achieved in the Setting 3. The relative error reduction ratio (as compare to baseline) is 3.24% for *CDR-Setting 3* and 17.26% for *IOP-Setting 3*.

The prediction performance against number of selected SNPs (feature dimension) is illustrated in Figure 4.3. For CDR prediction, setting 2 with top 50 related SNPs yield good result, but more SNPs only introduces noises. For IOP prediction, the top related 50 SNPs do not yield optimal result and extra dimension gives better performance. The group SNPs (Setting 3) modeling outperforms Setting 2 with a reasonable dimension introduced.

L2-regularized linear SVR defined by Equation (4.1) models a situation where features are consider as individual factors. In Equation (4.1), parameter C determines the trade-off between the training error and model complexity. Increasing C allows a more complex model, or more SNPs selected. On the other hand, L1,2-norm based regularized learning defined in Equation (4.2) models the situation where the features work in groups jointly affecting the outcome. In Equation (4.2) one can adjust λ to control the group sparsity, a larger λ tends to reduce the number of selected groups

Table 4.1 Optimal regression results for CDR and IOP

Experiment		Dimension (optimal)	Training error (%)	Testing error (%)	Relative error reduction ratio (%)
CDR	Setting 1	All SNPs	1.46	12.62	baseline
	Setting 2	50 SNPs	12.20	12.76	-1.1
	Setting 3	290 SNPs	12.14	**12.21**	**3.24**
IOP	Setting 1	All SNPs	0.27	3.94	baseline
	Setting 2	400 SNPs	2.89	4.14	-5.1
	Setting 3	381 SNPs	3.25	**3.26**	**17.26**

and a smaller λ would allow more groups being selected. Accordingly, the x-dimension of Figure 4.2 can be interpreted as the direction of decreasing λ and increasing C, both yielding a larger number of selected SNPs.

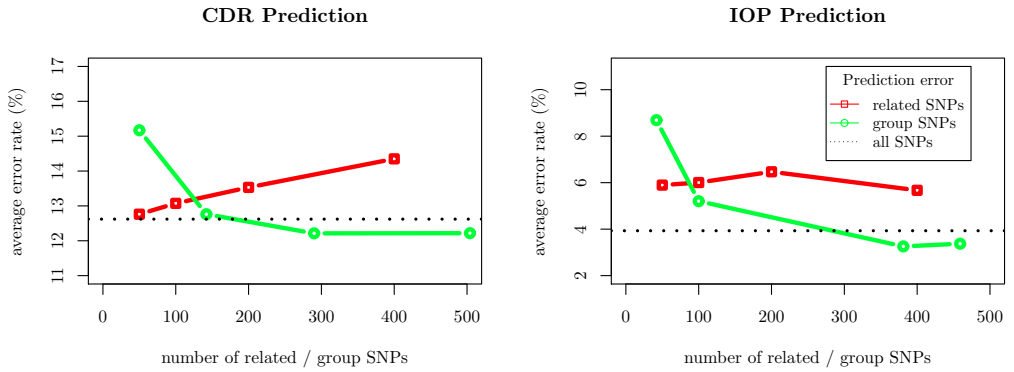


Fig. 4.3 Prediction error rate on different number of SNPs. Setting 1 - all SNPs; Setting 2 - related SNPs; Setting 3 - group SNPs

To explore the underline biological function of the selected SNPs, we further analyze the 290 SNPs identified for CDR prediction in Setting 3. We match the SNPs back to genetic region, and look for a list of keywords in their gene page from NCBI

gene database [316]. The relevant keyword list includes: Glaucoma, ocular, optic, macular, ciliary and retinal. From 290 SNPs, evidence shows that at least 21 genes are ocular-related, which provides a strong support for our approach.

Table 4.2 Ocular related genes identified from selected SNPs

SNP Ids	Gene Symbol	Ocular-related description
rs17851391	CCNL2	Retinoblastoma
rs34315387	SCAMP3	Retinoic
rs10515929	SLC4A10	Glaucoma
rs17469794	LRP1B	Optic
rs17701917	SPAG16	Cilia
rs11916441	KAT2B	Retinoic acid
rs7433024	OTOL1	Maculae
rs7614429	SUMF1	Retinal
rs10071548	GPR98	Ocular ,Retinal
rs7733024	GPR98	Ocular ,Retinal
rs17847865	PPARD	Retinal
rs397576	BRD2	Retinoblastoma
rs10486537	BBS9	Macular, Retinal
rs11982601	DGKI	Retina
rs1078907	CHD7	Ocular
rs16913039	TYR	Ocular,Macular,Retina,Retinal
rs1362629	ZNF423	Retinoic
rs6500767	RBFOX1	Retinopathy
rs11651398	DNAH17	Ciliary
rs12449302	RNF135	Retinoic
rs12449302	NF1	Optic

4.5 Discussion

Sparse learning in high dimensional problems conducts feature selection naturally. The reduced feature space keeps the subset of features that are more relevant and informative. This leads to a simplified model for faster prediction and may improve prediction accuracy and model stability, as noise features have been removed from the model. More importantly, in this context, a small set of SNPs is desirable for further biological interpretation.

In a common GWAS, all SNPs are treated equally and assumed to work individually, which is often not true given our prior knowledge about candidate genes and biological pathways. The hypothesis is that, with more biological knowledge embedded, a learning mechanism yields higher predictive power. Our experiment results support the hypothesis. Our method differs from the previously proposed approaches [247–249] that we incorporate prior knowledge into SNP grouping prior and then use sparse learning to identify the groups, the object function is optimized by improving the predictive power. The grouping priors can carry information on genes, protein-protein interactions or biological pathways. Our experiment demonstrates that the sparse feature space composed by gene-related SNP groups possesses higher predictive power in learning as compare to the whole-genome single-SNP or top-related SNP feature space. We believe that, with more complete prior knowledge and larger sample size, the approach can expect better result.

We will continue our work on SNPs grouped by Gene-Gene interaction as well as pathway. The predicted SNP groups can be used as a preliminary genome-wide scanning for further replication and validation. It can be treated as a feature selection step in dimension reduction for the proposed AODI framework.

Chapter 5

A Unified Multiple Kernel Learning Framework for the Detection of Ocular Diseases Using Multiple Data Sources

In Chapter 3, we developed AODI-PM framework using traditional MKL method for PM detection. In this chapter, we propose AODI-MKL^{clm}, an extension to AODI-PM that explores the prior knowledge learned from each individual data source into the framework, e.g., incorporates pre-learned SVM classifiers trained from individual data source. These classifiers are used to generate prediction outputs (known as *classemes*) for the data from each domain. The *classemes* are utilized as additional features together with the original features to construct base kernels in AODI-MKL^{clm}, which helps improve the classification performance of the detection of each ocular disease. We validate our AODI-MKL^{clm} framework by conducting extensive experiments for three leading ocular diseases: Glaucoma, AMD and PM. Experimental results show that AODI-MKL^{clm} achieves AUCs of 0.935 ± 0.015, 0.822 ± 0.024 and 0.946 ± 0.010 as well as specificity of 0.853 ± 0.022, 0.732 ± 0.038 and 0.882 ± 0.024 at 85% sensitivity, for the three diseases respectively. Statistical test shows that AODI-MKL^{clm} is significantly better than the standard SVMs using data from individual domains as well as the traditional MKL method.

5.1 Background

Multiple Kernel Learning (MKL) techniques, in which a kernel is defined as a combination of multiple pre-defined base kernels, have been successfully applied to the fusion of multiple types of data features in some computer vision applications [291, 320, 321] such as in image classification and object recognition. In particular, these methods construct each pre-defined base kernel by utilizing one type of features and attempt to learn the combination coefficients of the base kernels through the maximum margin framework of MKL. With its demonstrated effectiveness, medical imaging researchers have investigated the use of MKL in the medical domain and reported several interesting results [73, 322–324]. Hinrichs *et al.*[323] tried to fuse multiple modalities (e.g., magnetic resonance and positron emission tomography) of image data for Alzheimer's disease classification. Gál *et al.*[322] proposed to combine both visual features and meta textual information from medical images to determine their modality. More recently, Liu *et al.*[324] applied MKL for automatic glaucoma detection by fusing heterogeneous data from three data sources (i.e., personal clinic data, genome information and ocular fundus images). Also Zhang [73] employed the same strategy for the assessment of pathological myopia. The promising results reported in [73, 324] demonstrates good potential for the use of machine learning methods clinically.

MKL-based methods are able to fuse information from different data sources and can be used to detect ocular diseases. However, we argue that the generalization ability of existing MKL methods [73, 324] is not optimal, as some of the the prior knowledge which can add more discriminative power to the methods has not been fully utilized. As shown in [325], the prediction outputs (a.k.a., *classemes*) of data from Support Vector Machine (SVM) classifiers, which encode the discriminative information of original data features, can help improve the classification performance. Therefore, we

propose to incorporate *pre-learned* SVM classifiers to generate classemes in this work, which can enhance the discriminative ability of the predictive model.

Motivated by MKL [73, 324] and classemes [325], we develop a unified MKL framework called AODI-MKL^{clm} to automatically detect ocular diseases by effectively fusing personal demographic/clinical data, genome information and visual information from retinal fundus images through the incorporation of pre-learned SVM classifiers. In AODI-MKL^{clm}, each pre-learned SVM classifier is trained using data from one source. The prediction outputs (i.e., classemes) of the data are later obtained from these pre-learned classifiers. The contributions of AODI-MKL^{clm} are two-fold: 1) The classemes are effectively integrated in the framework to augment the original data features, which helps boost the classification performance of the detection of each ocular disease. 2) AODI-MKL^{clm} is the first reported unified framework that can be applied on three leading ocular diseases (i.e., glaucoma, AMD and PM). Further, it can be readily applied to the detection of other ocular diseases.

5.2 Data Processing and Feature Extraction

For genetic data, we employ the same SNP selection process described in Section 3.3.1 to get SNPs associated with the three ocular diseases (glaucoma, AMD and PM). The data cleaning process for demographic and clinical data are described in Section 3.3.2.

5.2.1 Retinal Fundus Images Processing

In order to assess the retinal fundus images, it is important for us to convert every image into a meaningful numerical representation which can retain the image properties such as intensity and scale.

For glaucoma detection, we first utilize the optic cup-to-disc ratio [326] as a reference parameter. This measurement is clinically used for glaucoma detection, which can be costly to obtain from manual delineation of the optic disc and cup. Inspired by the ranking based automatic cup localization method in [327], we extract similar histogram-based features from each automatically segmented optic disc by using the active shape model (ASM) method [328], which then forms the feature vector for this fundus image. The details of this procedure are described as follows:

- Each optic disc is normalized to a circle with a diameter of 256 pixels, with the cup and disc centers aligned. Virtual cups with CDR values $\rho \in \{0.2, 0.4, 0.6, 0.8\}$ are then generated.

- Histograms are obtained from the green channel with different bin numbers $\beta \in \{8, 16, 32, 64\}$ such that each bin has an equal (or as equal as possible due to quantization) number of pixels.

- For each β and each virtual cup with CDR ρ, we obtain two types of features: 1) L1-normalized histogram of the virtual cup region and 2) L1-normalized histogram of the virtual rim (i.e., non-cup) region within the disc.

- The final representation of each fundus image is obtained by concatenating the above two types of features over four CDR values and four bin numbers into a lengthy feature vector with its number of dimensions as $2 \times 4 \times (8+16+32+64) = 960$.

For AMD detection, we propose to represent each fundus image by using the bag-of-words model, because of its good performance [73]. Specifically in this work, we first extract local features by using the Difference-of-Gaussian detector and the Scale-Invariant Feature Transform (SIFT) descriptor [329], as SIFT has been widely used and has shown promising results in many applications such as object recognition and

classification. Note that the feature extraction is conducted on the green channel of images only, since the retinal images are less differentiable in other channels (blue and red). Given SIFT features extracted from all the fundus images, we perform k-means clustering to group them into 1000 clusters. The centroid of each cluster represents a visual word, thus a codebook of 1000 visual words is generated. Finally, for the SIFT features in each image, we quantize them into 1000 bins, and then each image is represented as a 1000-bin histogram with each bin value as the frequency of SIFT features having the same visual word. We use the same way to generate the bag-of-words representation of the fundus images for PM detection.

In our framework, feature normalization is performed for each of the features from the personal demographic/clinical data, the genetic features and the retinal image features. Specifically, for each feature dimension, the corresponding value is normalized to the range of $[0, 1]$ to refrain possible bias among feature dimensions due to magnitude differences among feature values.

5.3 A Unified MKL Framework by Using Multiple data sources

As the information from each data source contains a different perspective in terms of the disease risk of patients, addressing the heterogeneity of different perspectives to make effective use of the discriminative domain knowledge becomes very important for automatic disease detection. Therefore, our goal is to design a novel framework that can effectively fuse the heterogeneous data from different data sources in order to learn robust ocular diseases prediction models.

Motivated by the previous work [73, 324], we propose an improved Multiple Kernel Learning (MKL) framework by constructing base kernels with the data from each het-

erogeneous informatics domain. Moreover, our AODI-MKL^{clm} framework is also able to incorporate existing pre-learned classifiers (e.g., binary SVM classifiers) which are considered as a useful prior that can be leveraged to further boost the classification performance. The flowchart of AODI-MKL^{clm} is illustrated in Figuure 5.1. Specifically, for a particular ocular disease, we first learn a binary SVM classifier with the training data from each data source. After that, each base kernel in our AODI-MKL^{clm} framework will be constructed from each data source and by using *augmented* features which is a concatenation of the decision values (referred to as *classemes* [325]) from the pre-learned classifiers as well as original features of the data from this data source. Finally, a predictive model will be learned based on the constructed base kernels in our framework and will be used for the automatic detection of each ocular disease.

5.3.1 Proposed Formulation by Using Pre-learned Classifiers

In traditional Multiple Kernel Learning (MKL) [278], it usually assumes that the learned kernel is a combination of a set of pre-computed base kernels. The utilization of multiple base kernels enhances the interpretability of MKL over other kernel methods (e.g., SVMs [261]) which use only one single kernel. Because of the demonstrated effectiveness of MKL for fusing different types of features [73, 291], we propose a unified framework based on MKL for the automatic detection of different ocular diseases. In our framework, each base kernel is constructed based on one type of data features.

Recently, the outputs of pre-learned class-specific classifiers (i.e., binary SVM classifiers learned for each class) have been effectively employed as additional features to help improve classification performance [325, 330]. Motivated by that, we make use of pre-learned SVM classifiers in our framework. Specifically, we train one pre-learned SVM classifier in advance by using the training data from each data source. We then use such a classifier learned from one domain to obtain a decision value (i.e., classeme)

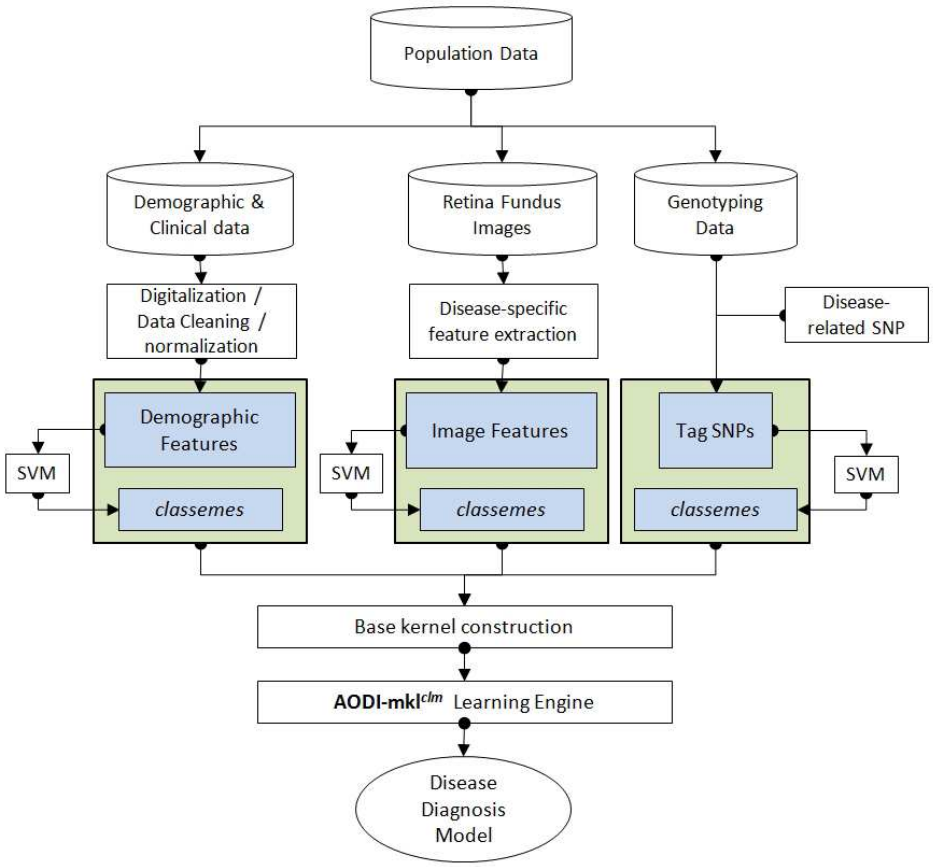


Fig. 5.1 Overview of AODI-MKL^{clm} framework using heterogeneous data from multiple data sources for the automatic detection of ocular diseases.

for each data sample from the same domain. Assuming there are P data sources (e.g., $P = 3$ in this work), we then have P pre-learned classifiers, and also P classemes will be generated for each patient x. It is worth noting that a classeme $f_p(\mathbf{x})$ of each $\mathbf{x}$ actually encodes the information from the corresponding (i.e., p-th) data source, and thus it can be a useful prior which helps to further boost the classification performance of our proposed framework. Also, the value of each classeme $f_p(\mathbf{x})$ can be interpreted as the probability of the presence of the ocular disease in the individual.[1].

In this work, we aim to learn a binary classifier by using multiple types of features from different data sources for the automatic detection of an ocular disease. Suppose

[1]The classeme value can be passed into some sigmoid function to produce a probability value within the range of $[0, 1]$.

we have M types of features (each represented as $\mathbf{x}_i^{\{m\}}$) for a patient $\mathbf{x}_i$, and we concatenate them into one lengthy feature $\mathbf{x}_i = [{\mathbf{x}_i^{\{1\}}}^\top, \ldots, {\mathbf{x}_i^{\{M\}}}^\top]^\top$. Let us denote $\{(\mathbf{x}_i, y_i)|_{i=1}^n\}$ as the set of all the training samples, where $y_i \in \{-1, 1\}$ is the class label. Here, $y_i = 1$ means that a patient $\mathbf{x}_i$ has such an ocular disease and $y_i = -1$ otherwise. Moreover, we denote the set of pre-learned classifiers by $\{f_p|_{p=1}^P\}$, each of which is learned from the standard SVM using the RBF kernel with one type of feature (see the experimental setup in Section 5.4.2 for more details). Moreover, for each patient $\mathbf{x}_i$, we augment its original feature with the classemes from P pre-learned classifiers, i.e., $\mathbf{f}_i = [f_1(\mathbf{x}_i), \ldots, f_P(\mathbf{x}_i)]^\top$, which leads to the final augmented feature representation $\mathbf{z}_i = [{\mathbf{x}_i}^\top, \mathbf{f}_i^\top]^\top$. By respectively introducing nonlinear feature mapping functions $\{\phi_m(\cdot)|_{m=1}^M\}$ and $\varphi(\cdot)$ for $\mathbf{x}_i^{\{m\}}$'s and $\mathbf{f}_i$, we map $\mathbf{z}_i$ from its original feature space into a higher dimensional space as $\psi(\mathbf{z}_i) = [\phi_1(\mathbf{x}_i^{\{1\}})^\top, \ldots, \phi_M(\mathbf{x}_i^{\{M\}})^\top, \varphi(\mathbf{f}_i)^\top]^\top$. With the above notations, we define the following parametric function of the classifier for our classification problem based on MKL and classemes:

$$\begin{aligned} f(\mathbf{z}_i) &= \mathbf{w}^\top \psi(\mathbf{z}_i) + b \\ &= \sum_{m=1}^{M} \mathbf{w}_m^\top \phi_m(\mathbf{x}_i^{\{m\}}) + \mathbf{w}_f^\top \varphi(\mathbf{f}_i) + b, \end{aligned} \tag{5.1}$$

where $\mathbf{w} = [\mathbf{w}_1^\top, \ldots, \mathbf{w}_M^\top, \mathbf{w}_f^\top]^\top$, $\mathbf{w}_m$ and $\mathbf{w}_f$ are the weights for the m-th feature and the classemes, respectively, and b is the bias term. We then formally present the formulation of our proposed AODI-MKL^{clm} framework by using the hinge loss to model the empirical loss as follows:

$$\min_{\substack{\mathbf{w}_m, \mathbf{w}_f, \\ \mathbf{d}, b, \xi_i}} \frac{1}{2}\left(\sum_{m=1}^{M} \frac{\|\mathbf{w}_m\|^2}{d_m} + \mu\|\mathbf{w}_f\|^2 + \theta\|\mathbf{d}\|^2\right) + C\sum_{i=1}^{n}\xi_i, \tag{5.2}$$

$$\text{s.t.} \quad y_i f(\mathbf{z}_i) \geq 1 - \xi_i,\ \xi_i \geq 0,\ i = 1, \ldots, n; \tag{5.3}$$

$$\sum_{m=1}^{M} d_m = 1,\ d_m \geq 0,\ m = 1, \ldots, M,$$

where $\mathbf{d} = [d_1, \ldots, d_M]^\top$ is a weight vector in which d_m controls the square norm of $\mathbf{w}_m$ [278], and ξ_i's are slack variables for the positive and negative training samples, respectively; $\mu, \theta > 0$ are pre-defined parameters which respectively regularize the complexity of $\mathbf{w}_f$ and $\mathbf{d}$, $C > 0$ is pre-defined as a trade-off parameter to balance the regularization on the weight variables and the averaged empirical errors. Note that we normalize the empirical errors on the positive and negative data in order to avoid the imbalance problem. We also note that the value of μ is to control the effect of prior information (i.e., classemes) in AODI-MKL^{clm}. Intuitively, when μ is small, $\mathbf{w}_f$ is encouraged be large, which means the prior information will have strong impact on the learning of the classifier. On the other hand, when μ is large, the prior information will have less impact.

To solve the optimization problem in Equation (5.2), we first introduce dual variables $\alpha_i \geq 0$ and $\beta_i \geq 0$ for each of the constraints in Equation (5.3). The Lagrangian of the optimization problem in Equation (5.2) can be obtained as follows:

$$\begin{aligned} L = \frac{1}{2}\left(\sum_{m=1}^{M}\frac{\|\mathbf{w}_m\|^2}{d_m} + \mu\|\mathbf{w}_f\|^2 + \theta\|\mathbf{d}\|^2\right) + C\sum_{i=1}^{n}\xi_i \\ - \sum_{i=1}^{n}\alpha_i\left(y_i f(\mathbf{z}_i) - 1 + \xi_i\right) - \sum_{i=1}^{n}\beta_i\xi_i \end{aligned}$$

By setting the derivatives of the above Lagrangian to zeros with respect to the primal variables $\mathbf{w}_m, \mathbf{w}_f, b$ and ξ_i (except for $\mathbf{d}$), we can get:

$$\mathbf{w}_m = d_m\sum_{i=1}^{n}\alpha_i\phi_m(\mathbf{x}_i^{\{m\}}),\ \mathbf{w}_f = \sum_{i=1}^{n}\alpha_i\varphi(\mathbf{f}_i),\ \sum_{i=1}^{n}\alpha_i y_i = 0;$$

$$\alpha_i + \beta_i = C,\ i = 1, \ldots, n;$$

After substituting the above equations back into the Lagrangian, we can obtain the dual problem. Therefore, the optimization problem (5.2) can be equivalently rewritten

as follows:

$$\min_{\mathbf{d}\in\mathcal{D}} \max_{\boldsymbol{\alpha}\in\mathcal{A}} G(\mathbf{d},\boldsymbol{\alpha}) = \frac{\theta}{2}\|\mathbf{d}\|^2 + \mathbf{1}^\top\boldsymbol{\alpha} - \frac{1}{2}(\boldsymbol{\alpha}\circ\mathbf{y})^\top\tilde{\mathbf{K}}(\boldsymbol{\alpha}\circ\mathbf{y}), \tag{5.4}$$

where $\mathcal{D} = \{\mathbf{d} \mid \sum_{m=1}^{M} d_m = 1,\ d_m \geq 0\}$ is the feasible set of $\mathbf{d}$, $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_n]^\top$ is a vector of the dual variables, $\mathcal{A} = \{\boldsymbol{\alpha} \mid \boldsymbol{\alpha}^\top\mathbf{y} = 0,\ 0 \leq \alpha_i \leq C\}$ is the feasible set of $\boldsymbol{\alpha}$, $\mathbf{y} = [y_1, \ldots, y_n]^\top$ is a vector of the training labels, the notation $\circ$ represents the operator of element-wise product, and $\tilde{\mathbf{K}} \in \mathbb{R}^{n\times n}$ is the kernel matrix with each element defined as below:

$$\tilde{k}(\mathbf{z}_i, \mathbf{z}_j) = \sum_{m=1}^{M} d_m k_m(\mathbf{x}_i^{\{m\}}, \mathbf{x}_j^{\{m\}}) + \frac{1}{\mu}\cdot k_f(\mathbf{f}_i, \mathbf{f}_j), \tag{5.5}$$

where $k_m(\cdot,\cdot)$ is a pre-defined base kernel function for the m-th type of features, which is induced by the nonlinear feature mapping $\phi_m(\cdot)$, i.e., $k_m(\mathbf{x}_i^{\{m\}}, \mathbf{x}_j^{\{m\}}) = \phi_m(\mathbf{x}_i^{\{m\}})^\top\phi_m(\mathbf{x}_j^{\{m\}})$; and $k_f(\cdot,\cdot)$ is also a kernel function for classemes, induced by the feature mapping $\varphi(\cdot)$, i.e., $k_f(\mathbf{f}_i, \mathbf{f}_j) = \varphi(\mathbf{f}_i)^\top\varphi(\mathbf{f}_j)$.

Similar to [278], we develop an alternating update procedure to solve the optimization problem in Equation (5.4) by iteratively optimizing $\mathbf{d}$ and $\boldsymbol{\alpha}$. First, given fixed $\mathbf{d}$, Equation (5.4) is reduced to a standard SVM problem with respect to $\boldsymbol{\alpha}$ and can be effectively solved by using existing software such as LIBSVM [331]. After that, we fix $\boldsymbol{\alpha}$ and solve the following quadratic programming problem of $\mathbf{d}$:

$$\min_{\mathbf{d}\in\mathcal{D}} \quad \frac{\theta}{2}\|\mathbf{d}\|^2 - \mathbf{h}^\top\mathbf{d}, \tag{5.6}$$

where $\mathbf{h} = [h_1, \ldots, h_M]^\top$ is a vector with each $h_m = \frac{1}{2}(\boldsymbol{\alpha}\circ\mathbf{y})^\top\tilde{\mathbf{K}}_m(\boldsymbol{\alpha}\circ\mathbf{y})$, and $\tilde{\mathbf{K}}_m$ is a modified base kernel with each element defined as $\tilde{k}_m(\mathbf{z}_i, \mathbf{z}_j) = k_m(\mathbf{x}_i^{\{m\}}, \mathbf{x}_j^{\{m\}}) + \frac{1}{\mu}\cdot k_f(\mathbf{f}_i, \mathbf{f}_j)$. We iterate the above alternating update procedure until meeting the stopping criterion that the value of the objective function in (5.4) converges or the

Input: labeled training samples $\{(\mathbf{x}_i, y_i)|_{i=1}^n\}$, pre-learned classifiers $\left\{f_p(\mathbf{x})|_{p=1}^P\right\}$ and pre-defined base kernel functions $\left\{k_m(\cdot,\cdot)|_{m=1}^M\right\}$
Initialization: $t \leftarrow 1$ and $\mathbf{d}_t \leftarrow \frac{1}{M}\mathbf{1}_M$
With fixed $\mathbf{d}_t$, update $\boldsymbol{\alpha}_t$ in a standard SVM problem (reduced from Equation (5.4)).
while $t < T_{\max}$ **do**
  With fixed $\boldsymbol{\alpha}_t$, update $\mathbf{d}_{t+1}$ in the quadratic programming problem (5.6).
  With fixed $\mathbf{d}_{t+1}$, update $\boldsymbol{\alpha}_{t+1}$ in a standard SVM problem (reduced from (5.4)).
  if $|G(\mathbf{d}_{t+1}, \boldsymbol{\alpha}_{t+1}) - G(\mathbf{d}_t, \boldsymbol{\alpha}_t)| \leq 10^{-5}$ **then**
    break;
  end
  $t \leftarrow t + 1$
end
Output: $\mathbf{d}_{t+1}$ and $\boldsymbol{\alpha}_{t+1}$

Algorithm 1: The alternating update procedure of AODI-MKL^{clm}

maximal number of iterations has been reached. The optimization procedure is detailed in Algorithm 1.

After obtaining the optimal $\mathbf{d}$ and $\boldsymbol{\alpha}$, we can rewrite the classifier of AODI-MKL^{clm} in Equation (5.1) as below and use it to perform the automatic detection on any test patient:

$$f(\mathbf{z}) = \sum_{i=1}^{n} \alpha_i y_i \left(\sum_{m=1}^{M} d_m k_m(\mathbf{x}_i^{\{m\}}, \mathbf{x}^{\{m\}}) + \frac{1}{\mu} \sum_{p=1}^{P} k_f(\mathbf{f}_i, \mathbf{f}) \right) + b,$$

where $\mathbf{z}$ is the augmented feature representation of a test patient, $\mathbf{x}^{\{m\}}$ is its m-th feature, and $\mathbf{f} = [f_1(\mathbf{x}), \ldots, f_P(\mathbf{x})]^\top$ is its classemes obtained from the pre-learned classifiers.

Discussions: 1) It is worth noting that a variety of methods (e.g., Support Vector Machines [261], Support Vector Regressions [332], AdaBoost [333], etc.) can be readily used as pre-learned classifiers in our framework in order to generate classemes for the data. 2) Pre-learned classifiers not only can be trained by using different learning methods, but also by using different types of features and kernels. Thus, the total number of pre-learned classifiers can be larger than that of different data sources or different types of features. But note that in our experiments, for each data source, we assume to have only one feature type, and one pre-learned classifier is trained by

using SVM with one RBF kernel constructed by the training data from this domain. So the total number of pre-learned classifiers is the same as that of the feature types (i.e., $P = M$) in this work.

5.4 Experiments

To thoroughly evaluate the performance of our proposed AODI-MKL^{clm} framework, we conducted extensive experiments for each of the three leading ocular diseases: Glaucoma, AMD and PM. We also consider using different types of features as well as their different combinations in the experiments, which gives us a total of seven settings: 1) demographic/clinic data (**D**); 2) genome data (**G**); 3) retinal image data (**I**); 4) **D+G**; 5) **D+I**; 6) **G+I**; 7) **D+G+I**. The notation '+' indicates a combination of two or three types of features in the corresponding setting.

5.4.1 Dataset Description

The detailed data description can be found in Chapter 3. Among 2,258 subjects obtained from SiMES [17], there are 100 with glaucoma, 122 with AMD and 58 with PM. The distribution of the subjects of each disease in the dataset is representative of the disease prevalence in the population.

5.4.2 Experiment

In this work, we compare our proposed AODI-MKL^{clm} with two baseline methods (i.e., standard SVM and the MKL method in [324]) under three sets of experiments. Each set *independently* evaluates all the methods for the automatic detection of one ocular disease (i.e., glaucoma, AMD or PM) under the aforementioned seven settings using different feature types as well as their combinations. For the standard SVM which can

Table 5.1 Means and standard deviations of Area Under Curve (AUC) and Balanced Accuracy ($\bar{P}$) of all methods using different feature combinations for glaucoma detection

	Single kernel SVM		MKL		AODI-MKL^{clm}	
	AUC	$\bar{P}$	AUC	$\bar{P}$	AUC	$\bar{P}$
1) **D**	0.586±0.030	0.546±0.030				
2) **G**	0.900±0.017	0.805±0.025	*N.A.*	*N.A.*	*N.A.*	*N.A.*
3) **I**	0.789±0.026	0.716±0.020				
4) **D+G**	0.894±0.022	0.804±0.041	0.898±0.019	0.796±0.031	0.900±0.021	0.813±0.034
5) **D+I**	0.654±0.033	0.559±0.040	0.811±0.022	0.715±0.019	0.814±0.023	0.719±0.025
6) **G+I**	0.904±0.018	0.810±0.025	0.914±0.020	0.826±0.029	0.931±0.015	0.815±0.024
7) **D+G+I**	0.897±0.016	0.810±0.026	0.916±0.023	0.829±0.028	**0.935±0.015**	**0.853±0.022**

†Intraocular pressure (IOP) only achieves 0.609±0.035 for AUC and 0.573±0.034 for $\bar{P}$.

Table 5.2 Means and standard deviations of Area Under Curve (AUC) and Balanced Accuracy ($\bar{P}$) of all methods using different feature combinations for age-related macular degeneration (AMD) detection.

	Single kernel SVM		MKL		AODI-MKL^{clm}	
	AUC	$\bar{P}$	AUC	$\bar{P}$	AUC	$\bar{P}$
1) **D**	0.692±0.039	0.600±0.039				
2) **G**	0.728±0.025	0.642±0.028	*N.A.*	*N.A.*	*N.A.*	*N.A.*
3) **I**	0.761±0.024	0.660±0.026				
4) **D+G**	0.767±0.026	0.666±0.035	0.748±0.043	0.651±0.043	0.759±0.038	0.659±0.041
5) **D+I**	0.710±0.036	0.610±0.038	0.768±0.032	0.667±0.033	0.773±0.035	0.664±0.029
6) **G+I**	0.769±0.021	0.682±0.026	0.806±0.020	0.705±0.032	0.808±0.022	0.714±0.037
7) **D+G+I**	0.779±0.026	0.674±0.039	0.798±0.036	0.700±0.042	**0.822±0.024**	**0.732±0.038**

Table 5.3 Means and standard deviations of Area Under Curve (AUC) and Balanced Accuracy ($\bar{P}$) of all methods using different feature combinations for pathological myopia (PM) detection.

	Single kernel SVM		MKL		AODI-MKL^{clm}	
	AUC	$\bar{P}$	AUC	$\bar{P}$	AUC	$\bar{P}$
1) **D**	0.613±0.037	0.552±0.035				
2) **G**	0.776±0.036	0.697±0.040	*N.A.*	*N.A.*	*N.A.*	*N.A.*
3) **I**	0.931±0.031	0.850±0.043				
4) **D+G**	0.767±0.036	0.688±0.045	0.772±0.042	0.694±0.052	0.785±0.034	0.705±0.045
5) **D+I**	0.923±0.035	0.847±0.049	0.928±0.039	0.853±0.049	0.929±0.028	0.856±0.045
6) **G+I**	0.927±0.020	0.843±0.032	0.932±0.038	0.853±0.046	0.941±0.016	0.860±0.030
7) **D+G+I**	0.924±0.028	0.851±0.040	0.936±0.029	0.858±0.036	**0.946±0.010**	**0.882±0.012**

only take one single kernel, one SVM classifier is trained by using the corresponding training data under each setting[2]. It is worth mentioning that the MKL method in

[2]Note that under settings 4)–7), we concatenate two or three types of features (i.e., a combination of **D**, **G** or **I**) to form one lengthy feature vector which is used as the final feature representation for each subject.

[324] and our AODI-MKL^{clm} framework are designed to deal with multiple types of features. Thus, we evaluate the performances of both methods under settings 4)–7) where multiple types of features are available. Recall that our AODI-MKL^{clm} method incorporates pre-learned classifiers which generates classemes for the training subjects. As such, in the experiments, we take the learned classifiers from the baseline standard SVM method as the pre-learned classifiers for AODI-MKL^{clm}. Note that each subject has three classemes with each obtained from one data source, and we do not use concatenated features (as in settings 4)–7) for the standard SVM) to generate classemes for each subject.

We use the RBF kernel (i.e., $k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2\right)$) for all the methods. The kernel parameter γ is set as the reciprocal of the mean of all the distances between any two training subjects. For each method, we set the trade-off parameter for the empirical error as $C \in \{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$. Two-fold cross validation is performed to select the optimal parameter for each method. After selecting the optimal parameter through cross-validation for each method, all the training data are used again to train the final predictive model which will be used for the detection of each ocular disease.

For comprehensive comparison, we conduct ten rounds of independent tests in the experiments by randomly partitioning all the subjects into one training and one testing subset of equal size. Specifically, we randomly sample half of the positive and negative subjects. These randomly sampled subjects are defined as the initial training subset, and the remaining subjects are used as the test subset. Each independent test will be conducted twice by swapping the training and testing subsets. For fair comparison, all the methods use the same partition of training and testing subsets for each round of tests. As a result, a total of twenty testing results for each method will be obtained

under one of the seven settings. The aggregate of these results will then be used for performance analysis.

We follow the evaluation metric as in [324] to evaluate the classification performance of each method. In particular, we utilize the area under the Receiver Operating Characteristic curve (AUC) and balanced accuracy ($\bar{P}$) with a fixed specificity of 85%, where ROC is plotted as a curve that shows the tradeoff between sensitivity (P_+) and specificity (P_-), which are defined as follows:

$$P_+ = \frac{TP}{TP+FN}, \quad P_- = \frac{TN}{TN+FP}, \quad \bar{P} = \frac{P_+ + P_-}{2}, \tag{5.7}$$

where TP, TN, FP and FN denote the numbers of true positives, true negatives, false positives and false negatives, respectively.

5.4.3 Result Analysis

In the experiments, the performance of different methods are evaluated under each of the seven settings. The means and standard deviations of the AUC and $\bar{P}$ of all the methods over ten rounds of independent tests are shown in Tables 5.1, 5.3 and 5.2. Note that in Table 5.1 we also report the performance of the current clinical assessment method for glaucoma using intra-ocular pressure (IOP). From the results in the three tables, we have the following observations:

- For standard SVM with a single kernel, using a combination of two or three features under settings 4)–7) generally achieves comparable or slightly worse performance than using the corresponding single type of features under settings 1)–3), except for the result under setting 4) in Table 5.2. This observation shows that the simple feature concatenation strategy generally cannot improve the overall classification performance.

- From the results for glaucoma detection in Table 5.1, the standard SVM, the MKL method [324] and our proposed AODI-MKL^{clm} framework all perform better than the current clinic assessment intra-ocular pressure (IOP) method under almost all the settings in terms of both AUC and $\bar{P}$ (except for only one case of the standard SVM under setting 1) using the demographic/clinic data). Moreover, the three machine learning methods also obtain a large performance gain over IOP, when fundus image data and genome information are presented. These results clearly show that machine learning methods using additional fundus image data and genome information can potentially be used for clinical usage.

- When using multiple types of features under each of settings 4)–7) in all the three tables, the two MKL based methods consistently achieve better performances than the standard SVM using concatenated features except for setting 4) in Table 5.2 where our AODI-MKL^{clm} framework is just slightly worse than the standard SVM. Importantly, we also observe that when using a combination of all the three feature types under setting 7), the results of both the MKL based methods are better than the results of the standard SVM under all the settings. All these observations clearly demonstrate that the MKL based methods can learn better combination coefficients to effectively fuse the information from different data sources.

- With the utilization of classemes, our proposed AODI-MKL^{clm} framework consistently achieves a better performance than the MKL method [324], which demonstrates the usefulness of classemes as well as the effectiveness of our AODI-MKL^{clm} to seamlessly incorporate classemes into the MKL based framework. Moreover, by fusing all three types of features, statistical tests (T-test) show that the improvement of AODI-MKL^{clm} framework over the other methods under all the settings are significant

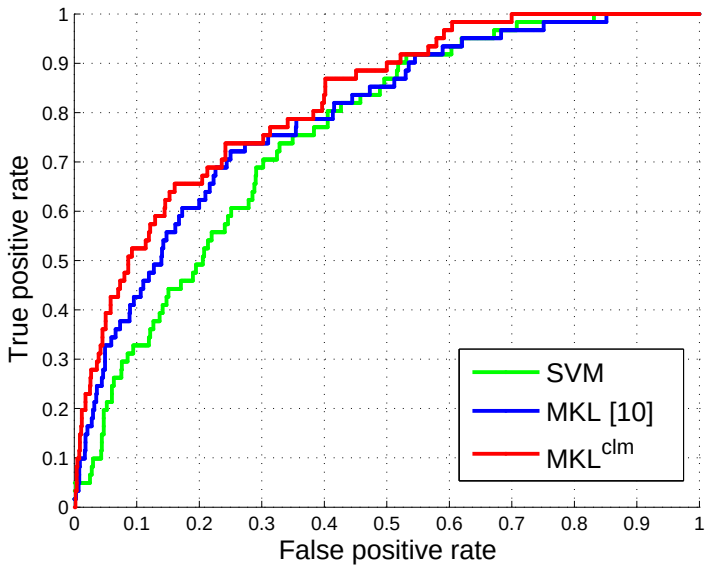


Fig. 5.2 ROC curves of all the methods for AMD detection.

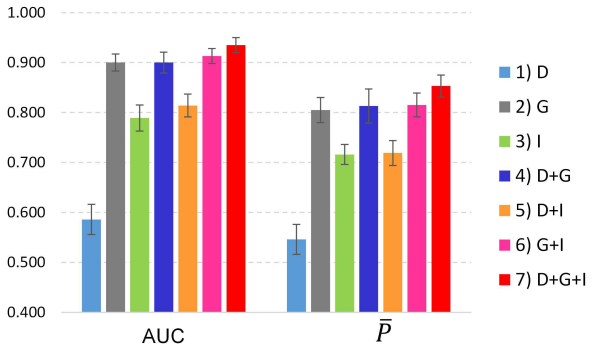
We also plot in Figure 5.2 the ROC curves of the three methods by using one round of individual test for AMD detection as a showcase. Specifically, the ROC curve of each method is plotted based on the setting when it achieves the best performance. In this case, setting 7) in Table 5.2 gives the best result for each method. In Figure 5.2, the ROC curves clearly show that our AODI-MKL^{clm} framework covers a larger area under the curve (AUC), which again indicate the superior performance of AODI-MKL^{clm} over the other two baseline methods and thus demonstrate its effectiveness. We also have the same observations for glaucoma and PM detection as well.

In addition, for each disease detection, we show in Figure 5.3 an overall illustration of how the performance changes with respect to different settings using different types of features and their combinations, with the means and stand deviations of both AUC and $\bar{P}$. In the figure, we use the results of the standard SVM for settings 1)–3); and for settings 4)–7), the best results of our AODI-MKL^{clm} framework are used. From the results, we conclude that using either genome information or fundus image data achieves much better performance compared to using demographic/clinic features. Moreover, using the combined multiple types of features can usually lead to better

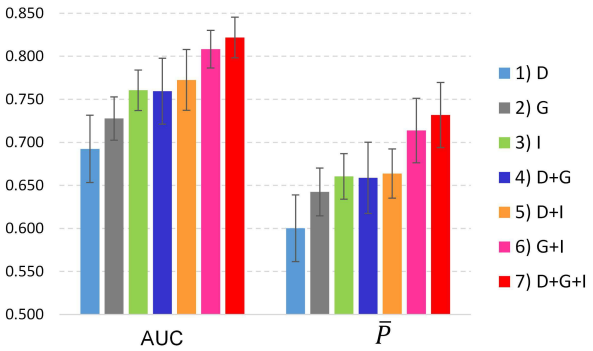
performance than using one of the corresponding single features. It is also interesting to note that even though the demographic data as in setting 1) always results in very poor performance, our AODI-MKL^{clm} framework can still successfully utilize the demographic data as a useful complement to further improve the performance.

5.5 Discussion

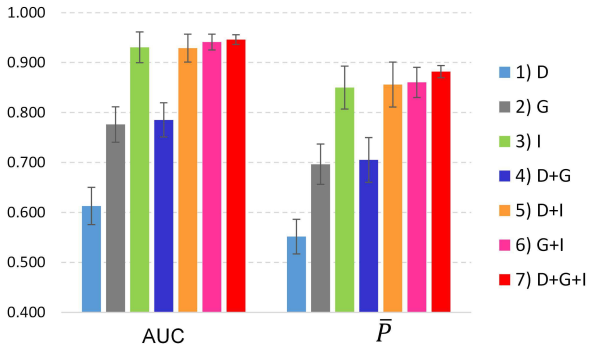
We have proposed a unified AODI-MKL^{clm} framework for the automatic computer aided detection of ocular diseases. AODI-MKL^{clm} effectively fuses the discriminative information from multiple heterogeneous data sources (i.e., demographic/clinical data, genome information and retinal fundus images) as well as simultaneously incorporates pre-learned SVM classifiers trained from individual data sources. We also have conducted independent experiments to evaluate our AODI-MKL^{clm} framework for the detection of three leading ocular diseases (i.e., glaucoma, AMD and PM). The good performance of AODI-MKL^{clm} demonstrates its capability of fusing heterogeneous data for ocular disease detection and also shows its potential for automatic disease detection in the clinic.



(a) Glaucoma



(b) AMD



(c) PM

Fig. 5.3 Performance variations w.r.t. different settings for each disease detection, in terms of Area Under Curve (AUC) and balanced accuracy ($\bar{P}$).

Chapter 6

Conclusions and Future Work

Research done in this thesis could very well serve as a 'first step' or pioneer work towards a general computational framework for automatic ocular diseases detection based on heterogeneous data sources. The conclusion to this thesis describes the progress made towards this goal in terms of the development of the AODI framework and its extension AODI-MKL^{clm}. I also suggest some future research directions that provide the next steps toward the proposed goal.

6.1 Conclusions

This thesis investigated the recent advancements in kernel learning and proposed a computational framework AODI based on multiple kernel learning algorithms. AODI innovatively integrates and draws inference from a collection of biomedical data. By intelligently combining imaging, SNP and clinical data, AODI and its extension AODI-MKL^{clm} are able to effectively boost the accuracy for automatic ocular disease diagnosis. This study fills in the blanks of conventional CAD research and development, which typically account for only one type of data.

Several common characteristics are found in ocular diseases: the diagnosis is heavily reliant on manual reading of various ocular images; the diseases are resulted of the interplay between environmental and genetic factors; and a number of ocular dis-

eases are asymptomatic until in later stages where treatment becomes less effective in preventing vision impairment. These characteristics make such diseases highly appropriate for the purpose of this study. To validate the AODI system, we conduct comprehensive experiments using fundus image, SNP and clinical data obtained from a holistic population study conducted in Singapore. Our statistical analysis shows that that use of AODI-MKL^{clm} provides significant improvements in terms of the accuracy of disease detection; up to 18.5% improvement compared to CAD using imaging alone, a 5.5% improvement than standard SVM using data from individual domains, and a 3% improvement compared to the traditional MKL method.

To the best of our knowledge, AODI is the first published work that employs MKL to integrate heterogeneous data for ocular disease screening/diagnosis. This study also tackles the challenge of feature selection from SNP data by grouping SNPs into functional groups (genes, interacting proteins and biological pathways), and thus embeds biomedical knowledge into sparse learning. Finally, the study incorporates classemes (pre-learned classifiers trained from individual informatics domains) into AODI, which further improves the performance of ocular disease detection.

6.2 Suggestions for Future Work

A multitude of possible research directions can be pursued in tandem with the goal of a truly general purpose framework for ocular disease detection in mind. One such research direction would be to allow automatic learning of the SNP selection, doing away with the current framework that requires the SNPs to be specified explicitly. Chapter 4 provides some insight for automatic SNP selection but novel computational models still need to be proposed to incorporate the known disease related SNPs. A second proposal would be to conduct statistical analysis on the association of ocular diseases and various image features based on larger data sets. This would improve

the interpretability of the outcome generated from AODI. Furthermore, current experiments are conducted by cross–validation within one data set from a population study. Nevertheless, as more of such data become available, subsequent research ought to focus on validating the system across multiple data sets.

6.3 Summary

In summary, I have developed a framework which allows various biomedical data to be integrated for a higher degree of accuracy in ocular disease detection. The promising results demonstrated in this thesis raises the possibility of using a clinical decision support systems such as AODI running parallel with existing clinical workflow to offer an objective, evidence-based diagnosis to clinicians as a second opinion. In clinics, the combination of multiple measurements is more often an art than a science, and are usually mastered by experienced ophthalmologists. The interpretation can be susceptible to inter-observer discrepancy due to differences in training and experience. A carefully designed disease screening program based on AODI can provide a faster, more cost–effective and more accurate detection of the disease. This will lead to improvements in ocular disease management and cost savings for both patients and public healthcare providers.

References

[1] B. E. Robinson, "Prevalence of asymptomatic eye disease," *Canadian Journal of Optometry*, vol. 65, no. 5, pp. 175–180, 2003.

[2] National Eye Institute, "Don't lose sight of diabetic eye disease : information for people with diabetes," *NIH publication*, vol. 04, p. 3252, 2004.

[3] H. Fujita, Y. Uchiyama, T. Nakagawa, D. Fukuoka, Y. Hatanaka, T. Hara, G. Lee, Y. Hayashi, Y. Ikedo, X. Gao, and X. Zhou, "Computer-aided diagnosis: The emerging of three cad systems induced by japanese health care needs," *Computer Methods and Programs in Biomedicine*, vol. 92, pp. 238–248, 2008.

[4] T. Wong, M. Knudtson, R. Klein, B. Klein, S. Meuer, and L. Hubbard, "Computer-assisted measurement of retinal vessel diameters in the beaver dam eye study: methodology, correlation between eyes, and effect of refractive errors.," *Ophthalmology*, vol. 111, no. 6, pp. 1183–90, 2004.

[5] C. Cheung, Y. Zheng, W. Hsu, M. Lee, Q. Lau, P. Mitchell, J. Wang, R. Klein, and T. Wong, "Retinal vascular tortuosity, blood pressure, and cardiovascular risk factors.," *Ophthalmology*, vol. 118, no. 5, pp. 812–8, 2011.

[6] N. Perumalsamy, N. Prasad, S. Sathya, and K. Ramasamy, "Software for reading and grading diabetic retinopathy: Aravind diabetic retinopathy screening 3.0," *Diabetes Care*, vol. 30, no. 9, pp. 2302–2306, 2007.

[7] A. Sommer, J. M. Tielsch, J. Katz, H. A. Quigley, J. D. Gottsch, J. Javitt, and K. Singh, "Relationship between intraocular pressure and primary open angle glaucoma among white and black americans. the baltimore eye survey.," *Archives of ophthalmology*, vol. 109, no. 8, pp. 1090–1095, 1991.

[8] T. Y. Wong, R. Klein, B. E. Klein, J. M. Tielsch, L. Hubbard, and F. J. Nieto, "Retinal microvascular abnormalities and their relationship with hypertension, cardiovascular disease, and mortality," *Survey of ophthalmology*, vol. 46, no. 1, pp. 59–80, 2001.

[9] A. Colenbrander, "Measuring vision and vision loss," *Duane's Ophthalmology. 15th ed. Philadelphia, Pa: Lippincott Williams & Wilkins*, 2009.

[10] O. A. Makeeva, V. V. Markova, and V. P. Puzyrev, "Public interest and expectations concerning commercial genotyping and genetic risk assessment," *Personalized Medicine*, vol. 6, no. 3, pp. 329–341, 2009.

[11] S. Weiss, C. Kulikowski, S. Amarel, and A. Safir, "A model-based method for computer-aided medical decision making," *Artificial Intelligence*, vol. 11, pp. 145–72, 1978.

[12] C. A. Kulikowski and S. M. Weiss, "Representation of expert knowledge for consultation: the casnet and expert projects," *Artificial Intelligence in medicine*, vol. 51, 1982.

[13] K. Chan, T.-W. Lee, P. A. Sample, M. H. Goldbaum, R. N. Weinreb, and T. J. Sejnowski, "Comparison of machine learning and traditional classifiers in glaucoma diagnosis," *IEEE Trans. Biomedical Engineering*, vol. 49, no. 9, pp. 963–974, 2002.

[14] B. Dimitrios, H. Anders, and B. Boel, "Integration and fusion of standard automated perimetry and optical coherence tomography data for improved automated glaucoma diagnostics," *BMC ophthalmology*, vol. 11, no. 1, p. 20, 2011.

[15] D. Kourkoutas, I. S. Karanasiou, G. Tsekouras, M. Moshos, E. Iliakis, and G. Georgopoulos, "Glaucoma risk assessment using a non-linear multivariable regression method," *Computer methods and programs in biomedicine*, vol. 108, no. 3, pp. 1149–59, 2012.

[16] K. Attebo, P. Mitchell, and M. Smith, "Visual acuity and the causes of visual loss in australia," *Ophthalmology*, pp. 103(3):357–64, 1996.

[17] A. Foong, S. Saw, J. Loo, S. Shen, S. Loon, M. Rosman, T. Aung, D. Tan, E. Tai, and T. Wong, "Rationale and methodology for a population-based study of eye diseases in Malay people: The Singapore Malay eye study (SiMES)," *Ophthalmic epidemiology*, vol. 14, no. 1, pp. 25–35, 2007.

[18] C.-W. Pan, T.-Y. Wong, L. Chang, X.-Y. Lin, R. Lavanya, and et al, "Ocular biometry in an urban indian population: The singapore indian eye study (sindi)," *Invest Ophthalmol Vis Sci.*, vol. 52, no. 9, pp. 6636–42, 2011.

[19] C. C. Sng, L. L. Foo, C. Y. Cheng, J. C. Allen, M. He, G. Krishnaswamy, M. E. Nongpiur, D. S. Friedman, T. Y. Wong, and T. Aung, "Determinants of anterior chamber depth: the singapore chinese eye study," *Ophthalmology*, vol. 119, no. 6, pp. 1143–50, 2012.

[20] M. Matsui, T. Tashiro, K. Matsumoto, and S. Yamamoto, "A study on automatic and quantitative diagnosis of fundus photographs. i. detection of contour line of retinal blood vessel images on color fundus photographs," *Nippon Ganka Gakkai Zasshi*, vol. 77, no. 8, p. 907, 1973.

[21] C. Baudoin, B. Lay, and J. Klein, "Automatic detection of microaneurysms in diabetic fluorescein angiography.," *Revue depidemiologie et de sante publique*, vol. 32, no. 3-4, p. 254, 1984.

[22] H. Narasimha-Iyer, A. Can, B. Roysam, C. Stewart, H. Tanenbaum, A. Majerovics, and H. Singh, "Robust detection and classification of longitudinal changes in color retinal fundus images for monitoring diabetic retinopathy," *IEEE Trans. Biomedical Engineering*, vol. 53, no. 6, pp. 1084–1098, 2006.

[23] G. Quellec, K. Lee, M. Dolejsi, M. K. Garvin, M. D. Abramoff, and M. Sonka, "Three-dimensional analysis of retinal layer texture: identification of fluid-filled regions in sd-oct of the macula," *IEEE Trans. Medical Imaging*, vol. 29, no. 6, pp. 1321–1330, 2010.

[24] J. Liu, D. Wong, J. Lim, H. Li, N. Tan, Z. Zhang, T. Wong, and R. Lavanya, "ARGALI: An automatic cup-to-disc ratio measurement system for glaucoma analysis using level-set image processing," in *Proceedings of 13th International Conference on Biomedical Engineering*, pp. 23:559–562, 2009.

[25] W. Huang, K. L. Chan, H. Li, J. H. Lim, J. Liu, and T. Y. Wong, "A computer assisted method for nuclear cataract grading from slit-lamp images using ranking," *IEEE Trans. Medical Imaging*, vol. 30, no. 1, pp. 94–107, 2011.

[26] X. Chen, M. Niemeijer, L. Zhang, K. Lee, M. D. Abràmoff, and M. Sonka, "Three-dimensional segmentation of fluid-associated abnormalities in retinal oct: Probability constrained graph-search-graph-cut," *IEEE Trans. Medical Imaging*, vol. 31, no. 8, pp. 1521–1531, 2012.

[27] G. R. Wilkins, O. M. Houghton, and A. L. Oldenburg, "Automated segmentation of intraretinal cystoid fluid in optical coherence tomography," *IEEE Trans. Biomedical Engineering*, vol. 59, no. 4, pp. 1109–1114, 2012.

[28] Z. Hu, M. Niemeijer, K. Lee, M. D. Abramoff, M. Sonka, and M. K. Garvin, "Automated segmentation of the optic disc margin in 3D optical coherence tomography images using a graph-theoretic approach," in *Proceedings of SPIE Conference on Medical Imaging*, p. 72620U, 2009.

[29] Z. Hu, M. Niemeijer, K. Lee, M. D. Abramoff, M. Sonka, and M. K. Garvin, "Automated segmentation of the optic canal in 3d spectral-domain oct of the optic nerve head (onh) using retinal vessel suppression," *Invest. Ophthalmol. Vis. Sci.*, vol. 50, no. 1, pp. 33–44, 2009.

[30] Y. H. Kwon, Z. Hu, M. D. Abramoff, K. Lee, and M. K. Garvin, "Automated segmentation of neural canal opening and optic cup in sd-oct images," in *Amer. Glaucoma Soc. 20th Annu. Meeting*, 2010.

[31] S. Yun and B. Bouma, *Wavelength swept lasers.* In Drexler WFujimoto JG, eds. Optical Coherence Tomography: Technology and Applications, New York: Springer, 2008.

[32] I. Grulkowski, J. J. Liu, B. Potsaid, V. Jayaraman, C. D. Lu, J. Jiang, A. E. Cable, J. S. Duker, and J. G. Fujimoto, "Retinal, anterior segment and full eye imaging using ultrahigh speed swept source oct with vertical-cavity surface emitting lasers," *Biomedical optics express*, vol. 3, no. 11, pp. 2733–2751, 2012.

[33] R. F. Spaide, M. Akiba, and K. Ohno-Matsui, "Evaluation of peripapillary intrachoroidal cavitation with swept source and enhanced depth imaging optical coherence tomography," *Retina*, vol. 32, no. 6, pp. 1037–1044, 2012.

[34] K. Lee, M. Niemeijer, M. K. Garvin, Y. H. Kwon, M. Sonka, and M. D. Abràmoff, "Segmentation of the optic disc in 3d-oct scans of the optic nerve head," *IEEE Transactions on Medical Imaging*, vol. 29, pp. 159–168, January 2010.

[35] K. Lee, M. Niemeijer, M. K. Garvin, Y. H. Kwon, M. Sonka, and M. D. Abramoff, "3D segmentation of the rim and cup in spectral-domain optical coherence tomography volumes of the optic nerve head," in *Proceedings of SPIE Conference on Medical Imaging*, pp. 7262–7283, 2009.

[36] M. D. Abramoff, K. Lee, M. Niemeijer, W. Alward, E. C. Greenlee, M. K. Garvin, M. Sonka, and Y. H. Kwon, "Automated segmentation of the cup and rim from spectral domain oct of the optic nerve head," *Invest. Ophthalmol. Vis. Sci.*, vol. 50, no. 12, pp. 5778–5784, 2009.

[37] K. Ohno-Matsui, M. Akiba, T. Modegi, M. Tomita, T. Ishibashi, T. Tokoro, and M. Moriyama, "Association between shape of sclera and myopic retinochoroidal lesions in patients with pathologic myopia," *Invest Ophthalmol Vis Sci.*, vol. 53, no. 1, pp. 6046–61, 2012.

[38] K. Ohno-Matsui, M. Akiba, M. Moriyama, N. Shimada, T. Ishibashi, T. Tokoro, and R. F. Spaide, "Acquired optic nerve and peripapillary pits in pathologic myopia," *Ophthalmology*, vol. 119, pp. 1685–1692, August 2012.

[39] Z. Hu, M. D. Abramoff, Y. H. Kwon, K. Lee, and M. K. Garvin, "Automated segmentation of neural canal opening and optic cup in 3D spectral optical coherence tomography images of the optic nerve head," *Invest. Ophthalmol. Vis. Sci.*, vol. 51, no. 11, pp. 5708–5717, 2010.

[40] H. Li, J. Lim, J. Liu, P. Mitchell, A. Tan, J. Wang, and T. Wong, "A computer-aided diagnosis system of nuclear cataract," *IEEE Trans. Biomedical Engineering*, vol. 57, no. 7, pp. 1690–1698, 2010.

[41] D. Duncan, O. Shukla, S. West, and O. Schein, "New objective classification system for nuclear opacification," *Journal of the Optical Society of America. A, Optics and image science and vision*, vol. 14, no. 6, pp. 1197–1204, 1997.

[42] S. Fan, C. Dyer, L. Hubbard, and B. Klein, "An automatic system for classification of nuclear sclerosis from slit-lamp photographs," *Proceedings of Int Conf MICCAI*, pp. 592–601, 2003.

[43] Nidek Co. Ltd, "Anterior eye segment analysis system: Eas-1000. operator's manual, nidek, japan," 1991.

[44] A. Gershenzon and L. Robman, "New software for lens retro-illumination digital image analysis," *Australian and New Zealand Journal of Ophthalmology*, vol. 27, no. 3-4, pp. 170–172, 1999.

[45] B. Klein, R. Klein, K. Linton, Y. Magli, M. Neider, *et al.*, "Assessment of cataracts from photographs in the beaver dam eye study.," *Ophthalmology*, vol. 97, no. 11, pp. 1428–1433, 1990.

[46] X. Gao, H. Li, J. H. Lim, and T. Y. Wong, "Computer-aided cataract detection using enhanced texture features on retro-illumination lens images," in *Proceedings of IEEE Int. Conf. Image Processing*, pp. 1565–1568, 2011.

[47] Y. C. Chow, X. Gao, H. Li, J. H. Lim, Y. Sun, and T. Y. Wong, "Automatic detection of cortical and psc cataracts using texture and intensity analysis on retro-illumination lens images," in *Conf Proceedings of IEEE Eng Med Biol Soc*, pp. 5044–5047, 2011.

[48] X. Gao, D. W. K. Wong, T.-T. Ng, C. Y. L. Cheung, C.-Y. Cheng, and T. Y. Wong, "Automatic grading of cortical and psc cataracts using retroillumination lens images," in *Proceedings of the 11th Asian conference on Computer Vision-Volume Part II*, pp. 256–267, 2012.

[49] M. Sehi, D. Guaqueta, W. Feuer, and D. Greenfield, "Scanning laser polarimetry with variable and enhanced corneal compensation in normal and glaucomatous eyes," *Am. J. Ophthalmol.*, vol. 143, no. 2, pp. 272–279, 2007.

[50] P. J. Lee, C. J. Liu, R. Wojciechowski, J. E. Bailey-Wilson, and C. Y. Cheng, "Structure-function correlations using scanning laser polarimetry in primary angle-closure glaucoma and primary open-angle glaucoma," *Am. J. Ophthalmol.*, vol. 149, no. 5, pp. 817–825, 2010.

[51] K. Vermeer, B. Lo, Q. Zhou, F. Vos, A. Vossepoel, and H. Lemij, "Event-based progression detection strategies using scanning laser polarimetry images of the human retina," *Computers in Biology and Medicine*, vol. 41, no. 9, pp. 857–864, 2011.

[52] F. Medeiros, L. Zangwill, C. Bowd, and R. Weinreb, "Comparison of the gdx vcc scanning laser polarimeter, hrt ii confocal scanning laser ophthalmoscope, and stratus oct optical coherence tomograph for the detection of glaucoma," *Archives of Ophthalmology*, vol. 122, no. 6, pp. 827–837, 2004.

[53] Z. Ben Sbeh, L. D. Cohen, G. Mimoun, and G. Coscas, "A new approach of geodesic reconstruction for drusen segmentation in eye fundus images," *IEEE Trans. Medical Imaging*, vol. 20, no. 12, pp. 1321–1333, 2001.

[54] T. P. Karnowski, V. P. Govindasamy, K. W. Tobin, E. Chaum, and M. Abramoff, "Retina lesion and microaneurysm segmentation using morphological reconstruction methods with ground-truth data," in *Conf Proceedings of IEEE Eng Med Biol Soc*, pp. 5433–5436, 2008.

[55] K. Rapantzikos, M. Zervakis, and K. Balas, "Detection and segmentation of drusen deposits on human retina: Potential in the diagnosis of age-related macular degeneration," *Medical Image Analysis*, vol. 7, no. 1, pp. 95–108, 2003.

[56] A. Hoover, V. Kouznetsoza, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 203–210, 2000.

[57] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, and L. Kennedy, "Measurement of retinal vessel widths from fundus images based on 2-d modeling," *IEEE Trans Biomed Eng*, vol. 23, no. 10, pp. 1196–1204, 2004.

[58] C. Heneghan, J. Flynn, M. OKeef, and M. Cahill, "Characterization of changes in blood vessel and tortuosity in retinopathy of prematurity using image analysis," *Med. Image Anal.*, vol. 6, no. 4, pp. 407–429, 2002.

[59] G. Joshi, J. Sivaswamy, and S. Krishnadas, "Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment," *IEEE Tran. Medical Imaging*, vol. 30, no. 6, pp. 1192–1205, 2011.

[60] Y. Hatanaka, A. Noudo, C. Muramatsu, A. Sawada, T. Hara, T. Yamamoto, and H. Fujita, "Automatic measurement of vertical cup-to-disc ratio on retinal fundus images," *Medical Biometrics, Lecture Notes in Computer Science*, vol. 6165, pp. 64–72, 2010.

[61] J. Liu, D. W. Wong, J. H. Lim, N. M. Tan, Z. Zhang, H. Li, F. Yin, B. Lee, S. M. Saw, L. Tong, *et al.*, "Detection of pathological myopia by pamela with texture-based features through an svm approach," *Journal of Healthcare Engineering*, vol. 1, no. 1, pp. 1–12, 2010.

[62] S. Tan, X. Cheng, Y. Wang, and H. Xu, "Adapting naive bayes to domain adaptation for sentiment analysis," in *Advances in Information Retrieval - European Conference on IR Research*, (Toulouse, France), pp. 337–349, Springer, April 2009.

[63] J. Cheng, D. Tao, J. Liu, D. W. K. Wong, N. M. Tan, T. Y. Wong, and S. M. Saw, "Peripapillary atrophy detection by sparse biologically inspired feature manifold," *IEEE Trans. Medical Imaging*, vol. 31, no. 12, pp. 2355–2365, 2012.

[64] R. T. Smith, J. K. Chan, T. Nagasaki, U. F. Ahmad, I. Barbazetto, J. Sparrow, M. Figueroa, and J. Merriam, "Automated detection of macular drusen using geometric background leveling and threshold selection," *Archives of ophthalmology*, vol. 123, no. 2, pp. 200–206, 2005.

[65] B. Lee and H. Adam, "Drusen detection in a retinal image using multi-level analysis," in *Proceedings of Int Conf MICCAI*, pp. 618–625, 2003.

[66] D. Freund, N. Bressler, and P. Burlina, "Automated detection of drusen in the macula," in *Proceedings of IEEE Int Symposium Biomedical Imaging*, pp. 61–64, 2009.

[67] S. Tamura, Y. Okamoto, and K. Yanashima, "Zero-crossing interval correction in tracing eye-fundus blood vessels," *Pattern Recogn*, vol. 21, no. 3, pp. 227–233, 1988.

[68] B. Kochner, D. Schulmann, M. Michaelis, G. Mann, and K. Englemeier, "Course tracking and contour extraction of retinal vessels from colour fundus photographs: most efficient use of steerable filters for model based image analysis," in *Proceedings of SPIE Conference on Medical Imaging*, pp. 755–761, 1988.

[69] B. Antal and A. Hajdu, "An ensemble-based system for microaneurysm detection and diabetic retinopathy grading," *IEEE Trans. Biomedical Engineering*, vol. 59, no. 6, pp. 1720–1726, 2012.

[70] D. W. K. Wong, J. Liu, J. H. Lim, X. Jia, F. Yin, H. Li, and T. Y. Wong, "Level-set based automatic cup-to-disc ratio determination using retinal fundus images in ARGALI," *Conf Proceedings IEEE Eng Med Biol Soc*, pp. 2266–2269, 2008.

[71] R. Bock, J. Meier, G. Michelson, L. Nyl, and J. Hornegger, "Classifying glaucoma with image-based features from fundus photographs," *Lecture Notes in Computer Science*, vol. 4713, pp. 355–364, 2007.

[72] J. Meier, R. Bock, G. Michelson, L. G. Nyl, and J. Hornegger, "Effects of preprocessing eye fundus images on appearance based glaucoma classification," *Lecture Notes in Computer Science*, vol. 4673, pp. 165–172, 2007.

[73] Z. Zhang, Y. Xu, J. Liu, D. W. K. Wong, C. K. Kwoh, S. M. Shaw, and T. Y. Wong, "Automatic diagnosis of pathological myopia from heterogeneous biomedical data," *PLoS ONE*, vol. 8, no. 6, p. e65736, 2013.

[74] J. Liu, D. Wong, N. Tan, Z. Zhang, S. Lu, J. Lim, H. Li, S. Saw, L. Tong, and T. Wong, "Automatic classification of pathological myopia in retinal fundus images using pamela," in *Proceedings of SPIE Conference on Medical Imaging*, p. 76240G, 2010.

[75] E. Barriga, V. Murray, C. Agurto, M. Pattichis, S. Russell, M. Abramoff, H. Davis, and P. Soliz, "Multi-scale AM-FM for lesion phenotyping on age-related macular degeneration," in *Proceedings IEEE Int Symp Computer-Based Medical Systems*, pp. 1–5, 2009.

[76] P. Soliz, S. R. Russell, M. D. Abramoff, S. Murillo, M. Pattichis, and H. Davis, "Independent component analysis for vision-inspired classification of retinal images with age-related macular degeneration," in *Proceedings of IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 65–68, 2008.

[77] Y. Zheng, B. Vanderbeek, E. Daniel, D. Stambolian, M. Maguire, D. Brainard, and J. Gee, "An automated drusen detection system for classifying age-related maculardegeneration with color fundus photographs," in *Proceedings on IEEE International Symposium on Biomedical Imaging*, pp. 1440–1443, 2013.

[78] B. Harangi, I. Lazar, and A. Hajdu, "Automatic exudate detection using active contour model and regionwise classification," in *Conf Proceedings of IEEE Eng Med Biol Soc*, pp. 5951–4, 2012.

[79] C. I. O. Martins, F. Medeiros, R. M. Veras, F. N. Bezerra, and R. Cesar, "Evaluation of retinal vessel segmentation methods for microaneurysms detection," in *Proceedings of IEEE Int. Conf. Image Processing*, 2009.

[80] H. F. Jaafar, A. K. Nandi, and W. Al-Nuaimy, "Detection of exudates in retinal images using a pure splitting technique," in *Conf Proceedings of IEEE Eng Med Biol Soc*, 2010.

[81] R. Bock, J. Meier, L. G. Nyúl, J. Hornegger, and G. Michelson, "Glaucoma risk index: automated glaucoma detection from color fundus images," *Medical Image Analysis*, vol. 14, pp. 471–481, June 2010.

[82] J. Cheng, D. Tao, J. Liu, D. Wong, B. Lee, B. Mani, T. Wong, and A. T., "Focal biologically inspired feature for glaucoma type classification," in *Proceedings of Int Conf MICCAI*, pp. 6893:91–98, 2011.

[83] Y. Xu, J. Liu, S. Lin, D. Xu, C. Cheung, T. Aung, and T. Wong, "Efficient optic cup detection from intra-image learning with retinal structure priors," in *Proceedings Int Conf MICCAI*, pp. 7510:58–65, 2012.

[84] J. Cheng, D. W. K. Wong, X. Cheng, J. Liu, N. M. Tan, M. Bhargava, C. M. G. Cheung, and T. Y. Wong, "Early age-related macular degeneration detection by focal biologically inspired feature," in *Proceedings of IEEE Int. Conf. Image Processing*, pp. 2805–2808, 2012.

[85] C. Köse, U. Sevik, O. Gencalioglu, C. Ikibas, and T. Kayikicioglu, "A statistical segmentation method for measuring age-related macular degeneration in retinal fundus images," *J Med Sys*, vol. 34, no. 1, pp. 1–13, 2010.

[86] G. Quellec, S. R. Russell, and M. D. Abràmoff, "Optimal filter framework for automated, instantaneous detection of lesions in retinal images," *IEEE Trans. Medical Imaging*, vol. 30, no. 2, pp. 523–533, 2011.

[87] T. Walter, P. Massin, A. Erginay, R. Ordonez, C. Jeulin, and J.-C. Klein, "Automatic detection of microaneurysms in color fundus images," *Medical Image Analysis*, vol. 11, no. 6, pp. 555–566, 2007.

[88] I. Lazar, A. Hajdu, *et al.*, "Retinal microaneurysm detection through local rotating cross-section profile analysis.," *IEEE Trans. medical imaging*, vol. 32, no. 2, pp. 400–407, 2013.

[89] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson, and P. F. Sharp, "Automated microaneurysm detection using local contrast normalization and local vessel detection," *IEEE Trans. Medical Imaging*, vol. 25, no. 9, pp. 1223–1232, 2006.

[90] H. Ohwada, M. Daidoji, S. Shirato, and F. Mizoguchi, "Learning first order rules from image applied to glaucoma diagnosis," in *Proceedings of Pacific Rim International Conference on Artificial Intelligence*, pp. 494–505, 1998.

[91] L. G. Nyul, "Retinal image analysis for automated glaucoma risk evaluation," in *Proceedings of SPIE Conference on Medical Imaging*, pp. 74971C1–9, 2009.

[92] R. McIntyre, M. I. Heywood, P. H. Artes, and S. S. R. Abidi, "Toward glaucoma classification with moment methods," in *Computer and Robot Vision*, pp. 265–272, 2004.

[93] K. Ujjwal, A. Chakravarty, and J. Sivaswamy, "Visual saliency based bright lesion detection and discrimination in retinal images," in *Proceedings IEEE 10th Int Symposium Biomedical Imaging: From Nano to Macro*, pp. 1428–1431, 2013.

[94] J. P. Medhi, M. K. Nath, and S. Dandapat, "Automatic grading of macular degeneration from color fundus images," in *Proceedings of World Congress on Information and Communication Technologies*, pp. 511–514, 2012.

[95] Z. Liang, D. W. Wong, J. Liu, K. L. Chan, and T. Y. Wong, "Towards automatic detection of age-related macular degeneration in retinal fundus images," in *Conf Proceedings of IEEE Eng Med Biol Soc*, pp. 4100–4103, 2010.

[96] M. Esmaeili, H. Rabbani, A. M. Dehnavi, and A. Dehghani, "A new curvelet transform based method for extraction of red lesions in digital color retinal images," in *Proceedings of IEEE Int. Conf. Image Processing*, 2010.

[97] S. Ravishankar, A. Jain, and A. Mittal, "Automated feature extraction for early detection of diabetic retinopathy in fundus images," in *Proceedings IEEE Conf. on Comp Vis Pattern Recognition*, 2009.

[98] B. Antal and A. Hajdu, "Improving microaneurysm detection using an optimally selected subset of candidate extractors and preprocessing methods," *Pattern recognition*, vol. 45, no. 1, pp. 264–270, 2012.

[99] J. Yu, S. S. R. Abidi, P. H. Artes, A. Mcintyre, and M. Heywood, "Automated optic nerve analysis for diagnostic support in glaucoma," in *Proceedings of In Symposium on Computer-Based Medical Systems*, pp. 97–102, 2005.

[100] Y. Xu, S. Lin, D. W. K. Wong, J. Liu, and D. Xu, "Efficient reconstruction-based optic cup localization for glaucoma screening," in *Proceedings of Int Conf MICCAI*, pp. 445–452, 2013.

[101] C. Muramatsu, T. Nakagawa, A. Sawada, Y. Hatanaka, T. Hara, T. Yamamoto, and H. Fujita, "Determination of cup and disc ratio of optical nerve head for diagnosis of glaucoma on stereo retinal fundus image pairs," in *Proceedings of SPIE Conference on Medical Imaging*, pp. 603–610, 2009.

[102] M. H. A. Hijazi, F. Coenen, and Y. Zheng, "Retinal image classification using a histogram based approach," in *IEEE International Joint Conference on Neural Networks*, pp. 3501–3507, 2010.

[103] H. M. H. Ahmad, C. Frans, and Z. Yalin, "Retinal image classification for the screening of age-related macular degeneration," in *Research and Development in Intelligent Systems XXVII*, 2011.

[104] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics gems IV*, pp. 474–485, Academic Press Professional, Inc., 1994.

[105] B. Lay, C. Baudoin, and J.-C. Klein, "Automatic detection of microaneurysms in retinopathy fluoro-angiogram," in *Proceedings of 27th Annual Techincal Symposium*, pp. 165–173, 1984.

[106] M. J. Cree, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester, "A fully automated comparative microaneurysm digital detection system," *Eye*, vol. 11, no. 5, pp. 622–628, 1997.

[107] J. Cheng, J. Liu, D. Tao, F. Yin, D. Wong, and T. Y. Wong, "Superpixel classification based optic cup segmentation," in *Proceedings of Int Conf MICCAI*, pp. 421–428, 2013.

[108] J. Cheng, J. Liu, Y. Xu, F. Yin, D. W. K. Wong, N.-M. Tan, D. Tao, C.-Y. Cheng, T. Aung, and T. Y. Wong, "Superpixel classification based optic disc and optic cup segmentation for glaucoma screening," *IEEE Trans. Med. Imaging*, vol. 32, pp. (6):1019–1032, 2013.

[109] Y. Xu, D. Xu, S. Lin, J. Liu, J. Cheng, C. Y. Cheung, T. Aung, and T. Y. Wong, "Sliding window and regression based cup detection in digital fundus images for glaucoma diagnosis," in *Proceedings Int Conf MICCAI*, pp. 1–8, 2011.

[110] H. M. H. Ahmad, J. Chuntao, C. Frans, and Z. Yalin, "Image classification for age-related macular degeneration screening using hierarchical image decompositions and graph mining," *Lecture Notes in Computer Science*, vol. 6912, pp. 65–80, 2011.

[111] Y. Zheng, M. H. A. Hijazi, and F. Coenen, "Automated "disease/no disease" grading of age-related macular degeneration by an image mining approach," *Invest Ophthalmol Vis Sci*, vol. 53, no. 13, pp. 8310–8318, 2012.

[112] R. Priya and P. Aruna, "Automated diagnosis of age-related macular degeneration from color retinal fundus images," in *Proceedings of the 3rd International Conference on Electronics Computer Technology*, pp. 2:227–230, 2011.

[113] T. Walter and J.-C. Klein, "Automatic detection of microaneurysms in color fundus images of the human retina by means of the bounding box closing," in *Proceedings the Third International Symposium on Medical Data Analysis*, pp. 210–220, 2002.

[114] M. Niemeijer, B. van Ginneken, J. Staal, M. S. Suttorp-Schulten, and M. D. Abràmoff, "Automatic detection of red lesions in digital color fundus photographs," *IEEE Trans. Medical Imaging*, vol. 24, no. 5, pp. 584–592, 2005.

[115] C. Sinthanayothin, J. Boyce, T. Williamson, H. Cook, E. Mensah, S. Lal, and D. Usher, "Automated detection of diabetic retinopathy on digital fundus images," *Diabetic Medicine*, vol. 19, no. 2, pp. 105–112, 2002.

[116] D. Wong, J. Liu, J. H. Lim, H. Li, X. Jia, F. Yin, and T. Y. Wong, "Automated detection of kinks from blood vessels for optic cup segmentation in retinal images," in *Proceedings of SPIE Conference on Medical Imaging*, vol. 7260, pp. 72603L1–8, 2009.

[117] G. D. Joshi, J. Sivaswamy, K. Karan, R. Prashanth, and R. Krishnadas, "Vessel bend-based cup segmentation in retinal images," in *Proceedings of Int. Conf. Pattern Recognition*, pp. 2536–2539, 2010.

[118] E. Chaum, T. P. Karnowski, V. P. Govindasamy, M. Abdelrahman, and K. W. Tobin, "Automated diagnosis of retinopathy by content-based image retrieval," *Retina*, vol. 28, no. 10, pp. 1463–1477, 2008.

[119] K. W. Tobin, M. Abdelrahman, E. Chaum, V. P. Govindasamy, and T. P. Karnowski, "A probabilistic framework for content-based diagnosis of retinal disease," in *Conf Proceedings IEEE Eng Med Biol Soc*, pp. 6743–6746, 2007.

[120] G. Gardner, D. Keating, T. Williamson, and A. Elliott, "Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool.," *British journal of Ophthalmology*, vol. 80, no. 11, pp. 940–944, 1996.

[121] M. U. Akram, S. Khalid, and S. A. Khan, "Identification and classification of microaneurysms for early detection of diabetic retinopathy," *Pattern Recognition*, vol. 46, no. 1, pp. 107–116, 2013.

[122] B. Zhang, F. Karray, L. Zhang, and J. You, "Microaneurysm (ma) detection via sparse representation classifier with ma and non-ma dictionary learning," in *Proceedings Int Conf Pattern Recognition*, pp. 277–280, 2010.

[123] J. Xu, O. Chutatape, E. Sung, C. Zheng, and P. Kuan, "Optic disc feature extraction via modified deformable model technique for glaucoma analysis," *Pattern Recognition*, vol. 40, no. 7, pp. 2063–2076, 2007.

[124] M. Abramoff, W. Alward, E. Greenlee, L. Shuba, C. Kim, J. Fingert, and Y. Kwon, "Automated segmentation of the optic disc from stereo color photographs using physiologically plausible features," *Invest Ophthalmol Vis Sci*, vol. 48, no. 4, pp. 1665–1673, 2009.

[125] E. Corona, S. Mitra, M.Wilson, T. Krile, Y. H. Kwon, and P. Soliz, "Digital stereo image analyzer for generating automated 3D measures of optic disc deformation in glaucoma," *IEEE Trans. Med. Imag.*, vol. 21, no. 10, pp. 1244–1253, 2002.

[126] M. Niemeijer, B. van Ginneken, S. R. Russell, M. S. Suttorp-Schulten, and M. D. Abramoff, "Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis," *Invest Ophthalmol Vis Sci*, vol. 48, no. 5, pp. 2260–2267, 2007.

[127] M. Mubbashar, A. Usman, and M. U. Akram, "Automated system for macula detection in digital retinal images," in *Proceedings of Int. Conf. on Information and Communication Technologies*, pp. 1–5, 2011.

[128] A. Hunter, J. A. Lowell, B. Ryder, A. Basu, and D. Steel, "Automated diagnosis of referable maculopathy in diabetic retinopathy screening," in *Conf Proceedings of IEEE Eng Med Biol Soc*, pp. 3375–3378, 2011.

[129] K. Ram, G. D. Joshi, and J. Sivaswamy, "A successive clutter-rejection-based approach for early detection of diabetic retinopathy," *IEEE Trans. Biomedical Engineering*, vol. 58, no. 3, pp. 664–673, 2011.

[130] B. Zhang, X. Wu, J. You, Q. Li, and F. Karray, "Detection of microaneurysms using multi-scale correlation coefficients," *Pattern Recognition*, vol. 43, no. 6, pp. 2237–2248, 2010.

[131] A. Guesalag, P. Irarrźabal, M. Guarini, and R. Álvarez, "Measurement of the glaucomatous cup using sequentially acquired stereoscopic images," *Measurement*, vol. 34, no. 3, pp. 207–213, 2003.

[132] M. B. Merickel, X. Wu, M. Sonka, and M. Abramoff, "Optimal segmentation of the optic nerve head from stereo retinal images," in *Proceedings of SPIE Conference on Medical Imaging*, pp. 1031–1038, 2006.

[133] S. Lu, J. Liu, J. H. Lim, Z. Zhang, T. N. Meng, W. K. Wong, H. Li, and T. Y. Wong, "Automatic fundus image classification for computer-aided diagonsis," in *Conf Proceedings of IEEE Eng Med Biol Soc*, pp. 1453–1456, 2009.

[134] X. Cheng, D. W. K. Wong, J. Liu, B.-H. Lee, N. M. Tan, J. Zhang, C. Y. Cheng, G. Cheung, and T. Y. Wong, "Automatic localization of retinal landmarks," in *Conf Proceedings of IEEE Eng Med Biol Soc*, pp. 4954–4957, 2012.

[135] N. Lee, J. Wielaard, A. Fawzi, P. Sajda, A. Laine, G. Martin, M. Humayun, and R. Smith, "In vivo snapshot hyperspectral image analysis of age-related macular degeneration," in *Conf Proceedings of IEEE Eng Med Biol Soc*, pp. 5363–5366, 2010.

[136] H. F. Jaafar, A. K. Nandi, and W. Al-Nuaimy, "Automated detection of red lesions from digital colour fundus photographs," in *Conf Proceedings of IEEE Eng Med Biol Soc*, pp. 584–592, 2011.

[137] C. Agurto, V. Murray, E. Barriga, S. Murillo, M. Pattichis, H. Davis, S. Russell, M. Abràmoff, and P. Soliz, "Multiscale AM-FM methods for diabetic retinopathy lesion detection," *IEEE Trans. Medical Imaging*, vol. 29, no. 2, pp. 502–512, 2010.

[138] N. Jain, S. Farsiu, A. A. Khanifar, S. Bearelly, R. T. Smith, J. A. Izatt, and C. A. Toth, "Quantitative comparison of drusen segmented on sd-oct versus drusen delineated on color fundus photographs," *Invest Ophthalmol Vis Sci*, vol. 51, no. 10, pp. 4875–4883, 2010.

[139] M. Niemeijer, B. van Ginneken, S. R. Russell, M. S. Suttorp-Schulten, and M. D. Abramoff, "Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis," *Invest Ophthalmol Vis Sci*, vol. 48, no. 5, pp. 2260–2267, 2007.

[140] H. F. Jelinek, A. Rocha, T. Carvalho, S. Goldenstein, and J. Wainer, "Machine learning and pattern classification in identification of indigenous retinal pathology," in *Conf Proceedings of IEEE Eng Med Biol Soc*, pp. 5951–5954, 2011.

[141] L. Tang, M. Niemeijer, J. Reinhardt, M. Garvin, M. Abramoff, *et al.*, "Splat feature classification with application to retinal hemorrhage detection in fundus images.," *IEEE Trans. medical imaging*, vol. 32, no. 2, pp. 364–375, 2013.

[142] A. F. M. Hani, N. F. Ngah, T. M. George, L. I. Izhar, H. Nugroho, and H. A. Nugroho, "Analysis of foveal avascular zone in colour fundus images for grading of diabetic retinopathy severity," in *Conf Proceedings of IEEE Eng Med Biol Soc*, pp. 5632–5635, 2010.

[143] F. Oloumi, R. M. Rangayyan, and A. L. Ells, "Computer-aided diagnosis of proliferative diabetic retinopathy," in *Conf Proceedings IEEE Eng Med Biol Soc*, pp. 1438–1441, 2012.

[144] M. Niemeijer, M. D. Abramoff, and B. van Ginneken, "Information fusion for diabetic retinopathy cad in digital color fundus photographs," *IEEE Trans. Medical Imaging*, vol. 28, no. 5, pp. 775–785, 2009.

[145] A. Rocha, T. Carvalho, H. F. Jelinek, S. Goldenstein, and J. Wainer, "Points of interest and visual dictionaries for automatic retinal lesion detection," *IEEE Trans. Biomedical Engineering*, vol. 59, no. 8, pp. 2244–2253, 2012.

[146] M. D. Abràmoff, J. C. Folk, D. P. Han, J. D. Walker, D. F. Williams, S. R. Russell, P. Massin, B. Cochener, P. Gain, L. Tang, *et al.*, "Automated analysis of retinal images for detection of referable diabetic retinopathy," *JAMA Ophthalmology*, vol. 131, no. 3, pp. 351–357, 2013.

[147] V. Murray, C. Agurto, S. Barriga, M. S. Pattichis, and P. Soliz, "Real-time diabetic retinopathy patient screening using multiscale AM-FM methods," in *Proceedings of IEEE Int. Conf. Image Processing*, pp. 525–528, 2012.

[148] C. Agurto, E. S. Barriga, V. Murray, S. Nemeth, R. Crammer, W. Bauman, G. Zamora, M. S. Pattichis, and P. Soliz, "Automatic detection of diabetic retinopathy and age-related macular degeneration in digital fundus images," *Invest Ophthalmol Vis Sci*, vol. 52, no. 8, pp. 5862–5871, 2011.

[149] G. Quellec, M. Lamard, P. M. Josselin, G. Cazuguel, B. Cochener, and C. Roux, "Optimal wavelet transform for the detection of microaneurysms in retina photographs," *IEEE Trans. Medical Imaging*, vol. 27, no. 9, pp. 1230–1241, 2008.

[150] M. Niemeijer, B. van Ginneken, and M. Abràmoff, "Automatic detection and classification of microaneurysms and small hemorrhages in color fundus photographs," *Proceedings Computer Aided Fundus Imaging and Analysis (CAFIA)*, 2003.

[151] Z. Burgansky-Eliash, G. Wollstein, R. Bilonick, H. Ishikawa, L. Kagemann, and J. Schuman, "Glaucoma detection with the heidelberg retina tomograph 3," *Ophthalmology*, vol. 114, no. 3, pp. 466–471, 2007.

[152] B. Chauhan, J. Blanchard, D. Hamilton, and R. LeBlanc, "Technique for detecting serial topographic changes in the optic disc and peripapillary retina using scanning laser tomography," *Invest. Ophthalmol. Vis. Sci.*, vol. 41, no. 3, pp. 775–782, 2000.

[153] S. Miglior, M. Guareschi, E. Albe, S. Gomarasca, M. Vavassori, and N. Orzalesi, "Detection of glaucomatous visual field changes using the moorfields regression analysis of the heidelberg retina tomograph," *Am. J. Ophthalmol.*, vol. 136, no. 1, pp. 26–33, 2003.

[154] G. Wollstein, D. Garway-Heath, and R. Hitchings, "Identification of early glaucoma cases with the scanning laser ophthalmoscope," *Ophthalmology*, vol. 105, no. 8, pp. 1557–1563, 1998.

[155] R. Bernardes, P. Serranho, and C. Lobo, "Digital ocular fundus imaging: a review," *Ophthalmologica*, vol. 226, no. 4, pp. 161–181, 2011.

[156] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, *et al.*, "Optical coherence tomography," *Science*, vol. 254, no. 5035, pp. 1178–1181, 1991.

[157] M. R. Hee, C. R. Baumal, C. A. Puliafito, J. S. Duker, E. Reichel, J. R. Wilkins, J. G. Coker, J. S. Schuman, E. A. Swanson, and J. G. Fujimoto, "Optical coherence tomography of age-related macular degeneration and choroidal neovascularization," *Ophthalmology*, vol. 103, no. 8, p. 1260, 1996.

[158] J. G. Fujimoto, "Optical coherence tomography for ultrahigh resolution in vivo imaging," *Nature biotechnology*, vol. 21, no. 11, pp. 1361–1367, 2003.

[159] G. Pardianto *et al.*, "Understanding diabetic retinopathy," *Mimbar Ilmiah Oftalmologi Indonesia*, vol. 2, pp. 65–6, 2005.

[160] H. J. Jelinek, M. J. Cree, D. Worsley, A. Luckie, and P. Nixon, "An automated microaneurysm detector as a tool for identification of diabetic retinopathy in rural optometric practice," *Clinical and Experimental Optometry*, vol. 89, no. 5, pp. 299–305, 2006.

[161] K. Noronha, U. R. Acharya, K. P. Nayak, S. Kamath, and S. V. Bhandary, "Decision support system for diabetic retinopathy using discrete wavelet transform," *Proceedings the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 227, no. 3, pp. 251–261, 2013.

[162] N. Larsen, J. Godt, M. Grunkin, H. Lund-Andersen, and M. Larsen, "Automated detection of diabetic retinopathy in a fundus photographic screening population," *Invest Ophthalmol Vis Sci*, vol. 44, no. 2, pp. 767–771, 2003.

[163] A. B. Hansen, N. V. Hartvig, M. S. Jensen, K. Borch-Johnsen, H. Lund-Andersen, and M. Larsen, "Diabetic retinopathy screening using digital non-mydriatic fundus photography and automated image analysis," *Acta Ophthalmologica Scandinavica*, vol. 82, no. 6, pp. 666–672, 2004.

[164] G. Michelson, S. Wrntges, J. Hornegger, and B. Lausen, "The papilla as screening parameter for early diagnosis of glaucoma," *Deutsches Aerzteblatt Int.*, vol. 105, pp. 34–35, 2008.

[165] M. Mookiah, U. Acharya, C. M. Lim, A. Petznick, and J. S. Suri, "Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features," *Knowledge-Based Systems*, vol. 33, pp. 73–82, 2012.

[166] T. Damms and F. Dannheim, "Sensitivity and specificity of optic disc parameters in chronic glaucoma," *Invest. Ophth. Vis. Sci.*, vol. 34, no. 7, pp. 2246–2250, 1993.

[167] D. Michael and O. D. Hancox, "Optic disc size, an important consideration in the glaucoma evaluation," *Clinical Eye and Vision Care*, vol. 11, no. 2, pp. 59–62, 1999.

[168] M. Mookiah, U. Acharya, C. Chua, L. Min, E. Ng, M. Mushrif, and A. Laude, "Automated detection of optic disk in retinal fundus images using intuitionistic fuzzy histon segmentation," *Proceedings Inst Mech Eng H.*, vol. 227, no. 1, pp. 37–49, 2013.

[169] N. Harizman, C. Oliveira, A. Chiang, C. Tello, M. Marmor, R. Ritch, and J. M. Liebmann, "The ISNT rule and differentiation of normal from glaucomatous eyes," *Arch Ophthalmol*, vol. 124, no. 11, pp. 1579–1583, 2006.

[170] J. Jonas, M. Fernandez, and G. Naumann, "Glaucomatous parapapillary atrophy: Occurrence and correlations," *Arch Ophthalmol*, vol. 110, pp. 214–222, 1992.

[171] R. Allingham, *Shields' Textbook of Glaucoma, 5th Edition.* Lippincott Williams & Wilkins, 2004.

[172] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 169–208, 2010.

[173] A. Sharma, A. Sobti, M. Wadhwani, and A. Panda, "Evaluation of retinal nerve fiber layer using scanning laser polarimetry," *Journal of Current Glaucoma Practice*, vol. 4, no. 3, pp. 240–251, 2010.

[174] J. De and T. Paulus, "Age-related macular degeneration," *New England Journal of Medicine*, vol. 355, no. 14, pp. 1474–1485, 2006.

[175] M. H. A. Hijazi, F. Coenen, and Y. Zheng, "Data mining techniques for the screening of age-related macular degeneration," *Knowledge-Based Systems*, vol. 29, pp. 83–92, 2012.

[176] D. Tax and R. Duin, "Support vector data description," *Machine Learning*, vol. 54, no. 1, pp. 45–66, 2004.

[177] D. W. Wong, J. Liu, X. Cheng, J. Zhang, F. Yin, M. Bhargava, G. C. Cheung, and T. Y. Wong, "Thalia-an automatic hierarchical analysis system to detect drusen lesion images for amd assessment," in *Proceedings of IEEE Int Symposium Biomedical Imaging*, pp. 884–887, 2013.

[178] T. Young, S. Ronan, A. Alvear, S. Wildenberg, W. Oetting, L. Atwood, D. Wilkin, and R. King, "A second locus for familial high myopia maps to chromosome 12q," *Am J Hum Genet.*, vol. 63, no. 5, pp. 1419–24, 1998.

[179] S. Saw, J. Katz, O. Schein, S. Chew, and T. Chan, "Epidemiology of myopia," *Epidemiol Rev.*, vol. 18, no. 2, pp. 175–187, 1996.

[180] Y. Xu, J. Liu, Z. Zhang, N. M. Tan, D. Wong, S. M. Saw, and T. Y. Wong, "Learn to recognize pathological myopia in fundus images using bag-of-feature and sparse learning approach," in *Proceedings IEEE 10th Int Symposium Biomedical Imaging*, pp. 888–891, April 2013.

[181] J. Cheng, D. Tao, J. Liu, D. Wong, N. Tan, T. Wong, and S. Saw, "Peripapillary atrophy detection by sparse biologically inspired feature manifold," *IEEE Trans. Medical Imaging*, vol. 31, no. 12, pp. 2355 – 2365, 2012.

[182] L. Lim, G. Cheung, and S. Lee, "Comparison of spectral domain and swept-source optical coherence tomography in pathological myopia," *Eye (Lond)*, vol. 28, no. 4, pp. 488–491, 2014.

[183] B. Potsaid, B. Baumann, D. Huang, S. Barry, A. Cable, J. Schuman, J. Duker, and J. Fujimoto, "Ultrahigh speed 1050nm swept source/fourier domain oct retinal and anterior segment imaging at 100,000 to 400,000 axial scans per second," *Optics Express*, vol. 18, no. 19, pp. 20029–48, 2010.

[184] R. Varma, W. Steinmann, G. Spaeth, and R. Wilson, "Variability in digital analysis of optic disc topography," *Graefes Arch Clin Exp Ophthalmol.*, vol. 226, no. 5, pp. 435–42, 1988.

[185] J. Jonas, P. Martus, W. Budde, and J. Hayler, "Morphologic predictive factors for development of optic disc hemorrhages in glaucoma," *Invest. Ophthalmol. Vis. Sci.*, vol. 43, no. 9, pp. 2956–2961, 2002.

[186] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, R. Fletcher, and L. Kennedy, "Optic nerve head segmentation," *IEEE Trans. Medical Imaging*, vol. 23, no. 2, pp. 256–264, 2004.

[187] D. Wong, J. Liu, N. Tan, F. Yin, and T. Wong, "Automatic detection of the optic cup using vessel kinking in digital retinal fundus images," in *Proceedings IEEE Int Symposium Biomedical Imaging*, pp. 1647–1650, 2012.

[188] G. Joshi, J. Sivaswamy, and S. Krishnadas, "Depth discontinuity-based cup segmentation from multiview color retinal image," *IEEE Trans. Biomedical Engineering*, vol. 59, no. 6, pp. 1523–1531, 2012.

[189] J. Cheng, J. Liu, D. W. K. Wong, N. M. Tan, C. Cheung, M. Baskaran, W. T. Y., and S. S. M., "Peripapillary atrophy detection by biologically inspired feature," in *Proceedings of Int Conf Pattern Recognition*, pp. 2063–2066, 2012.

[190] Y. Wang, J. Shen, W. Liao, and L. Zhou, "Automatic fundus images mosaic based on sift feature," in *Proceedings the 3rd International Congress on Image and Signal Processing*, pp. 6:2747 – 2751, 2010.

[191] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[192] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proceedings Int Conf Computer Vision*, pp. 1:10–17, 2003.

[193] G. Mori, X. Ren, A. Efros, and J. Malik, "Recovering human body configurations: Combining segmentation and recognition," in *Proceedings IEEE Conf Computer Vision and Pattern Recognition*, 2004.

[194] A. Radhakrishna, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels," *Technical Report 149300, EPFL*, 2010.

[195] R. Caruana, "Multitask learning," *Machine Learning*, vol. 28, pp. 41–75, July 1997.

[196] S. Thrun, "Is learning the n-th thing any easier than learning the first?," *In Advances in Neural Information Processing Systems, MIT Press*, vol. 8, pp. 640–646, 1996.

[197] L. Mihalkova, T. Huynh, and R. Mooney, "Mapping and revising markov logic networks for transfer learnin," in *Proceedings the 22nd AAAI Conference on Artificial Intelligence*, pp. 608–614, 2007.

[198] N. A. Holtzman, P. D. Murphy, M. S. Watson, and P. A. Barr, "Predictive genetic testing: from basic research to clinical practice," *Science*, vol. 278, no. 5338, pp. 602–605, 1997.

[199] P. Sanfilippo, A. Hewitt, C. Hammond, and D. Mackey, "The heritability of ocular traits," *Survey of Ophthalmology*, vol. 55, no. 6, pp. 561 – 583, 2010.

[200] R. Plomin, J. DeFries, and G. McClearn, *Behavioral Genetics*. New York: World Publishers, 2001.

[201] A. M. Herskind, M. McGue, N. V. Holm, T. I. Sörensen, B. Harvald, and J. W. Vaupel, "The heritability of human longevity: a population-based study of 2872 danish twin pairs born 1870-1900," *Human genetics*, vol. 97, no. 3, pp. 319–323, 1996.

[202] D. Karasik, S. Demissie, L. A. Cupples, and D. P. Kiel, "Disentangling the genetic determinants of human aging: biological age as an alternative to the use of survival measures," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 60, no. 5, pp. 574–587, 2005.

[203] R. Schnoll, T. Johnson, and C. Lerman, "Genetics and smoking behavior," *Curr Psychiatry Rep*, vol. 9, no. 5, pp. 349–57, 2007.

[204] J. Vink, G. Willemsen, and D. Boomsma, "Heritability of smoking initiation and nicotine dependence.," *Behav Genet*, vol. 35, no. 4, pp. 397–406, 2005.

[205] W. Jablonski, "A contribution to the heredity of refraction in human eyes," *Arch Augenheilk*, vol. 91, pp. 308–28, 1922.

[206] F. Fajnkuchen and S. Cohen, "Update on the genetics of age-related macular degeneration," *French Journal of Ophtalmology*, vol. 31, no. 6 Pt 1, pp. 630–7, 2008.

[207] K. Antoniak, W. Bienias, and J. Nowak, "Age-related macular degeneration-a complex genetic disease," *Klin Oczna*, vol. 110, no. (4-6), pp. 211–8, 2008.

[208] H. Scholl, M. Fleckenstein, I. P. Charbel, C. Keilhauer, F. Holz, and B. Weber, "An update on the genetics of age-related macular degeneration," *Molecular Vision*, vol. 7, no. 13, p. 196–205, 2007.

[209] J. Seddon, J. Cote, W. Page, S. Aggen, and M. Neale, "The us twin study of age-related macular degeneration: relative roles of genetic and environmental influences.," *Arch Ophthalmol*, vol. 123, no. 3, pp. 321–7, 2005.

[210] C. Hammond, A. Webster, H. Snieder, A. Bird, C. Gilbert, and T. Spector, "Genetic influence on early age-related maculopathy: a twin study.," *Ophthalmology*, vol. 109, no. 4, pp. 730–6, 2002.

[211] I. C. Munch, B. Sander, L. Kessel, J. L. Hougaard, N. C. B. B. Taarnhoj, T. I. Sorensen, K. O. Kyvik, and M. Larsen, "Heredity of small hard drusen in twins aged 20-46 years," *Invest Ophthalmol Vis Sci*, vol. 48, no. 2, pp. 833–838, 2007.

[212] T. Toh, S. Liew, J. MacKinnon, A. Hewitt, J. Poulsen, T. Spector, C. Gilbert, J. Craig, C. Hammond, and D. Mackey, "Central corneal thickness is highly heritable: the twin eye studies," *Invest Ophthalmol Vis Sci*, vol. 46, no. 10, pp. 3718–22, 2005.

[213] J. Charlesworth, P. Kramer, T. Dyer, V. Diego, J. Samples, J. Craig, D. Mackey, A. Hewitt, J. Blangero, and M. Wirtz, "The path to open-angle glaucoma gene discovery: endophenotypic status of intraocular pressure, cup-to-disc ratio, and central corneal thickness.," *Invest Ophthalmol Vis Sci*, vol. 51, no. 7, pp. 3509–14, 2010.

[214] B. Klein, R. Klein, and K. Lee, "Heritability of risk factors for primary open-angle glaucoma: the beaver dam eye study," *Invest Ophthalmol Vis Sci*, vol. 45, pp. (1):59–62, 2004.

[215] M. Dirani, A. Islam, S. Shekar, and P. Baird, "Dominant genetic effects on corneal astigmatism: the genes in myopia (gem) twin study.," *Invest Ophthalmol Vis Sci*, vol. 49, no. 4, pp. 1339–44, 2008.

[216] N. Congdon, K. Broman, H. Lai, B. Munoz, H. Bowie, D. Gilbert, R. Wojciechowski, and S. West, "Cortical, but not posterior subcapsular, cataract shows significant familial aggregation in an older population after adjustment for possible shared environmental factors.," *Ophthalmology*, vol. 112, no. 1, pp. 73–77, 2005.

[217] C. Hammond, D. Duncan, H. Snieder, M. de Lange, S. West, T. Spector, and C. Gilbert, "The heritability of age-related cortical cataract: the twin eye study.," *Invest Ophthalmol Vis Sci*, vol. 42, no. 3, pp. 601–5, 2001.

[218] J. Teikari, "Genetic factors in open-angle (simple and capsular) glaucoma. a population-based twin study.," *Acta Ophthalmol*, vol. 65, no. 6, pp. 715–20, 1987.

[219] P. Alsbirk, "Anterior chamber depth and primary angle-closure glaucoma. ii. a genetic study.," *Acta Ophthalmol*, vol. 53, no. 3, pp. 436–49, 1975.

[220] Y. Tu, Z. Yin, H. Pen, and C. Yuan, "Genetic heritability of a shallow anterior chamber in chinese families with primary angle closure glaucoma.," *Ophthalmic Genet*, vol. 29, no. 4, pp. 171–6, 2008.

[221] J. Teikari, M. Koskenvuo, J. Kaprio, and J. O'Donnell, "Study of gene-environment effects on development of hyperopia: a study of 191 adult twin pairs from the finnish twin cohort study.," *Acta Genet Med Gemellol*, vol. 39, no. 1, pp. 133–6, 1990.

[222] M. Lee, S. Cho, H. Kim, Y. Song, K. Lee, J. Kim, D. Kim, T. Chung, Y. Kim, J. Seo, D. Ham, and J. Sung, "Epidemiologic characteristics of intraocular pressure in the korean and mongolian populations: the healthy twin and the gendiscan study.," *Ophthalmology*, vol. 119, no. 3, pp. 450–7, 2012.

[223] E. Forsman, R. Cantor, A. Lu, A. Eriksson, J. Fellman, I. Jrvel, and H. Forsius, "Exfoliation syndrome: prevalence and inheritance in a subisolate of the finnish population.," *Acta Ophthalmol Scand*, vol. 85, no. 5, pp. 500–7, 2007.

[224] F. Carbonaro, T. Andrew, D. Mackey, T. Young, T. Spector, and C. Hammond, "Repeated measures of intraocular pressure result in higher heritability and greater power in genetic linkage studies.," *Invest Ophthalmol Vis Sci*, vol. 50, no. 11, pp. 5115–9, 2009.

[225] M. Heitmann, H. Hamann, R. Brahm, H. Grussendorf, C. Rosenhagen, and O. Distl, "Analysis of prevalence of presumed inherited eye diseases in entlebucher mountain dogs," *Vet Ophthalmol*, vol. 8, no. 3, pp. 145–51, 2005.

[226] C. Hammond, H. Snieder, T. Spector, and C. Gilbert, "Genetic and environmental factors in age-related nuclear cataracts in monozygotic and dizygotic twins.," *N Engl J Med*, vol. 342, no. 24, pp. 1786–90, 2000.

[227] N. Lyhne, A. Sjlie, K. Kyvik, and A. Green, "The importance of genes and environment for ocular refraction and its determiners: a population based study among 20-45 year old twins.," *Br J Ophthalmol*, vol. 85, no. 12, pp. 1470–6, 2001.

[228] M. Tsai, L. Lin, V. Lee, C. Chen, and Y. Shih, "Estimation of heritability in myopic twin studies.," *Jpn J Ophthalmol*, vol. 53, no. 6, pp. 615–622, 2009.

[229] B. Gilmartin, "Myopia: precedents for research in the twenty-first century.," *Clin Experiment Ophthalmol*, vol. 32, no. 3, pp. 305–24, 2004.

[230] M. P. Stefan, "Genetic linkage analysis," *Arch Neurol*, vol. 56, no. 6, pp. 667–672, 1999.

[231] Wellcome Trust Case Control Consortium, "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.," *Nature*, vol. 447, no. 7145, pp. 661–678, 2007.

[232] McKusick-Nathans Institute of Genetic Medicine and Johns Hopkins University (Baltimore, MD), "Online mendelian inheritance in man, omim, http://omim.org/."

[233] J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, *et al.*, "The sequence of the human genome," *science*, vol. 291, no. 5507, pp. 1304–1351, 2001.

[234] National Human Genome Research Institute, "Fact sheets: Genome-wide association studies, http://www.genome.gov/20019523," 2013.

[235] R. Klein, C. Zeiss, E. Chew, R. Tsai, J.Y.and Sackler, and C. Haynes, "Complement factor h polymorphism in age-related macular degeneration," *Science*, vol. 308, pp. 385–9, April 2005.

[236] J. D. Cooper, D. J. Smyth, A. M. Smiles, V. Plagnol, N. M. Walker, J. E. Allen, K. Downes, J. C. Barrett, B. C. Healy, J. C. Mychaleckyj, *et al.*, "Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci," *Nature genetics*, vol. 40, no. 12, pp. 1399–1401, 2008.

[237] H.-C. Fung, S. Scholz, M. Matarin, J. Simon-Sanchez, D. Hernandez, A. Britton, J. R. Gibbs, C. Langefeld, M. L. Stiegert, J. Schymick, *et al.*, "Genome-wide genotyping in parkinson's disease and neurologically normal controls: first stage analysis and public release of data," *The Lancet Neurology*, vol. 5, no. 11, pp. 911–916, 2006.

[238] M. Larson, L. Atwood, E. Benjamin, L. A. Cupples, R. D'Agostino, C. Fox, D. Govindaraju, C.-Y. Guo, N. Heard-Costa, S.-J. Hwang, *et al.*, "Framingham heart study 100k project: genome-wide associations for cardiovascular disease outcomes," *BMC medical genetics*, vol. 8, no. Suppl 1, p. S5, 2007.

[239] A. Scuteri, S. Sanna, W.-M. Chen, M. Uda, G. Albai, J. Strait, S. Najjar, R. Nagaraja, M. Orru, G. Usala, *et al.*, "Genome-wide association scan shows genetic variants in the fto gene are associated with obesity-related traits," *PLoS genetics*, vol. 3, no. 7, p. e115, 2007.

[240] C. Kooperberg, M. LeBlanc, and V. Obenchain, "Risk prediction using genome-wide association studies," *Genetic Epidemiology*, vol. 34, no. 7, pp. 643–652, 2010.

[241] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. Ferreira, and D. Bender, "Plink: a toolset for whole-genome association and population-based linkage analysis," *American Journal of Human Genetics*, vol. 81, no. 3, pp. 559–75, 2007.

[242] J. Marchini and B. Howie, "Genotype imputation for genome-wide association studies," *Nature Reviews Genetics*, vol. 11, no. 7, pp. 499–511, 2010.

[243] X. Wan, C. Yang, Q. Yang, H. Yang, H. Xue, and X. Fan, "Boost: A fast approach to detecting gene-gene interactions in genome-wide case-control studies," *Am J Hum Genet*, vol. 87, no. 3, pp. 325–340, 2010.

[244] X. Zhang, S. Huang, F. Zou, and W. Wang, "Team: efficient two-locus epistasis tests in human genome-wide association study," *Bioinformatics*, vol. 26, no. 12, pp. i217–227, 2010.

[245] J. Wu, B. Devlin, and S. Ringquist, "Screen and clean: A tool for identifying interactions in genome-wide association studies," *Genetic Epidemiology*, vol. 34, no. 3, pp. 275–285, 2010.

[246] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal. Statist. Soc. B.*, vol. 58, no. 1, pp. 267–288, 1996.

[247] T. Wu, Y. Chen, T. Hastie, E. Sobel, and K. Lange, "Genome-wide association analysis by lasso penalized logistic regression," *Bioinformatics*, vol. 25, no. 6, pp. 714–21, 2009.

[248] A. Wu, C. Aporntewan, D. Ballard, J. Lee, J. Lee, and H. Zhao, "Two-stage joint selection method to identify candidate markers from genome-wide association studies," *BMC Proc*, vol. 3, no. 7, p. s29, 2009.

[249] C. Hoggart, J. Whittaker, M. De Iorio, and D. Balding, "Simultaneous analysis of all snps in genome-wide and re-sequencing association studies," *PLoS Genet*, vol. 4, no. 7, p. e1000130, 2008.

[250] G. D'Angelo, D. Rao, and C. Gu, "Combining least absolute shrinkage and selection operator (lasso) and principal-components analysis for detection of gene-gene interactions in genome-wide association studies," *BMC Proc*, vol. 3, no. 7, p. S62, 2009.

[251] C. Li, M. Li, E. Lange, and R. Watanabe, "Prioritized subset analysis: Improving power in genome-wide association studies," *Human Heredity*, vol. 65, no. 3, pp. 129–141, 2008.

[252] E. H. Shortliffe, "Mycin: A knowledge-based computer program applied to infectious diseases," in *Proceedings of Annu Symp Comput Appl Med Care*, pp. 66–9, 1977.

[253] R. Miller, H. J. Pople, and J. Myers, "Internist-i, an experimental computer-based diagnostic consultant for general internal medicine," *N Engl J Med*, vol. 307, no. 8, pp. 468–476, 1982.

[254] M. Drent, M. A. van Nierop, F. A. Gerritsen, E. F.Wouters, and P. G. Mulder, "A computer program using BALF-analysis results as a diagnostic tool in interstitial lung diseases," *Am J Respiratory Critical Care Medicine*, vol. 153, no. 2, pp. 736–41, 1996.

[255] S. Raza, Y. Sharma, Q. Chaudry, A. N. Young, and M. D.Wang, "Automated classification of renal cell carcinoma subtypes using scale invariant feature transform," in *Conf Proceedings IEEE Eng Med Biol Soc*, pp. 6687–690, 2009.

[256] P. Haug, P. D. Clayton, P. Shelton, T. Rich, I. Tocino, P. R. Frederick, R. O. Crapo, W. J. Morrison, , and H. R. Warner, "Revision of diagnostic logic using a clinical database," *Medical Decision Making*, vol. 9, no. 2, pp. 84–90, 1989.

[257] C. Xu, D. Tao, and C. Xu, "A survey on multi-view learning," *Corrabs/1304.5634*, 2013.

[258] K. B.Wagholikar, V. Sundararajan, and A. W. Deshpande, "Modeling paradigms for medical diagnostic decision support: A survey and future directions," *J Med Syst*, vol. 36, no. 5, pp. 3029–49, 2012.

[259] B. Boser, I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pp. 144–152, 1992.

[260] V. Vapnik and A. Lerner, "Pattern recognition using generalized portrait method," *Automation and Remote Control*, vol. 24, pp. 774–780, 1963.

[261] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[262] I. Vapnyarskii, "Lagrange multipliers," *in Hazewinkel, Michiel, Encyclopedia of Mathematics, Springer, ISBN 978-1-55608-010-4*, 2001.

[263] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, pp. 273–297, 1995.

[264] B. Scholkopf, A. Smola, and P. Williamson, R.C.and Bartlett, "New support vector algorithms," *Neural Computation*, vol. 12, no. 5, pp. 1207–1245, 2000.

[265] J. Shawe-Taylor and N. Cristianini, "Kernel methods for pattern analysis," *Cambridge University Press*, 2004.

[266] M. Aizerman, E. Braverman, and L. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition learning," *Automation and Remote Control*, vol. 25, pp. 821–837, 1964.

[267] G. Lanckriet, N. Cristianini, P. Bartlett, L. Ghaoui, and M. Jordan, "Learning the kernel matrix with semidefinite programming," *Journal of Machine Learning Research*, pp. 27–72, 2004.

[268] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee, "Choosing multiple parameters for support vector machines," *Machine Learning*, vol. 46, no. 1-3, pp. 131–159, 2002.

[269] K. Duan, S. Keerthi, and A. Poo, "Evaluation of simple performance measures for tuning svm hyperparameters," *Neurocomputing*, vol. 51, pp. 41–59, 2003.

[270] C. Ong, A. Smola, and R. Williamson, "Hyperkernels," *NIPS*, pp. 478–485, 2002.

[271] C. Ong and A. Smola, "Machine learning with hyperkernels," *ICML*, pp. 568–575, 2003.

[272] C. Ong, A. Smola, and R. Williamson, "Learning the kernel with hyperkernels," *Journal of Machine Learning Research*, vol. 6, pp. 1043–1071, 2005.

[273] I. W. Tsang and J. T. Kwok, "Large-scale sparsified manifold regularization," in *Advances in Neural Information Processing Systems 19*, pp. 1401–1408, MIT Press, 2007.

[274] G. Lanckriet, N. Cristianini, P. Bartlett, L. Ghaoui, and M. Jordan, "Learning the kernel matrix with semi-definite programming," *ICML*, 2002.

[275] C. Cortes, M. Mohri, and A. Rostamizadeh, "Learning non-linear combinations of kernels," *In Advances in Neural Information Processing Systems (NIPS)*, pp. 396–404, 2009.

[276] F. Zhuang, P. Luo, H. Xiong, Q. He, Y. Xiong, and Z. Shi, "Exploiting associations between word clusters and document classes for cross-domain text categorization," *Statistical Analysis and Data Mining*, vol. 4, pp. 100–114, February 2011.

[277] S. Sonnenburg, G. Ratsch, and C. Schafer, "A general and efficient multiple kernel learning algorithm," *Journal of Machine Learning Research*, vol. 7, pp. 1531–1565, 2006.

[278] A. Rakotomamonjy, F. R. Bach, S. Canu, and Y. Grandvalet, "Simplemkl," *Journal of Machine Learning Research*, vol. 9, pp. 2491–2521, November 2008.

[279] L. Bottou and C. Lin, "Support vector machine solvers," *Large Scale Kernel Machines*, pp. 1–28, 2007.

[280] G. Lanckriet, T. De-Bie, N. Cristianini, M. Jordan, and W. Noble, "A statistical framework for genomic data fusion," *Bioinformatics*, vol. 20, no. 16, pp. 2626–35, 2004.

[281] F. R. Bach, G. R. G. Lanckriet, and M. I. Jordan, "Multiple kernel learning, conic duality, and the smo algorithm," in *Proceedings of International Conference on Machine Learning*, (Banff, Alberta, Canada), pp. 6–13, ACM, July 2004.

[282] Z. Xu, R. Jin, K. Rong, L. King, and M. Lyu, "An extended level method for efficient multiple kernel learning," *In Advances in Neural Information Processing Systems 21 (NIPS)*, pp. 1825–1832, 2009.

[283] F. Bach, "Consistency of the group lasso and multiple kernel learning," *Journal of Machine Learning Research*, vol. 9, pp. 1179–1225, 2008.

[284] Z. Xu, R. Jin, H. Yang, I. King, and M. Lyu, "Simple and efficient multiple kernel learning by group lasso," *Proceeding of the 29th International Conference of Machine Learning*, 2010.

[285] M. Kloft, B. Marius, U. Brefeld, P. Laskov, and S. Sonnenburg, "Non-sparse multiple kernel learning," *In NIPS workshop on Kernel Learning: Automatic Selection of Optimal Kernels*, 2008.

[286] N. Saketha, G. Dinesh, S. Raman, C. Bhattacharyya, A. Ben-Tal, and K. Ramakrishnan, "On the algorithmics and applications of a mixed-norm based kernel learning formulation," *In Advances in Neural Information Processing Systems*, vol. 22, pp. 844–852, 2009.

[287] M. Szafranski, Y. Grandvalet, and A. Rakotomamonjy, "Composite kernel learning," pp. 1040–1047, 2008.

[288] M. Varma and B. Babu, "More generality in efficient multiple kernel learning," in *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 1065–1072, 2009.

[289] S. Sonnenburg, G. Ratsch, and C. Schafer, "Learning interpretable svms for biological sequence classification," *In RECOMB 2005, LNBI 3500, pp 389-407. Springer-Verlag Berlin Heidelberg*, 2005.

[290] Z. Harchaoui and F. Bach, "Image classification with segmentation graph kernels," in *Proceedings of Computer Vision and Pattern Recognition*, pp. 1–8, 2007.

[291] A. Vedaldi, V. Gulshan, M. Varma, and A. Zisserman, "Multiple kernels for object detection," in *Proceedings of the International Conference on Computer Vision (ICCV)*, 2009.

[292] Z. Zhang, J. Cheng, J. Liu, C. Yeo, C. Chui, and S. Saw, "Pathological myopia detection from selective fundus image features," *Proceedings 6th IEEE Conference on Industrial Electronics and Applications*, pp. 1–2, 2012.

[293] website, *http://www.genome.gov/sequencingcosts/*. 2012.

[294] C. Laurie, K. Doheny, D. Mirel, E. Pugh, L. Bierut, B. T., B. F., C. N.E., and C. M. et al., "Quality control and quality assurance in genotypic data for genome-wide association studies," *Genetic Eqidemiology*, no. 34, pp. 591–602, 2010.

[295] J. Wigginton, D. Cutler, and G. Abecasis, "A note on exact tests of hardy-weinberg equilibrium," *American Journal of Human Genetics*, vol. 76, no. 5, pp. 887–93, 2005.

[296] B. Curtin, *The Myopias: Basic Science and Clinical Management*. Philadelphia: Harper & Row, 1985.

[297] T. Young, "Ophthalmic genetics/inherited eye disease," *Curr Opin Ophthalmol*, vol. 14, no. 5, pp. 296–303, 2003.

[298] D. Stambolian, G. Ibay, L. Reider, D. Dana, and C. Moy, "Genomewide linkage scan for myopia susceptibility loci among ashkenazi jewish families shows evidence of linkage on chromosome 22q12," *Am J Hum Genet.*, vol. 75, no. 3, pp. 448–59, 2004.

[299] R. Wojciechowski, C. Moy, E. Ciner, G. Ibay, L. Reider, J. Bailey-Wilson, and D. Stambolian, "Genomewide scan in ashkenazi jewish families demonstrates evidence of linkage of ocular refraction to a qtl on chromosome 1p36," *Hum Genet.*, no. 4, pp. 389–99, 2006.

[300] Y. Li, L. Goh, and C. Khor, "Genome-wide association studies reveal genetic variants in ctnnd2 for high myopia in singapore chinese," *Ophthalmology*, vol. 118, no. 2, pp. 368–75, 2010.

[301] Q. Fan, V. Barathi, C. Cheng, X. Zhou, A. Meguro, and I. Nakata, "Genetic variants on chromosome 1q41 influence ocular axial length and high myopia," *PLoS Genetics*, no. 6, 2012.

[302] A. Solouki and V. Verhoeven, "A genome-wide association study identifies a susceptibility locus for refractive errors and myopia at 15q14," *Nat Genet.*, vol. 42, no. 10, pp. 897–901, 2010.

[303] H. Nakanishi, R. Yamada, N. Gotoh, and *et al.*, "A genome-wide association analysis identified a novel susceptible locus for pathological myopia at 11q24.1," *PLoS Genetics*, vol. 5, p. e1000660, September 2009.

[304] Z. Li and J. Qu, "A genome-wide association study reveals association between common variants in an intergenic region of 4q25 and high-grade myopia in the chinese han population," *Hum Mol Genet.*, vol. 15, pp. 2861–8, July 2011.

[305] Catalog of published Genome Wide Association Studies *WWW: http://www.genome.gov/gwastudies/*, 2013.

[306] T. I. H. Consortium, "A second generation human haplotype map of over 3.1 million snps," *Nature*, vol. 449, pp. 851–861, 2007.

[307] C. Carlson, M. Eberle, M. Rieder, Q. Yi, L. Kruglyak, and D. Nickerson, "Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium," *Am. J. Hum. Genet*, vol. 74, pp. 106–120, 2004.

[308] F. Li and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *Proceedings of IEEE Conf. on Comp Vis Pattern Recognition*, pp. 524–531, June 2005.

[309] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of IEEE Conf. on Comp Vis Pattern Recognition*, pp. 886–893, 2005.

[310] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, I. Kadir, and L. van Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, pp. 43–72, November 2005.

[311] R. Fan, K. Chang, and C. Hsieh, "Liblinear: A library for large linear classification," *Journal of Machine Learning Research*, vol. 9, pp. 1871–4, 2008.

[312] T. Chang, N. Congdon, R. Wojciechowski, B. Munoz, D. Gilbert, P. Chen, D. Friedman, and S. West, "Determinants and heritability of intraocular pressure and cup-to-disc ratio in a defined older population," *Ophthalmology*, vol. 112, no. 7, pp. 1186–91, 2005.

[313] C. C. Khor, W. D. Ramdas, E. N. Vithana, B. K. Cornes, X. Sim, W. T. Tay, and S. M. Saw, "Genome-wide association studies in asians confirm the involvement of atoh7 and tgfbr3, and further identify card10 as a novel locus influencing optic disc area," *Hum Mol Genet.*, vol. 20, no. 9, pp. 1864–72, 2011.

[314] W. Ramdas, "A genome-wide association study of optic disc parameters," *Plos Genetics.*, p. 6:e1000978, 2010.

[315] L. van Koolwijk and M. Ramdas, "Common genetic determinants of intraocular pressure and primary open-angle glaucoma," *PLoS Genet.*, vol. 8, no. 5, p. e1002611, 2012.

[316] S. Sherry, W. Ward, M. Kholodov, J. Baker, L. Phan, W. Smigielski, and K. Sirotkin, "dbsnp: the ncbi database of genetic variation," *Nucleic Acids Res.*, vol. 29, no. 1, pp. 308–11, 2001.

[317] C. Ho and C. Lin, "Large-scale linear support vector regression," *Technical report*, 2012.

[318] R. Cantor, K. Lange, and J. Sinsheimer, "Prioritizing gwas results: A review of statistical methods and recommendations for their application," *Am J Hum Genet*, vol. 86, no. 1, pp. 6–22, 2010.

[319] J. Liu, S. Ji, and J. Ye, *SLEP: Sparse Learning with Efficient Projections*. Arizona State University, http://www.public.asu.edu/ jye02/Software/SLEP, 2009.

[320] S. S. Bucak, R. Jin, and A. K. Jain, "Multiple kernel learning for visual object recognition: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. preprints, 2013.

[321] M. Guillaumin, J. Verbeek, and C. Schmid, "Multimodal semi-supervised learning for image classification," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 902–909, IEEE Computer Society, June 2010.

[322] V. Gál, E. Kerre, and M. Nachtegael, "Multiple kernel learning based modality classification for medical images," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition Workshops*, pp. 76–83, IEEE Computer Society, June 2012.

[323] C. Hinrichs, V. Singh, G. Xu, and S. Johnson, "Mkl for robust multi-modality ad classification," *Med Image Comput Comput Assist Interv.*, vol. 12, no. Pt 2, pp. 786–794, 2009.

[324] J. Liu, Z. Zhang, D. Wong, Y. Xu, F. Yin, J. Cheng, N. Tan, C. Kwoh, D. Xu, Y. Tham, T. Aung, and T. Wong, "Automatic glaucoma diagnosis through medical imaging informatics," *J Am Med Assoc*, vol. 20, no. 6, pp. 1021–7, 2013.

[325] L. Torresani, M. Szummer, and A. Fitzgibbon, "Efficient object category recognition using classemes," in *Proceedings of the European Conference on Computer Vision*, pp. 776–789, 2010.

[326] J. B. Jonas, W. M. Budde, and S. Panda-Jonas, "Ophthalmoscopic evaluation of the optic nerve head," *Survey of Ophthalmology*, vol. 43, pp. 293–320, January–February 1999.

[327] Y. Xu, D. Xu, S. Lin, J. Liu, J. Cheng, C. Y. Cheung, T. Aung, and T. Y. Wong, "Efficient optic cup detection from intra-image learning with retinal structure priors," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 58–65, 2012.

[328] F. Yin, J. Liu, S. H. Ong, D. Sun, D. W. K. Wong, N. M. Tan, M. Baskaran, C. Y. Cheung, T. Aung, and T. Y. Wong, "Model-based optic nerve head segmentation on retinal fundus images," in *Proceedings of the International Conference on Engineering in Medicine and Biology Society*, pp. 2626–2629, 2011.

[329] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, November 2004.

[330] L. Duan, D. Xu, I. W. Tsang, and J. Luo, "Visual event recognition in videos by learning from web data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 1667–1680, September 2012.

[331] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines." Software available at http://www.csie.ntu.edu.tw/cjlin/libsvm, 2001.

[332] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, pp. 199–222, August 2004.

[333] R. E. Schapire, "A brief introduction to boosting," in *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 1401–1406, 1999.