Fuzzy Inference System for Simulating Ophthalmologists in the Diagnosis of Diabetic Retinopathy

by

JiHo Han

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science
(Artificial Intelligence)
in the University of Michigan-Dearborn
2023

Master's Thesis Committee:

Professor Adnan K. Shaout, Chair
Assistant Professor Fred Feng
Lecturer Khalid Kattan

Dedication

This thesis is dedicated to my father, who is also experiencing diabetic retinopathy.

"In this world, you will have trouble.
But take heart! I have overcome the world."

John 16:33

Acknowledgements

First and foremost, I would like to thank my thesis advisor, Dr. Adnan K. Shaout, for their guidance, patience, and insightful feedback. Their expertise, mentorship, and availability have been instrumental in shaping my research and helping me reach this milestone.

I would also like to extend my appreciation to the University of Michigan staff members who helped me acquire the necessary information for the thesis throughout my study. Completion of this research would never have been possible without their efforts.

On a personal note, I would like to thank my parents for their unwavering support, encouragement, and love throughout my academic journey. Their constant guidance, motivation, faith, and sacrifices have been invaluable and have been the driving force behind my success. Especially, I am particularly driven to present this research in honor of my father, who, facing a period of ailment, has instilled within me a strong desire to delve into related subjects and further my understanding.

Finally, I thank God for all His blessings.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

AI: Artificial intelligence

CAD: Computer-assisted diagnosis

FP: Fundus Photography

DR: Diabetic Retinopathy

NPDR: Non-proliferative Diabetic Retinopathy

PDR: Proliferative Diabetic Retinopathy

FIS: Fuzzy Inference System

Abstract

With the rapid advancement of artificial intelligence (AI) technology, Computer-Assisted Diagnosis (CAD) has emerged as a prominent player in modern medical diagnostics, revolutionizing the way diseases are detected, diagnosed, and assessed. Despite the enthusiasm surrounding CAD, however, there is a noticeable gap in the research landscape - to be specific, a scarcity of thorough studies that rigorously compare its diagnostic capabilities and viewpoints with those of medical experts. This study aims to fill this void by subjecting the innovative fuzzy method, designed to replicate the methodologies of ophthalmologists in diagnosing diabetic retinopathy (DR), one of the leading causes of blindness.

Using general case studies, a comprehensive review is conducted regarding the advantages and limitations inherent in each respective methodology. Through this process, the current problem of diagnosis is identified, and by using multiple diagnostic criteria for diabetic retinopathy, a diagnosis system based on a fuzzy inference system is created. Testing using FGADR and APTOS datasets acquired an accuracy of 77.59% and 98% respectively – overall achieving an accuracy of 87.04%. Although its value is lower in comparison to state-of-the-art performance (sensitivity of 91.9% and specificity of 91.3%), the advantages it could provide show promising alternatives that could be researched. While testing through datasets, edge case testing is conducted to check whether the proposed system correctly identifies misidentified data, validating the objectives of the research.

This study will provide fresh insights into critically analyzing both the traditional diagnostic approach by human ophthalmologists and the innovative next-generation diagnostic technologies driven by AI, within their respective contexts. The availability of technology that mimics human doctors' decision-making will ultimately facilitate the analysis of the current state of the technology.

Keywords: Diabetic Retinopathy, Fundus Photography, Computer-Assisted Diagnosis, Fuzzy Inference Systems

CHAPTER 1 Introduction

This chapter concentrates on elucidating the intricate structures of the fundus, delineating the manifestations of diabetic retinopathy, and emphasizing its global significance in ocular health. Subsequently, the discussion turns to the distinctive attributes of fundus photography, chosen as the primary diagnostic modality for this study. The chapter navigates through the rationale behind this selection, encompassing factors such as accessibility, cost-effectiveness, and its established role in clinical practice. The overarching aim is to contribute substantial insights to the domain of ophthalmology, enriching our understanding of diabetic retinopathy for the betterment of clinical outcomes.

1.1 Understanding Fundus Anatomy

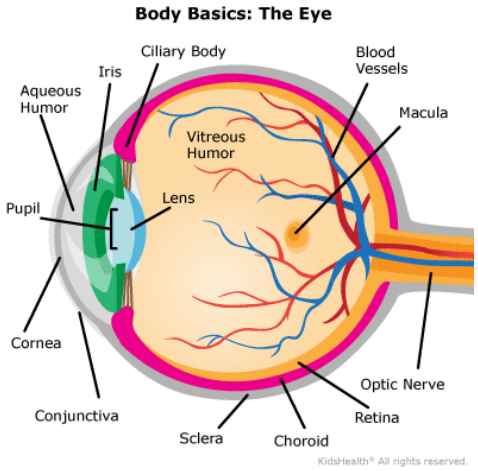


Figure 1.1: Basic Anatomy of Human Eyes [1]

Although we take it for granted, the human eye is a complex organ. Understanding the role of its parts, especially organs that are located within the fundus (e.g., retina, macula) is crucial for understanding the diagnosis and treatment of eye diseases.

The fundus in ophthalmology refers to the interior surface of the eye opposite the lens, as shown in Figure 1.1. This region comprises intricate components for visual function, including the retina, optic disc, macula, and blood vessels that supply the retina.

At the back of the eye lies the retina, a layer of tissue containing photoreceptor cells that convert light into neural signals. The retina plays a critical role in vision, as it is responsible for transmitting visual information to the brain.

The macula, a small central area of the retina, is responsible for sharp, detailed vision. It contains a high concentration of photoreceptor cells called cones, which are responsible for color vision and visual acuity. The macula is essential for activities such as reading, driving, and recognizing faces.

1.2 Manifestation of Diabetic Retinopathy

Diabetic retinopathy, a complication stemming from diabetes, exerts its impact on ocular health by primarily targeting the small blood vessels. The escalation of blood sugar levels in uncontrolled diabetes contributes to the thickening and fragility of these vessels, leading to vessel leakage or blockage. In the early stages of diabetic vascular damage, microaneurysms emerge as small, round, red dots, signifying protrusions in the retinal capillaries. Additionally, exudates such as cotton wool spots or hard exudates manifest, serving as indicators of localized

areas of retinal ischemia induced by nerve fiber layer infarcts. To make matters worse, the damaged blood vessel may result in the growth of new, abnormal blood vessels on the surface of the retina, potentially leading to retinal detachment and blindness, also known as neovascularization, as shown in Figure 1.2.

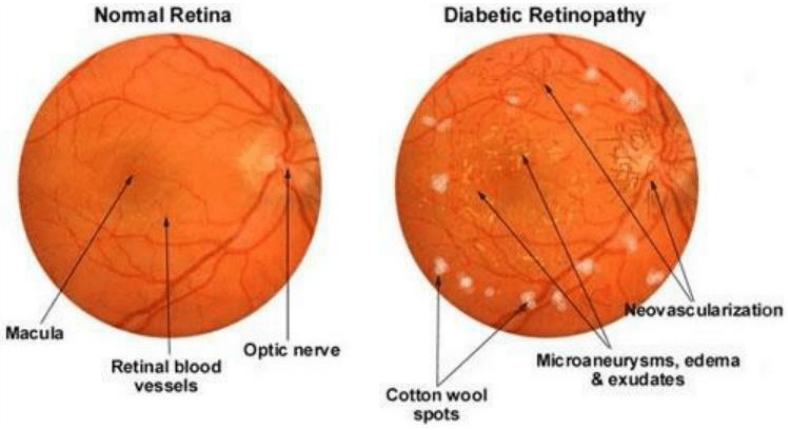


Figure 1.2: Difference between Normal Retina and Diabetic Retinopathy Fundus Image [2]

1.2.1 Stages of Diabetic Retinopathy

There are multiple ways of staging the severity of diabetic retinopathy. In this thesis, we mainly used the diagnostic criteria for diabetic retinopathy according to the International Clinical Diabetic Retinopathy (ICDR) Severity Scale for DR. The ICDR not only successfully combines the findings of the Early Treatment of Diabetic Retinopathy Study (ETDRS) and the Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR), but it's also the standard scale that is used widely for the staging of the disease. Table 1.1 shows the ICDR Severity Scale for DR.

Proposed disease severity level	Findings observable by dilated ophthalmoscopy
No diabetic retinopathy	No abnormalities
Mild NPDR	Microaneurysms only
Moderate NPDR	More than just microaneurysms but less than severe NPDR diabetic retinopathy
Severe NPDR	Any of the following: - more than 20 intraretinal hemorrhages in each of 4 quadrants; definite venous beading in 2 quadrants. - Prominent intraretinal microvascular abnormalities in 1 quadrant and no signs of proliferative retinopathy
PDR	One or more of the following: - Neovascularization - vitreous/pre-retinal hemorrhage

Table 1.1: ICDR Severity Scale for Diabetic Retinopathy [3]

1.3 Significance of the Problem

Diabetic Retinopathy stands as a formidable global concern, ranking prominently among the major causes of blindness. It holds the distressing distinction of being a leading contributor to visual impairment and blindness in adults on a worldwide scale. As a progressive complication of diabetes, this ocular ailment can initiate a cascade of pathological changes within the delicate structures of the eye's retina. If left unchecked, it has the potential to compromise visual acuity to a profound extent, ultimately culminating in significant vision loss or even complete blindness. A case study from TensorFlow showed that diabetic retinopathy is one of the fastest-growing causes of blindness, as over 415 million people suffer from diabetes [4].

This alarming prevalence underscores the pressing need for early detection, vigilant management, and ongoing monitoring to counteract its potentially devastating impact on visual health. Thankfully, a nationwide diabetic eye screening (DES) program was carried out in the

United Kingdom, and its result proved that extensive screening plays a pivotal role in the prevention and treatment of DR-induced blindness [5].

In the next chapters, two main methodologies employed in DR diagnosis – human DR diagnosis and CAD – and its pros and cons will be examined. With it, a solution to remedy the cons of each side will be proposed.

CHAPTER 2

Literature Review of Diabetic Retinopathy and its Diagnosis Methods

In this chapter, a comprehensive exploration of the landscape surrounding DR is undertaken, tracing its historical trajectory to lay a foundation for subsequent discussions. The literature review will provide valuable insight into the evolution of understanding of DR. Following the historical overview, a critique of human DR diagnosis and CAD is undertaken, dissecting its merits and demerits in detecting DR. A novel approach is then suggested to maximize the benefits of both methods while minimizing the drawbacks of both methods.

2.1 History of Diabetic Retinopathy Diagnosis

Despite the suspicion that the relationship between diabetes and retinopathy was suggested in the 1850s, it was 1968 when the Airlie House Symposium evaluated the knowledge of diabetic retinopathy (DR), and established the first standards for defining one quantitatively [12]. Some studies indicate that early pioneering work on screening for DR was conducted in the Baltic Sea region of Northern Europe in the late 1900s [13]–[15]. Such standards spread throughout Europe starting from the UK to USA and Singapore.

In the past decade, the steady increase in diabetes cases worldwide fueled the advancement of automated screening systems – many of which leverage AI technology. Early AI

software in diagnosing retinopathy was focused on identifying specific image features, such as microaneurysms [16]. However, as extracting vessels before detecting DR with machine learning is time-inefficient in both preparing the dataset for training and diagnosing and as the field of machine learning has experienced tremendous growth over time in both algorithms and hardware, another paradigm of AI arose – a data-driven deep learning AI system which distinguishes severity of DR by using nothing but retinal images without performing vessel segmentation. A wide variety of deep learning methods (e.g., transfer learning [17] or multi-layer convolutional neural networks (CNN) [18]–[20]) have been applied for DR detection, diagnosing, and staging [6]. On top of that, various preprocessing methods (e.g., normalization, data augmentation) were performed to further enhance the accuracy of the diagnosis.

A state-of-the-art review of CAD in the detection of DR in 2023 revealed that from 2020 and beyond, there is a significant drop in the number of studies that use preprocessing, as shown in Figure 2.2 [21]. The study pointed out that the advancement of CNN architectures allowed the model to track the minor details within the image without the need for preprocessing overall.

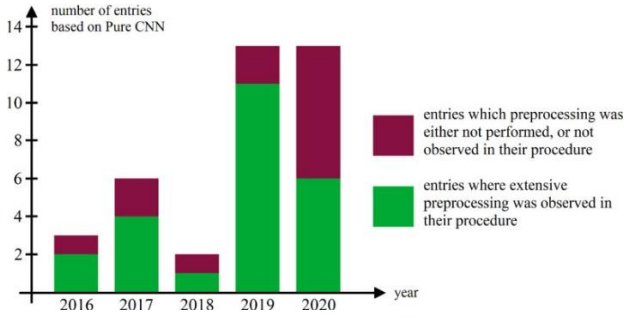


Figure 2.1: The Number of Entries That Had Preprocessing Stage Implemented [21]

In the same research with comparative studies ranging from 2016 to 2020, an accuracy ranging from 85.96% to 99.49% has been claimed. However, it is noteworthy that the author also

pointed out that the quantitative comparisons of accuracy do not conclude one model outperforms the other, as there remain many degrees of freedom involved in testing. This study shows that the area of diagnosis in DR is everchanging, and numerous attempts are being made not only to improve accuracy but also to improve the efficiency and ease of use of the diagnosis system.

Unequivocally, the modern cutting-edge methods for diabetes screening demonstrate impressive accuracy and efficiency in identifying individuals at risk or those already affected by this chronic condition. However, like any approach, they are not without their limitations. In the following sections, we will delve into the various challenges and drawbacks associated with these cutting-edge technologies. By examining these limitations, the study aims to gain a comprehensive understanding of the evolving landscape of diabetes screening and the pressing need for different approaches in making a diagnosis.

2.2 Advantages of Human Diabetic Retinopathy Diagnosis

One of the biggest advantages of human diagnosis comes from human interaction. Human diagnosticians excel at interpreting unstructured data, such as patient narratives, emotional cues, and context. Each patient is unique in their way, and so must be their diagnosis and treatment; in this sense, unstructured information can be vital in diagnosis but poses a challenge for the general CAD system to fully utilize in the diagnosing process.

2.3 Limitations of Human Diabetic Retinopathy Diagnosis

Despite the efforts, the human DR diagnosis poses many challenges to overcome. Firstly, ophthalmologists and medical practitioners often employ a multifaceted diagnostic approach to ensure accurate detection and assessment of the condition. This usually ensures a more accurate diagnosis and assessment of the given condition, but such a multifaceted approach might not only increase the overall difficulty in diagnosis but also require more time and resources, which could impact the efficiency of diagnosis and treatment. Research from Dr. Yin showed that the low rate of fundus examination due to limitations of medical resources delays the diagnosis and treatment of diabetic retinopathy – which validates the need for an automated diabetic retinopathy screening system.

Moreover, there are multitudes of tables that ophthalmologists can use to reference for diagnosing, which leads to a rather vague diagnosing system. Although there was an attempt to create an international clinical DR scale, some discrepancy remains, as different treatments may be required even though they are categorized in the same severity [22].

On top of that, certain studies discuss that ophthalmologists are inconsistent with their diagnosis. In the case study of TensorFlow in Medicine, Dr. Peng noted that even the most renowned doctors are surprisingly variable when it comes to categorizing the stages of DR. Despite the contestants being US board-certified ophthalmologists and well-known guidelines do exist, the lack of conciseness was evident in the study as shown in Figure 2.2. He did point out that there were good agreements amongst the experts about normal and PDR, but in between, there were many variabilities, disagreements, and fuzziness about where each disease should be categorized [4].

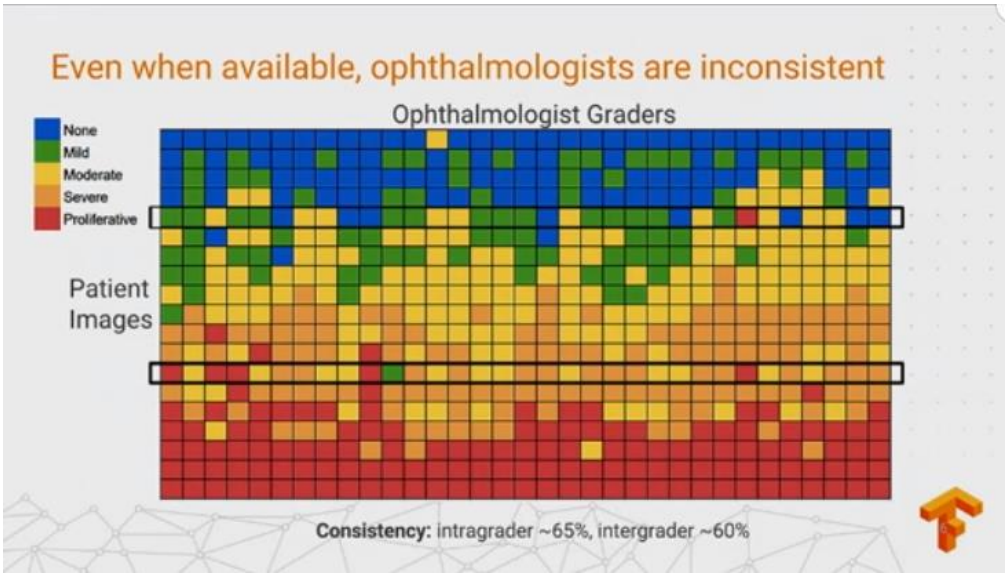


Figure 2.2: Case Study: TensorFlow in Medicine - Retinal Imaging [4]

2.4 Advantages of the Computer-Assisted Diagnosis in Diabetic Retinopathy Diagnosis

The utmost advantage of using CAD comes from its speed, availability, and consistency. Computer-assisted diagnosis systems can process and analyze vast amounts of data quickly while adhering to clinical guidelines, and the diagnoses deducted from it can be easily accessed from places where human experts aren't present. For instance, the automated system implemented during the study in Portugal showed the potential for a human grading burden reduction of 48.42% [23].

Additionally, with the advent of neural networks, through large-scale modeling and data analysis, AI-powered CAD systems can identify certain patterns and correlations that weren't evident in experts' observations, which may lead to finding additional diagnosis factors, improving the accuracy of the diagnosis.

2.5 Limitations of the Computer-Assisted Diagnosis in Diabetic Retinopathy Diagnosis

One of the biggest and most prominent issues of using CAD systems for diagnosis is that computers lack clinical judgment from critical thinking. Therefore, if healthcare professionals grow excessive reliance on CADs, it may result in deskilling them, potentially leading to increased misdiagnosis, especially if the conditions display rare or unusual conditions that were non-existent in the CAD database. Food and Drug Administration (FDA) Guidance on Non-Device Clinical Decision Support recommends that the medical information, including the output of CAD software, provide a list of preventive, diagnostic, or treatment options with logic or methods to provide such options in plain language description rather than providing a specific diagnostic or treatment plan [24], [25].

2.6 Types of Medical Image Modalities

Various medical imaging modalities crucial for the diagnosis and management of DR exist as shown in Figure 1.3. These modalities offer unique perspectives on ocular structures, aiding clinicians in comprehending the intricacies of the disease progression.

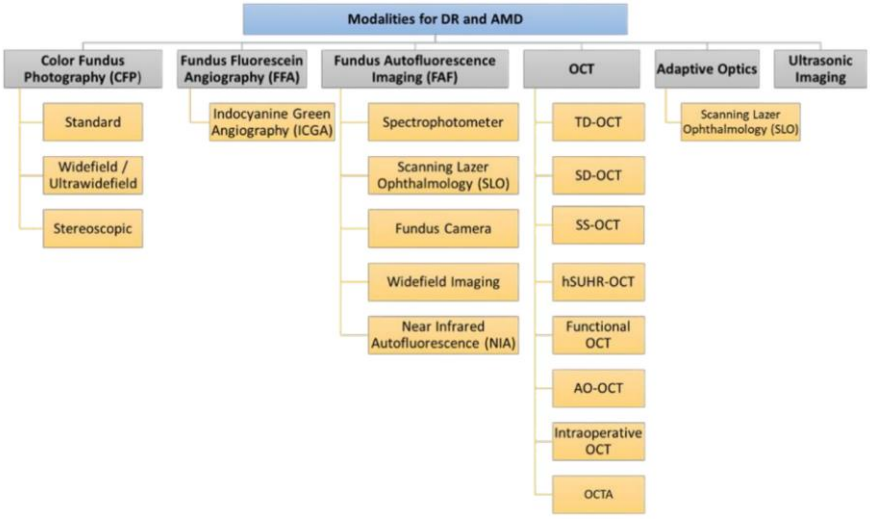


Figure 2.3: Medical Image Modalities for the Detection/Diagnosis/Staging of DR and AMD [6]

Fundus Photography (FP) stands as a fundamental modality, capturing detailed images of the retina. FP provides a comprehensive view of the ocular fundus, highlighting abnormalities such as hemorrhages, microaneurysms, and exudates. FP is valued for its non-invasive nature, widespread availability, and cost-effectiveness. However, its limitations include surface-level insights with restricted in-depth information, and an inability to visualize minuscule vascular changes, making it less suitable for detailed assessments.

In such a situation, Fundus Fluorescein Angiography (FFA) stands out for its ability to visualize blood flow dynamics and identify vascular abnormalities. The invasive nature of FFA, involving the injection of a fluorescent dye, is a significant drawback. Additionally, the procedure carries the risk of potential allergic reactions, which is another trade-off between diagnostic capability and patient discomfort.

Due to the invasive nature of FFA, Fundus Autofluorescence Imaging (FAF) introduces a non-invasive alternative by capturing the natural fluorescence emitted by the retinal pigment epithelium (RPE). Its non-invasive nature and widespread availability are significant advantages. However, limitations might include a reduced ability to provide detailed vascular information compared to other modalities.

Unlike the previous methodologies, Optical Coherence Tomography (OCT) offers distinctive cross-sectional images with high-resolution structural information, allowing precise assessments of retinal layers and providing precise assessment of retinal layers. However, since OCT relies on light waves, certain conditions such as vitreous hemorrhage, dense cataracts, or corneal clouding can introduce media opacities, thereby constraining the effectiveness of the

OCT scan in such cases [7]. One must keep in mind, however, that methods such as FP also get impacted by such symptoms.

Ocular Ultrasound, though less commonly used for DR, can be employed to evaluate the posterior segment of the eye. This is especially relevant in cases where other modalities may be challenging, such as in the presence of dense cataracts.

Table 1.2 shows the Pros and Cons of Ophthalmological Medical Image Modalities.

Imaging Modality	**Pros**	**Cons**
Fundus Photography (FP)	• Non-invasive • Widespread availability • Cost-effective	• Limited in-depth information • Inability to visualize miniscule vascular changes
Fluorescein Angiography (FFA)	• Visualizes blood flow • Easily identify vascular abnormalities	• Invasive • Potential allergic reaction
Fundus Autofluorescence Imaging (FAF)	• Non-invasive: no external dye used	• Non-invasive • Widespread availability
Optical Coherence Tomography (OCT)	• Cross-sectional images • High-resolution structural information • Precise assessment of retinal layers	• Distinctive in comparison to other modalities • Can be clouded by certain conditions
Ocular Ultrasound	• Dynamic Imaging	• Limited resolution

Table 2.1: Pros and Cons of Ophthalmological Medical Image Modalities

2.6.1 Why Fundus Photography?

Fundus photography, also known as fundoscopy or ophthalmoscopy, has emerged as a critical tool for diagnosing a wide range of ophthalmologic conditions. Due to its widespread

availability, it is increasingly being employed in CAD using AI. There are several compelling reasons why fundus photography has become an area of interest.

First and foremost, fundus photography allows for non-invasive visualization and documentation of the retina, optic nerve head, and surrounding vasculature.

It provides an overview of color image of the back of the eye in good resolution, allowing for detailed examination of the retina and its blood vessels. Moreover, its nature of non-invasiveness allows its procedure to be painless, making it adequate for a wide range of patients.

Fundus photography's capability to capture an overview of the eye also signifies that it provides a permanent record of the eye's appearance, which can be used for comparison with future images to monitor changes over time. This can be especially useful for tracking the progression of certain eye conditions and assessing the effectiveness of treatments. The availability of visual documentation leads to empowering patients to understand their current condition, and the importance of the treatment, successfully enhancing their commitment to managing their blood sugar levels in a healthy range and ultimately lowering the overall risk [8].

Research from the International Journal of Ophthalmology showed that fundus fluorescein angiography outperforms fundus photography in detection accuracy [9]. Due to the nature of differences in angiography and photography, it is quite natural that angiography shows better results in diseases where blood vessel assessment is crucial. However, many non-numerical indicators, including but not limited to the cost, complexity of the procedure as well as the level of patient discomfort show that fundus photography has its advantages.

Overall, fundus photography is a valuable tool for ophthalmologists in the diagnosis and management of various eye conditions, and its non-invasive nature makes it a safe and patient-friendly procedure.

2.7 Proposed Solution

To merge the benefits of both human diabetic retinopathy and CAD as can be seen from similar studies that were mentioned earlier, fuzzy logic and FIS are widely used in medical classification/staging methods.

The core concept of integrating fuzzy logic in diagnosis is to model the imprecise aspects of the behavior of the system through fuzzy sets and fuzzy rules [26]. Unlike conventional logic sets with crisp boundaries, a fuzzy set allows a gradual transition between two different sets, characterized through membership function. These characteristics are advantageous for diagnosing medical conditions for several reasons:

- Medical diagnosis often involves experts (ophthalmologists in our case) making medical decisions based on their knowledge and experience. Fuzzy logic is useful in emulating those human reasoning processes.
- Fuzzy logic can handle uncertainties and variabilities that can exist in the patient data effectively, ultimately resulting in more robust diagnoses.
- Fuzzy logic is suitable for adapting to changes in patient data and adjusting its diagnostic decisions accordingly.

2.7.1 Design Process for the Proposed Solution

Unlike other typical FIS, the proposed solution introduces two key design innovations aimed at enhancing the efficiency and credibility of the diagnostic system.

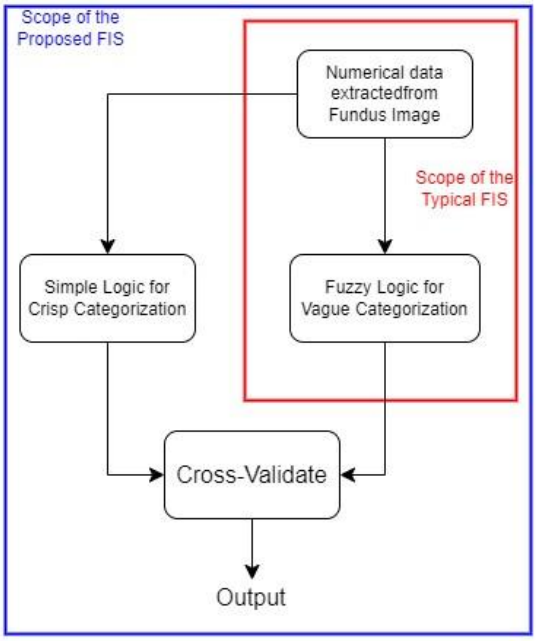


Figure 2.4: Difference in Scope of Proposed FIS in comparison to the Typical FIS

Firstly, the incorporation of simple logic alongside the conventional fuzzy logic in the inputs allows a notable reduction in the number of fuzzy rules. In addition, the system employs a cross-validation mechanism that compares the output of two distinct classification stages, successfully increasing credibility and validity while maintaining a low number of rules, as shown in Figure 2.3.

All these efforts result in a judiciously low number of rules. Only 63 rules were required to fully implement the system, which is 98.43% lower rules than the theoretical maximum number of rules available – 4000 (10 classifications of IRH * 10 classifications of MA * 2 classifications of HE * 5 classifications of IRMA * 2 classifications of CWS * 2 classification of NV/VH/PRH). This low number of rules will not only ease the difficulty of maintaining the system for doctors but also let patients easily understand the reasoning behind their diagnosis.

2.8 Previous Attempts

A few similar studies have been conducted and published for diagnosing DR accurately. In 2017, the Diabetic Retinopathy Detection System (DRDS) based on the Mamdani fuzzy inference system (FIS) was created [10], while in 2020, fuzzy based decision-making system for the Detection of Diabetic Retinopathy was proposed [11]. However, the implementation of DRDS utilized different inputs that are more related to the cause of diabetes such as visual field, fasting blood sugar, HDL, and LDL. In the case of the study in 2020 as well, the study used glucose levels in blood, intra-ocular pressure, etc. Both of those studies' input parameters require additional testing other than FP itself.

Unlike these approaches, our aim is to devise a diagnostic system whose decisions rely solely on parameters extractable from a single fundus photograph, making it a valuable resource for healthcare professionals. Attempting to make a diagnosis system whose decisions are based on parameters, all of which can be acquired from a single image, with human-explainable reasoning from professionals, will prove to be an asset in the future of the diagnosis system.

CHAPTER 3

DATASETS AND METHODS

This chapter introduces the foundational elements of the datasets and methods used in creating and evaluating the proposed diagnostic solution. An overview of the Fuzzy Inference System (FIS) will first be provided with a detailed explanation of the fuzzy subsystem that lies within. Afterwards, the parameters of the inputs and outputs for the system will be briefly explained. Two different datasets – FGADR and APTOS – will then be introduced, explaining why they were suitable for testing the overall system. The goal of this chapter is to establish a clear and transparent understanding of the technical aspects involved in the development and assessment of the system.

3.1 Fuzzy Inference System

To simulate the human ophthalmologists' diagnoses, the dataset was modified to have additional data required for diagnosis.

The proposed FIS model consists of a few fuzzy logic sub-systems, with generic inputs that can be acquired from the FP. Like the Patient-Reported Outcome Measure (PROM) questionnaires that are generally used in hospitals for triage, the inputs are numerically done, from the lowest being mild to the highest being severe. The FIS inputs are defined in Appendix A (Diagnostic Questionnaire for Fuzzy Inference System).

The rubrics are inspired by incorporating common aspects of DR severity scale tables [3], [22]. Note that the following inputs and categories are tentative and can be altered with the consultation of the experts, as with the betterment of the technology, research occurs that questions the validity of the table itself, while proposing a newer scale system.

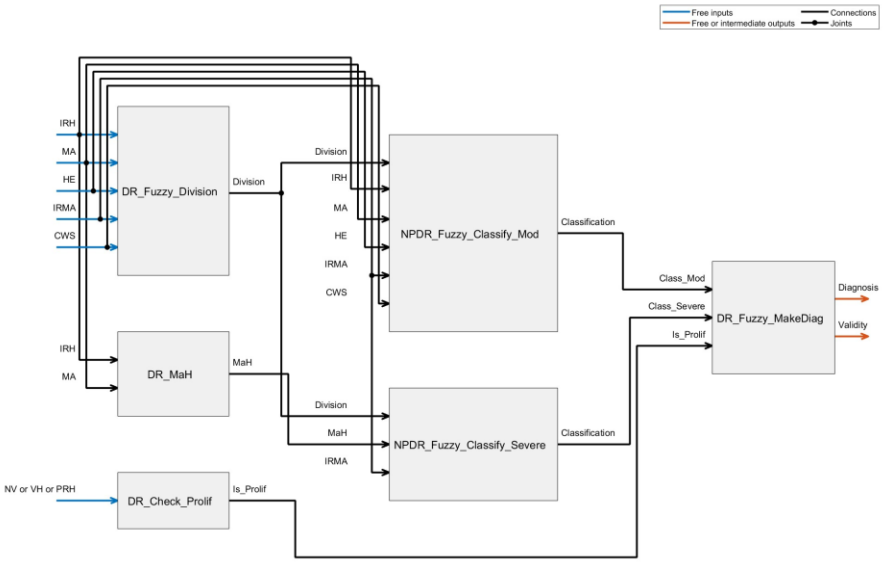


Figure 3.1: FIS Tree Model of Implemented System

The proposed fuzzy inference system (FIS) tree model consists of two simple logics – "*DR_Check_Prolif*" and "*DR_MaH*", and four fuzzy logics – "*DR_Fuzzy_Division*", "*NPDR_Fuzzy_Classify_Mod", "NPDR_ Fuzzy_Classify_Severe", and "DR_Fuzzy_MakeDiag"* as shown in Figure 3.1.

"*DR_Check_Prolif*" checks if the current state of DR is proliferative or non-proliferative. Based on the ICDR severity scale for DR [3], a DR is considered proliferative if the image shows NV or VH / PRH, no matter what other symptoms imply.

"DR_Fuzzy_Division" serves as the foundational layer of the diagnostic process, vaguely categorizing the current stage of the eye into 4 different membership outputs as division – normal, mild, moderate, and severe. This sets up the groundwork of subsequent fuzzy logic for more detailed diagnosis. Table 1 shows the fuzzy rules for the "*DR_Fuzzy_Division"*.

#	**Fuzzy Rule of *DR_Fuzzy_Division***	**Weight**
1	IRH==absent & MA==absent & HE~=present & IRMA==absent & CWS~=present => Division=normal	1
2	IRH==absent & MA==present in 1 quad & HE~=present & IRMA==absent & CWS~=present => Division=mild	1
3	MA==present in 4 quads | IRMA==present in 4 quads as mild or 1 quad as prominent => Division=severe	1
4	IRH==present in 1 quad & MA==absent & HE~=present & IRMA==absent & CWS~=present => Division=mild	0.5
5	HE==present | IRMA==present in 1 quad as mild | CWS==present => Division=moderate (0.5)"	0.5
6	HE==present | IRMA==present in 1 quad as mild | CWS==present => Division=severe	0.5
7	IRH==present in 2 quads | MA==present in 2 quads => Division=moderate	1
8	IRH==present in 3 quads | MA==present in 3 quads => Division=moderate	0.5
9	IRH==present in 3 quads | MA==present in 3 quads => Division=severe	0.5
10	IRH==present in 4 quads | MA==present in 4 quads => Division=severe	1
11	IRH==present in 3 quads & MA==present in 1 quad => Division=severe	1
12	IRH==present in 1 quad & MA==present in 3 quads => Division=moderate	1

Table 3.1: Fuzzy Rule of Logic *"DR_Fuzzy_Division"*

Rules 1-3 show the base case of normal/mild/severe division respectively based on the ICDR severity scale.

- If there are no symptoms at all, the eye is normal.

- If there are only microaneurysms (MA) and no other abnormalities, the eye has a mild NPDR.
- If the 4-2-1 rule is met (MA is prominent in four quadrants or intraretinal microvascular abnormalities (IRMA) are found in one of the quadrants of the retina), the eye has a severe NPDR.

Rule 4-6, 8, and 9 fill in the gap within Rule 1-3 for symptoms that can present in multiple stages of severity. For instance, if only intraretinal hemorrhage (IRH) is shown in one of the quadrants, it is diagnosed as mild based on the scale, but only given a weight of 0.5 instead of 1, as IRH usually is discoverable together with MA. This is because the weakened walls of the retinal blood vessels by MA usually induce the IRH as DR progresses.

Rule 7, 10-12 denotes the diagnostic cases based on the severity of MA and IRH of the eye. This is because the proliferation of the MA and IRH are directly related to diagnosing the severity of moderate NPDR to severe NPDR. Moreover, as the scale of MA and IRH is set much finer (0-10) in comparison to other metrics, more rules were defined to avoid random edge cases which result in misdiagnosis.

Using the output from "*DR_Fuzzy_Division"* together with necessary inputs for further classification, 2 fuzzy logic subsystems have been created – "*NPDR_Fuzzy_Classify_Mod"* and "*NPDR_ Fuzzy_Classify_Severe"*. Each system has the output of "*DR_Fuzzy_Division"* as a foundation of finer diagnosis. In the case of "*NPDR_ Fuzzy_Classify_Severe"*, several inputs other than necessary inputs for fine diagnosis have been pruned to ensure that the number of rules required for the diagnosis is reduced while maintaining all the necessary functions. The detailed diagnosis is branched out from the initial elementary diagnosis based on the table provided in Review of Optometry [22].

The fuzzy rules for the "*NPDR_Fuzzy_Classify_Mod"* and "*NPDR_Fuzzy_Classify_Severe"* can be found in Tables 3.2 and 3.3, respectively. In the case of "*NPDR_Fuzzy_Classify_Mod"*, the output parameters have been specified and pruned into normal, mild, moderate, moderately severe, and beyond moderate. By doing so, we were able to effectively reduce the number of rules needed, while effectively segmenting the symptoms into finer categories.

#	**Fuzzy Rule of *NPDR_Fuzzy_Classify_Mod***	**Weight**
1	Division==normal & IRH==absent & MA==absent & HE~=present & IRMA==absent & CWS~=present => Classification=normal	1
2	Division==mild & IRH==present in 1 quad & MA==absent & HE~=present & IRMA==absent & CWS~=present => Classification=mild	1
3	Division==mild & IRH==absent & MA==present in 1 quad & HE~=present & IRMA==absent & CWS~=present => Classification=mild	1
4	Division==mild & IRH==absent & MA==absent & HE==present & IRMA==absent & CWS~=present => Classification=normal	1
5	Division==mild & IRH==present in 1 quad & MA==present in 1 quad & HE~=present & IRMA==absent & CWS~=present => Classification=mild	0.5
6	Division==mild & IRH==present in 1 quad & MA==absent & HE==present & IRMA==absent & CWS~=present => Classification=mild	0.5
7	Division==mild & IRH==absent & MA==present in 1 quad & HE==present & IRMA==absent & CWS~=present => Classification=mild	1
8	Division==mild & IRH==present in 1 quad & MA==present in 1 quad & HE==present & IRMA==absent & CWS~=present => Classification=mild	0.5
9	Division==moderate & MA==present in 2 quads => Classification=moderately severe	0.5
10	Division==moderate & IRMA==present in 2 quads as mild => Classification=moderately severe	1
11	Division==moderate & MA==absent & HE~=present & IRMA==present in 1 quad as mild => Classification=moderate	1
12	Division==moderate & MA==absent & HE==present & IRMA==absent => Classification=moderate	1
13	Division==moderate & MA==present in 1 quad & HE~=present & IRMA==absent => Classification=moderate	1

14	Division==moderate & MA==absent & HE==present & IRMA==present in 1 quad as mild => Classification=moderate	0.5
15	Division==moderate & MA==present in 1 quad & HE~=present & IRMA==present in 1 quad as mild => Classification=moderate	0.5
16	Division==moderate & MA==present in 1 quad & HE==present & IRMA==absent => Classification=moderate	0.5
17	Division==moderate & MA==present in 1 quad & HE==present & IRMA==present in 1 quad as mild => Classification=moderate	0.25
18	Division==moderate & IRH==present in 2 quads => Classification=moderately severe	1
19	Division==moderate & IRH==present in 1 quad & MA==absent & HE~=present & IRMA==absent & CWS~=present => Classification=mild	0.5
20	Division==normal & IRH==absent & MA==present in 1 quad & HE~=present & IRMA==absent & CWS~=present => Classification=mild	0.25
21	Division==normal & IRH==absent & MA==absent & HE==present & IRMA==absent & CWS~=present => Classification=mild	0.25
22	IRH==present in 3 quads | MA==present in 3 quads | IRMA==present in 3 quads as mild => Classification=beyond moderate	1
23	IRH==present in 4 quads | MA==present in 4 quads | IRMA==present in 4 quads as mild or 1 quad as prominent => Classification=beyond moderate	1

Table 3.2: Fuzzy Rule of Logic *"DR_Fuzzy_Classify_Mod"*

Rules 1 and 4 allow the system to diagnose the current state as normal when there are no symptoms present. This may sound unorthodox since all the normal states must be diagnosed by "*DR_Fuzzy_Division*" beforehand. However, this is to prevent any potential false trues occurring when new rules are added to the "*DR_Fuzzy_Division*".

For symptoms that were directly listed in the ICDR scale, a fuzzy rule with a weight of 1 has been designated. This includes rules 2, 3, and 7 for mild NPDR, rules 4, 11-13 for moderate NPDR, 10 for moderately severe, and rules 22 and 23 for beyond moderate NPDR.

The rest of the rules are for the symptoms that could be categorized into more than two categories at the same time. In such cases, multiple rules with the equivalent condition have been

created with lower weights. This not only allows a natural transition between the categories but also enables finer, consistent diagnosis for the images with niche symptoms which are hard to diagnose with the currently available severity scales.

Similarly, in "*NPDR_Fuzzy_Classify_Severe"*, the output has been altered from standard normal, mild, moderate, and severe into below severe, moderately severe, severe, and extremely severe. Since the MA and IRH are virtually considered as a single parameter for diagnosing severe NPDR in the ICDR severity scale, the status of MA and IRH has been combined through a simple OR statement within the simple logic of "*DR_MaH"*. By doing this, we can effectively half the number of rules needed for *"NPDR_Fuzzy_Classify_Severe"*.

#	**Fuzzy Rule of *NPDR_Fuzzy_Classify_Severe***	**Weight**
1	IRMA==present in 2 quads as mild (1 VB) => Classification=moderately severe	1
2	MaH==present in 4 quads | IRMA==present in 4 quads as mild or 1 quad as prominent (multiple VB) => Classification=severe	1
3	IRMA==present in 4 quads as mild or 1 quad as prominent (multiple VB) => Classification=extremely severe	1
4	MaH==present in 4 quads & IRMA==present in 4 quads as mild or 1 quad as prominent (multiple VB) => Classification=extremely severe	1
5	IRMA==present in 4 quads as mild or 1 quad as prominent (multiple VB) => Classification=extremely severe	1
6	IRMA==present in 4 quads as mild or 1 quad as prominent (multiple VB) => Classification=extremely severe	1
7	Division==moderate & MaH==present in 3 quads => Classification=moderately severe	0.5
8	Division==severe & MaH==present in 3 quads => Classification=severe	1
9	Division==less than moderate & MaH==present in 1 quad => Classification=below severe	1
10	Division==moderate & MaH==present in 1 quad & IRMA==absent => Classification=below severe	1
11	Division==less than moderate & MaH==absent & IRMA==absent => Classification=below severe	1

12	Division==less than moderate & MaH==absent & IRMA==absent => Classification=below severe	1

Table 3.3: Fuzzy Rule of Logic *"DR_Fuzzy_Classify_Severe"*

Finally, "*DR_Fuzzy_MakeDiag"* makes a final diagnosis based on the output of "*DR_Check_Prolif"*, "*NPDR_Fuzzy_ Classify_Mod"*, and "*NPDR_Fuzzy_Classify_Severe"*. While doing so, "*DR_Fuzzy_MakeDiag"* also checks the validity of the diagnosis – although it is almost impossible to have an invalid diagnosis since both "*NPDR_Fuzzy_Classify_Mod"* and "*NPDR_Fuzzy_Classify_Severe"* use similar inputs, it is a good practice to have a model to cross-validate the diagnosis. This becomes more evident when medical experts expand the model by adding more fuzzy inputs and their linguistic values, increasing the chance of misdiagnosis. Table 3.4 shows the fuzzy rules for "*DR_Fuzzy_MakeDiag"*.

#	**Fuzzy Rule of *DR_Fuzzy_MakeDiag***	**Weight**
1	Class_Mod==normal & Class_Severe==below severe => Diagnosis=normal, Validity=valid	1
2	Is_Prolif==present => Diagnosis=proliferative, Validity=valid	1
3	Class_Mod==moderately severe & Class_Severe==moderately severe & Is_Prolif~=present => Diagnosis=moderately severe, Validity=valid	1
4	Class_Mod==beyond moderate & Class_Severe==severe & Is_Prolif~=present => Diagnosis=severe, Validity=valid	1
5	Class_Mod==beyond moderate & Class_Severe==extremely severe & Is_Prolif~=present => Diagnosis=extremely severe, Validity=valid	1
6	Class_Mod==moderate & Class_Severe==below severe & Is_Prolif~=present => Diagnosis=moderate, Validity=valid	1
7	Class_Mod==beyond moderate & Class_Severe==below severe => Validity~=valid	0.25
8	Class_Mod==normal & Class_Severe~=below severe => Validity~=valid	1
9	Class_Mod==mild & Class_Severe~=below severe => Validity~=valid	0.5
10	Class_Mod==moderate & Class_Severe~=below severe => Validity~=valid	0.25

11	Class_Mod~=beyond moderate & Class_Severe==moderately severe => Validity~=valid	0.25
12	Class_Mod~=beyond moderate & Class_Severe==severe => Validity~=valid	0.5
13	Class_Mod~=beyond moderate & Class_Severe==extremely severe => Validity~=valid	1
14	Class_Mod==normal & Is_Prolif==present => Diagnosis=normal, Validity~=valid	1
15	Class_Mod==mild & Class_Severe==below severe & Is_Prolif~=present => Diagnosis=mild, Validity=valid	1
16	Class_Mod==mild & Class_Severe==below severe & Is_Prolif~=present => Diagnosis=mild, Validity=valid	1

Table 3.4: Fuzzy Rule of Logic *"DR_Fuzzy_MakeDiag"*

Rules 1-6, 15, and 16 cover the case where the output of "*NPDR_Fuzzy_ Classify_Mod*", and "*NPDR_Fuzzy_Classify_Severe*" points towards the same direction in diagnosis. On the other hand, rules 7-14 cover the case where the two outputs show a discrepancy, signifying the diagnosis can't be trusted. The bigger the discrepancy is, the higher the assigned weight is for those rules.

3.1.1 Implementation and Membership Functions

The entire system has been implemented through MATLAB. The implementation starts with reading the subsystems using "readfis()" Each subsystem was implemented based on the ruleset provided in chapter 3.1, and its actual implementation can be viewed in Appendix B.

- *"DR_Fuzzy_Division"*
    - HE and CWS are crisp input that go into the system. Henceforth, it has a shape of step function.

- IRH and MA have a shape of trapezoid, as the categorization of the progress of IRH and MA are both linear.
- IRMA, on the other hand, uses Gaussian membership function, for the diagnose of each category shows distinctive features.
- The membership functions of the output will form a shape of trapezoid, as the progress of disease itself is linear. However, notice that the early stages of disease have their output much denser. That's because the accurate, precise classification of the early stages of disease matters a lot more in comparison to later stages in the disease, where the treatment becomes vital instead.

- *"NPDR_Fuzzy_Classify_Mod" & "NPDR_Fuzzy_Classify_Severe"*
  - An output of *"DR_Fuzzy_Division"* is directly parsed as an input, together with other inputs used in the *"DR_Fuzzy_Division"* for reduction in number of rules, as stated in the end of chapter 2.6.
  - The outputs are in the form of Gaussian function instead of linear functions such as trapezoid or triangular, since the based rubric used to create this system employed crisp categorization, and changing it abruptedly into linear output resulted in increased error. This has been mitigated through using narrow Gaussian membership functions with increased numbers of categories.
- *"DR_Fuzzy_MakeDiag"*
  - A new crisp input, Is_Prolif, comes into play to check if the disease is proliferative or not. This is because PDR has distinct symptoms in comparison to NPDR and removing them in the staging process vastly reduces the number of rules and unnecessary complexity.

- Other than diagnosis, additional linear S-shaped function “Validity” is outputted to check if the output generated makes sense.

Afterwards, all necessary connections for each fuzzy subsystem are made. Using the specified inputs and outputs, a FIS tree is created using “fistree()”.

```
fis_check_prolif = readfis("DR_Check_Prolif");
fis_division = readfis("DR_Fuzzy_Division");
fis_mah = readfis("DR_MaH");
fis_class_moderate = readfis("NPDR_Fuzzy_Classify_Mod");
fis_class_severe = readfis("NPDR_Fuzzy_Classify_Severe");
fis_makeDiag = readfis("DR_Fuzzy_MakeDiag");

%mamfis("Name", "DR_Check_MakeDiag", "NumInputs", 2, "NumOutputs", 1);

con1 = ["DR_Check_Prolif/Is_Prolif" "DR_Fuzzy_MakeDiag/Is_Prolif"];
con2 = ["DR_Fuzzy_Division/Division" "NPDR_Fuzzy_Classify_Mod/Division"];
con3 = ["DR_Fuzzy_Division/Division" "NPDR_Fuzzy_Classify_Severe/Division"];
con4 = ["NPDR_Fuzzy_Classify_Mod/Classification" "DR_Fuzzy_MakeDiag/Class_Mod"];
con5 = ["NPDR_Fuzzy_Classify_Severe/Classification" "DR_Fuzzy_MakeDiag/Class_Severe"];

con_input_01 = ["DR_Fuzzy_Division/IRH" "NPDR_Fuzzy_Classify_Mod/IRH"];
con_input_02 = ["DR_Fuzzy_Division/MA" "NPDR_Fuzzy_Classify_Mod/MA"];
con_input_03 = ["DR_Fuzzy_Division/IRH" "DR_MaH/IRH"];
con_input_04 = ["DR_Fuzzy_Division/MA" "DR_MaH/MA"];
con_input_05 = ["DR_Fuzzy_Division/HE" "NPDR_Fuzzy_Classify_Mod/HE"];
con_input_06 = ["DR_Fuzzy_Division/IRMA" "NPDR_Fuzzy_Classify_Mod/IRMA"];
con_input_07 = ["DR_Fuzzy_Division/IRMA" "NPDR_Fuzzy_Classify_Severe/IRMA"];
con_input_08 = ["DR_Fuzzy_Division/CWS" "NPDR_Fuzzy_Classify_Mod/CWS"];
con_input_09 = ["DR_MaH/MaH" "NPDR_Fuzzy_Classify_Severe/MaH"];


%fuzzTree = fistree([fis_check_prolif fis_check_severity], [])
fuzzTree = fistree([fis_division fis_mah ...
    fis_class_moderate fis_class_severe fis_check_prolif fis_makeDiag], ...
    [con1;con2;con3;con4;con5; ...
    con_input_01;con_input_02;con_input_03;con_input_04;con_input_05; ...
    con_input_06;con_input_07;con_input_08;con_input_09]);
```

Figure 3.2: FIS implementation in MATLAB

Now that the FIS tree is created, an input from each dataset is parsed through the system, and its output is saved for analysis.

```
%FGADR
data_table_FGADR = readtable("DR_Seg_Grading_Label - Annotated.csv");
data_table_FGADR.Hemorrhage = num2cell(data_table_FGADR.Hemorrhage);
data_table_FGADR.Microaneurysms = num2cell(data_table_FGADR.Microaneurysms);
data_table_FGADR.HardExudates = num2cell(data_table_FGADR.HardExudates);
data_table_FGADR.IRMA = num2cell(data_table_FGADR.IRMA);
data_table_FGADR.SoftExudate = num2cell(data_table_FGADR.SoftExudate);
data_table_FGADR.Neovascularization = num2cell(data_table_FGADR.Neovascularization);
data_table_FGADR.GroundTruth = num2cell(data_table_FGADR.GroundTruth);
data_table_FGADR.ModifiedGroundTruth = num2cell(data_table_FGADR.ModifiedGroundTruth);
data_table_FGADR.Output = num2cell(data_table_FGADR.Output);
data_table_FGADR.Validity = num2cell(data_table_FGADR.Validity);
data_FGADR = table2array(data_table_FGADR);

% plot fuzzy inference system diagram
% plotfis(fuzzTree, Legend="on")

% fuzzTree.Inputs
% IRH: Intraretinal Hemorrhage
% MA: Microaneurysm
% HE: Hard Exudate
% IRMA:
% CWS: Cotton Wool Spot
% NV or VH or PRH or PRP Scar:

for c = 1:60
    output_test = evalfis(fuzzTree, cell2mat(data_FGADR(c, 2:7)));
    data_FGADR(c, 10) = num2cell(output_test(1,1));
    data_FGADR(c, 11) = num2cell(output_test(1,2));
end
```

Figure 3.3: Code Snippet for Testing FGADR Dataset

```
%APTOS
data_table_APTOS = readtable("DR_Seg_Grading_Label_APTOS - Annotated.csv");
data_table_APTOS.Hemorrhage = num2cell(data_table_APTOS.Hemorrhage);
data_table_APTOS.Microaneurysms = num2cell(data_table_APTOS.Microaneurysms);
data_table_APTOS.HardExudates = num2cell(data_table_APTOS.HardExudates);
data_table_APTOS.IRMA = num2cell(data_table_APTOS.IRMA);
data_table_APTOS.SoftExudate = num2cell(data_table_APTOS.SoftExudate);
data_table_APTOS.Neovascularization = num2cell(data_table_APTOS.Neovascularization);
data_table_APTOS.GroundTruth = num2cell(data_table_APTOS.GroundTruth);
data_table_APTOS.ModifiedGroundTruth = num2cell(data_table_APTOS.ModifiedGroundTruth);
data_table_APTOS.Output = num2cell(data_table_APTOS.Output);
data_table_APTOS.Validity = num2cell(data_table_APTOS.Validity);
data_APTOS = table2array(data_table_APTOS);

% plot fuzzy inference system diagram
% plotfis(fuzzTree, Legend="on")

% fuzzTree.Inputs
% IRH: Intraretinal Hemorrhage
% MA: Microaneurysm
% HE: Hard Exudate
% IRMA:
% CWS: Cotton Wool Spot
% NV or VH or PRH or PRP Scar:

for c = 1:50
    output_test = evalfis(fuzzTree, cell2mat(data_APTOS(c, 2:7)));
    data_APTOS(c, 10) = num2cell(output_test(1,1));
    data_APTOS(c, 11) = num2cell(output_test(1,2));
end


% GUI
run("DR_GUI.mlapp");
```

Figure 3.4: Code Snippet for Testing APTOS Dataset

The detailed functions of membership functions within each .fis files and its justification can be found in Appendix B. Now that the FIS tree is created, an input from each dataset is parsed through the system, and its output is saved for analysis.

3.2 Parameters

For diagnosis, 6 different parameters were set as an input, listed as follows:

- Microaneurysm (MA)
- Intraretinal Hemorrhage (IRH)
- Hard Exudates (HE)
- Intraretinal Microvascular Abnormality (IRMA)
- Cotton Wool Spots (CWS, also known as Soft Exudates)
- Neovascularization (NV) or Hemorrhage (VH or PRH)

A detailed explanation of each parameter is demonstrated in Appendix A.

The proposed Fuzzy Inference System (FIS) tree model consists of a few Fuzzy Logic sub-systems, with generic inputs that can be acquired from the FP. Like the Patient-Reported Outcome Measure (PROM) questionnaires that are generally used in hospitals for triage, the inputs are numerically done, from the lowest being mild to the highest being severe. The FIS inputs are defined in Appendix A.

The rubrics for interpreting FIS outputs are the following:

- Around 0: Normal.
- Around 0.5: Mild.
- Around 1: Moderate.
- Around 1.5: Moderately Severe. Starting from here, a minimum of two or four months' follow-up with retinal referral is advised.
- Around 2: Severe.

- Around 2.5: Extremely Severe.
- 3 or higher: Proliferative. A retinal referral in one week is advised.

3.3 Dataset Used

The system is tested through some of the images from the Fine-Grained Annotated Diabetic Retinopathy (FGADR) dataset [27]. This dataset was suitable for validating our system, as each image showed good consistency while having its grading labels annotated by three ophthalmologists. To increase the credibility, each annotation was doubly checked after processing the images through fuzzy-based retinal image contrast enhancement, and any missing or mislabeled lesions were redefined properly [28]. To simulate the human ophthalmologists' diagnoses, the dataset was modified to have additional data required for diagnosis.

To provide analysis with a more generally available dataset, an Asia Pacific Tele-Ophthalmology Society (APTOS) dataset was utilized [29]. As the APTOS dataset successfully provides a diverse and comprehensive collection of well-annotated retinal images, it's been widely used in the research community, promoting the comparability between different studies.

For both datasets, similar number of images were chosen for testing. This is to prevent any bias in accuracy that may occur in testing, as the system may outperform in certain datasets over others, based on how well they follow the ICDR severity scale.

Also, images with the key features that showed good correlation but weren't present in the ophthalmologists' DR scales were not included in the system. A good example is images with previous laser marks (PLM) caused by pan-retinal photocoagulation (PRP). As almost every image with PLM was diagnosed proliferative additional rules might have been set to classify

those edge cases. However, those are excluded for a few different reasons. Firstly, the PLM signifies that PRP has already been performed on the patient's eye, which is followed by the diagnosis. Therefore, using this system to diagnose an eye that already has been diagnosed by the experts and treatment is practiced wouldn't make logical sense. Secondly, PLM was not discussed as part of the observable findings in the ICDR severity scale for DR. Finally, different datasets took different approaches in diagnosing such images – some asserted that such images must diagnosed as PDR since it was proliferative before the procedure, whereas others asserted that such images must be diagnosed as NPDR as it is no longer proliferative after the procedure. It is possible to edit the proposed fuzzy rules based on the diagnosis made by each dataset but to prevent confusion, such images were omitted.

CHAPTER 4

ACCURACY ANALYSIS OF INDUCTED FUZZY CLASSIFIERS

In this chapter, we conduct a numerical evaluation of the proposed system's diagnostic capabilities. Subsequently, we transition to an analysis of the edge cases, where diagnostic inaccuracies manifest. Here, we scrutinize specific instances where the diagnosis of the proposed system deviates from the ground truth provided by the expert and attempt to dissect the origins of those discrepancies in classification.

4.1 Numerical Evaluation of the Proposed Fuzzy System

4.1.1 FGADR

Out of 58 images randomly chosen within the dataset for the test, 45 images were diagnosed correctly within the error range of 0.5 (77.59% accuracy), and 52 images were diagnosed correctly within the error range of 1 (89.66% accuracy). Detailed data has been provided in Appendix C.

Undoubtedly, the level of accuracy observed in this system does not surpass that of other contemporary state-of-the-art DR diagnosis systems. It is noteworthy, however, that the error can be caused by the discrepancy with the diagnostic scale the ophthalmologists have chosen during the annotation, and its error ranges are further increased due to the fuzzification. In other words,

some of the “errors” found are technically not an error, but an adjustment to make a fine judgment about the state of the eye.

4.1.2 APTOS

Out of 50 images randomly chosen, 49 images were diagnosed correctly within the error range of both 0.5 and 1 (98% accuracy). Amongst the 49 images, 1 of the images accurately reported that the result has low validity. Most of this discrepancy originated from distinguishing between mild and moderate NDPR, and its details will be explained in the next section. Detailed data has been provided in Appendix D.

Dataset	**Total Images**	**Accuracy (Error Range 0.5)**	**Accuracy (Error Range 1)**
FGADR	58	45/58 (77.59%)	52/58 (89.66%)
APTOS	50	49/50 (98.00%)	49/50 (98.00%)
FGADR + APTOS	108	94/108 (87.04%)	101/108 (93.52%)

Table 4.1: Combined Accuracy Analysis for Datasets FGADR and APTOS

Overall, an accuracy of 87.04% has been achieved, with 93.52% accuracy within the error range of 1 as shown in Table 4.1. It is noteworthy that the inaccuracy from the testing simply signifies that its diagnosis differed from the opinion of experts. In other words, the proposed system could shed light on the grey areas where the diagnosis may differ from doctor to doctor. A detailed explanation with examples is covered in the next section.

4.2 Analysis of the Edge Cases

The main reason behind the diagnostic inaccuracy comes from the discrepancy in classification standards within the dataset. Both images in Figure 4.1 display signs of exudates (marked as a red circle). Though minuscule, such findings should result in a classification of moderate or higher based on the ICDR Severity Scale (Table 1.1). However, unlike the standard scale, the APTOS dataset classified both images as 1.

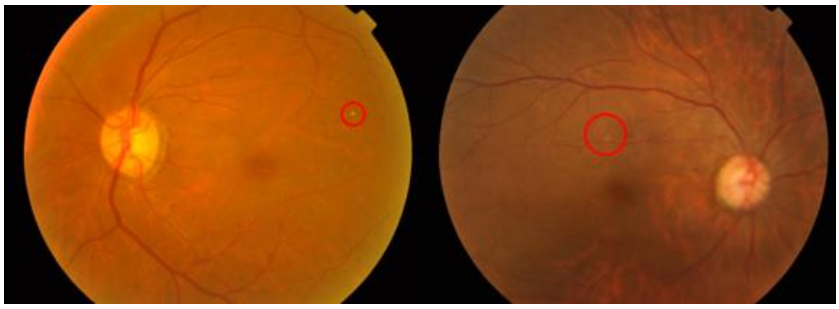


Figure 4.1: Sample Images of Misdiagnosed APTOS Images

This does not signify either the ground truth of the APTOS dataset or the ICDR Severity Scale is wrong – it instead displays that the experts who diagnosed the ground truth for the APTOS dataset put more variables into consideration when diagnosing the images (i.e., the severity of CWS/HE). Analysis was conducted to figure out the similarities between the outliers that the system had trouble diagnosing, as shown in Table 4.2.

Input that was classified inaccurately	IRH	MA	HE	IRMA	CWS	NV/VH/ PRH
#1	1	0	0	0	0	0
#2	1	0	1	0	0	0
#3	3	1	1	0	0	0
#4	1	0	1	0	0	0
#5	1	1	0	0	0	0
#6	3	1	1	0	0	0
#7	1	2	1	1	0	0
#8	2	1	0	0	0	0
#9	1	1	0	0	0	0

Table 4.2: Input Parameters of APTOS Images Which had a Mild Error (< 0.5)

A few notable similarities were speculated through the analysis of the data that were misdiagnosed.

- No signs of severe symptoms (IRMA or NV/VH/PRH) were found.
- Early stages of IRH/MA were detected.
- Exudates were present, but very local and not widespread through the retina.

Through minor adjustments in fuzzy rules regarding IRH/MA and by subdividing HE/CWS membership functions more specifically, better results customized to the classification of APTOS classification standards may be yielded.

Similarly, the FGADR dataset may have different scales utilized for diagnosing the severity of the DR in the image.

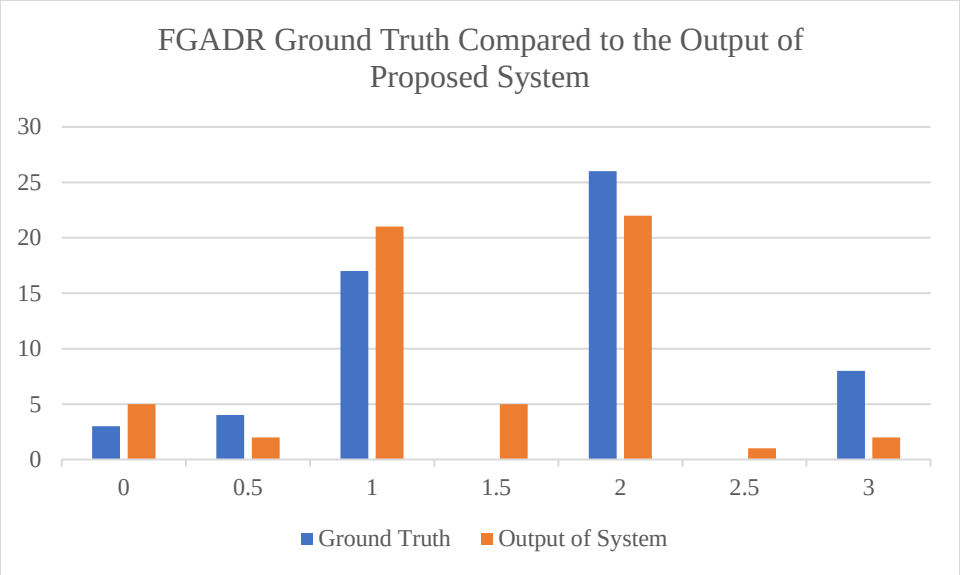


Figure 4.2: FGADR Ground Truth Compared to the Output of Proposed System

As depicted in Figure 4.2, it is quite evident that the diagnosis made in FGADR tended to mark the condition of the eyes as more serious.

CHAPTER 5

DISCUSSION

In this chapter, we delve into a comprehensive discussion that navigates the improvements the proposed diagnostic system achieved. We then pivot into discussing objective, transparent analysis of the challenges inherent in its design and application. With it, a discussion of potential countermeasures will be made, laying down the groundwork for strategies to mitigate limitations and addressing critical perspectives.

5.1 Comparison with the State of the Art

Based on the leaderboard available for APTOS 2019 in Kaggle, the strategy and accuracy acquired are as follows:

Ranking	Strategy	Accuracy
1	Neural Network using Generalized Mean Pooling	Public: 0.850 Private: 0.936
2	Neural Network with Preprocessing (Crop + Resize) and Augmentation (Blur, Flip, RandomBrightnessContrast, ShiftScaleRotate, ElasticTransform, Transpose, GridDistortion, HueSaturationValue, CLAHE, CoarseDropout)	Public: 0.841 Private: 0.936
4	Neural Network with Preprocessing (Crop + Image Type Categorization) and Augmentation (Dihedral, RandomCrop, Rotation, Contrast, Brightness, Cutout, PerspectiveTransform, CLAHE)	Public: 0.840 Private: 0.933

Table 5.1: Strategy and Accuracy of Leaderboard Competitors in APTOS 2019 Blindness Detection Competition [29]

It is important to note that those accuracies cannot be the direct comparison of the proposed system, as the APTOS 2019 Blindness Detection Competition conducted by Kaggle only provides image data to achieve the results. As any methods proposed in this competition rely solely on image recognition and classification, their resultant accuracy is naturally lower than that of the proposed system, which relies on experts' perspectives on the image as input to begin with. Nonetheless, it is noteworthy that the overall accuracy that our system obtained (87.04% in FGADR + APTOS, 98.00% for APTOS only) is competitive when compared to a few of the best methods proposed out there.

When the scope is broadened for more direct comparison, the state-of-the-art DR detection methods show much more impressive accuracy overall. A systematic review and meta-analysis conducted in 2020 demonstrated a pooled sensitivity of 91.9% (95% CI: 89.6% to 94.3%) and specificity of 91.3% (95% CI: 89.0% to 93.5%) over 24 studies using the Rutter and Gatsonis hierarchical summary receiver operating characteristics (HSROC) model [30]. This shows that our proposed model has room for further improvements in improving the overall accuracy of diagnosis. However, there were certain advantages that no other models presented in the study could replicate.

5.2 Improvements

The proposed model offers a set of advantages. First and foremost, the usage of fuzzy logic enables gradient changes in the severity of the diagnosis. Usage of the crisp diagnosis scale had inconsistencies within the same categories – despite being classified as the same stages of the same disease, some sample data showed relative differences in the severity of the disease. The usage of the FIS tree not only allows experts to have crisp categorizations, but through a

simple change in the defuzzification methods, experts can also acquire detailed severity of the disease within the same category. Figure 5.1, which displays how the diagnosis is made when all the processed inputs are entered at the final fuzzy subsystem, exemplifies how a diagnosis can be "smooth", in comparison to other diagnostic tables usually used by ophthalmologists which can be found in Table 1.1 [3].

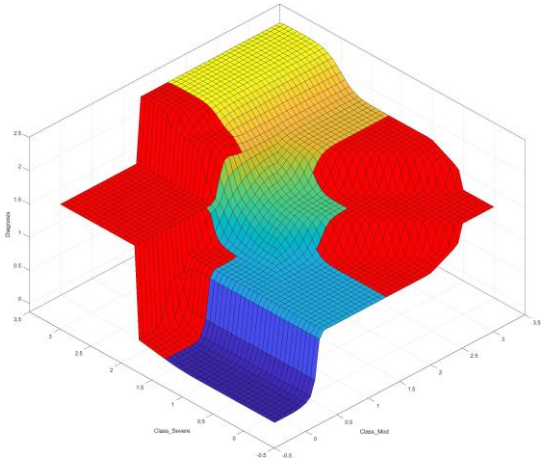


Figure 5.1: Diagram of "*DR_Fuzzy_MakeDiag*" When "*Is_Prolif*" Is Not Present. The Area Marked Red Is Where the Validity Is Low.

In Figure 5.2, both images are classified as mild NPDR in the APTOS Dataset – which is equivalent to 1 in the proposed system. However, the proposed system employs a finer granularity in its classification, assigning the final classification of the left image as 1.057, and the right image as 0.954, as the left image showed more widespread dot hemorrhages overall. The enhanced sensitivity of the proposed system allows for a nuanced evaluation, capturing minor details that contribute to a more refined and precise classification of DR severity.

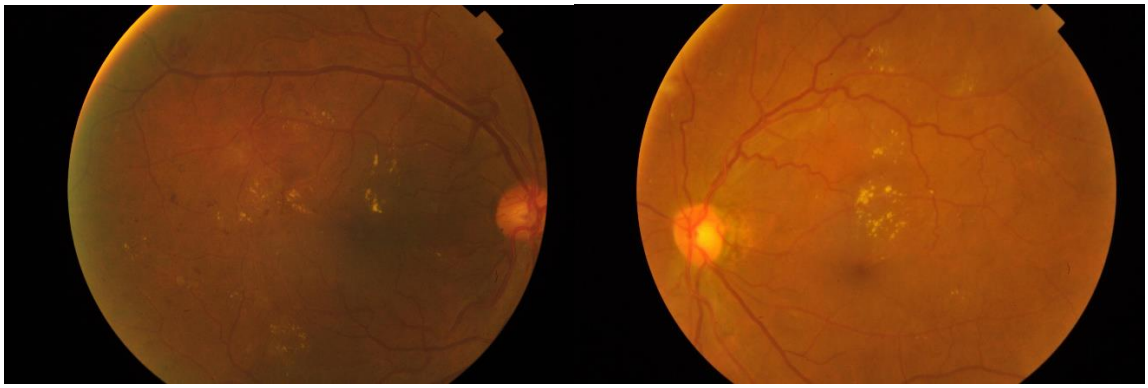


Figure 5.2: Images That Are Diagnosed *Exactly* Same in APTOS, but Differently in the Proposed System

The proposed methodology not only ensures a meticulous evaluation of retinal images but also facilitates the detection of potential diagnostic errors that might be overlooked by human practitioners. For instance, both images in Figure 5.3 are classified as moderate NPDR in the APTOS dataset. The rule-based analysis of the system, however, diagnoses the left image as more severe, with a point of 2.005, in comparison to the right image, with a point of 0.954. Although these variances might initially appear as significant inaccuracies, closer scrutiny reveals that the left image exhibits more severe symptoms, such as hard/soft exudates and widespread hemorrhages in comparison to the right image.

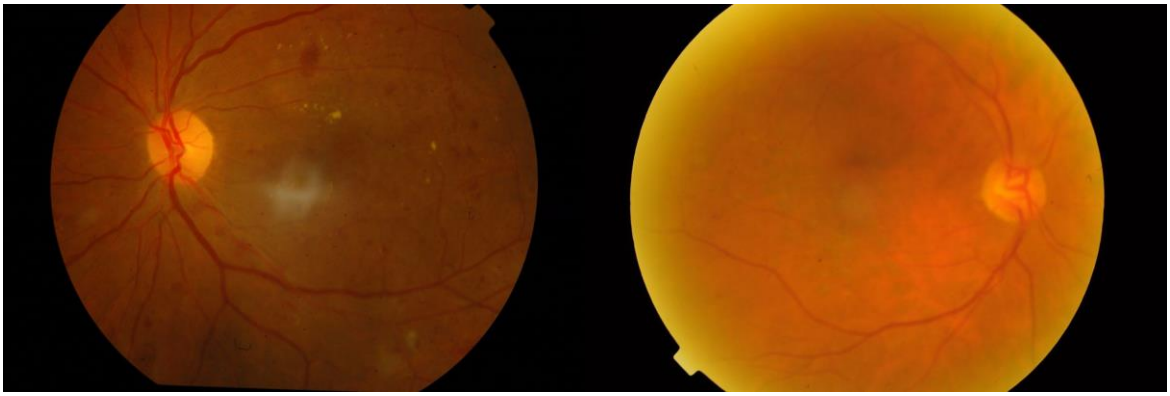


Figure 5.3: Images That Are Diagnosed Same in APTOS, but Found to be a Potential Diagnostic Error in the Proposed System

Another standout feature of the proposed model is its ability to incorporate an expert's diagnosis into the model. The customizable nature of the FIS ensures that the FIS remains a relevant and effective diagnostic tool in the dynamic landscape of healthcare while promoting transparency in the results. Healthcare providers and patients can better understand how a diagnosis is reached, as the model's decision-making process is based on clearly defined rules, enhancing trust in the diagnostic outcomes. Although many studies used CAD to improve the accuracy of diagnosis, only a few kept an eye on expert recognizability, as the nature of modern neural networks allowed diagnosis based on numerous correlations that cannot be distinguished normally in human eyes.

Additionally, a noteworthy feature of the proposed FIS tree is its usability – thanks to its easily customizable yet transparent nature while providing real-time support. As each logic within the model employs a fuzzy rule-based approach (FRBS), ophthalmologists can easily modify and refine the rules through programmers to adapt to ever-evolving medical knowledge

and criteria, while not hindering both the experts and the patients in understanding how a diagnosis is reached after an update.

This could be even more useful if hospitals can utilize the patients' private data that can't be publicly shared but obtained through medical procedures to add further correlation, leading to improvement in diagnostic accuracy overall. The prime example is the analysis conducted in chapter 4.2 – despite not being aware of the set of rules used for diagnosis in the APTOS dataset specifically, a proper hypothesis was deduced based on analysis conducted on the outliers.

On top of that, although it may not be as fast as, or as accurate as other state-of-the-art AI-driven CAD systems, the proposed model is still capable of providing real-time decision support with reasonable accuracy, addressing the pressing need for timely interventions and improved patient care in the actual field. Such features combined ensure that the model remains relevant in the industry longer. Moreover, the proposed system can be valuable in the fields of academia as well as industry due to the aforementioned reasons as well.

5.3 Limitations and Countermeasures

The biggest limitation of the proposed system currently is, paradoxically, the need for expert availability. Both proper operation and maintenance require experts to attend aside. For instance, there is no doubt that the FIS tree-based models are very expandable. However, as the model is relevant to medical industries, far more rigorous validation and testing must be practiced beforehand, as the standard requirement of accuracy in those fields is crucial. While trying to meet the accuracy and tuning the weights of the fuzzy logic, one might face an

overfitting issue, and some might end up adding too many fuzzy rules, eventually hindering not only the accuracy but also the efficiency of the system.

To minimize such limitations, and for continuous improvement in the system, fostering a collaborative feedback loop between ophthalmologists and AI scientists will be crucial. Medical practitioners can provide constant feedback about the system, while the AI scientists work on updating the model based on real-world clinical outcomes.

Input having a range of severity can also induce niche problems. Although having a gradient input allows output to be more intricate and precise rather than a giant category, each doctor may have different standards in deciding the severity of the symptoms and lesions in numerical values. This may even lead to some doctors modifying the input value slightly so that they can get the desired outcome out of the system, which is easier to do in comparison to other CAD systems due to the transparency, modifiability, and dependency it offers to the doctors. However, such problems can, and must be resolved through proper ethics training of the experts.

Another drawback of the model is that it shares many constraints that FP has. Some of the information, such as checking whether a malformation of blood vessel seen in the FP is a prominent IRMA or NV, or inspecting if the red dots that are shown in the image are MA, IRH, or device noise.

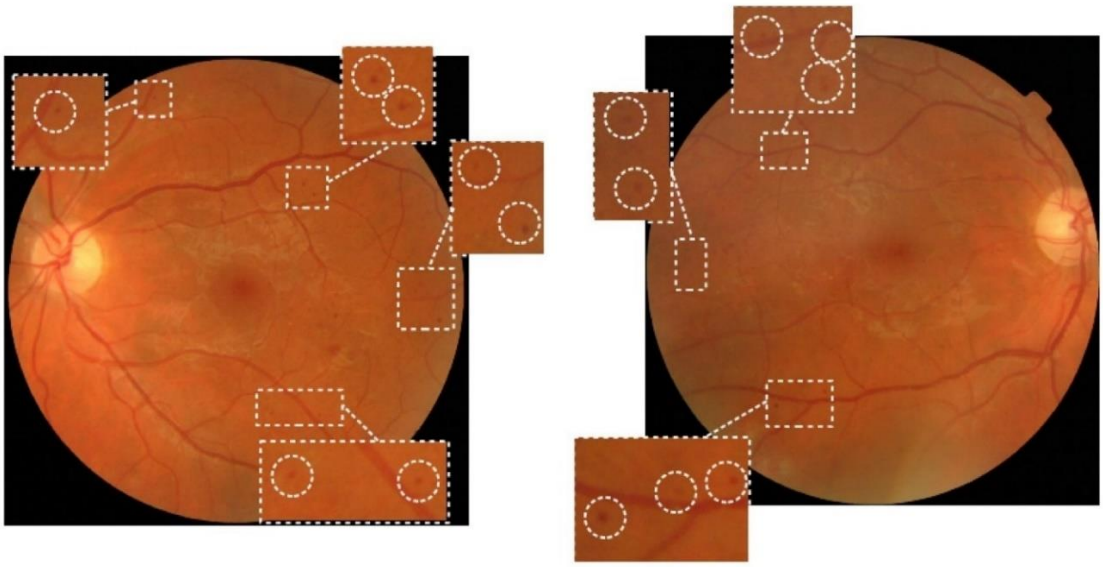


Figure 5.4: Images of Several Microaneurysms, One of the Observable Findings Used for DR Diagnosis [31]

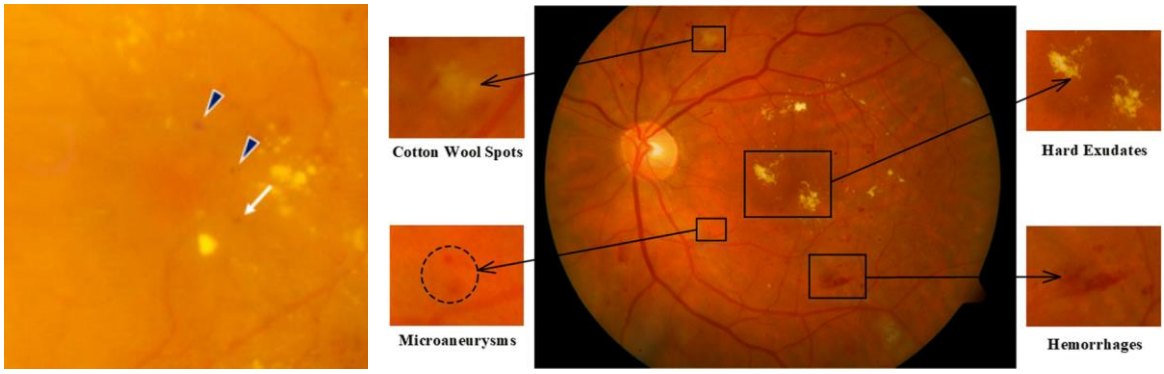


Figure 5.5: Two Figures Each Displaying Similarities of Distinct Symptoms. Microaneurysms (Black Arrowheads) and Dot Hemorrhages (White Arrows) in FP (Left) [6], Sample Image Which Contains All Intermediate-Level DR Features (Right) [32]

In the case of Figure 5.4, it is quite evident that some of the microaneurysms are barely visible to human eyes. Some may argue that doctors are trained to distinguish such findings, but

it is unequivocal that such issues may lead to misdiagnosis. In the left image of Figure 5.5, microaneurysms and dot hemorrhages are detected by fluorescein angiography and optical coherence tomography angiography (OCTA), both of which show a more precise discrepancy in blood vessel assessment at the cost of its price and complexity. However, in FP, two different features are both shown as nearly identical red dots. Similarly, in the right image of Figure 5.5, microaneurysms and hemorrhages, or cotton wool spots and hard exudation, both showed striking similarities.

Unlike some of the more advanced examination methods such as fluorescein angiography, fundus photography has some limitations in acquiring enhanced visualization of blood flow and vascular abnormalities at the cost of its convenience. To further improve the accuracy of diagnosis and reduce human error, preprocessing of the fundus photography is highly advised.

One can also use other examination methods such as fluorescein angiography or optical coherence tomography (OCT) when a more precise assessment is required. For instance, IRMA blood vessels are patent, whereas neovascular vessels are occluded. In these cases, the usage of angiography will ensure only IRMA blood vessels exhibit fluorescence [33].

On top of that, one study shows that microaneurysms and dot hemorrhages are clinically indistinguishable, so they are referred to as hemorrhages and/or microaneurysms (H/Ma) as well [34]. Correlation analysis in the proposed model, depicted in Figure 5.6, supports this fact as IRH and MA show similar outcomes in making classification – IRH is considered a slightly more severe symptom, but misclassification of it will not result in drastic diagnostic error. This not only proves that some features are not as important as other features in the staging of DR but also opens new possible studies – as FIS is relatively easy to modify, it will allow clinicians to

edit or merge the inputs for hemorrhages and microaneurysms as one using simple or fuzzy logic, while diversifying types of hemorrhages, utilizing them as a new parameter for improving diagnosis. This validates how the proposed system mimics how human diagnosticians would describe and make their decisions.

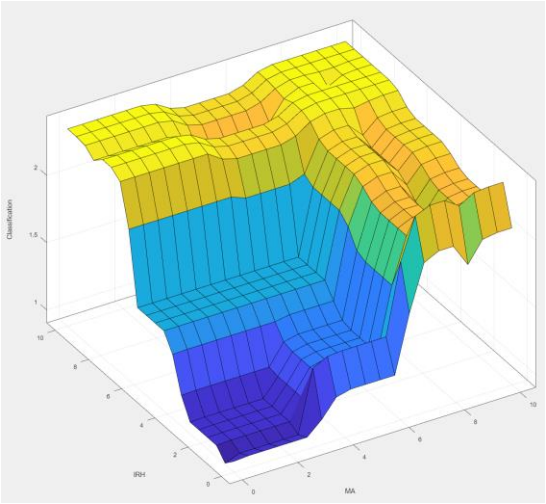


Figure 5.6: Diagram of the Correlation Analysis of IRH and MA in Output of "*NPDR_Fuzzy_Classify_Mod*"

Finally, since the proposed system is rule-based, the system failed to diagnose images with symptoms that are not ruled in the system. A good example of this is images with PLM caused by PRP treatment or images with marks on the retina which can signify other possible diseases, such as vein occlusion, macular degeneration, non-diabetic retinopathy, etc.

Although it may seem like a big issue, it is also easily fixed. PRP is a treatment that comes "after" the diagnosis of DR. In other words, images containing the PLM imply that the patient has been diagnosed with DR before, and treatment has been made, which means that the patient already has the medical record at the hospital. Moreover, if the correlation of PLM and

the severity of the PRP is found, one can simply add it to the system – in the current model, the PLM has not been considered as input, as its data was inaccessible from the FGADR dataset yet

CHAPTER 6

CONCLUSIONS

In this thesis, we have developed a novel FIS method that allows accommodating experts' opinions and understandings while maintaining the consistency and reproducibility of the CAD system. The proposed system was able to acquire an overall accuracy of 87.04%. Through ongoing refinement and collaboration, we hope the proposed model to make a meaningful impact in both the healthcare industry and academic research.

The future works that can be added is to attempt replicating an agreeable severity scale that could be utilized locally based on the doctors' inputs, or creating a system that allows patients to diagnose themselves with the help of smartphone fundus photography in the poor countries where both the equipment and the doctors are scarce [35].

APPENDIX A.

Diagnostic Questionnaire for Fuzzy Inference System

This is the PROM questionnaire that was utilized to acquire a numerical dataset from the image-based dataset. This table is given to personnel who were responsible for data acquisition after training on how to identify each symptom within the image. The people who were responsible for processing the data had ample experience in converting real-life data into quantitative data through their experience in the AI lab, and to promote fairness, they were not told what the purpose of this data refining was, or any information regarding the proposed FIS.

(Intraretinal) Hemorrhage (IRH) 0~10:

0 = absent

1~3 = present in 1, over 20 hemorrhage major per quad (>5 in image)

4~6 = present in 2 quads, over 20 major hemorrhages per quad (>10 in image)

7~8 = present in 3 quads, over 20 major hemorrhages per quad (>15 in image)

9~10 =present in 4 quads, over 20 major hemorrhages per quad (>20 in image)

Microaneurysm (MA): 0~10

0 = absent

1~3 = present in 1 quad

4~6 = present in 2 quads

7~8 = present in 3 quads

9~10 = present in 4 quads

Hard Exudates (HE): 0, 1

0 = absent

1 = present

Intraretinal Microvascular Abnormality (IRMA): 0~4

0 = absent

1 = present in 1 quad

2 = present in 2 quads, or vascular beading (VB) in 1 quad

3 = present in 3 quads

4 = present in 4 quads, VB in 2 quads, or prominently present in 1 quad

Cotton Wool Spots (CWS, Soft Exudates): 0, 1

0 = absent

1 = present

Neovascularization (NV) or Hemorrhage (VH or PRH): 0, 1

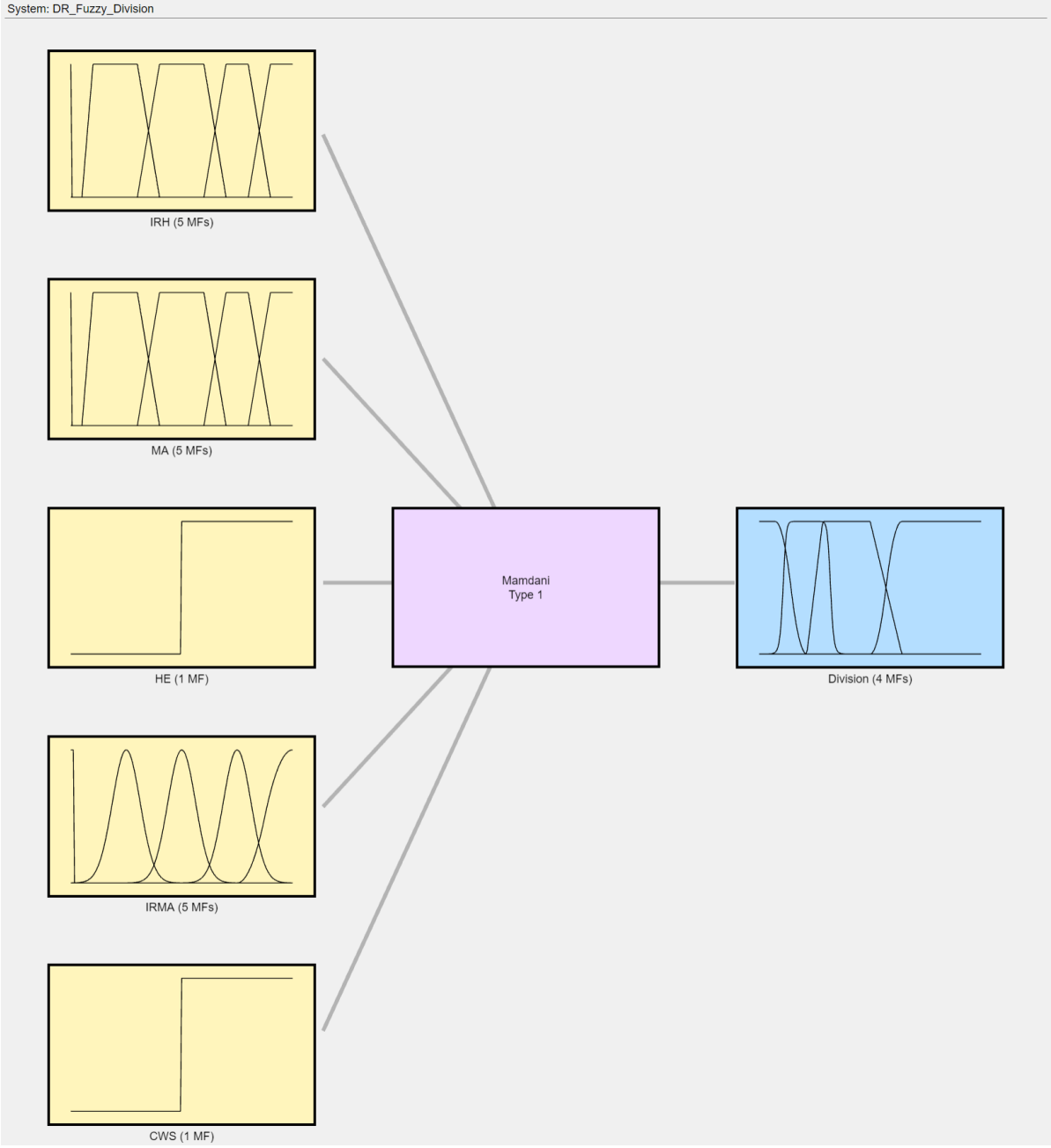
0 = absent

1 = present

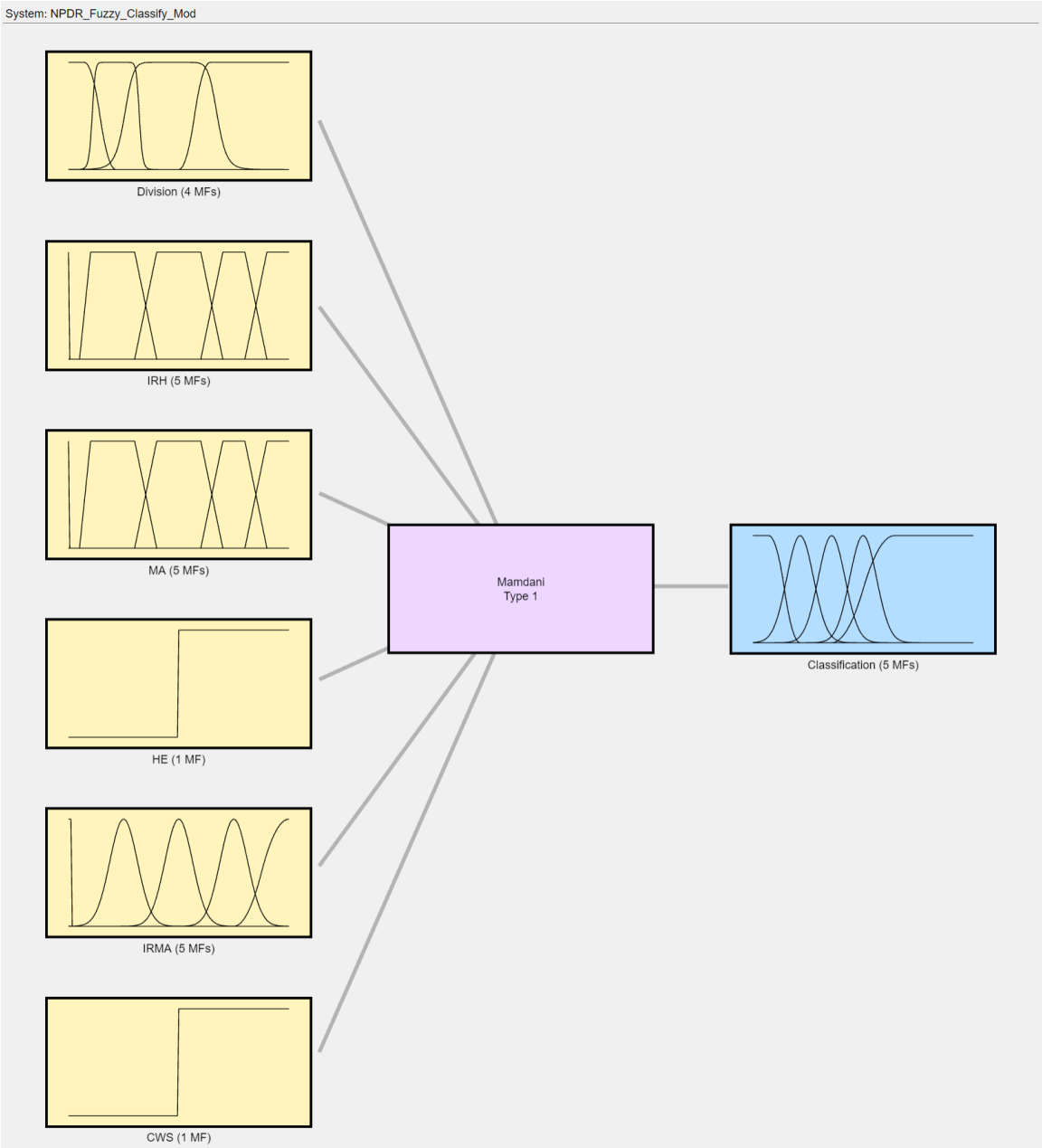
APPENDIX B.

Fuzzy Subsystems Implemented in MATLAB

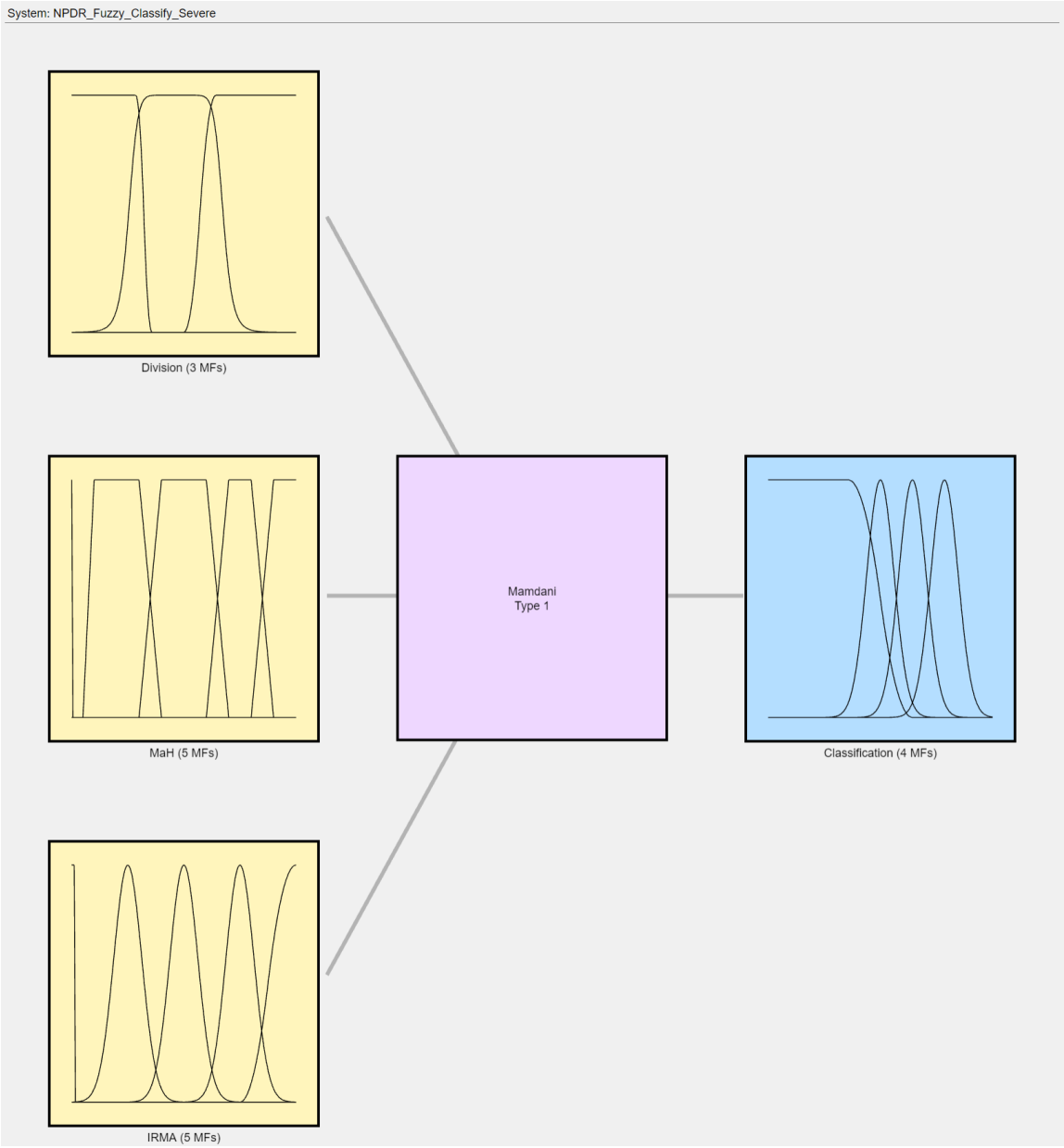
1. DR_Fuzzy_Division



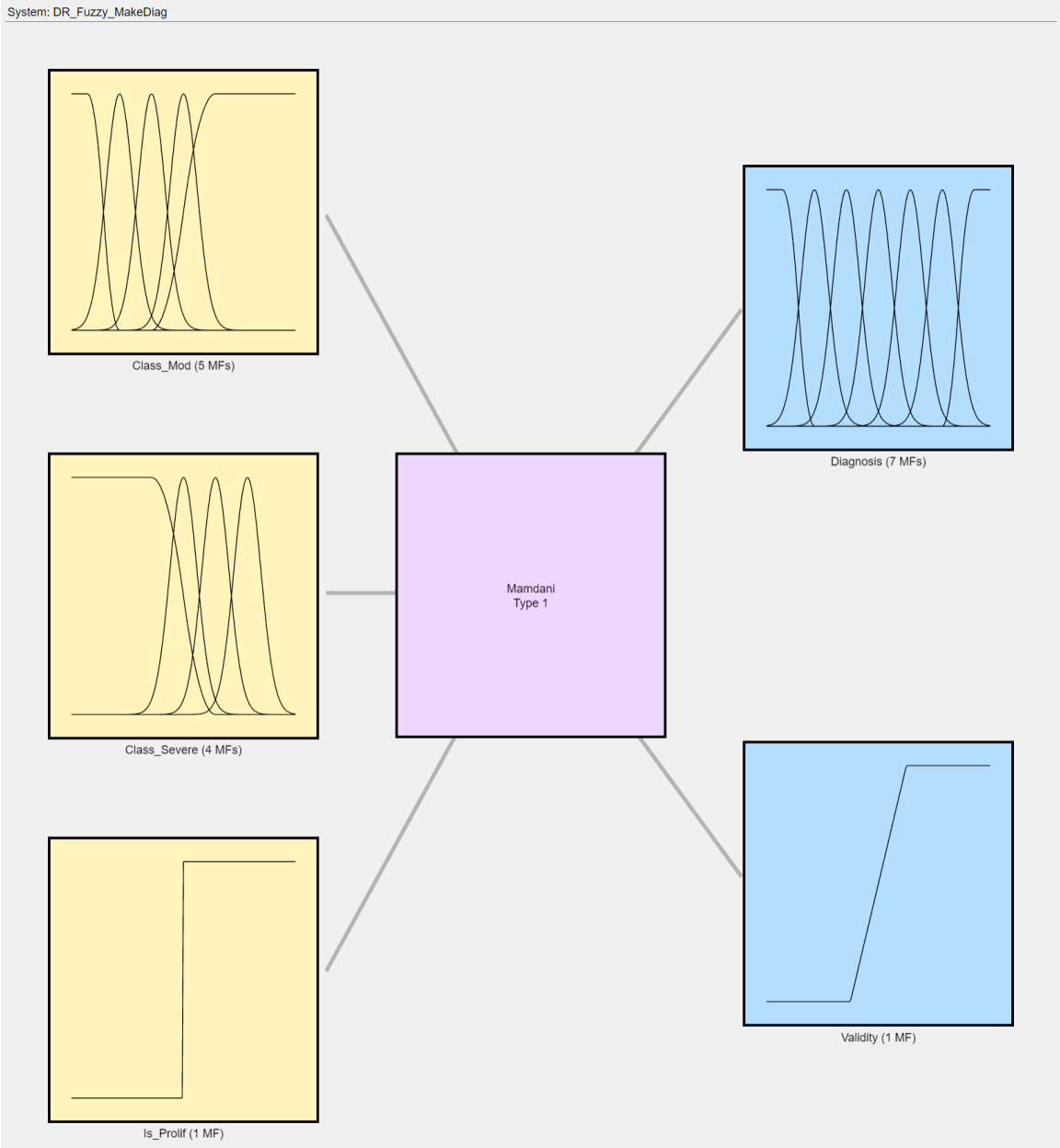
2. NPDR_Fuzzy_Classify_Mod



3. NPDR_Fuzzy_Classify_Severe



4. DR_Fuzzy_MakeDiag



APPENDIX C.

Dataset - FGADR

#	H	MA	HE	IRMA	CWS	NV/VH/ PRH	Ground Truth	Modified Ground Truth	Output	Validity
1	3	1	1	0	0	0	2	1	0.95	0.99
2	7.5	2	0	0	0	0	3	2	1.99	0.96
3	1	1	0	0	0	0	2	1	0.95	0.99
4	5	3	1	0	1	0	4	3	1.06	0.92
5	5	2	1	0	0	0	2	1	1.06	0.92
6	0	1	0	0	0	0	1	0.5	0.60	0.98
7	7	1	0	0	0	0	2	1	1.99	0.96
8	8	2	1	0	0	0	4	3	1.99	0.96
9	3	7	1	2	0	0	3	2	1.93	0.92
10	5	2	0	0	1	0	2	1	1.03	0.94
11	1	1	0	0	1	0	2	1	0.95	0.99
12	10	3	1	0	1	0	3	2	2.04	0.99
13	7	4	1	0	0	0	3	2	1.93	0.92
14	2	5	0	0	0	0	2	1	1.34	0.90
15	7	4	1	0	0	0	3	2	1.93	0.92
16	7.5	1	1	1	1	0	3	2	1.99	0.96
17	10	5	1	2	1	0	3	2	1.97	0.94
18	1	1	0	0	0	0	3	2	0.95	0.99
19	0	1	1	0	0	0	2	1	0.95	0.99
20	7.5	3	1	0	0	0	2	1	1.99	0.96
21	3	2	1	0	0	0	2	1	0.95	0.99
22	2.5	0	1	0	1	0	4	3	0.95	0.99
23	0	2	1	0	0	0	2	1	0.95	0.99
24	8	7	1	0	1	0	3	2	1.97	0.94
25	9	3	1	0	1	0	3	2	2.04	0.99
26	8	3	1	0	0	0	3	2	1.99	0.96
27	8	5	1	0	1	0	3	2	1.93	0.92
28	0	0	0	0	0	0	0	0	0.06	0.99
29	1	1	1	0	1	0	2	1	0.95	0.99
30	2.5	2	1	0	0	0	2	1	0.95	0.99
31	8	4	1	0	1	1	4	3	3.00	0.99
32	8.5	3	1	0	1	1	4	3	3.00	0.99
33	5	2	1	0	0	0	3	2	1.06	0.92

34	0	0	0	0	0	0	2	1	0.06	0.99
35	0	0	0	0	0	0	0	0	0.06	0.99
36	0	0	0	0	0	0	1	0.5	0.06	0.99
37	0	1	0	0	1	0	2	1	0.95	0.99
38	7	2	1	0	0	0	3	2	1.99	0.96
39	8	3	1	0	0	0	3	2	1.99	0.96
40	5.5	3	1	0	1	0	3	2	1.06	0.92
41	0	0	0	0	0	0	0	0	0.06	0.99
42	1	2	1	0	0	0	2	1	0.95	0.99
43	9	4	1	0	0	0	4	3	2.04	0.99
44	5.5	2	1	2	0	0	3	2	1.50	0.90
45	5	0	0	2	1	0	3	2	1.50	0.90
46	1	1	0	0	0	0	1	0.5	0.95	0.99
47	3	2	1	0	0	0	3	2	0.95	0.99
48	5	7	1	0	1	0	3	2	1.93	0.92
49	7	6	1	3.5	0	0	4	3	1.93	0.92
50	3	5	0	4	0	0	4	3	2.25	0.94
51	1	2	1	2	0	0	3	2	1.50	0.90
52	2	0	1	0	0	0	3	2	0.95	0.99
53	0	1	0	0	0	0	1	0.5	0.60	0.98
54	7	5	1	2	0	0	3	2	1.93	0.92
55	0	0	1	2	0	0	3	2	1.50	0.90
56	3	3	0	0	0	0	2	1	0.95	0.99
57	7	5	1	0	0	0	3	2	1.93	0.92
58	5	2	1	2	1	0	3	2	1.50	0.90

5.

APPENDIX D.

Dataset - APTOS

#	H	MA	HE	IRMA	CWS	NV/VH/ PRH	Ground Truth	Modified Ground Truth	Output	Validity
1	3	1	1	0	0	0	2	1	0.95	0.99
2	7.5	2	0	0	0	0	3	2	1.99	0.96
3	1	1	0	0	0	0	2	1	0.95	0.99
4	5	3	1	0	1	0	4	3	1.06	0.92
5	5	2	1	0	0	0	2	1	1.06	0.92
6	0	1	0	0	0	0	1	0.5	0.60	0.98
7	7	1	0	0	0	0	2	1	1.99	0.96
8	8	2	1	0	0	0	4	3	1.99	0.96
9	3	7	1	2	0	0	3	2	1.93	0.92
10	5	2	0	0	1	0	2	1	1.03	0.94
11	1	1	0	0	1	0	2	1	0.95	0.99
12	10	3	1	0	1	0	3	2	2.04	0.99
13	7	4	1	0	0	0	3	2	1.93	0.92
14	2	5	0	0	0	0	2	1	1.34	0.90
15	7	4	1	0	0	0	3	2	1.93	0.92
16	7.5	1	1	1	1	0	3	2	1.99	0.96
17	1	1	1	0	0	0	4	3	0.95	0.99
18	10	5	1	2	1	0	3	2	1.97	0.94
19	1	1	0	0	0	0	3	2	0.95	0.99
20	0	1	1	0	0	0	2	1	0.95	0.99
21	7.5	3	1	0	0	0	2	1	1.99	0.96
22	3	2	1	0	0	0	2	1	0.95	0.99
23	2.5	0	1	0	1	0	4	3	0.95	0.99
24	0	2	1	0	0	0	2	1	0.95	0.99
25	8	7	1	0	1	0	3	2	1.97	0.94
26	9	3	1	0	1	0	3	2	2.04	0.99
27	8	3	1	0	0	0	3	2	1.99	0.96
28	8	5	1	0	1	0	3	2	1.93	0.92
29	0	0	0	0	0	0	0	0	0.06	0.99
30	1	1	1	0	1	0	2	1	0.95	0.99
31	2.5	2	1	0	0	0	2	1	0.95	0.99
32	8	4	1	0	1	1	4	3	3.00	0.99
33	8.5	3	1	0	1	1	4	3	3.00	0.99

34	5	2	1	0	0	0	3	2	1.06	0.92
35	0	0	0	0	0	0	2	1	0.06	0.99
36	0	0	0	0	0	0	0	0	0.06	0.99
37	0	0	0	0	0	0	1	0.5	0.06	0.99
38	0	1	0	0	1	0	2	1	0.95	0.99
39	7	2	1	0	0	0	3	2	1.99	0.96
40	8	3	1	0	0	0	3	2	1.99	0.96
41	5.5	3	1	0	1	0	3	2	1.06	0.92
42	0	0	0	0	0	0	0	0	0.06	0.99
43	1	2	1	0	0	0	2	1	0.95	0.99
44	9	4	1	0	0	0	4	3	2.04	0.99
45	5	3	1	0	1	0	4	3	1.06	0.92
46	5.5	2	1	2	0	0	3	2	1.50	0.90
47	5	0	0	2	1	0	3	2	1.50	0.90
48	1	1	0	0	0	0	1	0.5	0.95	0.99
49	3	4	1	0	1	0	4	3	1.34	0.89
50	3	2	1	0	0	0	3	2	0.95	0.99

REFERENCES

[1] *Body Basics: The Eye*. [Online]. Available: https://kidshealth.org/en/kids/eyes.html

[2] V. Pittu, S. R. Avanapu, and J. Sharma, "'Diabetic Retinopathy - Can Lead to Complete Blindness'.," *Int. J. Sci. Invent. Today*, vol. 2, pp. 254-265., Jan. 2013.

[3] C. P. Wilkinson, F. L. Ferris, R. E. Klein, P. P. Lee, C. D. Agardh, M. Davis, D. Dills, A. Kampik, R. Pararajasegaram, and J. T. Verdaguer, "Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales," *Ophthalmology*, vol. 110, no. 9, pp. 1677–1682, Sep. 2003, doi: 10.1016/S0161-6420(03)00475-5.

[4] "Case Study: TensorFlow in Medicine - Retinal Imaging (TensorFlow Dev Summit 2017) - YouTube." Accessed: Aug. 28, 2023. [Online]. Available: https://www.youtube.com/watch?v=oOeZ7IgEN4o

[5] "Diabetic eye screening: programme overview," GOV.UK. Accessed: Nov. 06, 2023. [Online]. Available: https://www.gov.uk/guidance/diabetic-eye-screening-programme-overview

[6] G. Saleh, N. Batouty, S. Haggag, A. Elnakib, F. Khalifa, Dr. F. Taher, M. Mohamed, R. Farag, H. Sandhu, A. Sewelam, and A. El-Baz, "The Role of Medical Image Modalities and AI in the Early Detection, Diagnosis and Grading of Retinal Diseases: A Survey," *Bioengineering*, vol. 9, p. 366, Aug. 2022, doi: 10.3390/bioengineering9080366.

[7] iproweb, “OCT Examination VS Fundus Photography: What to Choose.” Accessed: Nov. 21, 2023. [Online]. Available: https://www.altris.ai/article/oct-examination-vs-fundus-photography-which-method-to-choose/

[8] “Diabetic Retinopathy | National Eye Institute.” Accessed: Aug. 28, 2023. [Online]. Available: https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/diabetic-retinopathy

[9] X.-Y. Li, S. Wang, L. Dong, and H. Zhang, “Comparison of fundus fluorescein angiography and fundus photography grading criteria for early diabetic retinopathy,” *Int. J. Ophthalmol.*, vol. 15, no. 2, pp. 261–267, Feb. 2022, doi: 10.18240/ijo.2022.02.11.

[10] T. Kaur, J. Singh, V. Rehani, A. Sharma, and A. John, “Diabetic Retinopathy Detection System (DRDS) ; A Novel GUI Based Approach for Diabetic Retinopathy Detection,” 2017. Accessed: Nov. 05, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Diabetic-Retinopathy-Detection-System-(-DRDS-)-%3B-A-Kaur-Singh/ad5ef21b7476a5be3e2ff5532718bd0eac3c4f6f

[11] N. Malhotra, V. Kapoor, and Ratish, “Fuzzy Based Decision Making System for the Detection of Diabetic Retinopathy,” *Int. J. Trend Sci. Res. Dev.*, Jul. 2020, Accessed: Nov. 05, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Fuzzy-Based-Decision-Making-System-for-the-of-Malhotra-Kapoor/3d714966d5c229ff284e9e4de8321330f53d629c

[12] M. F. Goldberg and L. M. Jampol, “Knowledge of diabetic retinopathy before and 18 years after the Airlie House Symposium on Treatment of Diabetic Retinopathy,” *Ophthalmology*, vol. 94, no. 7, pp. 741–746, Jul. 1987, doi: 10.1016/s0161-6420(87)33524-9.

[13] J. K. Kristinsson, E. Stefánsson, F. Jónasson, I. Gíslason, and S. Björnsson, “Screening for eye disease in type 2 diabetes mellitus,” *Acta Ophthalmol. (Copenh.)*, vol. 72, no. 3, pp. 341–346, Jun. 1994, doi: 10.1111/j.1755-3768.1994.tb02770.x.

[14] J. K. Kristinsson, E. Stefánsson, F. Jónasson, I. Gíslason, and S. Björnsson, “Systematic screening for diabetic eye disease in insulin dependent diabetes,” *Acta Ophthalmol. (Copenh.)*, vol. 72, no. 1, pp. 72–78, Feb. 1994, doi: 10.1111/j.1755-3768.1994.tb02741.x.

[15] E. Agardh, C. D. Agardh, and C. Hansson-Lundblad, “The five-year incidence of blindness after introducing a screening programme for early detection of treatable diabetic retinopathy,” *Diabet. Med. J. Br. Diabet. Assoc.*, vol. 10, no. 6, pp. 555–559, Jul. 1993, doi: 10.1111/j.1464-5491.1993.tb00120.x.

[16] “Automated detection of microaneurysms in digital red-free photographs: a diabetic retinopathy screening tool - PubMed.” Accessed: Nov. 05, 2023. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/11073180/

[17] Li M., Wang G., Xia H., Tang X., Feng Z., Yao Y., Huang Y., Fan W., Yuan Z., and Yuan J., “Clinical evaluation of artificial intelligence system based on fundus photograph in diabetic retinopathy screening,” *Chin. J. Exp. Ophthalmol.*, pp. 663–668, 2019.

[18] M. D. Abràmoff, P. T. Lavin, M. Birch, N. Shah, and J. C. Folk, “Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices,” *NPJ Digit. Med.*, vol. 1, p. 39, Aug. 2018, doi: 10.1038/s41746-018-0040-6.

[19] M. D. Abràmoff, Y. Lou, A. Erginay, W. Clarida, R. Amelon, J. C. Folk, and M. Niemeijer, “Improved Automated Detection of Diabetic Retinopathy on a Publicly Available

Dataset Through Integration of Deep Learning," *Invest. Ophthalmol. Vis. Sci.*, vol. 57, no. 13, pp. 5200–5206, Oct. 2016, doi: 10.1167/iovs.16-19964.

[20] F. K. Malerbi, R. E. Andrade, P. H. Morales, J. A. Stuchi, D. Lencione, J. V. de Paulo, M. P. Carvalho, F. S. Nunes, R. M. Rocha, D. A. Ferraz, and R. Belfort, "Diabetic Retinopathy Screening Using Artificial Intelligence and Handheld Smartphone-Based Retinal Camera," *J. Diabetes Sci. Technol.*, vol. 16, no. 3, pp. 716–723, Jan. 2021, doi: 10.1177/1932296820985567.

[21] G. Selvachandran, S. G. Quek, R. Paramesran, W. Ding, and L. H. Son, "Developments in the detection of diabetic retinopathy: a state-of-the-art review of computer-aided diagnosis and machine learning methods," *Artif. Intell. Rev.*, vol. 56, no. 2, pp. 915–964, 2023, doi: 10.1007/s10462-022-10185-6.

[22] B. J. T. OD OD, Anna Kathryn Bedwell, OD, Daniel Bollier, OD, and Brad Sutton, "My Patient Has Diabetic Retinopathy...Now What?," Review of Optometry: Leadership in clinical care. Accessed: Sep. 01, 2023. [Online]. Available: https://www.reviewofoptometry.com/article/my-patient-has-diabetic-retinopathynow-what

[23] L. Ribeiro, C. M. Oliveira, C. Neves, J. D. Ramos, H. Ferreira, and J. Cunha-Vaz, "Screening for Diabetic Retinopathy in the Central Region of Portugal. Added Value of Automated 'Disease/No Disease' Grading," *Ophthalmol. J. Int. Ophtalmol. Int. J. Ophthalmol. Z. Augenheilkd.*, Nov. 2014, doi: 10.1159/000368426.

[24] C. for D. and R. Health, "Clinical Decision Support Software." Accessed: Aug. 25, 2023. [Online]. Available: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-decision-support-software

[25] *Software Thought Leadership Series - FDA REGULATION OF CLINICAL DECISION SUPPORT SOFTWARE AND AI/ML*, (Feb. 08, 2023). Accessed: Aug. 25, 2023. [Online Video]. Available: https://www.youtube.com/watch?v=YMU_UAqQoSo

[26] S. Alayón, R. Robertson, S. K. Warfield, and J. Ruiz-Alzola, "A Fuzzy System for Helping Medical Diagnosis of Malformations of Cortical Development," *J. Biomed. Inform.*, vol. 40, no. 3, pp. 221–235, Jun. 2007, doi: 10.1016/j.jbi.2006.11.002.

[27] Y. Zhou, B. Wang, L. Huang, S. Cui, and L. Shao, "A Benchmark for Studying Diabetic Retinopathy: Segmentation, Grading, and Transferability," *IEEE Trans. Med. Imaging*, vol. 40, no. 3, pp. 818–828, 2021, doi: 10.1109/TMI.2020.3037771.

[28] A. Shaout and J. Han, "A Novel Retinal Image Contrast Enhancement - Fuzzy Based Method." University of Michigan (Unpublished internal document).

[29] "APTOS 2019 Blindness Detection." [Online]. Available: https://www.kaggle.com/competitions/aptos2019-blindness-detection/data

[30] S. Wang, Y. Zhang, S. Lei, H. Zhu, J. Li, Q. Wang, J. Yang, S. Chen, and H. Pan, "Performance of deep neural network-based artificial intelligence method in diabetic retinopathy screening: a systematic review and meta-analysis of diagnostic test accuracy," *Eur. J. Endocrinol.*, vol. 183, no. 1, pp. 41–49, Jul. 2020, doi: 10.1530/EJE-19-0968.

[31] V. Raudonis, A. Kairys, R. Verkauskiene, J. Sokolovska, G. Petrovski, V. J. Balciuniene, and V. Volke, "Automatic Detection of Microaneurysms in Fundus Images Using an Ensemble-Based Segmentation Method," *Sensors*, vol. 23, no. 7, Art. no. 7, Jan. 2023, doi: 10.3390/s23073431.

[32] S. Long, J. Chen, A. Hu, H. Liu, Z. Chen, and D. Zheng, “Microaneurysms detection in color fundus images using machine learning based on directional local contrast,” *Biomed. Eng. OnLine*, vol. 19, no. 1, p. 21, Apr. 2020, doi: 10.1186/s12938-020-00766-3.

[33] M. Arya, O. Sorour, J. Chaudhri, Y. Alibhai, N. K. Waheed, J. S. Duker, and C. R. Baumal, “Distinguishing Intraretinal Microvascular Abnormalities From Retinal Neovascularization Using Optical Coherence Tomography Angiography,” *Retina Phila. Pa*, vol. 40, no. 9, pp. 1686–1695, Sep. 2020, doi: 10.1097/IAE.0000000000002671.

[34] M. Torres Soriano, G. García-Aguirre, M. Gordon, and V. Kon-Graversen, *Diagnostic Atlas of Retinal Diseases Volume 1*. 2018. doi: 10.2174/97816810835751160101.

[35] U. Iqbal, “Smartphone fundus photography: a narrative review,” *Int. J. Retina Vitr.*, vol. 7, no. 1, p. 44, Jun. 2021, doi: 10.1186/s40942-021-00313-9.