Fuzzy Inference System for Human-Perceptible Diagnosis of Diabetic Retinopathy

JiHo Han
College of Engineering & Computer Science
University of Michigan
Michigan, USA
jihohan@umich.edu

Adnan Shaout
College of Engineering & Computer Science
University of Michigan
Michigan, USA
shaout@umich.edu

***Abstract*—Despite the enthusiasm surrounding CAD, 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). Upon testing the proposed model using FGADR and APTOS datasets, an accuracy of 77.59% and 98% are acquired respectively. An edge case testing is conducted to check whether the proposed system correctly identifies misidentified data, validating the objectives of the research.**

Keywords—diabetic retinopathy; fundus photography; computer-assisted diagnosis; fuzzy inference system;

I. INTRODUCTION

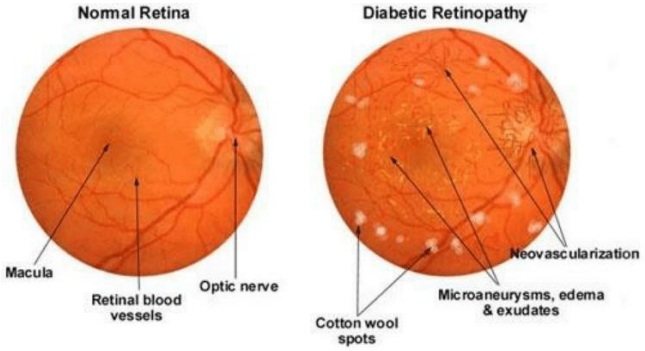

Fig. 1. *Difference between Normal Retina and Diabetic Retinopathy Fundus Image*

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 (Fig. 1.).

A. Stages of Diabetic Retinopathy

There are multiple ways of staging the severity of diabetic retinopathy. In this journal, 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 I. ICDR SEVERITY SCALE FOR DIABETIC RETINOPATHY

Proposed disease severity level	Findings observable by dilated ophthalmoscopy
No diabetic retinopathy	No abnormalities
Mild Nonproliferative Diabetic Retinopathy (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 1quadrant and no signs of proliferative retinopathy
Proliferative Diabetic Retinopathy PDR	One or more of the following: - Neovascularization - vitreous/pre-retinal hemorrhage

II. Literature Review of Diabetic Retinopathy And Its Diagnosis Methods

A. 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.

B. 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 [1], [2].

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 (Figure 3). 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 [3].

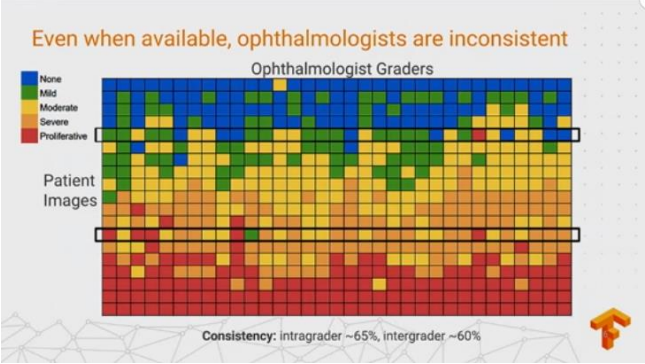

Fig. 2. *Case Study: TensorFlow in Medicine - Retinal Imaging [3]*

C. Advantages of the CAD in DR 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% [4].

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.

D. Limitations of the CAD in DR 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 [5], [6].

III. Proposed Method

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 [7]. 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.

A. Fuzzy Inference System

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

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 rubrics are inspired by incorporating common aspects of DR severity scale tables [1], [2]. 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.

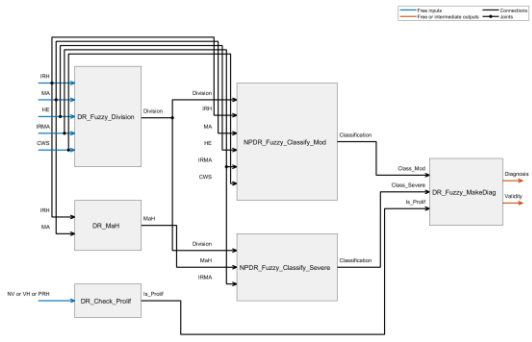

Fig. 3. *FIS Tree Model of Implemented System*

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"* (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 [1], 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"*.

TABLE II. FUZZY RULE OF "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

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 middle 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 [2].

The fuzzy rules for the "*NPDR_Fuzzy_Classify_Mod"* and "*NPDR_Fuzzy_Classify_ Severe"* can be found in Tables III and IV. 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.

TABLE III. FUZZY RULE OF "NPDR_FUZZY_CLASSIFY_MOD"

#	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 &	0.25

	IRMA==present in 1 quad as mild => Classification=moderate	
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

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"*.

TABLE IV. FUZZY RULE OF "DR_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

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 V shows the fuzzy rules for "*DR_Fuzzy_MakeDiag"*.

TABLE V. FUZZY RULE OF "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

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.

B. Parameters

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 fundus photography (FP). Similar to 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 as follows:

- Intraretinal Hemorrhage (IRH): 0-10, based on the severity.
  - 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
- Microaneurysm (MA): 0-10, based on the severity.
  - 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)
- Hard Exudates (HE): 0 if absent, 1 if present.
- Intraretinal Microvascular Abnormality (IRMA): 0-4. 0 denotes the absence of IRMA, ranging up to 4. Any mild to moderate IRMA (e.g. Venous beading) will be denoted around 2, whereas a "prominent" IRMA will be noted as 4.
- Cotton Wool Spots (CWS, also known as Soft Exudates (SE)): 0 if absent, 1 if present.
- Neovascularization (NV) or Vitreous / Preretinal Hemorrhage (VH / PRH): 0 if absent, 1 if present.

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.

The rubrics are inspired by incorporating common aspects of renowned DR severity scale tables. 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.

C. Dataset Used

The system is tested through some of the images from the Fine-Grained Annotated Diabetic Retinopathy (FGADR) dataset [8]. 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 [9]. To simulate the human ophthalmologists' diagnoses, the dataset was modified to have additional data required in a crisp numerical scale for diagnosis.

To provide analysis with a more generally available dataset, an Asia Pacific Tele-Ophthalmology Society (APTOS) dataset was utilized [10]. 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, 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.

IV. RESULTS

In the case of the FGADR dataset, 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), 52 images were diagnosed correctly within the error range of 1 (89.66% accuracy).

For the APTOS dataset, 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. Including the image that showed low validity, 9 of the 49 images showed an error margin of 0.5. Most of this discrepancy originated from distinguishing between mild and moderate NDPR.

Overall, an accuracy of 87.04% has been achieved, with 93.52% accuracy within the error range of 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.

The main reason behind the diagnostic inaccuracy comes from the discrepancy in classification standards within the dataset. Both images in Figure 4 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 I). However, unlike the standard scale, the APTOS dataset classified both images as 1.

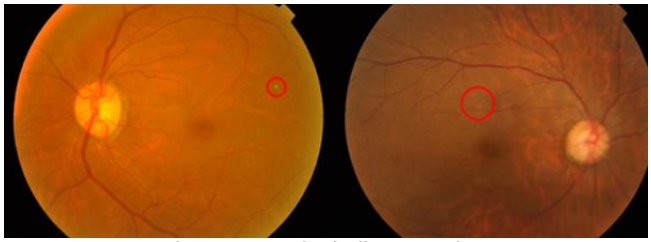

Fig. 4. *Sample Images of Misdiagnosed APTOS Images*

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 (Table VI).

TABLE VI. INPUT PARAMETERS OF MISDIAGNOSED APTOS IMAGES

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

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

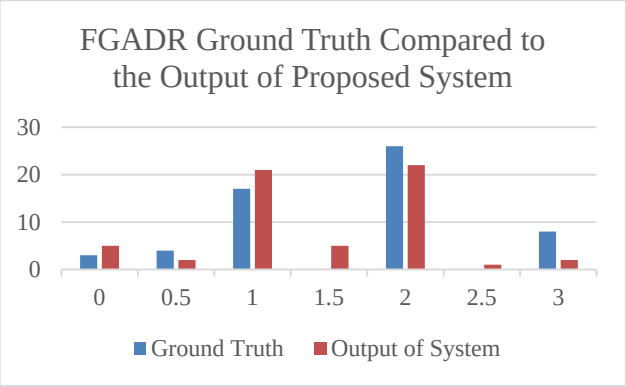

Fig. IV5. *FGADR Ground Truth Compared to the Output of Proposed System*

As depicted in Fig 5., it is quite evident that the diagnosis made in FGADR tended to mark the condition of the eyes as more serious

V. DISCUSSION

A. *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. Fig. 6., 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 I [1].

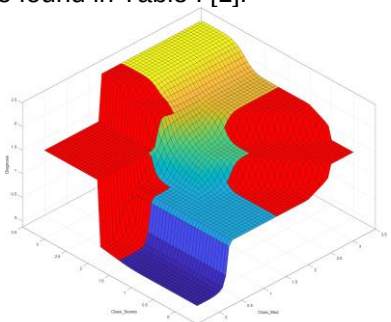

Fig. 6. *Diagram of “DR_Fuzzy_MakeDiag” When “Is_Prolif” Is Not Present. The Area Marked Red Is Where the Validity Is Low.*

In Fig. 7., 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 right 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.

Fig. 7. *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 Fig. 8 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.

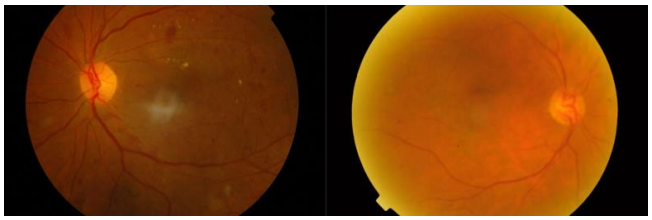

Fig. 8. *Images That Are Diagnosed Same in APTOS, but Found to be a Potential Diagnostic Error in the Proposed System*

Another standout feature of the proposed model is its ability to incorporate an expert's diagnosis into the model. Another noteworthy feature of the FIS Tree is its customizable nature. This adaptability ensures that the FIS Tree 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.

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 the results section – 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.

B. 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.

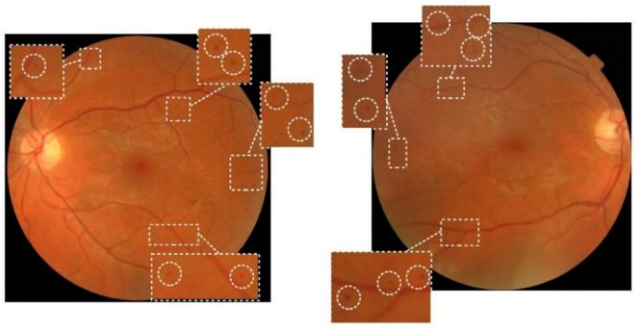

Fig. 9. *Images of Several Microaneurysms, One of the Observable Findings Used for DR Diagnosis*

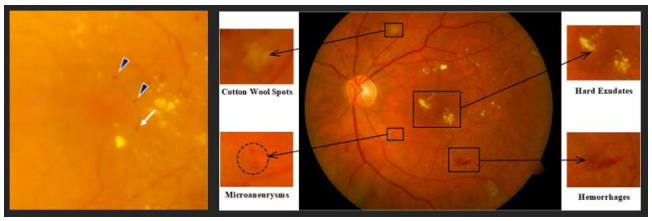

Fig. 10. *Two Figures Each Displaying Similarities of Distinct Symptoms. Microaneurysms (Black Arrowheads) and Dot Hemorrhages (White Arrows) in FP (Left) [12], Sample Image Which Contains All Intermediate-Level DR Features (Right)[13]*

In the case of Fig. 9., 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 Fig. 9., 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 Fig. 10., 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 [14].

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 [15]. Correlation analysis in the proposed model supports this fact (Fig. 11.), 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.

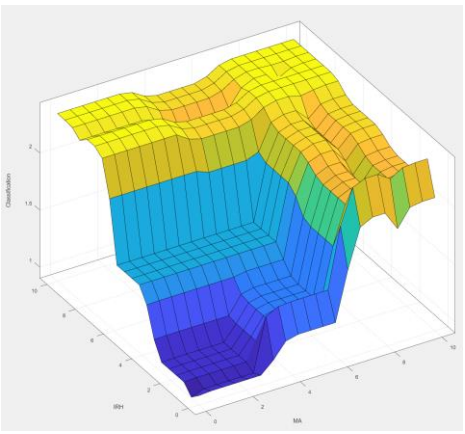

Fig. 11. *Diagram of the Correlation Analysis of IRH and MA in Output of "NPDR_Fuzzy_Classify_Mod"*

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.

VI. Conclusion

In this paper, 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 [16].

Acknowledgment

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.

REFERENCES

[1] C. P. Wilkinson *et al.*, “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.

[2] 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

[3] “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

[4] 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.

[5] 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

[6] *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

[7] 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.

[8] 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.

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

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

[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] G. Saleh *et al.*, “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.

[13] 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.

[14] S. Ruia and K. Tripathy, “Fluorescein Angiography,” in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2023. Accessed: Nov. 25, 2023. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK576378/

[15] M. Torres Soriano, G. García-Aguirre, M. Gordon, and V. Kon-Graversen, *Diagnostic Atlas of Retinal Diseases Volume 1*. 2018. doi: 10.2174/9781681083575116010l.

[16] 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.