Few-Shot Learning of Diagnostic Rules for Neurodegenerative Diseases Using Inductive Logic Programming

Dany Varghese[1(✉)], Roman Bauer[1,2], and Alireza Tamaddoni-Nezhad[1]

[1] Department of Computer Science, University of Surrey, Guildford, UK
{dany.varghese,r.bauer,a.tamaddoni-nezhad}@surrey.ac.uk
[2] School of Computing, Newcastle University, Newcastle upon Tyne, UK

Abstract. Traditional machine learning methods heavily rely on large amounts of labelled data for effective generalisation, posing a challenge in few-shot learning scenarios. In many real-world applications, acquiring large amounts of training data can be difficult or impossible. This paper presents an efficient and explainable method for few-shot learning from images using inductive logic programming (ILP). ILP utilises logical representations and reasoning to capture complex relationships and generalise from sparse data. We demonstrate the effectiveness of our proposed ILP-based approach through an experimental evaluation focused on detecting neurodegenerative diseases from fundus images. By extending our previous work on neurodegenerative disease detection, including Alzheimers disease, Parkinsons disease, and vascular dementia disease, we achieve improved explainability in identifying these diseases using fundus images collected from the UK Biobank dataset. The logical representation and reasoning inherent in ILP enhances the interpretability of the detection process. The results highlight the efficacy of ILP in few-shot learning scenarios, showcasing its remarkable generalisation performance compared to a range of other machine learning algorithms. This research contributes to the field of few-shot learning using ILP and paves the way for addressing challenging real-world problems.

1 Introduction

Few-shot learning [2] is a challenging task in machine learning that aims to enable models to generalise and make accurate predictions with only a limited amount of labelled training data available. Traditional machine learning models typically rely on large amounts of labelled data for training to achieve high accuracy. However, when faced with scenarios where only a few training examples are available, these models often struggle to generalise effectively. In contrast, humans exhibit a remarkable capability for one-shot or few-shot learning, where they can quickly grasp new concepts and make accurate predictions with minimal exposure to data. The key reason for the disparity between machine learning models and human performance in few-shot learning lies in the inherent differences in their learning mechanisms. Machine learning models, especially those based on

E. Bellodi et al. (Eds.): ILP 2023, LNAI 14363, pp. 109–123, 2023.
https://doi.org/10.1007/978-3-031-49299-0_8

deep neural networks, rely on a data-driven approach where patterns and representations are learned through optimisation processes. These models require substantial amounts of labelled data to capture the complexity of the underlying problem and generalise effectively. In contrast, humans possess innate cognitive abilities that enable them to reason, abstract, and leverage prior knowledge when encountering novel tasks or scarce data.

Human learners can draw upon their vast prior knowledge and existing mental frameworks to make inferences and generalise from limited examples. They can recognise commonalities, abstract underlying concepts, and adapt previous knowledge to new situations. These cognitive abilities, coupled with an innate capacity for transfer learning, allow humans to excel at few-shot learning tasks. Additionally, humans possess a rich set of prior experiences, enabling them to leverage contextual cues, background knowledge, and intuition, often lacking in machine learning models.

Inductive Logic Programming (ILP) [16] offers a unique perspective and approach to address the limitations of traditional machine learning in few-shot learning scenarios. By leveraging logical representations and reasoning, ILP can capture and exploit domain-specific knowledge and prior assumptions. This ability allows ILP models to learn from sparse data by generalising from a small number of examples [5,29]. ILP is a subfield of machine learning that combines the principles of logic programming and inductive reasoning to learn logical rules from examples. Unlike traditional machine learning approaches that focus on statistical patterns in data, ILP incorporates logical representations and reasoning to capture complex relationships and generalise knowledge.

The logical nature of ILP enables it to represent complex relationships and dependencies explicitly. Using logical rules and constraints, ILP models can reason and make inferences beyond the observed examples, providing a strong foundation for few-shot learning. ILP also benefits from incorporating background knowledge, including prior domain expertise, into the learning process. This prior knowledge helps guide the learning process and facilitates better generalisation, even with limited training data. Furthermore, ILP's capability to handle structured data, such as relational databases or ontologies, is advantageous in few-shot learning tasks that involve complex relationships and hierarchical structures. By representing data logically, ILP can effectively exploit the inherent structure and dependencies in the data, enabling more effective learning from a few examples.

In a previous paper [26] we introduced an ILP approach called One-Shot Hypothesis Derivation (OSHD) and we used this for one-shot learning from retinal images to detect neurodegenerative diseases. Building upon the previous work, we propose a novel methodology that utilises a histogram-based binning method for improving interpretability and accuracy in detecting neurodegenerative diseases from retinal images. We also extend the previous study by using state-of-the-art ILP systems PyGol and Metagol as well as comparing with a range of other non-ILP machine learning methods. In the experiments, we focus on the challenging task of learning diagnostic rules for neurodegenerative diseases, including Alzheimer's, Parkinson's, and vascular dementia, from small

number of training examples (fundus images). With limited labelled data available for each disease, we investigated whether ILP could successfully learn the discriminative features and accurately classify the diseases based on the fundus images. Through these experiments, we aim to showcase the effectiveness of ILP in addressing few-shot learning challenges. By leveraging its logical representations, reasoning mechanisms, and the incorporation of prior knowledge, ILP holds promise in enhancing the capabilities of machine learning models for few-shot learning tasks.

2 Few-Shot Learning

Few-shot Learning (FSL) is a subfield of machine learning that aims to address the challenge of learning new concepts or tasks with limited labelled data. One of the most widely accepted definitions of FSL is the one provided by Wang et al. in 2020 [30], which defines it in terms of the experience, task, and performance of machine learning [1]. If the program's performance on some classes of tasks T as assessed by some performance measure P improves with the addition of experience E, then we say that the program has learned from its experience. It's important to stress that E is negligibly low in FSL.

To fully understand FSL, it is important to explain the idea of the N-way-K-shot problem. The N-way-K-shot problem is a way to describe different problems in FSL. In this problem setting, the support set is made up of a small set of data that is used for training and then used as a reference for the testing step. Most of the time, the number of categories (N) and the number of samples per category (K) in the reference set are used to describe the N-way-K-shot problem. So, the whole job comprises only $N \times K$ samples. For example, N-way-1-shot is a type of one-shot learning in which the reference set has N categories, but each category only has one sample. Different taxonomies of FSL models has been listed below;

1. Active Learning [8,19]
2. Transfer Learning [31,32]
3. Meta-Learning [21,32]
4. ILP Methods [5,26,27,29]

While AI approaches have made significant progress, they still struggle to generalise quickly from limited samples. Successful AI applications often rely on learning from large datasets. In contrast, individuals can quickly learn new activities by using their past experiences and expertise. For computers to match human skills, they must solve the FSL problem. Similar to human learning, computer programs can learn from supervised instances and pre-trained notions like parts and relationships. Another significant scenario where FSL plays a vital role is when acquiring examples with supervised information becomes challenging or even impossible due to concerns regarding privacy, safety, or ethical considerations.

Furthermore, FSL offers the advantage of reducing the data-gathering effort necessary for data-intensive applications. By leveraging only a few labelled examples, FSL techniques can effectively generalise and make accurate predictions,

even with limited data. This ability to learn from a small number of examples helps alleviate the burden of collecting and annotating massive amounts of data, making FSL an efficient and practical approach for data-intensive tasks.

3 Feature Extraction for Neurodegenerative Disease Detection from Retinal Images

The human retina, an extension of the central nervous system, provides valuable insights into various neurological conditions [13]. Retinal images capture intricate details, such as vessel abnormalities, optic nerve changes, and retinal layer thickness alterations as potential biomarkers for neurological disorders. By analysing retinal images, medical professionals can gain valuable insights into conditions like diabetic retinopathy, glaucoma, multiple sclerosis, and even neurodegenerative diseases like Alzheimer's and Parkinson's [3,15,24].

In the field of medical diagnostics, accurate and timely identification of neurological conditions is crucial for effective treatment and management. Retinal imaging has emerged as a valuable tool in this endeavour, offering a non-invasive and accessible means of examining the intricate structures within the eye. To further enhance diagnostic capabilities, the application of few-shot learning techniques to retinal images has gained traction, enabling efficient and accurate identification of neurological disorders even with limited labelled data. Retinal imaging and few-shot learning present a powerful approach to improving neurological diagnosis. Using few-shot learning algorithms can effectively recognise distinct patterns associated with different neurological conditions, even with limited labelled data.

3.1 Feature Extraction from Retinal Images

Retinal images have become a valuable source of information for various medical applications, including disease diagnosis, monitoring, and treatment. Extracting informative features from retinal images is a crucial step in leveraging the potential of these images for accurate and efficient analysis. In this study, we delve into the realm of feature extraction from retinal images, exploring the advantages and applications of both handcrafted and learned features.

Learned Features. The use of deep learning and convolutional neural networks (CNNs) has led to a move towards learning features directly from data. CNN architectures use hierarchical feature extraction to learn representations, often known as deep features or embeddings. The abstract characteristics depict subtle patterns and variances in retinal pictures. CNN models trained on big datasets can automatically learn discriminative features suited for certain tasks. Learned features excel in retinal image processing tasks such as illness categorization, lesion identification, and picture segmentation. Adaptable networks can extract task-specific characteristics, revealing subtle patterns that humans may miss.

Fig. 1. Demonstration of processing steps for vessel segmentation and artery *vs.* vein classification

We mainly use learned feature techniques for optic-disc localisation and artery/vein classification using Haar-discrete wavelet transform [12] and a pre-trained CNN model [7]. Optic disc localisation plays a crucial role in automated retinal image analysis, as it serves as a vital landmark for various diagnostic tasks. The Haar wavelet is a simple, orthogonal wavelet transform that captures variations in an image at different scales. It decomposes the image into low-frequency (approximation) and high-frequency (detail) components. On the other hand, discrete wavelet transform (DWT) extends the Haar wavelet concept to more complex wavelet functions, enabling more sophisticated analysis of image features.

Artery/vein classification in retinal images plays a vital role in understanding the vascular structure and dynamics of the human eye. Figure 1 shows the processing steps for vessel segmentation and artery/vein classification. Accurate identification and differentiation of arteries and veins provide valuable insights into various ocular and systemic diseases. In recent years, deep learning techniques, particularly CNNs, have emerged as a powerful approach for automated artery/vein classification. CNN models have revolutionised artery/vein classification in retinal image analysis, providing a robust and efficient approach for the automated identification of vascular structures.

Handcrafted Features. Handcrafted features are designed to capture individual retinal traits or patterns. The features are manually designed using domain expertise and expert insights. Early retinal image analysis has commonly utilised handcrafted characteristics, which have proven beneficial in several applications. Handcrafted traits include vessel width, curvature and tortuosity. The benefits of handcrafted features are their interpretability and explicit representation of domain-specific knowledge. Capturing complicated and subtle retinal image alterations is limited by these methods. The study used handcrafted retinal vascular characteristics, as shown in Table 1, and distinct retinal zones were analysed in Fig. 2.

The following summary describes the calculations and measurements involved:

– **Vascular Calibres:** The calibres of the six most extensive arterioles and six largest venules were calculated. These measurements represent the width of the vessels.

Table 1. Retinal Vascular Features (RVFs) with the retinal zone of interest

Features	Description	Retinal Zone
CRAE	Central Retinal Arteriolar Equivalent	B
CRVE	Central Retinal Venular Equivalent	B
AVR	Arteriole-Venular ratio	B
FDa	Fractal Dimension arteriole	C
FDv	Fractal Dimension venular	C
BSTDa	Zone B Standard Deviation arteriole	B
BSTDv	Zone B Standard Deviation venular	B
TORTa	Tortuosity arteriole	C
TORTv	Tortuosity venular	C

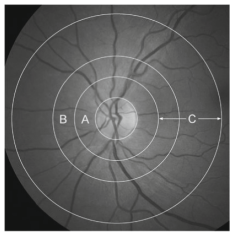

Fig. 2. Retinal zones considered in this study [6].

- **Standard Deviation of Width in Zone B (BSTD)**: The standard deviation of the vessel width was calculated for both the arteriolar and venular networks within zone B. This measurement quantifies the variation in vessel width within the specified zone.
- **Vascular Equivalent Calibre:** Summary measures of vascular equivalent calibre were computed using an improved version of the Knudston-Parr-Hubbard formula [10,11]. This formula provides estimates of the equivalent single-vessel parent calibre (width) for the six arterioles (CRAE) and six venules (CRVE).
- **Arteriole-to-Venule Ratio (AVR):** The arteriole-to-venule ratio (AVR) was calculated by dividing the CRAE (arteriolar equivalent calibre) by the CRVE (venular equivalent calibre). This ratio provides insight into the relative size differences between arterioles and venules.
- **Fractal Dimension (FD):** The fractal dimension of the retinal vascular network was determined using the box-counting method [14]. The fractal dimension describes the self-similarity or branching pattern of the vascular network across different scales. Higher values indicate a more complex branching pattern.
- **Retinal Vascular Tortuosity:** Vascular tortuosity refers to the curvature and bending of blood vessels. In this study, the retinal vascular tortuosity was quantified by calculating the integral of the curvature squared along the vessel path, normalized by the total path length [9]. The tortuosity values

were averaged across the measured vessels. Smaller tortuosity values indicate straighter vessels.

By extracting and analysing these retinal vascular features, valuable information can be obtained regarding vessel calibres, variations, equivalent calibres, arteriole-to-venule ratio, fractal dimension, and tortuosity. These features provide insights into the structural characteristics of the retinal vascular network and can be utilised in various medical and research applications related to retinal vascular analysis and disease diagnosis.

4 Histogram-Based Binning Method

Inductive Logic Programming (ILP) is a powerful framework that combines logic programming and machine learning techniques to learn hypotheses from examples. ILP traditionally operates on discrete and symbolic data, relying on logical representations and rules. It excels at capturing patterns and relationships in categorical or discrete domains, making it well-suited for symbolic reasoning tasks. However, the inherent nature of ILP poses obstacles when it comes to handling continuous data. There are certain challenges when we try to include numerical data in the context of ILP:

- **Representation:** ILP traditionally operates on discrete and symbolic data, which requires a conversion process to represent numerical data appropriately. Representing continuous values as discrete symbols may lead to loss of information and introduce discretisation errors.
- **Expressiveness:** Logic programming languages typically lack built-in support for numerical operations and comparisons. This limitation hampers the direct handling of numerical data and restricts the expressive power of ILP models.
- **Scalability:** Numerical data often introduces increased computational complexity due to continuous value ranges and arithmetic computations. This can significantly impact the scalability of ILP algorithms and hinder their efficiency.
- **Sensitivity to Scaling:** ILP algorithms can be sensitive to the scaling of numerical features. Differences in the magnitude or range of numerical values can significantly impact ILP's ability to extract meaningful patterns or relationships. Inconsistent scaling across features may lead to biased or misleading results.

We introduce a histogram-based binning method for numerical or continuous data to address the above-mentioned issues.

Histograms are graphical representations that illustrate the distribution of continuous data. They are highly valuable for exploratory analysis as they unveil insights about datasets that cannot be captured solely through summary statistics. Histograms visually depict the data's shape, spread, and central tendencies. By organising the data into bins or intervals along the x-axis and representing the

frequency or count of observations in each bin on the y-axis, histograms enable us to discern patterns, identify outliers, and understand the overall distribution of the data.

The advantages of histograms lie in their ability to showcase the underlying characteristics of sample data. Unlike summary statistics such as mean or standard deviation, histograms reveal the specific values and frequencies within each interval, allowing us to grasp the range and concentration of data points at different levels. This level of detail aids in understanding the skewness, kurtosis, multimodality, or presence of gaps in the data distribution, which may not be apparent from mere summary statistics. Histograms serve as a powerful exploratory tool, providing a comprehensive overview of the data and highlighting features such as clusters, peaks, or outliers that might influence subsequent analysis.

Our proposed binning method for ILP takes advantage of histograms' inherent flexibility and interpretability, allowing for accurate representation of data distributions while preserving relevant statistical properties. The key principle of our binning method is to dynamically determine optimal bin widths based on the characteristics of the dataset. By employing advanced statistical techniques, such as kernel density estimation or adaptive binning algorithms, we ensure that the resulting histograms capture the underlying structure of the data with greater precision. This approach mitigates issues related to subjectivity and arbitrary bin widths choices while maintaining the original data's integrity.

Now we define the notions used in the histogram-based binning method.

Definition 1. ***Number of Bins*** **(*k*).** *The number of Bins, denoted as k represents the desired number of equally spaced bins to divide the data range into.*

The number of bins determines the level of granularity in the histogram representation.

Definition 2. ***Width of Bin*** **(*w*).** *The width of each bin, denoted as w, represents the size of the interval for which the occurrences are counted. It determines the level of granularity in the histogram representation.*

Let R be the data range then bin width is calculated by

$$w = \frac{max(R) - min(R)}{k} \tag{1}$$

Definition 3. ***Bin Edges*** **(B_k).** *The bin edges, denoted as* $B_k = [b_0, b_1, \cdots, b_k]$*, represent the boundaries of the bins used in the histogram. The bin edges can be calculated as* $B_k = \{min(R) + i \times w : i \in \{0, 1, 2, \ldots, k\}\}$.

5 Empirical Evaluation

In this section we evaluate the effectiveness of the ILP systems, and the binning method described in the previous section, in generating interpretable and accurate rules for detecting neurodegenerative diseases such as Alzheimer's, Parkinson's, and vascular dementia from retinal images. We compare different ILP

approaches with a range of statistical machine learning and neural network models. We also demonstrate that by leveraging the binning method, the ILP methods can capture meaningful patterns and relationships within the retinal images, enabling the development of more accurate and explainable diagnostic rules. The data, codes and configuration files used in the experiments in this paper are available from https://github.com/hmlr-lab/FSL_Fundus_Images.

5.1 Materials

The data used for this study is extracted from the UK Biobank resources [23]. The UK Biobank is a large-scale project that recruited 500,000 individuals between the ages of 40 and 69 to undergo various tests and have their health monitored over their lifetimes. It is worth noting that only a subset of these participants, specifically 84,767 individuals, had their retinas imaged as part of the study. Retinal imaging was performed using the TOPCON 3D OCT 1000 Mk2 device, which combines optical coherence tomography (OCT) with fundus photography. The imaging procedure focused on capturing images of the macula, the central region of the retina. The resulting images have a 45-degree field of view and dimensions of 2,048 by 1,536 pixels. The information regarding the participants in this study was collected and organised in a large CSV (Comma Separated Values) file. Each row in the CSV file represents a participant, while each column represents a specific data point. The UK Biobank online system provides detailed explanations for the codes used in the column names and the associated data, ensuring transparency and clarity in the dataset. In terms of diagnoses, the dataset follows the International Classification of Diseases, Tenth Revision (ICD-10) coding system.

Through a comprehensive analysis of the participant data file, we identified a specific subset of individuals who satisfied two conditions: (1) they had fundus images captured, and (2) they were diagnosed with one of three conditions: Alzheimer's disease, Parkinson's disease, or vascular dementia. Within this subset, we found 18 cases of Alzheimer's, 133 cases of Parkinson's, and 54 cases of vascular dementia. In addition to the fundus images from these individuals with neurodegenerative conditions, we included images from 528 participants who were confirmed to be healthy concerning these three conditions. It is important to note that only fundus images of the left eye were used in this study, ensuring consistency in the dataset and simplifying the analysis process. We extracted artery/vein information using learned features and then derived handcrafted features from this information. Later, the structured data in the form of CSV was converted into 100 different bins and encoded into logical rules.

5.2 Methods

This section outlines the methodology employed to conduct our study on few-shot learning. We utilised a dataset consisting of images from four distinct classes, each containing 18 images. The dataset was divided into training and test data, with a split ratio of 6:4. To perform the N-way-K-shot learning, we

Table 2. Machine learning algorithms used in this study

ILP models	Other learning models from Scikit-learn [20]
1) Meta Inverse Entailment (MIE) - PyGol [25,28]	1) Decision Tree (DT)
2) Meta-Interpretive Learning (MIL) - Metagol$_{NT}$ [4,18]	2) Naive Bayes (NB)
3) One-Shot Hypothesis Derivation (OSHD) - TopLog [26]	3) Linear Discriminant Analysis (LDA)
4) Inverse Entailment (IE) - Aleph [22]	4) Support Vector Machine (SVM)
	5) Logistic Regression (LR)
	6) Random Forest (RF)
	7) Perceptron (Per)
	8) Multilayer Perceptron (MLP)
	9) K Nearest Neighbors (KNN)

employed various learning models from different domains, including Inductive Logic Programming (ILP), statistical machine learning, and neural network models. Specifically, we utilised 13 different learning models to compare their performance in the context of our study. These include 4 ILP models (IE [17,22], MIL [4,18], OSHD [26] and MIE [25,28]) and 9 non-ILP models from Scikit-learn as listed in Table 2.

Language Bias. Next, we describe the methodology used to analyse the background knowledge and generate mode declarations and metarules for ILP systems Aleph, TopLog, and Metagol. These ILP systems heavily rely on user-defined mode declarations and metarules, which are crucial in guiding the learning process. It is important to note that the manual generation of mode declarations and metarules is a user-intensive and highly domain-specific task. The mode declarations and metarules used in the experiment are listed in Table 3.

To begin, we carefully examined the background knowledge available for our study. This process involved a comprehensive review of domain-specific information, including the relationship between predicates, as well as the potential types of hypothesis structures that could be learned from the available data. First, we focused on identifying the relationship between predicates within the problem domain. We examined how different predicates could be combined to form meaningful rules and how these rules could be interconnected to represent the underlying knowledge in a logical manner. Simultaneously, we explored the potential hypothesis structures that could be learned from the available data. We considered the possible combinations and arrangements of predicates to form hypotheses that accurately represented the underlying patterns and relationships within the data.

In our experiment, we also include the novel ILP system PyGol, and analyse its ability to learn language biases without relying on user-defined mode declarations. PyGol offers a promising approach by automating the process of learning language biases, reducing the need for extensive user interaction and manual input. By excluding user-defined mode declarations, PyGol aimed to automatically learn the language biases solely from the available data. This approach

Table 3. Language biases used in ILP models IE (Aleph), OSHD (TopLog) and MIL (Metagol$_{NT}$)

Mode Declarations	Metarules
modeh(1, diagnosis(A, alzheimers)(+image))	P(A) :- Q(A,B), R(B,C)
modeb(1, crae(+image, -group))	P(A,B) :- Q(A,C), R(A)
modeb(1, crve(+image, -group))	P(A) :-Q(B,A), R(A,C)
modeb(1, avr(+image, -group))	P(A) :-Q(A,B), R(B,C)
modeb(1, bstda(+image, -group))	P(A) :-Q(A,B), R(B)
modeb(1, bstdv(+image, -group))	P(A) :-Q(A,B), R(A,C)
modeb(1, fda(+image, -group))	
modeb(1, fdv(+image, -group))	
modeb(1, torta(+image, -group))	
modeb(1, tortv(+image, -group))	
modeb(*, lteq(+group, #float))	
modeb(*, gteq(+group, #float))	

enabled us to assess PyGol's ability to capture and represent the inherent biases and patterns present in the dataset without any additional user intervention.

In the experimental methodology, the N-way-K-shot algorithm described in Sect. 2 was utilised. The value of K was varied, specifically set to 2, 4, 6, 8, and 10 to represent the number of training-relevant positive examples. Concurrently, a fixed number of five negative examples from the other three classes were chosen. For example, if 2 positive instances were selected from the Alzheimer's class, 5 negative instances were selected from the vascular dementia, Parkinson's, and healthy data sets combined. In addition, each experimental episode included *twenty iterations* (N). We imposed a maximum length restriction on hypotheses of five literals, allowing for a maximum of four conditions in the body of each hypothesis. In addition, during the testing phase, an equal number of positive and negative examples were chosen to sustain a 50% of default accuracy. It is important to note that the same instance was used for both training and testing in all assessed models.

5.3 Results and Discussions

In this section, we present the results of our experiments on Alzheimer's, dementia, and Parkinson's diseases using various models. We compare the performance of 13 different models, focusing on the number of positive examples used for training, which ranges from 2 to 10. The results are visualised in three separate graphs (Fig. 3) corresponding to each disease. From the results obtained in our experiments, it becomes evident that the ILP models, specifically PyGol and OSHD, exhibit superior performance compared to the other models. As depicted in the accuracy analysis, the ILP models consistently outperform the alternative models as the number of positive examples increases from 2 to 10. Figure 4 shows example rules learned using PyGol.

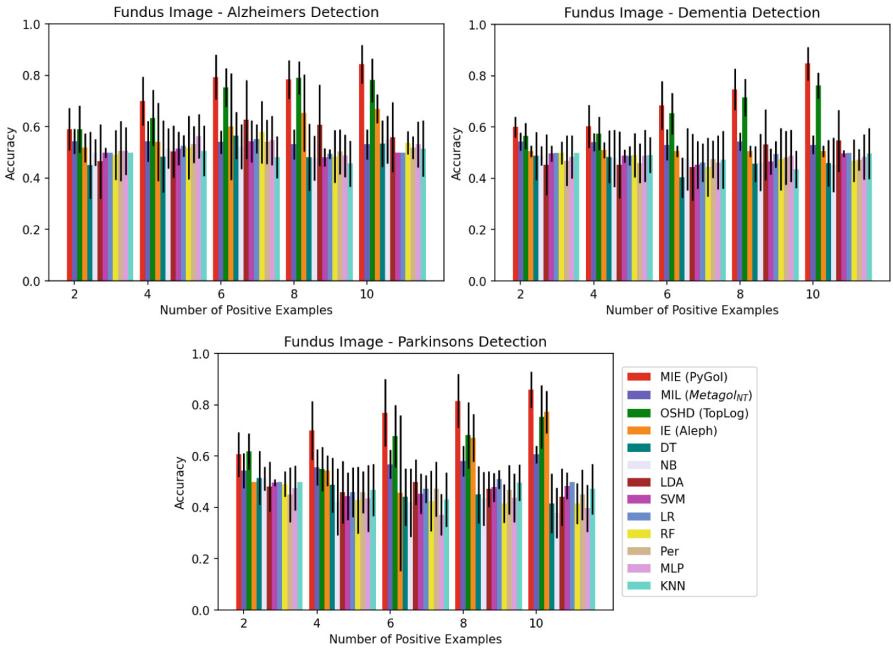

Fig. 3. Comparing the performance of various ILP (MIE, MIL, OSHD, IE) and non-ILP algorithms for learning diagnostic rules for Alzheimer's, Parkinson's, and vascular dementia

Among the ILP models evaluated, both PyGol and OSHD (TopLog) consistently demonstrate improvement in accuracy across all three diseases as the number of positive examples increases. However, Aleph did not exhibit a strong performance, specifically in the experiment related to vascular dementia detection. Notably, PyGol emerges as the frontrunner with the highest accuracy among all the models, showcasing its exceptional ability to learn from a limited number of positive examples effectively. These results highlight the remarkable effectiveness of ILP models, particularly PyGol, in addressing the challenges associated with learning from a small set of positive examples, thereby underscoring their potential for accurate disease detection.

The MIL model did not demonstrate significant performance across the experiments. This could be attributed to a couple of reasons. Firstly, our dataset containing continuous values may have introduced noise, making it challenging for MIL to find generalised hypotheses. Secondly, it is possible that the metarules we utilised in the MIL model did not provide sufficient expressive power to generate effective hypotheses. The limitations of the metarules may have constrained the model's ability to capture the complex patterns and relationships present in the data.

Furthermore, it is evident from our results that both statistical and neural network models struggle to learn efficiently from a small number of examples.

```
diagnosis(A, alzheimers) :- crve(A, B),
                            gteq(B, 0.0),
                            tortv(A, C),
                            gteq(C, 18477.29).
diagnosis(A, alzheimers) :- torta(A, B),
                            gteq(B, 26396.13),
                            lteq(B, 27715.93).
```

Fig. 4. Sample diagnostic rules learned by PyGol for Alzheimer's disease

As the number of positive examples increases, the performance of these models does improve to some extent, but they generally fall short compared to the ILP models, particularly PyGol and OSHD (TopLog). This limitation can be attributed to the inherent complexity and flexibility of statistical and neural network models, which typically require a larger amount of data to capture the underlying patterns and relationships effectively.

The rules shown in Fig. 4 exhibit higher interpretability and accuracy than those obtained in our previous work [26], where the range of the numerical values were explicitly fixed. The binning mechanism, which effectively converts the continuous data, is crucial in improving the interpretability of these rules. By converting the numerical data into bins, the ILP models can capture the underlying patterns more effectively. Moreover, the binning approach contributes to the improved accuracy of the ILP models.

6 Conclusions

The experiments have provided valuable insights into the performance of various models for few-shot learning for neurogenerative disease detection. The results demonstrate the effectiveness of ILP models, particularly PyGol and OSHD, in learning from a small number of positive examples. These ILP models consistently outperformed statistical and neural network models, showcasing their ability to address the challenges of few-shot learning. Additionally, the histogram-based binning approach proved to be a valuable technique for enhancing the interpretability and accuracy of ILP models. By discretising the continuous data, the binning mechanism enabled the ILP models to capture meaningful thresholds and ranges, leading to more interpretable rules. The binning approach also contributed to improved accuracy by effectively capturing important features and patterns in the data. The histogram-based binning mechanism offers a practical solution for enhancing the interpretability and accuracy of ILP models.

In conclusion, our study demonstrates the efficacy of ILP models, particularly PyGol, in addressing the challenge of disease detection from limited image data. The utilisation of ILP models, coupled with the histogram-based binning mechanism, provides a powerful and promising approach for accurate and interpretable disease detection. Also, our study highlights the potential of PyGol in leveraging limited training data for accurate and interpretable disease detection. Applying

the histogram-based binning mechanism further enhances the performance of ILP models, paving the way for advancements in similar disease detection from small data using ILP.

Acknowledgments. The first author would like to acknowledge the Vice Chancellor's PhD Scholarship Award at the University of Surrey. The third author would like to acknowledge the EPSRC Network Plus grant on Human-Like Computing (HLC) and the EPSRC grant on human-machine learning of ambiguities. This research has been conducted using the UK Biobank Resource (Application No 1969).

References

1. Machine Learning. McGraw Hill, New York (1997)
2. Chen, W.Y., Liu, Y.C., Kira, Z., Wang, Y.C., Huang, J.B.: A closer look at few-shot classification. In: International Conference on Learning Representations (2019)
3. Cheung, C.Y.l., Ikram, M.K., Chen, C., Wong, T.Y.: Imaging retina to study dementia and stroke. Prog. Retin. Eye Res. **57**, 89–107 (2017)
4. Cropper, A., Muggleton, S.H.: Metagol system (2016). https://github.com/metagol/metagol
5. Dai, W.Z., Muggleton, S., Wen, J., Tamaddoni-Nezhad, A., Zhou, Z.H.: Logical vision: one-shot meta-interpretive learning from real images. In: ILP (2017)
6. Frost, S., Kanagasingam, Y., Sohrabi, H., Vignarajan, J., Bourgeat, P., et al.: Retinal vascular biomarkers for early detection and monitoring of Alzheimer's disease. Transl. Psychiatry **3**, e233 (2013)
7. Galdran, A., Meyer, M., Costa, P., MendonÇa, Campilho, A.: Uncertainty-aware artery/vein classification on retinal images. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 556–560 (2019)
8. Garcia, V., Bruna, J.: Few-shot learning with graph neural networks. arXiv preprint arXiv:1711.04043 (2017)
9. Hart, W.E., Goldbaum, M., Côté, B., Kube, P., Nelson, M.R.: Measurement and classification of retinal vascular tortuosity. Int. J. Med. Inform. **53**(2), 239–252 (1999)
10. Hubbard, L.D., Brothers, R.J., King, W.N., et al.: Methods for evaluation of retinal microvascular abnormalities associated with hypertension/sclerosis in the atherosclerosis risk in communities study. Ophthalmology **106**(12), 2269–2280 (1999)
11. Knudtson, M., Lee, K.E., Hubbard, L., Wong, T., et al.: Revised formulas for summarizing retinal vessel diameters. Curr. Eye Res. **27**, 143–149 (2003)
12. Lalonde, M., Beaulieu, M., Gagnon, L.: Fast and robust optic disc detection using pyramidal decomposition and Hausdorff-based template matching. IEEE Trans. Med. Imaging **20**, 1193–200 (2001)
13. London, A., Benhar, I., Schwartz, M.: The retina as a window to the brain - from eye research to CNS disorders. Nat. Rev. Neurol. **9** (2012)
14. Mainster, M.: The fractal properties of retinal vessels: Embryological and clinical implications. Eye **4**, 235–241 (1990)
15. McGrory, S., Taylor, A.M., Kirin, et al.: Retinal microvascular network geometry and cognitive abilities in community-dwelling older people: the Lothian birth cohort 1936 study. Ophthalmology **101**(7), 993–998 (2017)
16. Muggleton, S.: Inductive logic programming. ACM **5**, 5–11 (1994)

17. Muggleton, S.: Inverse entailment and progol. New Gener. Comput. **13**, 245–286 (1995). https://doi.org/10.1007/BF03037227
18. Muggleton, S., Lin, D., Tamaddoni, N.A.: Meta-interpretive learning of higher-order dyadic datalog: predicate invention revisited. MLJ **100**, 49–73 (2015)
19. Müller, T., Pérez-Torró, G., Basile, A., Franco-Salvador, M.: Active few-shot learning with FASL. arXiv preprint arXiv:2204.09347 (2022)
20. Pedregosa, F., Varoquaux, G., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
21. Ren, M., et al.: Meta-learning for semi-supervised few-shot classification. arXiv preprint arXiv:1803.00676 (2018)
22. Srinivasan, A.: A learning engine for proposing hypotheses (aleph) (2001). https://www.cs.ox.ac.uk/activities/programinduction/Aleph/aleph.html
23. Sudlow, C., et al.: UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLOS Med. (2015)
24. Tian, J., Smith, G., Guo, H., Liu, B., Pan, Z., et al.: Modular machine learning for Alzheimer's disease classification from retinal vasculature. Sci. Rep. **11**(1) (2021)
25. Varghese, D., Barroso-Bergada, D., Bohan, D.A., Tamaddoni-Nezhad, A.: Efficient abductive learning of microbial interactions using meta inverse entailment. In Proceedings of the 31st International Conference on ILP (2022, in press)
26. Varghese, D., Bauer, R., Baxter-Beard, D., Muggleton, S., Tamaddoni-Nezhad, A.: Human-like rule learning from images using one-shot hypothesis derivation. In: Katzouris, N., Artikis, A. (eds.) ILP 2021. LNCS, vol. 13191, pp. 234–250. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-97454-1_17
27. Varghese, D., Patel, U., Krause, P., Tamaddoni-Nezhad, A.: Few-shot learning for plant disease classification using ILP. In: Garg, D., Narayana, V.A., Suganthan, P.N., Anguera, J., Koppula, V.K., Gupta, S.K. (eds.) IACC 2022. Communications in Computer and Information Science, vol. 1781, pp. 321–336. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-35641-4_26
28. Varghese, D., Tamaddoni-Nezhad, A.: PyGol. https://github.com/PyGol/
29. Varghese, D., Tamaddoni-Nezhad, A.: One-shot rule learning for challenging character recognition. In: Proceedings of the 14th Interantional Rule Challenge, pp. 10–27, August 2020
30. Wang, Y., Yao, Q., Kwok, J.T., Ni, L.M.: Generalizing from a few examples: a survey on few-shot learning, **53**(3) (2020)
31. Yu, Z., Chen, L., Cheng, Z., Luo, J.: TransMatch: a transfer-learning scheme for semi-supervised few-shot learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12856–12864 (2020)
32. Zhu, P., Zhu, Z., Wang, Y., Zhang, J., Zhao, S.: Multi-granularity episodic contrastive learning for few-shot learning. Pattern Recogn. 108820 (2022)