Mark E. Whiting
Department of Mechanical Engineering,
Carnegie Mellon University,
Pittsburgh, PA 15213
e-mail: mwhiting@alumni.cmu.edu

Joseph Mettenburg
Department of Radiology,
University of Pittsburgh,
Pittsburgh, PA 15213
e-mail: mettenburgjm@upmc.edu

Enrico M. Novelli
Department of Medicine,
University of Pittsburgh,
Pittsburgh, PA 15213
e-mail: noveex@upmc.edu

Tales Santini
Department of Bioengineering,
University of Pittsburgh,
Pittsburgh, PA 15260
e-mail: santini.tales@pitt.edu

Tiago Martins
Department of Bioengineering,
University of Pittsburgh,
Pittsburgh, PA 15260
e-mail: t.martins@pitt.edu

Tamer S. Ibrahim
Department of Bioengineering,
University of Pittsburgh,
Pittsburgh, PA 15260
e-mail: tibrahim@pitt.edu

Philip R. LeDuc
Fellow ASME
Department of Mechanical Engineering,
Carnegie Mellon University,
Pittsburgh, PA 15213
e-mail: prl@andrew.cmu.edu

Jonathan Cagan
Fellow ASME
Department of Mechanical Engineering,
Carnegie Mellon University,
Pittsburgh, PA 15213
e-mail: cagan@cmu.edu

Inducing Vascular Grammars for Anomaly Classification in Brain Angiograms

As machine learning is used to make strides in medical diagnostics, few methods provide heuristics from which human doctors can learn directly. This work introduces a method for leveraging human observable structures, such as macroscale vascular formations, for producing assessments of medical conditions with relatively few training cases, and uncovering patterns that are potential diagnostic aids. The approach draws on shape grammars, a rule-based technique, pioneered in design and architecture, and accelerated through a recursive subgraph mining algorithm. The distribution of rule instances in the data from which they are induced is then used as an intermediary representation enabling common classification and anomaly detection approaches to identify indicative rules with relatively small data sets. The method is applied to seven-tesla time-of-flight angiography MRI (n = 54) of human brain vasculature. The data were segmented and induced to generate representative grammar rules. Ensembles of rules were isolated to implicate vascular conditions reliably. This application demonstrates the power of automated structured intermediary representations for assessing nuanced biological form relationships, and the strength of shape grammars, in particular for identifying indicative patterns in complex vascular networks. [DOI: 10.1115/1.4053424]
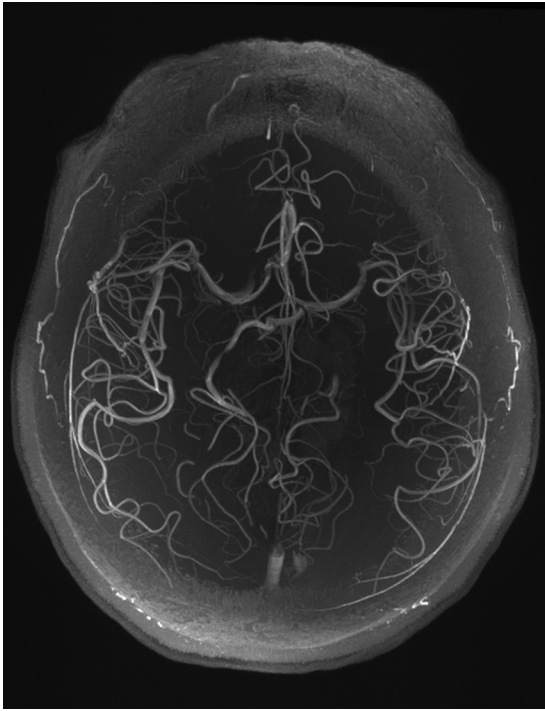
Introduction

Radiological angiograms of the brain (Fig. 1) enable the diagnosis of conditions by observing the detailed features of the vascular network. These features, including vascular tortuosity and angiogenesis have been associated with malignancy [1], and metastasis [2] in sarcomas and other types of tumors. Additionally, vascular anomalies in the brain are a telling indicator of a variety of health outcomes in patients [3]. While some diseases can be identified through direct observation of certain feature such as tortuosity, others are more subtle and, to the physician, are not directly obvious. Whether obvious or not, a computational tool that would identify nuanced variations that indicate a possible diagnosis could provide a significant advancement in the early diagnosis of medical conditions. This paper introduces one such approach using machine learning methods to induce a shape-based grammatical representation.

Machine learning methods used in medical diagnostics often require large data sets—outperforming dermatologists in identifying cancerous skin lesions taken of the order of millions of samples [4]. Yet, datasets of this magnitude are only available in

Fig. 1 Time-of-flight angiogram of the vascular system in the brain

some situations and are particularly scarce for medical conditions that are not traditionally diagnosed radiologically. In addition, angiography is more often used as a procedural aid than as a diagnostic indicator so the volume of shared clinical data is small compared to the corpus available for soft tissue tumors. Yet, vascular tortuosity and angiogenesis have been associated with malignancy [1], and metastasis in sarcomas and other types of tumors [2]. Additionally, vascular anomalies in the brain are a strong indicator of a variety of health outcomes [3]. However, the lack of data means that few traditional statistical learning methods
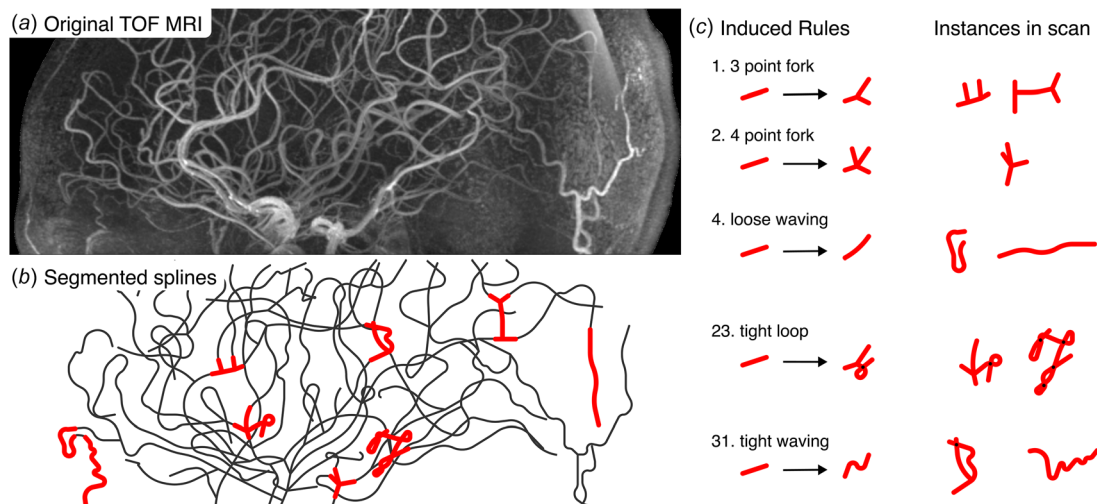
can be applied to improve diagnosis in these situations. This work aims to provide an alternative approach to this situation—drawing insights from smaller data sets through a preprocessing step that turns one angiogram into many hundreds of spatially derived features, affording computational comparison and analysis.

The specific preprocessing step explored in this work builds on shape grammars [5] and graph grammars [6]—techniques initially used in structuring urban planning, architecture, and industrial and mechanical design problems. These approaches facilitate representing complex structural relationships by identifying repeated subgraphs and breaking the data into generalized rules. In contrast to traditional machine learning techniques, grammar methods can require less data and are more interpretable. They closely mirror an intuitive human information structuring approach of associating meaning with visual patterns across multiple scales [7] and produce rules that reflect actual situations in the data (not a statistical summary). Here, automated grammar induction [8], is applied to angiographic data to generate a rich preprocessed representation of the vessel networks in the brain, enabling the use of traditional classification methods despite few individual patients. The method affords to identify differences between individuals by extracting spline networks of vasculature from time-of-flight MRI data, inducing grammars from the spline networks (Fig. 2), and studying the distribution of rules occurrences between individuals. The rule distribution serves as a feature for differentiating indicative factors using traditional statistical learning techniques such as those leveraging bag-of-features models [9].

In this work, leveraging grammars to scaffold automated diagnostics is shown to be a potentially viable technique and one that has a number of desirable properties. For example, because of the intermediary representation approach, this grammar-based process requires significantly less data than many others in use today. Further, because the grammar rules are induced over signals currently used by radiologists for diagnosis, the rules capture information that is meaningful to them and could be used as training aids in the future.
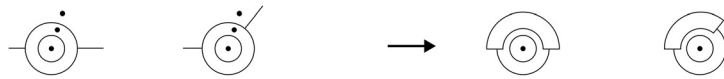
Related Work

This section will address two bodies of related work: radiological analysis, and grammar-based methods. First, the context and

Fig. 2 A depiction of the pipeline from MRI to grammar rules. (*a*) A time-of-flight MRI was shown from the midsagittal plane and rendered with a maximum intensity projection of the full brain. (*b*) Segmentation from the shown MRI aligned at the same orientation with parts of the vessel structure marked indicating distinct areas found in highlighted rules. (*c*) A subset of the induced rules is shown with their index and description as well as instances found in the scan. This pipeline indicates the major processing steps in preparing rules for classification and then finding them in the data from a given scan to establish the bag-of-rules representation.
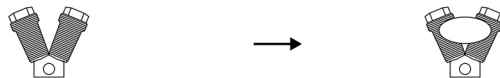
Rule 30. Fenders: Draw fenders & discarde point lables
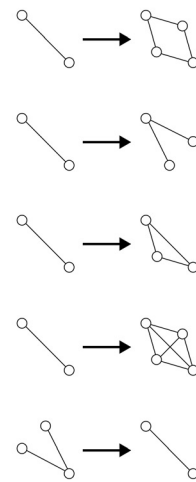
Rule 41. Seat: Create headlight profile
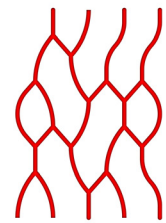
Rule 45. Engine Trim: Add Air Filter

(a)

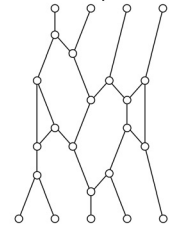Rules   Vessels

Graph

(b)

Fig. 3 Shape grammar can be applied in mechanical and biological contexts because of its common treatment of complexity in many types of systems. (*a*) Selected shape grammar rules for a motorcycle, which can be parameterized to articulate the Harley Davidson brand [17]. (*b*) A collection of grammar rules is used to generate vessel structures in the form of node graphs. In both (*a*) and (*b*), the rules are depicted with a left-hand side (LHS), which can be detected in context, and a right-hand side (RHS) which replaces the LHS providing a specific augmentation. In other words, where one of the subgraphs on the LHS of a rule in (*b*) is found in a graph, it can be replaced by the corresponding RHS subgraphs. To afford variability in the output, some rules have the same LHS, while having different RHS, such as the first four rules depicted here, which use the same subgraph as the LHS, so the same LHS can be converted to different RHS depending on context and rule selection. In (*a*), the augmentations influence overall design shapes, whereas with (*b*), the augmentations adjust graph connections, such as the example shown depicting an abstracted vessel graph.

opportunities of grammar-based techniques will be discussed to provide background and familiarity with the associated challenges to this unconventional computational technique. Then, a brief section addressing how the angiograms that this work uses as a case study are examined currently, and what other computational techniques are being applied in this context.
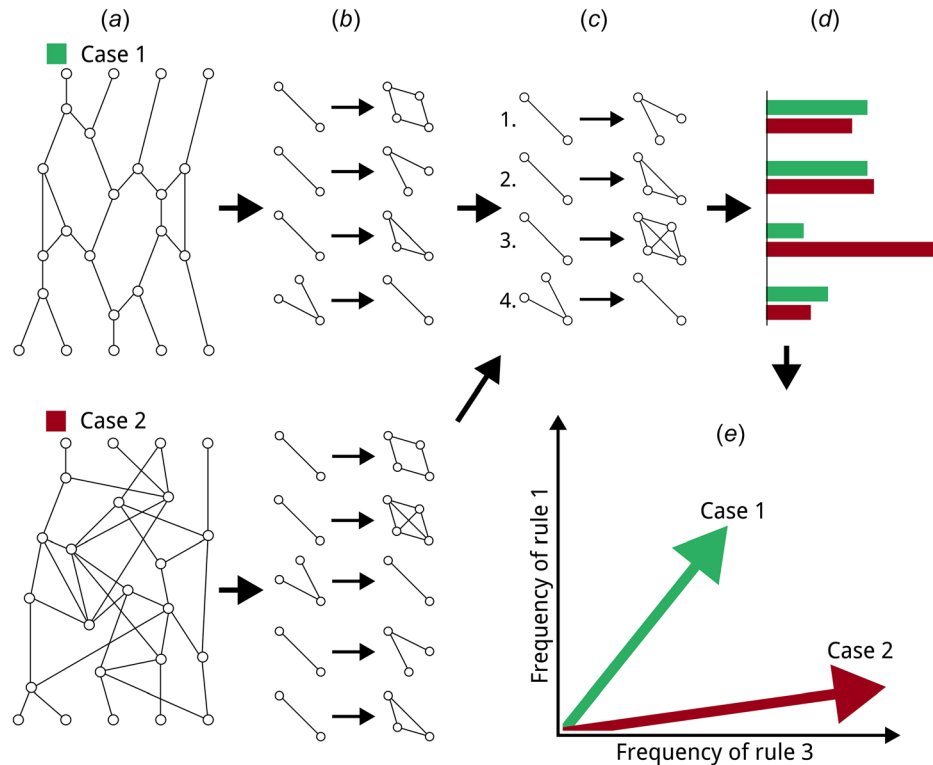
Grammar Methods. Establishing the legitimacy of art or writing, establishing who built what [10], and establishing if a designed product is a brand name or counterfeit [11] are all examples of this overarching need to identify anomalies in semistructured data. Traditional techniques for detecting these differences include decomposing complex systems into their fundamental elements, and classifying instances by identifying features unique to them [12]. Broadly, this is an anomaly detection problem in the field of design, which has motivated the development of many shape and graph grammar-based tools. Conveniently, these methods, which will be discussed in detail next, can be generalized to other kinds of anomaly detection problems, such as those in between distinguishing healthy and problematic medical situations.

Grammar methods use token pairs with left-hand side (LHS) inputs, and right-hand side (RHS) outputs (Fig. 3), where output can replace an input in the data. The process of replacing one component with another compatible component in this way enables systematically changing the complexity of a representation. These structured representations transcend scale and orientation, and even degrees of abstraction when leveraging metarules [13]. Thus, grammars are amenable to generating output within a stylistic language [14,15] and analyzing existing data through comparison of rules [16]; they have been applied in analyzing architecture [5], comprehending the underlying brand principles of modern vehicles (Fig. 3) [17,18], and, in an abstract form, aiding in the design of complex mechanical systems [6,19]. However, despite this tremendous utility, these applications had been only partially automated [20,21] until recent progress in automated grammar

induction [8]. As a consequence, grammar methods have had their primary impact in design alone, and have not been leveraged as an intermediate representation for broad scale computational analysis, or used in domains like medical data analysis, where noncomputational methods would be too slow due to the complexity of the data.

Shape grammars [5] have recently enabled the representation of vascular systems [22] through a three rule grammar that captured distinct features related to the complexities of retinal fundus vascularization. However, this grammar was determined manually by the researchers and recreated the vascular network but did not directly identify irregular conditions that could result in a diagnosis. We build on this work by inducing vasculature rules from examples and associate the resulting rules with different conditions. To accomplish this a far richer grammar is required, one that includes more detailed and nuanced rules.

Classification With Grammars. Classification and analysis based on the frequency of features (e.g., grammar rules), such as support vector machines [23] and bag-of-features [9] models, provide a context for converting a distribution of features into a meaningful signal about underlying data. A key principle of these methods is that particular relationships between features may not need to be considered to uncover a reliable signal [24]. For example, bag-of-words models of text use only the presence and frequency of word use to classify its content, while ignoring word order. A strength of this approach is that it affords analysis without true comprehension of a complex dataset. On the other hand, a weakness is those small datasets can be confounding because word order plays a significant role on the sentence level. Treating grammar representable data as bags-of-rules enables this kind of analysis to be conducted with contextually derived features, i.e., grammar rules. A grammar-based representation also has the strength that it is not simply a bag-of-words, but a bag-of-rules, encompassing decomposed structural information, making analysis with relatively small datasets possible. Additionally, syntactic

Fig. 4 An automated pipeline for detection of grammar differences. (*a*) Input 2 cases of generated graphs with similar extremities and different average degree on internal nodes; (*b*) automatically induce grammars for each case leveraging the grammar induction mechanism from Ref. [8]; (*c*) find a minimal set of representative rules by eliminating duplicates and removing composite rules, for example, the first rule in both induced grammars is removed; (*d*) evaluate the frequency of each rule in the original cases; and (*e*) map rule frequency to values in vectors for each case, generating a vector space of conditions. For this example, only two rules are visualized, rule 3 on the *x* axis and rule 1 on the *y* axis.

pattern recognition has formalized issues of using grammar style representation as a tool for pattern identification [25,26], making it possible to apply a wide range of classification techniques.

Distance metrics use a heuristic to compute the distance in some vector space between cases so as to formalize the cases difference. Initially theorized to establish the degree of difference between biological samples based on genetic feature maps [27], this technique has also seen use in measuring the distances between designs, both analogically [28], and based on structural features. These vector space techniques are also formalized under Hilbert spaces [29]. For these methods to work, and established threshold distance indicates a threshold to be considered problematic. Sensing thresholds in high dimensions is also a field of rich discussion, however in this work only elementary methods are sufficient (e.g., k/-nearest neighbors (KNN) [30], so a more in-depth exploration has not been implemented.

Frequency-based approaches rely on detecting differences in frequency distributions of particular features in a sample [31]. Methods utilizing this type of detection have been a center point in outlier detection in the data mining and machine learning communities [12]. In particular, techniques such as frequent subgraph mining [32], typified by gSpan [33] and Auto-Part [34], have been used with great success to find graph anomalies and outliers. Further, techniques leveraging statistical learning, such as support vector-based approaches [35] and principal component analysis-based approaches [36,37] have proven effective when dimensional reduction is apposite.

Automatic Grammar Methods. As with Yeh et al.'s vascular grammar [22], shape grammars have been used to provide a

classification for product designs in a predominantly manual pipeline [20]. This generally involves first building an overarching grammar, then establishing if the grammar can be manipulated to represent a challenging case. Due to the manual nature of this process, human perception of rules is a potential source of inaccuracy, but additionally, a large amount of time it takes to conduct such a procedure makes comprehensive analysis impossible. As a consequence, statistical shape base analysis of designs [38,39] has been leveraged as an aid in generating concepts, but this approach requires sufficient data for statistical analysis using principle component analysis.

Grammar induction has been automated for a range of types of data in the computational design and machine translation literature. A distinguishing factor of approaches within these domains is that they generally assume coded knowledge of the roles and relationships represented in the data being induced. For example, using a precoded lexicon to produce lingual grammars [40]. On the other hand, Sequitur [41] can interpret lingual data with no added information, constructing character-level grammars. In design, precoded visual design data, such as semantically coded parts of a website, have been used in automatically inducing Bayesian grammars of website design [21], while other approaches to statistically deconstructing visual information without semantic coding were also explored [42]. Statistical shape grammar techniques have also been applied in the analysis of automotive design around applications like characterizing classes of vehicles based on design patterns [39]. An automated, nonstatistical shape grammar induction technique for uncoded design and graph data has also been introduced [8], allowing inducing grammars for almost any kind of structured data with a graph

Table 1   Automated grammar induction algorithm

Steps	
1	Initialize graph into list of remaining tokens
2	while remaining tokens exist do
3	chose a random token
4	create subgraph of proximal neighboring tokens
5	if subgraph is unique in graph then
6	ignore subgraph
7	store token as rule
8	else
9	store subgraph as new token
10	replace token in graph as a new token
11	end if
12	end while

representation. This final technique serves as a basis for the rule frequency analysis pipeline that we utilize.

Rule Equitability. Frequency has served as a foundational indicator in information processing techniques (e.g., using a Fourier transform for spectroscopic classification of biological or chemical elements [43]). However, to facilitate measures of frequency, equitability must be assessable over the elements for comparison. In other words, if rules cannot be differentiated, and equated, then frequency for rules between cases (e.g., designs or medical conditions) being compared cannot be derived.

Equating rules is nuanced because in many situations rules can be used to make other rules that are already within the grammar. To address this challenge, isomorphism techniques are required for identifying and reducing rules that are hard to otherwise compare. Markov equivalence classes [44] provide a mechanism for the formal identification of unique subgraphs by establishing an essential graph that embodies the core ontological relationship of a particular subgraph component. This approach, though not traditionally used in this way, is useful in identifying rule similarities because rules can be treated as subgraphs. Similarly, sets of rules can be identified as a combination of subgraphs. When a rule and a set of rules have the same essential graph, they can conduct the same ontological function as rules in the grammar.

Our work uses the frequency of rules, established through automated grammar induction, to facilitate the classification of anomalous vascular subgraphs with minimal datasets using data from the brain.

Methods

This work introduces an automated approach to classification based on grammar rule frequencies and applies it to vascular graphs in brains as an analytical tool. This section has two main parts: first defining the main method presented by the work, and second presenting the dataset and preparation techniques used to convert angiograms into usable data.

Computational Pipeline. This work leverages shape grammar induction [8], rule deduplication using Markov equivalence classes [44], multiscale rule frequency checking, and case representation based on Hilbert spaces [45], and enabling the use of many further classification techniques. These steps are integrated and evaluated in aggregate. Each step in this pipeline (Fig. 4) is described in detail in the sections Automated Grammar Induction, Removing Duplicate Rules, Ranking Rules by Frequency, and Applying Classification Approaches. Together, these steps are designed to use graph-based data, such as vascular graphs, and provide rich rule frequency information which can be used to assess anomalies and diagnose disease.

Automated Grammar Induction. In this work a general mechanism for inducing grammars from uncoded and unlabeled structured data is used as the underlying approach to establish grammars for processing [8]. This approach has been used because it offers flexible and generic grammar induction, not requiring precoding of induced data, and being agnostic to both data complexity and structure—as long as the data has a graph representation incorporating nodes and edges, it can be leveraged by this approach, without it needing to adhere to any other particular formalism or information type. The algorithm breaks the data into tokens, generally the smallest meaningful units of the data, then recursively examine the tokens' relationships to find the most commonly repeated patterns in the graph and define rules based on those patterns (Table 1). As more tokens are processed the number of rules iteratively grows, and the related elements of the graph are replaced with the new rules. Because this happens recursively, earlier rules are often referenced by later rules, and as a consequence, a network of rules emerges that can generalize the structure of the induced data.

Our implementation uses a random starting point and a random walk, moving through the graph choosing the next node to evaluate at random from the connected nodes, to explore the unevaluated parts of the graph. Additionally, forming groups of parsed tokens and rules based on proximity within the graph facilitates faster rule generation by providing rule chunking [8]. Together these techniques constitute the first stage of rule frequency-based classification, establishing the set of rules across all cases, which will then be deduplicated between cases, and have their frequency assessed.

Removing Duplicate Rules. After using a representative sample of test cases to induce grammar with the previously described automated method, it is necessary to establish a minimal set of representative rules. This is done so that a functionally similar part of two cases will be identified as similar when comparing with the grammar rules, which is needed to perform accurate comparisons of sets of rules from different cases.

Repeated rules are easy to identify; if the LHS and RHS in each case match, then the rules are considered identical. However small groups of rules that have a similar collective function but are made up of unique rules require a more nuanced process to computationally assess their composed similarity.

Markov equivalence classes identify groups of elements with shared members through an adjacency matrix representation [44]. Groups are formed for chains of rules that share inputs and outputs. In this way, chains of rules found in one case, which compare identically to chains of rules found in another case, may be treated as similar metarules and removed, even when the individual rules making up these chains do not match exactly. With this approach, a single rule that can be represented with two separate rules is discouraged, as separate rules provide more versatility and nuance in classification.

This process involves checking each possible subgroup of rules against its counterparts, essentially creating a higher-level rule for each subgroup of rules. These higher-level rules can be compared rapidly in a pairwise fashion, but the process of checking each subgroup is computationally intensive. In practice, and in the examples conducted for this work, grammars are generally much smaller than 100,000 rules, and at this scale, the delay is unsubstantial for a standard modern computer. In a case where many rules must be compared further optimization can be achieved by introducing filtering criteria to rules, for example, if a particular rule is never observed outside of a particular scale range, it should only be considered as a relevant option with rule combinations near that scale.

Ranking Rules by Frequency. Having established a set of shared rules, the instances of each rule are counted in each case to be compared. This is straightforward with simple rules in which both the left and right-hand sides are particular elements or configurations. As each right-hand side is counted, updating the graph to contain the left-hand side allows for ongoing counting at

increasing levels of abstraction. Metarules [13], which are rules containing other rules on either side and thereby encompassing high level, abstract relationships in the data, are only applicable when all the standard rules have already been applied. For this reason, all the standard rules are counted by applying them to the data. Then the applicable metarules can be counted by applying them to the remaining combination of data and rules that have already been applied. In this way, all the rules and metarules can be counted for each case. This can lead to a process with more rule evaluations, as such, counting individual rules within a meta-rule when evaluating it allows for a substantial increase in speed, combinatorial simplification, and sensitivity to more granular features of the data.
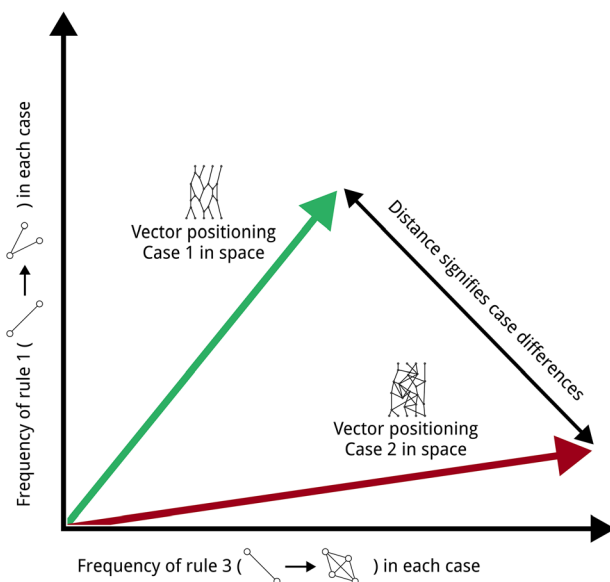
Rule frequency for each case is used as the core representation for learning and classification. Because groups of comparable cases are likely to share a majority of rules, after the initial induction process, further induction is not necessary except when there is a situation in which a rule is missing. For example, if in the process of checking rule frequency on a particular case, there is a part of the data with which no rule can be paired, this is an indication that the ruleset being used does not completely define the relevant design space. In this situation, the entire new case should be induced and frequency counting should be repeated for any previously evaluated cases, to avoid biases due to the earlier, partial rule set. In general, this will not influence metarules, because the newly identified rules will only be newly identified if they are atomic and are not constructed of metarules. Further, if novel metarules do exist in new cases, this only slows down the counting process and does not contribute to incorrect frequencies of underlying rules. In practice this is an uncommon situation because needing to reevaluate rules tends to indicate that the differences between the cases of data are significant, and may be obvious without utilizing a grammar-based approach.

Applying Classification Approaches. Given the convenient rule frequency abstraction, many classification approaches could be applied. In this work, we focus on demonstrating our method, so we use a common vector space-based classification mechanism, but others could be more appropriate for specific use cases or particular computational pipelines. In practice classifier selection in applied settings would depend on clinical details such as the condition base rate, feature heterogeneity, inferential context, and computational resources that can be applied to the specific classification problem. The vector space approach was deemed suitable in this work because it demonstrates a familiar representation that is compatible with many types of machine learning methods.

Treating each rule as a dimension in a vector representing a particular case, and the corresponding frequency of that rule in that case as its value, a Hilbert space [45] of designs is derived, extending traditional vector space computation into high dimensions. Treating each rule as a dimension may mean that the space is thousands of dimensions but the Hilbert space representation affords the use of standard distance metrics such as Euclidean distance with many dimensions, making the data amenable to detecting differences between induced cases using methods leveraging this representation.

As a note of caution, more entropy in the induced data will lead to more rules being needed to represent it. Providing more training samples than distinguishing rules minimizes potential overfitting. As such, reaching a minimal singular set of rules and metarules for vector space representation is strongly recommended. This way, only a solvable number of combinations of rules must be evaluated. If this precaution is impossible—due to high entropy data, for example, removing metarules can help, however, in very high entropy data, where compression by abstraction is effectively impossible, grammar methods will not offer a parsimonious representation.

Figure 5 demonstrates a simplified example with two dimensions and two cases. The x axis indicates the normalized

Fig. 5 Treating cases as vectors in a high dimensional space. The depiction demonstrates two dimensions associated with two rules in a simplified example. The length of the difference between vectors indicates how differently each case utilizes the grammar. This can serve as a proxy for a more general difference metric and is akin to the Euclidean distance within the space of cases. While only two dimensions are demonstrated for clarity, actual comparisons tend to integrate hundreds if not thousands of dimensions, one per rule in the deduplicated rule set.
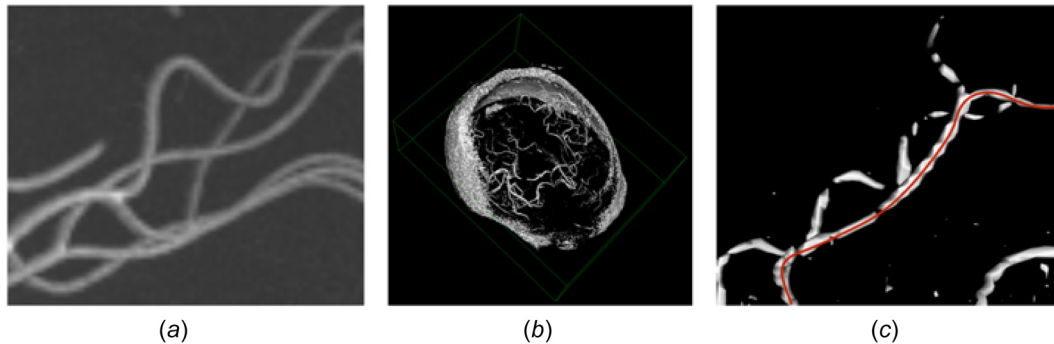
frequency of rule 3, while the y axis indicates the normalized frequency of rule 2 in each case, based on the rules defined in Fig. 2. In this way, the two vectors show where each case would be positioned in this space, due to their differing compositions of rules. The distance between these positions in space is the difference between the cases in this representation.

The vector space representation also lends itself to many more sophisticated statistical comparison techniques. For example, KNN [30] could be used to establish a nuanced classification boundary if there were many cases to train with. KNN establishes a classification boundary based on the a priori classification of the K nearest training cases to the test case. Nearness is defined contextually, but the Euclidian distance serves as the purest interpretation of conceptual distance in this space. Other statistical learning and classification techniques are also facilitated by the discrete vector representation of cases proposed, however in this work only the KNN approach is applied for classifying cases in the vector space representations due to the simplicity of the model.

Although many classification techniques require parameter tuning, for example, determining the smallest distance considered significant, aspects of the current system require minimal intervention because there are no integrated parameters for adjusting the grammar induction and vector space representation approaches. Additionally, once a case domain has been established, further classification comes at a very low computational cost, requiring only deriving a rule histogram and then performing the preferred distance classification technique with the resulting case vector.

Dataset

Our dataset consists of brain scans of 54 individuals with sickle cell disease using multimodal MRI, from a larger study conducted by several of the authors at the University of Pittsburgh Medical Center. Individuals had varied clinical histories, including stroke and other conditions that can be observed in brain vasculature. The images were acquired using an in-house developed radio

Fig. 6 Data segmentation process: (*a*) MRI is input, (*b*) 3D models from original MRI data are formed and cleaned by hand, and (*c*) centerline splines are extracted for processing

frequency system, composed of a 16-channel tic-tac-toe transmit array in conjunction with a 32-ch receive-only array and a 32-channel receive insert [46–49]. Ultrahigh-resolution 7T (Siemens Magnetom, Erlangen Germany) time-of-flight imagery was used, as it is a modality of MRI that provides coherent signals of arterial flow in the brain (shown in Fig. 2). The scans were taken with 320 μm isotropic resolution over the full brain. An advantage of time-of-flight imagery is that at this resolution, the arterial flow is very clear and assumed to be close to physically accurate in terms of position and scale, due to minimal blooming in the time-of-flight modality.
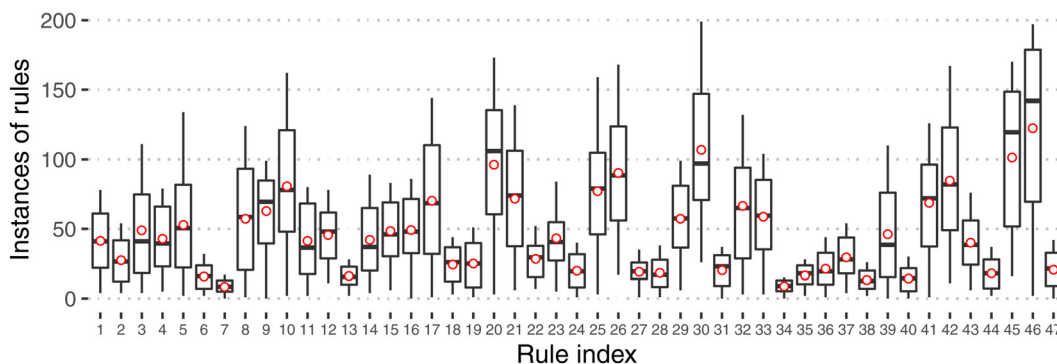
Sic radiologically identifiable conditions with key vascular indicators were the diagnostic focus in this work:

(1) Infarction and microvascular ischemia are correlated with areas of reduced flow or uneven flow [50,51] and can result in permanent tissue injury [52].

(2) Arteriolosclerosis [53], a risk factor for infarction caused by buildups of plaque, can be identified by signals such as uneven changes in arterial radius.

(3) Hypertension, a condition brought on by extended high blood pressure and associated with tortuosity in arterial vasculature [54,55].

(4) Aneurysm is identified with areas of distended vasculature and may have serious health impacts including rupture and sudden death [56].

(5) Signals like abruptly terminating vessels in the brain are indicators of vessel occlusion [57] and may result in infarction in the clinical setting of stroke.

(6) Finally, smooth reductions in arterial radius can indicate displacement, chronically reduced flow, or an underlying vasculopathy [58].

With these indicators and associated conditions six categories of vessel patterns are identified and are used in assessing individuals in our analysis.

Data Preparation. Vascular surfaces were extracted from the time-of-flight scans and exported to the stereolithography file format using Horos's default threshold standard (Fig. 6(*a*)). Extraneous data such as isolated venous artifacts and skull fragments were manually removed en masse in Blender (Fig. 6(*b*)). The center splines of the vascular surfaces were identified by averaging three-dimensional (3D) points' locations until a singular point remained at every cross section using the blender application programming interface (Fig. 6(*c*)). This process was conducted semimanually to provide added flexibility when dealing with this small dataset; however, analogous approaches have been fully automated, achieving similar resulting vessel graphs [43,59].

A distance threshold used to average vessels to splines was recorded at each spline point as an approximation for vessel diameter. A graph was formed to include position, diameter, and relative point cloud data at every point along all the resulting splines. Normalized changes in vascular diameter were recorded as metadata in the grammar representation. Using the point cloud data of a vessel, a similar approach was taken to represent information about the tortuosity and curvature of vessel segments when the vascular graph was generated. For example, a highly tortuous vessel with no branching would be recorded as a single edge in the abstract graph with a point cloud in the grammar representation for positioning rules and visualization. Additionally, splines that, at some point, pass closer than the raw width value of the proximal vessels were considered intersecting for implementing graph

Fig. 7 Grammar rule counts shown across all individuals. Markers indicate the mean frequency of each rule. Rule indices are assigned arbitrarily based on the order in the rules are finalized by the algorithm.

abstraction, so a node would be added at that point in the graph encapsulating an intersection.

The grammar induction process was conducted repeatedly with the entire set of 54 individuals and with subgroups of the population to evaluate how grammar induction results were impacted by the number of individuals induced. Forty-seven rules were identified when the entire dataset was processed and deduplicated (see Appendix). Even inducing just three individual cases established a robust library of rules, enabling rule counting from other individuals' data. Adding more individuals' data to the induction process had a minimal impact on the quantity of resulting rules. As a consequence, and to leverage a standardized representation throughout the classification analysis, the set of 47 rules is used in our approach here.
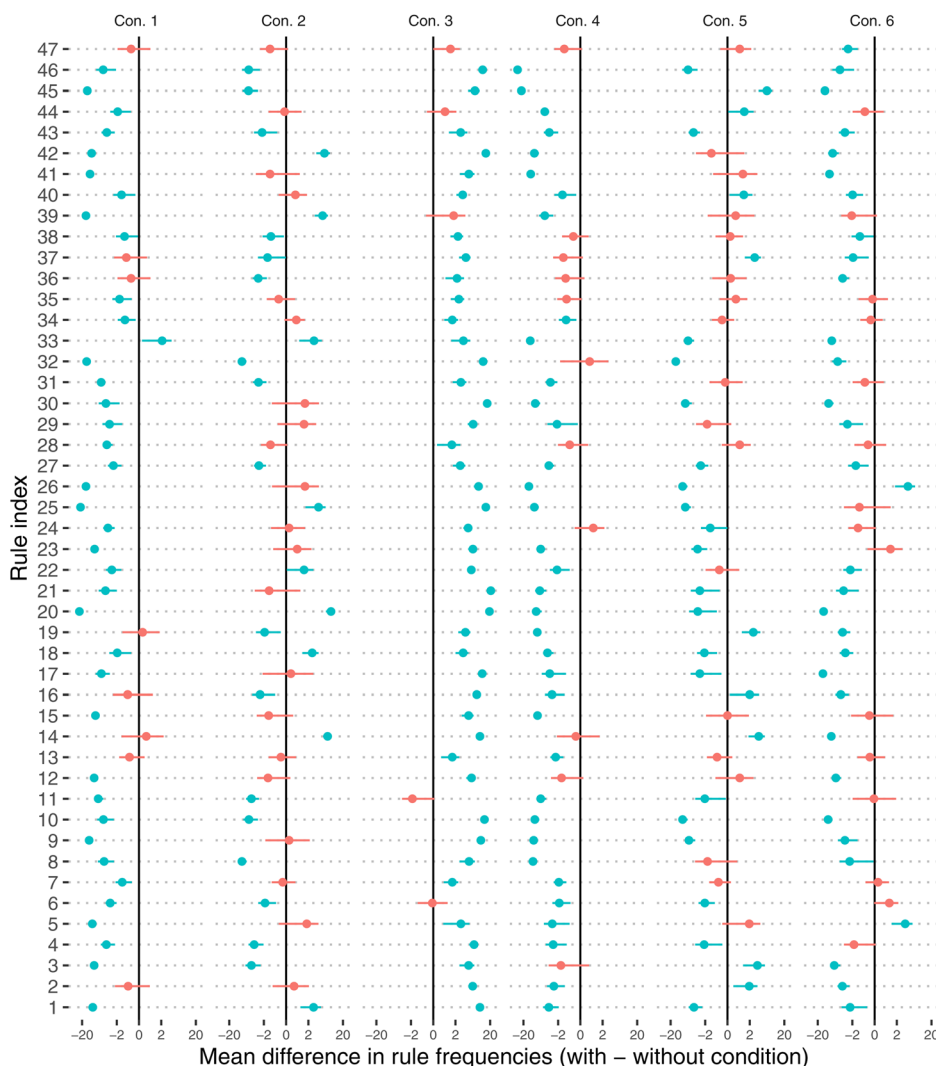
Results

We evaluate our method by applying it to identify a range of condition indicators in individuals using automatically induced grammar rule frequencies in an out-of-sample, in-distribution prediction design. In other words, this work aims to check if grammar methods can predict the correct condition indicators among individuals within a clinical context—sickle cell disease—to demonstrate broader uses in identifying novel conditions, and in supporting radiologists.

First, rule frequencies across cases and conditions are analyzed, then significant differences in rule frequencies are used to highlight which rules serve as condition predictors, and finally, out-of-sample prediction accuracy is computed.
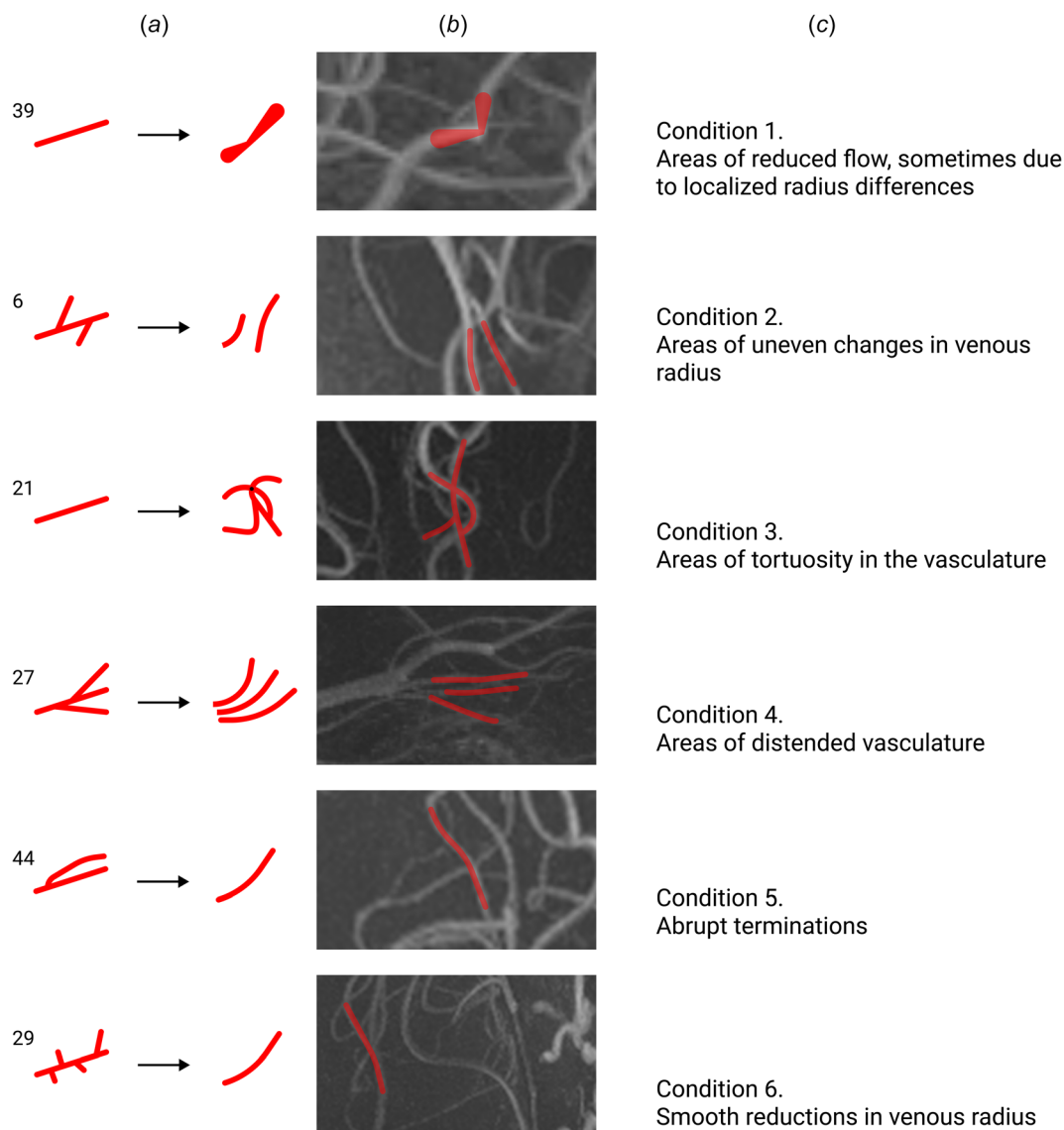
Identifying Condition Indicators From Rules. Condition indicators are treated as categorical features and rules are counted and reported as the raw count of the number of instances observed in a given individual (Fig. 7 shows box plots of each rule count distribution). Rules ranged in frequency substantially across the population ($\mu = 46.37$, $\sigma = 40.59$), the least common on average being rule 7 ($\mu = 8.48$, $\sigma = 5.07$), and the most being rule 46 ($\mu = 122.57$, $\sigma = 62.60$). The stark variability in the base rates of rules is a consequence of integrating rules across patients in the induction process but potentially adds sensitivity—smaller differences in rare rule counts may also serve as condition indicators.

Mean differences between the number of occurrences of a rule for individuals with conditions and the population mean for that rule provide a straightforward metric for identifying groups of rules that serve as condition indicators for each condition (Fig. 8 highlights predictive rules). Confidence intervals that do not intersect with 0 difference indicate that a given rule is a significant predictor of a condition. However, because some patients have no

Fig. 8 The mean difference between rule frequencies of patients with and without a condition are shown with 95% confidence intervals. Intervals not intersecting vertical lines are significant predictors.

Fig. 9 Rules in situ versus diagnostic indicators. (*a*) key rules from the grammar with their rule index, (*b*) key rules in situ-based on the detection conducted in the classification process, and (*c*) descriptions of each condition indicator based on the radiological assessment of a vessel pattern. Condition indicators 1 and 3 are visually recapitulated by the grammar rules, however, the others do not visually corroborate the expected diagnostic indicators, even though these are each the strongest individual indicator of each condition.

occurrences of some indicative rules, ensembles of rules are required to create reliable predictions. Notably, several conditions have quite notable directional trends in the rule count mean difference measure. Condition 1, 4, and 6 are below the population means for the majority of the rules. On the other hand, condition 3 is higher than the population means for all significant features.

Evaluating rules in context, some display patterns in the data that align with informal visual diagnostic indicators, while others do not (Fig. 9). This reinforces the notion that rules capture human observable patterns through induction and rule frequency analysis, but also, that it identifies patterns that have no discernible meaning to human judges, without losing interpretability.

Predicting Condition Indicators in Aggregate. To evaluate the predictive strength of rules as an ensemble, logistic regression models are fitted to predict individual conditions given rule count mean differences. This produces one model per condition. The

models are evaluated with 10-fold cross validation to estimate predictive accuracy (Table 2).

Conditions 1, 3, and 4 achieve moderate accuracy with this approach, however, condition 2, has extremely low predictive

Table 2 Tenfold cross validation of evaluation models for each condition indicator

Condition	Predictive accuracy
1	0.72
2	0.00
3	0.72
4	0.58
5	0.21
6	0.41

The cross-validation accuracy values are reported (larger is better). Rule mean difference rule counts for every rule (Fig. 8) are used to generate the model.

accuracy when attempting to model all rules simultaneously. Conditions 5 and 6 appear to encounter a similar effect to a lesser degree. As is shown in Fig. 8, it is visually apparent that these conditions have fewer significant predictor rules. However, there are single rules or smaller groups of rules that can be used to directly detect these conditions accurately. This may indicate that some conditions have more underlying noise, in terms of rule count distributions, and as a consequence, different classification approaches should be applied for those conditions. While accurate classifications are possible, with this approach, tuning classifiers for specific conditions is almost certainly required for most practical uses of such a system. For example, it is possible that condition 2 could be identified if the plaque was more clearly distinguishable, but without that, changes in venous radius alone are not a strong indicator to be predictive. Similarly, conditions 5 and 6 could potentially be identified by rules that differentiate terminations from vessels with low flow.

Discussion

Through qualitatively comparing algorithmically identified key rules to the vessel patterns that are often sought by radiologists for each condition indicator, it is clear that some rules highlight expected features, while others leverage signals that do not appear in the human indicators at all. Figure 9 shows each rule in situ with the related diagnostic indicator. The key rules for condition indicators 1 and 3 show a pattern that can be visually interpreted similarly to the condition indicator. However, many of the other rules found to predict an indicator well do not actually appear visually similar to the signals radiologists use. This does not degrade the accuracy of prediction but speaks to a challenge that radiologists face when assessing conditions: traditional indicators remain relatively nonspecific, and, due to a lack of structured computational analysis, it is hard for them to formally refine the indicators for which they look. On the other hand, radiologists look for much more than a few graph rules when diagnosing an individual, and the wealth of other data they incorporate plays an instrumental role in achieving high-quality diagnoses.

Critically, this work studies a consistent clinical context and therefore does not allow for a general prediction of these cases or a clinical diagnostic. Instead, it aims to contribute a high-level method that could be applied to other datasets in which grammar induction and classification are suitable. As a consequence, we have not conducted the usual comparative statistical analyses needed to validate this method as a diagnostic for a particular condition but instead introduce it as a potential alternative exploratory mechanism in settings where the number of data samples is small, and the specific data features that can serve as predictors are relatively unknown.

Additionally, the data used in this evaluation is higher resolution and lower noise than most of the MRI data in clinical use today. This level of resolution is important to adequately identify the vascular graph that is necessary for rule induction. To achieve widespread clinical use of our approach, high-resolution MRI must become more widespread, or alternative approaches for extracting the vascular graph data must be developed.

This work has also not addressed a temporal change in individuals, so in this format, it cannot be used to predict future conditions, however, with longitudinal training samples, this limitation could be overcome. Inducing grammars on the same individual data at several time points would allow for analyzing trends and predicting future states across patients.

Radiologically this work demonstrates that: (1) data requirements are drastically decreased by utilizing inherent structural information and learning over structural features, making the bag-of-features approach assessable on a diversity of alternative radiological situations and (2) specific alternative representations, such as design-based grammar models, afford structured symbolic abstraction of graph-structured data. The combined implication of these is that many traditional applications of machine learning in medicine may be revisited, and many new avenues of analysis may be unlocked by both considering structural information as a key indicator, and identifying derivable high-level representations which can be learned as preprocesses for further classification or analysis using statistical methods. Further, methods like this are fundamentally interpretable—an increasingly desirable property of model-driven decision systems [60].

Conclusion

Our work introduced a pipeline for using grammar rules in anomaly detection by studying rule frequency-based representations of data samples. The technique was applied to evaluating anomalies in brain vascular graphs and associating them with a series of conditions. Many rules that indicate particular conditions shared features that clinicians use to identify those conditions by eye.

By integrating automated shape grammar methods in a statistical learning pipeline, the ability of clinical analysis to potentially achieve greater efficiency in processing and greater reliability in isolating indicators for particular conditions holds tremendous promise. In the context of this work, the introduced approach is also likely to be suitable for many other types of analysis in the medical domain such as oncology, neurology, hematology, and epidemiology, in addition to fields outside of medicine that implicitly or explicitly utilize graph-structured data.

Appendix: All Induced Rules

The full set of 47 induced and simplified grammar rules from the patient data that was processed (Fig. 10). Note, the grammar rules have been simplified for visualization purposes. For instance, many left-hand side tokens of rules here are depicted as a linear section of a vessel which indicates a relatively arbitrary segment of the vessel. Additionally, feature scale is not explicitly defined in this grammar approach so scale and orientation may change more than is visually obvious in a singular representation of each rule. These are two-dimensional projections of 3D vessel networks, so some rules include vessels that overlap in this visualization but do not intersect; these situations are marked with a small black dot at the overlap. Vessel radius changes are represented as wider lines, e.g., rules 9, 10, 13, 16, 22, 33, and 39. While indistinguishable regions—which tend to be a very tight bundle of vessels—are depicted with blobs, e.g., rules 12, 16, and 36. Because of the scale-independent nature of the rules, changes in tortuosity or curvature are proportionate to a venous radius. In this depiction, rules have been oriented to maximize the visibility of the modified properties.

Fig. 10   All induced rules

References

[1] Bullitt, E., Zeng, D., Gerig, G., Aylward, S., Joshi, S., Smith, J. K., Lin, W., and Ewend, M. G., 2005, "Vessel Tortuosity and Brain Tumor Malignancy: A Blinded study1," Acad. Radiol., **12**(10), pp. 1232–1240.

[2] Bergers, G., and Benjamin, L. E., 2003, "Tumorigenesis and the Angiogenic Switch," Nat. Rev. Cancer, **3**(6), pp. 401–410.

[3] Arvanitakis, Z., Capuano, A. W., Leurgans, S. E., Buchman, A. S., Bennett, D. A., and Schneider, J. A., 2017, "The Relationship of Cerebral Vessel Pathology to Brain Microinfarcts," Brain Pathol., **27**(1), pp. 77–85.

[4] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S., 2017, "Dermatologist-Level Classification of Skin Cancer With Deep Neural Networks," Nature, **542**(7639), pp. 115–118.

[5] Stiny, G., 1980, "Introduction to Shape and Shape Grammars," Environ. Plann. B, **7**(3), pp. 343–351.

[6] Rozenberg, G., 1997, "Handbook of Graph Grammars and Computing by Graph Transformation," Handbook Graph Grammars, Vol. 1, World Scientific, Singapore, pp. 1–8.

[7] Gobet, F., Lane, P., Croker, S., Cheng, P., Jones, G., Oliver, I., and Pine, J., 2001, "Chunking Mechanisms in Human Learning," Trends Cognit. Sci., **5**(6), pp. 236–243.

[8] Whiting, M. E., Cagan, J., and LeDuc, P., 2018, "Efficient Probabilistic Grammar Induction for Design," AI Edam, **32**(2), pp. 177–188.

[9] Csurka, G., Dance, C., Fan, L., Willamowski, J., and Bray, C., 2004, "Visual Categorization With Bag of Keypoints," Workshop on Statistical Learning in Computer Vision, ECCV (Vol. 1, No. 1–22), Prague, Czech Republic, May 11–14.

[10] Jones, M., Craddock, P. T., and Barker, N., 1990, "Fake?: The Art of Deception," Book, UC Press, Berkeley, CA, p. 312.

[11] Bamossy, G., and Scammon, D. L., 1985, "Product Counterfeiting - Consumers and Manufacturers Beware," Adv. Consumer Res., **12**, pp. 334–339.

[12] Hodge, V. J., and Austin, J., 2004, "A Survey of Outlier Detection Methodologies," Artif. Intell. Rev., **22**(2), pp. 85–126.

[13] Sondheimer, N. K., and Weischedel, R. M., 1980, "A Rule-Based Approach to Ill-Formed Input," The 8th International Conference on Computational Linguistics, Tokyo, Japan, Sept. 30–Oct. 4, Vol. 1, pp. 46–53.

[14] Trescak, T., Rodriguez, I., and Esteva, M., 2009, "General Shape Grammar Interpreter for Intelligent Designs Generations," Proceedings of the Sixth International Conference on Computer Graphics, Imaging and Visualization: New Advances and Trends, CGIV2009, Tianjin, China, Aug. 11–14, pp. 235–240.

[15] D'Ulizia, A., Ferri, F., and Grifoni, P., 2011, "A Survey of Grammatical Inference Methods for Natural Language Learning," Artif. Intell. Rev., **36**(1), pp. 1–27.

[16] Stiny, G., and Mitchell, W. J., 1978, The Palladian Grammar, Environment and Planning B: Planning and Design, **5**(1), pp. 5–18.

[17] Pugliese, M. J., and Cagan, J., 2002, "Capturing a Rebel: Modeling Harley-Davidson Brand through a Motorcycle Shape Grammar," Res. Eng. Design, **13**(3), pp. 139–156.

[18] McCormack, J. P., Cagan, J., and Vogel, C. M., 2004, "Speaking the Buick Language: Capturing, Understanding, and Exploring Brand Identity With Shape Grammars," Des. Stud., **25**(1), pp. 1–29.

[19] Starling, A. C., and Shea, K., 2003, "A Grammatical Approach to Computational Generation of Mechanical Clock Designs," DS 31: Proceedings of ICED 03, the 14th International Conference on Engineering Design, Stockholm, Sweden, Aug. 19–21, pp. 445–446.

[20] Gips, J., 1999, "Computer Implementation of Shape Grammars," NSF/MIT Workshop Shape Computation, Boston, MA, Apr., pp. 1–11.

[21] Talton, J., Yang, L., Kumar, R., Lim, M., Goodman, N., and Měch, R., 2012, "Learning Design Patterns With Bayesian Grammar Induction," Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology—UIST '12, New York, Oct., p. 63.

[22] Yeh, R. Y., Nischal, K. K., LeDuc, P., and Cagan, J., 2020, "Written in Blood: Applying Shape Grammars to Retinal Vasculatures," Transl. Vision Sci. Technol., 9(9), pp. 36–36.

[23] Cortes, C., and Vapnik, V., 1995, "Support Vector Machine," Mach. Learn., 20(3), pp. 273–297.

[24] Wallach, H. M., 2006, "Topic Modeling: Beyond Bag-of-Words," Proceedings of the 23rd International Conference on Machine Learning, ACM, Pittsburgh, PA, June 25–29, pp. 977–984.

[25] Fu, K.-S., and Swain, P., 2012, "On Syntactic Pattern Recognition," SEN Report Series Software Engineering, Vol. 2, Springer, Berlin, pp. 155–182.

[26] Albus, J. E., Anderson, R., Brayer, J., DeMori, R., Feng, H.-Y., Horowitz, S., Moayer, B., 2012, Syntactic Pattern Recognition, Applications, Vol. 14, Springer, Berlin.

[27] Hattemer, H. H., 1982, "Genetic Distance Between Populations," TAG Theor. Appl. Genet., 62(3), pp. 219–223.

[28] Christensen, B. T., and Schunn, C. D., 2007, "The Relationship of Analogical Distance to Analogical Function and Preinventive Structure: The Case of Engineering Design," Memory Cognit., 35(1), pp. 29–38.

[29] Wootters, W. K., 1981, "Statistical Distance and Hilbert Space," Phys. Rev. D, 23(2), pp. 357–362.

[30] Cover, T., and Hart, P., 1967, "Nearest Neighbor Pattern Classification," IEEE Trans. Inf. Theory, 13(1), pp. 21–27.

[31] Akoglu, L., Tong, H., and Koutra, D., 2015, "Graph Based Anomaly Detection and Description: A Survey," Data Mining Knowl. Discov., 29(3), pp. 626–688.

[32] Jiang, C., Coenen, F., and Zito, M., 2013, "A Survey of Frequent Subgraph Mining Algorithms," Knowl. Eng. Rev., 28(1), pp. 75–105.

[33] Yan, X., and Han, J., 2002, "Gspan: Graph-Based Substructure Pattern Mining," Proceedings of IEEE International Conference on Data Mining, IEEE, Maebashi City, Japan, Dec. 9–12, pp. 721–724.

[34] Chakrabarti, D., 2004, "AutoPart: Parameter-Free Graph Partitioning and Outlier Detection," J. F. Boulicaut, F. Esposito, F. Giannotti, and D. Pedreschi, eds., Knowledge Discovery in Databases: PKDD 2004, Lecture Notes in Computer Science, Vol. 3202, Springer, Berlin, Heidelberg.

[35] Hearst, M. A., Dumais, S. T., and Osman, E., 1998, "Support Vector Machines," IEEE Intell. Syst. Appl., 1(1), pp. 1–8.

[36] Candès, E. J., Li, X., Ma, Y., and Wright, J., 2011, "Robust Principal Component Analysis?," J. ACM (JACM), 58(3), pp. 1–37.

[37] Wold, S., Esbensen, K., and Geladi, P., 1987, "Principal Component Analysis," Chemom. Intell. Lab. Syst., 2(1–3), pp. 37–52.

[38] Orsborn, S., Boatwright, P., and Cagan, J., 2008, "Identifying Product Shape Relationships Using Principal Component Analysis," Res. Eng. Des., 18(4), pp. 163–180.

[39] Orsborn, S., Cagan, J., and Boatwright, P., 2008, "A Methodology for Creating a Statistically Derived Shape Grammar Composed of Non-Obvious Shape Chunks," Res. Eng. Des., 18(4), pp. 181–196.

[40] Sánchez-Martínez, F., and Pérez-Ortiz, J. A., 2010, "Philipp Koehn, Statistical Machine Translation," Mach. Transl., 24(3–4), pp. 273–278.

[41] Nevill-Manning, C. G., 1996, "Inferring Sequential Structure," Doctoral dissertation, University of Waikato, Hamilton, New Zealand.

[42] Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., and Xiao, J., 2015, "3D Shapenets: A Deep Representation for Volumetric Shapes," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, June 7–12, pp. 1912–1920.

[43] Helmberger, M., Pienn, M., Urschler, M., Kullnig, P., Stollberger, R., Kovacs, G., Olschewski, A., Olschewski, H., and Bálint, Z., 2014, "Quantification of Tortuosity and Fractal Dimension of the Lung Vessels in Pulmonary Hypertension Patients," PLoS One, 9(1), p. e87515.

[44] Andersson, S. A., Madigan, D., and Perlman, M. D., 1997, "A Characterization of Markov Equivalence Classes for Acyclic Diagraphs," Ann. Stat., 25(2), pp. 505–541.

[45] Smola, A., Gretton, A., Song, L., and Schölkopf, B., 2007, "A Hilbert Space Embedding for Distributions," M. Hutter, R. A. Servedio, and E. Takimoto, eds., Algorithmic Learning Theory. ALT. Lecture Notes in Computer Science, Vol. 4754, Springer, Berlin, Heidelberg.

[46] Santini, T., Kim, J., Wood, S., Krishnamurthy, N., Farhat, N., Maciel, C., Raval, S., Zhao, T., and Ibrahim, T. S., 2018, "A New rf Transmit Coil for Foot and Ankle Imaging at 7t MRI," Magn. Reson. Imag., 45, pp. 1–6.

[47] Kim, J., Krishnamurthy, N., Santini, T., Zhao, Y., Zhao, T., Bae, K. T., and Ibrahim, T. S., 2016, "Experimental and Numerical Analysis of b1+ Field and Sar With a New Transmit Array Design for 7t Breast MRI," J. Magn. Reson., 269, pp. 55–64.

[48] Kim, J., Santini, T., Bae, K. T., Krishnamurthy, N., Zhao, Y., Zhao, T., and Ibrahim, T. S., 2017, "Development of a 7 trf Coil System for Breast Imaging," NMR Biomed., 30(1), p. e3664.

[49] Smagula, S. F., Karim, H. T., Rangarajan, A., Santos, F. P., Wood, S. C., Santini, T., Jakicic, J. M., Reynolds, C. F., Cameron, J. L., Vallejo, A. N., Butters, M. A., Rosano, C., Ibrahim, T. S., Erickson, K. I., and Aizenstein, H. J., 2018, "Association of Hippocampal Substructure Resting-State Functional Connectivity With Memory Performance in Older Adults," Am. J. Geriatr. Psychiatry, 26(6), pp. 690–699.

[50] Rivas, F., Cobb, F. R., Bache, R. J., and Greenfield, J. C., 1976, "Relationship Between Blood Flow to Ischemic Regions and Extent of Myocardial Infarction. Serial Measurement of Blood Flow to Ischemic Regions in Dogs," Circu. Res., 38(5), pp. 439–447.

[51] O'Sullivan, M., Lythgoe, D. J., Pereira, A. C., Summers, P. E., Jarosz, J. M., Williams, S. C. R., and Markus, H. S., 2002, "Patterns of Cerebral Blood Flow Reduction in Patients With Ischemic Leukoaraiosis," Neurology, 59(3), pp. 321–326.

[52] Ay, H., Koroshetz, W. J., Vangel, M., Benner, T., Melinosky, C., Zhu, M., Menezes, N., Lopez, C. J., and Sorensen, A. G., 2005, "Conversion of Ischemic Brain Tissue Into Infarction Increases With Age," Stroke, 36(12), pp. 2632–2636.

[53] Ross, R., 1986, "The Pathogenesis of Atherosclerosis—An Update," New Engl. J. Med., 314(8), pp. 488–500.

[54] Hiroki, M., Miyashita, K., and Oda, M., 2002, "Tortuosity of the White Matter Medullary Arterioles is Related to the Severity of Hypertension," Cerebrovasc. Diseases, 13(4), pp. 242–250.

[55] Dobrin, P. B., Schwarcz, T., and Baker, W., 1988, "Mechanisms of Arterial and Aneurysmal Tortuosity," Surgery, 104(3), pp. 568–571.

[56] Chaikof, E. L., Blankensteijn, J. D., Harris, P. L., White, G. H., Zarins, C. K., Bernhard, V. M., Matsumura, J. S., May, J., Veith, F. J., Fillinger, M. F., Rutherford, R. B., and Kent, K. C., 2002, "Reporting Standards for Endovascular Aortic Aneurysm Repair," J. Vasc. Surg., 35(5), pp. 1048–1060.

[57] Moore, P. M., 1989, "Diagnosis and Management of Isolated Angiitis of the Central Nervous System," Neurology, 39(2), pp. 167–167.

[58] Tang, D., Yang, C., Kobayashi, S., and Ku, D. N., 2001, "Steady Flow and Wall Compression in Stenotic Arteries: A Three-Dimensional Thick-Wall Model With Fluid–Wall Interactions," ASME J. Biomech. Eng., 123(6), pp. 548–557.

[59] Zhang, Y., Bazilevs, Y., Goswami, S., Bajaj, C. L., and Hughes, T. J. R., 2006, "Patient-Specific Vascular NURBS Modeling for Isogeometric Analysis of Blood Flow," Computer Methods Appl. Mech. Eng., 196(29–30), pp. 2943–2959.

[60] Rudin, C., 2019, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead," Nat. Mach. Intell., 1(5), pp. 206–215.