

# Clinical Careflows aided by Uncertainty Representation Models

Tiago Oliveira<sup>1</sup>, João Neves<sup>2</sup>, Ernesto Barbosa<sup>1</sup>, and Paulo Novais<sup>1</sup>

<sup>1</sup> CCTC/DI, University of Minho, Braga, Portugal

{toliveira,pjon}@uminho.pt

pg18744@alunos.uminho.pt

<sup>2</sup> Hospital of Braga, Braga, Portugal

joao.neves@hospitaldebraga.pt

**Abstract.** Choosing an appropriate support for Clinical Decision Support Systems is a complicated task, and dependent on the domain in which the system will intervene. The development of wide solutions, which are transversal to different clinical specialties, is impaired by the existence of complex decision moments that reflect the uncertainty and imprecision that are often present in these processes. The need for solutions that combine the relational nature of declarative knowledge with other models, capable of handling that uncertainty, is a necessity that current systems may be faced with. Following this line of thought, this work introduces an ontology for the representation of Clinical Practice Guidelines, with a case-study regarding colorectal cancer. It also presents two models, one based on Bayesian Networks, and another one on Artificial Neural Networks, for colorectal cancer prognosis. The objective is to observe how well these two ways of obtaining and representing knowledge are complementary, and how the machine learning models perform, attending to the available information.

**Keywords:** Clinical Decision Support Systems, Computer-Interpretable Guidelines, Clinical Uncertainty, Machine Learning.

## 1 Introduction

Currently, the penetration of Clinical Decision Support Systems (CDSSs) in daily healthcare delivery is becoming a reality. There is even evidence that the use of such systems can contribute positively to the improvement of healthcare services, namely in the prevention of medication errors [2], and the improvement of practitioner performance [3]. The main goal of these systems is to help healthcare professionals to make decisions by dealing with clinical data and knowledge. The advent of CDSSs occurred in the middle of the 1960s and the early 1970s. Through the years, CDSSs evolved into three main types [1]: (i) tools for information management (e.g. Electronic Medical Record (EMR) systems); (ii) tools for focusing attention (e.g. alert systems); and (iii) tools for providing patient-specific recommendations. This paper focuses on the last which are tools

that provide custom assessments based on sets of patient data. Different techniques have been used to support the decision making process of CDSSs, they range from mathematical modelling, pattern recognition and statistical analysis of large databases to specific algorithms represented as flowcharts [1].

This work follows a hybrid approach consisting in specific algorithms combined with models obtained through machine learning processes. The basis for the algorithmic part will be provided by Clinical Practice Guidelines (CPGs) [4], which are systematically developed statements that provide healthcare professionals with instructions regarding specific clinical circumstances. These documents are the preferred support for the delivery of medical information to physicians, but their current format shows some fundamental limitations that, sometimes, prevent them from fully achieving their goals. The referred limitations include ambiguity of instructions and large textual descriptions that are difficult to consult during healthcare delivery [5]. A machine-readable format would provide a solution for some of the limitations of CPGs and, at the same time, a proper support for a CDSS. As such, this work proposes an ontology model for the representation of CPG tasks combined with classification models for specific cases where uncertainty is more evident.

The paper is organized as follows. The next section contains a description of the primitives used in the CPG ontology along with a proper case study featuring colorectal cancer (CRC) diagnosis and treatment. In Europe, this is one of the most common forms of cancer (only second to breast cancer) and it affects predominantly the western countries, a group in which Portugal is included [6]. Section three introduces a moment in CRC management that is usually clouded with uncertainty, the prognosis after surgery, as well as a set of models based on Bayesian Networks (BNs) and Artificial Neural Networks (ANNs) for mortality prediction. The last section presents some conclusions about the work done so far and points out to future directions.

## 2 Clinical Practice Guideline Representation

The approach followed for CPG representation includes an ontology developed in Ontology Web Language (OWL) *McGuinness2004*. When the objective is to create a standard of knowledge representation for use in different applications, an ontology is arguably the best way to convey information. OWL-DL (Description Logics) is a highly expressive language comprised of classes (sets of individuals having certain properties), individuals (objects of the domain) and properties (binary relationships between individuals or between individuals and data). The developed ontology is called *CompGuide* and presents a formalisation of guidelines as linked lists of tasks. This approach was based on Computer-Interpretable Guideline (CIG) [8] formalisms that follow the Task Network Model (TNM), representing CPGs as networks, or workflows, of tasks [9]. Such formalisms include the Guideline Interchange Format (GLIF)[10], PROforma [11] and the Standards-based Sharable Active Guideline Environment (SAGE) [12], just to name a few. The following subsections will present the main class primitives and

the properties that enable the definition of the order between tasks, as well as temporal and clinical constraints.

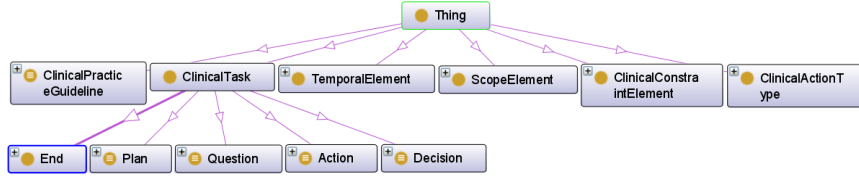


Fig. 1. Representation of the main primitive classes of *CompGuide*.

## 2.1 Task Primitives

In *CompGuide* a GPG is represented as an instance of the class *ClinicalPracticeGuideline*. To sanction the nesting of classes, it was considered that all tasks of a guideline are contained in a broader task called *Plan*, to which an individual of *ClinicalPracticeGuideline* is linked through the *hasPlan* object property. Figure 1 shows a graph containing the top classes of *CompGuide*

Each guideline has only one *Plan*, and every *Plan* has a variable number of tasks, including other *Plans*. The main classes are subclasses of *ClinicalTask* and consist in *Plan*, *Action*, *Question* and *Decision*. These tasks have properties, or are linked to individuals from other classes in order to express different procedures. Starting with the *Action* class, it is used for steps in the guidelines that must be performed by a healthcare agent, thus encompassing clinical procedures, clinical exams, medication and non-medication recommendations. When some statement concerning a patient has to be asserted, the *Decision* task is used to produce it based on the verification of previously specified conditions and the selection of defined options. The association of conditions to options is done via object properties that link individuals from *Decision* to individuals from *ClinicalConstraintElement*. Feeding this decision process is possible through a *Question* task, which collects all the information necessary for applying a guideline. The individuals that belong to this class have data properties to specify the clinical parameters, and the units under which they should be expressed. Finally, instances of the *End* class are used to signal the end of a careflow.

The definition of a relative order between the tasks is achieved through a set of object properties. A *Plan* is linked to the first of its tasks by the *hasFirstTask* object property, and the task that follows it is connected to the previous by the *nextTask* property. The property ensures the sequential execution of tasks, but leaves out cases where they should be carried out at the same time or alternatively. For these special cases, one uses the *parallelTask* and the *alternativeTask* object properties, respectively.

Figure 2 shows a simplified excerpt from a guideline for diagnosis and management of CRC from the National Comprehensive Cancer Network (NCCN),

represented according to *CompGuide*. The main *Plan* of the guideline starts with a *Question* task with the objective of obtaining some specific clinical parameters (e.g. change of bowel habits, occurrence of weight loss and vomiting, among others). Then a *Decision* task is proposed to assess the need for complementary means of diagnosis, based on the answers to the previous task. There are two alternative tasks for the next step, selected according to trigger conditions concerning the result of the Decision. If the *Action* task is the one selected, a set of exams is proposed based on which the next *Decision* task will assess the need for CRC surgery. Again, two tasks are shown as an alternative and the selection is carried out the same way as in the previous situation, using trigger conditions according to the possible outcomes of the *Decision*. If the surgery route is followed, then the next step would be a *Plan* for adjuvant therapy selection, i.e., choosing the most appropriate chemotherapy and/or radiotherapy scheme to be applied. In the meantime, there is a prognosis stage whose representation falls outside of the capabilities of the ontology. As such, the integration of the careflow provided by the ontology with a classification model, derived from data, is necessary, as it will be discussed further here-in.

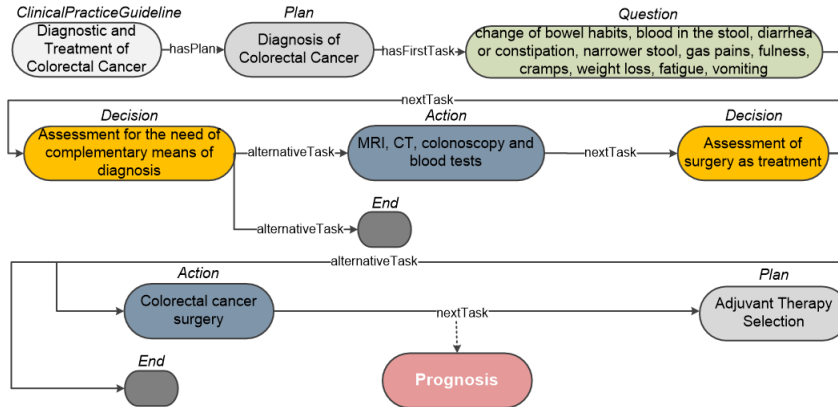


Fig. 2. Excerpt of a guideline for the diagnosis and treatment of colorectal cancer.

## 2.2 Temporal Constraints

Besides the relative order by which they are executed, clinical procedures are also bound by temporal constraints, such as duration and cyclic repetitions. In this case, the duration indicates for how long the task should stay active. Hence, to express this, there is the *Duration* class under *TemporalElement*. This class is defined only for *Plans* and *Actions* and each of its individuals has a *decimal* data property, the *DurationValue*, and a *hasTemporalUnit* object property that connects it to an individual from *TemporalUnit*. The available temporal units

are *second*, *minute*, *hour*, *day*, *week*, *month* and *year*. In *Loop*, also under *TemporalElement*, one defines the repetition cycles for both *Actions* and *Plans*. An individual from *Loop* has an *integer* data property named *RepetitionValue* where it is possible to specify the number of repetitions of the referred tasks. Another feature is an object property, *hasPeriodicity*, linking to individuals of a class called *Periodicity*. The individuals from this class possess two constructors to define periodic executions of tasks, namely the *PeriodicityValue* data property and the *hasTemporalUnit* object property.

### 2.3 Clinical Constraints

The execution of tasks depends on the verification of conditions. In a *Decision* task there is a choice between two or more options which are represented by individuals of the *Option* class under *ClinicalConstraintElement*. The *Option* class is defined by properties that enable the expression of the *Parameter* the option reports to and the value to be asserted to the patient state which might be either a *NumericalValue* or a *QualitativeValue*. For option selection, the definition of conditions is essential. This is done through the *ConditionSet* classes whose instances represent sets of conditions which, in turn, are created as instances of the *Condition* class. The later has appropriate properties to specify the clinical parameter the condition reports to and the value of that parameter that should be checked.

CompGuide also models other types of conditions, namely *TriggerConditions*, *PreConditions* and *Outcomes*, all of them defined under *ClinicalConstraintElement*. A *TriggerCondition* is used to choose the next task in the clinical careflow when they are connected to the previous task by the *alternativeTask* object property. This is accomplished using the *ConditionSet* class in a manner that is similar to *Option*. A *PreCondition* is slightly different in the sense that it represents conditions that must be checked before the application of tasks. Finally, the *Outcome* indicates the expected result after a *Plan*, or an *Action*, that will only be accomplished if their results are met.

## 3 Management of Clinical Uncertainty

Uncertainty may be defined as something that is not certain and transmits doubts, being an important concept in the medical domain. Indeed, a symptom may be viewed as an uncertain indication of a disease, since it may occur or not together with a certain health condition [13]. The prediction of the expected outcome of a treatment process is one of the responsibilities of healthcare professionals, and is also the moment of the clinical encounter in which uncertainty affects more the decisions. Prognosis may be defined as the prediction of the future course of a disease process that depends on the patient's health history [14]. In this case, the fundamental objective is to predict mortality after 30 days of CRC surgery. This is a critical aspect for surgeons because the death of a patient during this period is considered their direct responsibility. Moreover, this period is particularly important for the posterior recovery of patients.

### 3.1 The Case of Colorectal Cancer Prognosis

CRC develops in the cells lining the colon when they suffer mutations that cause their uncontrollable growth. They begin to invade healthy tissues, yielding malignant tumours and may also spread to other parts of the body by entering the bloodstream or the lymphatic system [20].

There are many variables that influence CRC prognosis, this being one of the reasons why this process is so problematic. The other reason is the interactions between the variables and the effect they have on the outcome, which are not entirely known and, as such, are difficult to deal with, even when one is rigorously following a CPG. For these situations, other models are necessary for the completion of guidelines and to build a complete solution for the management of diseases and treatments. From the literature it was possible to isolate a set of variables considered important for CRC prognosis and group them under two classes [15, 16]: physiological factors and operative severity factors.

The physiological factors describe the physical condition of a patient, thus including [15, 16]: age, sex, cardiac signal, respiratory signal, ElectroCardioGram (ECG) findings, systolic blood pressure, diastolic blood pressure, cardiac frequency, levels of substances in the blood (e.g. haemoglobin, leukocytes, sodium and potassium), urea levels, Dukes cancer classification and the American Society of Anaesthesiologists (ASA) physical status classification.

On the other hand, the operative severity factors include elements related with the surgery that affect the patient's recovery [15, 16]. This class consists of: pathology type, surgical urgency, surgical approach, operative severity (as classified by the surgeon), total blood loss, contamination of the peritoneal cavity, type of CRC procedure and cancer resection status (i.e., if the tumour is technically removable or not).

These were the factors used for the construction of the models presented in the next section. They were used as inputs for the models to predict the outcome expressed as 30-day mortality after surgery.

### 3.2 Developed Models

The term machine learning may be regarded as the capability of machines to identify patterns in data, use them to build a memory and then perform future tasks based on the memory they have built. This work configures a case of supervised learning, since one is trying to infer a function and make a generalization based on labelled data. The data set used corresponds to a sample of 230 patients that received surgical treatment for CRC at the Hospital of Braga. The attributes in the data set are the factors presented in the previous section, regarded as inputs for a classifier. The outcome of the classifier was considered to be the mortality within 30 days after surgery, with the possible values *yes* or *no*.

There is a number of machine learning models for supervised learning, according to the way they represent information. For this work, Bayesian Networks

(BNs) and Artificial Neural Networks (ANNs) were considered as potential solutions for modelling CRC prognosis. A Naïve Bayes classifier is a probabilistic model that uses Bayes rule and has a graphical representation in the form of a directed graph, it is characterized by the assumption that all attributes (inputs) are independent [17]. Although the independence assumption is a simplistic one for real life, this type of classifier usually performs well in actual data sets. As for ANNs, they are a mathematical approach inspired in biological neuron networks that consist in an interconnected group of artificial neurons, each one having a specific activation function. Arguably the most well known form of ANN is the Multi-Layer Perceptron which is an ANN that trains using backpropagation and consists in multiple layers containing neurons, namely an input layer connected to a variable number of hidden layers, which in turn are connected to an output layer.

Using the Classify tab in the Weka Explorer interface, a *NaiveBayes* and a *MultilayerPerceptron* classifiers were obtained. For testing purposes, 5-fold cross-validation was performed, producing the results shown in Table 1.

**Table 1.** 5-fold cross-validation results for the CRC prognosis classifiers.

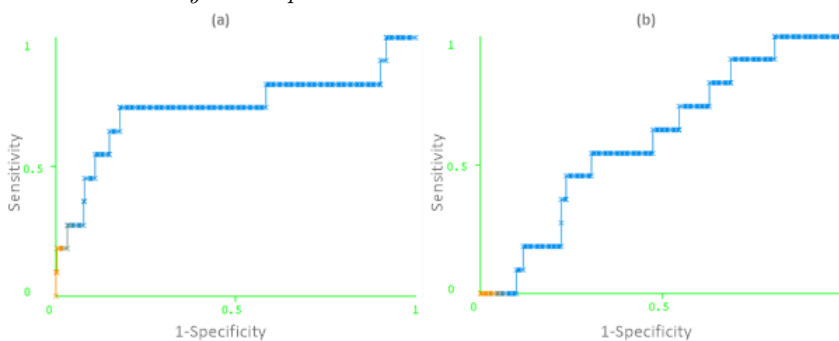
<i>Classifier</i>	<i>Kappa statistic</i>	<i>Mean abs. error</i>	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>
NaiveBayes	0.192	0.0715	0.927	0.939	0.932
MultilayerPerceptron	-0.0452	0.093	0.904	0.913	0.908

The *Kappa statistic* measures the agreement between two raters, in this case, between each classifier and the true classes in the data set [18]. This measure removes the probability of chance agreement and if a classifier has a value higher than 0, as it is the case of the *NaiveBayes*, it means that said classifier is performing better than chance. On the other hand, a value inferior to 0 means that agreement occurs less than it was predicted by chance. This happens with the *MultilayerPerceptron*, revealing a poor correspondence with reality. In turn, the *mean absolute error* summarises how close forecasts are to eventual outcomes [19] and, in this parameter, the *NaiveBayes* is associated with a lower error than its counterpart, thus indicating a minor deviation from the real values of the labels, though not by much.

*Precision* and *recall* are measures that are usually used in pattern recognition to assess model performance [17]. The first corresponds to the fraction of instances classified as positive that are true positives, while the second represents the fraction of positives that were correctly classified. The *F-measure* is the harmonic mean of the previous two, a combined score [17]. The values of Table 1 are the weighted averages of these three measures, in which the *NaiveBayes* shows an overall better performance. However, these values may be misleading once the occurrence of an actual death is a rare phenomenon, which translates into the classification classes not being approximately equally represented. In

fact, the values for *precision*, *recall* and *F-measure* for class *yes* were abnormally low, whilst for class *no* they were significantly higher, which results in a high weighted average. This class imbalance results from the difficulty to obtain data of deceased patients, evident in the sample studied for this work where there are only 11 cases of death out of 230 instances.

The results of the the Receiver Operating Characteristics (ROC) of Figure 3 are in consonance with the ones already shown. A ROC curve is a graphical plot of *sensitivity*, also called true positive rate, against *1-specificity*, or probability of false alarm, that evaluates the performance of a binary classifier. The desired result is to have low values of *1-specificity* for high values of sensitivity, i.e, the biggest possible are under the curve. As the graphics show, the *NaiveBayes* classifier is the one with a bigger area under curve, with a value of 0.795, against 0.668 of the *MultilayerPerceptron*.



**Fig. 3.** Receiver operating characteristic for (a) the *NaiveBayes* and (b) the *MultilayerPerceptron* classifiers, regarding the class *yes* represents a death during the 30 day period.

## 4 Conclusions and Future Work

This work suggests an alternative to the purely rule-based methods for clinical decision, addressing the limitations of explicit knowledge. This enables the system to tackle problems such as high complexity situations and uncertainty.

The *CompGuide* ontology deals with the definition of clinical tasks, their ordering and scheduling, in a care flow with different plans. Care flow management systems with an underlying ontology allow an advanced reasoning and the sharing of a standard representation. However, the representation of clinical information requires an inherent flexibility, given the variability of decision making processes that one may find in different medical domains. CRC prognosis is one of such cases, where healthcare professionals require more powerful tools than



simple CPG algorithms. This calls for the inclusion of models capable of representing complex and uncertain information in the procedural logics of the CIG execution engine.

Two classifiers were produced to forecast the outcome of the prognosis after CRC surgery. The *NaiveBayes* classifier was the one that showed a better performance. Being a graphical model, it is also better at delivering information to healthcare professionals. The belief network enables the users to selectively condition each entry variable and verify its impact on the outcome variable, in the form of a probability adjustment. This is more advantageous over the opaqueness of ANNs and thus the *MultilayerPerceptron*, where it is possible to view the system in terms of inputs and outputs, but not its internal workings. Healthcare professionals consider the inference process as equally valuable as the outcome. As so, it may be concluded that the Bayesian model is the best choice for integration with the care flow modelled by the ontology. Being so, it is also noticeable that the model needs refinement by extracting the most relevant features in order to make better generalizations and achieve better performances. Moreover, some pre-processing with techniques to adjust imbalanced data sets, such as the (Synthetic Minority Oversampling TEchnique) SMOTE [21], is needed in order to see if they increase the *NaiveBayes* performance.

The next steps include the improvement of the current classifiers and the development of other models of the same type for other key moments of the guideline depicted in Figure 2, such as the prediction of patient response to adjuvant therapy. The goal is to build a general solution capable of providing personalized recommendations.

### Acknowledgements

This work is funded by National Funds through the FCT - Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project PEst-OE/EEI/UI0752/2011". The work of Tiago Oliveira is supported by a doctoral grant by FCT (SFRH/BD/85291/2012).

### References

1. Musen M A , Shahar Y, Shortliffe E H (2006) Clinical decision-support systems. *Biomedical Informatics*, pp 698736.
2. Kaushal R, Shojania K G, Bates D W (2003) Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review. *Archives of internal medicine*, vol. 163, no. 12, pp. 140916.
3. Adhikari N K J, Beyene J, Sam J, Haynes R B(2005) Effects of Computerized Clinical Decision Support Systems on Practitioner Performance, vol. 293, no. 10, pp. 12231238 .
4. Rosenbrand K, Croonenborg J, Wittenberg J (2008) Guideline Development. In Teije A, Miksch S, Lucas P (eds) *Computer-based Medical Guidelines and Protocols: A Primer and Current Trends*, pp. 3-22.

5. Codish S, Shiffman R N (2005) A Model of Ambiguity and Vagueness in Clinical Practice Guideline Recommendations. *AMIA Annual Symposium proceedings*, pp. 146150.
6. Ferlay J, Autier P, Boniol M, Heanue M, Colombet M, Boyle P (2007) Estimates of the cancer incidence and mortality in Europe in 2006. *Annals of oncology: official journal of the European Society for Medical Oncology*, vol. 18, no. 3, pp. 58192.
7. McGuinness D L, Van Harmelen F (2004) OWL Web Ontology Language Overview. *W3C recommendation*, vol. 10, pp. 119.
8. Isern D, Moreno A (2008) Computer-based execution of clinical guidelines: a review. *International journal of medical informatics*, vol. 77, no. 12, pp. 787-808.
9. Oliveira T, Novais P, Neves J (2013) Development and implementation of clinical guidelines: An artificial intelligence perspective. *Artificial Intelligence Review*, doi: 10.1007/s10462-013-9402-2.
10. Ohno-Machado L. et al. (1998) The guideline interchange format. *Journal of the American Medical Informatics Association*, vol. 5, no. 4, p. 357.
11. Vier E, Fox J, Johns N, Lyons C, Rahmanzadeh A, Wilson P (1997) PROforma: systems. *Computer Methods and Programs in Biomedicine*, vol. 2607, no. 97.
12. Tu S et al. (2007) The SAGE Guideline Model: achievements and overview. *Journal of the American Medical Informatics Association*, vol. 14, no. 5. pp. 589-98.
13. Straszecka E (2006) Combining uncertainty and imprecision in models of medical diagnosis. *Information Sciences*, vol. 176, no. , pp. 30263059.
14. Lucas P, Boot H, Taal B (1998) A decision-theoretic network approach to treatment management and prognosis. *Knowledge-Based Systems*, vol. 11, no. 5, pp. 321330.
15. Horzic M, Kopljar M, Cupurdija K, Bielen D V, Vergles D, Lackovic Z (2007) Comparison of P-POSSUM and Cr-POSSUM scores in patients undergoing colorectal cancer resection. *Archives of surgery*, vol. 142, no. 11, pp. 10438.
16. Senagore A J, Warmuth A J, Delaney C P, Tekkis P P, Fazio V W (2004) POSSUM, p-POSSUM, and Cr-POSSUM: implementation issues in a United States health care system for prediction of outcome for colon cancer resection. *Diseases of the colon and rectum*, vol. 47, no. 9, pp. 143541.
17. Witten I H, Frank E, Hall M (2011) *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
18. Lantz C A, Nebenzahl E (1996) Behavior and interpretation of the statistic: Resolution of the two paradoxes. *Journal of clinical epidemiology*, vol. 49, no. 4, pp. 431434.
19. Willmott C J, Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, vol. 30, no. 1, pp. 79-82 (2005).
20. Winawer S, Fletcher R, Rex D, Bond J, Burt R, Ferrucci J, Ganiats T, Levin T, Woolf S, Johnson D, Kirk L, Litin S, Simmang C (2003) Colorectal cancer screening and surveillance: Clinical guidelines and rationale Update based on new evidence. *Gastroenterology*, vol. 124, no. 2, pp. 544560.
21. Chawla N V (2005) Data mining for imbalanced datasets: An overview. *Data mining and knowledge discovery handbook*, Springer, pp. 853867