# WAYS OF USING ARTIFICIAL INTELLIGENCE

Zyzygy Inc

# TO IMPROVE HEALTHCARE REVENUE CYCLE MANAGEMENT

**PREPARED BY**
**Edgardo Promenzio**
**Ana Maguitman**
**Axel Soto**

# Ways of Using Artificial Intelligence to Improve Healthcare Revenue Cycle Management

An AI approach can combine electronic health records, patient level information, previous claims, and publicly available data to define a tool for predicting and explaining payers' response. Besides predicting denial or rejection probability, an AI model can also provide denial reason codes, identify questionable fields in the claim and suggest possible repairs. In developing such a tool, several fundamental decisions must be taken based on the quantity and quality of the data and knowledge available to build a solution.

## Data-driven approach

A data-driven or machine learning system can extract features from claims data and other sources to build payers' response predictive models. However, very limited research work has addressed the denial or rejection prediction problem from a purely data-driven perspective. This is probably due to the fact that the required data to implement a machine learning solution is usually only accessible to a limited group of authorized people. As a consequence the lack of transparency and low reproducibility are major obstacles that prevent the advance on this research area.

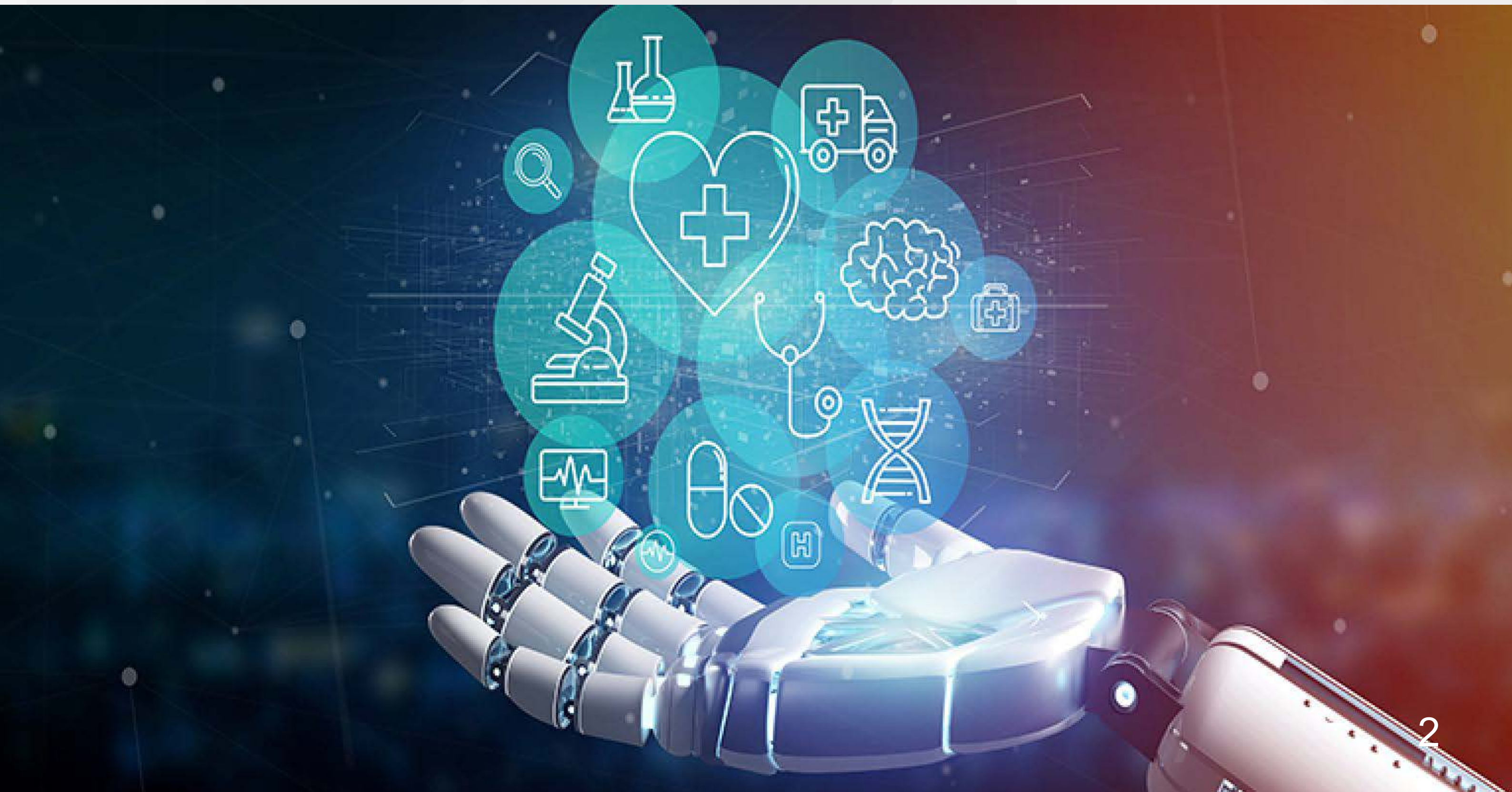
A review of the scientific literature reveals only a handful of systems that apply a data-driven approach to support the Healthcare Revenue Cycle Management process or similar processes (Kumar et al. 2010; Saripalli et al 2017; Kim et al. 2020). For instance, in (Kim et al. 2020) a deep learning approach named Deep Claim is proposed to predict payers' response to claims. The proposed framework exploits payers' raw claims data, reducing the need for expert domain knowledge and significant processing for feature extraction. Besides being able to predict payers' response, Deep Claim has the ability to identify the claim's most questionable fields that should be reviewed. Hence, the framework not only offers prediction but also explanations.

# Knowledge-driven approach

A knowledge-driven approach relies on domain-specific knowledge to perform reasoning that allows to derive conclusions. The principal power of such a system is derived from the knowledge the system possesses rather than from the adopted reasoning mechanism. The quantity and quality of such knowledge often determines the success of a knowledge-driven approach. High quality knowledge not only refers to expert knowledge that is credible, salient and legitimate, but also to knowledge represented in such a way that can be usefully exploited by a system.

Developing a knowledge-driven system to support the Healthcare Revenue Cycle Management process requires codifying expert knowledge associated with the medical billing domain as facts and rules. As discussed in (Abdullah, Ahmed, Asghar, & Zafar, 2015) the knowledge engineering needed to implement thousands of such rules manually would be a laborious task and such a system would be very hard to maintain. To address these issues our strategy exploits expert knowledge available in Current Procedural Terminology (CPT) ontologies to generate facts and rules that a system can use to anticipate rejections and denials. The CPT UMLS Metathesaurus, for instance, contains the list of codes that health care professionals use for billing with a description of the code. These codes can refer to therapeutic, diagnostic or laboratory procedures. In addition, relationships among the codes are defined, such as do_not_code_with, or has_patient_type, etc.

Likely, due to the existing constraints for the development of data-driven solutions mentioned earlier, existing systems developed to support the Healthcare Revenue Cycle Management process typically apply a knowledge-driven approach rather than a data-driven one. For instance, scrubber-based approaches (Umair et al., 2009) are traditionally applied to verify charge information in order to verify coherence in the light of the corresponding contracts between providers and payers.



## A hybrid approach

Combining data- and knowledge-driven solutions can offer some additional benefits. In particular, it is possible to apply a data-driven approach to discover new rules from past claims to complement a knowledge-driven approach. The extracted rules can be used to screen new claims with the purpose of detecting irregularities that may help anticipate rejections and denials. Also, rules learned from data offer a more dynamic way to anticipate rejections and denials than the knowledge codified in an ontology.

Another dynamic way of exploiting past claims to analyze new ones is by applying a case-based reasoning (CBR) approach. In CBR, the knowledge is not represented by rules but by stored cases recording previous experiences. As a consequence, new solutions are not generated by combining rules but retrieving the most relevant previous cases and adapting them to the new situation.

The applicability of a CBR approach to the Healthcare Revenue Cycle Management process relies on the premise that many regularities can be found in this overall process. In particular, similar claims will have similar responses. CBR also relies on the fact that the types of errors that can be encountered in a claim tend to recur. Accordingly, the outcome for similar previous claims are a useful starting point for predicting the response and suggesting a possible repair for a new claim.

Rather than building a model from a large amount of data, a CBR system indexes previous cases (past claims) and uses them as expert-knowledge to help processing new cases (claims under analysis). For a CBR approach to be successful it is key to implement an intelligent information retrieval system to be able to retrieve useful past claims given a new claim.

## Challenges and Opportunities

Each of these approaches encounters a number of challenges, which include the need not only to predict but also to explain potential claim rejections and denials, the scarcity of labeled information, the sparsity of patient data, the need to extract features from unstructured or semi-structured data, and the lack of statistically relevant data to build robust models, among others.

Some of the solutions that can help address these challenges include:

- Prediction interpretability: Determining which aspect of the claim can lead to a denial or rejection is critical in providing explanations for the prediction task. To that end, rule-based approaches offer a natural way to provide informed explanations of the reasons that can lead to a claim denial or rejection. However, providing explainability is more challenging for data-driven approaches. This is in part due to the fact that machine learning approaches often rely on non-interpretable features or non-linear models to generate predictions. In those cases where interpretable features are available, explainability relies on identifying those input features that contribute most to a denial or rejection prediction by computing specialized scores that reflect the sensitivity of the prediction outcome to the input features.

The lack of explanation of recommendations coming from AI applications is acceptable when the impacts of the decisions are minor. However, there are several cases where knowing the rationale behind a recommendation or prediction is vital or adds high value to the solution. This is particularly important in those domains where the cost of poor decisions is high and human participation is involved. Hence, in the specific scenario of Healthcare Revenue Cycle Management process, a prediction on whether a claim will be denied or rejected alone is not enough. Knowing the reasons behind this prediction is essential to provide feedback to those issuing a claim or to evaluate the possibility of repairing a claim by applying the necessary amendments.

Transparency of AI decisions are required to improve the human-AI system synergies. In particular, an AI system that opens up its reasoning to scrutiny by providing human-friendly explanations is especially useful at the moment of empowering humans to take corrective actions. In a rule-based approach, explanations come out naturally as they can usually be derived in a rather straightforward way from the facts and rules that support a derived conclusion. In the case of machine-learning approaches, there is usually an accuracy-explainability trade-off as those models that offer the higher accuracy, such as those based on neural networks (e.g., deep learning) or ensemble methods (e.g., random forest, AdaBoost and gradient boosting) tend to achieve high performance but have less explanation power. Even when we understand the underlying mechanisms of these models it is often very difficult to get insight into why a result was achieved. On the other hand, machine learning approaches such as decision trees have lower accuracy but higher interpretability.

To overcome some of these issues for a data-driven approach, some existing solutions include providing explanations at the data level or at the model level. An explanation at the data-level can include a comparison with other similar examples to justify a decision. For instance, an explanation for a rejected or denied claim can rely on presenting other similar claims that were rejected in the past. On the other hand, an explanation at the model level requires that interpretable rules be derived from a trained model or that the model itself be augmented with some form of semantic knowledge.

- Active learning: Machine learning approaches require labeled information to train a model for response prediction. Ideally, a large volume of claims with their associated response should be available to learn such a model in a supervised fashion. Because labeled information is hard to obtain, active learning can be applied to integrate data labeling and model training in a semi-supervised fashion. Active learning is an iterative process where the system strategically chooses the examples for the expert to label. This allows to reduce the labeling budget while at the same time obtaining a more useful sample of labeled data.

For an active learning approach to be applicable in the Healthcare Revenue Cycle Management process is necessary to depart from a data-driven system that is trained on some data (e.g., claims labeled as approved, rejected or denial), a knowledge-driven system that achieves a certain level of accuracy (e.g., a rule-based system that derives rules from a CPT ontology), or a hybrid system. Also, it is necessary to include a human expert in the loop, who will be in charge of labeling claims strategically selected by the system, to amend the system prediction, or to confirm it. To thoroughly exploit expert feedback, an active learning approach could request the human not only to label data, but also to provide a justification for the labeling decision. For instance, if the human determines that a claim should be denied, not only the claim must be labeled as such but the elements of the claim form that had the higher relevance for such a decision should be highlighted as this information can also be usefully exploited by the learning strategy. This approach, known as dual active learning, gives the system the additional option to ask the expert to label a particular feature as being indicative of a prediction.

- Word Embeddings: Patient representation derived from electronic health records is a way to exploit external information sources to boost prediction. Cutting-edge embedding technology can be used to build similar representations for similar patients. In this way, using embeddings for patient representation offers a vehicle to address the patient sparsity problem.

- Feature learning from unstructured or semi-structured data: While much of the data available for claim analysis is structured and can be directly used as input of an AI system, there is a significant amount of data that requires some form of natural language processing. This is the case of unstructured and semi-structured data, which is commonly found in EHR, CPT descriptors, and other relevant files. Information extraction techniques from natural language processing can be applied to identify useful features from textual data. Features coming from structured, semi-structured and unstructured data can be combined to improve the performance of prediction and explanation of an AI system.

Two main information extraction tasks are relevant in the Healthcare Revenue Cycle Management process domain, namely entity extraction and relation extraction. Entity extraction requires recognizing codes, dates, ages, and other concepts that conform a CPT ontology in the texts that are part of a claim. On the other hand, relation extraction requires identifying connections between entities. Another relevant information extraction task is coreference resolution, which allows to find links. between the detected entities (e.g. a treatment is described with two different names).

- Indexing for intelligent retrieval: A previously processed claim can be used to more effectively process a new similar claim. If errors were identified in the past claim and required rework, previous experience can help avoid problems in the new situation. When the amount of data is not statistically sufficient to build robust models, this form of case-based-reasoning approach can be usefully applied to retrieve relevant past solutions and reuse them by adapting the solution to the new case. This approach requires building highly specialized indexes for intelligent case retrieval.

# References

3M Audit Expert System. (2013, October 1). Retrieved November 1, 2015, from Health Information System Website:

http://multimedia.3m.com/mws/media/218379O/3m-audit-expert-system-fact-sheet.pdf

Abdullah, U., Ahmed, A., Asghar, S., & Zafar, K. (2015). Data mining driven rule-based expert system for medical billing compliance: A case study. In Improving knowledge discovery through the integration of data mining techniques (pp. 267–296). Hershey, Pennsylvania, USA: IGI Global.

Current Procedural Terminology (CPT). American Medical Association, 2020,Chicago, IL

Kim, B. H., Sridharan, S., Atwal, A., & Ganapathi, V. (2020). Deep Claim: Payer Response Prediction from Claims Data with Deep Learning. arXiv preprint arXiv:2007.06229.

Kumar, M., Ghani, R., & Mei, Z. S. (2010, July). Data mining to predict and prevent errors in health insurance claims processing. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 65-74).

Pocatello, I. (2015, March 10). Claim inspector. Retrieved January 15, 2016, from http://www.advancedmd.com:
http://www.advancedmd.com/uploads/files/cs-tmj-sleepcenter.pdf (non available)
(https://www.advancedmd.com/company/new-products-features/claim-inspector-pm/)

Saripalli, P., Tirumala, V., & Chimmad, A. (2017, October). Assessment of healthcare claims rejection risk using machine learning. In 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom) (pp. 1-6). IEEE.

Umair, A., Mohammad, J. S., and Aftab, A. (2009) Comparative study of medical claim scrubber and a rule based system. In Proceedings of the 2009 International Conference on Information Engineering and Computer