Claim categorization using Artificial Intelligence: a proof of concept

Two-Sentence Overview

The goal of this project is to develop an **automatic claim categorization module** for openIMIS based on state-of-the-art Artificial Intelligence (AI) algorithms and methodologies which will drastically reduce the manpower, resources and time required to review a reimbursement claim. As a contribution towards achieving Universal Health Coverage, the Swiss Tropical and Public Health Institute (Swiss TPH) designed and developed the health insurance software 'openIMIS'. Other partners joined later such as the IT firm SolDevelo for specific work areas, thus bringing together state of the art Public Health and Software development expertise.

High-Level Budget Summary

	Work Package 1 Specification and development of the AI algorithm	**Work Package 2 openIMIS AI module development**	**Work Package 3 openIMIS AI module implementation integration**	**Total Cost (USD)**
Total Project Costs				

Executive Summary

This project aims to develop an Artificial Intelligence (AI) -based claim categorization prototype application to automatically update the claim's status to accepted, rejected or to be further analyzed by a medical expert. The process will be divided into three phases:

- Research phase: Swiss TPH will work closely with GIZ Nepal to undertake the research and development of the AI algorithm for claim categorization based on anonymized openIMIS claim data and associated entities (insuree, health facility, diagnostics, medical items and services). It will be crucial to have access to a database with already categorized digitized claims (by a Medical Officer) and to identify the key input variables to be processed by the AI algorithms. The overall objective of the AI component is to obtain a claim categorization performance equal or above that of a human expert. Several classification methods, supervised and unsupervised, will be considered to fit the specific AI model to be elaborated.
- Development phase: SolDevelo will develop an openIMIS module that will integrate the AI algorithm and will create the requisite links and transitions for its activation.
- Testing phase: GIZ Nepal will run the AI module in real case scenarios to validate the previous developments.

Building on solid implementation and software maintenance expertise in the past 7 years, and extending now the partnership to Nepal Health Insurance Board/GIZ Nepal, an existing openIMIS Implementer, who will make categorized claims data available, we are confident that the automatic claims categorization module can be built and validated in the proposed time frame. The Digital Square investment would allow

us to realize the automated (and AI enhanced) claims categorization that will add substantial benefit to the health insurance system community.

Consortium Team

Swiss TPH (prime organization)

Swiss TPH is a leading institute in global health with a particular focus on low- and middle-income countries with a staff strength of over 850 staff from 80 different nations, currently active in 300 projects across 100 countries.

Swiss TPH will be responsible for technical project management, the expertise on health financing, openIMIS, the development of the AI algorithm, drafting of business and technical specifications, supporting the system and architecture design.

Swiss TPH's relevant experience includes involvement in the design and implementation of the Insurance Management Information System (which is the genesis of the openIMIS Initiative) since its inception in Tanzania and has supported its implementation in a number of countries. Swiss TPH is currently implementing two projects at scale for the deployment of insurance schemes through openIMIS, in Tanzania and Cameroon, and two openIMIS pilots in Chad and Democratic Republic of Congo. In addition, Swiss TPH is actively involved in the development of openIMIS as part of the Implementers and Developers Committees of the openIMIS Initiative. Further, Swiss TPH is also implementing projects with AI based on supervised machine learning to improve Clinical Decision Support Systems.

Qualifications of key members of the proposed project team:

- Dragos Dobre, PhD
    - IT System Architect at Swiss Tropical and Public Health Institute
    - Management of software life cycle (from specification to development to deployment)
    - Design, development and maintenance of open source applications
    - openIMIS design, development and implementation experience since 2018 in implementation sites of openIMIS - Tanzania, DRC and Chad
    - Project Coordinator experience across international teams for GIZ openIMIS mandate
    - OMG-Certified Systems Modeling Professional™

SolDevelo

SolDevelo is a dynamic IT company based in 2009 (+80 staff) and focused on delivering high-quality software and innovative solutions.

SolDevelo will be responsible for software integration of the AI module into openIMIS.

SolDevelo is currently involved in several openIMIS projects, including HL7 FHIR module development, openIMIS integration with OpenMRS, and maintenance, support and enhancing the security of the Microsoft solution. SolDevelo has been involved in many opportunities that required skill sets relevant to this particular project, especially through opportunities like OpenMRS (core contributors), HL7 FHIR implementation (OpenMRS Sync 2.0 module), nationwide micro-service based implementations (OpenLMIS), nationwide OpenHIE architecture based implementations (National Health Infrastructure project with such components like OpenELIS, DHIS2, OpenMRS and many other HIE compatible applications, health standards-based workflows for the Client Registry, Facility Registry, Health Management Information System, Shared Health Record, and Interoperability Layer).

Qualifications of key members of the proposed project team:

- Kamil Madej
    - Senior-level Java Developer/Team Leader
    - Working in international teams for various projects/clients, like:
        - openIMIS
        - OpenMRS
        - MOTECH
        - Terre des hommes
        - Connect for Life
    - Performing code review
    - Creating high level designs using tools for wireframing
    - Leading several frontend and backend development teams

Nepal Health Insurance Board / GIZ Nepal

Nepal Health Insurance Board is the implementer of openIMIS in Nepal from 2016.
Together with GIZ Nepal, they will be responsible for the appropriation of categorized claim data, the validation of the anonymized data set for AI analytics and the testing of the openIMIS AI module developed under this project.

Background or Problem Statement

Substantial manpower, time and resources are deployed in the process of claims management for cost refunds or claim audits. Considering the respective insurer policies and regulations, this process results in a) a payment of the claim, b) (partial) rejection of the demand or c) further analysis by a Medical Officer.

Currently, static validation rules based on insurance product configuration are implemented in openIMIS. They are 'hardcoded' in the system and applied on claims submission. However, these validation rules only allow claims to be accepted or rejected and require the subsequent review by a professional Medical Officer.

For example, the openIMIS implementation in Nepal receives up to 7.000 claims per day. From all these claims, the team composed of five Medical Officers reviews only 1.000 claims. This increases the risk factor for abuse, and perhaps fraud, in healthcare.

Once the database of categorized claims is available, the automated claims categorization (based on AI) enables the encoding of the Medical Officer's knowledge into a model that can automatically check future claims, and thus allowing the Medical Officer to concentrate on those claims that really need to be reviewed such as inconsistent, erroneous and fraud/abuse claims.

For the AI algorithm development, several methods will be considered [1,2,3,4]:

- Supervised methods will attempt to discover relationships between claim entities (e.g. price of a drug or service, the age and sex of the patient, the diagnosis, etc.) and the output variable. Methods such as neural networks, decision trees, Support Vector Machine are known to be applied in the insurance sector.
- Unsupervised learning methods will allow to assess one claim attribute in relation to other claims to determine the similarities between them and thus will be able to distinguish anomaly records or group of similar records using methods like 'anomaly detection', 'association rules' or 'clustering'.

The work will take place in different places (Nepal, Switzerland, Poland) and will have a common test environment in Nepal infrastructure. It is an exploratory project that will improve openIMIS and increase adherence to new implementation sites.

Digital Health Technologies

Nowadays, AI algorithms can address decision-making and detection tasks related to different health issues (diabetes, health problems, asthma, etc.) and is highly favorable to patient engagement. Furthermore, it can improve the access to care by enhancing the dynamic communication between patients and physicians. In that direction, the automatic process of claim management could improve access to care, by reducing the payment delay.

The aim of this project is to improve the open source Insurance Management Information System (openIMIS) digital tool with the use of AI.

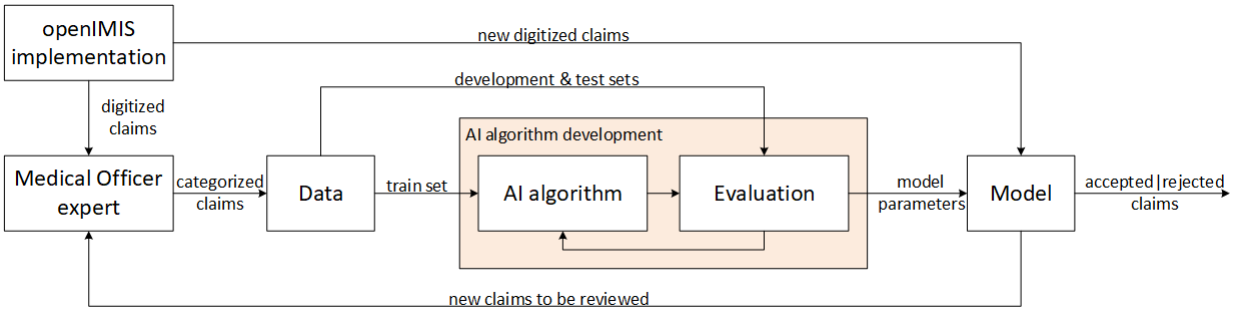
openIMIS is a comprehensive system for managing a health insurance scheme (enrollment, renewal, claims management, feedback, reporting). We will focus on Claim review process and will do the development on the new modular architecture developed in Python with Django framework for the backend and Javascript with React framework for the frontend.

The new developments will integrate the AI based algorithm (will be defined in the research phase) which will classify digital claims, recorded in openIMIS. The AI algorithm will be developed in Python and will be able to build, train and validate models.

Together with GIZ Nepal, we will follow the Principles for Digital Development and involve end-users into the specification, development and testing of the proposed solution.

Use Cases and User Stories

The following use case will be covered by the developments for this proposal.



Process steps:

1. Once the openIMIS software processes insurance data, it will start the claim adjudication process based on the scheme-specific configuration.
2. A Medical Officer expert will manually categorize (review) the digitized claims as accepted or rejected.
3. When sufficient claims have been gathered, the openIMIS AI module will be activated and subsequently splits the data into training, development and test sets to feed into the existing AI module for training and evaluation. This will result in assessment parameters (output) specific to the (implementation) usage.
4. The resulting adapted AI model will be then used to categorize new digitized claims to accepted, rejected and to be reviewed by a Medical Officer expert. Moreover, these manually reviewed claims will then be fed back into the system and improve the new AI model (steps 2 and 3).

Objectives and Activities

To achieve the described objectives, the activities will be grouped into three work packages.

Work package 1: Specification and development of the Artificial Intelligence algorithm (research phase)

Objective 1.1: Data gathering and preparation

Activity 1.1.1: Required data fields selection

Discussions are mandatory between partners in order to understand the claim review process, openIMIS existing database and necessary entry models and attributes necessary to feed to the AI algorithm. Exploration of the quality of the categorized digital claim database, to avoid duplicates, incorrect/ inconsistent/missing values.

Output: AI input data model structure

Activity 1.1.2: Claims data anonymization

Based on the required data and on the established data privacy standards, we will develop a script (code) that will anonymize the implementer's already categorized claims data being used as the input data during the AI algorithm development.

Output: anonymization script and anonymized AI input data

Activity 1.1.3: Data preparation

Visualization of the anonymized database, normalization, randomization and splitting the database into train/development/test sets.

Output: normalized input data sets

Objective 1.2: Implementation of the AI algorithm

Activity 1.2.1: AI algorithm selection

Several algorithms specific to health insurance applications are documented in the literature and are relevant for the proposed goal. Unsupervised and supervised methods can be applied and the choice will depend on the existing database and specific objectives. Several models can be selected to compare their performances based on a selected evaluation metric. This step includes the definition of the output model.

Output: AI methods, model outputs and evaluation metric

Activity 1.2.2: Building the AI algorithm

This step will be dedicated to the development of the AI algorithm (based on the selected methods, model outputs, and evaluation metric), train the model (estimation of the model parameters based on the training data set), tuning the hyperparameters (on the development data set) and make predictions (on the test data set). A comparison of the performance obtained on the train/development/test sets will allow deciding if there are improvements to be made (in terms of dimensions of the model, hyperparameters values, optimization algorithms, etc.).

Output: AI code and model parameters

Work package 2: openIMIS AI module development
(development phase)

Objective 2.1: Specification of AI algorithm integration

Activity 2.1.1: Claim adjudication process update

The AI algorithm and model will be integrated with the openIMIS 'Claim Module' and will replace the manual review process. During this activity, we will define the new openIMIS Claim adjudication process (including the automated claim categorization).

Output: AI based Claim adjudication process

Activity 2.1.2: openIMIS AI module specification

Specification of the openIMIS AI code integration will be developed: the database structure extension, the openIMIS AI module specification, the events that will activate the AI code, etc.

Output: AI module specification

Objective 2.2: openIMIS AI module development

Activity 2.2.1: openIMIS AI module development

Based on the AI module specification, a new module will be implemented into the new openIMIS modular architecture based on Python with Django and JavaScript with React software development frameworks.

Output: openIMIS AI module

Activity 2.2.2: openIMIS AI module validation

The new openIMIS AI module will be validated on new claims with Medical Officers involved in the research phase.

Output: validated openIMIS AI module (code and model parameters)

Work package 3: openIMIS AI module implementation integration
(testing phase)

Objective 3.1: User Acceptance Testing (UAT) of the AI module

Activity 3.1.1: Installation of AI-based openIMIS

We will install the AI-based openIMIS implementation in the Implementer's testing environment and connect to the Implementer's testing database. The initial parameters (output from *Activity 2.2.2*) for the AI categorization model will be configured in the AI module.

Output: AI based openIMIS testing environment

Activity 3.1.2: Organization of the UAT

The UAT process will be specified (testing plans) and the required resources will be defined. Any issue reported will be solved.

Output: UAT test plans and reports

In all these phases, Swiss TPH will closely monitor the advancement of the project and report periodically to Digital Square and to the openIMIS Initiative.

Community Feedback

The topic of Artificial Intelligence is still underused by digital development and open source communities. We want to change that, by providing easy-accessible regular reports from our research phase. Our experiences and findings should help other initiatives to start working with this technology. In addition, through our contact email and comment section, we would be able to hear the opinion of others on our work and eventually incorporate some of their feedback into our project. We would also like to create a case study at the end of the project, that would be published on various blogs and websites for digital development specialists (like, for example, ICTWorks).

Schedule

The following is a high-level work plan.

Activity	Team	Month								
		Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
		1	2	3	4	5	6	7	8	9
1.1.1. Required data fields selection	Swiss TPH	x	x							
1.1.2. Claims data anonymisation	Swiss TPH		x							
1.1.3. Data preparation	Swiss TPH			x						
1.2.1. AI algorithm selection	Swiss TPH			x						
1.2.2. Building the AI algorithm	Swiss TPH				x	x	x			
2.1.1. Claim adjudication process update	Swiss TPH							x		
2.1.2. openIMIS AI module specification	Swiss TPH							x		
2.2.1. openIMIS AI module development	SolDevelo							x	x	
2.2.2. openIMIS AI module validation	Swiss TPH								x	
3.1.1. Installation of AI based openIMIS	SolDevelo								x	
3.1.2. Organisation of the UAT	Implementer									x

Deliverables

Deliverable	Month Due
AI input data model structure	January 2020
Anonymization script and anonymized AI input data	January 2020
Normalized input data sets	February 2020
AI methods, model outputs and evaluation metric	February 2020
AI code and model parameters	May 2020
AI-based Claim adjudication process	June 2020
AI module specification	June 2020
openIMIS AI module	July 2020
Validated openIMIS AI module (code and model parameters)	July 2020
AI-based openIMIS testing environment	July 2020
UAT test plans and reports	August 2020

Global Good Maturity Model Assessment

Please review the updated link.

References

[1] Li, J., Huang, KY., Jin, J. and Shi, J. A survey on statistical methods for health care fraud detection, Health Care Manage Science, 11 (3), pp. 275-287, 2008
[2] Ortega, P.A., Figueroa, C.J., and Ruz, G.A. A medical claim fraud/abuse detection system based on data mining: A case study in Chile, DMIN'06, The 2006 International Conference on Data Mining, Las Vegas, Nevada, USA, June 26-29, pp. 224-231, 2006
[3] Obenshain, Mary K. Application of data mining techniques to healthcare data. Infection control and hospital epidemiology, 25(8), pp. 690-695, 2004
[4] Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M., & Arab, M. Using data mining to detect health care fraud and abuse: a review of literature. Global journal of health science, 7(1), pp. 194–202, 2014