



HAL
open science

A SmartNextGISSA for Monitoring and Predicting of Comorbidities

Daniel Fiuza, Francisco Junior, Fabio Gomes, Ivana Cristina, Odorico Monteiro, Mauro Oliveira

► **To cite this version:**

Daniel Fiuza, Francisco Junior, Fabio Gomes, Ivana Cristina, Odorico Monteiro, et al.. A Smart-NextGISSA for Monitoring and Predicting of Comorbidities. 10th International Workshop on ADVANCEs in ICT Infrastructures and Services (ADVANCE 2023), Federal University of Ceara, University of Evry, Feb 2023, Fortaleza-Jericoacoara, Brazil. 10.48545/advance2023-fullpapers-3_1 . hal-04083150

HAL Id: hal-04083150

<https://hal.science/hal-04083150v1>

Submitted on 27 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A SmartNextGISSA for Monitoring and Predicting of Comorbidities

Daniel Fiuza¹, Francisco Junior¹, Fabio Gomes², Ivana Cristina³, Odorico Monteiro⁴, Mauro Oliveira²

¹ Federal Institute of Ceará - IFCE, Fortaleza, CE, Brazil,

² Federal Institute of Ceará - IFCE, Aracati, CE, Brazil,

³ Oswaldo Cruz Foundation - Fiocruz, CE, Brazil,

⁴ Federal University of Ceara - UFC, Fortaleza, CE, Brazil,

jwebsys@gmail.com, danielfiuzacm@gmail.com,
prof.fabiojose@gmail.com, ivana.barreto@fiocruz.br, odorico@ufc.br,
amauroboliveira@gmail.com

Abstract

A casual or late diagnosis given the lack of knowledge of the severity of a problem or asymptatology of patients with comorbidity can cause serious consequences for the patient. Therefore, there is an urgent need for decision-making systems based on continuous monitoring in real time. This research presents SmartNextGISSA, an intelligent mechanism used in NextGISSA, an architecture developed with the objective of assisting decision-making in a situation room in monitoring patients with comorbidity, based on the crossing of data collected in real time from this patient and his health history. NextGISSA predictively and continuously monitors in real time patients with comorbidities (Pregnant Women, Hypertensives, Diabetics and Bedridden) of GISSA, an intelligent governance platform for decision-making in Digital Health, a product already operational in several municipalities by the startup “Avicenna Intelligent Governance”. With the intention of validating the machine learning module of the architecture proposed in NextGISSA, the SmartNextGISSA module was implemented, as an emphasis of this article, which is characterized by a composition of the NextGISSA architecture responsible for the inference about the data of an individual classifying it with different comorbidity risk levels through machine learning algorithms. Thus, we present a development pipeline that consist of a comparative analysis between binary classification algorithms, such as Naive Bayes, Decision Tree and Random Forest, using the CRISP-DM methodology, a concept applied to data mining processes and, later, the model was implemented in a web application.

1 Introduction

Scientific and technological development has provided means of using digital information and communication technologies (TDIC), widely disseminated in the most diverse instances of society. Thus, many areas, such as health, have taken advantage of the benefits of these technologies, such as Machine Learning, applied to management, evaluation and analysis systems.

In the context of the Covid pandemic, the term comorbidity has become commonplace on a daily basis. The name is given to a set of causes that aggravate a disease and, in this way, can worsen a clinical condition. This study is justified, since a casual or late diagnosis given the lack of knowledge of the seriousness of the problem or even asymptatology of patients with comorbidity urges the taking of measures with a fruitful and continuous follow-up in real time.

Machine learning can target different users with some comorbidity. With the intention of validating the machine learning module of the architecture proposed in NextGISSA, the SmartNextGISSA module was implemented, as an emphasis of this article, which is characterized by a composition of the NextGISSA architecture responsible for the inference about the data of an individual classifying it with different comorbidity risk levels through machine learning algorithms.

As for the choice of the research universe - population with comorbidity, it is known that this public is one of the priorities in care within the Primary Care policy of the public health network.

In view of this, the object of study is problematized through the following guiding question of this research: How to validate the machine learning module of the architecture proposed in NextGISSA from a scenario to monitor predictively and continuously in real time and in a way smart the patients with comorbidity found in the report of enrolled people (Pregnant Women, Hypertensive, Diabetic and Bedridden) of GISSA?

Thus, this article details the machine learning process applied to a dataset of individuals with hypertension to evaluate the application of supervised machine learning methods, specifically, binary classification algorithms, and implement them in the SmartNextGISSA module in order to validate the architecture proposed in NextGISSA.

Therefore, the main contribution of this article includes validating the proposed implementation of the SmartNextGISSA module embedded in the NextGISSA architecture in order to provide a predictive model capable of inferring an individual's risk of comorbidity.

Thus, this article is structured as follows: Section two presents related works that use artificial intelligence are presented. In section three, the methodology used in the data mining process for the elaboration of the model. Section four presents an application scenario for SmartNextGISSA. In section five, the results obtained are presented and discussed and, finally, in section six, the conclusion and future aspirations of this article.

As for the choice of the research universe - population with comorbidity, it is known that this public is one of the priorities in care within the Primary Care policy of the public health network.

In view of this, the object of study is problematized through the following guiding question of this research: How to validate the machine learning module of the architecture proposed in NextGISSA from a scenario to monitor predictively and continuously in real time and in a way smart the patients with comorbidity found in the report of enrolled people (Pregnant Women, Hypertensive, Diabetic and Bedridden) of GISSA?

Therefore, the main contribution of this article includes validating the proposed implementation of the SmartNextGISSA module embedded in the NextGISSA architecture in order to provide a predictive model capable of inferring an individual's risk of comorbidity.

Thus, this article is structured as follows: Section two presents the methodology used in the data mining process for the elaboration of the model. In section three, related works that use artificial intelligence are presented. Section four presents an application scenario for SmartNextGISSA. In section five, the results obtained are presented and discussed and, finally, in section six, the conclusion and future aspirations of this article.

2 Related Works

BRAGA, Oton Crispim (2017) proposed an intelligent solution based on classifiers as an inference mechanism, capable of helping health professionals during the process of clinical management of diseases transmitted by the *Aedes Aegypti* mosquito, identifying the diagnosis based on symptoms and test results. The work was divided into two steps: one focused on pre-diagnosis, considering symptoms and clinical history, anamnesis; and another focused on the final diagnosis, also considering the results of specific tests, such as serology tests. The study uses a methodology based on Data Mining to guarantee knowledge based on examples. After several tests and adjustments in machine learning algorithms, it was possible to define two learning models capable of inferring the probability of a patient being infected with a certain disease, with an accuracy of up to 91.6%. From these models, they were able to build an intelligent API to support decision-making during the clinical management of dengue and chikungunya. The solution allows several applications to access the learning models. Among them, a popular consultation mobile application for the identification of dengue and chikungunya, MARCIA, an interoperable system for the clinical management of chikungunya.

Costa Filho (2021) proposes Smart GISSA, which is an architecture built from the Intelligent Governance System in Health Systems – GISSA, an innovative solution in the health area already highlighted in this work. This solution provides managers, servers and users with intelligent governance by promoting the systemic integration of information. Smart GISSA's architecture, in addition to extracting, transforming and loading data into dashboards, includes machine learning models to specialize decision-making in public health, in which two new Data Mining methodologies are proposed, focusing on risk analysis of death and in epidemiological surveillance to predict epidemics. It was possible to verify that this architecture already proposed to GISSA by Costa Filho (2021) does not use IOT and is not aimed at the hypertensive public.

In this way, the GISSA platform has undergone continuous evolution since its conception in 2009, as a result of research, development and innovation. Thus, research has been incorporated into GISSA, notably in the recent master's research (Francisco Junior, 2022).

From the study carried out on the related works and the proposed solutions, it was observed that almost all of the works directly related to the subject under study, Machine Learning, are aimed at different situations of case studies. Thus, this study arose from the need and possibility of converging Machine Learning and any comorbidity (hypertensive, diabetic, cardiac) on an existing platform, adding value to the architecture, in this case, of GISSA, in an expanded architecture.

3 Methodology

The research takes place from the performance of GISSA (Gissa, 2022), having as spatial delimitation the city of Fortaleza. The GISSA platform operates in a datacenter, in the cloud, provided by the company Avicena Software e Serviços Ltda. It can be accessed through the web, by computer terminals, smartphones and tablets and proactively communicating through the use of the system, accessing panels (dashboards), with access to summarized and analytical information, filterable, weighted, classified according to risks and compliance with standards quality for services provided, reporting and sending alerts and categorized messages.

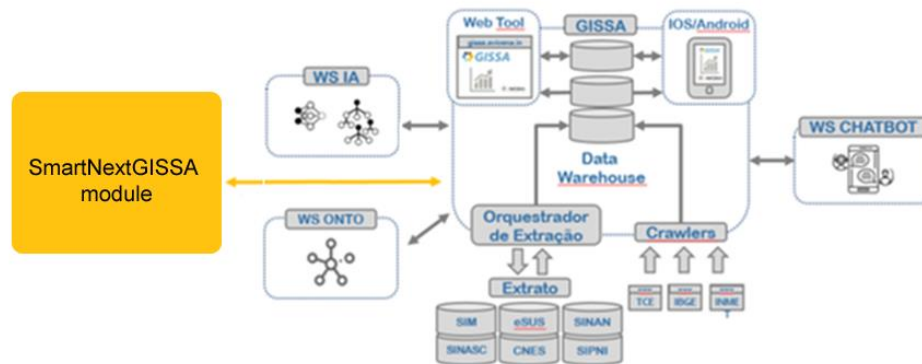


Figure 1: Proposal of the SmartNextGISSA System in the GISSA architecture

Considering that the pipeline was developed based on the CRISP-DM methodology (SCHRÖER, Christoph; KRUSE, Felix; GÓMEZ, Jorge M., 2021), the process was segmented into phases as proposed by the concept:

1. Understanding the Business: This step was characterized by understanding the problem based on the research carried out associated with the business domain of the NextGISSA platform, leading to the search for a dataset that corresponds to the scenario covered by NextGISSA.
2. Understanding the Data: Considering the definition of the problem and the compatibility of the data with the NextGISSA platform, the dataset was composed with health data from individuals characterized with 11 predictive variables (age, gender, height, weight, blood pressure, cholesterol, glucose level, smoker, alcoholic and active) and are classified into two groups of individuals (hypertensive and non-hypertensive). Data were obtained from resources available on the Kaggle platform and then exploratory data analysis was applied in order to describe and qualify them through statistical measurements. Among the analyses, the quality of the data was verified (presence of missing values, outliers, amount of data, among others) and the balance of data, indicating that they are approximately equally distributed between the two groups, with 34,979 cases of cardiac and 35,021 non-cardiac cases, totaling a sample of 70,000 cases.
3. Data preparation: This step consist of pre-processing the data, summarized by the steps of selecting columns and rows of interest (filtering), cleaning the data (removing missing values and outliers) and formatting the data (conversion of values categorical values into numerical values). After the pre-processing step, the data were normalized in order to transform the attributes into a similar scale.
4. Modeling: The process of defining the model is based on the pipeline prepared by Costa Filho (2021), since binary classification algorithms can be applied to the data set used in this article. Thus, some steps were performed to identify and evaluate the model to be defined, among them, the data loading and pre-processing, exploratory data analysis, hyperparameter optimization. The developed flow is presented in Figure 2. In the training stage, the binary classification algorithms, Logistic Regression (LR), Random Forest (RF) and Decision Tree (DT) were applied using the Scikit-learn library. After training, the model was evaluated based on the main metrics, accuracy, precision, recall, f1 and the AUROC curve (Area Under ROC Curve). The features of the dataset are (F1) age in days of the individual, (F2) gender, (F3) height, (F4) weight, (F5) blood pressure, (F6) cholesterol level, (F7) glucose level, (F8) smoker, (F9) alcoholic and (F10) physically active.

- Evaluation: In this stage, the performance of the supervised classifiers used in the training stage was evaluated. Then, the GridSearch strategy associated with the K-Fold Cross Validation (CV) method was applied, thus allowing the identification of the best parameters from combinations of hyperparameters.

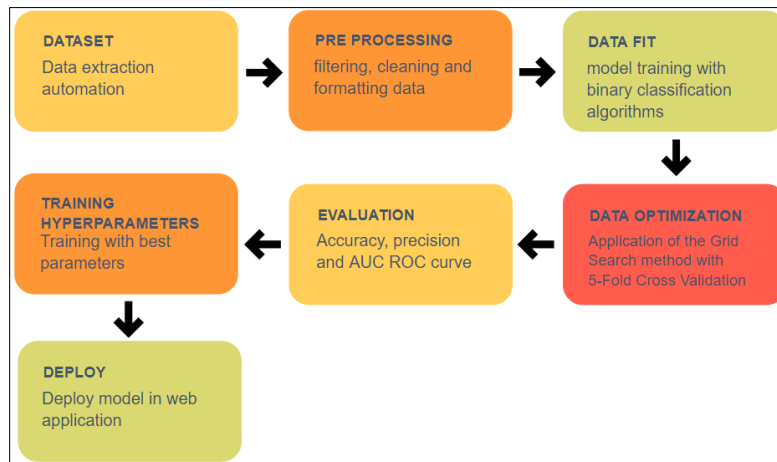


Figure 2: General flow for defining the model.

For the RF and DT classifiers, hyperparameters, number of estimators and maximum number of features were used, as shown in Table 1, while in LR penalty, C and solver from the Scikit-learn library were used, according to Table 2.

Parameters	Description	Tested Values
n_estimators	Number of trees in the forest	10, 60, 110, 160
max_features	Number of features to consider when looking for the best split	6, 11, 16, 21

Table 1 - Random Forest and Decision Tree classifier evaluation parameters

Parameters	Description	Tested Values
penalty	Specify the norm of the penalty	l1, l2
C	Inverse of regularization strength	0.0001, 0.01, 1, 100, 10000
solver	Algorithm to use in the optimization problem	liblinear

Table 2 – Logistic Regression classifier evaluation parameters

- Implementation: After evaluating the model from the AUROC curves and performance metrics, its parameters were serialized and made available in a Rest API destined to a web application for validating the SmartNextGISSA module. In this context, this module's web service consumes the API that performs the inference based on the trained model. For this, an endpoint was made available that receives a POST request according to the format:

```
POST https://<host>/predict
```

```
{
  "input": "[ 40, 140, 289, 172, 0, 0, 1, 0, 1, 0, 0,1, 0, 0, 1, 0,
1,0,0,0,1]"
}
```

The API returns, through the POST response, the value of the class (0 for non-hypertensive and 1 for hypertensive).

```
Response: [{"result": 1}]
```

Intelligent mechanisms are used to monitor vital data (weight, blood pressure, heart rate) and check these data against the patient's history through artificial intelligence algorithms.

4 SmartNextGISSA Architecture

This work concerns the use of Intelligent Systems in dealing with comorbidities, expanding the functionality of GISSA, a Public Health Governance product already in operation in several municipalities in Northeast Brazil (GISSA Manual, 2021). GISSA is a computing platform that automatically collects the databases of the Ministry of Health systems: E-SUS, CNES, SIM, SINASC, SI-PNI, SINAN through data extractor robots, analyzes this information using various intelligent and transforming them into integrated technologies information , useful for decision-making at different levels of health management.

Named SmartNextGISSA, the proposal presented here includes an intelligent module on the GISSA platform for monitoring and predicting cases of comorbidities, notably patients enrolled in the GISSA system (diabetic, hypertensive), in the city of Fortaleza-Ceará, as shown in figure 3.

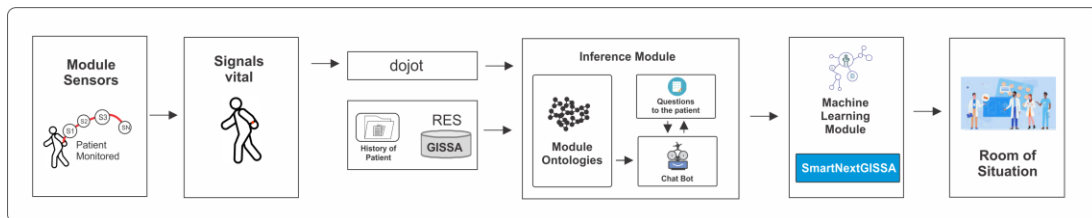


Figure 3: SmartNextGISSA Architecture

The SmartNextGISSA is proposed as an additional component to GISSA. In this context, the service will operate in the Machine Learning module of the NextGISSA scenario, receiving the history of patients from the databases monitored by the GISSA platform in order to filter individuals at potential risk through the inference of the trained model, thus allowing, that these cases move to the next stage of this scenario.

5 Results

Several experiments were carried out following the CRISP-DM methodology in order to optimize the model's performance. In the comparative performance analysis between the algorithms used, the

Logistic Regression (RL) classifier obtained better accuracy in relation to the others, as shown in Figure 4.

	Accuracy	Precision
Logistic Regression	0.719905	0.678717
Random Forest	0.716810	0.704431
Decision Trees	0.636381	0.643609

Figure 4: Accuracy and Precision

Accuracy represents the percentage of examples correctly categorized and is associated with the evaluation of the model's performance in all classes. This metric is accepted for the purpose of this work, since the dataset data are balanced and it is expected that the model presents assertive results.

In the evaluation, the AUROCC curve was also used, which represents the general performance of an estimator, since this metric considers all computed values of sensitivity and specificity. The greater the capacity of the estimator to discriminate individuals with and without hypertension, the more the curve approaches the upper left corner and the AUROCC will approach 1. In the application of the Grid Search optimization strategy associated with the Cross Validation (CV) technique, the Regression Logistic (RL) classifier was obtained as the best performance with the following parameters: $C= 1.0$, $penalty = l1$ e $solver = liblinear$.

After training with the mentioned values, improvements were obtained in the following metrics: accuracy = 0.720381, precision: 0.678812

The area under the ROC curve for this parameterization is shown in Figure 5 and presented as a value of 0.7220.



Figure 5: The area under the ROC curve

6 Conclusion

AI-based technologies using machine learning have the potential to transform healthcare. In practice, applications are used for more accurate diagnosis, identification of physiological patterns, decision support for processes in personalized medicine, etc. In the case of study, it is proposed to use it for patients with comorbidities, such as hypertension.

This article presented SmartNextGISSA, an intelligent system for the prediction of comorbidities, via monitoring of the patient's vital health signs, from their clinical history. SmartNextGISSA offers a monitoring environment by automating some services such as obtaining historical patient data, in addition to expanding the range of GISSA functionalities by including an intelligent module for monitoring and predicting comorbidities. To carry out the proof of concept of the proposed solution, a prototype will be built to carry out the tests in order to evaluate the effectiveness of SmartNextGISSA in carrying out online monitoring, integrating solutions that complement each other to signal alerts to the patient about their health status in case of any intercurrentence.

As future work, we will assess monitoring for other comorbidities in addition to the report of people enrolled in GISSA, aiming to expand its scope and also identify improvements in the proposed solution.

References

- AL-AZZAM, M. K.; ALAZZAM, M. B. Smart city and Smart-Health framework, challenges and opportunities. *International Journal of Advanced Computer Science and Applications*, v. 10, n. 2, p. 171–176, 2019.
- BRAGA, Oton Crispim. Solução Inteligente Baseada em Aprendizado de Máquina para a Classificação de Doenças transmitidas pelo vetor *Aedes Aegypti*. Trabalho de Conclusão de Curso (TCC) apresentado ao curso de Bacharelado em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE - Campus Aracati. 2017.
- COSTA FILHO, Raimundo Valter. Smart-Gissa, Um Sistema Para Governança em Saúde Digital Baseado em Aprendizado de Máquina. Tese apresentada ao Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Ciências e Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Doutor em Engenharia de Teleinformática. FORTALEZA.2021.
- EL ATTAOUI, A., Largo, S., Jilbab, A. et al. Wireless medical sensor network for blood pressure monitoring based on machine learning for real-time data classification. *J Ambient Intell Human Comput* 12, 8777–8792 (2021). <https://doi.org/10.1007/s12652-020-02660-1>
- FREITAS, Nicodemos. Predição de Casos de Arboviroses na Cidade de Fortaleza-CE. Utilizando Internet das Coisas e Aprendizado de Máquina/ Nicodemos Freitas. – 2021. Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Ciência da Computação, Fortaleza, 2021.
- STEPANIAN, Robert S.H. and LACAL, Jose C. Emerging Mobile Communication Technologies for Health: Some Imperative notes on m-health. Conference: Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE. October, 2003.
- JUNG, E.-Y. et al. Mobile healthcare application with EMR interoperability for diabetes patients. *Cluster Computing*, Sep. 2014. Volume 17, PP 871-880..
- Gissa - Avicena - Governança Inteligente. Avicena, 2022. Disponível em: <https://avicena.in/gissa/>. Acesso em: 20, Novembro, 2022.

SCHRÖER, Christoph; KRUSE, Felix; GÓMEZ, Jorge M. A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*. v. 181, p. 526-534, 2021.