








# Supplementary Material - AmazonForest: In-silico meta-prediction of pathogenic variants

Helber Palheta <sup>1</sup> , Wanderson Gonçalves e Gonçalves <sup>1,2</sup> , Leonardo Brito <sup>1</sup> , Arthur Ribeiro dos Santos <sup>1</sup> , Marlon dos Reis Matsumoto <sup>1</sup> , Ândrea Ribeiro-dos-Santos <sup>1,2</sup>  and Gilderlanio Santana Araújo <sup>1,\*</sup> 

<sup>1</sup> Laboratory of Human and Medical Genetics, Institute of Biological Sciences, Graduate Program of Genetics and Molecular Biology, Federal University of Pará, Belém 66075-110, Brazil; gilderlanio@gmail.com

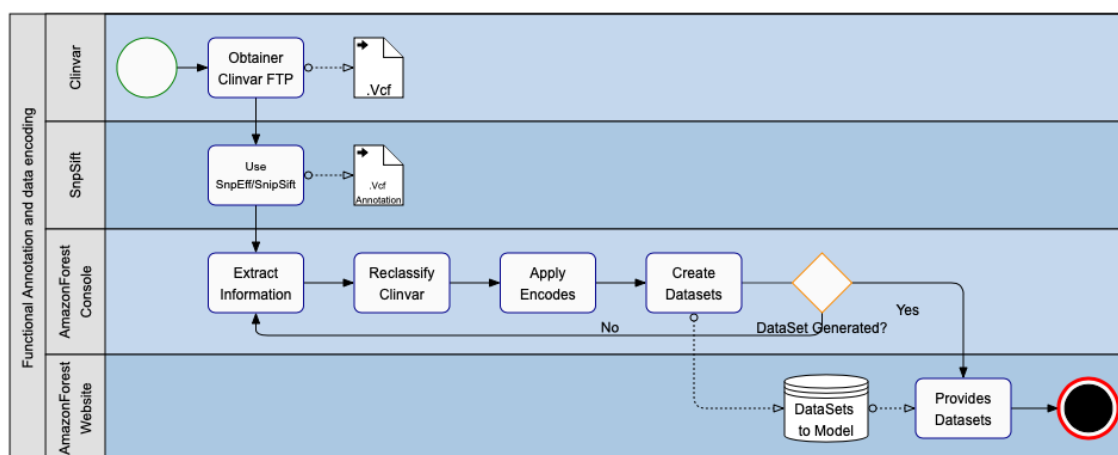
<sup>2</sup> Research Center on Oncology, Graduate Program of Oncology and Medical Science, Federal University of Pará, Belém 66073-000, Brazil; akelyufpa@gmail.com

\* Correspondence: gilderlanio@gmail.com

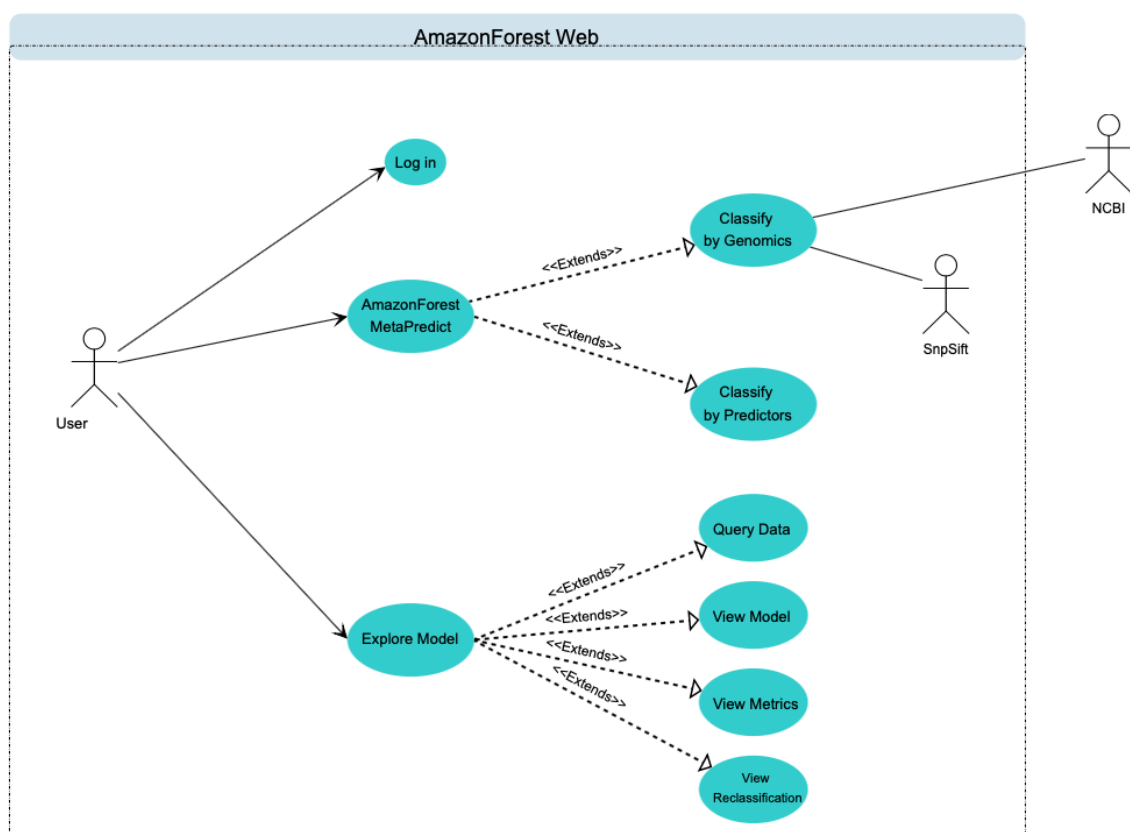
Version November 18, 2020 submitted to Journal Not Specified

**Abstract:** Here we detailed the business process model and notation, use cases and usage examples of AmazonForest.

## 1. AmazonForest: use cases, web architecture and business process model and notation



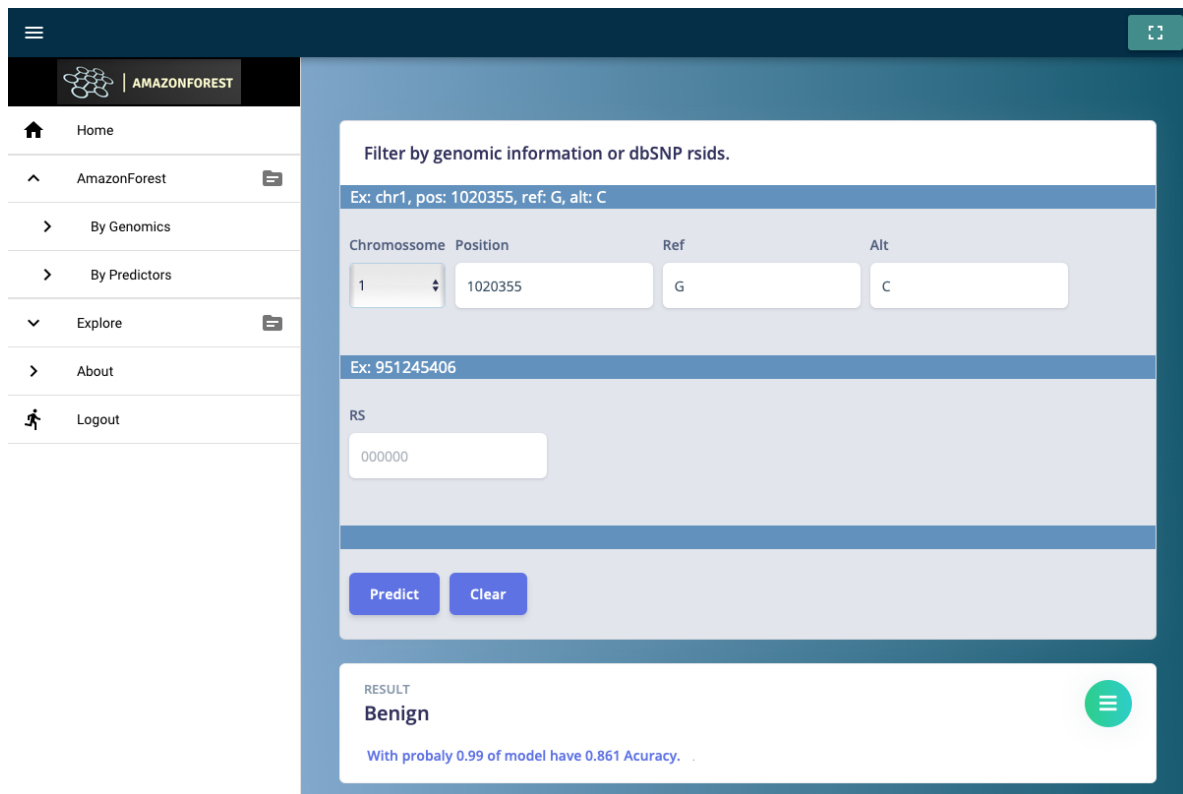
**Figure 1.** AmazonForest data flow starts by extracting ClinVar .vcf file available via File Transfer Protocol (FTP). SnpSift/SnpEff is used for functional variant annotation of ClinVar .vcf file for eight in silico predictors. AmazonForest extracts information from .vcf annotated file and performs relabel of ClinVar pathogenic classification. Data transformation is performed by encoding strategies. Finally, the datasets are available on the web system for further training and reclassification analysis.



**Figure 2.** Use Case Diagram of AmazonForest.

## 2. Querying AmazonForest

The web interface allows querying variant pathogenicity based on chromosome, position, or variant identifier from dbSNP. For example, the user can set the following data: chromosome 1, position 1020355, in reference allele textbox the user can set 'G', and in alternative allele text box set 'C' to perform prediction (see figure 3). Also, AmazonForest allows prediction by inputting rsID (see Figure 4) or a combination of in silico predictors (see Figure 5).



The screenshot displays the AmazonForest web application interface. On the left is a navigation sidebar with links: Home, AmazonForest, By Genomics, By Predictors, Explore, About, and Logout. The main content area features a form titled "Filter by genomic information or dbSNP rsids." with an example: "Ex: chr1, pos: 1020355, ref: G, alt: C". The form includes input fields for Chromosome (set to 1), Position (1020355), Ref (G), and Alt (C). Below these is a section for "Ex: 951245406" with an "RS" input field containing "000000". At the bottom of the form are "Predict" and "Clear" buttons. A "RESULT" box at the bottom shows the prediction "Benign" with a note: "With probaly 0.99 of model have 0.861 Accuracy.".

**Figure 3.** Example of query view on AmazonForest using genomic information.

© 2020 by the authors. Submitted to *Journal Not Specified* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Filter by genomic information or dbSNP rsids.

Ex: chr1, pos: 1020355, ref: G, alt: C

Chromosome	Position	Ref	Alt
Seleccione	POS	REF	ALT

Ex: 951245406

RS

951245406

Predict Clear

RESULT  
**Benign**

With probaly 0.99 of model have 0.861 Accuracy.

Figure 4. Example of query view on AmazonForest using rsID.

Fill with results of preditors by snpSift

FATHMM	LRT	MutaAss	MutaTaster	PROVEAN	Pph2 HDIV	Pph2 HVAR	SIFT
TOLERATED	Unknown	neutral	polymorphism	Neutral	Benign	Benign	TOLERATED

Predict Clear

RESULT  
**Benign**

With probaly 0.99 of model have 0.861 Accuracy.

Figure 5. Example of query view on AmazonForest using in silico categorical predictors.