

# 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

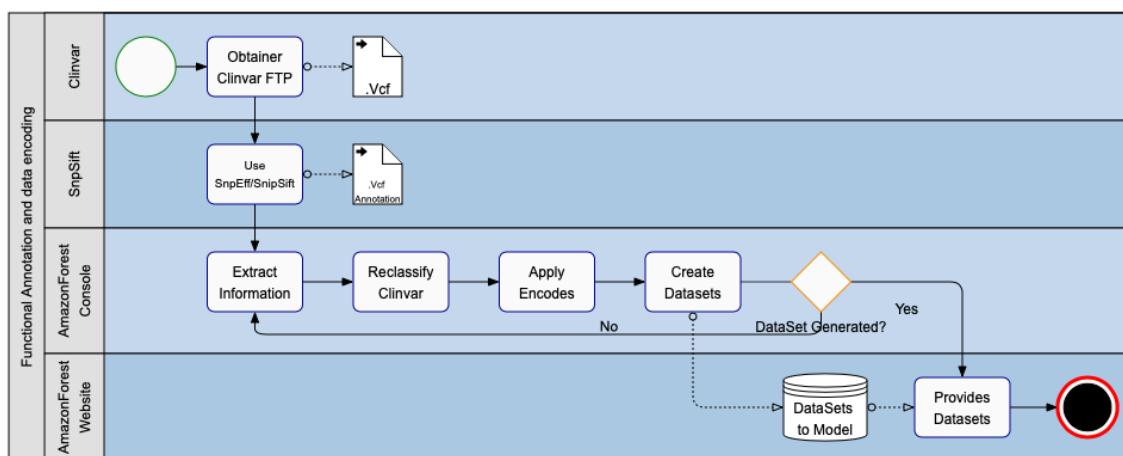
<sup>2</sup> Research Center on Oncology, Graduate Program of Oncology and Medical Science, Federal University of Pará, Belém 66073-000, Brazil; akelyufpa@gmail.com

\* Correspondence: gilderlanio@gmail.com

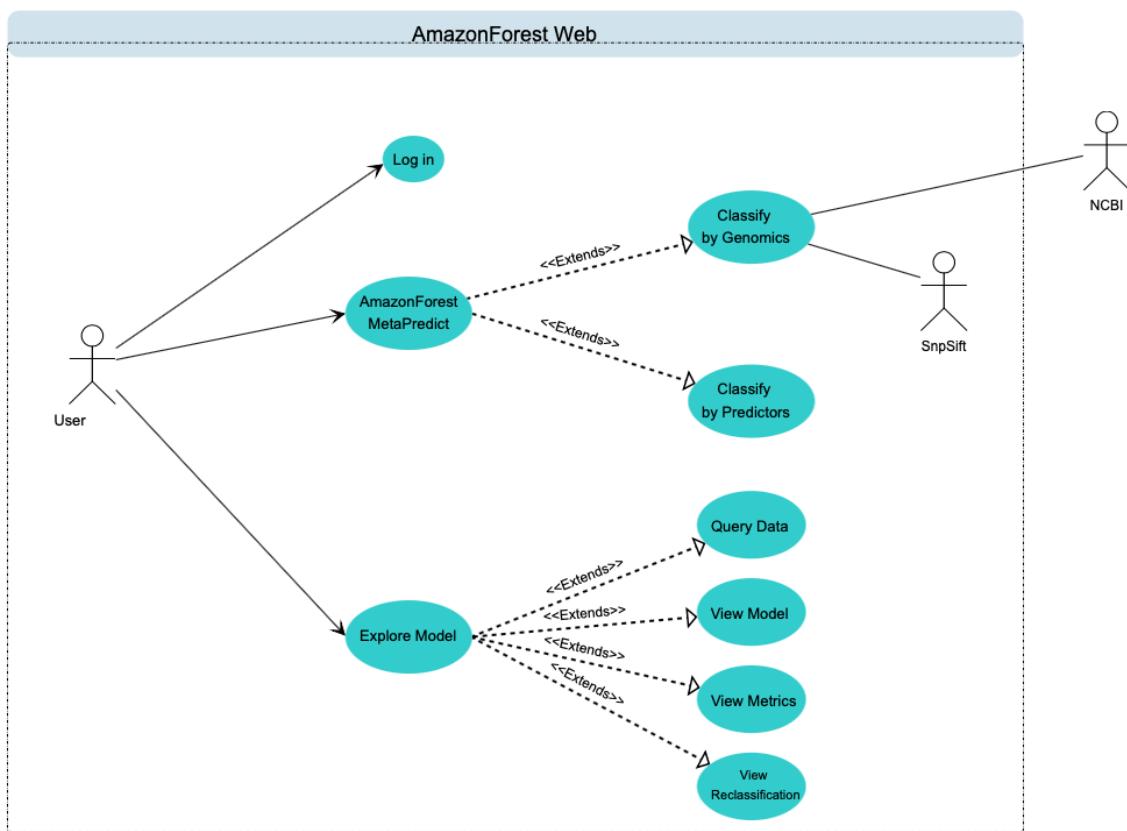
Version November 18, 2020 submitted to Journal Not Specified

<sup>1</sup> **Abstract:** Here we detailed the business process model and notation, use cases and usage examples of AmazonForest.

<sup>3</sup> **1. AmazonForest: use cases, web architecture and business process model and notation**



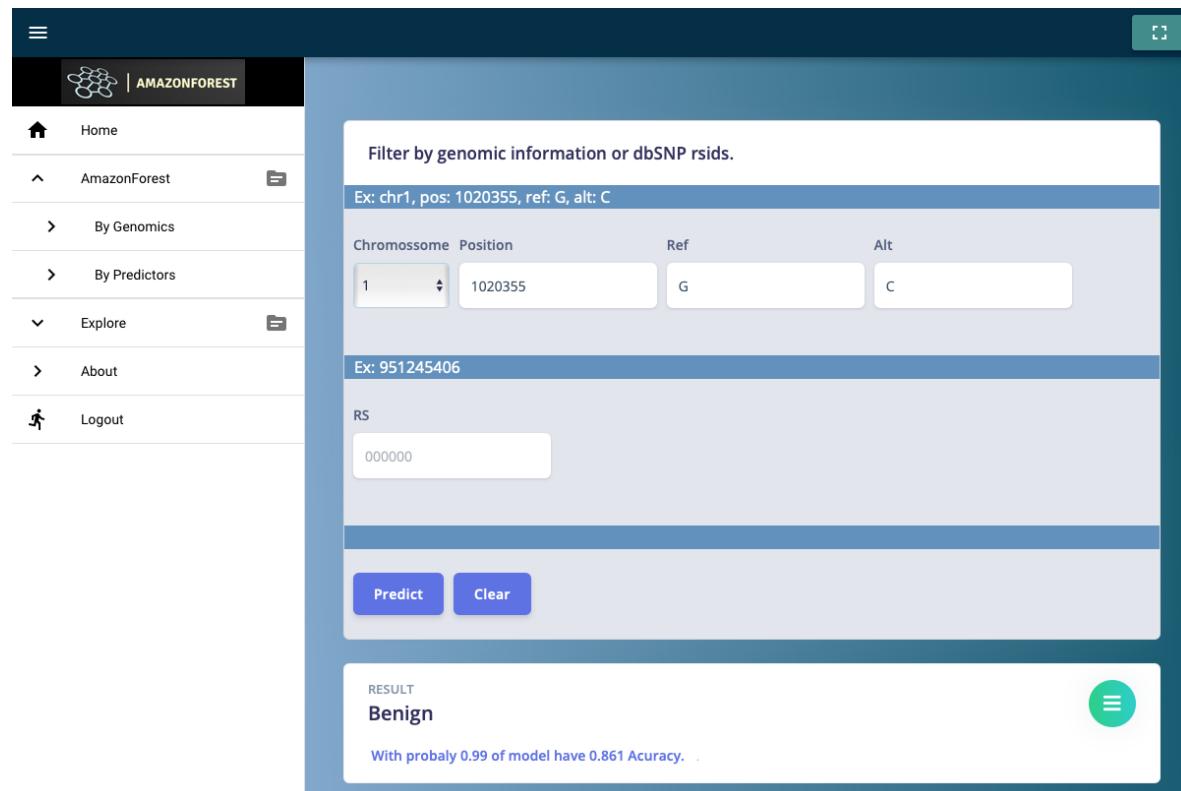
**Figure 1.** AmazonForest data flow starts by extracting ClinVar *.vcf* file available via File Transfer Protocol (FTP). SnpSift/SnpEff is used for functional variant annotation of ClinVar *.vcf* file for eight in silico predictors. AmazonForest extracts information from *.vcf* annotated file and performs 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.

#### 4 2. Querying AmazonForest

5 The web interface allows querying variant pathogenicity based on chromosome, position, or  
6 variant identifier from dbSNP. For example, the user can set the following data: chromosome 1, position  
7 1020355, in reference allele textbox the user can set 'G', and in alternative allele text box set 'C' to  
8 perform prediction (see figure 3). Also, AmazonForest allows prediction by inputting rsID (see Figure  
9 4) or a combination of in silico predictors (see Figure 5).



Filter by genomic information or dbSNP rsids.  
Ex: chr1, pos: 1020355, ref: G, alt: C

Chromosome	Position	Ref	Alt
1	1020355	G	C

Ex: 951245406

RS  
000000

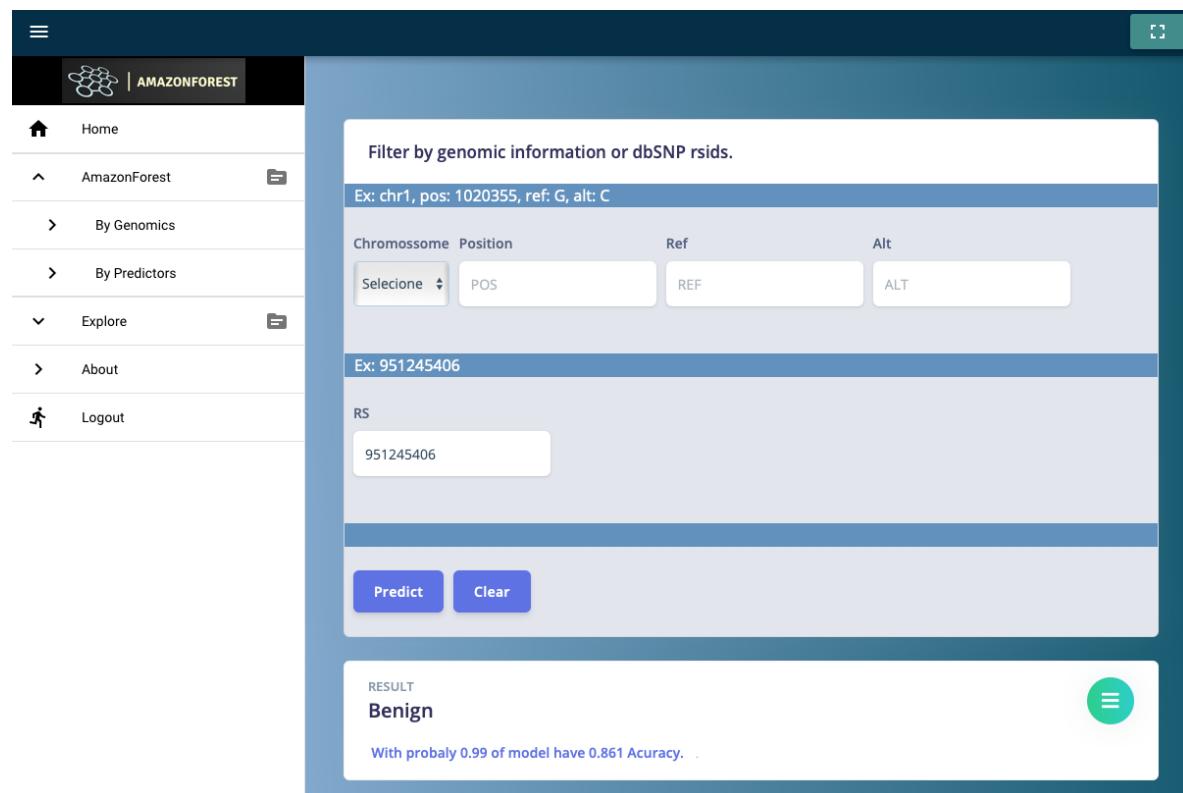
Predict Clear

RESULT  
Benign

With probaly 0.99 of model have 0.861 Acuracy.

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

10 © 2020 by the authors. Submitted to *Journal Not Specified* for possible open access publication  
11 under the terms and conditions of the Creative Commons Attribution (CC BY) license  
12 (<http://creativecommons.org/licenses/by/4.0/>).



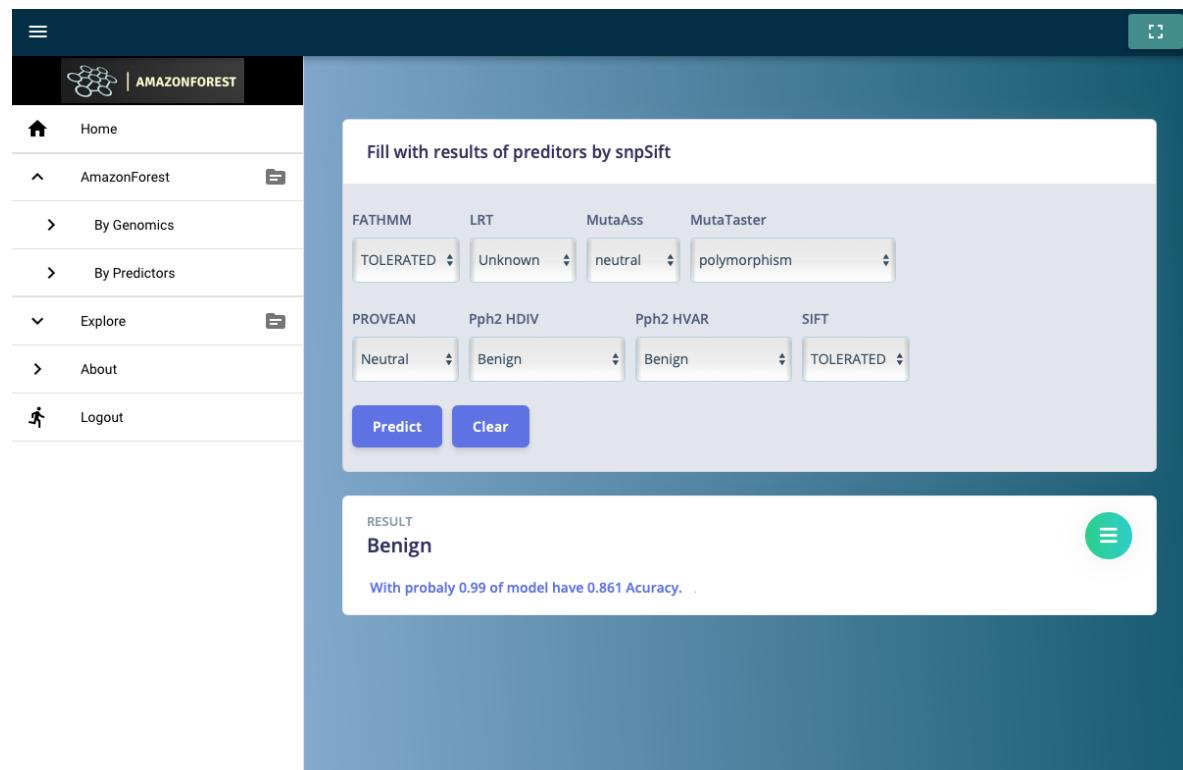
The screenshot shows the AmazonForest web application interface. On the left is a sidebar with a navigation menu:

- Home
- AmazonForest
  - By Genomics
  - By Predictors
- Explore
- About
- Logout

The main content area is titled "Filter by genomic information or dbSNP rsids." It contains a form with fields for "Chromosome" (dropdown: "Seleccione"), "Position" (input: "POS"), "Ref" (input: "REF"), and "Alt" (input: "ALT"). Below this is an example input "Ex: 951245406" and a "RS" input field containing "951245406". At the bottom are "Predict" and "Clear" buttons.

Below the form is a "RESULT" section with the text "Benign" and a note "With probaly 0.99 of model have 0.861 Accuracy. .".

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



The screenshot shows the AmazonForest web application interface, similar to Figure 4 but with different predictor selection. The sidebar and main layout are identical.

The main content area is titled "Fill with results of predictors by snpSift". It contains a form with dropdowns for various predictors:

FATHMM	LRT	MutaAss	MutaTaster
TOLERATED	Unknown	neutral	polymorphism

PROVEAN	Pph2 HDIV	Pph2 HVAR	SIFT
Neutral	Benign	Benign	TOLERATED

Below these are "Predict" and "Clear" buttons.

Below the form is a "RESULT" section with the text "Benign" and a note "With probaly 0.99 of model have 0.861 Accuracy. .".

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