

1 **Artificial neural networks to predict virological and immunological success in HIV**
2 **patients under antiretroviral therapy from a nationwide cohort in Colombia, using the**
3 **SISCAC database.**

4 **Predicting HIV Success with Neural Networks in Colombia**

5 Alberto Buitrago-Gutierrez¹ , Alexandra Porras-Ramirez²

6

7 1. Physician, infectious disease specialist at Hospital San José and Clinica Sanitas,
8 Bogota, Colombia. <https://orcid.org/0000-0003-2300-0540>

9 2. Group of Community Medicine and Public Health, Universidad El Bosque, Bogota,
10 Colombia. <https://orcid.org/0000-0002-0800-1388>

11

12 * AUTOR DE CORRESPONDENCIA:

13 Alexandra Porras-Ramirez

14 ORCID: 0000-0002-0800-1388

15 Address: Ak. 9 #131a-40, Usaquén, Bogotá, Cundinamarca. Postal Code: 110121

16 Mail: rporrasalexandra@unbosque.edu.co

17

18 Abstract

19 **Objective:** This study aimed to develop predictive models both for viral suppression and
20 immunological reconstitution using a standard set of reported variables in a nationwide
21 database system (SISCAC) from a cohort of patients living with HIV in Colombia.

22 **Materials and Methods:** We included 2.182 patients with no missing data related to the
23 outcomes of interest, during a 12 month follow up period. We randomly assigned a 0,7
24 proportion of this cohort to de training dataset for 2 different predictive models (logistic
25 regression, artificial neural networks). The AUC/ROC results were compared with those
26 obtained through the construction of artificial neural networks with the specified parameters.

27 **Results:** From a cohort of 2182 patients, 85,79% were male and at HIV diagnosis, the mean
28 value of the CD4 count was 342 x mm³. The logistic regression models obtained AUC/ROC
29 accuracy for the outcomes “suppressed viral load” 0,7, “undetectable viral load” of 0,66 and
30 “immunological reconstitution” 0,83; whereas the artificial neural network perceptron multilayer
31 obtained AUC/ROC of 0,77, 0.69 and 0,87 for the same outcomes.

32 **Conclusions:** The selection of specific variables from a nationwide database in Colombia with
33 quality control purposes allowed us to generate predictive models with an initial evaluation of
34 performance regarding three predefined outcomes for virological and immunological success.

35 **Keywords:** HIV/AIDS, machine learning, logistic regression, artificial neural networks,
36 dataset, prediction.

37 Introduction

38 Human immunodeficiency virus (HIV) infection and AIDS continue to generate a significant
39 burden of morbidity and mortality worldwide, with developing countries contributing the most
40 cases.

41 At a global level, the strategy of the Joint United Nations Program on HIV and AIDS (UNAIDS)
42 was generated to direct and prioritize efforts to control the disease. Thus, for 2020, the goal
43 of 90-90-90 was proposed, which refers to achieving a diagnosis of 90% of people living with

44 HIV, ensuring access to antiretroviral therapy for 90% of those diagnosed and achieving at
45 least 90% of virologic suppression in those on treatment. Additionally, and more recently, that
46 goal was revised to bring it to a goal of 95% in a horizon set until the year 2030 (1).

47 In Colombia exists a database in the high cost account (“Cuenta de alto costo”, in spanish),
48 this institution has the task of monitoring high cost diseases under the regulations defined by
49 the Ministry of Health and Social Protection of Colombia, it thereby reports the data generated
50 by the different health insurers and providers of the country's social security system, including
51 those referring to the care of people living with HIV (PLWH) in the country; latest results
52 showing an increase in the number of annual cases from 35,000 in 2012 to 123,490 cases by
53 2020 (1).

54 It is essential to introduce technological and computational tools in the clinical environment
55 and with data obtained from our own population of patients evaluate classification and
56 prediction models that can support the clinical approach; in our country the national SISCAC
57 system can provide a robust database to build classification and prediction models that
58 facilitate these clinical decisions and the creation of data driven algorithms and protocols.
59 Proposals for the use of machine learning in health care are supported with several goals,
60 such as enhancing the patient experience, optimizing population health, reducing costs, and
61 improving the working conditions of health workers. (2–15)

62 Studies such as that of Robbins GK et al, through the collection of data from electronic medical
63 record records in two tertiary care centers in Boston (United States); the Massachusetts
64 General Hospital and the Brigham and Women's Hospital, evaluated through a multivariable
65 logistic model which variables were associated with a greater risk of virological failure in
66 patients who entered care of the HIV clinic within the first year from the start of care and
67 treatment (10).

68 Wang D. et al. developed artificial neural network models to predict the response to therapy
69 using data from resistance genotypes and clinical information, they compared the results
70 obtained through the artificial neural network model with other methodologies such as random

71 forests and support vector machine algorithms. Using data from 1,204 episodes of treatment
72 changes for antiretroviral therapy, obtained from the RDI (HIV Resistance Response Database
73 Initiative) database, they sought to predict the behavior of viral load in the different models,
74 observing a similar correlation between the RF and SVM models with the artificial neural
75 network (17).

76 The results showed r^2 correlations in the predicted viral load with the observed value that
77 varied between 0.318-0.546 for the neural network, 0.590-0.751 for the random forests model
78 and 0.300 to 0.720 for the support vector machine model. Subsequently, by combining the
79 outputs of the 3 models, improvements were achieved in the correlations of the virological
80 predictions obtained and those observed. The authors concluded that the introduced models
81 (RF's and SVM) demonstrated similar performance to the neural network but that the
82 correlation results could improve with their combined application. (17).

83 Neural networks use a structure of neurons or nodes to analyze the interactions between a
84 group of covariates to predict an outcome; these neurons are organized by layers that are
85 associated in an input layer, intermediate hidden layers, and an output layer. The weighted
86 sum of neural inputs is somewhat analogous to the coefficients in linear and logistic
87 regressions, and the connection weights are iteratively adjusted by the learning algorithm to
88 minimize error and improve predictions (18). It is considered a strength for neural networks
89 its ability to fit interactions between variables in non-linear associations without any prior user
90 specification (11,19).

91 Our objective was to evaluate models for the prediction of virological and immunological
92 success with three predefined outcomes; viral load category "suppressed" or viral load <200
93 cop/mL, viral load category "undetectable" or viral load <20 cop/mL; and the immunological
94 outcome considering a CD4 lymphocyte count at the end of the 12-month follow-up period
95 above 350 cells x mm³ as indicative of "immunological reconstitution". Using the clinical and
96 laboratory variables obtained from the SISCAC registries for HIV/AIDS and with particular
97 interest in evaluating the inclusion of an integrase inhibitor in the antiretroviral therapy. We

98 sought to determine which model offers better performance through a comparison of a
99 multivariable logistic regression model and artificial neural networks.

100 **Materials and Methods**

101 This study pursued the training and validation of models using anonymous data from a cohort
102 of patients from a specialized HIV care provider in Bogotá (Col), using the SISCAC database
103 of the high-cost account. (<https://cuentadealtocosto.org/site/general/comunicado-siscac/>)
104 (21–23).

105 The analysis included only data obtained from adult patients (18 years +) reported within the
106 period of 12 months of follow-up between January 1st 2022 and January 1st 2023, the data
107 was previously uploaded to the high-cost account following a structured set of 99 variables, of
108 which the final analysis took 15, either in their primary registry or recoded to be included in the
109 modelling.

110 This study was approved by the Ethics and Research committee of Vidamedical IPS on March
111 17th, 2023. All methods followed national (Resolution 8430 of 1993, stated by the Colombian
112 Health Ministry) and international (The Declaration of Helsinki) standards. It was a secondary
113 data use, observational and retrospective study; therefore it did not require informed consent
114 from participants.

115 **Variables and definitions**

116 The demographic variables from the SISCAC notification record included age, sex, weight and
117 height, the diagnosis of pregnancy at the time of notification; the presence of comorbidities
118 such as active tuberculosis, hepatitis B and C virus coinfection, chronic kidney disease,
119 coronary disease and neoplasms not associated with AIDS.

120 Additionally, variables related to the characterization of HIV infection were taken, such as the
121 years of diagnosis at the time of data notification, the mechanism of transmission of HIV
122 infection, the stage at the time of diagnosis, the CD4 lymphocyte count and viral load at
123 diagnosis and also at the time of initiation of antiretroviral therapy; also at the end of the 12-
124 month follow-up period (January 2022 to January 2023). Changes in antiretroviral therapy, the

125 presence of treatment failures, the number of infectious disease specialty consultations within
126 the 12-month follow-up (>3 or <3 visits in the 12-month period) were recorded; the months of
127 effective dispensing of antiretroviral treatment, the presence of an integrase inhibitor
128 (raltegravir, elvitegravir, dolutegravir) was recorded in the initial scheme and the final follow-
129 up scheme for the 12-month period.

130 The variables were included in the models in a dichotomous way to facilitate their analysis.
131 For example, the measurement of the months of dispensing therapy in the last year was
132 categorized into patients with compliance greater than 95% and those below this goal,
133 considering 100% dispensing in the total 12 months of the year, The variable on frequency of
134 consultation by the infectious disease specialty was dichotomized according to the distribution
135 found in the observations in >3 or <3 visits (21).

136 **Data source**

137 The source of data were the records of the variables from the notification to the SISCAC
138 database, and we included those considered relevant to the purpose of this investigation. The
139 information is previously verified in accordance with the instructions for reporting information
140 to the high-cost account for HIV/AIDS, regulated under Resolution 273 of 2019 of the Ministry
141 of Health and Social Protection of Colombia (23) (Chrome-
142 extension://efaidnbmnnibpcajpcgclefindmkaj/https://www.minsalud.gov.co/sites/rid/Lists/Bibli
143 otecaDigital/RIDE/DE/DIJ/resolucion-273-de-2019.pdf) (23)

144 The data from ninety-nine variables corresponding to the administrative follow-up of the
145 SISCAC was initially extracted (except cut-off date/Var.98 and serial code/Var.99) for a cohort
146 of 6526 patients, belonging to an HIV service care provider in Colombia (Vidamedical IPS,
147 <https://www.vidamedicalips.com/>). Patients included came from three cities (Bucaramanga,
148 Cúcuta and Bogotá).

149 We discarded the records corresponding to patients under 18 years of age and the
150 observations or patients that had missing data regarding the outcomes of interest at diagnosis,
151 initiation of antiretroviral treatment, and at the end of follow-up (last 12 months between

152 January 2022 and January 2023). By filtering for lost data and recoding them, a total of fifteen
153 variables were included in the definitive analysis (See figure No. 1 Patient selection flowchart).

154 **Outcomes evaluated.**

155 We established three dependent variables or outcomes: suppressed viral load with a value
156 <200 cop/ml, undetectable viral load with a value of <20 cop/ml and CD4 lymphocyte count
157 >350 x mm³, all considered at the end of the twelve-month follow-up period. The CD4 count
158 value was set arbitrarily by the investigators for the immunological reconstitution outcome,
159 considering the mean value for the entire cohort in this parameter at the time of diagnosis.

160 **Ethical aspects**

161 The study was performed through the secondary use of data previously collected in a
162 periodically and regulated manner under resolution 273/2019 from the Ministry of Health and
163 Social Protection of Colombia. (Chrome-
164 extension://efaidnbnmnibpcajpcglclefindmkaj/https://www.minsalud.gov.co/sites/rid/Lists/Bibli
165 otecaDigital/RIDE/DE/DIJ/resolucion-273-de-2019.pdf).

166 The study followed the recommendations for the secondary use of clinical data (24–26). The
167 protocol was approved by the Vidamedical IPS research and ethics committee
168 (<https://www.vidamedicalips.com/>). Data extraction was performed by computing personnel of
169 the provider's data analysis area and delivered to the researchers in the excel program format
170 (Microsoft® Excel®).

171 To preserve the anonymity of the patients and since it was a secondary analysis of data
172 already collected for the purposes of notification to the high-cost account (CAC), each patient
173 had a natural consecutive number assigned, discarding the identification numbers and names.

174 **Statistical analysis**

175 We performed a bivariate analysis to obtain crude association measures in the training or
176 retrospective cohort, obtaining the respective OR's (See tables 1, 2 and 3). The analysis then
177 included the multivariate logistic regression model to adjust the variables for potential
178 confusion and interaction, using the statistical program Stata Version 18.0 B. E, with the

179 construction of the respective AUC//ROC curves for each of the outcomes; calculating in the
180 multivariate model the O.R's with the corresponding confidence intervals.

181 For the creation of the neural network, we used IBM SPSS Statistics Version 28.0.0.0
182 program, which implements the "multilayer perceptron" network with a partition of 70% of the
183 data in the training model and 30% in the testing; with a hyperbolic tangent hidden layer
184 activation function and softmax output layer activation function (18)

185 In the construction of the artificial neural network we included ten factors given by the
186 dichotomized predictive variables: age, sex, stage at diagnosis, start of antiretroviral therapy
187 with any integrase inhibitor, coinfection with hepatitis B virus, coinfection with hepatitis C virus,
188 active tuberculosis, changes in antiretroviral therapy, antiretroviral therapy failures,
189 consultations by the I.D specialist >3 in the period of follow-up, chronic kidney disease,
190 coronary disease, non-AIDS associated neoplasm and adherence to the antiretroviral therapy
191 dispensing above 95%.

192 Additionally, as covariates we included viral load at the beginning of treatment and CD4
193 lymphocyte count at the beginning of treatment plus the value of the body mass index (BMI)
194 (18).

195 The distribution of the factors and the quantitative covariates was done to maintain the
196 assumption of independence, using the multilayer perceptron network, as a supervised
197 learning technique, with feedback architecture, hyperbolic tangent activation function in the
198 hidden layers and softmax in the hidden layer output, plus cross entropy error function (18).

199 **Results**

200 From the 2182 patients included we identified a proportion of 85.79% for the male sex and
201 14.21% for the female sex. Regarding age, we observed for the entire cohort a mean of 33
202 years old (SD +/- 11.6 years), with an equal median, considering a normal distribution of this
203 variable. For this cohort, the mean HIV diagnosis time was 3.3 years, with a minimum follow-
204 up time from diagnosis of 3 months and a maximum of 18 years.

205 The reported mechanism of HIV transmission corresponded in 98.63% to sexual, with 1.37%
206 grouping other different mechanisms and those not reported. At the time of diagnosis of HIV
207 infection, the mean value of the CD4 lymphocyte count was 342 x mm³. The proportion of
208 those coinfecting with hepatitis B at the start of treatment was only 1.51%, while hepatitis C
209 coinfection was lower with 0.27%. The proportion of patients coinfecting with active
210 tuberculosis was 1.15%.

211 During the initiation of antiretroviral therapy, the mean dispensing time of the antiretroviral
212 therapy was 6.5 months (SD+/-4.45). The mean of therapy changes was 1.7 during de 12-
213 month follow-up period, and the diagnosis of antiretroviral therapy failures was done on
214 average 1.95 times with a minimum of one change and a maximum of three modifications in
215 the antiretroviral therapy.

216 Regarding the outcome of “suppressed viral load” (<200 cop/ml), the following variables were
217 observed as statistically significant: HIV stage at diagnosis (X² 4.507; p=0.105); hepatitis C
218 virus infection at diagnosis (OR 0.188 95% CI 0.038-0.934; p=0.022); changes in antiretroviral
219 therapy (OR 1.906 95% CI 1.409-2.577; p<0.001); adherence to antiretroviral therapy
220 dispensing above 95% (OR 8.152 95%CI 4.63-14.33; p<0.001) and the record of >3 attentions
221 for infectious diseases in the last year (OR 2.246 95%CI 1.68-2, 99; p<0.001) (See table 1).

222 We used the bivariate and multivariate logistic regression analysis to generate the
223 correspondent AUC/ROC curve for the outcome “suppressed viral load”, obtaining an area
224 under de ROC curve of 0.7017.

225 When estimating the associations for the outcome variable “undetectable viral load” (<20
226 cop/ml) it was found that the variables: HIV stage at diagnosis (X² 28.926; p<0.001); viral load
227 at the start of ART >100,000 cop/ml (OR 0.637 95% CI 0.527-0.770; p<0.001); hepatitis C
228 virus infection at diagnosis (OR 0.08 95% CI 0.009-0.687; p 0.003); changes in antiretroviral
229 therapy during the follow-up period (O,R 1.586 95% CI 1.265-1.989; p<0.001); adherence to
230 the dispensing of antiretroviral therapy above 95% (OR 2.618 IC95% 1.991-3.443; p<0.001)

231 and likewise the record of >3 attentions for the I.D specialist in the follow-up period (OR 1.648
232 IC95% 1.277-2.128; p<0.001) were statistically significant (See table 2).

233 We evaluated the accuracy for the outcome of “undetectable viral load” for the multivariable
234 logistic regression model, with the variables previously listed, obtaining an AUC/ROC of
235 0,6605. (See figure 2).

236 Finally we obtained crude and adjusted O.R's for the outcome immunological reconstitution at
237 the end of the follow-up period (CD4 count>350 x mm³).The variables HIV stage at diagnosis
238 (X² 556.626; p<0.001); viral load at the start of antiretroviral therapy >100,000 cop/ml (OR
239 0.582 95% CI 0.483-0.702; p<0.001); active tuberculosis at diagnosis (OR 0.288 95% CI
240 0.129-0.645; p=0.001); the inclusion of an integrase inhibitor in the initial antiretroviral therapy
241 (OR 0.601 95% CI 0.481-0.751; p<0.001); changes in antiretroviral therapy (OR 1.794 95%CI
242 1.432-2.249; p<0.001); adherence to antiretroviral therapy dispensing above 95% (OR 1.534
243 95% CI 1.208-1.946; p<0.001); the record of >3 consultations by the I.D specialist in the period
244 of follow-up (OR 1.539 IC95% 1.194-1.984; p<0.001) met the criterion of significance. (See
245 table No. 3).

246 The performance for immunological reconstitution of the multivariable logistic regression
247 model obtained an AUC/ROC of 0,8364, showing better predictive accuracy than for the other
248 previous outcomes (See figure 3).

249 For the artificial neural network, the model for the “suppressed viral load” outcome achieved
250 an overall correct percentage in the training phase of 84.5% and in the tests of 84% (See
251 figure 4). An AUC/ROC of 0.777 for this outcome. In this same model it showed 15.5% of
252 incorrect forecasts in the training phase and 16% in the tests. The relationships of the neural
253 network with its synaptic weighting can be observed in figure 4. Evaluating against its
254 predictive capacity, this neural network showed an accuracy of 80.9% and a Kappa value of
255 67.3%.

256 The accuracy of the neural network obtained through the AUC/ROC curve was 0.777 for this
257 outcome.

258 In predicting the outcome of undetectable viral load, the neural network obtained an accuracy
259 through the evaluation of AUC/ROC of 0.69, reflecting a similar performance when we
260 compare it with the logistic regression model, and being consistently inferior for both models
261 when compared with de suppressed viral load outcome. (See figure 5).

262 The best performance observed in the neural network models was observed for the outcome
263 of immunological reconstitution, obtaining an AUC/ROC of 0.878, consistent with the same
264 finding for the performance of the multivariable logistic regression models. (See figure 6).

265 When we compared the results of both modelling strategies for the three mentioned outcomes
266 we see similar performances, favoring by a slight difference the neural network models and
267 being consistent in better results for the prediction of immunological reconstitution in both
268 cases. (See table 4).

269 **Discussion**

270 In this study we applied predictive models using the multivariable logistic regression and
271 artificial neural networks for three predefined outcomes related to virological and
272 immunological success in the treatment of people living with HIV (PLWH), the variables
273 included those routinely collected through the SISCAC database in Colombia, including
274 demographic information, clinical characteristics, and laboratory parameters. The SISCAC
275 database purpose is to centralize the registration and validation in real time of quality and
276 assistance indicators from patients of the high-cost account of the Colombian health system
277 (<https://cuentadealcosto.org/general/comunicado-siscac/>).

278 We conducted a statistical analysis with a predictive intention, mainly referencing in the study
279 by Wang et al, who developed machine learning algorithms through artificial neural networks,
280 random forests, and support vector machines, to predict viral load results from 1204 episodes
281 of change in antiretroviral therapy; incorporating genotype mutations input and treatment
282 history variables (17).

283 Our population consisted of mainly men (86%), observing 75% of all below de 42 years of age,
284 and with a fairly good level for CD4 lymphocyte count both at diagnosis and the initiation of
285 antiretroviral therapy; suggesting earlier introductions of antiretroviral therapy for this cohort.
286 We found previously known variables predictive of success such as the stage for HIV/AIDS at
287 diagnosis, the presence of therapy failures during follow-up and the proportion of adherence
288 to overall antiretroviral dispensation by the provider of care. No association was found both
289 for the suppressed viral load and undetectable viral load outcomes regarding the initial therapy
290 with the inclusion of an INSTI. An interesting finding was the association of more than three
291 consultations with the I.D specialist during the follow-up period of twelve months with better
292 results in all three outcomes.

293 A limitation already established neural networks is the necessity of a great amount of data for
294 the training, as much as flexibility is required from the neural network bigger sets of data are
295 needed.

296 The advantages of neural networks compared to other deep learning or machine learning
297 techniques are:

- 298 • Non-linear system: By being able to process information in a non-linear way, it allows
299 the processing of more “chaotic” data.
- 300 • Fault tolerance: Since it processes the information in parallel between the different
301 nodes, neural networks are quite tolerant to single or multiple node failures.
- 302 • Adaptability: A neural network could adjust its parameters in real time depending on
303 the inputs it receives and its characteristics, while other models do not have this
304 adaptability.

305

306 Another weakness of neuronal networks is that once the layers increase, the interpretation of
307 the model becomes more difficult, until it becomes a true black box. In addition, gradients that
308 explode (that is, they grow without limit) or that vanish (that is, they progressively approach

309 zero), constitute aspects that make it difficult to train this architecture and reduce its ability to
310 process long-duration sequences.

311 Although neural networks significantly reduce this inconvenience, they do not eliminate it
312 and this problem becomes more evident as the sequence to process becomes increasingly
313 extensive. The presence of exploding and fading gradients limits the long-term memory of
314 these architectures.

315 Finally, regarding the results obtained with neural networks, the three models present
316 acceptable results, with high precision and moderate Kappa values. The positive points of the
317 networks neural networks is that they are very flexible and can work with very complex data
318 to find patterns that other techniques cannot; and while in the other techniques you have more
319 or less information about the patterns found in the data and how the variables affect the result,
320 with neural networks this information is lost and you have only one input and one output.

321 Regarding the comparison of the models, we found similar performance both for the
322 multivariable logistic regression and the neural networks, in both cases the accuracy was
323 superior in predicting the immunological reconstitution outcome, and in this matter the
324 inclusion of an INSTI for the initial antiretroviral therapy was statistically significant.

325 Also, we must consider the risk for selection bias, given that we obtained data from only one
326 HIV care provider in the country, which can affect external validity, but as this institution
327 allowed access to data from patients from three different cities in different circumstances of
328 health insurance, access to care and socioeconomical conditions, the bias can be minimized.

329 **Conclusions**

330 This is an important exercise from already verified data uploaded to a nationwide database,
331 allowing access to several variables from people living with HIV (PLWH) in Colombia; it allows
332 an evolving process of integrating artificial intelligence algorithms to provide a growing
333 framework directed to support health care providers with new tools towards their patients.

334 **Competing interests**

335 The authors of this manuscript declare they have no potential conflicts of interest with respect
336 to the research, authorship, and/or publication of this article.

337 **Ethical responsibilities**

338 The authors manifest that this study followed the recommendations for the secondary use of
339 clinical data (24–26). The protocol was reviewed and approved by the Vidamedical IPS
340 research and ethics committee in the corresponding session of march 17th, 2023.

341 **Acknowledgements**

342 We thank Vidamedical IPS (<https://www.vidamedicalips.com/>) for supporting this study at its
343 HIV Clinics in Colombia

344 **Funding**

345 This work was executed with funding from the investigators, personal computing equipment
346 and statistical programs with individual licensing from them. No external nor sponsor funding
347 was used.

348

349 **Data Availability Statement**

350 The data that support the findings of this study are available on request from the corresponding
351 author. The data are not publicly available due to privacy or ethical restrictions.

352 **List of abbreviations**

353 AIDS: Acquired immune deficiency syndrome

354 ANN: Artificial neural network

355 ART: antiretroviral therapy

356 I.D: Infectious Diseases

357 INSTI: Integrase strand transfer inhibitors

358 PLWH: people living with HIV

359 S.D: Standard deviation

360 **References**

- 361 1. De C, Costo A. Situación del VIH/SIDA en Colombia 2020 CUENTA DE ALTO COSTO
362 Fondo Colombiano de Enfermedades de Alto Costo.
363 https://cuentadealtocosto.org/vih/situacion-del-vih-2020-v_0-1/
- 364 2. Bodenheimer T, Sinsky C. From triple to Quadruple Aim: Care of the patient requires care
365 of the provider. *Ann Fam Med*. 2014 Nov 1;12(6):573–6. doi: 10.1370/afm.1713.
- 366 3. Lin J, Mauntel-Medici C, Heinert S, Baghikar S. Harnessing the power of the electronic
367 medical record to facilitate an opt-out HIV screening program in an urban academic
368 Emergency Department. *Journal of Public Health Management and Practice*. 2017;23(3):264–
369 8. doi: 10.1097/PHH.0000000000000448.
- 370 4. Roth JA, Radevski G, Marzolini C, Rauch A, Günthard HF, Kouyos RD, et al. Cohort-Derived
371 Machine Learning Models for Individual Prediction of Chronic Kidney Disease in People Living
372 with Human Immunodeficiency Virus: A Prospective Multicenter Cohort Study. *Journal of*
373 *Infectious Diseases*. 2021 Oct 1;224(7):1198–208. doi: 10.1093/infdis/jiaa236.
- 374 5. Oliveira A, Faria BM, Gaio AR, Reis LP. Data Mining in HIV-AIDS Surveillance System:
375 Application to Portuguese Data. *J Med Syst*. 2017 Apr 1;41(4). doi: 10.1007/s10916-017-
376 0697-4.
- 377 6. Kamal S, Urata J, Cavassini M, Liu H, Kouyos R, Bugnon O, et al. Random Forest machine
378 learning algorithm predicts virologic outcomes among HIV infected adults in Lausanne,
379 Switzerland using electronically monitored combined antiretroviral treatment adherence. *AIDS*
380 *Care - Psychological and Socio-Medical Aspects of AIDS/HIV*. 2021;33(4):530–6. doi:
381 10.1080/09540121.2020.1751045.
- 382 7. Xiang Y, Du J, Fujimoto K, Li F, Schneider J, Tao C. Application of artificial intelligence and
383 machine learning for HIV prevention interventions. Vol. 9, *The Lancet HIV*. Elsevier Ltd; 2022.
384 p. e54–62. doi: 10.1016/S2352-3018(21)00247-2.
- 385 8. Alehegn M. Application of machine learning and deep learning for the prediction of
386 HIV/AIDS. *HIV and AIDS Review*. 2022;21(1):17–23. doi.org/10.3389/fpubh.2022.967681.

- 387 9. Benitez AE, Musinguzi N, Bangsberg DR, Bwana MB, Muzoora C, Hunt PW, et al. Super
388 learner analysis of real-time electronically monitored adherence to antiretroviral therapy under
389 constrained optimization and comparison to non-differentiated care approaches for persons
390 living with HIV in rural Uganda. 2020; Available from:
391 <http://onlinelibrary.wiley.com/doi/10.1002/jia2.25467/full>
- 392 10. Robbins GK, Johnson KL, Chang Y, Jackson KE, Sax PE, Meigs JB, et al. Predicting
393 virologic failure in an HIV clinic. *Clinical Infectious Diseases*. 2010 Mar 1;50(5):779–86. doi:
394 10.1086/650537.
- 395 11. Bi Q, Goodman KE, Kaminsky J, Lessler J. What is machine learning? A primer for the
396 epidemiologist. *Am J Epidemiol*. 2019 Dec 31;188(12):2222–39. doi: 10.1093/aje/kwz189.
- 397 12. Peiffer-Smadja N, Rawson TM, Ahmad R, Buchard A, Pantelis G, Lescure FX, et al.
398 Machine learning for clinical decision support in infectious diseases: a narrative review of
399 current applications. Vol. 26, *Clinical Microbiology and Infection*. Elsevier B.V.; 2020. p. 584–
400 95. doi: 10.1016/j.cmi.2019.09.009.
- 401 13. Hinton G. Deep learning-a technology with the potential to transform health care. Vol. 320,
402 *JAMA - Journal of the American Medical Association*. American Medical Association; 2018. p.
403 1101–2. doi: 10.1001/jama.2018.11100.
- 404 14. Ridgway JP, Lee A, Devlin S, Kerman J, Mayampurath A. Machine Learning and Clinical
405 Informatics for Improving HIV Care Continuum Outcomes. Vol. 18, *Current HIV/AIDS Reports*.
406 Springer; 2021. p. 229–36. doi: 10.1007/s11904-021-00552-3.
- 407 15. Yang X, Zhang J, Chen S, Weissman S, Olatosi B, Li X. Utilizing electronic health record
408 data to understand comorbidity burden among people living with HIV: a machine learning
409 approach. *AIDS*. 2021 May 1;35:S39–51. doi: 10.1097/QAD.0000000000002736.
- 410 16. Gebrezgi MT, Fennie KP, Sheehan DM, Ibrahimou B, Jones SG, Brock P, et al.
411 Development and Validation of a Risk Prediction Tool to Identify People with HIV Infection
412 Likely Not to Achieve Viral Suppression. *AIDS Patient Care STDS*. 2020 Apr 1;34(4):157–65.
413 doi: [10.1089/apc.2019.0224](https://doi.org/10.1089/apc.2019.0224).

- 414 17. Wang D, Larder B, Revell A, Montaner J, Harrigan R, De Wolf F, et al. A comparison of
415 three computational modelling methods for the prediction of virological response to
416 combination HIV therapy. *Artif Intell Med.* 2009 Sep;47(1):63–74. doi:
417 10.1016/j.artmed.2009.05.002.
- 418 18. IBM SPSS Neural Networks New tools for building predictive models [Internet]. Available
419 from: www.ibm.com/spss/devcentral.
- 420 19. Sardari S, Sardari D. Applications of Artificial Neural Network in AIDS Research and
421 Therapy. Vol. 8, *Current Pharmaceutical Design*. 2002. doi: 10.2174/1381612024607199.
- 422 20. Pedrero V, Reynaldos-Grandon K, Ureta-Achurra J, Cortez-Pinto E. Generalidades del
423 Machine Learning y su aplicación en la gestión sanitaria en servicios de urgencia. *Rev Med*
424 *Chile* 2021; 149:248-54. <http://dx.doi.org/10.4067/s0034-98872021000200248>.
- 425 21. Consenso basado en Evidencia. Indicadores Mínimos para Evaluar Resultados de Gestión
426 y Clínicos en Instituciones de Atención a Personas Viviendo con VIH en Colombia. Cuenta de
427 Alto Costo. Fondo Colombiano de Enfermedades de Alto Costo. Junio-diciembre de 2014.
428 [https://cuentadealtocosto.org/publicaciones/indicadores-minimos-para-evaluar-resultados-](https://cuentadealtocosto.org/publicaciones/indicadores-minimos-para-evaluar-resultados-de-gestion-y-clinicos-en-instituciones-de-atencion-a-personas-viviendo-con-vih-en-colombia/)
429 [de-gestion-y-clinicos-en-instituciones-de-atencion-a-personas-viviendo-con-vih-en-colombia/](https://cuentadealtocosto.org/publicaciones/indicadores-minimos-para-evaluar-resultados-de-gestion-y-clinicos-en-instituciones-de-atencion-a-personas-viviendo-con-vih-en-colombia/)
- 430 22. De C, Costo A. Situación del VIH/SIDA en Colombia 2021 CUENTA DE ALTO COSTO.
431 *Fondo Colombiano de Enfermedades de Alto Costo, Cuenta de Alto*
432 *Costo (CAC). Situación del VIH/SIDA en Colombia 2020*; Bogotá D.C. 2021.
- 433 23. Instructivo para el reporte de información según resolución 273 del 2019. Cuenta de Alto
434 Costo. Ministerio de Salud y Protección Social. República de Colombia. Version 02. 2021.
435 [chrome-extension://efaidnbmninnibpcajpcgclclefindmkaj/https://www.cajacopieps.com/wp-](chrome-extension://efaidnbmninnibpcajpcgclclefindmkaj/https://www.cajacopieps.com/wp-content/uploads/Instructivo_Reporte_VIH_2022.pdf)
436 [content/uploads/Instructivo_Reporte_VIH_2022.pdf](chrome-extension://efaidnbmninnibpcajpcgclclefindmkaj/https://www.cajacopieps.com/wp-content/uploads/Instructivo_Reporte_VIH_2022.pdf)
- 437 24. Jungkunz M, Köngeter A, Mehlis K, Winkler EC, Schickhardt C. Secondary use of clinical
438 data in data-gathering, non-interventional research or learning activities: Definition, types, and
439 a framework for risk assessment. *J Med Internet Res.* 2021 Jun 1;23(6). doi: [10.2196/26631](https://doi.org/10.2196/26631).

- 440 25. Wiesenauer M, Johner C, Röhrig R. Secondary use of clinical data in healthcare providers
441 - An overview on research, regulatory and ethical requirements. In: Studies in Health
442 Technology and Informatics. IOS Press; 2012. p. 614–8. PMID: 22874264
- 443 26. Cumyn A, Dault R, Barton A, Cloutier AM, Ethier JF. Citizens, Research Ethics Committee
444 Members and Researchers' Attitude Toward Information and Consent for the Secondary Use
445 of Health Data: Implications for Research Within Learning Health Systems. Journal of
446 Empirical Research on Human Research Ethics. 2021 Jul 1;16(3):165–78. doi:
447 10.1177/1556264621992214

Table 1. Bivariate and multivariate regression analysis for the outcome "suppressed viral load."

Variables	Crude OR† 95%CI‡	X ² p-value	Adjusted OR 95%CI	P value
Age	1,462 1,160-1,842	10,442 p=0,001	1,394 1,078-1,801	p=0,01
Sex	0,833 0,629-1,239	0,520 p=0,471	1,144 0,795-1,646	p=0,467
Female	0,898 0,669-1,205	-		
Male	0,883 0,629-1,239	-		
Viral load at the start of ART§ (>o<100.000 cop/ml)	1,007 0,794-1,278	0,004 p=0,952	1,077 0,829-1,400	p=0,574
HIV stage at diagnosis		4,507 p 0,105	Stage 2 0,805 0,582-1,111	p=0,188
			Stage 3 0,592 0,415-0,845	p=0,004
HBV infection	0,076 0,005-0,664	0,014 p=0,905	1,236 0,450-3,390	p=0,680
HCV infection	0,188 0,038-0,934	5,230 p 0,022	0,149 0,024-0,915	p=0,04
Active tuberculosis at diagnosis	0,754 0,281-2,022	0,317 p 0,573	- -	-
ART antiretroviral therapy with INSTI's¶ included	0,920 0,689-1,227	0,324 p 0,569	1,360 1,004-1,842	p=0,04
ART Changes	1,906 1,409-2,577	17,976 p <0,001	2,150 1,513-3,055	p<0,001
Adherence due to ART dispensing (>95%/11 months)	8,152 4,63-14,33	73,156 p<0,001	8,23 4,629-14,656	p<0,001
Failure with current ART	0,931 0,546-1,589	0,068 p 0,794	0,241 0,241-0,843	p=0,013
Chronic kidney disease	0,841 0,825-0,856	0,948 p 0,330	- -	-
Coronary heart disease	0,841 0,825-0,856	0,758 p 0,384	- -	-
Neoplasm not associated with AIDS	0,841 0,825-0,825	1,138 p 0,286	- -	-
Consultations with I.D specialist in the period (>3 vs <3)	2,246 1,68-2,99	31,687 p <0,001	2,098 1,554-2,830	p<0,001

medRxiv preprint doi: <https://doi.org/10.1101/2024.10.26.24316181>; this version posted October 28, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

†: Odds Ratio
 ‡ Confidence interval
 §: Antiretroviral therapy
 ¶: Integrase strand transfer inhibitor

Table 2. Bivariate and multivariate regression analysis for the outcome "undetectable viral load"

Variables	Crude OR† 95%CI	X ² p-value	Adjusted OR 95%CI	P value
Age	1,182 0,982-1,423	3,131 p=0,077	1,226 0,997-1,506	p=0,05
Sex	0,999 0,766-1,302	0,000 p=0,991	1,191 0,896-1,583	p=0,226
Female	0,999 0,795-1,254	-		
Male	1,000 0,963-1,039	-		
Viral load at the start of ART§ (>o<100.000 cop/ml)	0,637 0,527-0,770	21,948 p<0,001	0,708 0,575-0,871	p=0,001
HIV stage at diagnosis		28,926 p<0,001	Stage2 0,783 0,602-1,019	p=0,06
			Stage3 0,520 0,391-0,693	p<0,001
HBV infection	1,076 0,498-2,329	0,035 p=0,852	1,186 0,524-2,686	p=0,681
HCV infection	0,080 0,009-0,687	8,758 p=0,003	0,760 0,008-0,676	p=0,021
Active tuberculosis at diagnosis	0,601 0,269-1,345	1,567 p=0,211	- -	-
ART antiretroviral therapy with INSTI's¶ included	0,851 0,675-1,075	1,835 p=0,176	1,154 0,901-1,477	p=0,256
ART Changes	1,586 1,265-1,989	16,164 p<0,001	1,783 1,373-2,316	p<0,001

Adherence due to ART dispensing (>95%/11 months)	2,618	1,991-3,443	50,036 p<0,001	2,854	2,135-3,815	p<0,001
Failure with current ART	0,750	0,492-1,142	1,811 p=0,178	0,475	0,293-0,769	p=0,002
chronic kidney disease	1,614	0,180-14,473	0,187 p=0,666	-	-	-
Coronary heart disease	0,712	0,693-0,731	1,616 p=0,204	-	-	-
Neoplasm not associated with AIDS	0,712	0,693-0,731	2,426 p=0,119	-	-	-
Consultations with I.D specialist in the period (>3 vs <3))	1,648	1,277-2,128	14,904 p<0,001	1,505	1,154-1,965	p=0,003

†: Odds Ratio
‡ Confidence interval
§: Antiretroviral therapy
¶ Integrase strand transfer inhibitor

Table 3. Bivariate and multivariate regression analysis for the outcome "immunological reconstitution"

medRxiv preprint doi: <https://doi.org/10.1101/2024.10.26.24316181>; this version posted October 28, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

Variables	Crude OR† 95%CI	X2 p-value	Adjusted OR 95%CI	p value
Age	0,634 0,527-0,762	23,605 p<0,001	0,986 0,778-1,250	p=0,912
Sex	1,238 0,960-1,596	2,718 p=0,099	1,181 0,863-1,617	p=0,297
Female	1,200 0,967-1,489			
Male	0,969 0,932-1,007			
Viral load at the start of ART‡ (>0<100.000 cop/ml)	0,582 0,483-0,702	32,678 p<0,001	1,402 1,107-1,776	p=0,005
HIV Stage at diagnosis		556,626 p<0,001	Stage 2 0,078 0,005-0,019 Stage 3 0,010	p<0,001 p<0,001
HBV infection	1,174 0,543-2,540	0,167 p=0,683	0,935 0,360-2,428	p=0,891
HCV infection	0,438 0,088-2,177	1,076 p=0,300	0,200 0,025-1,547	p=0,123
Active tuberculosis at diagnosis	0,288 0,129-0,645	10,362 p=0,001	-	-
ART antiretroviral therapy with INSTI's¶ included	0,601 0,481-0,751	20,298 p<0,001	0,661 0,498-0,877	p=0,004
ART Changes	1,794 1,432-2,249	26,203 p<0,001	2,594 1,927-3,492	p<0,001
Adherence due to ART dispensing (>95%/11 months)	1,534 1,208-1,946	12,484 p<0,001	2,123 1,586-2,843	p<0,001
Failure with current ART	0,749 0,495-1,135	1,865 p=0,172	0,471 0,277-0,801	p=0,005
Chronic kidney disease	0,694 0,675-0,714	2,202 p=0,138	-	-
Coronary heart disease	0,439 0,062-3,120	0,717 p=0,397	-	-
Neoplasm not associated with AIDS	2,201 0,257-18,872	0,545 p=0,460	-	-
Consultations with I.D specialist in the period (>3 vs <3))	1,539 1,194-1,984	11,203 p<0,001	1,392 1,017-1,905	p=0,038

†: Odds Ratio
‡ Confidence interval
§: Antiretroviral therapy
¶ Integrase strand transfer inhibitor

Table 4. AUC/ROC Comparison Logistic regression models vs multilayer perceptron neural networks

Predictive models	AUC/ROC
Multivariate logistic regression	
Suppressed viral load	0,70
Undetectable viral load	0,66
Immunological reconstitution	0,83

Multilayer Neural Network/Perceptron	
Suppressed viral load	0,77
Undetectable viral load	0,69
Immunological reconstitution	0,87

Figure 1. Population selection flowchart

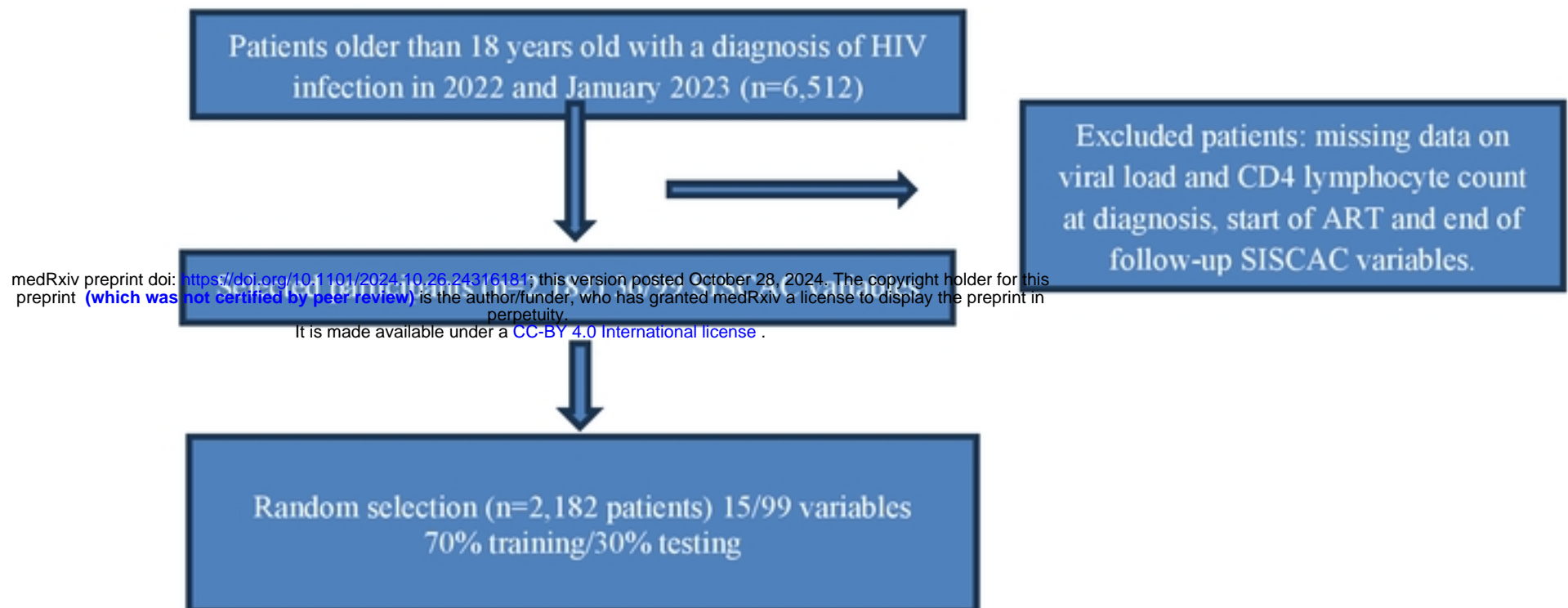


Figure 2. AUC/ROC for “undetectable viral load” multivariable logistic regression

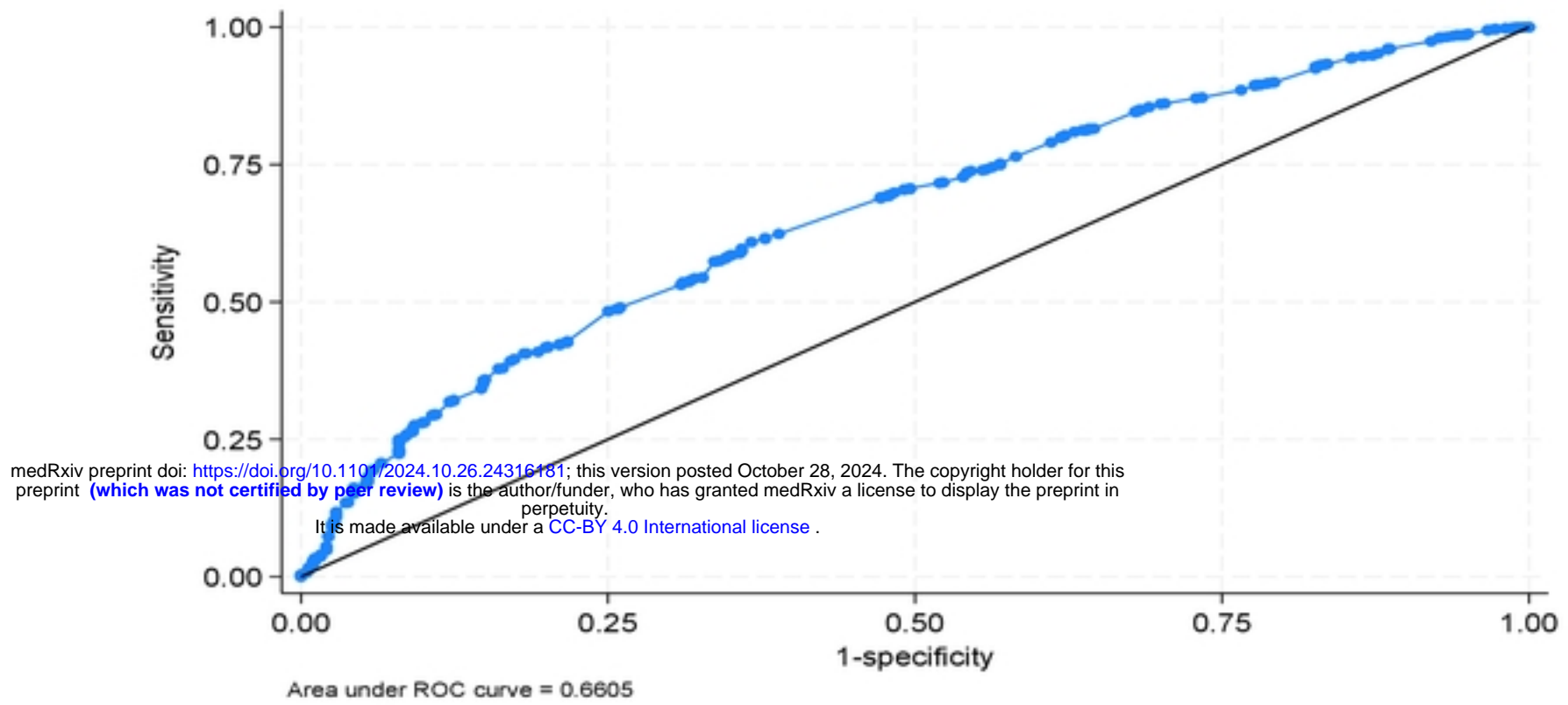


Figure 3. AUC/ROC for “immunological reconstitution” multivariable logistic regression

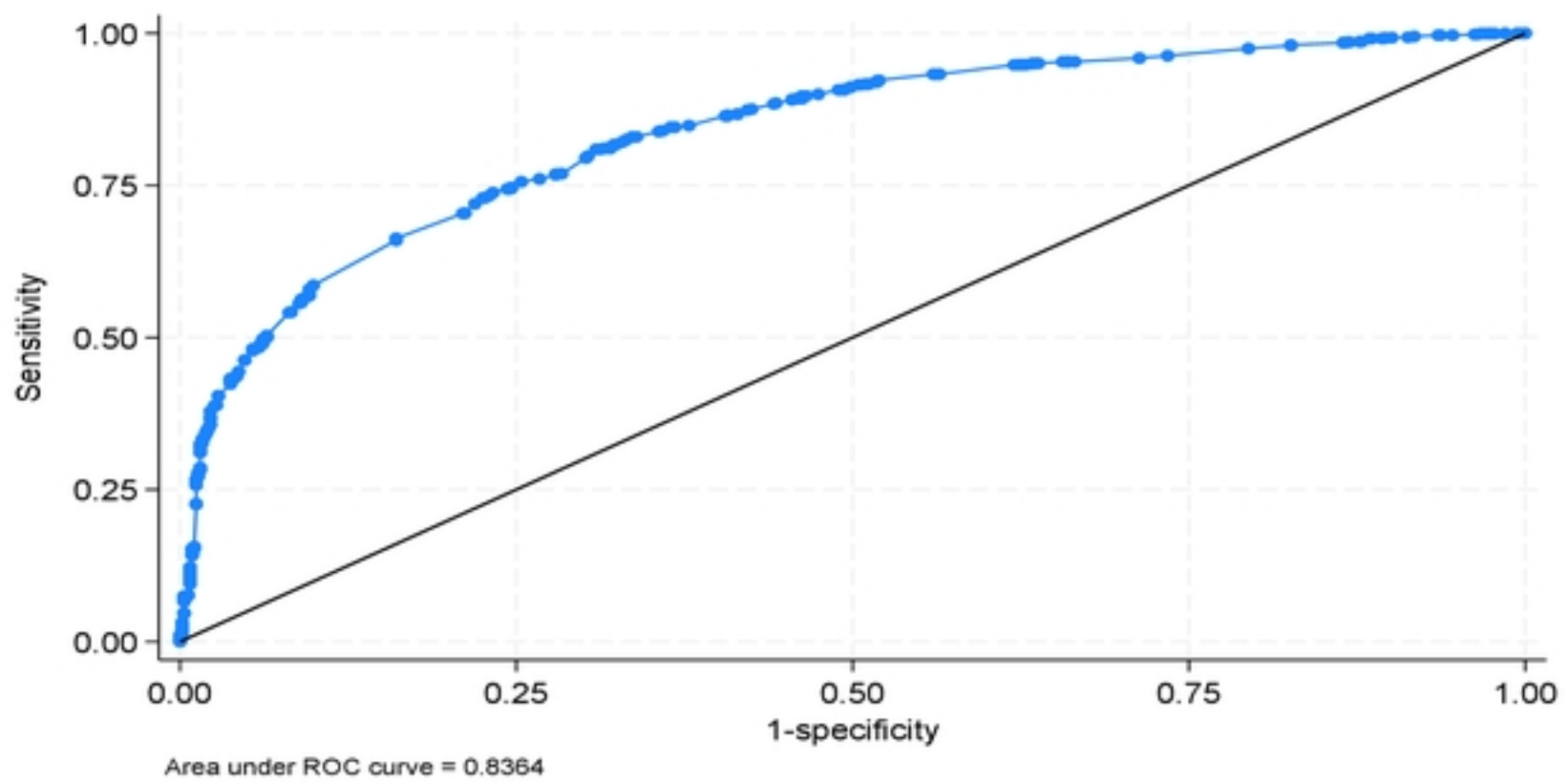


Figure 4. Artificial neural network for suppressed viral load.

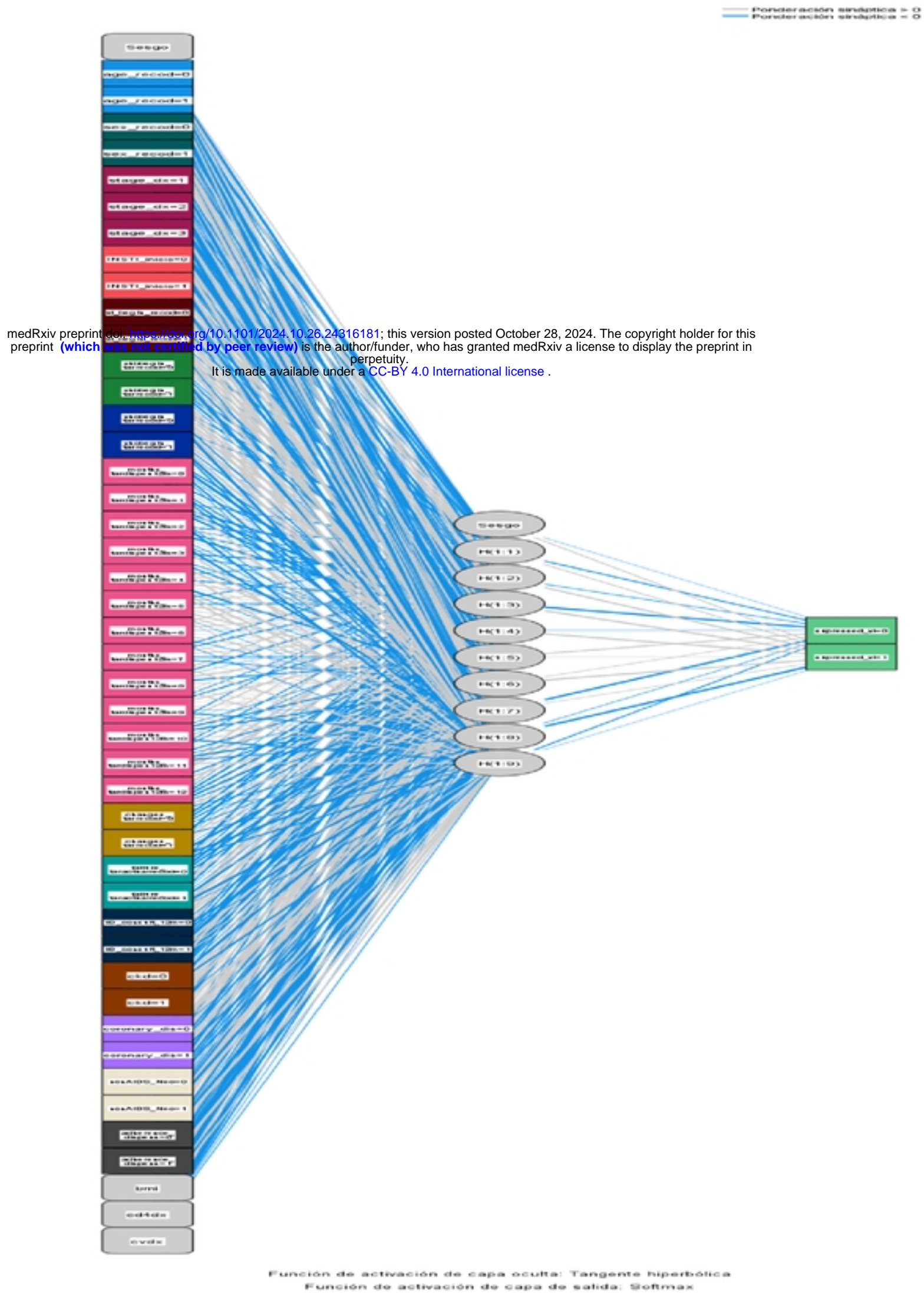
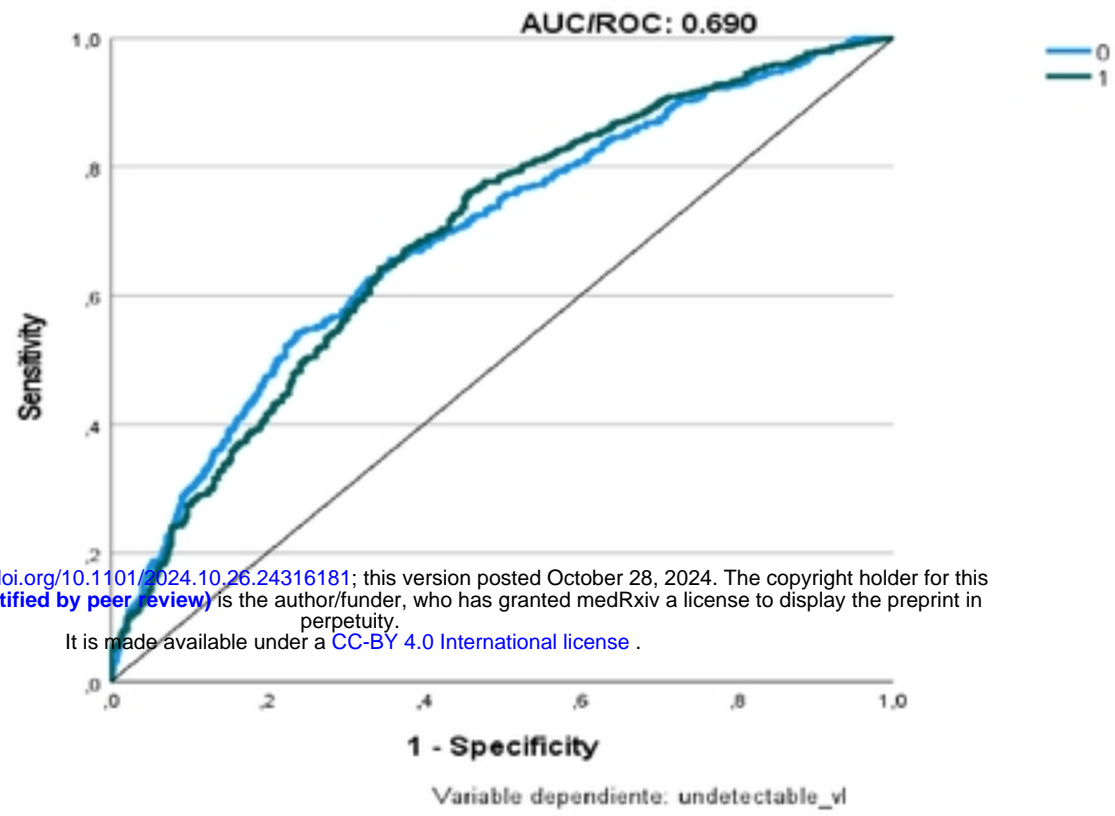


Figure 5. AUC/ROC curve for “undetectable viral load” ANN



medRxiv preprint doi: <https://doi.org/10.1101/2024.10.26.24316181>; this version posted October 28, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Figure 6. AUC/ROC curve for “immunological reconstitution” ANN

