Journal of Clinical Virology 118 (2019) 36–40

Contents lists available at ScienceDirect

Journal of Clinical Virology

journal homepage: www.elsevier.com/locate/jcv

Analytical performance evaluation and enhancement of the ADVIA Centaur® HIV Ag/Ab Combo assay

Baptiste Demey[a,b], Cédric Usureau[c], Paul Boillod[a], Sandra Bodeau[d], Catherine François[a,b], Catherine Roussel[a,b], Gilles Duverlie[a,b], Sandrine Castelain[a,b], Etienne Brochot[a,b,*]

[a] Department of Virology, Amiens University Medical Center, Amiens, France
[b] AGIR Research Unit, EA 4294, Jules Verne University of Picardie, Amiens, France
[c] Department of Oncobiology, Amiens University Medical Center, Amiens, France
[d] Laboratory of Pharmacology and Toxicology, Department of Clinical Pharmacology, Amiens University Hospital, Amiens, France

ARTICLE INFO

Keywords:
HIV
ADVIA centaur
Diagnosis
Immunoassay
Machine learning

ABSTRACT

Background: Fourth-generation immunoassays (such as the ADVIA Centaur® HIV Ag/Ab Combo (CHIV) assay) have improved the early diagnosis of human immunodeficiency virus (HIV), and their sensitivity and specificity usually exceed 99%. In regions with a low prevalence of HIV infection, however, the regular occurrence of false positives interferes with a medical laboratory's workflow. The additional reagent and staff costs associated with false positives can nevertheless be avoided or reduced by gaining a better knowledge of the CHIV assay's performance.

Objectives/study design: To improve our HIV diagnosis strategy, we retrospectively analyzed all the Centaur® CHIV assays and confirmatory tests performed at Amiens University Medical Center between 2012 and 2018. We used open-source machine learning software to process this large database, develop a predictive model, and identify a new cut-off for Centaur® CHIV index interpretation.

Results: A total of 56,682 HIV serological assay results were analyzed. The results of the CHIV assay were initially reactive or indeterminate for 449 samples. After p24 antigen and/or immunoblotting, there were 171 (38%) false positives and 278 (62%) confirmed true positives. The application of a cut-off of 2.12 led to reclassification of 130 of the 171 false positives as true negatives. Combining our predictive model with medical record analysis reduced the number of false positive CHIV assay results from 171 to 12.

Conclusions: The efficiency of the Centaur® CHIV assay can be increased by adjusting its cut-off for positivity. This adjustment may reduce the number of unnecessary confirmatory tests and accelerate the delivery of HIV test results.

1. Background

Human immunodeficiency virus types 1 and 2 (HIV-1 and HIV-2; genus: Lentivirus; subfamily: *Orthoretroviridae*; family: *Retroviridae*) are responsible for a global pandemic [1]. In the absence of treatment, HIV infection progresses to acquired immune deficiency syndrome (AIDS), which caused 940,000 deaths worldwide in 2017. In France, 40% of individuals living with HIV are not aware of their infection, and the median time interval between infection and diagnosis is 3.3 years [2]. Since 2009, and with a view to reducing this diagnostic delay, the French health authorities (*Haute Autorité de Santé*, HAS) have recommended (i) periodic HIV screening for high-risk individuals, and (ii) at least one test in a lifetime for the general population [3]. Hence, HIV testing accounts for a significant proportion of a medical laboratory's activity - especially in a University Medical Center such as ours. We use the ADVIA Centaur® HIV Ag/Ab Combo (CHIV) assay (Siemens Healthcare Diagnostics Inc., Tarrytown, NY), which simultaneously detects HIV p24 antigen (Ag) and HIV-1/2 antibodies (Abs) and expresses the result as an index [4]. Like other modern HIV screening assays, the CHIV assay is designed to be highly sensitive (> 99%) and specific (> 98%). On the basis of its own study populations, the manufacturer has defined a cut-off index of 1.0 [5]. However, the Positive Predictive Value (PPV) is not mentioned by the manufacturer in the performance evaluation of the assay. When a test gives an index above 1.0, the serum must be retested in duplicate after centrifugation. If the test is still positive, confirmation of the HIV infection using a different

* Corresponding author at: Laboratoire de Virologie, Centre de Biologie Humaine - CHU Amiens, Avenue René Laënnec, Salouel, F-80054 Amiens Cedex 1, France.
E-mail address: etienne.brochot@u-picardie.fr (E. Brochot).

https://doi.org/10.1016/j.jcv.2019.07.007
Received 14 June 2019; Received in revised form 22 July 2019; Accepted 24 July 2019

test (such as an immunoblot) is a legal requirement in France. These retests and confirmatory tests generate additional reagent and staff costs. For example, an immunoblot requires technician time, slows the transmission of the patient's test results to the clinician, and significantly increases the cost.

Thus, the primary objective of the present study was to evaluate the Centaur® CHIV assay in a large cohort of patients from 2012 to 2018 in a region with a low prevalence of HIV infections. The study's secondary objective was to use machine learning software (Orange Data Mining (6)) to refine our decision model and thus provide clinicians and laboratory diagnostic staff with more accurate, cheaper, and more rapidly available HIV screening test results.

We hypothesized that an artificial-intelligence (AI)-based analysis of reactive HIV serological assays in our population could (i) increase the Centaur® CHIV assay's cut-off index aiming an improved PPV without loss of sensitivity, and (ii) generate a decision tree for easy application in routine clinical practice.

2. Study design

2.1. Diagnostic tests

We retrospectively analyzed each Centaur® CHIV testing episode (defined as an initial Centaur® CHIV assay and, if necessary, a post-centrifugation duplicate retest and confirmatory test for a single serum sample) performed at Amiens University Medical Center between July 2012 and November 2018. All serum samples were tested with the Centaur® CHIV assay. According to the recommendations of the French accreditation standard (ISO15189), we have calculated an uncertainty zone around the cut-off, based on standard deviation of internal quality controls. Because of a variability of 20% in our laboratory for this assay, we considered that a measurement of 1.0 could be either 0.8 or 1.2. Thus, we defined three categories of test result: a non-reactive test (CHIV index < 0.8), an indeterminate test (CHIV index ≥ 0.8 to ≤ 1.20 and a reactive test (CHIV index > 1.2). Samples that gave reactive and indeterminate results were retested in duplicate after centrifugation at 10,000 x g for 10 min. If at least one of the retests was reactive or indeterminate, at least one additional confirmatory test was performed: a p24 Ag enzyme immunoassay (EIA) with confirmatory neutralization (VIDAS HIV P24 II, bioMérieux, Marcy l'Etoile, France) and/or an immunoblot assay (INNO-LIA HIV I/II Score, Fujirebio, Ghent, Belgium)). The confirmatory tests were carefully chosen on the basis of the clinical information and the HAS's (*Haute Autorité de Santé*) diagnostic algorithms [3].

Depending on the results of the additional tests, reactive and indeterminate CHIV assay results were then classified as true positives (i.e. at least one positive confirmatory test) or false positives (i.e. negative confirmatory tests). As recommended by the manufacturer, all the non-reactive CHIV assays (CHIV index < 0.8) were considered to be true negatives for both the Ag and the Abs.

2.2. Ethics

Consent is not needed for retrospective interpretation of anonymous and confidential routine care data in accordance with the French Jarde law.

2.3. Statistical analyses

For the analyses of continuous variables, CHIV indexes outside the linear range were converted from ' < 0.05' to '0' and from ' > 12' to '12'. All statistical analyses (including sensitivity and specificity analyses and receiver operating characteristic (ROC) curve analyses) were performed using GraphPad Prism software (version 5, GraphPad Software, San Diego, CA, USA).

2.4. Data mining

Our data were managed and visualized using Orange Data Mining, an open-source software package for visual programing based on Python [6]. To perform data mining on positive results and improve PPV, we used case and control groups of equivalent size. In order to balance our groups, an equal number of non-reactive results and reactive/indeterminate results were randomly selected through a command of Orange Data Mining. Next, we used Scikit-learn [7] (a Python AI package for machine learning with Orange Data Mining) to build a classification tree that featured a sensitivity of 100% below the newly defined CHIV cut-off index. To evaluate the classification tree we performed a stratified 10-fold cross validation on the data set.

3. Results

3.1. Testing episodes

Between July 2012 and November 2018, 56682 Centaur® CHIV testing episodes were recorded at Amiens University Medical Center. In all, 56233 tests were non-reactive (and were considered as true negatives (TNs)), 51 were repeatedly indeterminate, and 398 were repeatedly reactive after centrifugation (Fig. 1). Of the 449 indeterminate and reactive testing episodes, 171 (38%) were found to be false positives (FPs) and 278 (62%) were confirmed to be true positives (TPs) after a p24 Ag EIA and/or an immunoblot.

3.2. Specificity and sensitivity analyses

False-positive Centaur® CHIV testing episodes accounted for 0.30% of the total, with one every two weeks on average. The median (range) CHIV index was 1.4 (0.84–12) for FPs, 12.0 (2.15–12) for TPs, and 0.10 (0-0.79) for TNs.

We next drew up a ROC curve for the whole set of 56,682 testing episodes; the area under the curve was 0.9999 (Fig. 2). As indicated by Siemens Healthcare Diagnostics Inc., the sensitivity was 100% for a CHIV index of 1.0. However, the analysis of our ROC curve revealed that the sensitivity was 100% up to a cut-off of 1.2 (the upper limit for an indeterminate result). More interestingly, we did not observe a loss of the sensitivity until a cut-off of 2.12 while the PPV was greatly improved. In fact, for cut-offs of 0.8, 1.0, 1.2 and 2.12, the PPV was respectively 61.92%, 65.72%, 71.83%, and 86.88%.

3.3. Cumulative frequency

The cumulative frequencies of the CHIV indexes for TPs, FPs and TNs are shown in Fig. 3; the graph highlights the differences in distribution between the three categories. Most of the TPs presented a very high CHIV index, close to or beyond the upper limit of the linear range. The TNs were characterized by a low CHIV index, with few results (0.10%) above the lower indeterminate cut-off of 0.8. Centaur® CHIV assays giving an index of between 0.8 and 1 accounted for only 17.54% of the FPs. In contrast, 38.60% of the FPs had a CHIV index below the upper indeterminate cut-off (1.2). More interestingly, 76.02% of the FPs and none of the TPs presented a CHIV index below 2.12.

3.4. Artificial-intelligence-based classification

By using an Orange Data Mining command, 450 CHIV indexes from TNs were randomly selected and compared with 449 CHIV indexes from indeterminate and reactive testing episodes. This sample population allowed us to specifically compare CHIV indexes from FPs with those from TPs and TNs, and thus to improve the machine learning software's performance for classification of FPs in either TPs or TNs.

Then, we configured the generation of a classification tree with the Centaur® CHIV indexes as features and the confirmatory tests results as

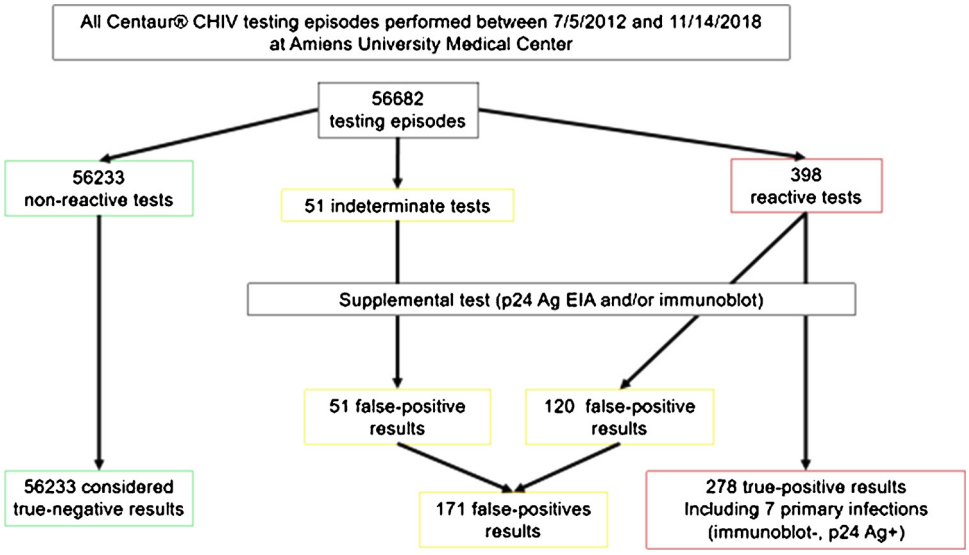

Fig. 1. The flow chart for all the Centaur® CHIV testing episodes analyzed. The results presented here correspond to the final Centaur® CHIV results after centrifugation and retesting in duplicate (if required). Ag: Antigen. EIA: enzyme immunoassay.

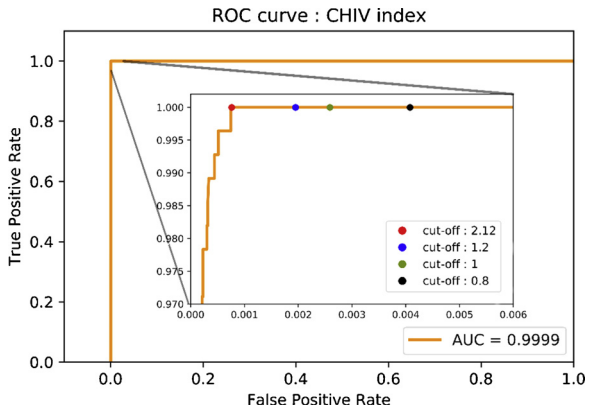

Fig. 2. A receiver operating characteristic (ROC) curve for the Centaur® CHIV index. The inset shows the inflection zone in more detail. AUC: area under the curve.

targets (Fig. 4A). The tree highlighted two major values to assist in the interpretation of the Centaur® CHIV index superior to the cut-off in our population. We found 284 CHIV index higher than 4.75 and 272 were TPs, which corresponds to a PPV of 95.8%. For the 250 CHIV index superior to 9.12, the actual positivity was confirmed for 247 or them (PPV: 98.8%).

The decision tree was evaluated using a stratified 10-fold cross-validation method from our (AI)-study sample (i.e. the 899 results). As shown in Fig. 4B, the area under the ROC curve for the predictive model was 0.991, and the accuracy was 98% (i.e. 881 correct predictions out of 899 iterations). Only 12 of the 899 Centaur® CHIV testing episodes (1.33% of the sample, PPV: 95.86%) were misclassified as predicted positive confirmatory tests and constituted FPs. Thus, the predictive tree yielded 14 times less FPs than in the initial data with CHIV index ranging from 0.8 to 1.2 (171 out of 899; 19.03% of the sample; PPV: 61.92%). The advantage of using a machine learning algorithm is to obtain suggestions of decision based on probabilities. The errors concerned indexes ranging from 4.75 and 12 with probability of positivity between 63.63% and 99.10%. Consequently, when the CHIV index is over 2.12, the classification tree offers predictions more efficient to exclude FPs but the final decision is still dependent of the laboratory diagnostic staff's interpretation.

4. Discussion

To the best of our knowledge, the present study constitutes yet the largest independent evaluation of the Centaur® CHIV assay's performance on non-selected inpatient sera. The previous studies were based on 831 to 3020 selected sera, and reported levels of performance

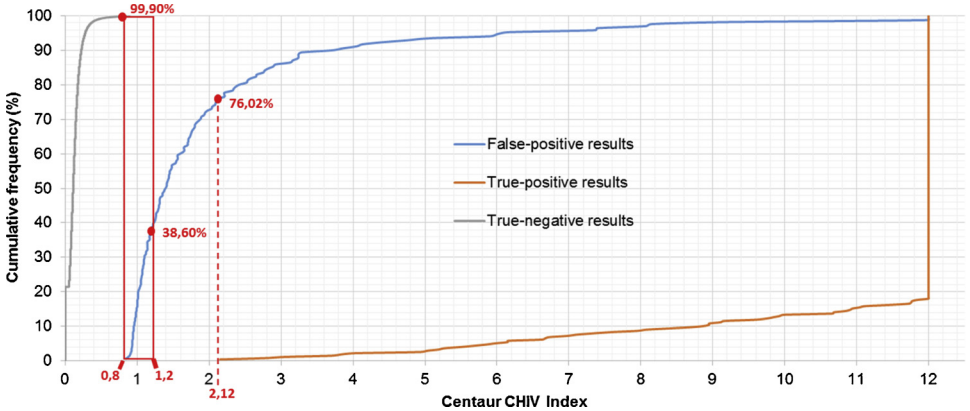

Fig. 3. Cumulative frequency of whole true-negative, false-positive and true-positive Centaur® CHIV testing episodes between 2012 and 2018, as a function of the Centaur® CHIV index.

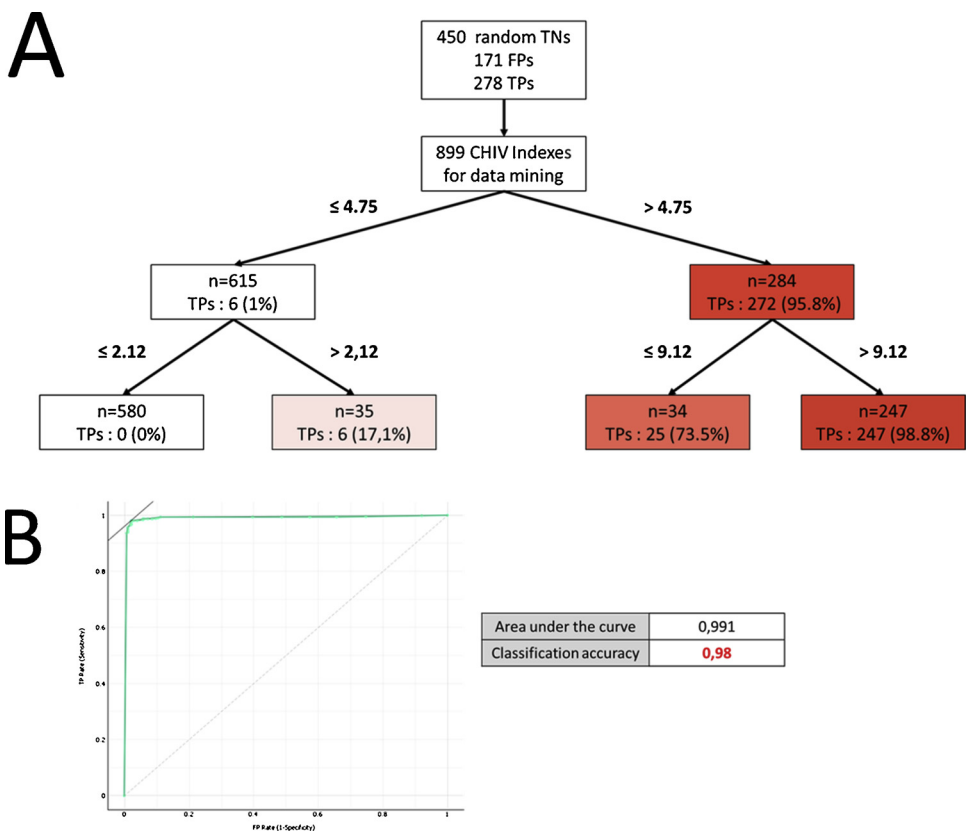

Fig. 4. A/ Illustration of the classification tree generated with the Orange Data Mining Toolbox, using the data on 449 indeterminate/reactive Centaur® CHIV testing episodes, 450 negative assays, and the associated confirmatory tests. The text boxes contain the details of the classification in terms of number and percentage for true positivity depending on the CHIV index. B/ ROC curve and accuracy for the classification tree according to the 10-fold cross validation. True Negatives: TNs. False Positives: FPs. True positives: TPs.

similar to those published by the manufacturer [8–10].

The frequency of FPs in our retrospective study (0.30%) was similar to that reported by the manufacturer (0.26%: 20 out of 8372 samples collected from individuals with HIV infection, blood donors, and hospitalized patients) [5]. This result is very low because the great majority of the tests were negative (93% in Pumarola et al.'s study, and 99% in the present study). In our center, a FP CHIV assay result occurs once every 14 days, on average. Even though Siemens Healthcare Diagnostics Inc. recommends the centrifugation and duplicate retesting of initially reactive samples, most of the latter remain positive and thus require the performance of confirmatory tests. Our present results demonstrated that a large proportion of these additional tests for a reactive Centaur® CHIV assay were avoidable. In fact, 38.6% of the patients with a reactive CHIV assay had an index below 1.2 and a negative immunoblot assay.

Given that the CHIV cut-off index of 1.0 was established by Siemens HealthCare Diagnostics Inc. based on ROC analysis following CLSI guidelines (package insert) in order to obtain a perfect sensitivity, we believe that Siemens could readjust this threshold of 1.0 while maintaining perfect sensitivity but at the same time increasing its specificity and especially its positive predictive value. In our study, a cut-off of 2.12 combined a sensitivity of 100% with greater specificity. The application of this value would make the interpretation of Centaur® CHIV results more efficient, by reducing the number of FPs by 76.02% which leads to a PPV improvement from 61.92% to 86.88%. Moreover, we used our data to build a classification tree that could help the laboratory diagnostic staff in CHIV index > 2.12 interpretation. The application of the algorithm highlighted a PPV up to 95.86%. This would have saved both time and money (additional test reagents, etc.). Overall, these improvements would accelerate the delivery of HIV test results for poorly reactive assays, and receiving a negative result faster could improve care.

We built a large database from the HIV serological assays performed in Amiens University Medical Center over the past 6 years, and used it to develop an easy-to-visualize, AI-based classification tree with several useful applications. The simplest approach is to check the CHIV index against the decision tree and thus infer the likelihood of a positive serological assay for HIV. More interestingly, new data could be incorporated into the algorithm, which could precisely predict the likelihood of a positive or negative sample. We made a cross-validation of the algorithm against our own data, which is commonly practiced in machine learning, but constitutes a source of bias which could be removed by using data from an external cohort. Moreover, the present predictive model was programmed to predict one of two categories if the probability was above 50% for "positive" or "negative". This involved the prediction of remaining FPs (12 out of 899). In such case, the likelihood of a positive sample ranged from 63.63% (with a CHIV index of 7.39) and 99.1% (with an index of 12). In particular, the single CHIV index of 12 with negative p24 Ag and immunoblot was for an 82-year old patient from oncology without HIV risk factor or background, and with other unexplained serological interferences (also with negative confirmatory tests). This value extracted from routine practice strongly influences the classifier but cannot be ignored. Thus, in the absence of a predictive probability of 100%, the model's result should be carefully interpreted by an expert laboratory diagnostic staff. Nevertheless, this new computer-based tool might be of value in clinical practice. Furthermore, the incorporation of more data (e.g. results from other centers, and additional variables from biochemical and hematological assays) would enable us to refine the predictive model and test more sophisticated designs (e.g. random forest and k-nearest neighbors methods) [11–13].

In France, however, the application of a revaluated CHIV cut-off value is limited by the legislation; the laboratory diagnostic staff is legally bound to follow the manufacturer's recommendations. Hence, a Centaur® CHIV testing episode with index greater than 1.0 must still always be initially considered as a positive result requiring further testing. Even when a second 4th-generation HIV serological assay is negative, the inconsistency between the two results means that the

laboratory is legally obliged to perform an additional test [14].

Despite technical progress in the automation of HIV testing, new-generation systems (such as the Atellica®) are still based on the Centaur® CHIV assay [15]. It seems that sharing data such as these would be important in working with the manufacturers of the assay. In this context, adjustment of the CHIV index on other analyzers might also increase the quality of HIV diagnosis.

Funding

Not applicable.

Author's contribution

BD and EB conceived of the presented idea.

BD, CU, SC, GD and EB wrote the manuscript with support from.

BD, PB, and SB performed the analytic calculations and performed statistical analysis.

SC, CF and CR supervised the project.

All authors discussed the results and contributed to the final manuscript.

Declaration of Competing Interest

No conflict of interest.

References

[1] UNAIDS, Global AIDS Monitoring 2018, Indicators for Monitoring the 2016 United Nations Political Declaration on Ending AIDS, (2018).

[2] HIV/AIDS Surveillance in Europe 2018–2017 Data [Internet], European Centre for Disease Prevention and Control, 2018 [cited 14 March 2019] Available from: http://ecdc.europa.eu/en/publications-data/hivaids-surveillance-europe-2018-2017-data.

[3] Haute Autorité de Santé, Réévaluation de la stratégie de dépistage de l'infection à VIH en France, Recommandations 25 (2017).

[4] Siemens Healthcare Diagnostics, Product Package Insert. ADVIA Centaur HIV Ag/Ab Combo (CHIV), Siemens Healthcare Diagnostics, Tarrytown, NY, USA, 2015.

[5] T. Pumarola, J. Freeman, E. Saxton, P. Dillon, T. Bal, J. van Helden, Performance evaluation of the ADVIA Centaur® HIV Ag/Ab combo assay, J. Virol. Methods 170 (December (1–2)) (2010) 16–20.

[6] J. Demšar, T. Curk, A. Erjavec, Č Gorup, T. Hočevar, M. Milutinovič, et al., Orange: data mining toolbox in python, J. Mach. Learn. Res. 14 (January (1)) (2013) 2349–2353.

[7] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., Scikit-learn: machine learning in python, J. Mach. Learn. Res. 12 (November) (2011) 2825–2830.

[8] B. Bahar, 4th Generation HIV testing false positivity by Siemens ADVIA Centaur® XP: a retrospective study, Turk. J. Biochem. 37 (4) (2012) 405–406.

[9] X. Qiu, L. Sokoll, P. Yip, D.J. Elliott, R. Dua, P. Mohr, et al., Comparative evaluation of three FDA-approved HIV Ag/Ab combination tests using a genetically diverse HIV panel and diagnostic specimens, J. Clin. Virol. 92 (July) (2017) 62–68.

[10] L. Vallefuoco, F. Aden Abdi, R. Sorrentino, D. Spalletti-Cernia, C. Mazzarella, S. Barbato, et al., Evaluation of the siemens HIV antigen-antibody immunoassay, Intervirology 57 (2) (2014) 106–111.

[11] G. Riddick, H. Song, S. Ahn, J. Walling, D. Borges-Rivera, W. Zhang, et al., Predicting in vitro drug sensitivity using random forests, Bioinformatics 27 (January (2)) (2011) 220–224.

[12] Y.V. Sun, L.F. Bielak, P.A. Peyser, S.T. Turner, P.F. Sheedy, E. Boerwinkle, et al., Application of machine learning algorithms to predict coronary artery calcification with a sibship-based design, Genet. Epidemiol. 32 (May (4)) (2008) 350–360.

[13] B. Wu, T. Abbott, D. Fishman, W. Mcmurray, G. Mor, K. Stone, et al., Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data, Bioinformatics 19 (2003) 1636–1643.

[14] JORF, Arrêté du 28 mai 2010 fixant les conditions de réalisation du diagnostic biologique de l'infection à virus de l'immunodéficience humaine (VIH 1 et 2) et les conditions de réalisation du test rapide d'orientation diagnostique dans les situations d'urgence. Sect. n° 31, 0.131, 9 June 9 2010 p. 10572.

[15] Chemistry and Immunoassay Menu, Siemens Healthcare Inc., 2018 [Internet]. [cited 14 March 2019]. Available from: https://static.healthcare.siemens.com/siemens_hwem-hwem_ssxa_websites-context-root/wcm/idc/groups/public/@global/@lab/documents/download/mda4/oda3/~edisp/30-18-11102-01-76_cross-prd_menu_ous_fnl_pgs-05988953.pdf.