Informatics in Medicine Unlocked 24 (2021) 100568

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: http://www.elsevier.com/locate/imu

A transfer learning approach to drug resistance classification in mixed HIV dataset

Moses E. Ekpenyong [a,*], Mercy E. Edoho [a], Ifiok J. Udo [a], Philip I. Etebong [a], Nseobong P. Uto [b], Tenderwealth C. Jackson [c], Nkem M. Obiakor [d]

[a] Department of Computer Science, University of Uyo, Nigeria
[b] School of Mathematics and Statistics, University of St Andrews, Scotland, UK
[c] Department of Pharmaceutics and Pharmaceutical Technology, University of Uyo, Nigeria
[d] College of Pharmacy, Gregory University, Nigeria

ARTICLE INFO

Keywords:
Adverse drug reaction
Treatment failure
Individualized therapy
HIV/AIDS
Semi-supervised learning
Transfer learning

ABSTRACT

As we advance towards individualized therapy, the 'one-size-fits-all' regimen is gradually paving the way for adaptive techniques that address the complexities of failed treatments. Treatment failure is associated with factors such as poor drug adherence, adverse side effect/reaction, co-infection, lack of follow-up, drug-drug interaction and more. This paper implements a transfer learning approach that classifies patients' response to failed treatments due to adverse drug reactions. The research is motivated by the need for early detection of patients' response to treatments and the generation of domain-specific datasets to balance under-represented classification data, typical of low-income countries located in Sub-Saharan Africa. A soft computing model was pre-trained to cluster $CD4^{+}$ counts and viral loads of treatment change episodes (TCEs) processed from two disparate sources: the Stanford HIV drug resistant database (https://hivdb.stanford.edu), or control dataset, and locally sourced patients' records from selected health centers in Akwa Ibom State, Nigeria, or mixed dataset. Both datasets were experimented on a traditional 2-layer neural network (NN) and a 5-layer deep neural network (DNN), with odd dropout neurons distribution resulting in the following configurations: NN (Parienti et al., 2004) [32], NN (Deniz et al., 2018) [53] and DNN [9 7 5 3 1]. To discern knowledge of failed treatment, DNN1 [9 7 5 3 1] and DNN2 [9 7 5 3 1] were introduced to model both datasets and only TCEs of patients at risk of drug resistance, respectively. Classification results revealed fewer misclassifications, with the DNN architecture yielding best performance measures. However, the transfer learning approach with DNN2 [9 7 3 1] configuration produced superior classification results when compared to other variants/configurations, with classification accuracy of 99.40%, and RMSE values of 0.0056, 0.0510, and 0.0362, for test, train, and overall datasets, respectively. The proposed system therefore indicates good generalization and is vital as decision-making support to clinicians/physicians for predicting patients at risk of adverse drug reactions. Although imbalanced features classification is typical of disease problems and diminishes dependence on classification accuracy, the proposed system still compared favorably with the literature and can be hybridized to improve its precision and recall rates.

1. Introduction

Over the last three decades, diverse drug therapies for the treatment of HIV/AIDS infection have been introduced, with six distinct therapy classes based on drug molecular mechanism and resistant profiles [1]. While ongoing research efforts have advanced towards simulation agents that target HIV reservoirs [2], interests in modeling drug resistance have increased [3], with mathematical models of antiretroviral therapy showing improved precision in risk assessment of drug resistance at individual and population levels [4]. However, state-of-the-art empirical models have failed to capture heterogeneities when assessing the risk of drug resistance among individuals due to model variable diversity [5,6], and its robustness to model real-life distribution. Many soft computing approaches have been studied in the past to provide solutions to diverse real-life problems in various sectors of the society. This quest has engaged scholars to evolve applications in different fields

* Corresponding author. Department of Computer Science, University of Uyo, P.M.B. 1017, 520003, Uyo, Akwa Ibom State, Nigeria.
E-mail address: mosesekpenyong@uniuyo.edu.ng (M.E. Ekpenyong).

https://doi.org/10.1016/j.imu.2021.100568
Received 26 December 2020; Received in revised form 30 March 2021; Accepted 31 March 2021
Available online 5 April 2021

namely, bioinformatics, image processing, medical diagnosis, and many more [7,8]. Studies on comprehensive analysis of nature-inspired meta-heuristics utilized in the domain of feature selection [9], queries optimization for serpentine database [10], spam detection in marketing [11] and financing using high-end performance analytics [12], are a few supporting studies, which results have proven the effectiveness of the approach. In these studies, the areas of application adopted either a single or combined machine learning algorithm(s), such as, decision tree, k-nearest neighbor, support vector machine, random forest, naive Bayes, and multi-layer perceptron. However, deep learning has evolved from Artificial Neural Networks (ANNs) to proffer higher classification performance to traditional machine learning, due to its ability to perform automatic feature extraction in high-dimensional datasets generated from high-end performance systems such as healthcare, engineering, e-commerce, cybernetics, security, power management, and more. Deep learning is widely applied in research on natural language processing, pattern recognition, robotics, information systems and so on, with a view to discovering intrinsic relations and patterns that are embedded within large datasets.

In traditional machine learning, the assumption is that the train and test data reside in same feature space or domain and have same data distribution characteristics. But a new task with different data distribution will require a new model based on the current data. Two outstanding limitations of this method are that a previous model cannot be deployed to model the intended system, and proprietary tools may be required to develop a new model. Inspired by the ability of humans to intelligently learn from experience and attempt to solve inactive problems, transfer learning has come to the rescue, as this approach accelerates the learning process for improved solutions. In contrast to traditional machine learning, transfer learning applies knowledge discovered from other sources to the target task, while tolerating variations that exist within disparate data distributions. Applications of transfer learning have covered the following areas: learning from simulations, adapting to new domains, and transferring knowledge across different domains. Three main problems that define research progress in transfer learning include what to transfer, how to transfer and when to transfer. Transfer learning as subclassified in Fig. 1, may be *inductive*–where there are disparate tasks, regardless of the similarities in the source and target domain [13–16]; *transductive*–where there are same tasks but diverse feature spaces or marginal probability distributions in both domains [17–19]; or *unsupervised*–where there are no labeled data to utilize for training [20–22].

To improve the performance of existing transfer learning methods and handle the knowledge transfer process in a more natural way (typical of real-world systems), nature inspired solutions that integrate transfer learning methods have emerged. These techniques include artificial neural network-based transfer learning [24]; evolution system-based transfer learning [25]; fuzzy system-based transfer learning [26], and hybrid or collaborative system-based transfer learning [27]. In real-world applications, training data for domain-specific predictions are limited but sizeable amount of control data can be exploited in different feature space or distribution. For such cases, transfer learning could improve the performance of predictive algorithms on the test data of the target task.

This paper proposes a transfer learning approach that pretrains known domain features for subsequent reuse in classifying new datasets. A domain here represents a feature space of inputs from a specific environment. Our approach permits adaptation to new datasets with labeled drug types for efficient modelling of drugs interaction and failed treatment classification. The proposed system can also serve as an expert support system for drug regimen recommendation since the developed models and algorithms could be made available as open-source tools with adaptive and replicable features for diverse domains/environments.

The remaining sections of the paper are organized as follows: Section 2 outlines the objectives of the research and contributions to knowledge.

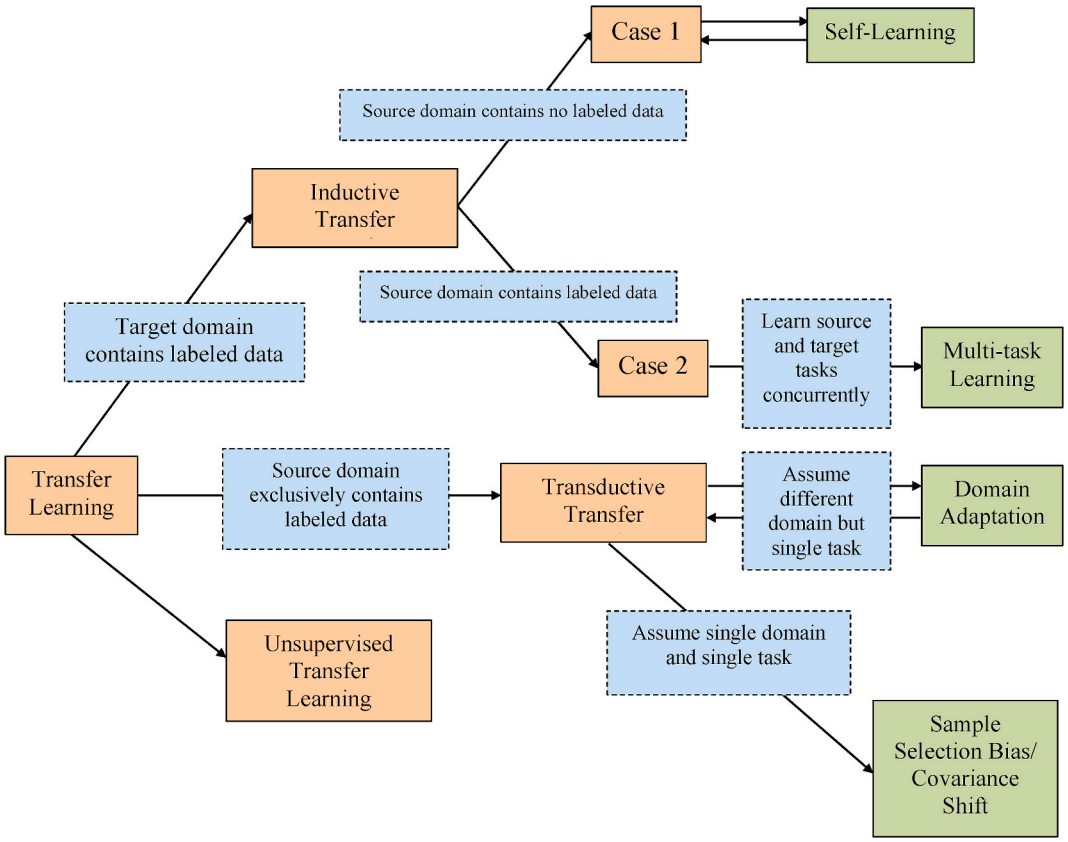


Fig. 1. Sub-categories of transfer learning (Adapted from Ref. [23]).

Section 3 reviews existing literature, exposing the research implications and gaps, with a justification and impact of the research. Section 4 discusses the materials and methods used. Section 5 presents the proposed system model. Section 6 discusses the results, scientific implications, and study limitation. Section 7 concludes on the study and offers future research perspectives.

2. Specific objectives and contributions to knowledge

The specific objectives of the study include, to.

- process input data from a publicly available database (control dataset) and local data from patients records (mixed dataset), for studying failed treatment in HIV patients.
- label classification targets of both datasets using available HIV prognostic markers (before and follow-up CD4 counts, before and follow-up viral loads) and patients' response to drugs (regimens) interaction derived from type-2 fuzzy logic system.
- learn the input datasets using traditional machine learning and transfer learning algorithms, for efficient classification and prediction of failed treatments.
- compare both machine and transfer learning algorithms using suitable performance metrics, for optimal system selection.

The contributions of this paper to knowledge include:

- *Real time diagnose of patients with failed treatment:* The risk associated with treatment failure is very high in low- and medium-income countries (LMICs) and very critical to the life of a patient. To the best of our knowledge no previous works have entertained the use of transfer learning for failed treatment prediction/classification. The proposed system provides a framework for implementing real time detection of failed treatment cases and offers useful decision support for the early diagnose of patients with failed treatment.
- *Informed therapy administration:* Aside the challenge of diagnosing patients with failed treatment, there is need for early administration of appropriate regimen. This paper has industrial applications in the growing field of medicine. To physicians, it resolves the puzzle of relying on similar experience when cases of failed treatment present themselves, as the possibility of predicting patient's response with certainty is guaranteed. In the field of drug delivery and administration, it illuminates the path towards individualized therapy–as the study of on-treatment variables influencing a set of study outcomes is achievable.
- *Generalizable performance across new datasets:* Our transfer learning model can generalize across new HIV datasets, as lower RMSE values were obtained using the transductive-deep learning classifier. Furthermore, a replicable and adaptive model is sure. It is replicable because available datasets can be exploited to produce similar results. It is adaptive because transfer learning methods provide opportunities to change the behavior of the system in response to its environment or domain.
- *Control database for a Sub-Saharan African (SSA) Country:* Access to clinical data is crucial for advancing the course of HIV research. This paper produces a very useful resource that will engender future research on HIV/AIDS in SSA.

3. Related works

In this section, a comprehensive review of previous research works is presented. These works are classified into two major groups namely: statistical and computational approaches to HIV treatment failure and transfer learning approaches to disease prediction/classification. The second group is predicated on the fact that, to the best of our knowledge no previous works have applied transfer learning methods to the treatment failure of any disease condition. Summaries of the two groups of related works documenting the Authors; Objectives; Methodologies, Tools, Data(bases); Findings and Drawbacks are presented in Table 1 and Table 2.

3.1. Statistical and computational approaches to HIV treatment failure: implications

Studies on the use of electronic health or medical records for predicting treatment failure lack complete datasets and are less accurate than clinical trials. These studies which are mostly limited to single cohorts [28] require further clinical confirmation, hence, resulting in incomplete, inconclusive, or inaccurate datasets [29]. Employing interview sessions in data collection can restrict the number of study variables, because patients may develop depression or the fear of being stigmatized; and variation in quality of healthcare services may differ from hospital to hospital. Understanding the use of predictors may guide programs to develop interventions for identifying patients with adverse drug reactions (ADRs) and for implementing prevention strategies. However, studies have employed feature selection techniques [30] to reveal contributors to ADRs. The Kaplan-Meier curve has been useful to investigate predictors of immunologic and virologic failures and has been the most widely used measure [31–33], but its use has failed to accurately estimate the actual CD4 count because of the lack of primary data, leading to uncertain estimates, reliance on self-reporting of virologic failures, and overestimation of therapy adherence. Besides the reason that secondary data have ignored key variables, most of the ones used in the literature rely on immunologic failure. Adopting genome sequencing [34] of patients with failed virologic antiretroviral therapy (ART) could not detect minority resistant viral strains thereby underestimating resistance cases.

Evaluating drug resistance patterns has mostly been achieved through simple descriptive statistics and limited to certain subtype test feature cutoffs [35]. Statistical methods such as binary, conditional, univariate, bivariate, or multivariate logistic regression [29,36–44] have dominated the literature for investigating, analyzing, or modeling risk factors, but such investigations are retrospectively done and are either burdened by under- or over-estimation of failure prevalence. However, combining regression and machine learning techniques [45], where the latter is used to train and test the models on practical datasets have proven to be efficient over traditional regression models. Consequently, comparing machine learning algorithms may enhance proper choice of improving prediction performance and avoid the over-fitting and under-fitting of regression models.

Convolutional learning models have been found to outsmart other state-of-the-art models [46], but they perform poorly with limited data. The sparsity of data does not enable true generalization of findings, as some samples-initiated ART are at an advanced stage of HIV due to resource-limited situation, comorbidity, mortality, and viral transmission. Although estimating relative risk using generalized linear models [47] does not allow for causal inference of failure rates because adherence data is usually self-reported, the use of soft computing techniques has helped to model the causal inference of failure by classifying patients at risk of ADRs [48,49]. Control datasets of failed treatment is therefore certain to improve the prediction of patients with failed treatment.

3.2. Transfer learning approach to disease prediction/classification: implications

Methodologies of most of the understudied works consider medical image extraction and classification [50–54], but these methods are not necessarily optimal in terms of predictive performance, as development of robust discriminative classifiers is still incipient. However, reported increase in accuracy usually comes at the expense of increased training time complexity [54] due to the use of multiple models such as deep convolutional network and their hybrid variants with transfer learning

Table 1
Previous works on statistical and computational approach to HIV treatment failure.

Reference	Objective	Method, tool, data(base)	Result/Finding	Drawback/Limitation
Parienti et al. [32]	To investigate virologic failure in HIV-infected patients.	Method, tool: Kaplan-Meier curves were constructed according to factors associated with virological failure, with proportional-hazard model used to assess the independent effects of each factor. To account for missing data, patients with viral load of <400 copies/mL were censored at last follow-up visit. Data(base): Previously identified consecutive HIV-infected patients with HAART controlled viral loads.	Virologic failure was significantly associated with previous nonadherence behaviors. Resistance to NNRTIs occurred at low levels of adherence.	Some (5/20) genotype data were missing. Self-reporting by patients overestimated adherence to therapy.
Dragsted et al. [31]	To investigate predictors of immunological failure after initial $CD4^+$ response.	Method, tool: Changes in $CD4^+$ cell count and proportion of patients with a pVL <400 copies/μL between starting HAART and the end of the study were calculated. Kaplan-Meier survival curves were used to describe the median time to a modified therapy or immunological failure. Data(base): EuroSIDA HIV type 1 cohort database (https://clinicaltrials.gov/ct2/show/NCT02699736).	The risk of immunological response to HAART diminished with long term treatment.	The study could not ascertain the actual number of $CD4^+$ cells at the time of failure.
Robbins et al. [28]	To identify predictors of ART treatment failure.	Method, tool: A retrospective analysis of HIV-infected patients was conducted with an HIV RNA measurement of ≤400 copies/mL on ART. Predictors of failure were assessed using proportional hazards modeling. Data(base): Electronic health records (EHRs) for patients in an urban HIV clinic	Poor adherence identified as the cause of treatment failures.	• Study was limited to a single cohort. • EHRs are incomplete and less accurate than clinical trial.
Hosseinipour et al. [35]	To evaluate drug resistance patterns among patients with failed first-line ART.	Method, tool: Patients with ART failure (s) were evaluated. Genotyping was performed for those with RNA >1000 copies/mL. Phenotyping was performed for complex genotype pattern. Simple descriptive statistics (mean, median, proportion, Student's *t*-test, Wilcoxon rank-sum, and chi-square) were used. Data(base): Patients with ART failure from January 2007 to July 2007.	96 confirmed ART failures. $CD4^+$ cell count criteria are associated with resistance profiles that markedly compromise the activity of second-line ART	Only phenotype cut-offs of patients infected with HIV subtype B were considered.
Hamers et al. [34]	To assess HIV-1 drug resistance mutations (DRMs).	Method, tool: HIV-1 sequences were generated for 142 participants with virologically failed ART. Group comparisons for categorical and continuous data were then performed using χ^2/Fisher exact test and Kruskal-Wallis test, respectively. Data: A total of 2588 antiretroviral-naive individuals-initiated ART at 13 clinical sites in 6 African countries.	Early failure detection limited the accumulation of resistance.	Population-based sequencing could not detect minor resistant viral strains, thus potentially underestimating resistance.
Haile et al. [33]	To investigate predictors of treatment failure.	Method, tool: Retrospective cohort study was done in 4 hospitals. Kaplan-Meier curve was used to describe the survival time of ART patients without treatment failure. Bivariate and multivariable Cox proportional hazards regression models were used to identify associated factors of treatment failure. Data(base): 4809 adult ART clients from 4 hospitals in the Bale Zone.	• Male ART clients were more likely to experience treatment failure compared to females. • Lower CD4 count (<100 m3/dl) at initiation of ART was significantly associated with higher odds of treatment failure.	Study was conducted based on secondary data analysis, hence ignored other essential variables.
Steinbrink et al. [39]	To identify variables associated with persistent viremia within academic practice.	Method, tool: HIV-infected patients with viral load of >200 copies/mL were studied. Multivariable data were collected. Conditional logistic regression was used for unadjusted/adjusted analysis. Final multivariable model was built using backwards elimination using likelihood ratio test. variable interactions were tested. Data(base): 66 viremic cases and 66 matched controls.	Hospitalization, underinsurance, and conflicting personal beliefs about disease were more common. Interaction was observed between psychiatric illness and the number of clinic visits.	Insufficient data led to poor generalization of findings.
Babo et al. [36]	To determine factors that predict first-line ART.	Method/tool: Case-control study was carried out on adult clients who failed	Lack of education, unemployment and socio-economic status of patients are	

(*continued on next page*)

Table 1 (*continued*)

Reference	Objective	Method, tool, data(base)	Result/Finding	Drawback/Limitation
		first line regimen and on active follow up, against controls of adult clients on non-failed first-line regimen. Binary logistic regression model was used to identify predictors of ART failure. Data(base): 59 cases of adult clients with failed first-line regimen and on active follow-up and 245 controls.	strong indicators to HIV treatment failures.	Treatment failure was assessed through clinical and/or immunological criteria.
Singh [30]	To improve the effectiveness of Agence Nationale de Recherches sur le SIDA (ANRS) HIV drug resistance using machine learning.	Method, tool: Association matrix was generated from feature selection techniques (ReliefF, MODTree filtering, FCBF filtering and CFS filtering), before incorporating same into ANRS algorithm rules. Tanagra–a software tool was then used to perform machine learning. Data(base): Sandford HIV drug resistance database with approximately 23000 protease and reverse transcriptase gene sequences.	• Feature selection revealed more predictors to HIV drug resistance for Protease and reverse transcriptase ART drugs. • Machine learning with ANRS showed improved HIV drug resistance prediction.	Small datasets limited performance
Dey et al. [46]	To identify and summarize drug compounds that have significant associations with ADRs.	Method, tool: Convolutional deep learning framework was used, to integrate a two-stage ADR prediction, feature creation and predictive model design. Data(base): SIDER database–1430 drugs and 6123 side-effects in 166123 unique drug-ADR associations.	• Proposed neural fingerprints model outperformed other state-of-the-art models in predicting ADRs. • Analysis on drug structures revealed important molecular substructures association with specific ADRs.	Promising solution for identifying risk components within molecular structures is required.
Sisay et al. [43]	To assess the incidence and risk factors of treatment failure among HIV/AIDS-infected children on ART.	Method, tool: Weibull regression model was employed to identify risk factors of treatment failure. Adjusted HRs (AHRs) with 95% CIs was used to declare statistical significance. Data(base): Children's medical charts and ART registration logbook. 824 children under the age of 15 who had started ART.	Incidence of treatment failure remains a significant public health concern.	• Data mainly depended on clinical and immunological criteria. • Virological failure was not used to detect treatment failure, hence, underestimating treatment failure.
Bisaso et al. [45]	To comparatively study logistic regression-based machine learning techniques in the prediction of early virologic suppression of antiretroviral therapy.	Method, tool: Cost functions optimization for multitask temporal logistic regression (MTLR) and simple logistic regression (SLR) was achieved using BFGS (Broyden-Fletcher-Goldfarb-Shanno). Nelder-Mead was used for patient specific survival prediction (PSSP). Data(base): Infectious disease institute (IDI) cohort data consisting of 559 HIV patients enrolled between April 2004 and April 2005.	• MTLR model showed an all-inclusive good calibration over PSSP and SLR models when tested with the IDI dataset. • MTLR and PSSP models gave ample accuracy and discrimination when tested with external EFV cohort data.	Integration of pharmacokinetics and pharmacodynamics would improve the modeling of drug therapy efficiency or resistance.
Ahmed et al. [42]	To identify virologic treatment failure predictors.	Method, tool: Binary logistic regression analysis was performed to determine the relationship between independent variable and outcome variable. Data(base): 9013 adult patients on first-line ART with 2 consecutive viral loads documented between August 2016 and February 28, 2018.	Low first line CD4 T-cell count, body mass index, and poor adherence to ART treatment, predicts virologic failure	The data collection method employed interviews, which restricted the number of variables studied.
Negash et al. [37]	To assess the effect of tuberculosis and other determinant factors of immunological response among HIV patients on HAART.	Method, tool: A retrospective follow up study was conducted. Interviewer-based questionnaire was deployed for data collection. Patient charts were used to extract clinical data and follow-up results of the $CD4^+$ T cell. Simple descriptive statistics and bivariate logistic regression were performed. Data(base): 393 participants on ART were enrolled.	High rate of $CD4^+$ T cells reconstitution failure among study participants was observed. Poor treatment adherence and tuberculosis infection were significantly associated with immunological failure.	The prevalence of failure could be underestimated because the study uses data collected retrospectively from patient charts.
Bezabih et al. [38]	To investigate risk factors for first-line antiretroviral failure using the virologic (plasma viral load) criteria	Method, tool: A case-control study was conducted on adult patients on ART for at least 6 months. Cases were selected from patients who were switched to second-line ART after first-line failure (viral load $\geq$1000 copies/mL). Controls were randomly selected from patients on first-line ART with viral load $<$50 copies/mL. Multivariate logistic regression was performed to identify risk	Findings underscored the importance of avoiding ART discontinuation. The risk of ART failure was high and comparable with duration of ART discontinuation.	Most independent variables were self-reported by the patients.

(*continued on next page*)

Table 1 (*continued*)

Reference	Objective	Method, tool, data(base)	Result/Finding	Drawback/Limitation
		factors for treatment failure. Data(base): 3238 patients with HIV/AIDS who had an ART follow-up in Asella Hospital, Ethiopia.		
Ekpenyong et al. [48]	To efficiently predict HIV patient response to ART.	Method, tool: Hybridized (two-stage) classification of patient response to ART, combining IT2FL and deep neural network (DNN) with multidimensional scaling (MDS), was performed. Data(base): 5780 individual treatment episodes excavated from the Stanford HIV drug resistance database. 3168 individual treatment episodes collected from 13 healthcare facilities in Akwa Ibom State (to form a mixed database).	DNN classification results showed best performance for both databases, with improved pattern predictions for experiment with MDS.	• High computational cost of type reduction process • Did not compare the performance of predicting cases of ADR/failed treatment in the mixed database.
Kiweewa et al. [47]	To identify factors contributing to viremia and virologic failure.	Method, tool: Generalized linear models were used to estimate relative risks with their 95% confidence intervals. Data(base): 2678 HIV-infected participants. 2577 (96.2%) had viral load data for the most recent visit.	Patients on second line, low CD4 count, and who skipped ART, including those with history of fever in the past week remain important predictors of virologic failure.	• Their analysis could not derive virologic failure • Adherence data was based on self-reporting
Pacheko et al. [44]	To analyze characteristics and factors associated with late initiation of ART.	Method, tool: Uni- and multi-variate analysis was performed using SPSS Data(base): 1371 HIV-infected treatment-naive participants-initiated ART from 2009 to 2012 in the public health system.	Late onset of ART was associated with higher mortality.	Half of their sample points, initiated ART at an advanced stage of HIV with high viral load.
Ekong et al. [29]	To understand socio-demographic, socio-economic and other related factors that could predict ADR.	Method, tool: Logistic regression model was used to estimate odds ratios (ORs) and 95% confidence intervals (CIs) for factors associated with ADR. Data(base): APIN electronic medical record system with 458 participants.	Understanding ADR predictors may guide programmes in developing interventions and for identifying patients at risk of developing ADR.	• Extracted data appear incomplete, inaccurate, inconclusive. • First factor contributing to ADR in a patient with several identified predictors unknown.
Feleke et al. [41]	To assess the magnitude and associated factors of ART failure.	Method, tool: Institution-based cross-sectional study was conducted using chart review data. Multivariable logistic regression. Data(base): HIV-positive adult patients between 45 and 54 years at stages 3 and 4 ART, poor drug adherence, and on antiretroviral therapy follow-up.	High rate of ART failure observed, with 21% prevalence of first-line antiretroviral treatment failure.	Prevalence of failure could be underestimated because the study uses data collected retrospectively from patient charts.
Ekpenyong et al. [49]	To model drugs interaction in treatment-enabled HIV patients and optimize their response to ART	Method, tool: IT2FL, weighted least-squares cost function and neural network (NN) were used to model and optimize patient response to ART. Data(base): 5780 individual treatment episodes excavated from the Stanford HIV drug resistance database; 3168 individual treatment episodes collected from 13 healthcare facilities in Akwa Ibom State (mixed database).	• Prognostic markers correlate suggests strong association between first line and follow-up CD4 counts • Moderately weak association was observed for first line and follow-up viral loads. • Correlation of physiological features gave very strong association between first line and follow-up body mass index in mixed database. • Improved RMSE and classification accuracy for both databases were noticed, when compared with existing works.	Study did not compare the performance of predicting cases with ADR/failed treatment in the mixed database.

capabilities. Sometimes the test datasets are not large enough to reliably analyze the performance difference between experimented features, hence, yielding non-generalizable models [55]. Although transfer learning may achieve better classification results with small datasets via knowledge learned from related tasks with larger datasets [56], performing experiments in similar settings is useful for the purpose of generalizing results to other methods/domains. Leveraging hybrid transfer learning methods on existing databases [57,59] have demonstrated superiority over baselines, but such models cannot handle or evaluate other/new prediction tasks.

3.3. Research justification and impact

From the foregoing literature, we observe that most outcomes of statistical and computational approaches rest on simulation studies and used datasets, accompanied by huge statistical analysis/reports. These datasets are strictly modeled without considering adaptation to new datasets. To ensure adaptability of predictive models to experimental datasets, research progress needs to be advanced towards the production of control datasets for domain-specific regions/environments – an active area that has not been adequately explored. Imbalanced classification is typical of any disease problem, as not all patients produce same response to the disease or respond equally to treatment. Recovery rates also vary from patient to patient, as some patients recover quickly than others. Hence, a 'one-size-fits-all' therapy is prone to possibilities of failed treatments (persistent symptoms). Treatment failure is associated with factors such as poor drug adherence, adverse side effect/reaction, co-infection, lack of follow-up, and drug-drug interaction. This research is therefore motivated by the need for clinicians/physicians to detect early patients' response to treatment and administer new drugs if the current regimen fails. The research is urgent due to the attendant risks of complications and even death. The use of nondomain-specific datasets to

Table 2
Previous works on application of transfer learning approaches to disease prediction/classification.

Reference	Objective	Methodology, tool, data(base)	Finding	Drawback/Limitation
Huynh et al. [50]	To extract tumor information from medical images via CNNs originally pretrained for nonmedical tasks.	Method, tool: This study compared support vector machine classifiers based on the CNN-extracted image features and prior computer-extracted tumor features, to distinguish between benign and malignant breast lesions. Data(base): 219 breast lesions obtained from 607 full-field digital mammographic images, under an Institutional Review Board-approved protocol from the University of Chicago Medical Center.	• Classifiers based on CNN-extracted features (with transfer learning) performed comparably to those using analytically extracted features. • Performance of ensemble classifiers based on both types was significantly better than that of either classifier type alone (AUC = 0.86 vs. 0.81)	Methodological choices were not necessarily optimal in terms of predictive performance.
Page et al. [51]	To apply advanced machine learning and hardware techniques to seizures detection.	Method, tool: Max-pooling CNN architecture consisting of 1–3 pairs of convolutional and max-pooling layers connected to 1–3 fully connected layers, with the last layer was deployed to implement a softmax classifier. Data(base): Scalp-based EEG recordings obtained from pediatric patients (collected from 23 patients) with intractable seizures.	Proposed system was able to detect all 184 seizure onsets from 24 cases with average latency of 1.47 s and 3.2 false alarms/day.	Personalized model performed much worse when applied to long-term EEG data.
Christodoulidis et al. [54]	To improve the accuracy and stability of a CNN on the task of lung tissue pattern classification.	Method, tool: 6 publicly available texture databases were used to pretrain data for a CNN architecture. Data(base): Amsterdam library of Textures (ALOT), Describable Textures Dataset (DTD), Flickr Material Database (FMD), Kylberg Texture Database (KTB), KTH-TIPS-2b and Ponce Research Group's Texture database (UIUC).	• Proposed approach resulted in an absolute increase of about 2% in the performance. • Results demonstrate the potential of transfer learning in the field of medical image analysis.	Reported increase in accuracy came at the expense of increased training time.
Samala et al. [55]	To develop a computer-aided detection (CAD) system for masses in digital breast tomosynthesis (DBT) volume.	Method, tool: Mass of interest on breast images was marked by an experienced breast radiologist as reference standard, and the dataset partitioned into training set and test set. Data(base): 2282 digitized film and digital mammograms. 324 digital breast tomosynthesis (DBT) volumes–from Department of Radiology at the University of Michigan Health System (UM), and University of South Florida (USF) digitized mammogram database	• DCNN-based CAD system outperformed the feature-based CAD system for breast-based performance with statistically significant results. • AUC improved from 0.81 (before transfer learning) to 0.91 (after transfer learning).	DBT test set was not large enough to reliably analyze the performance difference between malignant and benign masses.
Li et al. [56]	To implement diabetic retinopathy fundus image classification using CNN-based transfer learning.	Method, tool: Fine-tune all network layers of the pre-trained CNN models. Fine-tune a pre-trained CNN model in a layer-wise manner. Use the pre-trained CNN models to extract features from fundus images. Training support vector machines using these features. Data(base): 1014 and 1200 fundus images from two publicly available (DR1 and MESSIDOR) datasets	CNN-based transfer learning achieved better classification results with small datasets.	Difficult to compare results with other methods because experiments were not carried out on same settings.
Turki et al. [57]	To improve drug sensitivity prediction using transfer learning	Method, tool: Two transfer learning approaches that combines auxiliary data from the related task with target training data was experimented on clinical trial data leveraged on 3 auxiliary datasets. Database: 9114 genes, and drug IC_{50} values.	Experimental results demonstrate superiority of proposed approaches over baselines when auxiliary data were incorporated.	Their model could not handle new/other drug sensitivity prediction tasks
Shtar et al. [59]	To detect drug–drug interactions using artificial neural networks and classic graph similarity measures.	Method, tool: ANNs and factor propagation over graph nodes were used to perform a retrospect analysis by pre-training the models on previously released DrugBank database. Database: DrugBank database with 1141 drugs and 45,296 drugs interaction.	The proposed model out-performed various ensembles created using 29 different predictors based on several datasets; proving that known drugs interaction is effective for identifying potential drugs interaction.	The proposed system cannot evaluate new drugs with no known interactions.
Cha et al. [52]	To develop a CAD system for bladder cancer treatment response assessment in computer tomography (CT) using Radiomics with Deep-Learning.	Method, tool: Three unique radiomics-based predictive models were assessed. Models used different design principles ranging from pattern recognition using deep-learning convolution neural network (DL-CNN), to a more deterministic radiomics feature-based	DL-CNN performed better when transfer learning with bladder images was used instead of natural scene images.	Transfer learning did not improve the treatment response estimation

(*continued on next page*)

Table 2 (*continued*)

Reference	Objective	Methodology, tool, data(base)	Finding	Drawback/Limitation
		approach. Data(base): Natural scene images or regions of interest (ROIs) inside and outside the bladder.		
Deniz et al. [53]	To compare the classification of transfer learning and deep feature extraction on breast cancer detection.	Method, tool: Transfer learning and deep feature extraction methods are deployed, to adapt pre-trained CNN model. Data(base): BreaKHis dataset covering a total of 9109 microscopic images (2480 benign and 5429 malignant samples)	Transfer learning produced better result than deep feature extraction and SVM classification.	The critical nature of breast cancer requires higher detection accuracy.

predict other datasets, could result in over-sampling and requires datasets from the under-represented class. To improve the under-represented class, collection of more samples is necessary to provide a balanced perspective on the classes. This paper initiates the development of a control dataset for SSA region. The dataset is essential because it will not only provide baseline for comparing variability within/between domains, but also engender understanding of the status of an infected group within the region. To demonstrate usability of the datasets, we rely on existing clinical trial data (the Stanford resistance database – a curated public database for representing, storing, and analyzing HIV drug resistance data), as the control dataset, and apply normal probability plots to measure the deviation of the control and mixed datasets from normality. The Kaplan-Meier curve has been found to be less accurate in predictions, hence, this paper justifies the use of an integrated classification framework with intelligent models for clustering, learning, and predicting with precision, HIV patients with failed treatment. A major innovation introduced in this paper is the use of existing multi-line treatment to assist useful decision support on new drug prescription for patients on first-line treatment.

This research will surely impact the field of medical informatics as it offers opportunities for achieving individualized therapy, and expert recommendation system. By isolating patients with improved therapy from a controlled dataset, knowledge of optimal drugs combination with very low or no interaction can be discerned and used for learning a mixed dataset of non-isolated and isolated regimens of high and very high drugs interaction.

4. Materials and method

4.1. Description of the domain datasets

Data for this study came from two disparate sources: a publicly available online domain data–the Stanford HIV drug resistance dataset: https://hivdb.stanford.edu (control dataset); and locally sourced data from healthcare facilities in Akwa Ibom State, Nigeria (mixed dataset). While the control dataset holds curated clinical trials of patients who experienced ADRs after first-line treatment and were subjected to multi-line follow-up treatments for up to a period of 48 weeks, the mixed dataset holds only first-line treatment episodes (initial 6 months). Each patient record on the control dataset stores treatment change episodes (TCEs) or treatment course of the patient, as separate XML file. The mixed dataset was extracted from existing patients' records/files under the supervision of a medical superintendent after due ethical approval was obtained. Additionally, each patient's data was validated for consistency before recording while questionable, inconsistent, or not properly documented records were dropped. Both datasets contain the following prognostic features: before and follow-up $CD4^+$ count (BCD4 and FCD4), before and after viral load (BRNA and FRNA); and drug type combination (DType). The $CD4^+$ count is a test that measures the number of CD4 cells in the blood. Also called T cells, they are white blood cells that fight infection and play vital role in one's immune system. HIV viral load is the number of viral particles found in each millimeter. The more HIV-1 particles in the blood, the faster the $CD4^+$ T-cells are destroyed. Twenty-four (24) drug types were administered during the Stanford clinical trials. Table 3 lists the various drugs administered indicating the drug code (DrugCode) and drug name (DrugName) in the control dataset. For the mixed dataset, only 4 drug types are administered namely, Zidovudine (AZT), Lamivudine (3 TC), Tenofovir (TDF), and Efavirenz (EFV), corresponding to serial numbers 9, 4, 11 and 12 on Table 3, in an ongoing community antiretroviral testing programme in collaboration with the Family Health International (FHI) 360 Project in Akwa Ibom State of Nigeria. The mixed dataset covers patients who registered for treatment at various health facilities from 2015 to 2018. A total of 1521 and 1506 unique patients' records filtered from 5780 to 3168 individual treatment change episodes (TCEs) were processed using the Python programming language for the control and mixed datasets, respectively.

4.2. Exploratory analysis of domain datasets

Many natural or real-life phenomena can be approximated by a bell-shaped frequency distribution known as the normal or Gaussian distribution. Before exploring the domain datasets, initial assumptions about the data are necessary to condition the model for fluctuations not seen in the training data. To understand the domain datasets, we deployed the normal probability plots, also called quantile-quantile or Q-Q plots as a diagnostic tool to identify substantive departure from normality (outliers, skewness, kurtosis, etc.). This visualization helped us find patterns and discover potential predictors for querying if the necessary assumptions are violated or not, and for determining the choice of an appropriate or reasonable model.

From the normal probability plots visualizing CD4 count in the control dataset (Fig. 2a), it is evident that there is strong likelihood of the existence, on the average, of a slight discrepancy in the BCD4 and FCD4 counts of patients who were examined. More patients seem to have, on the average, a slightly higher BCD4 counts than the FCD4 counts for count values from about 40 to 840 units, indicating that more patients in this dataset may have had compromised immune system ($CD4 \leq 500$ cells/mm^3) prior to the clinical trials, and are gradually improving, immunologically. However, there seem to be no significant evidence of any discrepancy in these prognostic parameters for count

Table 3
Drugs administered in the Stanford experiment (control dataset).

DrugNo	DrugCode	DrugName	DrugNo	DrugCode	DrugName
1	RTV	Ritonavir	13	DDI	Didanosine
2	IDV	Indinavir	14	LPV	Lopinavir
3	D4T	Stavudine	15	APV	Amprenavir
4	3 TC	Lamivudine	16	NVP	Nivarapine
5	SQV	Squatonavir	17	DRV	Darunavir
6	T20	Nfoviritide	18	FTC	Emtricitabine
7	FPV	Fosamprenavir	19	ATV	Atazanavir
8	NFV	Nelfinavir	20	TPV	Tipranavir
9	AZT	Zidovudine	21	RAL	Raltenovir
10	ABC	Abacavir	22	ETR	Etravirine
11	TDF	Tenofovir	23	MVC	Maraviroc
12	EFV	Efavirenz	24	DLV	Delavirdine

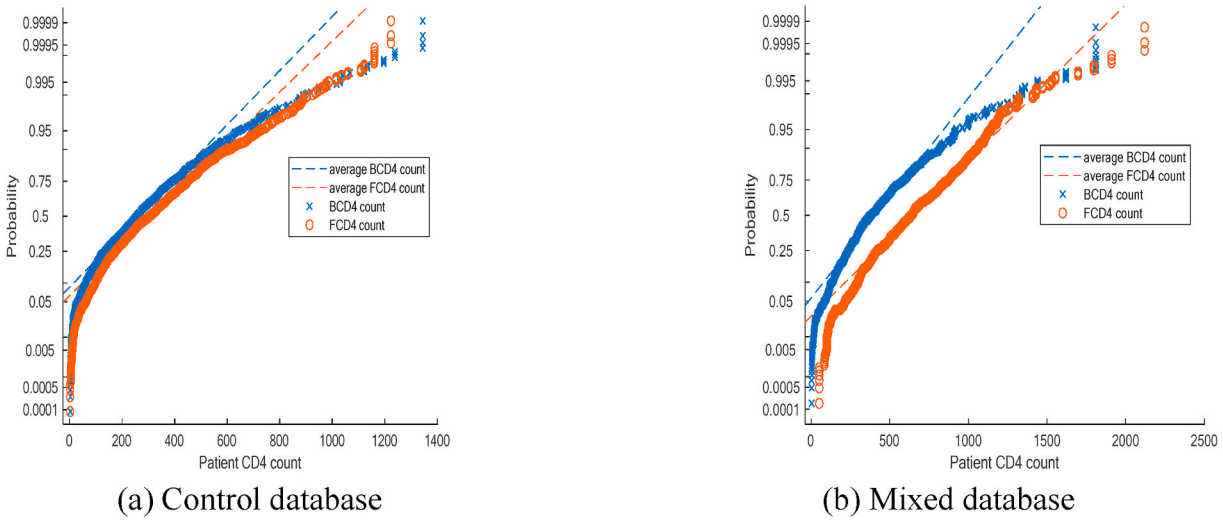


Fig. 2. Normal probability plots visualizing CD4 counts in the control and mixed datasets.

values greater than 840 up to about 1160. But, towards the right tail, for count values higher than about 1160 up to about 1280, the FCD4 counts of the patients tend to be higher, on the average, than their corresponding BCD4 counts; with a break in the BCD4 counts from values above 1280 up to a value just before 1400, showing a possibility of the prognostic parameters not falling in that range. This implies that few patients showed no immunological improvements over time. Also, there appears to be no evidence of both parameters higher than about 1400 units, indicating the candidate upper universe of discourse (UoD) for these parameters. No discrepancy in the prognostic markers exists for count values below 40 units.

Normal probability plots visualizing CD4 counts in the mixed dataset (Fig. 2b) show that there is a wide discrepancy, between the prognostic parameters. For count values in the range of 125–1400, the BCD4 counts of the patients, is observed to be significantly higher than their corresponding FCD4 counts, indicating evidence of significant immunological improvements. Towards the right tail, for count values higher than about 1500, there are some breaks, showing that there may have been some patients without some prognostic parameter values in that range. For count values of about 1900, there seem to be more patients with a higher FCD4 counts than BCD4 counts; and there may not be patients with BCD4 counts above this value. Furthermore, there appears to be few patients with FCD4 counts slightly lower than their BCD4 counts at the lower tail, indicating poor immunological recovery.

Suppressed viral load indicates good response to treatment/therapy or efficacy of the administered drugs. Viral load normal probability plots of the control dataset (Fig. 3a) show evidence of a much wider discrepancy between BRNA and FRNA copies. Particularly, for lower count values of FRNA below about 5.8, progressive improvements are noticed in the TCEs. Furthermore, no patient with evidence of BRNA values below 3.0 is found. The BRNA and FRNA copies patients are however found to be nearly identical for some count values towards the right tail; and at count values of approximately 6.0, there is a break on the BRNA and FRNA plots, showing evidence of patients with specified RNA values not beyond 6.0. From the viral load normal probability plots of the mixed dataset (Fig. 3b), it is obvious that there is a wide discrepancy between FRNA copies exceeding their corresponding BRNA copies, indicating evidence of suppressive therapy. The discrepancy appears to be much wider for RNA copies from about 1.8 to 6.0, hence, narrowing down towards the right tail of the plot. There also appears to be some patients with BRNA values at the right tail with no corresponding FRNA values. Similarly, towards the left tail, a couple of FRNA values with no corresponding BRNA values is observed. The much deviation of the FRNA plot from the central line shows the presence of

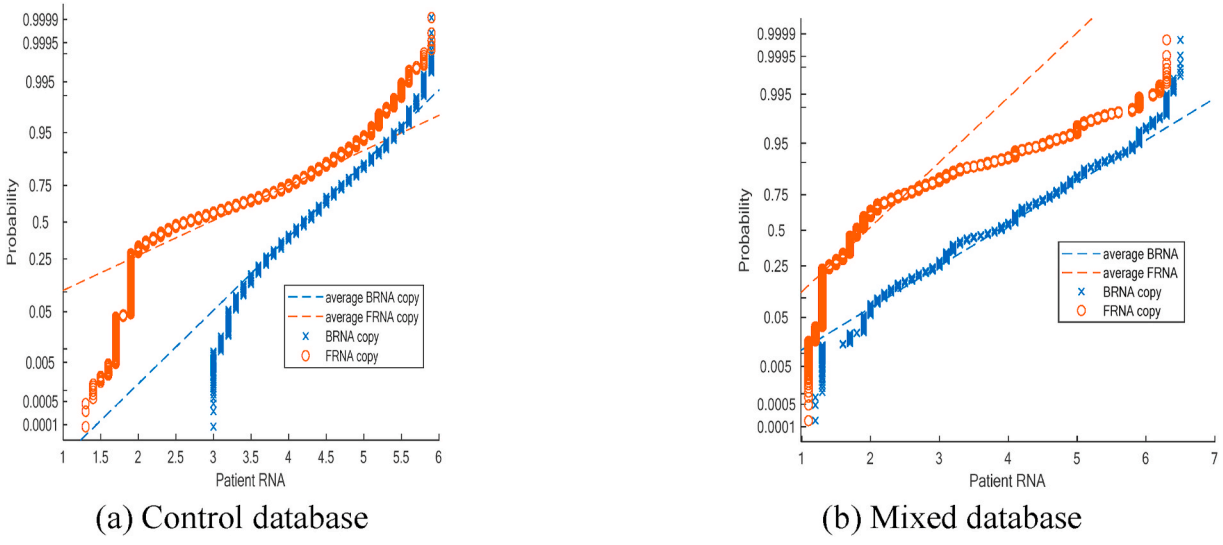


Fig. 3. Normal probability plots visualizing viral loads in the control and mixed datasets.

many outliers, as such, patients with FRNA values of higher than about 3.5 tend to deviate from the mean (below the average). Consequently, patients with BRNA values of above 1.8 tend to be close to the average, as the plots do not deviate much from the center.

4.3. Classification target labeling of patient response

For the control and mixed datasets, inputs to our classification algorithms include (i) prognostic markers processed from the Stanford resistance dataset and medical records of patients from a mixed dataset, respectively; and (ii) patients' response to drug regimens derived from output membership grades of an interval type-2 fuzzy logic (IT2FL) system [48,49]. The classification target is segmented into five output classes, C1–C5 (C1=NI: no interaction, C2=VLI: very low interaction, C3 = LI: low interaction; C4=HI: high interaction; C5=VHI: very high interaction). On Table 4 and Table 5, a list of the first 15 patients from both datasets is provided to aid readers understanding of the class labeling. Table 6 shows an analysis of the extent of immunological and virologic changes obtained from both datasets.

5. Proposed system model

5.1. Neural network algorithm

Suppose that a network has L layers, with layers 1 and L being the input and output layers, respectively. Suppose that layer $l : l = 1, 2, 3, \ldots, L$ contains n_l neurons, where n_1 is the dimension of the input data. Generally, the network maps from $\mathbb{R}^{n_1}$ to $\mathbb{R}^{n_L}$; and we denote $W^{[l]} \in \mathbb{R}^{n_l \times n_{l-1}}$ as the weights' matrix at layer l. Precisely, $w_{jk}^{[l]}$ is the weight that neuron j at layer l applies to the output from neuron k at layer $l-1$. Similarly, $b^{[l]} \in \mathbb{R}^{n_l}$ is the vector of biases for layer l, so neuron j at layer l uses the bias $b_j^{[l]}$. In this paper, our experiment architectures are the traditional 2 layers NN, and 5 layers deep neural network (DNN) with neurons distribution obeying an odd number dropout fashion as follows: NN [32], NN [53] and DNN [9 7 5 3 1]. These architecture configurations and their connecting nodes with a sample weight highlighted, are presented in Fig. 4. The defined parameters are given as follows: For NN [32], $L = 2$; $n_1 = 3, n_2 = 2$.For NN [53], $L = 2$; $n_1 = 5, n_2 = 3$. For DNN [9 7 5 3 1], $L = 5$; $n_1 = 9,\ n_2 = 7,\ n_3 = 5, n_4 = 3,\ n_5 = 1$. Hence, the weights and corresponding biases matrices at various layers of these architectures are given as follows:

(a) $W^{[2]} \in \mathbb{R}^{2\times 3}$, $b^{[2]} \in \mathbb{R}^2$;
(b) $W^{[2]} \in \mathbb{R}^{2\times 3}$, $b^{[2]} \in \mathbb{R}^2$
(c) $W^{[2]} \in \mathbb{R}^{7\times 9}$, $W^{[3]} \in \mathbb{R}^{9\times 7}$, $W^{[4]} \in \mathbb{R}^{7\times 5}$, $W^{[5]} \in \mathbb{R}^{1\times 3}$, $b^{[2]} \in \mathbb{R}^7$, $b^{[3]} \in \mathbb{R}^9$, $b^{[4]} \in \mathbb{R}^7$, $b^{[5]} \in \mathbb{R}^1$.

Given an input $x \in \mathbb{R}^{n_1}$, the action of the network can be neatly summarized by letting $o^{[l]}$ denote the output, or activation, from neuron j at layer l. So, we have:

$$o^{[1]} = x \in \mathbb{R}^{n_1}, \tag{1}$$

$$o^{[l]} = \sigma\left(W^{[l]} o^{[l-1]} + b^{[l]}\right) \in \mathbb{R}^{n_l}, \quad \text{for } l = 1, 2, 3, \ldots, L. \tag{2}$$

It becomes clear that (1) and (2) represent an algorithm for feeding input data through the network in order to produce the output $o^{[l]} \in \mathbb{R}^{n_L}$; and is required in any approach for training a network. For N data samples or training points in $\{x^{\{i\}}\}_{i=1}^{N} \in \mathbb{R}^{n_L}$, the target outputs are $\{y(x^{\{i\}})\}_{i=1}^{N} \in \mathbb{R}^{n_L}$. A total of 8948 TCEs (i.e., combined TCEs of the control dataset: 5780 samples plus TCEs of the mixed dataset: 3168 samples) were used in the classification tasks, save the DNN2 [9 7 5 3 1] classifier, which used only TCEs of those at risk of drug interaction (i.e., HI and VHI) in the mixed dataset. But while the traditional machine learning approach randomly partitions the entire datasets according to a 75%-train, 15%-validation, and 15%-test ratio with a safe finetune of the network to maintain acceptable validation errors; the transfer learning approach strictly preserves the mixed dataset for test purposes only, while splitting the control dataset into 85%-train and 15%-validation ratio.

The activation function adopted in this paper is the rectified linear unit (ReLU) function, which has become the most widely used activation function for neural networks especially in convolutional neural networks (CNNs). It is said to be a universal function approximation, and is defined as:

$$\phi(X) = \begin{cases} x; & x > 0 \\ 0; & otherwise. \end{cases} \tag{3}$$

ReLU provides a solution to the vanishing gradient problem (a condition that prevents a network to learn further or become drastically slow due to extremely low values) with identity derivative.

5.2. Transfer learning algorithm

Our transfer learning algorithm is transductive and is described

Table 4
Input linguistic variables and target classes for control dataset.

Input linguistic variable							Target class				
PID	BCD4	FCD4	BRNA	FRNA	DType	PR	C1	C2	C3	C4	C5
1	380	337	5.1	5.1	3 TC + D4T + IDV + RTV	33.25	0	1	0	0	0
2	39	65	5.3	5.5	D4T + RTV + SQV	30.00	1	0	0	0	0
3	86	58	5.1	5.2	3 TC + FPV + RTV + T20	30.00	1	0	0	0	0
4	59	49	4.7	4.9	3 TC + AZT + NFV	30.00	1	0	0	0	0
5	35	33	4.6	4.5	3 TC + ABC + AZT + RTV + SQV + T20 + TDF	30.00	1	0	0	0	0
6	86	113	5.3	2.7	ABC + AZT + EFV + RTV + SQV	44.82	0	1	0	0	0
7	144	104	4.7	2.3	3 TC + ABC + D4T	50.00	0	1	0	0	0
8	114	98	3.7	4.7	DDI + RTV + SQV	30.00	0	1	0	0	0
9	388	331	4.5	3.5	3 TC + D4T + NFV	33.81	0	0	1	0	0
10	790	444	4	2.9	3 TC + D4T + NFV	59.61	0	0	0	0	1
11	78	73	5	4.9	3 TC + ABC + EFV + NFV	30.00	1	0	0	0	0
12	114	110	4.3	4.6	3 TC + DDI + LPV + T20 + TDF	30.00	1	0	0	0	0
13	276	239	3.4	3	3 TC + D4T + RTV + SQV	43.13	0	0	1	0	0
14	56	125	5.4	5.1	3 TC + ABC + NFV	30.00	1	0	0	0	0
15	68	18	5.9	5.6	ABC + EFV + RTV + SQV	30.00	1	0	0	0	0
16	11	21	5.9	5	D4T + DDI + EFV	30.00	1	0	0	0	0
17	151	168	4	4.6	D4T + IDV + RTV	30.00	0	1	0	0	0
18	115	148	5.4	2.9	3 TC + AZT + RTV + SQV	41.07	0	1	0	0	0
19	209	382	5.1	4.5	3 TC + RTV + SQV	33.38	0	1	0	0	0
20	190	211	4.3	3.1	3 TC + APV + D4T + DDI + RTV	37.97	0	1	0	0	0

Table 5
Input linguistic variables and target classes for mixed dataset.

Input linguistic variable										Target class				
PID	CD4_B	CD4_F	VL_B	VL_F	Dr_T	PR	Sex	BWt	FWt	C1	C2	C3	C4	C5
1	148	106	3	1.3	TDF+3 TC + EFV	53.56	F	42	43	0	0	1	0	0
2	145	378	2.5	1.3	AZT+3 TC + NVP	55.33	F	57	60	0	0	0	1	0
3	78	131	4.1	1.7	AZT+3 TC + NVP	50	M	70	75	0	1	0	0	0
4	295	574	4.4	1.9	AZT+3 TC + NVP	50	M	64	66	0	0	1	0	0
5	397	792	1.9	1.3	AZT+3 TC + NVP	76	F	52	55	0	0	0	0	1
6	155	280	4.2	1.7	TDF+3 TC + EFV	50	F	59	56	0	1	0	0	0
7	303	679	4.2	1.3	AZT+3 TC + NVP	53.68	F	62	60	0	0	1	0	0
8	370	615	5.1	1.7	TDF+3 TC + EFV	52.62	M	78	68	0	0	0	1	0
9	210	242	5.1	4.1	AZT+3 TC + NVP	30	F	82	82	1	0	0	0	0
10	120	278	2.7	1.7	AZT+3 TC + NVP	50	M	85	80	0	0	1	0	0
11	450	347	3.1	1.3	TDF+3 TC + EFV	57.78	M	48	50	0	0	0	1	0
12	365	302	3	1.5	AZT+3 TC + NVP	52.62	F	50	52	0	0	0	1	0
13	450	500	4.1	2.5	TDF+3 TC + EFV	63.2	M	60	60	0	0	0	1	0
14	198	324	5.1	1.8	TDF+3 TC + EFV	50	F	48	50	0	1	0	0	0
15	254	631	4.1	1.7	TDF+3 TC + EFV	50	M	70	73	0	0	1	0	0
16	265	654	4.2	1.7	AZT+3 TC + EFV	50	M	77	77	0	0	1	0	0
17	432	765	4.1	1.9	TDF+3 TC + EFV	58.29	M	60	60	0	0	0	1	0
18	584	652	4.3	2.1	AZT+3 TC + EFV	71	F	57	59	0	0	0	1	0
19	280	584	3.2	1.3	AZT+3 TC + EFV	55.33	M	60	62	0	0	0	1	0
20	287	110	4.2	4.6	AZT+3 TC + NVP	30	M	57	59	1	0	0	0	0

Table 6
Immunological and virologic changes in experiment dataset.

Metric/measure	Statistic	
	Stanford	Akwa-Ibom
Total number of unique patients with TCEs>2 (total TCEs)	1490 (5780)	1056 (3168)
Total number of drugs administered	24	5
Total number of drug combinations	559	5
Patients with improved immunity (FCD4>BCD4)	1046	894
% of patients with improved immunity	68.91	84.66
Patients with FRNA < BRNA (total TCEs)	1248 (4774)	897 (2691)
Patients with improved RNA ([BRNA-FRNA] ≤ 0.5), (total TCEs)	1085 (4171)	821 (2463)
% RNA decrease (≤0.5)	71.48	77.75
% RNA increase (>0.5)	28.52	22.25
Unique drug combinations that yield improved treatment (i.e., with [BRNA-FRNA] ≤ 0.5)	446	5
Ineffective drugs combination	106	0

within the context of the problem as follows:

Let $F = \{(x_1, y_1), \ldots, (x_m, y_m)\}$ and $T = \{t_1, \ldots, t_r\}$, be the target training and target testing sets, respectively. In this case, the target training set is the control dataset and contains individual TCEs of patients who failed treatment; and the target testing set is the mixed dataset, which contain individual TCEs of patients on first-line treatment (i.e., contains patients with failed and successful/improved treatment). Note that both datasets are clustered by patient's response to treatment–into five target classes C1–C5, using [48]. In the target training set, $x_i \in R_p$ is the ith target training exemplar with p features or prognostic markers plus drug combinations; $y_i \in R$ is the corresponding label of x_i, representing patient's response to treatment. Similarly, $t_i \in R_p$ is the ith target testing exemplar with p features of the mixed datasets. The target training set and target testing set are considered disjoint, where m and r are the numbers of training and testing examples, respectively, in the target task. There exists a set of isolated patients $S = \{(s_1, u_1), \ldots, (s_f, u_f)\}$, with improved treatments sieved from the control dataset; where $s_i \in R_n$ is the ith example (i.e., TCEs of optimal drug combinations) with n prognostic markers and drug combinations (or features); $u_i \in R$ is the corresponding label of s_i; and n, the number of features in the isolated

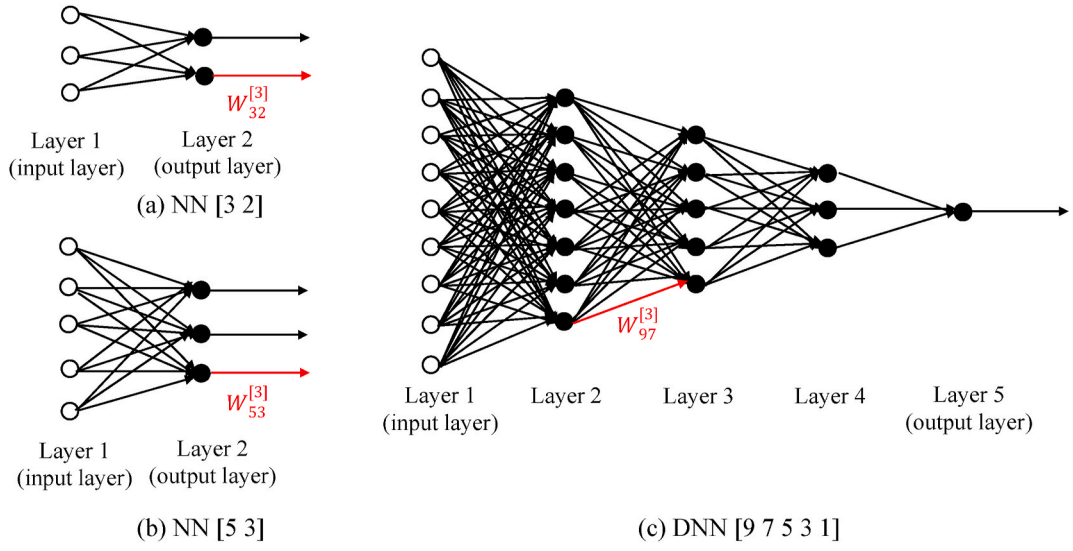


Fig. 4. Experimented NN architectures. The edges corresponding to the weights $W_{32}^{[2]}$, $W_{53}^{[3]}$ and $W_{97}^{[7]}$ are highlighted.

dataset, is different from p, the number of features in the target task. Our goal is to optimally predict alternative/new drug combinations or performance on the target testing set T (of the target task) via learning an accurate model using the isolated set S and the target training set F. The target testing set T is known at training time, such that the learning algorithm is fully aware of which exemplars it will be evaluated on after the training. This is a great asset to the algorithm, as it enables it to shape its decision function to match and exploit properties found in T – a special case of semi-supervised learning in which $S = T$.

A 4-step process workflow implementing the proposed approach is presented in Fig. 5. In the first step, the target training set and isolated datasets are preprocessed to eliminate outliers and inconsistencies in the TCEs. To remove outliers, a data normalization procedure is adopted in this paper. To eliminate inconsistencies, a fuzzy-logic system [48,49] is implemented. The outcome is a normalized and consistent data with non-fuzzy patient response. In the second step, TCEs of the target training set are clustered according to the potential risk of failed treatments. A clustered dataset with optimized TCEs is obtained as a result. In step 3, a learning algorithm is applied to learn the model, and the obtained model used to predict new drug combinations for patients with failed treatment (step 4). To quicken the computation process, there is the need to repurpose the trained model given the similarities in parameters/features of both datasets. Since a pre-trained model already exists, there is no need retraining the model from the scratch, but rather deploying this model on the new dataset with a new task for isolating patients with failed treatment. TCEs with high similarity measures are then matched to locate patients requiring new drug combinations.

6. Results and discussion

6.1. Visualization of classification performance

MATLAB 2020a was used to develop the necessary classification models. Confusion matrices were then generated to visualize the system performance. The overall confusion matrices in Fig. 6 and Fig. 7., show the performance of traditional machine learning and transfer learning approaches, respectively. In these matrices, rows correspond to the predicted or output class, while columns correspond to the true or target class. The diagonal cells indicate correctly classified observations while the off-diagonal cells correspond to misclassified observations. The

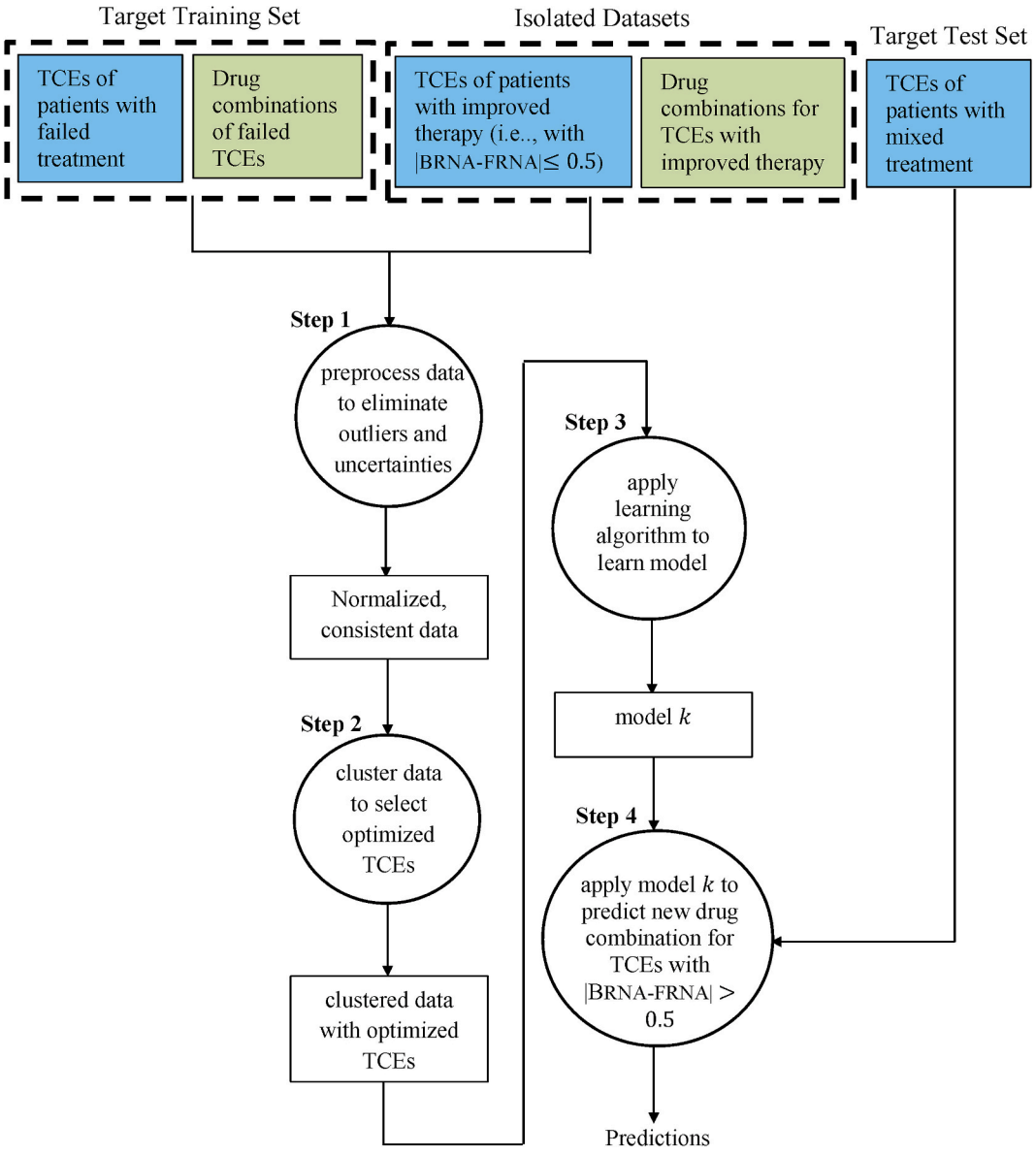
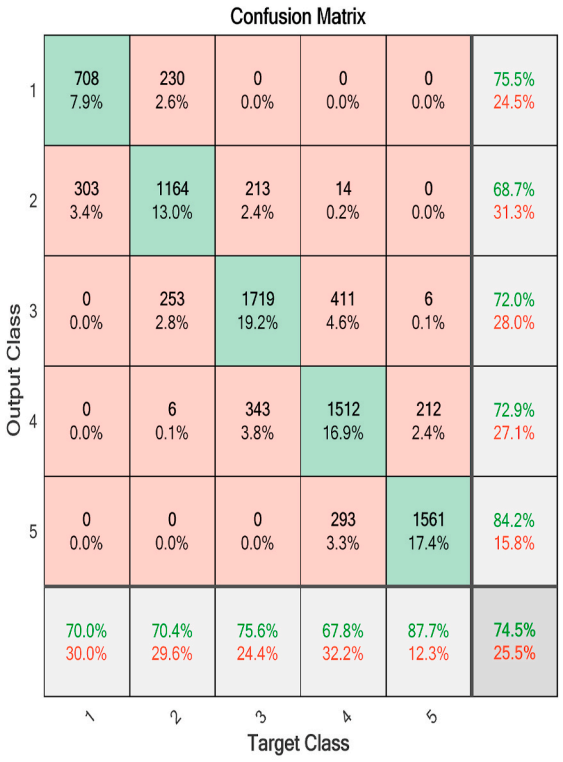
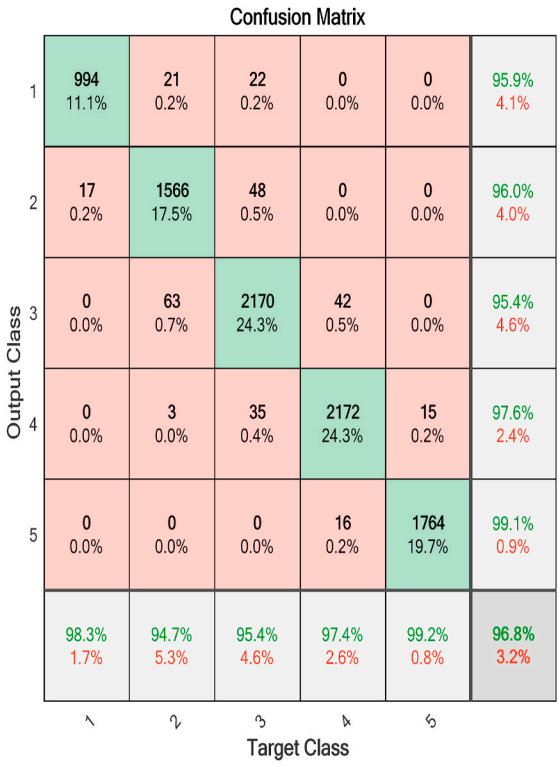


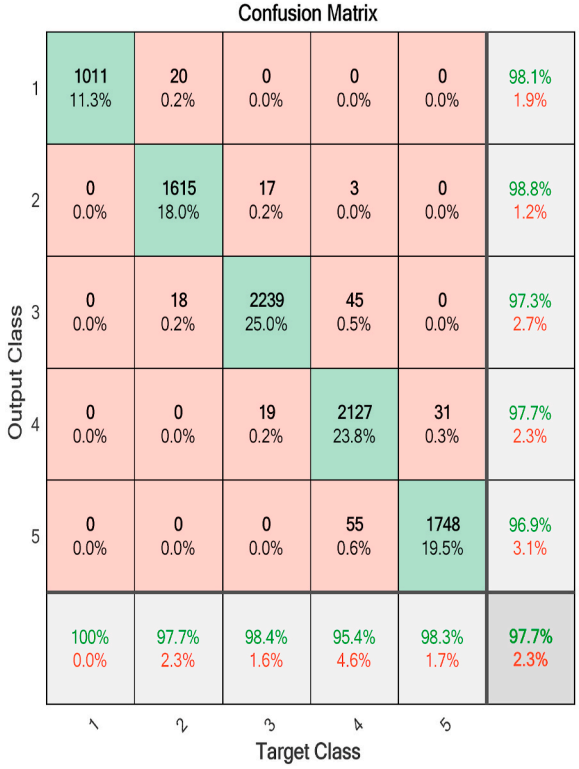
Fig. 5. Process workflow implementing the proposed approach.



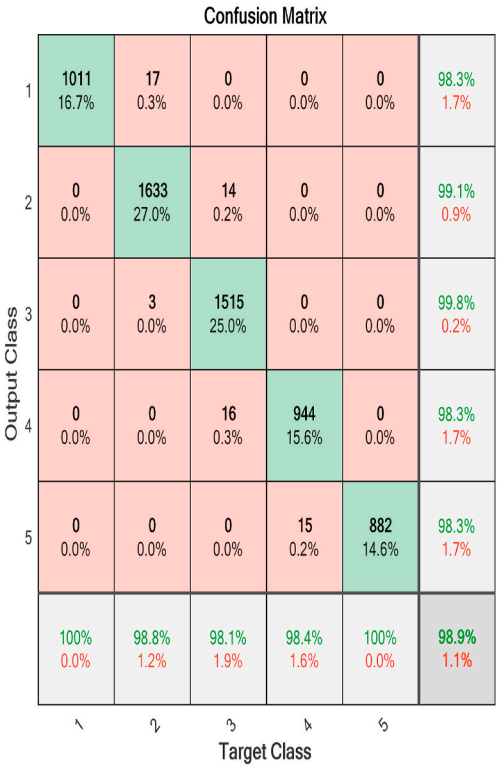
(a) NN [3 2]



(b) NN [5 3]



(c) DNN1 [9 7 5 3 1]



(d) DNN2 [9 7 5 3 1]

Fig. 6. Confusion matrices for traditional machine learning approach.

Fig. 7. Confusion matrices for transfer learning approach.

columns on the far right show the percentage of all samples predicted to belong to each class that are correctly and incorrectly classified, often called 'precision', i.e.: TP/(TP + FP), where TP is the true positive and FP is the false positive. The rows at the bottom show the percentage of all samples belonging to each class that are correctly and incorrectly classified, often called 'recall', i.e.: TP/(TP + FN), where FN is the false negative.

In Fig. 6, classification results of the traditional machine learning

approach across different network architectures are presented. It is observed that the DNN1 [9 7 5 3 1] configuration, which combines TCEs of control and mixed datasets correctly classified cases of patients response to drug regimens as follows: No Interaction {1011 (11.3%)}, Very Low Interaction {1615 (18.0%)}, Low Interaction {2239 (25.0%)}, High Interaction {2127 (23.8%)} and Very High Interaction {1748 (19.5%)}, representing about 97.7% of the overall classification performance; while the DNN2 [9 7 5 3 1] configuration, which used only TCEs of those at risk of drug interaction (i.e., High Interaction and Very High Interaction) as mixed dataset correctly classified patients response to drug regimens as follows: No Interaction {1011 (16.7%)}, Very Low Interaction {1633 (27.0%)}, Low Interaction {1515 (25.0%)}, High Interaction {944 (15.6%)} and Very High Interaction {882 (19.5%)} representing about 98.9% of the overall classification performance.

In Fig. 7, results of classification performance of transfer learning approach indicate that the DNN1 [9 7 5 3 1] configuration correctly classified patients response to drug regimens as follows: No Interaction {1011 (11.3%)}, Very Low Interaction {1611 (18.0%)}, Low Interaction {2213 (24.7%)}, High Interaction {2212 (24.7%)} and Very High Interaction {1779 (19.9%)} representing about 98.6% of the overall classification performance; while the DNN2 [9 7 5 3 1] configuration correctly classified patients response to drug regimens as follows: No Interaction {1011 (16.7%)}, Very Low Interaction {1640 (27.1%)}, Low Interaction {1531 (25.3%)}, High Interaction {954 (15.8%)} and Very High Interaction {878 (14.5%)}, representing 99.4% of the overall classification performance.

Visualizations of the receiver operating characteristic (ROC) curves for the traditional machine learning and transfer learning approaches are presented in Fig. 8 and Fig. 9, respectively. ROC curves show the trade-offs between sensitivity (or true positive rate: TPR) and specificity (1 − [false positive rate: FPR]). Classifiers which curves are neatly aligned to the top-left corner of the plot indicate better performing classifiers. On the average, better the DNN classifier variants showed better performance, with the HI and VHI curves aligning more closely to the top-left corner of the plots, indicating the robustness of the classifiers at predicting patients at risk of drug interaction or with high and very high interaction response to drug regimens.

6.2. Network performance evaluation

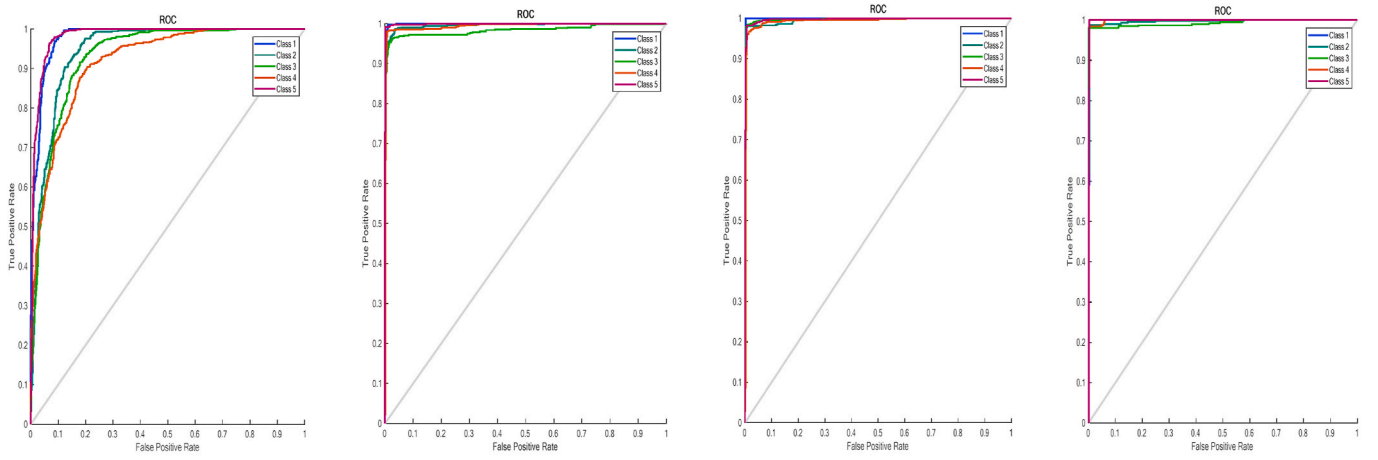
In this section, results of both classification approaches and how well they generalize across new datasets are evaluated for the neural network architectures considered in this study. Table 7 shows a summary of the root mean squared error (RMSE) and other classification metrics computed from the confusion matrices obtained in Figs. 7 and 8. (i.e., classification accuracy, precision, recall and F1-measure/score). From these results, a general improvement is observed as the number of neurons increases, but better performance when the DNN architecture was deployed. As the number of neurons in the 2-layer NN increases from Refs. [32–53], appreciable improvements in RMSE test performances were obtained for the traditional machine learning approach ([32] = 0.2530 [53], = 0.1044) and transfer learning approach ([32] = 0.2045 [53], = 0.1625); compared to the RMSE train performances obtained from traditional machine learning approach ([32] = 0.2602 [53], = 0.1040) and transfer learning approach ([32] = 0.2027 [53], = 0.1466). However, these results indicate not so good generalization or poor (regression) curve fitting, as the RMSE test results were slightly higher than the RMSE train results for both approaches, save the [53] configuration result of the traditional machine learning approach, which had a higher RMSE value.

Generally, a test error is expectedly higher than a training error, but should not be significantly higher, otherwise, the model overfits and generalizes poorly. However, this notion is not completely universal, as our DNN architecture exhibit good generalizations. Hence, using traditional machine learning approach, RMSE values of 0.0112 and 0.1861 were obtained as test and train errors, respectively, for DNN1 [9 7 5 3 1] configuration, compared to DNN2 [9 7 5 3 1] configuration, which gave 0.0102 and 0.0640 as RMSE values for test and train errors, respectively. Similarly, using transfer learning approach, RMSE values of 0.0082 and 0.0566 were obtained as test and train errors, respectively, for DNN1 [9 7 5 3 1] configuration, compared to DNN2 [9 7 5 3 1] configuration, which gave 0.0056 and 0.0510 as RMSE values for test and train errors, respectively.

Table 8 compares our approach with state-of-the-art literature. Observe that the transfer learning approach gave the best classification accuracy, but its recall rate is slightly lower than [48]. Although imbalanced features classification is typical of disease problems and diminishes dependence on accuracy as a good measure for assessing models' performance, the proposed system still compares favorably with the literature and can be hybridized to improve its precision and recall rates.

6.3. Scientific implications of the research and study limitation

Instead of estimating relative risk of patients using generalized linear/logistic regression or statistical models [29,36–42], we employ machine learning approach to model fitting, by training a fraction of the datasets with the model and then validating the trained datasets before



(a) NN [3 2] (b) NN [5 3] (c) DNN1 [9 7 5 3 1] (d) DNN2 [9 7 5 3 1]

Fig. 8. ROC curves for traditional machine learning method.

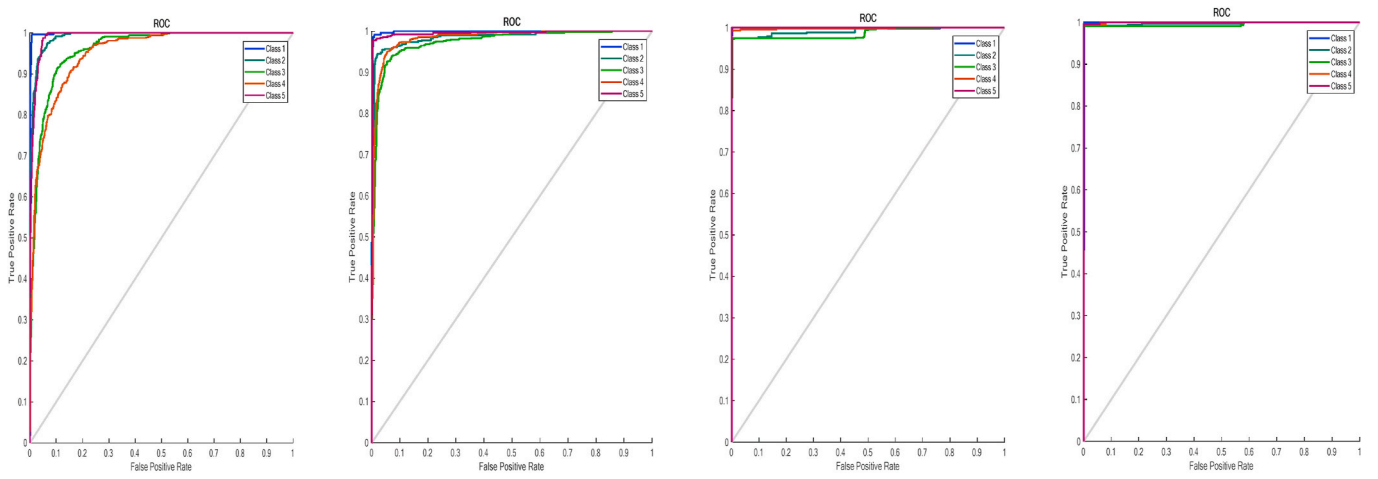


Fig. 9. ROC curves for transfer learning method.

Table 7
Summary of performance metrics.

Performance Metric		Traditional Machine Learning Approach				Transfer Learning Approach			
		NN [32]	NN [53]	DNN1 [9 7 5 3 1]	DNN2 [9 7 5 3 1]	NN [32]	NN [53]	DNN1 [9 7 5 3 1]	DNN2 [9 7 5 3 1]
RMSE	Training	0.2602	0.1040	0.1861	0.0640	0.2027	0.1466	0.0566	0.0510
	Validation	0.2648	0.1145	0.1100	0.0686	0.2007	0.1546	0.0768	0.0520
	Testing	0.2530	0.1044	0.0112	0.0102	0.2045	0.1652	0.0082	0.0056
	Overall	0.2593	0.1076	0.1024	0.0476	0.2032	0.1543	0.0472	0.0362
Classification Accuracy (%)		74.50	96.80	97.10	98.90	85.70	92.60	98.60	99.40
Precision		0.7566	0.9680	0.9776	0.9876	0.8686	0.9336	0.9846	0.9936
Recall		0.7430	0.9610	0.9796	0.9906	0.8756	0.9350	0.9880	0.9946
F1-Score		0.7448	0.9645	0.9786	0.9891	0.8721	0.9343	0.9863	0.9941

Table 8
Comparison of proposed model with previous related works.

Reference	Method; Drug Type; Data(sets)	Precision	Recall	F1-Score	Classification Accuracy (%)	RMSE
[42]	Fuzzy-Multidimensional Deep learning; HIV drugs; Control dataset; 1521 unique patient records.	0.9889	0.9919	0.9904	97.00	0.1153
[42]	Fuzzy-Multidimensional Deep learning; HIV drugs; Mixed dataset; 1301 unique patient records.	0.9974	0.9984	0.9979	98.87	0.0616
[43]	Hybrid technique (fuzzy logic, least squares optimization, NN); HIV drugs; Control dataset; 1527 unique patient records.	–	–	9.8920	98.90	1.83
[43]	Hybrid technique (fuzzy logic, least squares optimization, NN); HIV drugs; Mixed dataset; 1056 unique patient records.	–	–	9.6780	98.40	1.10
Shtar et al. [59]	Adjacency matrix factorization (AMF) and Adjacency matrix factorization with propagation (AMFP); HIV drugs; DrugBank; 1440 drugs and 248 drug–drug interactions.	–	–	–	56.00	–
This Study	Traditional machine learning DNN2 [9 7 5 3 1]; HIV drugs	0.9876	0.9906	0.9891	98.90	0.0473
This Study	Transfer learning DNN2 [9 7 5 3 1]; HIV drugs	0.9936	0.9946	0.9941	99.40	0.0362

testing the predicted values. The scientific implication of this approach is to aid manage the tradeoff between the variance (how much the model changes in response to the training data) and the bias (the strength of assumptions made about the data). By validating the training data, we can discover cases of underfitting or overfitting, or simply put, a degraded model performance, which can be fine-tuned further, using regularization techniques. Cross-validation (the process where data is split into different subsets) is an excellent way of dealing with this defect. The availability of control and experiment datasets can guide further tuning of the variance and bias in response to the training data (for better generalization). This is opposed to the deduced inference supported by statistical models without verifiable datasets. Control and experiment datasets also have positive consequence for advancing the progress in HIV/AIDS research, as cases of under estimation or over estimation can easily be quelled to avoid false alarms. Soft computing methods are therefore necessary for evaluating discovered patterns in datasets with potentials of increased accuracy and performance. In this paper, patterns were easily established and preserved from clinical trial datasets compared to medical images, which to a great extent may render unexpected outliers due to low quality or distortions. Sparsity of non-image datasets such as HIV may yield poor generalization of findings, because many samples-initiated ART data are at advanced stages of HIV due to resource-limited situation, comorbidity, mortality, and viral transmission. To evaluate new drugs with no known interaction, the labeled datasets could be modeled to recognize drug patterns with improved therapy and use this knowledge to discern unknown

interactions.

The proposed approach can only predict known (test) set of learned domain-specific datasets or feature-space and cannot adapt to domain features outside the input datasets. Sparse training dataset constitute another limitation of this study. Although limited datasets have always hindered the need for training data, methods such as transfer learning, online learning and the use of high-fidelity models have evolved to generate inexpensive datasets [58].

7. Conclusion and future research direction

Although ingenious and active ART have been the effectual measure for mitigating associative risk of HIV infection, treatment failure is inevitable, as prescribed drug regimens do not achieve maximal therapeutic effect. To a large extent, this challenge inhibits the proper management of HIV/AIDS patients and remains a major health burden in low- and medium-income countries. Factors that can cause treatment failure in ART include drug resistance, drug toxicity, poor or non-adherence to therapy, and late discovery of HIV status leading to the late initiation of treatment. Knowledge on these factors is therefore required by clinicians to make better decisions about patient diagnosis and treatment options, while understanding the possible outcomes and cost for each one. The value of machine learning is in its ability to learn experiential knowledge from existing datasets using nature-inspired algorithms, and then reliably convert the resulting analysis into clinical insights, ultimately producing better outcomes, lowered costs of care, and increased patient satisfaction. Using existing HIV datasets coded by experienced domain experts; machine learning methods could learn patterns associated with treatment failure, HIV risk, patient behavior, etc. Major outcomes of machine learning in healthcare include optimal clinical trials and new treatment options that have prolonged the lives of patients.

This paper has proposed a machine learning approach for the classification of failed treatment in mixed HIV dataset. Using supervised deep transductive-transfer learning, higher accuracy and generalization of the model were obtained compared to traditional machine learning. Our transfer learning approach maintained high accuracy and reduced performance errors for the DNN model architecture–indicating acceptable isolation of failed cases and good model generalization compared to the NN model architecture. Although convolutional learning models used in mining medical images are found to outsmart other state-of-the-art models [44], they perform poorly with limited (image) datasets, and are largely unsupervised. Furthermore, tuning of the available data into image-like structure is required. A future direction of this paper is the deployment of our proposed model as an expert system, for early and efficient detection of failed treatment and effectual drugs regimen prescription to patients on ART.

Funding

This research is funded by the Tertiary Education Trust Fund (TETFund), Nigeria.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We acknowledge the anonymous reviewers for their invaluable comments that have contributed to improving the quality of this paper.

References

[1] Arts EJ, Hazuda DJ. HIV-1 antiretroviral drug therapy. Cold Spring Harbor Perspect Med 2012;2(4):1–23. https://doi.org/10.1101/cshperspect.a007161.
[2] Archin NM, Sung JM, Garrido C, Soriano-Sarabia N, Margolis DM. Eradicating HIV-1 infection: seeking to clear a persistent pathogen. Nat Rev Microbiol 2014;12(11):750–64. https://doi.org/10.1038/nrmicro3352.
[3] Hurt CB, Eron Jr JJ, Cohen MS. Pre-exposure prophylaxis and antiretroviral resistance: HIV prevention at a cost? Clin Infect Dis 2011;53(12):1265–70. https://doi.org/10.1093/cid/cir684.
[4] Baggaley RF, Powers KA, Boily MC. What do mathematical models tell us about the emergence and spread of drug-resistant HIV? Curr Opin HIV AIDS 2011;6(2):131. https://doi.org/10.1097/COH.0b013e328343ad03.
[5] Supervie V, Barrett M, Kahn JS, Musuka G, Moeti TL, Busang L, Blower S. Modeling dynamic interactions between pre-exposure prophylaxis interventions & treatment programs: predicting HIV transmission & resistance. Sci Rep 2011;1:185. https://doi.org/10.1038/srep00185.
[6] Bershteyn A, Eckhoff PA. A model of HIV drug resistance driven by heterogeneities in host immunity and adherence patterns. BMC Syst Biol 2013;7(1):11. https://doi.org/10.1186/1752-0509-7-11.
[7] Gautam R, Sharma M. Prevalence and diagnosis of neurological disorders using different deep learning techniques: a meta-analysis. J Med Syst 2020;44(2):1–24. https://doi.org/10.1007/s10916-019-1519-7.
[8] Sharma S, Singh G, Singh D. Role and performance of different traditional classification and nature-inspired computing techniques in major research areas. EAI Endorsed Trans Scalable Informat Syst 2019;6(21):1–17. https://eudl.eu/doi/10.4108/eai.13-7-2018.158419.
[9] Sharma M, Kaur P. A comprehensive analysis of nature-inspired meta-heuristic techniques for feature selection problem. Arch Comput Methods Eng 2020:1–25. https://doi.org/10.1007/s11831-020-09412-6.
[10] Sharma M, Singh G, Singh R. A review of different cost-based distributed query optimizers. Progr Artificial Intell 2019;8(1):45–62. https://doi.org/10.1007/s13748-018-0154-8.
[11] Aswani R, Kar AK, Ilavarasan PV. Detection of spammers in twitter marketing: a hybrid approach using social media analytics and bio inspired computing. Inf Syst Front 2018;20(3):515–30. https://doi.org/10.1007/s10796-017-9805-8.
[12] Thulasiram RK, Thulasiraman P, Prasain H, Jha GK. Nature-inspired soft computing for financial option pricing using high-performance analytics. Concurrency Comput Pract Ex 2016;28(3):707–28. https://doi.org/10.1002/cpe.3360.
[13] Gavves E, Mensink T, Tommasi T, Snoek CG, Tuytelaars T. Active transfer learning with zero-shot priors: reusing past datasets for future tasks. Proc. IEEE Int Conf Comput Vis 2015:2731–9. https://doi.org/10.1109/ICCV.2015.313.
[14] Jie L, Tommasi T, Caputo B. Multiclass transfer learning from unconstrained priors. In: 2011 International Conference on Computer Vision. IEEE; 2011. p. 1863–70. https://doi.org/10.1109/ICCV.2011.6126454.
[15] Kuzborskij I, Orabona F, Caputo B. From n to n+1: multiclass transfer incremental learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2013. p. 3358–65. https://doi.org/10.1109/CVPR.2013.431.
[16] Srivastava N, Salakhutdinov RR. Discriminative transfer learning with tree-based priors. Adv Neural Inf Process Syst 2013;26:2094–102.
[17] Rohrbach M, Ebert S, Schiele B. Transfer learning in a transductive setting. In: Advances in neural information processing systems; 2013. p. 46–54.
[18] Xie L, Deng Z, Xu P, Choi KS, Wang S. Generalized hidden-mapping transductive transfer learning for recognition of epileptic electroencephalogram signals. IEEE Trans Cybernet 2018;49(6):2200–14. https://doi.org/10.1109/TCYB.2018.2821764.
[19] Kobylarz J, Bird JJ, Faria DR, Ribeiro EP, Ekárt A. Thumbs up, thumbs down: non-verbal human-robot interaction through real-time EMG classification via inductive and supervised transductive transfer learning. J Ambient Intell Humanized Comput 2020:1–11. https://doi.org/10.1007/s12652-020-01852-z.
[20] Zhu Y, Chen Y, Lu Z, Pan SJ, Xue GR, Yu Y, Yang Q. Heterogeneous transfer learning for image classification. AAAI 2011;11:1304–9.
[21] Du B, Zhang L, Tao D, Zhang D. Unsupervised transfer learning for target detection from hyperspectral images. Neurocomputing 2013;120:72–82. https://doi.org/10.1016/j.neucom.2012.08.056.
[22] Chang H, Han J, Zhong C, Snijders AM, Mao JH. Unsupervised transfer learning via multi-scale convolutional sparse coding for biomedical applications. IEEE Trans Pattern Anal Mach Intell 2017;40(5):1182–94. https://doi.org/10.1109/TPAMI.2017.2656884.
[23] Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2009;22(10):1345–59. https://doi.org/10.1109/TKDE.2009.191.
[24] Naseer A, Rani M, Naz S, Razzak MI, Imran M, Xu G. Refining Parkinson's neurological disorder identification through deep transfer learning. Neural Comput Appl 2020;32(3):839–54. https://doi.org/10.1007/s00521-019-04069-0.
[25] Ma X, Chen Q, Yu Y, Sun Y, Ma L, Zhu Z. A two-level transfer learning algorithm for evolutionary multitasking. Front Neurosci 13,1408. https://doi.org/10.3389/fnins.2019.01408.
[26] Zhang XX, Li HX, Cheng C, Ma S. Transfer learning based 3D fuzzy multivariable control for an RTP system. Appl Intell 2020;50(3):812–29. https://doi.org/10.1007/s10489-019-01557-7.
[27] Cheng L, Wang K, Tsung F. A hybrid transfer learning framework for in-plane freeform shape accuracy control in additive manufacturing. IISE Trans 2020:1–5. https://doi.org/10.1080/24725854.2020.1741741.
[28] Robbins GK, Daniels B, Zheng H, Chueh H, Meigs JB, Freedberg KA. Predictors of antiretroviral treatment failure in an urban HIV clinic. J Acquir Immune Defic

Syndr 1999;44(1):30. https://doi.org/10.1097/01.qai.0000248351.10383.b7. 2007.
[29] Ekong E, Ndembi N, Okonkwo P, Dakum P, Idoko J, Banigbe B, Okuma J, Agaba P, Blattner W, Adebamowo C, Charurat M. Epidemiologic and viral predictors of antiretroviral drug resistance among persons living with HIV in a large treatment program in Nigeria. AIDS Res Ther 2020;17(1):1–8. https://doi.org/10.1186/s12981-020-0261-z.
[30] Singh Y. Machine learning to improve the effectiveness of ANRS in predicting HIV drug resistance. Healthcare Informat Res 2017;23(4):271–6. https://doi.org/10.4258/hir.2017.23.4.271.
[31] Dragsted UB, Mocroft A, Vella S, Viard JP, Hansen AB, Panos G, Mercey D, Machala L, Horban H, Lundgren JD, EuroSIDA Study Group. Predictors of immunological failure after initial response to highly active antiretroviral therapy in HIV-1-infected adults: a EuroSIDA study. J Infect Dis 2004;190(1):148–55. https://doi.org/10.1086/420786.
[32] Parienti JJ, Massari V, Descamps D, Vabret A, Bouvet E, Larouzé B, Verdon R. Predictors of virologic failure and resistance in HIV-infected patients treated with nevirapine-or efavirenz-based antiretroviral therapy. Clin Infect Dis 2004;38(9):1311–6. https://doi.org/10.1086/383572.
[33] Haile D, Takele A, Gashaw K, Demelash H, Nigatu D. Predictors of treatment failure among adult antiretroviral treatment (ART) clients in bale zone hospitals, south eastern Ethiopia. PloS One 2016;11(10):e0164299. https://doi.org/10.1371/journal.pone.0164299.
[34] Hamers RL, Sigaloff KC, Wensing AM, Wallis CL, Kityo C, Siwale M, Mandaliya K, Ive P, Botes ME, Wellington M, Osibogun A. Patterns of HIV-1 drug resistance after first-line antiretroviral therapy (ART) failure in 6 sub-Saharan African countries: implications for second-line ART strategies. Clin Infect Dis 2012;54(11):1660–9. https://doi.org/10.1093/cid/cis254.
[35] Hosseinipour MC, van Oosterhout JJ, Weigel R, Phiri S, Kamwendo D, Parkin N, Fiscus SA, Nelson JA, Eron JJ, Kumwenda J. The public health approach to identify antiretroviral therapy failure: high-level nucleoside reverse transcriptase inhibitor resistance among Malawians failing first-line antiretroviral therapy. AIDS (London, England) 2009;23(9):1127. https://doi.org/10.1097/QAD.0b013e32832ac34e.
[36] Babo YD, Alemie GA, Fentaye FW. Predictors of first-line antiretroviral therapy failure amongst HIV-infected adult clients at Woldia Hospital, Northeast Ethiopia. PloS One 2017;12(11):e0187694. https://doi.org/10.1371/journal.pone.0187694.
[37] Negash H, Legese H, Tefera M, Mardu F, Tesfay K, Gebresilasie S, Fseha B, Kahsay T, Gebrewahd A, Berhe B. The effect of tuberculosis on immune reconstitution among HIV patients on highly active antiretroviral therapy in Adigrat general hospital, eastern Tigrai, Ethiopia; 2019: a retrospective follow up study. BMC Immunol 2019;20(1):1–9. https://doi.org/10.1186/s12865-019-0327-7. 45.
[38] Bezabih YM, Beyene F, Bezabhe WM. Factors associated with first-line antiretroviral treatment failure in adult HIV-positive patients: a case-control study from Ethiopia. BMC Infect Dis 2019;19(1):1–8. https://doi.org/10.1186/s12879-019-4170-5. 537.
[39] Steinbrink J, Imlay H, Rao K, Riddell J. Identifying causes of antiretroviral treatment failure in HIV-infected patients. In: Open Forum Infectious Diseases, vol. 3. Oxford University Press; 2016. p. 1547. https://doi.org/10.1093/ofid/ofw172.1248. suppl_1.
[40] Pacheco PR, Zara AL, Silva e Souza LC, Turchi MD. Late onset of antiretroviral therapy in adults living with HIV in an urban area in Brazil: prevalence and risk factors. J Trop Med 2019:1–8. https://doi.org/10.1155/2019/5165313.
[41] Feleke R, Geda B, Teji Roba K, Weldegebreal F. Magnitude of antiretroviral treatment failure and associated factors among adult HIV-positive patients in Harar public hospitals, Eastern Ethiopia. SAGE Open Med 2020;8:1–7. https://doi.org/10.1177/2050312120906076.
[42] Ahmed M, Merga H, Jarso H. Predictors of virological treatment failure among adult HIV patients on first-line antiretroviral therapy in Woldia and Dessie hospitals, Northeast Ethiopia: a case-control study. BMC Infect Dis 2019;19(1):1–7. https://doi.org/10.1186/s12879-019-3924-4. 305.
[43] Sisay MM, Ayele TA, Gelaw YA, Tsegaye AT, Gelaye KA, Melak MF. Incidence and risk factors of first-line antiretroviral treatment failure among human immunodeficiency virus-infected children in Amhara regional state, Ethiopia: a retrospective follow-up study. BMJ Open 2018;8(4):e019181. https://bmjopen.bmj.com/content/8/4/e019181.
[44] Pacheco PR, Zara AL, Silva e Souza LC, Turchi MD. Late onset of antiretroviral therapy in adults living with HIV in an urban area in Brazil: prevalence and risk factors. J Trop Med 2019. https://doi.org/10.1155/2019/5165313.
[45] Bisaso KR, Karungi SA, Kiragga A, Mukonzo JK, Castelnuovo B. A comparative study of logistic regression based machine learning techniques for prediction of early virological suppression in antiretroviral initiating HIV patients. BMC Med Inf Decis Making 2018;18(1):1–10. 77.
[46] Dey S, Luo H, Fokoue A, Hu J, Zhang P. Predicting adverse drug reactions through interpretable deep learning framework. BMC Bioinf 2018;19(21):1–13. https://doi.org/10.1186/s12911-018-0659-x.
[47] Kiweewa F, Esber A, Musingye E, Reed D, Crowell TA, Cham F, Semwogerere M, Namagembe R, Nambuya A, Kafeero C, Tindikahwa A. HIV virologic failure and its predictors among HIV-infected adults on antiretroviral therapy in the African Cohort Study. PloS One 2019;14(2):1–16. https://doi.org/10.1371/journal.pone.0211344.
[48] Ekpenyong ME, Etebong PI, Jackson TC. Fuzzy-multidimensional deep learning for efficient prediction of patient response to antiretroviral therapy. Heliyon 2019;5(7):e02080. https://doi.org/10.1016/j.heliyon.2019.e02080. 1-14.
[49] Ekpenyong ME, Etebong PI, Jackson TC, Udofa EM. Modeling drugs interaction in treatment-experienced patients on antiretroviral therapy. Soft Comput 2020;24:17349–64. https://doi.org/10.1007/s00500-020-05024-1.
[50] Huynh BQ, Li H, Giger ML. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. J Med Imag 2016;3(3):1–6. https://doi.org/10.1117/1.JMI.3.3.034501.
[51] Page A, Shea C, Mohsenin T. Wearable seizure detection using convolutional neural networks with transfer learning. In: 2016 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE; 2016. p. 1086–9. https://doi.org/10.1109/ISCAS.2016.7527433.
[52] Cha KH, Hadjiiski L, Chan HP, Weizer AZ, Alva A, Cohan RH, Caoili EM, Paramagul C, Samala RV. Bladder cancer treatment response assessment in CT using radiomics with deep-learning. Sci Rep 2017;7:8738. https://doi.org/10.1038/s41598-017-09315-w.
[53] Deniz E, Şengür A, Kadiroğlu Z, Guo Y, Bajaj V, Ü Budak. Transfer learning based histopathologic image classification for breast cancer detection. Health Inf Sci Syst 2018;6(1):1–7. https://doi.org/10.1007/s13755-018-0057-x.
[54] Christodoulidis S, Anthimopoulos M, Ebner L, Christe A, Mougiakakou S. Multisource transfer learning with convolutional neural networks for lung pattern analysis. IEEE J Biomed Health Informat 2016;21(1):76–84. https://doi.org/10.1109/JBHI.2016.2636929.
[55] Samala RK, Chan HP, Hadjiiski L, Helvie MA, Wei J, Cha K. Mass detection in digital breast tomosynthesis: deep convolutional neural network with transfer learning from mammography. Med Phys 2016;43(12):6654–66. https://doi.org/10.1118/1.4967345.
[56] Li X, Pang T, Xiong B, Liu W, Liang P, Wang T, Li X, Pang T, Xiong B, Liu W, Liang P. Wang T. Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification. In: 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). IEEE; 2017. p. 1–11. https://doi.org/10.1109/CISP-BMEI.2017.8301998. 2017.
[57] Turki T, Wei Z, Wang JT. Transfer learning approaches to improve drug sensitivity prediction in multiple myeloma patients. IEEE Access 2017;5:7381–93. https://doi.org/10.1109/ACCESS.2017.2696523.
[58] Li W, Gu S, Zhang X, Chen T. Transfer learning for process fault diagnosis: knowledge transfer from simulation to physical processes. Comput Chem Eng 2020;139:106904. https://doi.org/10.1016/j.compchemeng.2020.106904.
[59] Shtar G, Rokach L, Shapira B. Detecting drug–drug interactions using artificial neural networks and classic graph similarity measures. PloS One 2019;14(8):1–21. https://doi.org/10.1371/journal.pone.0219796.