Predicting First-Episode Psychosis Associated with Cannabis Use with Artificial Neural Networks and Deep Learning

Daniel Stamate[1], Wajdi Alghamdi[1*], Daniel Stahl[2], Ida Pu[1],
Fionn Murtagh[3], Danielle Belgrave[4], Robin Murray[5] and Marta di Forti[6]

[1] Data Science & Soft Computing Lab, and Department of Computing, Goldsmiths, University of London.
[2] Department of Biostatistics & Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London.
[3] School of Computing and Engineering, University of Huddersfield.
[4] Microsoft Research Cambridge.
[5] Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London.
[6] MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London.
[*] Joint first author, Corresponding author, Email: map01wa@gold.ac.uk

Abstract. In recent years, a number of researches started to investigate the existence of links between cannabis use and psychotic disorder. More recently, artificial neural networks and in particular deep learning have set a revolutionary wave in pattern recognition and machine learning. This study proposes a novel machine learning approach based on neural network and deep learning algorithms, to developing highly accurate predictive models for the onset of first-episode psychosis. Our approach is based also on a novel methodology of optimising and post-processing the predictive models in a computationally intensive framework. A study of the trade-off between the volume of the data and the extent of uncertainty due to missing values, both of which influencing the predictive performance, enhanced this approach. Furthermore, we extended our approach by proposing and encapsulating a novel post-processing k-fold cross-testing method in order to further optimise, and test these models. The results show that the average accuracy in predicting first-episode psychosis achieved by our models in intensive Monte Carlo simulation, is about 89%.

Keywords: First-episode psychosis, Precision medicine, Cannabis use, Prediction modelling, Classification, Neural Network, Deep Learning, Post-processing, Monte Carlo simulation, Missing data based uncertainty.

1 Introduction

An estimated 183 million people consumed cannabis in 2014 [1] making it the most popular illicit drug in the world. Legalising cannabis, especially in countries such as the Netherlands and Uruguay, and in some states of USA, and the increasing lobby for

making cannabis use legal in other countries such as Canada, is an important contributing factor for the popularity of this drug. On the other hand evidence shows that the increase in cannabis consumption is proportionate to the increase in the proportion of people seeking treatment for psychotic disorders [1]. While there is some evidence that consuming cannabis is a risk factor for several types of psychotic disorders [2], the link between these two factors needs to be better quantified.

These days, researchers attempted to understand whether specific patterns of cannabis use such as potency or age are associated with a higher risk of developing psychotic disorders. One study concluded that nearly a quarter of all new psychosis patients in South London (UK) could be associated with the use of high-potency, skunk-like cannabis [3]. Another study [4] estimated that if a person uses cannabis daily for more than six months, then there is a 70% likelihood that this person will suffer from psychotic disorders.

There are few such studies based on risk prediction modelling using advanced machine learning algorithms establishing a link between cannabis use and first-episode psychosis – in fact we are not aware of the existence of other studies apart our recent work [4]. Most studies so far rely only on explanatory research strategies and are mainly based on a number of conventional statistical techniques such as hypotheses formulation and verification via statistical tests, logistic regression modelling, etc. These techniques are well-recognised and used in medical research, but in many situations, they do not match the high potential of machine learning methods. The domain of machine learning has developed at an enormous speed in recent years, with advanced predictive techniques being expanded and improved upon. In particular artificial neural networks and especially deep networks, which are state of the art in prediction, have proven their abilities in many pattern recognition and machine learning applications. One such field of implementation is the domain of medical research [5] [6].

On the one hand, artificial neural networks have been successfully used in understanding the heterogeneous manifestations of asthma [7], diagnosing tuberculosis [8], classifying leukaemia [9], detecting heart conditions in ECG data [10], etc. These studies show that neural networks have been proven to be capable of dealing with complicated medical data such as the ambiguous nature of the ECG signal data, where neural networks show some outstanding results compared to other methods.

On the other hand, recently, deep networks have attracted widespread attention, mainly by defeating alternative machine learning methods such as support vector machines in numerous critical applications such as classifying Alzheimer's disease [11], classifying AD/MCI patients [12], and improving palliative care [13]. While support vector machines are still popular techniques within the machine learning community [4] [14], the family of deep learning techniques are gaining considerable attention [15]. Deep learning methods are types of representation learning methods, which can automatically identify the optimal representation of raw data without requiring prior feature selection.

In this study, we propose a novel machine learning approach based on neural networks and deep learning techniques to develop predictive models for the onset of first-episode psychosis. The dataset that we based our study upon was collected by psychiatry practitioners, and used in previously conducted studies such as [3] [4]. It comprises

an extensive set of variables including demographics, drug-related, and several other variables with specific information on the participants' history of cannabis use as seen in Table 1.

Our approach features a gradual control of the limitation of the uncertainty present in the data due to missing values which are usually inherent in clinical datasets due to patients missing appointments, patients not reporting all details, etc. This feature involves considering different thresholds for allowed levels of missingness (per attributes and per records) in the data sets, that we call cutting points, in order to examine how the prediction models' performances may vary with these thresholds. Our approach is based also on a novel methodology of optimising and post-processing the predictive models in a computationally intensive framework. Furthermore, we extended our approach by proposing and encapsulating a novel post-processing k-fold cross-testing method in order to further optimise, and test these models. The results show that the accuracy in predicting first-episode psychosis achieved by our best models in intensive Monte Carlo simulation, falls between 85.13% and 91.54%, with an average of about 89%.

Table 1. Cannabis use attributes among other attributes in the analysed dataset

Attribute	Description
lifetime_cannabis_user	Ever used cannabis: yes or no
age_first_cannabis	Age when first used cannabis: 7 to 50
age_first_cannabis_under15	Age less than 15 when first used cannabis: yes, no or never used
age_first_cannabis_under14	Age less than 14 when first used cannabis: yes, no or never used
current_cannabis_user	Current cannabis user: yes or no
cannabis_fqcy	Pattern of cannabis use: never used, only at weekends, or daily
cannabis_measure	Cannabis usage measure: none, hash less than once per week, or hash at weekends, hash daily, skunk less than once per week, or skunk at weekends, skunk daily
cannabis_type	Cannabis type: never used, hash, or skunk
duration	Cannabis use duration: 0 to 41 (months)

2 Methods

2.1 The clinical data

The data used to develop our novel approach to predict the first-episode psychosis is a part of a case-control study at the inpatient units of the South London and Maudsley (SLaM) NHS Foundation Trust in United Kingdom [3]. The clinical data consists of 1106 records, including 489 patients, 370 controls and 247 unlabelled records. Those described as patients were patients of the Trust who at one time presented with first-episode psychosis; controls were healthy people recruited from the local area. Each

record refers to a participant in the study and has 255 possible attributes, which were divided into four categories. The first category consists of demographic attributes which represent general features such as gender, race, and level of education. Secondly, drug-related attributes contain information on the use of non-cannabis drugs such as tobacco, stimulants and alcohol. The third category is formed of genetic attributes which were removed from the analysis for the purpose of this study. The final category contains cannabis-related attributes such as the duration of use, initial date of use, frequency, cannabis type, etc (see Table 1).

2.2 Rationalisation and refinement

The goal of this stage is to perform a high-level simplification of the dataset, and it embraces several steps. First, records that were missing critical data were removed from the dataset. This included records with missing labels as well as records with missing values on all cannabis-related variables. Secondly, certain variables were removed from the dataset. This primarily involved variables that were deemed to be irrelevant to the study (such as those related to individual IDs of the study participants), and also variables which were outside the scope of the current study (for example, certain gene-related variables). In addition, any numeric predictors that had zero or near-zero variance were dropped. Thirdly, we sought to make the encoding of missing values consistent across the dataset. Prior to this step, values including 66, 99, and -99 all represented cases with missing values – so all such indicators were replaced with a consistent missing value indicator, NA. Fourthly, some variables were re-labelled to provide more intuitive descriptions of the data contained within. Finally, since in multiple situations some variables had a similar meaning, yet there were often missing values for some records in some of these variables, a process of imputation was used to effectively combine the information from related variables into one. For example, two variables described alcohol use but were inconsistently present across the records and presented missing values. These were combined in a way that created one single variable with consistent and as complete as possible values. Such a process was used to generate value-reacher and value-consistent variables related to alcohol use, tobacco use, employment history, and subjects' age.

2.3 A trade-off between the extent of missing values and the dataset size

A trade-off between the extent of missing values present in the dataset, and the dataset size, needed to be investigated from the point of view of the predictive power of the models that can be built on the dataset. The intuition is that by using a larger subset of the available dataset in the analysis, one would obtain a positive effect on the performance of predictive models (since more data is used to build the models). But this larger subset may also encapsulate more uncertainty due to the presence of more missing values, which usually has a negative effect on the predictive models (even with imputation). Therefore, different cutting points, defined as the thresholds for the percentage of missing values (or level of missingness) allowed in attributes and records, respectively, were considered in order to study the variation of the predictive power of

subsets of the dataset. Attributes and records presenting some levels of missingness up to the respective cutting points or thresholds, respectively, were kept in the dataset, and the remaining ones were removed. The considered cutting points for the records were 10%, 20%, …, 100%. For instance 30% in this grid means that we keep in the dataset only the records that have up to 30% missing values (and 100% means practically that all records are kept in the dataset). Moreover, the cutting points for the attributes were identified by first determining the percentage of missing values for each attribute, and then ordering these percentages and splitting them into twenty equal groups. The extreme values in each group formed the cutting points for the attributes.

Overall, these cutting points were applied to the dataset and compared with respect to the performance of single-layer neural network tuned models, in an attempt to determine optimal cutting points which were those for which these models had the highest accuracy. Once these cutting points were determined, they were applied, and a final dataset was thus obtained as the outcome of a trade-off between the extent of missing values present in the dataset, and the dataset size.

How did we exactly proceed to obtain this final dataset? Note that we don't do a full optimisation on all pairs of cutting points for attributes and records to determine this final dataset (because training and tuning neural networks is a computationally expensive procedure), but we just apply a heuristic in our framework. Initially we search for an optimal value among all the attribute cutting points, and we apply it on the dataset. In our case this was 92%. Then, on the resulting dataset, we applied different record cutting points following the grid mentioned above, and we determined the best cutting point, which was 70% in our case. To compare the cutting points and select the best ones, the criterion was the accuracy of the single-layer neural networks which have been tuned on the training set (70% of the data), in a 5-fold cross-validation procedure, on a 10x10 grid for the number of hidden units, and decay values to prevent overfitting with regularisation methods. Random forest imputations of missing values were applied. The models' performances consisting of accuracy and kappa were estimated on the test set (30% of the data).

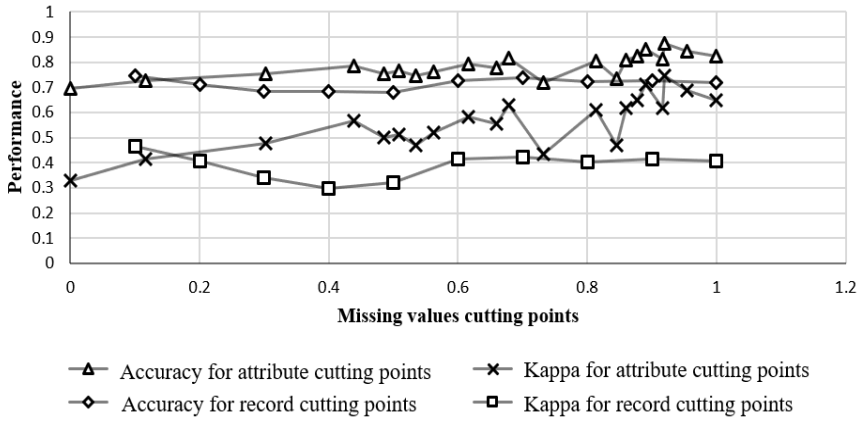


Fig. 1. Model performance for record and attribute cutting points

Fig. 1 illustrates the process, in which we observed a decrease in the performance when all the attributes were included or when the cannabis attributes were not present in the obtained dataset.

By applying the determined 92% cutting point for the attributes and 70% cutting point for the records to the original dataset, we obtained 107 attributes and 628 records divided into 360 patients and 268 controls, on which the main phase of predictive modelling with various algorithms was developed, and presented in what follows. We note that the proportion of controls and patients in the final dataset are approximately the same as in the original dataset, so the current dataset is representative.

2.4 Imputation

Missing values' presence in clinical data is rather common due to reasons explained above, and this is the case also of our dataset. The predictive power in the data may depend significantly on the way missing values are treated. While some machine learning algorithms, such as decision trees [16], have the capability to handle missing data outright, most machine learning algorithms do not. In many situations missing values are imputed using a supervised learning technique such as k-Nearest Neighbour (KNN) after suitable scaling to balance the contribution of the numeric attributes. These imputation techniques do not have theoretical formulations but have been much implemented in practice [4][6]. In this work, we considered different imputations such as the KNN imputation, the tree bagging imputation from the caret package [16], and the random forest imputation from the randomForest package [17]. The last method led to the best results in terms of the performance of the predictive models finally built, although it was more computationally expensive.

2.5 Training and optimizing (tuning) predictive models

For the purpose of developing optimised predictive models for the first-episode psychosis, the values of the parameters for each of the considered algorithms have been controlled by chosen grids. Predictive models have been fitted, in a 5-fold cross-validation procedure, on each training set after pre-processing techniques were applied on the same training set, and have been tested on each test set. Models based on neural networks with a single-layer, neural networks with multi-hidden-layers, and deep networks, were optimized (tuned) based on maximizing AUC, the area under the ROC curve.

The single-layer neural networks were tuned over 10 values of the size (i.e. the number of hidden units) and 10 values of the decay (i.e. the weight decay), which is the parameter in the penalization method for model regularization to avoid overfitting, similar to the penalization method in ridge regression, based on the L2 norm [16]. The optimal values were 3 and 0.01, respectively. The neural network with multi-hidden layers were tuned over 10 values for each of the 3 hidden layers (i.e. 10 values for the number of hidden units in each layer), and 10 values for the decay. The optimal values were 5, 5, 5 for the 3 layers, and 0.01 for decay, respectively.

As for the deep networks, we employed the H2O's deep learning, which is based on a multi-layer feedforward artificial neural network that is trained with stochastic gradient descent using back-propagation [20]. The deep networks usually contain a large number of hidden layers consisting of neurons with *tanh*, *rectifier*, and *maxout* activation functions. This type of models has many parameters, but it was designed to reduce the number of parameters that the researcher has to specify by applying feature selection and early stopping techniques. We used deep networks with the method of Gedeon [19] to select the best attributes. In our experiments, the early stopping was set to let it stop automatically once the area under the curve AUC does not continue improving, in particular, when AUC does not improve by at least 1% for 10 consecutive scoring events.

Also, a grid optimisation was used with the parameters that need to be tuned such as the activation function, the number and sizes of the hidden layers, the number of epochs, and the 2 parameters corresponding to the L1 and L2 regularisations for preventing overfitting.

The models were tuned over all activation functions, and over 3, 4, …, 25 layers and 30, 35, …, 50 layers. The number of units in each layer had the values 50, 100,…, 250. Also, we used the values 2, 3, 5, and 10 for tuning the number of epochs. Finally, the parameters for the L1 and L2 regularisations were each tuned over the values 10^{-1}, 10^{-2} ,…, 10^{-10}.

After performing the proposed techniques, the optimal values selected for the deep learning model are *rectifier* as an activation function, 5 epochs, and 8 hidden layers of 200 neurons each. As for the L1 and L2 parameters, the optimal values were 10^{-4} and 10^{-5}, respectively.

2.6 Sampling and post-processing k-fold cross-testing

When there is a priori knowledge of a class imbalance, one direct method to reduce its influence on model training is to select training set samples to have roughly equal event rates [16]. Treating data imbalance usually leads to better predictions models and better trade-off between sensitivity and specificity.

In this study, we considered three sampling approaches to subsample the training data in a manner that mitigates the imbalance problem. The first approach is down-sampling in which we sampled (without replacement) the majority class to be the same size as the minority class. The second method is up-sampling in which we sampled (with replacement) the minority class to be the same size as the majority class. The last approach we used is the synthetic minority over-sampling technique (SMOTE) [21]. SMOTE selects a data point randomly from the minority class, and the K-nearest neighbours to that point are determined and used to generate new synthetic data points by slight alterations to these data points. Five neighbours are used in our analysis. The results show that the up-sampling procedure had no real improvement on AUC or the accuracy performances. Simple down-sampling of the data also had no positive effect on the model performances. However, SMOTE with neural network models has led to an increase in AUC and accuracy.

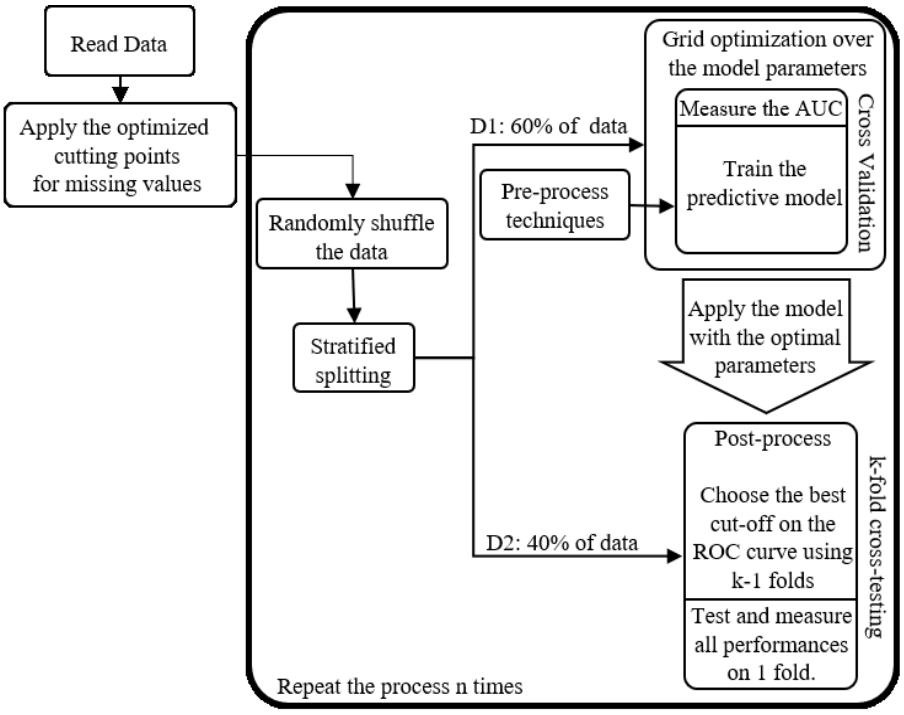


Fig. 2. Summary of the implemented methodology, with k-fold cross-testing method

Fig. 2 gives an overall description of the methodology followed here, based on pre-processing, model optimisation, and post-processing. The dataset is randomly split, with stratification, in 60% and 40% parts denoted here by D1 and D2, respectively. D1 is used for training and for optimising the model, as explained in Subsection 2.5, in a cross-validation fashion, with AUC as optimisation criterion, with and without class balancing. Different pre-processing methods such as missing values imputation and sampling methods that we have explained above, were appropriately integrated into the cross-validation. The optimal model obtained on D1 was then applied to score D2 accounting for the remaining 40% of the dataset. In order to further enhance the model performance, a specially designed post-processing procedure that we introduce here, was applied with the optimised model using D2 dataset. We call it the *k-fold cross-testing method*. In this procedure, we produce k post-processed model variants of the original optimised model. First, we create k stratified folds of D2 dataset. Then, k-1 folds are used to find an alternative probability cut-off on the ROC curve such as the cut-off associated with the largest accuracy. The remaining one-fold is scored with the post-processed model based on the newly found cut-off point. Finally, the whole procedure is repeated until all folds are used for scoring at their turn, then the predictions are integrated, and the model performance is measured on the whole scored dataset D2. We note here as an important remark that in each such iteration of the procedure, the

ROC optimisation data (the k-1 folds) and the scored data (the remaining fold) are always distinct, so the data for model post-processing and the data for scoring are always distinct.

2.7 Monte Carlo simulations

Due to expected potential variations of the predictive models' performance, depending on the datasets for training and testing, but in particular due to the uncertainties introduced by the missing values in the data, we conducted extensive Monte Carlo simulations to study these variations, and the stability of the models. In particular, the simulations for each single-layer neural networks, multi-layers neural networks and deep networks consisted of 2,000 iterations of the procedure included in the bold contour box of Figure 2. The models' performances consisting of accuracy, sensitivity, specificity, and kappa were evaluated in each iteration. The aggregation of all iterations formed various distributions of the above performance measures. These distributions were visualised using histograms to capture the performance capability and stability of models, as shown in the Results section.

2.8 Hardware and software

The Monte Carlo simulations that we conducted as explained above are computationally very expensive procedures, therefore a robust framework was required. Parallel processing was performed on a data analytics cluster of 11 servers with Xeon processors and 832GB fast RAM. The R software was used with a number of packages, including *caret, pROC, e1071, randomForest, ggplot2, plyr, DMwR, AppliedPredictiveModeling, doParallel* and *H2O*.

3 Results

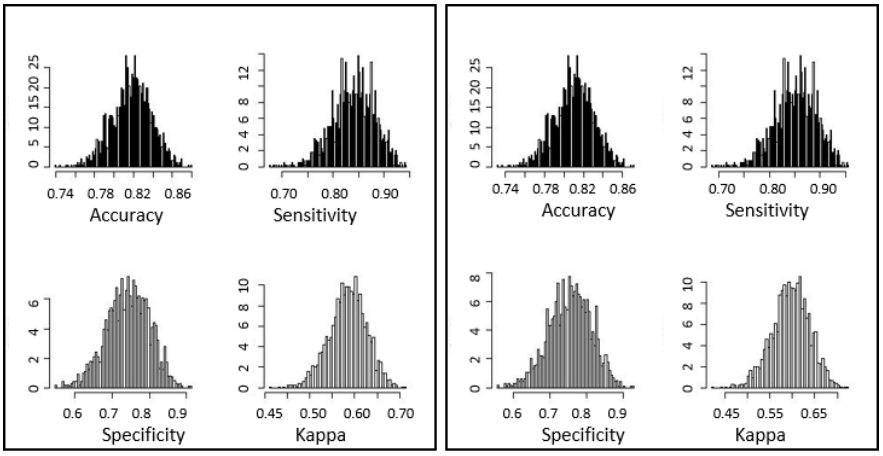
We present here the performances obtained with our approach to predicting first-episode psychosis, investigated with Monte Carlo simulations, as explained above. We should note that, due to lack of space, in this section we only report results regarding models which either are not post-processed, or are post-processed with ROC optimisation based on the largest accuracy cut-off methodology.

Table 2. Estimations of the predictive models' performances.

Model	Accuracy	Kappa	Sensitivity	Specificity
Single-layers neural networks	0.80	0.59	0.84	0.74
Multi-layers neural networks	0.81	0.60	0.85	0.75
Deep networks	0.89	0.76	0.83	0.93

The results show that the single-layer neural network scored a mean accuracy of 0.80 (95% CI [0.76, 0.84]) and a mean sensitivity of 0.84 (95% CI [0.76, 0.91]). Also, the

multi-layers neural networks achieved a mean accuracy of 0.81 (95% CI [0.77, 0.85]) and a mean sensitivity of 0.85 (95% CI [0.77, 0.92]). Figure 3 shows histogram plots of the Monte Carlo simulations for single and multi-layer neural networks with post-processing and performances evaluated with our k-fold cross-testing method. Results indicate that the difference between single and multi-layer neural networks is not significant regarding the 4 performances.

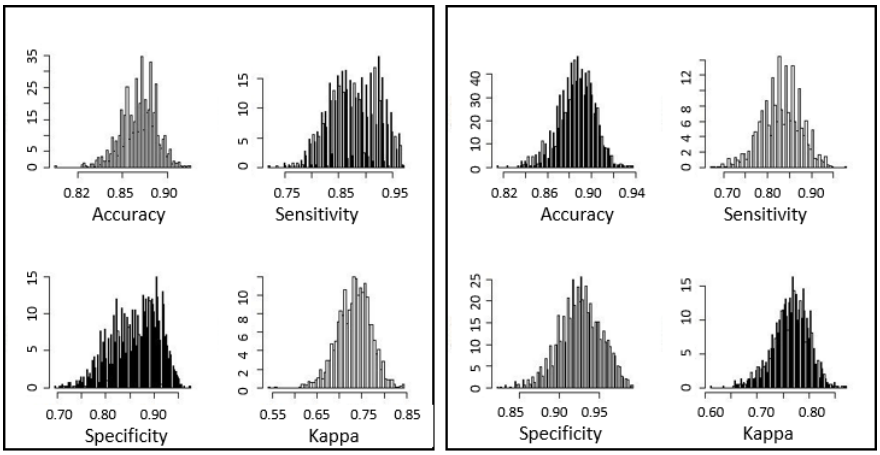


(a) Multi-layers neural networks (b) Single layer neural networks

Fig. 3. 2000 Monte Carlo simulation for neural networks.

As for deep learning, the results show significantly better performances. Figure 4 illustrates histogram plots of the 2000 Monte Carlo simulations for models based on deep networks without the post-processing (left) and with post-processing (right). The results for the latter show a mean accuracy of 0.89 (95% CI [0.85, 0.92]) and a mean sensitivity of 0.83 (95% CI [0.74, 0.92]).

Overall, we remark a good predictive power and stability of these models, based on an acceptable level of variation of their performance measures evaluated across extensive Monte Carlo experiments. As mentioned before, a significant proportion of this variation may be explained by the uncertainties due to the presence of missing values in the dataset.



(a) Deep network without post-processing (b) Deep network with post-processing

Fig. 4. 2000 Monte Carlo simulation for deep networks.

4 Conclusion and directions for further work

The aim of this work has been to propose a novel machine learning approach to developing predictive models for the onset of the first-episode psychosis with neural networks and deep learning. To our knowledge, previous studies on the link between cannabis use and first-episode psychosis investigated this highly important relationship via conventional statistical methodologies and techniques and did not tackle the predictability of this condition in relation to the cannabis use. An exception is [4] which is the first study to predict first episode-psychosis using machine learning based on support vector machines, bagged trees, boosted classification trees, eXtreme gradient boosting and random forests. However, the accuracy performances in [4] were slightly under 80%, and as such, under all neural and deep network models' performances achieved in this work.

In this paper, we successfully classified first-episode psychosis from normal control with 89% accuracy using deep learning. This solution proves the high potential of applicability of machine learning, in particular deep learning, in Psychiatry, and enables researchers and doctors to evaluate the risk for and predict first-episode psychosis.

Our approach features a gradual control of the limitation of the uncertainty present in the data by investigating a trade-off between the extent of missing values entailing uncertainty, and the dataset size. Moreover, due to expected potential variations of the predictive models' performance due to the uncertainties entailed by the remaining missing values in the data, we conducted extensive Monte Carlo simulations to study these variations, and the stability of the models.

A potential work direction concerns including genotype data in the study for prediction purposes, and redefining the predictive modelling approach by taking into account the particularities of the newly introduced data, such as the high dimensionality.

References

1. United Nations Office on Drugs and Crime, World Drug Report, (United Nations publication, Sales No. E.16.XI.7), 2016.
2. Radhakrishnan, R., Wilkinson, S., Dsouza, D.: Gone to Pot: A Review of the Association between Cannabis and Psychosis. Frontiers in Psychiatry, vol. 5 (2014).
3. Di Forti, M., Marconi, A., et al.: Proportion of Patients in South London with First-Episode Psychosis Attributable to Use of High Potency Cannabis: a Case-Control Study. The Lancet Psychiatry, vol. 2, no. 3, pp. 233-238, (2015).
4. Alghamdi, W., Stamate, D., et al.: A Prediction Modelling and Pattern Detection Approach for the First-Episode Psychosis Associated to Cannabis Use, IEEE ICMLA 15, pp.825-830, (2016).
5. Zhou, H., Tang, J., Zheng, H.: Machine Learning for Medical Applications. The Scientific World Journal, vol. 20, pp. 1-1, (2015).
6. Iniesta, R., Stahl, D., McGuffin P.: Machine learning, Statistical Learning and the Future of Biological Research in Psychiatry, Psychological Medicine, (2016).
7. Stamate, D., Cassidy, R., Belgrave: Predictive Modelling Strategies to Understand Heterogeneous Manifestations of Asthma in Early Life, IEEE ICMLA 16, (2017).
8. Elveren, E., Yumuşak, N.: Tuberculosis Disease Diagnosis Using Artificial Neural Network Trained with Genetic Algorithm, J Med Syst, 35: 329–332, (2011).
9. Adjouadi, M., Ayala, M., et al.: Classification of Leukemia Blood Samples Using Neural Networks. Ann Biomed Engineering, 38(4):1473-82, (2010).
10. Yan, Y., Qin, X., et al.: A Restricted Boltzmann Machine Based Two-Lead Electrocardiography Classification, in Proc. 12th Int. Conf. Wearable Implantable Body Sens, (2015).
11. Sarraf, S., Tofighi, G.: Classification of Alzheimer's Disease Using fMRI Data and Deep Learning Convolutional Neural Networks, arXiv:1603.08631, (2016).
12. Li, F., Tran, L., et al.: Robust Deep Learning for Improved Classification of AD/MCI Patients. Int. Workshop on Machine Learning in Medical Imaging, pp 247, (2014).
13. Avati, A., Jung, K., et al.: Improving Palliative Care with Deep Learning, (2017).
14. Katrinecz, A., Stamate, D., et al.: Predicting Psychosis Using the Experience Sampling Method with Mobile Apps, IEEE ICMLA 16, (2017).
15. Ravì, D., Wong, C., et al.: Deep Learning for Health Informatics. in IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 1, pp. 4-21, Jan. (2017).
16. Kuhn, M., Johnson, K.: Applied Predictive Modelling. Springer, (2013).
17. Liaw, A., Wiener, M.:Classification and Regression by randomForest, R News 2(3), (2002).
18. Arel, I., Rose, D., Karnowski, T.: Deep Machine Learning a New Frontier in Artificial Intelligence Research, Computational Intelligence Magazine, IEEE, (2010).
19. Aiello, S., Eckstrand, E., et al.: Machine Learning with R and H2O, (2016).
20. Candel, A., Parmar, V., LeDell, E., and Arora, A.: Deep Learning with H2O, http://h2o.ai/resources, (2015).
21. Qazi, N., Raza, K.: Effect of Feature Selection, SMOTE and Under Sampling on Class Imbalance Classification. 2012 UKSim 14th, pp. 145-150, (2012).