Can Artificial Neural Networks Predict Psychiatric Conditions Associated with Cannabis Use?

Daniel Stamate[1], Wajdi Alghamdi[1(✉)], Daniel Stahl[2],
Alexander Zamyatin[3], Robin Murray[4], and Marta di Forti[5]

[1] Data Science and Soft Computing Lab, Department of Computing,
Goldsmiths, University of London, London, UK
`map01wa@gold.ac.uk`
[2] Department of Biostatistics and Health Informatics, Institute of Psychiatry,
Psychology and Neuroscience, King's College London, London, UK
[3] Faculty of Informatics, Department of Applied Informatics, National Research
Tomsk State University, Tomsk, Russia
[4] Department of Psychosis Studies, Institute of Psychiatry, Psychology
and Neuroscience, King's College London, London, UK
[5] MRC Social, Genetic and Developmental Psychiatry Centre, Institute
of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

Abstract. This data-driven computational psychiatry research proposes a novel machine learning approach to developing predictive models for the onset of first-episode psychosis, based on artificial neural networks. The performance capabilities of the predictive models are enhanced and evaluated by a methodology consisting of novel model optimisation and testing, which integrates a phase of model tuning, a phase of model post-processing with ROC optimisation based on maximum accuracy, Youden and top-left methods, and a model evaluation with the k-fold cross-testing methodology. We further extended our framework by investigating the cannabis use attributes' predictive power, and demonstrating statistically that their presence in the dataset enhances the prediction performance of the neural network models. Finally, the model stability is explored via simulations with 1000 repetitions of the model building and evaluation experiments. The results show that our best Neural Network model's average accuracy of predicting first-episode psychosis, which is evaluated with Monte Carlo, is above 80%.

Keywords: Machine learning · Neural networks · Prediction modelling
ROC optimisation · Monte carlo · Computational psychiatry · Cannabis
Psychosis

1 Introduction

Policy changes and the legalisation of cannabis across countries, especially the Netherlands, Uruguay and some states in the USA, indicate that cannabis is gaining greater global acceptance. It has been reported that cannabis has been the most popular

illicit drug in the world in the last decade, with an estimated 183 million annual users [1]. However, studies show that countries with higher cannabis consumption also have higher proportions of people seeking treatment for psychotic disorders, including first-episode psychosis, schizophrenia, anxiety and substance use disorders [2]. This makes it imperative to understand the likely consequences of cannabis use, even if these consequences affect only a minority of users. Therefore, the link between cannabis consumption and the risk of experiencing psychotic disorders must be thoroughly scrutinised.

Recent researches have 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 a person who uses cannabis daily for more than six months has a 70% likelihood of suffering from psychotic disorders.

Few studies have used risk prediction modelling or advanced machine learning algorithms to establish a link between cannabis use and first-episode psychosis. In fact, apart from our recent work [4], we are not aware of the existence of any such studies. Most prior studies have relied solely on explanatory research strategies and been based on various conventional statistical techniques, such as hypothesis formulation and verification via statistical tests, logistic regression modelling, etc. These methods are well-recognised and widely used in medical research, but they rarely match the high potential of machine learning methods. The domain of machine learning has continued to develop for many years, and advanced predictive techniques have been expanded and improved constantly. These advanced predictive techniques, in turn, are prompting extensive usage of computers and artificial intelligence techniques in many domains, such as medicine [5, 6]. Artificial neural networks have proven their remarkable ability to detect predictive patterns in different types of datasets of various complexities, and often showed their superiority compared with other machine learning techniques.

Artificial neural networks are computational models that simulate the way biological neural networks process information in the human brain [7]. They are usually composed of several highly interconnected groups of artificial neurons that work together to solve specific problems. These computational models are typically used to model complex relationships between inputs and outputs, such as those in prediction models. Just like learning in biological neural networks, artificial neural networks process information by 'tuning' existing connections among neurons in order to process information.

While techniques involving support vector machines are still popular within the machine learning community [4, 8], artificial neural networks are gaining considerable attention again. Recently, artificial neural networks have been successfully used in understanding the heterogeneous manifestations of asthma [9], diagnosing tuberculosis [10], classifying leukaemia [11], detecting heart conditions in electrocardiogram (ECG) data [12], etc. These studies show that neural networks are capable of handling complex medical data, such as ambiguous ECG signal data, and achieving outstanding results not yet produced by other methods.

In this study, we offer a novel machine learning approach that uses neural networks to develop predictive models for the onset of first-episode psychosis. The dataset on which we based our study was collected by psychiatry practitioners and has been used in previously conducted studies, such as [3, 4]. It comprises an extensive set of variables, including demographic, drug-related and other variables, with specific information on participants' histories of cannabis use, as seen in Table 1.

Table 1. Cannabis use attributes in the analysed dataset [4].

Attribute	Description
lifetime_cannabis_user	Ever used cannabis: yes or no
age_first_cannabis	Age upon the first use of 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 on weekends or daily
cannabis_measure	Cannabis usage measure: none, hash less than once per week, hash on weekends, hash daily, skunk less than once per week, skunk on weekends, skunk daily
cannabis_type	Cannabis type: never used, hash or skunk
duration	Cannabis use duration: 0 to 41 (months)

Our paper proposes a novel data-driven computational psychiatry and machine learning approach to developing predictive models for the onset of first-episode psychosis, based on feed-forward artificial neural networks. The performance capabilities of the predictive models are enhanced and evaluated by a methodology consisting of model optimisation and testing, which integrates a phase of model tuning, a phase of model post-processing with ROC optimisation based on maximum accuracy, Youden and top-left methods, and a model evaluation with the k-fold cross-testing methodology. We further extend our framework by investigating the cannabis use attributes' predictive power, and demonstrating statistically that their presence in the dataset enhances the prediction performance of the neural network models. Finally, the model stability is explored via simulations with 1000 repetitions of the model building and evaluation experiments. The results show that our best models' accuracies in predicting first-episode psychosis in intensive Monte Carlo simulation fall between 75.03% and 85.13%, with an average of about 81%.

The rest of the paper is organised as follows. Section 2 presents our methodology for predicting the first-episode psychosis, based on experimenting with artificial neural networks, and our novel methodology for model optimisation and post-processing, and evaluation with optimized cut-off point selection on the ROC curve. The section also investigates the outcomes of the extensive Monte Carlo simulations in order to study the variation of the models' performance. In Sect. 3, we build optimised prediction models without the cannabis attributes to study if there is a statistically significant

difference with respect to the performances of the models using the cannabis attributes. Finally, the conclusion and the directions for future work are presented in Sect. 4.

2   Building Prediction Models

2.1   Data Preparation

The data we used to build our predictive models were a part of a case-control study [3]. The clinical data comprise 1106 records divided into 489 patients, 370 controls and 247 unlabelled records. The patients were individuals who presented with first-episode psychosis to the inpatient units of the South London & Maudsley Mental Health National Health Service (NHS) Foundation Trust. The controls were healthy people recruited from the same area served by the Trust. The control samples were similar to the patient samples in age, gender, ethnicity, educational qualifications and employment status.

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

The goal of this stage is to perform a high-level simplification of the dataset and to prepare the dataset for use in our novel approach to predict first-episode psychosis. This stage involved several steps. First, records that were missing critical data were removed from the dataset. This included both records with missing labels and records with missing values on all cannabis-related variables. Second, certain variables were removed from the dataset. This primarily included variables that were deemed to be irrelevant to the study (e.g. those related to the individual IDs of the study participants) and variables that fell outside the scope of the study (e.g. certain gene-related variables). In addition, any numeric predictors with zero or near-zero variance were dropped. Third, we sought to standardise the encoding of missing values across the dataset. Prior to this step, values including 66, 99 and –99 all represented cases with missing values; thus, we replaced all such indicators with a consistent missing value indicator: NA. Fourth, some variables were re-labelled to provide more intuitive descriptions of the data they contained. In multiple situations, some variables had similar meanings but also records with missing values. Therefore, we conducted an imputation process to effectively combine the information from all the related variables into one. For example, two variables described alcohol use, but were inconsistently present and presented missing values across the records. These were combined in a way that created one single variable with consistent values that were as complete as possible. This process was used to generate value-reacher and value-consistent variables related to alcohol use, tobacco use, employment history and age.

Finally, any attribute that had more than 50% missing values was removed from the study. We then dropped any record for which more than 70% of the remaining attributes contained missing values. The resulting dataset, after the transformations above, contained 783 records and 78 attributes. The records are divided into 451 patients and 332 controls. A summary of some of these fields—specifically, those that relate to cannabis use, such as type, the age of first use and duration—can be seen in Table 1.

2.2 Missing Values Treatment

Although the data set was pre-processed and attributes with more than 50% missing values were removed, the final dataset still contained several missing values. Of the 783 records, only 22.8% were complete cases. This volume of missing information makes modelling more challenging but is often the reality of medical and social research.

The predictive power of the data may depend significantly on the way that missing values are treated. Some machine learning algorithms, such as decision trees [7], have the capability to handle missing data outright. However, most of machine learning algorithms do not have the capability to handle missing data. In many situations, missing values are imputed using a supervised learning technique, such as k-Nearest Neighbour (KNN. These imputation techniques do not have theoretical formulations but are often applied in practice [4, 6]. Several imputations techniques, such as the KNN imputation, the tree bagging imputation from the *caret* package [7] and the random forest imputation from the *randomForest* package [13] were considered in this work. The last method, although it was the most computationally expensive, produced the best results regarding the performance of the final predictive models.

2.3 Training and Tuning Feed-Forward Artificial Neural Networks

To develop optimised predictive models for first-episode psychosis, we controlled the values of the parameters for each of the considered algorithms using chosen grids. Predictive models have been fitted in a five-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 single-layer and multi-hidden-layer neural networks were optimised (tuned) to maximise AUC, the area under the ROC curve.

To avoid overfitting, the single-layer neural networks were tuned over 10 values for the size (i.e. the number of hidden units) and 10 values for the decay (i.e. the weight decay), which is the parameter in the penalisation method for model regularisation. This approach is like the penalisation method in ridge regression and is based on the L2 norm [7]. The optimal values were 17 and 0.01, respectively. Multi-layer neural networks were tuned over 10 values for each of the three hidden layers (i.e. 10 values for the number of hidden units in each layer) and over 10 values for the decay. The optimal values were 10, 10 and 10 for the three layers and 0.001 for the decay.

2.4   Treating Unbalanced Classes

When there is a priori knowledge of a class imbalance, one direct method to reduce the imbalance's influence on model training is to select training set samples with roughly equal event rates [7]. Treating data imbalances usually leads to better prediction models and a better trade-off between sensitivity and specificity.

In this study, we considered three sampling approaches to sub-sample the training data in a manner that mitigated the imbalance problem. The first approach was down-sampling, in which we sampled (without replacement) the majority class to be the same size as the minority class. The second method was up-sampling, in which we sampled (with replacement) the minority class to be the same size as the majority class. The last approach was the synthetic minority over-sampling technique (SMOTE) [14]. SMOTE selects a data point randomly from the minority class, determines the K nearest neighbours to that point and then uses these neighbours to generate new synthetic data points using slight alterations. Our analysis used five neighbours. The results show that the up-sampling procedure yielded no real improvement in the AUC or the accuracy performances. A simple down-sampling of the data also had no positive effect on the model performances. However, SMOTE with neural networks models led to an increase in both the AUC and the accuracy.

As mentioned, data balancing supports a good trade-off between sensitivity and specificity. Another method that helps to balance sensitivity and specificity, or a good trade-off between the two performances, is model post-processing through the determination of new cut-off points on the ROC curves [7]. Our framework used three such methods, which can be seen as post-processing optimisations of the models. The first method found the point on the ROC curve closest to the top-left corner of the ROC plot, which represents the perfect model (100% sensitivity and 100% specificity). The second method is Youden's J index [15], which corresponds to the point on the ROC curve farthest from the main diagonal of the ROC plot. The third method, which is "maximum accuracy" found the cut-off, which is the point with the highest accuracy.

In order to further improve the model performance, a specially designed post-processing procedure and model evaluation were adapted in our modelling procedure. First, the dataset is stratified splattered randomly into in 60% training data and 40% evaluation data. Then, the training data is used for training and for optimising the model, as explained in Subsect. 2.3, 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 the training data was then applied to the evaluation dataset in a specially designed post-processing procedure, which we call the *k-fold cross-testing method*.

In the *k-fold cross-testing method*, we produce k post-processed model variants of the original optimised model. First, we create k stratified folds of the evaluation dataset. Then, k-1 folds are used to find an alternative probability cut-off on the ROC curve with one of the three specific methods presented above (top-left, Youden, and largest accuracy), obtaining a post-processed model variant. The remaining one-fold is scored with the post-processed model variant 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 evaluation dataset. We note here as an essential 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.

2.5   Increasing Model Performance via Optimized Cut-off Point Selection on the ROC Curve

The ROC curve is a graphical technique for evaluating the ability of a prediction model to discriminate between two classes, such as patients and controls. ROC curves allow visual analyses of the trade-offs between a predictive model's sensitivity and specificity regarding various probability cut-offs. The curve is obtained by measuring the sensitivity and specificity of the predictive model at every cutting point and plotting the sensitivity against 1-specificity. The left image in Fig. 1 shows the ROC curves obtained for both the single-layer neural networks and the multi-layer neural networks. The curve shows that multi-layer neural network performs better regarding the evaluation dataset.
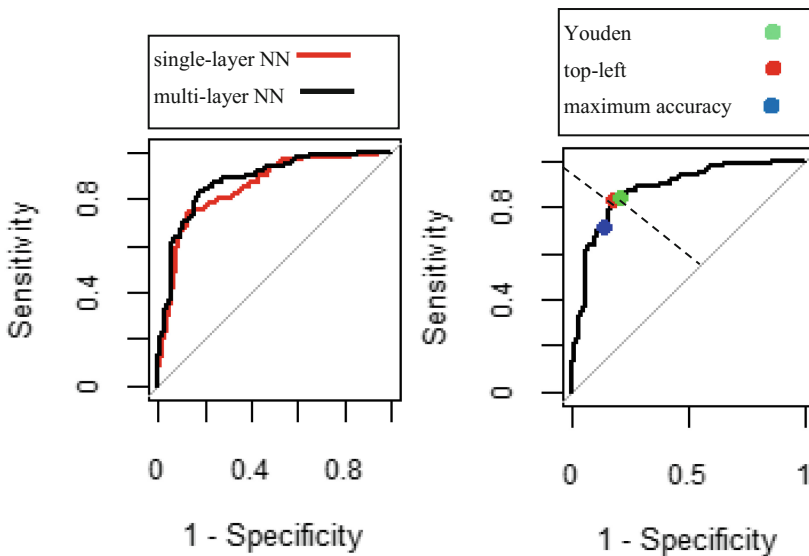

Fig. 1. Left: The ROC curves for 2 of our optimised neural networks (NN) models: single-layer NN and multi-layer NN. Right: ROC optimisation post-processing of the multi-layer NN model, with 3 optimal cutting points: maximum accuracy, Youden and top-left methods. (Color figure online)

Numerous methods exist for finding a new cut-off. First, one can find the point on the ROC curve that is closest to the perfect model (100% sensitivity and 100% specificity), which is the point with shortest distance value from the point (0,1) as

shown in the left image in Fig. 1. To find the shortest distance, $[(1 - sensitivity)^2 + (1 - specificity)^2]$ was calculated and minimised [16]. Another approach for finding an optimal cut-off point on the ROC curve is finding the largest distance from the diagonal to the ROC curve as shown in the right image in Fig. 1. This is the point with the largest value for the Youden index which is defined as (sensitivity + specificity -1) [15]. These are the two most popular methods for establishing the optimal cut-off [7, 17]. We used both of these methods, as well as the maximum accuracy approach, which determines the point on the ROC curve corresponding to the greatest accuracy (the red point in Fig. 1, right). In our analysis, the optimal cutting point was derived from independent sets, rather than from the training set or the evaluation sets, as shown previously. This is particularly important, especially, for small datasets.

2.6   Monte Carlo and Models' Stability

Due to the uncertainties introduced by the missing values in the data and due to expected variations of the predictive models' performance, depending on the datasets that were chosen for training and testing, we perform extensive Monte Carlo simulations to study the performances' variations and the models' stability. The simulations for each NN consisted of 1,000 iterations of the proposed procedure. The models' performances concerning accuracy, sensitivity, specificity and kappa were evaluated for each iteration on separate a testing dataset. The aggregation of all iterations yielded various distributions of the concerned performance measures. These distributions were then visualised using box plots in Fig. 2 to capture the models' performance capability and stability. The subfigures in Fig. 2 were grouped by their performances' measures into four subfigures. Each subfigure contained sex box plots for single-layer and multi-layer neural networks with several optimized cut-off points on the ROC curve such as top-left, Youden index and the maximum accuracy. Also, estimations of the predictive neural networks' performances regarding means and standard deviation (SD) are shown in Table 2. The results as shown in Table 2 are regarding the models' performances when applied with the ROC optimisation techniques.

On the one hand, single-layer neural networks with ROC optimisations based on Youden or top-left scored a mean accuracy of 0.76. The performance of the single-layer neural networks slightly improved when a ROC optimisation based on maximum accuracy was applied, resulting in a mean accuracy of 0.79 (95% CI [0.75, 0.83]) and a mean sensitivity of 0.83 (95% CI [0.76, 0.9]). On the other hand, the multi-layer neural networks with an ROC optimisation based on top-left cutting point achieved the best results with a mean accuracy of 0.81 (95% CI [0.77, 0.86]) and a mean sensitivity of 0.84 (95% CI [0.77, 0.88]), comparable to the results achieved by multi-layer neural networks with Youden and maximum accuracy.

This procedure is very computationally costly; therefore, a robust framework was essential. Parallel processing was performed on a data analytics cluster of 11 servers with Xeon processors and 832 GB fast RAM. We used the R software with some packages, including *caret*, *pROC*, *e1071*, *randomForest*, *ggplot2*, *plyr*, *DMwR*, *Applied Predictive Modeling* and *doParallel* [18].

In general, we detect that the proposed models have good predictive power and stability, based on an acceptable level of variation in their performance measures
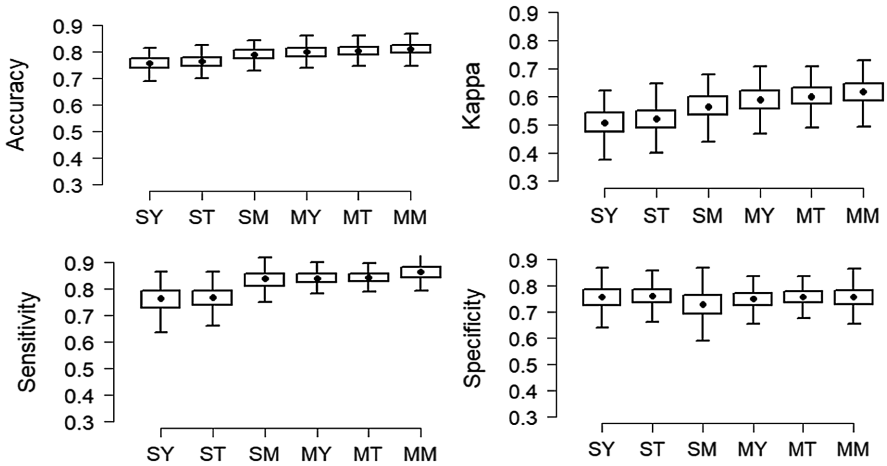
Fig. 2. 1000 Monte Carlo simulation for artificial neural networks. Where "SY" is single-layer NN with Youden cutting point, "ST" is single-layer NN with top-left cutting point, SM" is single-layer NN with maximum accuracy cutting point, "MY" is multi-layer NN with Youden cutting point, "MT" is multi-layer NN with top-left cutting point, and "MM" is multi-layer NN with maximum accuracy cutting point.

Table 2. Estimations of the predictive neural networks' performances.

Model	Accuracy		Kappa		Sensitivity		Specificity	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Single-layers NN (Youden)	0.76	0.03	0.52	0.05	0.76	0.05	0.75	0.05
Single-layers NN (top-left)	0.76	0.02	0.52	0.05	0.77	0.04	0.76	0.04
Single-layers NN (max accuracy)	0.79	0.02	0.56	0.05	0.83	0.04	0.73	0.05
Multi-layers NN (Youden)	0.8	0.02	0.59	0.05	0.84	0.02	0.75	0.04
Multi-layers NN (top-left)	0.81	0.02	0.61	0.04	0.84	0.02	0.76	0.03
Multi-layers NN (max accuracy)	0.80	0.08	0.59	0.13	0.87	0.04	0.74	0.08

evaluated across extensive Monte Carlo experiments. However, the results indicate that the performance differences between the two types of neural networks with different methods for selecting the ROC cutting points are not significant regarding the 4 performances.

3   Cannabis Use Attributes' Predictive Power

After performing the Monte Carlo simulations, we further investigated the predictive models in order to better comprehend the predictive power of the cannabis-related attributes over first-episode psychosis. Moreover, we investigated the association

between cannabis-related attributes and first-episode psychosis via statistical tests and attribute-ranking techniques.

3.1   Student's t-Test

In this subsection, we comprehend the predictive power of the cannabis-related attributes over first-episode psychosis via statistical tests by re-fitting our performing models but with the cannabis-related attributes, represented in Table 1, removed from the dataset. Then, we compared the performances with and without the cannabis-related attributes using Student's t-test. We thereby demonstrated the predictive value of cannabis-related attributes with respect to first-episode psychosis by showing that there is a statistically significant difference between the performances of the predictive models built with and without the cannabis variables.

Our analysis also showed that the accuracy decreased by 5% for single-layer neural networks and by 6% for the multi-neural networks, if the cannabis-related attributes were removed from the process of building the predictive models. Then, we compared the accuracies of the single-layer neural networks models built on the data sets with and without the cannabis-related attributes using one-tailed t-test. The p-value obtained for the t-test was 5.51×10^{-195}. As for the multi-layer neural networks models built on the data sets with and without the cannabis use attributes, the p-value obtained for the one-tailed t-test was and 2×10^{-16}. This means that the predictive models with cannabis attributes have higher predictive accuracy than the models that were built without the cannabis attributes. In other words, the additional cannabis variables jointly account for predictive information over first-episode psychosis. These results are consistent with findings from [4].

3.2   Ranking Attributes' Importance with the ROC Curve Approach

This subsection proposes the use of the ROC curves to determine the relevant variables affecting first-episode psychosis as introduced in [7]. We measure the individual importance of every attributes in the dataset to discover the attributes that yield significant improvements in the model predictivity power. To do so, the ROC curve is considered on each attribute. Then, a series of cutoffs is applied to the data to predict the class. The sensitivity and specificity are calculated for each cutoff, and the ROC curve is computed. Finally, the area under the curve is used as a measure of variable importance. Table 3 shows the top 10 attributes ranked by the ROC curve approach.

The results in Table 3 support prior evidence that cannabis attributes, such as the type of the cannabis used and the frequency of usage, have significant power in predicting first-episode psychosis. For example, the results in Table 3 support findings from [3] by associating the type of cannabis, especially high-potency cannabis, with the onset of psychosis. In addition, *bindur.3* in Table 3, which represents the duration of cannabis use, is consistent with findings from [4].

Table 3. ROC curve attribute importance

Attribute	Importance
typefreq. 1	100.00
cantype2.1	90.38
type_use.hash	88.53
totfreq. 1	87.74
bindur.3	85.84
education.university.professiolqualifications	85.55
frequenc.2	84.24
bullying.no	78.55
white	72.89
homeless.1	72.86

4   Conclusion and Future Work

This paper proposes a novel machine learning approach to developing predictive models for the onset of first-episode psychosis using artificial neural networks. We explored two types of artificial neural networks, each of which was able to recognise patterns differentiating patients from controls at an acceptable level. We based our approach on a novel methodology for optimising and post-processing predictive models. We also proposed several sampling methods and several methods for choosing the optimal cutting point on the ROC curve to improve the prediction models' performances. The models were then further tested using Monte Carlo experiments, and they consistently yielded adequate predictive power and stability.

The best-performing models were multi-layer neural networks, which achieved accuracies as high as 88% in some cases and an average accuracy of 81% in Monte Carlo simulations with 1000 repetitions. The scored performances were above all performances achieved in previous studies such as [3]. This paper extends on previous work as [3] by proposing a new machine learning framework based on a novel methodology in which models are post-processed based on optimized cut-off point selection on the ROC curve and evaluated with the recent method of k-fold cross testing which we adapt after [8]. Moreover, in this new framework, we developed optimized models with other powerful techniques such as artificial neural networks not addressed in [3]. Also, the predictive power of cannabis-use attributes was tested via statistical tests and ranking methods to demonstrate statistically that their presence in the dataset enhances the prediction performance of the neural network models. The proposed approach proves the high potential applicability of machine learning and, particularly, artificial neural networks in psychiatry and enables researchers and doctors to predict and evaluate risks for first-episode psychosis.

One possible direction for future work is to further investigate how this prediction performance variation evolves by limiting the uncertainty in the data, represented by the high proportion of missing values. The second possible work direction is redefining the predictive modelling approach by considering more high-dimensionality data, such as genotype data. A third future work direction, which we are currently investigating,

involves enhancing the power to predict first-episode psychosis using deep learning approaches.

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. Front. Psychiatry **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. Lancet Psychiatry **2**(3), 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. In: 15th IEEE International Conference on Machine Learning and Applications, pp. 825–830 (2016)
5. Zhou, H., Tang, J., Zheng, H.: Machine learning for medical applications. Sci. World J., **20**, 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. Kuhn, M., Johnson, K.: Applied Predictive Modelling. Springer (2013)
8. Katrinecz, A., Stamate, D., et al.: Predicting psychosis using the experience sampling method with mobile apps. In: 16th IEEE International Conference on Machine Learning and Applications (2017)
9. Belgrave, D., Cassidy, R., Stamate, D., et al.: Predictive modelling strategies to understand heterogeneous manifestations of asthma in early life. In: 16th IEEE International Conference on Machine Learning and Applications (2017)
10. Elveren, E., Yumuşak, N.: Tuberculosis disease diagnosis using artificial neural network trained with genetic algorithm. J. Med. Syst. **35**, 329–332 (2011)
11. Adjouadi, M., Ayala, M., et al.: Classification of leukaemia blood samples using neural networks. Ann. Biomed. Eng. **38**(4), 1473–1482 (2010)
12. Yan, Y., Qin, X., et al.: A restricted Boltzmann machine based two-lead electrocardiography classification. In: Proceedings of 12th International Conference on Wearable Implantable Body Sensor Networks (2015)
13. Liaw, A., Wiener, M.: Classification and regression by random forest. R News **2**(3), 18–22 (2002)
14. Qazi, N., Raza, K.: Effect of feature selection, SMOTE and under sampling on class imbalance classification. In: 2012 UKSim 14th International Conference on Computer Modelling and Simulation, pp. 145–150 (2012)
15. Bohning, D., Bohning, W., et al.: Revisiting Youden's index as a useful measure of the misclassification error in a meta-analysis of diagnostic studies. Stat. Methods Med. Res. **17**, 543–554 (2008)
16. Pepe, M.: The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press, New York (2003)
17. Perkins, N., Schisterman, F.: The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. Am. J. Epidemiol. **163**(7), 670–675 (2006)
18. Cran.r-project.org.: The Comprehensive R Archive Network (2018). https://cran.r-project.org/. Accessed 2 Jan 2018