

Deep Neural Network to Identify Patients with Alcohol Use Disorder

Ali EBRAHIMI^{a,1}, Uffe Kock WIL^a, Marjan MANSOURVAR^a, Amin NAEMI^a,
Kjeld ANDERSEN^b and Anette Søgaard NIELSEN^b

^aSDU Health Informatics and Technology, The Maersk Mc-Kinney Møller Institute,
University of Southern Denmark, Odense, Denmark

^bUnit for Clinical Alcohol Research, Clinical Institute, University of Southern
Denmark, Odense, Denmark

Abstract. This paper presents an application of deep neural networks (DNN) to identify patients with Alcohol Use Disorder based on historical electronic health records. Our methodology consists of four stages including data collection, preprocessing, predictive model development, and validation. Data are collected from two sources and labeled into three classes including Normal, Hazardous, and Harmful drinkers. Moreover, problems such as imbalanced classes, noise, and categorical variables were handled. A four-layer fully-connected feedforward DNN architecture was designed and developed to predict Normal, Hazardous, and Harmful drinkers. Results show that our proposed method could successfully classify about 96%, 82%, and 89% of Normal, Hazardous, and Harmful drinkers, respectively, which is better than classical machine learning approaches.

Keywords. Alcohol Use Disorder, Deep Neural Networks, Multiclass Classification.

1. Introduction

Alcohol Use Disorder (AUD) has a high prevalence rate and is related to many clinical conditions, which makes it one of the main causes of death [1] with almost 3.3 million deaths worldwide related to AUD [2]. In some Western societies, like Denmark, this rate is around 5.5% of recorded deaths every year related to AUD [3]. Several studies have been conducted related to detection and prediction of patients with AUD using traditional machine learning (ML) and deep learning methods based on Electronic Health Records (EHRs). Ngo et al. [4] developed a risk predictive model using Linear Regression (LR) based on an electronic database consisting of student's enrollment and their medical records. In their study, they considered a filter feature selection method as the feature engineering task to reduce the number of risk factors. Wang et al. [5] proposed a 10-layer Convolutional Neural Network (CNN) to detect automatically alcoholic and nonalcoholic patients from brain Magnetic Resonance Imaging (MRI). In our previous work [6], we proposed a predictive ML method using Support Vector Machine (SVM) to classify patients as either AUD positive or AUD negative. In that study, we applied a filter feature selection method to reduce the number of features, and a supervised ML algorithm to construct the predictive model.

¹ Corresponding Author, Ali Ebrahimi; E-mail: aleb@mmmi.sdu.dk.

Although previous studies in prediction of patients with AUD achieved their goals successfully, several factors distinguish their work from the current study. One of the main challenges of employing traditional ML methods such as SVM and LR is the feature reduction task, which needs to be conducted by human experts [4, 6]. Moreover, all the mentioned studies on AUD use binary classification approaches. To the best of our knowledge, prediction of patients with AUD based on their historical EHRs through deep learning algorithms to classify patients into several classes has not yet been explored. To address this gap, this study develops an automated multi-class classification model using a Deep Neural Networks (DNNs) algorithm to classify historical EHRs into Normal, Hazardous, and Harmful drinkers. AUD patient classification into three classes will enhance treatment by assisting medical professionals to improve the prognosis of patients by using previously stored data in EHRs.

2. Method

2.1. Data collection and Preprocessing

The data collection for this study was based on a mixture of two sources, the Relay Study [7] and EHRs from Odense University Hospital (OUH), Region of Southern Denmark. In the Relay Study, alcohol relevant data were obtained from 2,571 patients admitted to OUH from Oct. 2013 to June 2016. The EHR dataset covers 13,648 records of the Relay participants' prior visits to OUH. The dataset contains clinical records including Personal Identification Number, age, gender, length of stay, hospital department where the patient belonged to, hospital department where the patient received treatment, admission type, attending the intensive care unit (ICU), transfer from the emergency department (ED), ICD treatments, cares, operations, and more importantly diagnostic codes (Danish version of ICD-10), and health-related conditions. More detail on the dataset can be found in [6]. Based on the result of the Alcohol Use Disorders Identification Test [8], and according to Babor et al. [9], patients were divided into three groups, 2,114 Normal, 148 Hazardous, and 309 Harmful drinkers. This categorization was used to label the collected EHR data from OUH as the target value for training the predictive models.

Among the 13,648 records in the EHR dataset, 11,352, 1,490, and 806 records belonged to Normal, Hazardous, and Harmful drinkers, respectively. Thus, imbalanced classed was one of the major challenges in the preprocessing step. Synthetic Minority Oversampling (SMOTE) [10] is a well-known technique to overcome this problem. SMOTE is an oversampling technique, which increases the number of samples in minority classes by producing synthetic samples. However, because some majority class samples might be invading minority class space, class clusters in this technique might not be well defined [11]. To handle this problem, Batista et al. [11] combined the Wilson's Edited Nearest Neighbor Rule (ENN) [12] with SMOTE, introducing SMOTEENN method. ENN removes majority class samples by dropping any samples whose class label vary from the class of at least two of its three nearest neighbors [11]. By using SMOTEENN, besides balancing the training data, noisy samples appearing on the wrong side of the decision border will be removed. Since in this study, we do not need to go through feature reduction process, SMOTEENN would help us reach an excellent predictive accuracy. Moreover, categorical data in our dataset was encoded into numerical values, via a one-hot-encoding technique, which converts a single variable with n observations and d distinct values, to d binary features with n observations each.

Additionally, the standard normalization technique was used to achieve a common scale for all variables. Finally, we ended up with a dataset consisting of 857 features including primary diagnosis (851 different ICD codes), gender, age (18-110 years), length of stay, admission type (inpatient or outpatient), ICU stay (yes or no), and transfer from ED (yes or no).

2.2. Model Development and Evaluation

By implementing a proper Neural Network (NN) architecture, like the selection of activation function, or the number of hidden layers and neurons in each layer, a deep non-linear relation between input variables and output can be modeled. However, there is no straightforward NN architecture for building a predictive model; it is developed through several experimental iterations. This study uses a feedforward artificial NN architecture for prediction of patients with AUD. This architecture consists of four layers of nodes including, input layer, hidden layers, and output layer in which each layer has a specific function. Each hidden and output layer is comprised of a neuron using a non-linear activation function, which demonstrates the complexity of the predictive model in terms of the relationship between features as input variables and target values as output.

For this classification task, the Mean Square Error (MSE) as the loss function was employed and it is calculated as: $MSE(y, \hat{Y}) = \frac{1}{n} \sum (y_i - \hat{Y}_i)^2$, where \hat{Y} is the actual value, y is predicted value, and n is number of data points. Adaptive Moment Estimation (Adam) was employed as an optimization algorithm to update network weights. The main advantages of using Adam in DNN tasks are the easy implementation, accelerated training, and adequacy for non-stationary or noisy objectives. A Rectified Linear Unit (ReLU) was employed as the activation function which is defined as: $ReLU(\hat{Y}) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$, where x is input data. As this is an issue of multiclass classification with three output labels, including Normal, Hazardous, and Harmful drinkers, we had to normalize the output of the ReLU activation function and convert them into probabilities and then compare them with the class labels, which in this case is in the interval of 0, 1, and 2. To overcome this problem, Softmax activation function was employed which is calculated as: $Softmax(\hat{Y}_i) = \frac{e^{y_i}}{\sum_{i=0}^n e^{y_i}}$.

In summary, the proposed DNN model for this study is a feedforward architecture model composed of four fully connected layers. The ReLU activation function is chosen for the first three layers (hidden layers), and Softmax activation function for the last layer (output layer). The input layer had 857 neurons, equivalent to the number of features, the

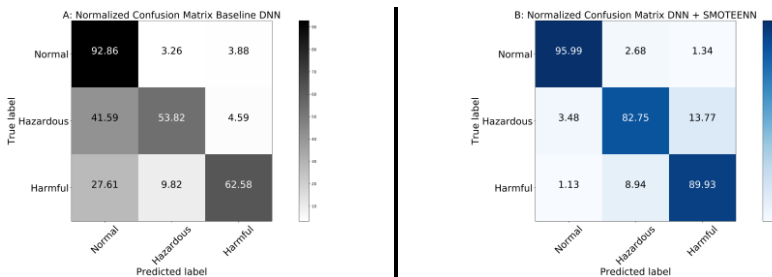


Figure 1. A: Normalized Confusion Matrix of Baseline DNN, B: Normalized Confusion Matrix of DNN based on SMOTEENN.

first, second, and third layers had 128, 64, and 16 neurons, respectively, and the output layer had three neurons. This was done with the Adam optimizer with 1000 epochs. Two models with same architecture were developed, one based on baseline dataset, the other benefitted from SMOTEENN. The performance of the proposed methods were evaluated according to precision, recall, and predictive accuracy (the usual metrics to measure the performance of learning models). Moreover, the receiver operating characteristics curve (ROC) and the area under the ROC curve (AUC) are used to evaluate the classification performance, which in multiclass classification problems, each curve evaluates the target class in comparison to all other classes. Additionally, macro- and micro-average curves are calculated and plotted. The values of these performance metrics can be computed by finding values for true positive (TP), false positive (FP), false negative (FN), and true negative (TN) [13]. The performances of the classifiers were evaluated using 5-fold cross-validation.

3. Results and Discussion

In terms of the classification approach, we had to deal with the multiclass classification task of predicting Normal, Hazardous, and Harmful drinkers. A four-layer DNN algorithm using a feedforward architecture was designed and developed using 857 EHR features. We ended up with 27,560 clinical records after handling the class imbalance in the dataset by SMOTEENN. This helped us on noise removal from Normal drinkers' class and decrease it to 6,796 records, while increasing records of Hazardous and Harmful drinkers' classes to 10,106 and 10,657 records, respectively. Figure 1a displays the normalized confusion matrix for the proposed DNN method based on the baseline dataset and Figure 1b display the same based on a SMOTEENN balanced dataset. As shown, all classes underwent an improvement by using SMOTEENN. The Normal drinker group by noise removal, and the other two groups by an increasing the number of samples. The ROC curves of both models are displayed in Figures 2a and 2b. As shown, improvements among all classes are evident. The AUCs for the micro and macro-average are 0.97 on the test set, conferring an excellent classification performance for our proposed DNN method based on SMOTEENN according to [14].

The construction of a successful DNN architecture for a given problem is not negligible. There seems to be no general way to decide the right number of neurons and layers for the DNN architecture, but the ideal configuration is generally tested by experimentation or familiarity with common challenges. However, our results are affected by the inherent lack of Harmful and Hazardous drinkers as well as the number of EHR for the training set, which resulted in high FN rates among Harmful and

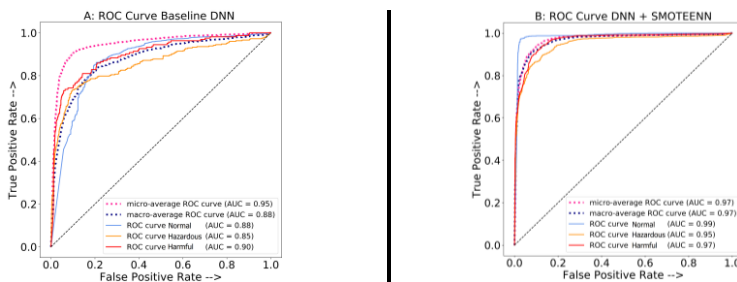


Figure 2. A: ROC Curve of Baseline DNN, B: ROC Curve of DNN based on SMOTEENN.

Hazardous classes. One of the main limitations of this study is the population of the dataset, which consists of patients from only three hospital departments, which may raise the risk of bias. Therefore, in our future study, we plan to validate our models based on a national Danish dataset with millions of records.

4. Conclusion

This study focused on the design and development of a predictive model for AUD. We proposed a feedforward DNN architecture consisting of four fully connected layers including two ReLU and a Softmax activation function to identify patients with AUD in three classes including Normal, Hazardous, and Harmful drinkers. Our previous work based on classical ML approaches resulted in a predictive accuracy of 80% [6]. Compared to this, the presented DNN approach with an overall predictive accuracy of 88% achieved better performance. This demonstrates the potential of deep learning in the prediction of AUD based on historical EHRs.

References

- [1] Laramée P, Leonard S, Buchanan-Hughes A, Warnakula S, Daeppen J-B, Rehm J. Risk of all-cause mortality in alcohol-dependent individuals: a systematic literature review and meta-analysis. *EBioMedicine*. 2015;2(10):1394-1404.
- [2] Ghani AMA, Faiz E, Nielsen A, Bilberg R. What is the cause of death, when alcohol dependent persons die prematurely? Drug and alcohol dependence. 2019;197:120-126.
- [3] Gottlieb Hansen AB, Hvidtfeldt UA, Grønbaek M, Becker U, Søgaard Nielsen A, Schurmann Tolstrup J. The number of persons with alcohol problems in the Danish population. *Scandinavian Journal of Social Medicine*. 2011;39(2):128-136.
- [4] Ngo DA, Rege SV, Ait-Daoud N, Holstege CP. Development and validation of a risk predictive model for student harmful drinking-A longitudinal data linkage study. *Drug and Alcohol Dependence*. 2019 Apr;197:102-107. doi:10.1016/j.drugalcdep.2019.01.016. Cited in: Pubmed; PMID WOS:0004644790000017.
- [5] Wang S-H, Muhammad K, Hong J, Sangaiah AK, Zhang Y-D. Alcoholism identification via convolutional neural network based on parametric ReLU, dropout, and batch normalization. *Neural Computing and Applications*. 2020;32(3):665-680.
- [6] Ebrahimi A, Wiil UK, Andersen K, Mansourvar M, Nielsen AS. A Predictive Machine Learning Model to Determine Alcohol Use Disorder. 2020 IEEE Symposium on Computers and Communications (ISCC); 2020. IEEE; 1-7 p.
- [7] Schwarz A-S, Nielsen B, Søgaard J, Nielsen AS. Making a bridge between general hospital and specialised community-based treatment for alcohol use disorder—A pragmatic randomised controlled trial. *Drug and alcohol dependence*. 2019;196:51-56.
- [8] Saunders JB, Aasland OG, Babor TF, De la Fuente JR, Grant M. Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption - II. *Addiction*. 1993;88(6):791-804.
- [9] Babor TF, Higgins-Biddle JC, Saunders JB, Monteiro MG. The alcohol use disorders identification test. Guidelines for use in primary care. 2001;2:1-41.
- [10] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*. 2002;16:321-357.
- [11] Batista GE, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter*. 2004;6(1):20-29.
- [12] Wilson DL. Asymptotic properties of nearest neighbor rules using edited data. *IEEE Transactions on Systems, Man, and Cybernetics*. 1972;(3):408-421.
- [13] Baeza-Yates R, Ribeiro-Neto B. Modern information retrieval. Vol. 463. ACM press New York; 1999.
- [14] Safari S, Baratloo A, Elfil M, Negida A. Evidence based emergency medicine; part 5 receiver operating curve and area under the curve. *Emergency*. 2016;4(2):111.