

# Using Edge AI and Language Understanding for Predictive Modeling of Acute Medical Intoxications

Strahil Sokolov <sup>1\*</sup> and Stanislava Georgieva <sup>2</sup>

<sup>1</sup> University of Telecommunications and Post/Information Technologies, Sofia, Bulgaria,

<sup>2</sup> Medical university of Varna/Toxicology, Varna, Bulgaria,

\*E-mail of corresponding author: strahil.sokolov@gmail.com

**Abstract:** This paper presents a new approach to processing and categorization of text from patient documents in Bulgarian language using Natural Language Processing and Edge AI. The proposed algorithm contains several phases - personal data anonymization, pre-processing and conversion of text to vectors, model training and recognition. The experimental results in terms of achieved accuracy are comparable with modern approaches.

**Received** 28 November 2021

**Accepted** 17 December 2021

**Published** 22 December 2021

**Keywords:** NLP, Edge AI, medical documents, medicinal intoxication.

## 1. Introduction

Poisoning is a significant public health problem. According to WHO data, in 2016 more than 1,206,000 people died as a result of accidental poisoning. Of these deaths, 84% occurred in low- and middle-income countries. The causes of poisoning are many: human error, industrial, intentional and accidental. The most common toxic agents worldwide are pesticides, drugs (sedatives, painkillers, antidepressants, etc.), chemicals (acids and copper sulfate), alcohol, plant toxins, and household chemicals. There are significant differences in patterns of poisoning between different countries. In developed countries, the annual prevalence of accidental or deliberate poisoning cases in humans varies between 0.2 - 9.3 exposures to poisoning per 1 000 people and continues to increase annually worldwide as case reporting and registration increases [1].

Acute poisoning is a condition of damage whose toxic effects occur almost immediately, within hours of exposure. It is a common cause worldwide for seeking emergency medical care and hospitalization, and its morbidity and mortality is becoming a major public health problem in many countries. The World Health Organization (WHO) estimates the total number of acute accidental poisonings worldwide at 2-3 million cases per year, of which 1 million severe poisonings, leading to 20,000 deaths per year. The estimated number of intentional poisonings is about 2 million, of which approximately 200,000 are suicides.

The high frequency of accidental or intentional poisonings with drugs from different groups are a prerequisite for the development of specific programs for diagnosis, treatment and prevention of poisoning in different countries. The main elements of these programs are the identification of toxic hazards that exist at the

local level (in order to establish preventive measures), the diagnosis of poisoning and the treatment of patients. Special attention should be paid to the relative share of poisonings to the total number of suicides. The incidence of intentional ingestion of suicide drugs is particularly high in the Nordic countries and the United Kingdom, and is relatively lower in Eastern Europe and Central and South America.

Data from the World Health Organization (WHO) show that the total number of acute accidental poisonings worldwide varies from 2 to 3 million cases per year, of which 1 million are severe poisonings, leading to 20,000 deaths per year, while the estimated annual number of intentional poisonings is about 2 million as a result of 200,000 suicides. It should be noted that patients with acute poisoning represent 15-20% of the total number of patients who are hospitalized annually. The exact number of cases may be higher, as most cases of poisoning actually go unreported. In Western Europe, twice as many patients are hospitalized for acute poisoning than for myocardial infarction, and the mortality rate for this pathology is higher than for infectious diseases and traffic accidents. Current statistics show an increase in the incidence of acute poisoning over the last decade, in contrast to the share of natural, man-made and socio-political emergencies.

Currently, in developed countries the most common acute exogenous intoxications are drugs (60%), which can be explained by the fact of their easy availability and their presence in every home. Both prescription and over-the-counter drugs are used worldwide for self-poisoning, mainly in urban areas. The most commonly used medicines are drugs which act on the central nervous system, such as antipsychotics, antidepressants, barbiturates and benzodiazepines; followed by analgesics mainly paracetamol, antiepileptic drugs such as carbamazepine, antiseptics and disinfectants, antimalarial drugs such as chloroquine and others.

Morbidity and mortality in each acute case of poisoning depends on a number of factors such as the



nature of the poison, the dose consumed, the availability of medical facilities, treatment by qualified personnel and the time interval between poisoning and the provision of medical care [2].

In recent years, Artificial Intelligence (AI) and Machine Learning (ML) technologies have seen unprecedented growth due to a number of factors. Image recognition, data classification and natural language processing (NLP) technologies are widely used to aid in human speech recognition; creation of robots for automated answering of questions and mood analysis are several of the areas of application [12, 15]. NLP is also one of the tools that serves to enrich the analysis in journalism to unlock additional functionality in the behavior of the system. For example, applications are known that support the word processing of log files to identify common clusters or data patterns [3]. The trend toward building intelligent autonomous systems on embedded hardware is known as Edge AI. There are examples of efficient applications that employ ML in various areas [4, 5].

This paper is organized as follows. Section 2 describes the approach for automated NLP analysis of about 300 medical documents of patients who underwent treatment for acute medicinal intoxications (AMI), 2016-2021, in Bulgarian language. Section 3 contains the proposed approach for building the mobile application. Section 4 contains the experimental results from the practical tests of the algorithms.

## 2. Approach for automated analysis for content extraction from natural language medical documents of AMI patients

This section describes the proposed approach for using NLP to create a mathematical model for assessing the risk of drug intoxication depending on the patient's medicinal products, dietary supplements in combination with or without ethanol. With the great advancement of the Pandas data processing libraries in the Python programming language and the rapid development of NLP libraries for natural language information processing, it has become possible to develop a computer application for the benefit of the patient, doctor and pharmacist. Pandas [7, 11] is an open source library that provides easy-to-use data structures and data analysis tools in the Python programming environment. There are a number of good open source libraries for NLP in Python. The scikit-learn library [9] was chosen because it has a broad set of machine learning tools for efficient data retrieval and word processing.

The practical application of the proposed approach is the possibility of creating a database of combined drug intoxications, which can be supplemented and changed.

### 2.1. Data acquisition and preprocessing

The data that are entered are text from part of the epicrisis, prescription and data from an outpatient list. Roles: Patient - mobile application, electronic file sharing; Doctor: integration with the coordination center system; Pharmacist: integration with a coordination center system. This workflow is in the described structural diagram in fig. 2. Pre-processing of the data

may include checking for spelling errors and searching for n-grams, which can bring normalization of the data at the input.

At the beginning of the proposed algorithm, a step for pre-processing of medical files has been introduced, which is in line with Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of individuals with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46 / EC (General Data Protection Regulation) [6]. Names, home and e-mail addresses, as well as any other information according to the above-mentioned regulation, which would be used to identify the patient, are removed from the medical files [7].

The next step in data preprocessing is to remove words and symbols that are not essential to the model. All numbers or special characters in the processed file are not useful in NLP and must be removed. Usually these words and symbols are part of the library used. This step is necessary due to the very nature of the data. To access this data in Python, an efficient data structure from the Pandas library called DataFrame is used. This is a two-dimensional tabular data structure that can be modified with labeled axes (rows and columns).

### 2.2. Feature extraction

In NLP, one of the most commonly used approaches to word processing is the n-gram generation. An n-gram is a sequence of n words used as an observation reference, for example, a unigram is unambiguously formulated, a bigram is an ambiguous phrase, and a trigram is a phrase formulated by three words. To determine n-grams of patient data, we use the scikit-learn [9, 10] TfidfVectorizer library. Vectorization of text data is the main process of converting a set of text documents into numeric vectors with attributes. A specific strategy (tokenization, counting and normalization) is called Bag-of-words or in the approach used: 'Bag of n-grams'. The presented approach should emphasize how NLP can be used to extract knowledge from documents, which complements the traditional analysis of files. Such knowledge is not only in the models of the types of acute drug poisonings, but can also facilitate the determination of the risk of poisoning depending on the available potential combination of drugs and ethanol in the patient. Because the format of the file available to the patient differs from one system to another, some pre-processing of the data is required. When processed data is available, Pandas can be used to manipulate that data using its highly efficient data structure. NLP can be used to find aggregated data of medical files via WordCloud and n-grams.

### 2.3. Algorithm for automated poisoning risk assessment

Once the classifiers are trained on a sufficient number of patient data, it is possible to create an automated mobile application. It is proposed that the data can be scanned with a mobile phone, processed in advance, sent via an encrypted connection to the cloud infrastructure of a coordination center, classified and transmitted for

feedback to a pharmacist, doctor and patient. An opportunity can be created for further training of the system with the help of new data provided by a doctor, pharmacist or patient.

#### 2.4. Classifier training

The scikit-learn library used in the proposed approach has a module for training classifiers. Since an open source Python library is used that comes freely as an installation package, it is imported directly into the program.

Two models were trained: Naive Bayes classifier (NB) and Support Vector Machines (SVM). NB classifier is a conventional and very popular method for solving the problem of document classification. It represents a supervised probabilistic classifier based on Bayes' theorem, assuming independence between each pair of input characteristics.

SVMs are trained in a supervised way in multiple classes that are known for their high effectiveness. The purpose of SVM is to separate some subset of a training sample of data from the others, called reference vectors (hyperplane separation boundary). The decision function of the SVM model, which predicts the class of test data, is based on support vectors and can be executed in a kernel.

Once the classifiers are trained, we check their accuracy against the test sample document projections. We extract a vector with word characteristics for each document in the test set and the trained classifier predicts its class (potential type of poisoning) with the trained NB classifier and SVM model.

#### 2.5. Training time and inclusion criteria

The toxicological part of the study was conducted in the city of Varna by the author S. Georgieva, using analysis on the medical histories, personal ambulatory cards of poisoned patients admitted for treatment, with the permission of the Chairman of the Ethics Commission at the Military Medical Academy in Varna.

#### 2.6. Overview of the proposed automated system for evaluation of potential acute drug intoxication

Once the classifiers are trained on a sufficient number of anonymized patient data, it is possible to create an automated mobile application. It is proposed that the data can be scanned with a mobile phone, processed in advance, sent via an encrypted connection to existing cloud infrastructure of a Coordination Center, classified and transmitted for feedback to a pharmacist, doctor and patient. An option can be created to further train the system with the help of new data provided by a doctor or patient.

It is necessary for the patient to scan part of the epicrisis / outpatient list with prescribed therapy for home treatment and / or packaging of medicinal products, food supplements, currently used medical devices and data on possible alcohol consumption. This data is shared with the GP or sent to focal point review with a risk assessment request. While being acquired on the mobile device, the data is subjected to an anonymization process.

The process of classifying the risk and providing feedback to the patient comes next. The information obtained can be discussed with a physician or pharmacist.

A doctor receives information from a patient through an application, with a scanned part of the epicrisis / outpatient list with prescribed therapy for home treatment by a specialist, containing the used nutritional supplements. Together with the part of the patient's file containing medicinal products for chronic use, it can be sent to the coordination center for assessment of the risk of the newly combined drug therapy and the risk of intoxication.

The pharmacist has the role of consultant to the patient or physician, and has the ability to annotate or correct a set of new or existing data as needed.

The presented statistical methods and approach for analysis of natural language documents with epicrisis of patients can be used to create a model / s of ADI, specific to a particular region, taking into account their features - qualitative and quantitative.

Modern approaches in the field of artificial intelligence are used for statistical processing and data analysis. The proposed application structure of Figure 1 would find practical realization in the eventual future construction of a Coordination Center for treatment of acute intoxications.

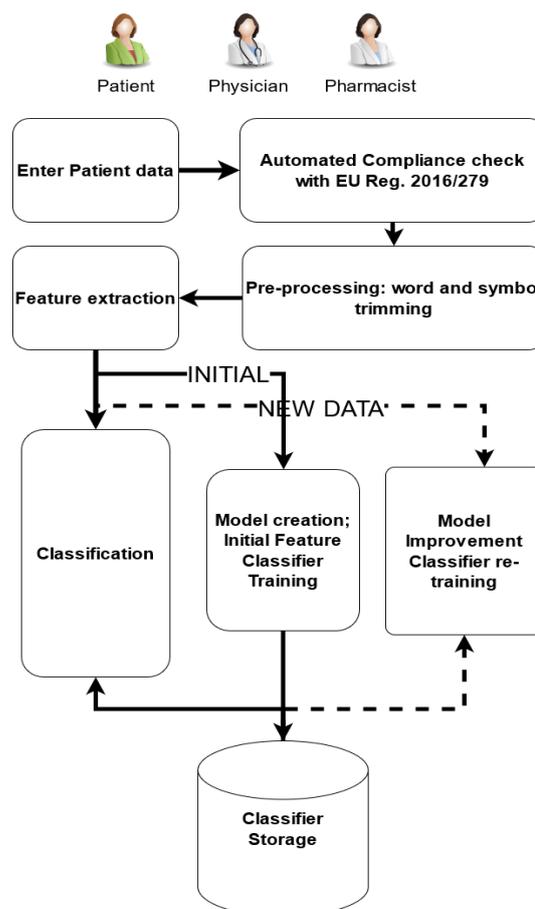


Figure 1. Workflow of the proposed system for automated evaluation of patient documents with ADI.

### 3. Results and Discussion

The pattern of acute poisoning changes over time and varies from country to country and even from geographical area to country. Between the years 1991-2020, 6977 patients with acute drug intoxications (ADI) were admitted to the Clinic for Intensive Treatment of Acute Poisoning and Toxic Allergies of the Military Medical Academy in Varna, Bulgaria. During the analyzed period of the preliminary study, the average frequency of cases per year was 232.

The close to linear trend of decreasing cases in Figure 2 of poisoning appears to be due to various factors:

- social, economic and cultural characteristics of the population of Varna region;
- demographic factors: increasing population in Varna region; population aging ( according to NSI) and permanent emigration of the young and active population;
- sufficient control of the dispensing and use of medicines and advanced training of pharmacists and assistant pharmacists in the Varna region;
- Prevention: especially in recent years there has been improved prevention of storage of drugs and food supplements at home.

The automated analysis of the structure of poisonings can be useful to justify measures to further and optimize medical care for the population, and the development of effective preventive measures, as well as the ways of rehabilitation aimed at reducing the impact of toxic factors as a cause of preventable morbidity and mortality. Assessment of poisoning cases or referral of intentional cases of poisoning to a psychologist or psychiatrist, specific educational programs There is a possibility that it can be proposed and developed a mechanism for automated evaluation of emergency care and the choice of therapeutically effective dosage forms in each a modern treatment center, as well as raising the qualification of pharmacists in order to prevent acute poisoning.

Prevention can be achieved by raising awareness and information about the risks of overdose of one or more medicinal products with / without alcohol co-administration. Educational and counseling campaigns could provide healthcare professionals (doctors,

pharmacists, nurses, etc.) with the necessary knowledge on risk assessment of overdose and early identification of high-risk patients.

The summary of typical toxicodromes in drug poisoning and treatment methods can serve as a basis for building a plausible decision-support network and for classifying non-discriminatory physical findings using artificial intelligence models.

The training of the classifiers has been performed on the dataset of preprocessed anonymized ambulatory cards in Bulgarian language. Pandas dataframes are used to store the anonymized text in Bulgarian language, normalized and projected values of the normalized text. The normalization level is as follows: Level 1 - removal of non-alphanumeric symbols, Level 2- removal of level-1 symbols and words that consist of less than three characters and Level 3- removal of all words with less than four characters.

**Table 1.** Accuracy of trained classifiers.

Classifier	Norm level1	Norm level2	Norm level3
NBayes	0.73	0.66	0.57
SVM	0.71	0.60	0.52

The results reported in Table 1 show that keeping variance of the data provides more promising results. Ongoing studies will be reported on keeping the whole variance of the text without removal of any characters compared to the normalized dataset usage.

### 4. Embedded system for Edge AI

In our approach we propose the usage of Edge AI for new data anonymization and text analysis with the trained NLP models. Embedded systems such as the one described in [13] allowed us to embed the text categorization algorithm in a relatively compact in size, power-efficient AI-enabled system. Edge AI systems are suitable in general for smaller applications, robots, and intelligent gateways that require analytics capabilities. The proposed system relies on an embedded device which is connected to the hospitals internal network. It would serve as an endpoint for encrypted http requests with outpatient files and will return the anonymized data, as well as a score for a particular ADI risk category.

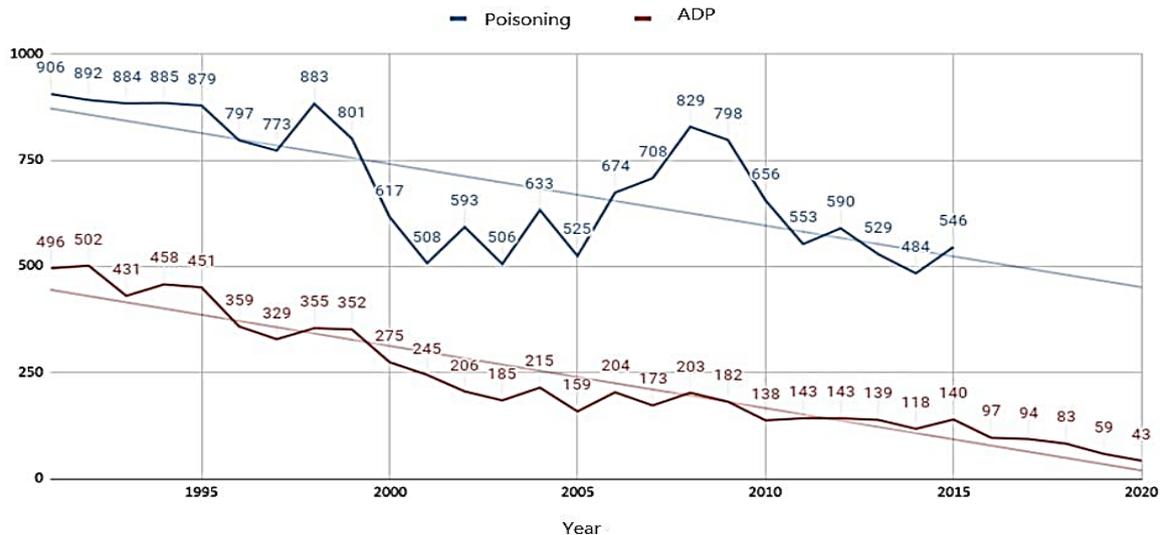


Figure 2. Graphical representation of ADI in relation to the total number of poisonings in Varna region for a 30-year period.

## 5. Conclusion

An approach has been presented in this paper for classifying the potential category of medical poisoning from patient medical records in Bulgarian language. The algorithm offers personal data anonymization which is essential for building applications nowadays. The text is preprocessed in the next step and unnecessary symbols and short words from a predefined list are filtered out. Text projection into vector space is performed with vectorizer based on bag-of words. Two classifiers are trained Naive Bayes and Support vector machines. The algorithm shows promising results and is a prerequisite for the implementation of a software application for visualization of clinical data in mobile devices.

The approach to training model and classifiers for patient documents containing text in natural language has been described, which has the ability to assess the possible risk of combined drug intoxication and proposed an algorithm for building a mobile application with trained models and classifiers to the Coordination Center. poisonings of international importance.

The full value of our approach is to help physicians successfully treat poisoned patients as well as to prevent poisoning in newly prescribed home therapy.

By incorporating the decision-making algorithm [16] for the treatment and prevention of poisoning, physicians would automate and select routine cases, freeing up time for processing and resolving more complex cases [14].

## 6. References

- [1] Georgieva S., Marinov P. (2019). Nutritional Toxicology – An Overview, *Proceedings of University of Ruse*, Vol. 58, Book 10.2, pp. 61-66 (in Bulgarian).
- [2] Georgieva S., Agova N. (2020). Risk Of Liver Injury During Use Of Dietary Supplements, *Medicine, Pharmacy, Public Health*, pp 48-51, 2020.
- [3] Abacha, Asma Ben, and Pierre Zweigenbaum (2015). MEANS: A medical question-answering system combining NLP techniques and semantic Web technologies, *Information processing & management* 51, no. 5, pp. 570-594.
- [4] Lee, Yen-Lin, Pei-Kuei Tsung, and Max Wu (2018). Technology trend of edge AI, in *2018 International Symposium on VLSI Design, Automation and Test (VLSI-DAT)*, pp. 1-2.

- [5] Shi, Y., Yang, K., Jiang, T., Zhang, J. and Letaief, K.B. (2020). Communication-efficient edge AI: Algorithms and systems, in *IEEE Communications Surveys & Tutorials*, 22(4), pp. 2167-2191.
- [6] Dhole, Gunjan, and Nilesh Uke. (2014). NLP based retrieval of medical information for diagnosis of human diseases, in *Int J Renew Energy Technol* 3, no. 10, 243e8.
- [7] Razno, M. (2019). Machine learning text classification model with NLP approach, *Computational Linguistics and Intelligent Systems* 2, pp. 71-73.
- [8] Afshar, Majid, Dmitriy Dligach, Brihat Sharma, Xiaoyuan Cai, Jason Boyda, Steven Birch, Daniel Valdez et al. (2019). "Development and application of a high throughput natural language processing architecture to convert all clinical documents in a clinical data warehouse into standardized medical vocabularies." *Journal of the American Medical Informatics Association* 26, no. 11: 1364-1369.
- [9] Nelli, F. (2018) "Machine Learning with scikit-learn.", *In Python Data Analytics*, Apress, Berkeley, CA, pp. 313-347.
- [10] Bisong, Ekaba. (2019) "Introduction to Scikit-learn." in *Building Machine Learning and Deep Learning Models on Google Cloud Platform*, Apress, Berkeley, CA, pp. 215-229.
- [11] Reiss, F., Cutler, B. and Eichenberger, Z., (2021) Natural Language Processing with Pandas DataFrames, *Proc. Of The 20th Python In Science Conf. (Scipy 2021)*, pp. 49-58.
- [12] Sokolov, S. (2018). Neural Network Based Multimodal Emotion Estimation." *ICAS 2018 - The 14th International Conference on Autonomic and Autonomous Systems* 12, pp. 4-7.
- [13] Cass, S. (2020). Nvidia makes it easy to embed AI: The Jetson nano packs a lot of machine-learning power into DIY projects-[Hands on]." *IEEE Spectrum* 57, no. 7, pp. 14-16.
- [14] Chary, Michael, Ed W. Boyer, and Michele M. Burns. (2021) Diagnosis of Acute Poisoning using explainable artificial intelligence." *Computers in Biology and Medicine* 134: 104469.
- [15] Özdikililer, E. and Göksel, Ç., (2018). Design a Model for Integrated Information Systems: DataOCEAN©, *Journal of Geomatics*, 3 (3), pp. 225-232. doi: 10.29128/geomatik.406294.
- [16] Sokolov, S. (2020). Deep Neural Network Based Processing of Medical Records in Cloud Environment", *Proceedings of 5<sup>th</sup> International Conference TIEM*, University of Telecommunications and Post, Sofia, pp. 104-108.

### **Acknowledgements**

The authors would like to express their gratitude to Mr. Angel Sokolov for his support during their research.