Polski Merkuriusz Lekarski

POLISH MEDICAL JOURNAL

ISSN 1426-9686

VOLUME LI, 2023, ISSUE 2, MAR – APR

EDITORIAL BOARD

© **ALUNA PUBLISHING**
Z.M. Przesmyckiego 29
05-510 Konstancin-Jeziorna, Poland
tel. +48 604 776 311
a.luczynska@wydawnictwo-aluna.pl

Aluna

www.polskimerkuriuszlekarski.pl

Distribution and subscription
Bartosz Guterman
tel. +48 22 245 10 55
prenumerata@wydawnictwo-aluna.pl

CONTENTS

Pol Merkur Lek, 2023; LI, 2: 128-134

ORIGINAL ARTICLE

DOI: 10.36740/Merkur202302104

IMPLEMENTATION OF THE DECISION TREE METHOD IN EXPERT ANALYSIS OF THE MEDICAL ERRORS IN OBSTETRIC PRACTICE

Valentyn V. Franchuk [1], Mykhailo S. Myroshnychenko [2], Mykhajlo S. Hnatjuk [1], Natalia M. Kalyniuk [1], Nadiia V. Humenna [1], Anna V. Narizhna [2], Ulyana Ya. Franchuk [1], Olena I. Hladii[1], Maksym V. Franchuk[1]

[1] I. HORBACHEVSKY TERNOPIL NATIONAL MEDICAL UNIVERSITY, TERNOPIL, UKRAINE
[2] KHARKIV NATIONAL MEDICAL UNIVERSITY, KHARKIV, UKRAINE

ABSTRACT

Aim: To identify expert patterns in cases of improper medical care in obstetric practice based on the analysis of the materials of judicial and investigative cases initiated against obstetrician-gynaecologists in cases of improper performance of their professional duties, using the decision tree method.

Materials and methods: A retrospective review of all alleged medical malpractice cases (a total 350) between 2007 and 2016 handled at Ternopil Regional Bureau of Forensic Medical Examination, Chernivtsi Regional Bureau of Forensic Medical Examination and Zhytomir Regional Bureau of Forensic Medical Examination (Ukraine) was performed.

Results: Expert commissions confirmed various shortcomings and omissions in provision of medical care in 232 (72.0%) of the investigated cases. Obstetricians were involved in claims in 82 (23.4%) cases. Application of intelligent data processing technology "Data Mining" with the use of the decision tree method revealed that inadequacies with regard to the medical records (attribute usage 100%) were the most informative attribute in the expert assessment of inappropriate medical care in obstetrics. Defects in the provision of obstetric care with a probability ($P = 0.71$) occur simultaneously both at pre-hospital and hospital levels and with a high probability ($P = 0.83$) result in severe consequences.

Conclusions: The use of modern technologies for data analysis and processing contributes to the formulation of mathematically substantiated statements that significantly enhance the reliability of expert opinions in cases of forensic medical examination attached to dereliction of duties by the medical practitioners.

KEY WORDS: forensic medical examination, inappropriate medical care, medical malpractice case, decision tree

INTRODUCTION

Analysis of medical care deficiencies in most scientific publications is usually based on well-known statistical data processing methods that provide absolute or relative numerical values of certain medical care defects.

However, in cases of non-obviousness, complexity and multifactorial nature of the system, when the strength of hidden information relationships among a large array of disaggregated data needs to be investigated, traditional methods of statistical processing of such data become insufficient [1]. Scientists have pointed out that applications where data mining tools are used in the fields of medicine are becoming more and more frequent [2-4]. Under such conditions, to solve the problems of expert assessment of the quality of medical care and predict defects in its provision, artificial intelligence technologies are applicable, in particular data processing and analysis technology Data Mining and one of its most popular methods - the decision tree method [5].

Due to many factors, both objective and subjective, including the increase in the level of passionarity of society, the growth of legal awareness and social activity of the public, the ability to access almost any information resource due to the widespread use of modern means of communication, etc., a persistent public demand for raising quality standards of medical and diagnostic care provided to citizens is being formed. So, patients' rights are occupying a prominent place in health care [6-8].

As some researchers report, according to the specialties of medical practitioners, the most criminogenic in many countries are obstetrics and gynecology [9-11]. In Ukraine the number of medical malpractice litigations against obstetricians is huge too [12]. In such cases, representatives of law enforcement agencies initiate criminal cases, the investigation of which requires forensic medical examination. The expert opinion in the investigation of medical malpractice cases is of the utmost importance for examination and is one of the main sources of evidence in the case [6, 12]. At the same time, in practice, in view of the objective complexity of the examination of diagnostic and treatment work, which is due to the specifics of professional medical activity, the expert assessment of medical malpractice cases is rarely definitive, and in most cases is presumptive. This is due to the lack of objective criteria that could substantiate the scientific validation and, accordingly, increase the reliability of expert opinions in cases where medical practitioners are held liable for professional offences in health care [13].

AIM

In this regard, the aim of this study was to identify expert patterns in cases of improper medical care in obstetric practice based on the analysis of the materials of judicial and investigative cases initiated against obstetrician-gynaecologists in cases of improper performance of their professional duties, using the decision tree method.

MATERIALS AND METHODS

Materials of forensic medical examination for medical malpractice cases between 2007 and 2016 handled at Ternopil Regional Bureau of Forensic Medical Examination, Chernivtsi Regional Bureau of Forensic Medical Examination and Zhytomir Regional Bureau of Forensic Medical Examination (Ukraine) were reviewed. In total, 350 such cases were studied, as well as 5056 related ones, both medical documents (case histories, outpatient cards, expert opinions, post-mortem examination reports, etc.) and proceedings (originals of criminal cases, on-site inspection reports, medical staff interrogation files, materials of internal audits of the quality of medical care, etc.). For each medical malpractice case, 70 different indicators were studied: the type of defect (diagnostic, medical-tactical, organisational, deontological, in medical records); the nature of improper medical care (its insufficiency, untimeliness, incorrect provision of or failure to provide medical care); the stage of medical care at which the defect occurred; the speciality of the medical practitioner who committed improper professional actions; reasons for the defect (objective and subjective); its consequences; cause and effect relationship. Thus, the collected data bank amounted to 24,500 indicators, the statistical processing and visualization of which was performed using the MS Excel 2016 spreadsheet.

The decision tree method is a decision-making tool for statistical processing and analysis of a certain database using artificial intelligence technologies. The essence of the method is to compile a mathematical model that establishes the dependence of the so-called target variable (the attribute (factor), the influence of which is being studied) on many other independent variables (attributes), and predicts the probability of a certain value of such variable [2-5].

If the target variable takes discrete values (e.g., "yes" or "no"), then using the decision tree method, its dependence on many other independent variables is established and the forecasting problem is solved.

In its simplest form, a decision tree is a way of representing rules in a hierarchical, sequential structure. The basis of such a structure is the answers "yes" or "no" (programming language) to certain questions. In other words, the presence or absence of each of the 70 indicators mentioned above was noted in each case of the medical malpractice case studied. Their discrete values in the form of "yes" or "no" answers (mathematically denoted "1" or "0", respectively) were entered into a general map and formatted in the MS Excel 2016 spreadsheet. The data obtained in such way were further processed mathematically by the C 4.5 algorithm specifically developed for the decision tree induction method, available in the licensed environment of the RStudio platform on the corresponding portal (https://www.rstudio.com/).

We have already covered in detail the theoretical foundations of the decision tree method, the technical details of creating an algorithm for the method, the mathematical problem of induction of the decision tree and its graphical model [13].

RESULTS AND DISCUSSION

The process of creating a decision tree in cases of forensic examination of medical care defects in obstetric practice starts with the establishment of a dependent (target) variable, i.e., the main parameter; the relationship of that variable with others should be established, the strength of such relationship should be mathematically proven and the probability of the value of the variable itself should be calculated.

As we established before [14] out of 350 materials of investigative cases analysed, various shortcomings and omissions during the provision of medical care were ascertained by expert commissions in 232 (72.0%) cases. Among them, law enforcement agencies initiated 82 cases directly against obstetrician-gynaecologists (23.4% of all studied). The vast majority of criminal cases was due to complaints of citizens about the improper provision of obstetric care (80 cases or 97.6%), and only 2 criminal cases were initiated against gynaecologists. Thus, using the phrase "obstetrician-gynaecologists" further in the article, we meant doctors who are mainly engaged in obstetric practice.

To ensure the correct operation of the algorithm, the target variable (obstetrician-gynaecologists) should receive the appropriate abbreviation in English. For our example, we shall call it "obstr" (short for English "obstetricians"). Similarly, it is necessary to provide English names for other attributes with which the informative relationship of defects caused by obstetrician-gynaecologists in professional activity, is being investigated.

The tree to be modelled answers the question of what the predicted value of the target variable will be, given the known values of other attributes. The implementation of this task consists in finding such attributes that are the most informative in relation to the target variable selected for the study (i.e., for obstetrician-gynaecologists). The following attributes were selected by the C4.5 algorithm as most significant parameters (from the 70 indicators studied for each medical malpractice case) (Table 1).

It is from them that the algorithm selects certain attributes as internal nodes of the tree, along which its further gradual branching is carried out. When building a tree, it is necessary that the target variable "obstr" takes on a value of exactly one type ("yes" or "no"); those are the leaf nodes. Each leaf node is characterized by the probability of obtaining a specific value for the target variable, i.e., the probability of belonging to class "obstr: yes" or "obstr: no".

Table 1. Attributes for constructing a decision tree regarding to the target variable "obstr" (obstetrician-gynaecologists)

Attribute designation	Attribute name	Attribute designation	Attribute name
A1 (obstr)	obstetrician-gynaecologists	A17 (emrgn)	provision of inadequate medical care by ambulance personnel
A2 (ansth)	anaesthesists	A18 (plclcs)	provision of inadequate medical care at outpatient stage
A3 (surg)	surgeons	A19 (hospit)	provision of inadequate medical care at hospital stage
A4 (pdtr)	paediatrician	A20 (direct)	direct cause and effect relationship
A5 (family)	family doctors	A21 (dif_dig)	medical care defect caused by diagnostic difficulties
A6 (traum)	traumatologists	A22 (atipic)	medical care defect due to atypical pathogenic mechanism
A7 (admsn)	admitting physicians	A23 (griev_)	improper provision of medical care led to serious consequences for the patient
A8 (diagn)	diagnostic pitfalls	A24 (comorb)	defect in medical care due to the presence of concomitant pathology
A9 (tactic)	medical-tactical defects	A25 (vidm)	adverse outcome due to the patient's refusal from hospitalization
A10 (istit)	organisational defects	A26 (latehsp)	adverse outcome due to late presentation of the patient
A11 (record)	defects in preparation of medical records	A27 (rapid)	adverse outcome due to the rapid progression of pathogenic mechanism
A12 (deont)	deontological errors	A28 (regum)	adverse outcome due to violation of the hospital regimen by the patient
A13 (unsuf_)	insufficient medical care	A29 (ridk)	defect in medical care due to the rarity of the pathology
A14 (untime_)	untimely medical care	A30 (unskill)	adverse outcome due to unskilled actions of medical personnel
A15 (improp_)	improper medical care	A31 (standr)	adverse outcome due to violation of the standards of medical care
A16 (care_abs_)	failure to provide medical care	A32 (Exitus)	medical care ended in a fatal outcome for the patient

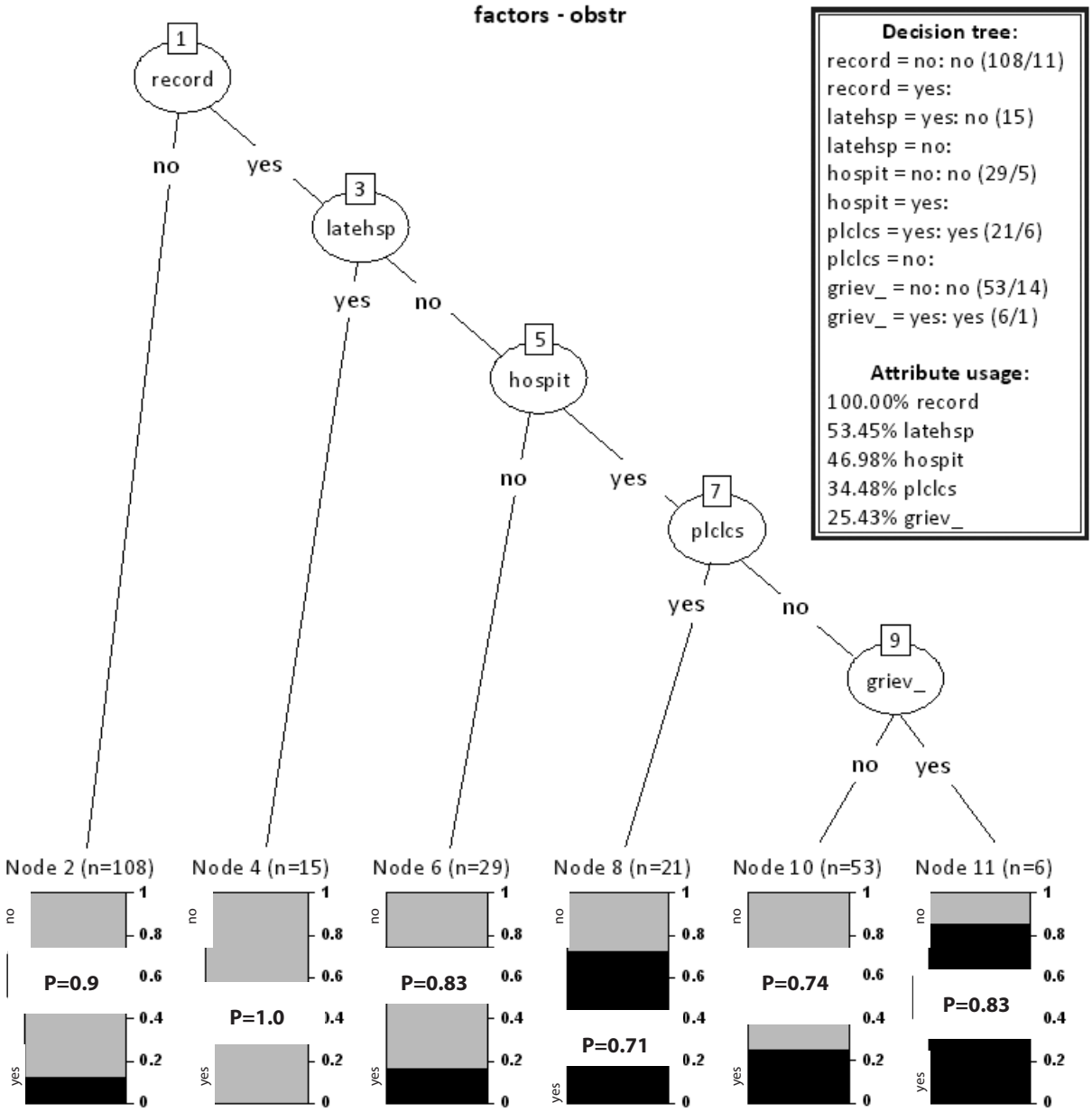

Fig 1. Induced decision tree for the target variable "obstr" (obstetrician-gynaecologists).

Further, the construction of the decision tree consists of two stages: construction of the classification model (the decision tree itself) and its interpretation.

In fact, at the model building stage, a classification tree is built, where the path from its root to one of the nodes is a set of rules for a specific case and is used to answer the question "What is the probability of the target variable "obstr"? (Fig. 1).

Namely: "What is the probability of the relationship of medical-tactical, diagnostic, organisational shortcomings during the provision of obstetric and gynaecological care, defects in the preparation of medical records, taking into account the characteristics of such inadequate medical care, the stages of its provision, negative consequences, objective and subjective factors that contributed to them." The rule in our case is a logical construction, represented as "if ..., then obstr: yes with the appropriate probability" or "if ..., then obstr: no with the appropriate probability". This allows to conclude: "if a criminal case is initiated against an obstetrician-gynaecologist in connection with the improper performance by them of their professional duties, then ..." or "if a criminal case is initiated against an obstetrician-gynaecologist in connection with the improper performance by them of their professional duties, then ..."

In the constructed decision tree, the following main elements are distinguished:

- root of tree: the algorithm-defined attribute that has the strongest informative association (attribute usage 100%) with the target variable studied. In our case, it is "record" (defects in the preparation of medical records);

- internal tree nodes: other attributes selected from the database under study, in accordance with their information content in relation to the main factor (target variable). These attributes are called splitting attributes. For the constructed decision tree (see Fig. 1), such attributes are: "latehsp" (adverse outcome due to late presentation of the patient), "hospit" (provision of inadequate medical care at hospital stage), "plclcs" (provision of inadequate medical care at outpatient stage), "griev_" (improper provision of medical care led to serious consequences for the patient);
- leaf node (or end node of the tree) is the value of the target variable "obstr", which suggests only two responses: "yes" or "no" – the so-called binary decision tree model (dichotomous classification). A leaf node is characterized by the probability of obtaining the target variable of a particular P value. The larger the P value, the higher the probability of a particular value (yes or no) of the main attribute under study.
- tree branch: sequence of response options (yes or no).

Each branch of the tree that comes from an internal node is marked with a splitting predicate. The latter may refer to only one splitting attribute of a given node. A characteristic feature of splitting predicates: each entry uses a unique path from the root of the tree to only one decision node. The combined information of splitting attributes and the splitting predicates at a node is called the splitting criterion. Selection of a specific attribute for splitting is the most important task of the decision tree derivation algorithm. The splitting criterion is formalised with certain mathematical expressions, with the help of which the informational entropy is analysed at each stage of the tree derivation, starting from the selection of a node and continuing to all final nodes. The best known of the existing splitting criteria are entropy and Gini index.

The decision tree construction process usually works in a top-down manner. During this process, the algorithm must find such a splitting criterion that can be used to split a set of data under study into subsets that would be associated with a certain internal node of the tree. Each such node must be marked with a certain attribute. There is a rule for attribute selection: it must split the input data set in such a way that the objects of the subsets resulting from such splitting are representatives of the same class. In order to identify the relationship between medical-tactical and diagnostic errors, defects in the preparation of medical records and organisational shortcomings during the provision of obstetric and gynaecological care with the features of such inadequate medical care, the stages of its provision, negative consequences, objective and subjective factors that contributed to them, etc., the algorithm processed 7424 separate data from 232 medical cases, in which defects in medical care were established by forensic expert commissions.

Along with the construction of a graphical model (the decision tree itself), the C4.5 algorithm also creates a so-called listing, which is a text output of a computer program written in the R programming language. The listing provides a list of attributes selected by the attribute splitting criterion, indicating their numerical values (presented as a table in the upper right corner of Fig. 1).

Therefore, according to the splitting criterion, which calculates the informational entropy at each stage, all independent attributes are reviewed. Algorithm C4.5 selected the "record" attribute as the root of the tree: defects in filling out medical records as an attribute with the highest value of the splitting criterion among all the others.

It means that the closest informative relationship (attribute usage is 100%) between adverse medical activities in the field of obstetrics and the 32 attributes studied, the algorithm chose the "record" attribute (defects in the preparation of medical records) and defined it as the root node of the induced tree (Fig. 1). In addition to that node, the program highlighted 4 internal nodes (splitting attributes): "latehsp" (adverse outcome of treatment associated with objective reason – late presentation), "hospit" (provision of inadequate medical care at hospital stage), "plclcs" (provision of inadequate medical care at follow-up medical care stage) , "griev_" (adverse outcome of improper medical care in the form of death of obstetric patient, foetus or newborn, severe birth trauma, loss of reproductive function, etc.). Thus, among the data processed by the decision tree derivation method, characterizing the features of inappropriate medical activity in obstetric practice, informative relationship was found only with those four attributes, while in relation to other studied attributes, the database intellectual processing program established no informative relationships.

The "record" attribute (defects in the preparation of medical records) divides the entire set of studied "medical cases" into two subsets. In one of them, the "record" attribute has the value "no" (108 such cases), in the other one it has the value "yes" (15+29+21+53+6=124 such cases). Thus, the analysis of the "no" branch of the "record" root node demonstrates that out of 108 cases of medical cases where no shortcomings in the preparation of medical records were found during the forensic medical examination, in 97 such cases with a probability of 97/108 (P = 0.90) those criminal cases were not related to obstetrician-gynaecologists and were initiated by the judicial and investigative authorities against medical practitioners of other specialities. This mathematical statement should be understood as follows: there is a high probability (P = 0.90) that there are no obstetrician-gynaecologists among those medical practitioners who do not make errors in compiling medical records. It means that the absence of shortcomings in medical records is typical not for them, but for representatives of other medical specialities.

Next, the subset formed by the "record=yes" branch is split by the splitting attribute "latehsp" (adverse outcome due to late presentation of the patient). That attribute, among all but "record", provided the maximum value of the splitting criterion among 124 cases for which "record=yes". In other words, among medical cases in which errors in the preparation of medical records were

stated (the value of the "record" attribute was "yes"), the attribute "latehsp" (late hospitalization) is most closely associated with the "record" attribute; the informative connection of "latehsp" with the studied target variable was 53.45%.

At the same time, according to the decision tree built by the algorithm and its listing (mathematical interpretation), the cases for which "record= yes" and "latehsp= yes" belong to class "obstr=no" with a probability of 15/15 (P = 1) (Fig. 1). In other words, among the 15 cases of late presentation, not a single case was associated with obstetric or gynaecological pathology ("record = yes: latehsp = yes: no (15)"). It means that this statement should be interpreted as follows: among 232 "medical cases" in which various shortcomings in the provision of medical care were established by expert commissions and where there were omissions in the preparation of medical records, such an objective reason for a defect committed by doctors as late hospitalization, is not typical for obstetric practice with probability (P = 1.0)

Further analysis of the induced decision tree indicates the presence of a close informative relationship between the nodes "hospit" and "plclcs". This indicates that among the "medical cases" with defects in medical records and where there was no untimely hospitalization of the patient, close informative relationship with the studied target variable (attribute usage 46.98%) demonstrates the shortcomings of medical care at the hospital stage. Under such conditions, allegations of improper performance of professional duties are made against obstetrician-gynaecologists with a probability of 15 cases out of 21 (P = 0.71).

Special attention needs to be paid to the last node of the constructed tree "griev_", the semantic understanding of which should be stated taking into account predicate "hospit-yes-plclcs-no-griev_-yes". So, in cases where there were shortcomings in medical records, where there was no late hospitalization, where improper medical care took place at the hospital stage and was not associated with omissions at polyclinic, in 5 cases out of 6 (P = 0.83) obstetrician-gynaecologists cause serious bodily harm by their erroneous actions.

Thus, based on the exceptional importance for the investigation of not the presupposition (probability), but the reliability (authenticity) of expert conclusions [8], the arguments presented in the article may be useful for forensic evaluation of defects in medical and diagnostic activities. In addition, the splitting attributes established as a result of the application of the decision tree method can be considered risk criteria for the occurrence of various defects during the provision of obstetric and gynaecological care. The obtained data may beneficial for health authorities in development of the necessary preventive measures aimed at preventing medical errors in obstetric practice, which, ultimately, will improve the quality of medical care for citizens.

CONCLUSIONS

The use of the "Data Mining" intelligent data processing technology using the decision tree method showed that the most informative attribute in the expert evaluation of obstetric care provided with various defects is the shortcomings in the preparation of medical records (attribute usage 100%).

Defects in the provision of obstetric care with a probability (P = 0.71) occur interrelatedly both at the outpatient and inpatient stages, and with a high probability (P = 0.83) lead to serious consequences.

The use of technologies for data analysis and processing contributes to the formulation of mathematically substantiated statements that significantly enhance the reliability of expert opinions in cases of forensic medical examination attached to dereliction of duties by the medical practitioners.

REFERENCES

1. Kolling ML, Furstenau LB, Sott MK et al. Data mining in healthcare: applying strategic intelligence techniques to depict 25 years of research development. Int J Environ Res Public Health. 2021;18(6):3099.
2. Stämpfli D, Winkler BA, Vilei SB, Burden AM. Assessment of minor health disorders with decision tree-based triage in community pharmacies. Res Social Adm Pharm. 2022;18(5):2867-2873.
3. Vera-Salmerón E, Domínguez-Nogueira C, Romero-Béjar JL et al. Decision-tree-based approach for pressure ulcer risk assessment in immobilized patients. International Journal of Environmental Research and Public Health. 2022; 19(18):11161.
4. Hajjej F, Alohali MA, Badr M et al. A comparison of decision tree algorithms in the assessment of biomedical data. Biomed Res Int. 2022;2022:9449497.
5. Nazari Nezhad S, Zahedi MH, Farahani E. Detecting diseases in medical prescriptions using data mining methods. BioData Mining.2022;15;29.
6. Shevchuk O, Matyukhina N, Davydenko S et al. Forensic examination in cases on the protection of human rights in the sphere of healthcare in Ukraine: legal issues. Juridical Tribune/Tribuna Juridica. 2022;12(4):552-565.
7. Vilchyk TB, Sokolova AK. Areas for further improvement of legislative regulation of patients' rights in Ukraine. Wiad Lek. 2019;72(7):1324-1330.
8. Maika N, Kalyniuk N, Sloma V et al. Basic of medicinal products reimbursement: a comparative-legal analysis to Ukraine: an update. Biomed Pharmacol J. 2021;14(2).
9. Almannie R, Almuhaideb M, Alyami F et al. The status of medical malpractice litigations in Saudi Arabia: Analysis of the annual report. Saudi J Anaesth. 2021;15:97-100.
10. Weiner CP. Reassuringly expensive – A commentary on obstetric emergency training in high-resource settings. Best Pract. Res. Clin. Obstet. Gynaecol. 2022;80:14-24.
11. Gornostay A, Ivantsova A, Mykhailichenko T. Medical error and liability for it in some post-soviet countries (Belarus, Kazakhstan, Moldova, Ukraine). Wiad Lek. 2019;72(5):877-882.

12. Stepaniuk R, Shcherbakovskyi M, Kikinchuk V et al. Problems of investigation of medical crimes in Ukraine. Amazonia Investiga. 2022;11(57):39-47.
13. Franchuk VV, Mikhaylichenko BV, Franchuk MV. Application of the decision tree method in forensic-medical practice in the analysis of «doctors proceedings». Sudebno-meditsinskaya ekspertiza. 2020;63(1):9-14.
14. Franchuk VV. The forensic medical aspects of the inappropriate medical care in the modern-day Ukraine. Sudebno-meditsinskaya ekspertiza. 2018;61(2):48-52.

ORCID AND CONTRIBUTIONSHIP *
Valentyn V. Franchuk: 0000-0001-8484-8049 [A, D]
Mykhailo S. Myroshnychenko: 0000-0002-6920-837[4 F]
Mykhajlo S. Hnatjuk: 0000-0002-4110-5568 [E]
Natalia M. Kalyniuk: 0000-0002-1613-835X [B]
Nadiia V. Humenna: 0000-0002-3838-3858 [B]
Anna V. Narizhna: 0000-0002-8583-7445 [C]
Ulyana Ya. Franchuk: 0000-0002-0338-2791 [B]
Olena I. Hladii: 0000-0003-1800-9591 [C]
Maksym V. Franchuk: 0000-0002-2708-3614 [C]

ADDRESS FOR CORRESPONDENCE
Mykhailo S. Myroshnychenko
Kharkiv National Medical University
4 Nauky Avenue, 61000 Kharkiv, Ukraine
tel: +380501699763
e-mail: msmyroshnychenko@ukr.net

RECEIVED: 29.08.2022
ACCEPTED: 02.03.2023

CONFLICT OF INTEREST
The Authors declare no conflict of interest.

* Contribution: A – Work concept and design, B – Data collection and analysis, C – Responsibility for statistical analysis, D – Writing the article, E – Critical review, F – Final approval.