

Системи за подршку одлучивању базирани на вештачкој интелигенцији у третману умерене форме билијарног панкреатитиса

Ања Буљевић, Александар Глуховић, Мирна Н. Капетина, Александар Кнежевић, Зоран Д. Јеличић

Апстракт—У оквиру овог рада представљено је једно решење система за подршку одлучивању у лечењу умерене форме билијарног панкреатитиса. На основу параметара и показатеља из клиничке праксе, развијен је систем за усмеравање лекара приликом избора методе и поступка лечења овог запаљења панкреаса. Због релативно малог скупа података, услед специфичних медицинских процедура, избор релевантних обележја потврђен је кроз два формализма: корелационом анализом и стаблом одлучивања. На основу одабраних обележја, до финалног решења се долази уз ослонац на теорију потпорних вектора. Предложено решење нашло је своју примену у клиничкој пракси.

Кључне речи: умерена форма билијарног панкреатитиса, корелациона анализа, стабло одлучивања, СВМ.

I. УВОД

У наставку текста представимо основне показатеље и параметре који карактеришу умерену форму билијарног панкреатитиса, али и конвенционалне и не тако конвенционалне начине лечења истог. Наиме, полазећи од претпоставке да се ради о мултидисциплинарном проблему са чистом медицинском применом разумевање медицинског дела је од највећег значаја не само за крајње кориснике, већ као и тумачење изабраних обележја и разумевања свих сложених феномена који се могу добити у математичком опису проблема.

Акутни панкреатитис представља ензимско инфламаторно обољење панкреаса, које може захватити како сам орган, тако и околна ткива. Инциденција обољења је око 17/100000 становника. Најчешћи

А. Буљевић (anjabuljevic@uns.ac.rs)*, А. Глуховић (aleksandar.gluhovic@gmail.com)**, М. Н. Капетина (mirna.kapetina@uns.ac.rs)*, А. Кнежевић (aknezevic021@gmail.com)***, З. Д. Јеличић (jelicic@uns.ac.rs)*

* Универзитет у Новом Саду, Факултет техничких наука, Департман за рачунарство и аутоматику, Трг Доситеја Обрадовића 6, 21000, Нови Сад, Р. Србија

** New Hospital, Алберта Ајнштајна, Нови Сад, Р. Србија

*** Клинички центар Војводине, Хајдук Вељкова 1-9, Нови Сад, Р. Србија

етиолошки чиниоци који се везују за ово стање су билијарна калкулоза (45%) и конзумација алкохолних пића (35%). Умерене форме акутног панкреатитиса, које су и предмет овог истраживања, јављају се у 80% случајева и имају благ клинички ток са стопом mortalитета од 1%. Акутни панкреатитис билијарне етиологије узрокован је калкулозом жучне кесе и/или жучних путева. Третман калкулозе билијарног стабла код умерене форме панкреатитиса, по актуелној препоруци Америчког удружења гастроентеролога и ендоскопских хирурга (American Assosiation of Gastroenterology and Endoscopic Surgeons - SAGES), своди се на уклањање жучне кесе лапароскопском холецистектомијом (ЛХ) са интраоперативном холангиографијом (ИОХ), у циљу превенције појаве новог напада болести. Уколико постоји сумња на присуство калкулуса у жучним каналима са повишеним вредностима билирубина, ради се ендоскопска ретроградна холангиопанкреатографија (ЕРЦП) са ендоскопском папилотомијом (ЕПТ), у циљу уклањања калкулуса и детритуса из жучних водова, и обезбеђивања нормалног протока жучи у дванаестопалачно црево. Ова процедура се углавном изводи пре ЛХ, мада се може радити и током ЛХ или након ње, [1].

На студији случајева Војводине, ЛХ је доступна у свим болницама, као и у клиничком центру. Међутим, ЕРЦП процедура могућа је само у Клиничком центру Војводине и то од стране једног лекара. По нашим најбољим сазнањима, укупан број лекара у Србији који изводе ову операцију је 3 или 4. Као кључно питање и задатак намеће се одабир скупа објективних параметара који ће лекара одредити да само изведе широко доступну ЛХ или да ипак пацијента пошаље и на додатну процедуру ЕРЦП. Важно је напоменути да су анализирани само објективне параметре јер субјективни параметри (нпр. ултразвучни преглед) не могу да гарантују доследност у резултатима. Други начини прегледа, попут компјутеризоване томографије (ЦТ) и магнетна резонанца, нису широко доступни. При томе, за око 80% пацијената довољна је ЛХ, док свега 20% пацијената захтева и ЕРЦП.

Кључно за анализе у систему за подршку одлу-

чивању је било да минимизујемо лажно негативне закључке у потреби извођења ЕРЦП што је и заиста у складу са добром клиничком праксом. Због малог скупа података, избор обележја је урађен кроз корелациону анализу и стабло одлучивања што је оправдано са математичког аспекта [2], а такав начин рада се користи и у сличним медицинским истраживањима [3]. Према нашим најбољим сазнањима, овакав приступ подршке одлучивању у лечењу умерене форме акутног панкреатитиса је оригиналан и први пут се презентује у овом раду.

Рад је организован на следећи начин. У поглављу II приказан је одабир дискриминантних обележја. У поглављу III приказано је балансирање података. Формирање система за подршку одлучивању дато је у поглављу IV, док је закључак дат у поглављу V.

Ова студија је одобрена од стране Етичког комитета Медицинског факултета у Новом Саду и Етичке комисије Клиничког центра Војводине.

II. КОРЕЛАЦИОНА АНАЛИЗА И СТАБЛО ОДЛУЧИВАЊА

Први корак за решавање овог реалног проблема била је идентификација дискриминантних обележја (медицинских параметара) који су од круцијалне важности за указивање присуства калкулуса у жучним водовима. За потребе овог пројекта коришћени су медицински подаци пацијаната оболелих од акутног панкреатитиса који су прикупљени у Клиничком центру Војводине. Приликом самог пријема пацијената прикупљено је преко 80 параметара. Неки од њих су били дескриптивни подаци о субјективном стању пацијената, затим о историји болести, као и објективни параметри попут налази крви и субјективни као што је ултразвучни преглед. Сви параметри су по правилу распоређени у временским серијама од 12 часова у првих 48 сати по пријему у болницу.

У добијеној бази података налазе се подаци за 100 пацијената лечених од умерене форме билијарног панкреатитиса. У литератури [1] проналазимо да се потенцијални кандидати за ЕРЦП могу узети у разматрање уколико је њихов Глазгов скор мањи од 3. Због ове чињенице, база се смањује на 96 пацијента. Од тих 96 пацијената, код њих 19 је рађен ЕРЦП, док је код 77 пацијената била довољна ЛХ.

Кроз нумеричке поступке издвајања обележја и пратећи начин размисљања лекара издвојили смо 5 објективних параметара: укупни билирубин, директни билирубин, алкална фосфатаза, ЦРП и гамаГТ. Субјективни параметри су показали извесну недоследност и нису могли бити узети у разматрање. Испитивање дискриминантности обележја рађено је на два начина:

- 1) коришћењем Пирсоновог коефицијента корелације и
- 2) коришћењем стабла одлучивања.

A. Одабир обележја коришћењем Пирсоновог коефицијента корелације

Поступак идентификације дискриминантних обележја започели смо корелационом анализом. Уколико желимо да испитамо да ли постоји зависност између два (или више) обележја, тада говоримо о утврђивању постојања корелације између тих обележја [4]. Коефицијент корелације је показатељ степена статистичке повезаности обележја и представља меру њихове линеарне зависности. Једна од најчешће коришћених мера повезаности два обележја јесте Пирсонов коефицијент корелације [4]. У табели I су приказани Пирсонови коефицијенти корелације за она обележја која имају највећи тражени коефицијент и они уједно представљају најзначајнија обележја за наш проблем.

Табела I
Пирсонов коефицијент корелација за дискриминантна обележја

| Назив обележја | $ \rho $ |
|--------------------|----------|
| укупни билирубин | 0.4051 |
| директни билирубин | 0.3837 |
| алкална фосфатаза | 0.3703 |
| ЦРП | 0.274 |
| гама ГТ | 0.2164 |

Из ове табеле можемо приметити да су чак и најзначајнија обележја по Пирсоновом коефицијенту корелације уствари у слабој корелацији са излазном променљивом. Како бисмо потврдили да су одабрана обележја заиста дискриминантна за наш проблем, урађена је и валидација добијених обележја коришћењем стабла одлучивања.

B. Одабир обележја коришћењем стабла одлучивања

Други начин одређивања дискриминантних обележја који је имплементиран у овом раду јесте стабло одлучивања. Стабло одлучивања представља графички модел за визуализацију процеса одлучивања када се решавање проблема одлучивања своди на доношење више сукцесивних одлука, [5]. На почетку се бира параметар чија вредност најбоље дели расположиве узорке. Као што је познато, стабло одлучивања се осим за одређивање дискриминантних обележја, може користити и за класификацију података. Због малог броја података, као и слабе корелације улазних параметара са излазом, у овом случају стабло одлучивања је коришћено искључиво за селекцију обележја, док ће се метода вектора носача користити као подршка одлучивању у третману умерене форме билијарног панкреатитиса. Стабло одлучивања, поред селекције, омогућава нам и увид у значајност изабраних обележја, као и сам „пут” селекције. Предност стабла одлучивања у односу на Пирсонов коефицијент корелације јесте аутоматска селекција обележја [6].

За потребе овог рада коришћен је ID3 алгоритам. Квантитативна мера коју ID3 алгоритам користи ка-

ко би одредио најбоља обележја јесте информацио-на добит [7]. Симулације су вршене са различитим конфигурацијама стабла. Дубина стабла је узимала вредности на интервалу од 3 до 10, док се број листова се налазио у опсегу од 2 до 10. Посматрајући добијена стабла одлучивања примећено је да се као заједнички садржаоци свих стабала издвајају следећа обележја: укупни билирубин, ЦРП, директни билирубин, алкална фосфатаза и гама ГТ. Осим тога, наведена обележја се у већини генерисаних стабала налазе ближе корену стабла, односно на мањим дубинама стабла. Упоредивши резултате које смо добили коришћењем стабла одлучивања и Пирсоновог коефицијента корелације, потврдили смо потпуну подударност између дискриминантних обележја добијених коришћењем ова два поступка.

III. БАЛАНСИРАЊЕ ПОДАТАКА

Следећи проблем који се намеће је небалансираност података. Небалансираност података подразумева ситуацију у којој се број узорака значајно разликује по класама у поступку класификације. Класификатори машинског учења тешко се носе са небалансираним скупом података за обуку, јер су осетљиви на пропорционалност различитих класа, па се као последица јавља тенденција алгоритама да фаворизују класу са највећим уделом испитаника што најчешће резултује „обмањујућом” тачношћу. То је посебно проблематично када је од интереса тачна класификација „ретке” (мањинске) класе, али налазимо висок проценат тачности који је заправо последица исправне класификације већинске класе. Ову чињеницу овде експлицитно наводимо, не само да нагласимо потребу за балансирањем података, већ и да уведемо посебну метрику за оцену квалитета предикције. С обзиром на то да алгоритми машинског учења имају за циљ да смање укупну стопу грешке, неће обраћати посебну пажњу на мањинску класу, и вероватно неће успети да направе тачно предвиђање за ову класу, јер о њој не садржи довољно података.

У бази података разматраног проблема, од укупно 96 пацијената, код само 19 испитаника је рађен ЕРЦП, док код 77 испитаника није рађен ЕРЦП. Као што је наведено у уводном поглављу, приоритет је да се минимизују лажно негативни закључци у потреби извођења ЕРЦП који у овом случају представља мањинску класу. Дакле, неопходно је било урадити балансирање података.

У литератури [8] се може пронаћи неколицина потенцијалних метода за решавање овог проблема, а као најбољи метод се наводи додавање нових података у класе са процентуално мањим бројем узорака. Међутим, у пракси је то веома тешко постићи, па се прибегава неким другим методама. Један од најчешће коришћених је да се из већинске класе избаце узорци како би се број узорака већинске класе изједначио са бројем узорака мањинске класе. При томе, мора се

водити рачуна да укупан број узорака мора бити бар 10 пута већи од броја изабраних обележја [9].

Ова препорука за балансирање података је искоришћена у овом раду на следећи начин. На случајан начин се од 77 испитаника код којих није рађен ЕРЦП изабере 40 и тих 40 испитаника улазе у процес обуке SVM алгоритмом, заједно са 19 испитаника код којих је рађен ЕРЦП. Примећујемо да подаци и даље нису најбоље балансирани, али због ограничења да укупан број узорака мора бити барем 10 пута већи од укупног броја обележја, ово је најбоље што смо могли да добијемо из коришћене базе података.

IV. SVM МАТЕМАТИЧКИ МОДЕЛ

Након што смо идентификовали параметре од интереса и избалансирали податке, неопходно је било да се на основу издвојених параметара формира систем за подршку одлучивању у третману умерене форме акутног билијарног панкреатитиса. Због заиста малог скупа података, нарочито малог за пацијенте којима је рађен ЕРЦП, определили смо се за математички модел уз ослонац на теорију потпорних вектора (SVM) [10]. Метода класификације базирана на векторима носачима представља један од модела машинског учења који се веома често користи како за класификациону, тако и за регресиону анализу. SVM алгоритам је довољно познат алгоритам, па неће бити детаљно извођен у овом раду, а његово детаљно математичко извођење можете пронаћи у литератури [11], [12].

Посебна пажња је била посвећена оптимизацији параметара SVM алгоритма. Будући да перформансе генерализације SVM алгоритма у великој мери зависе од параметара C (хиперпараметар који прави компомис између сложености модела и степена до кога се толеришу одступања модела), ε (хиперпараметар који контролише ширину неосетљиве зоне, а његова вредност утиче на број вектора носача) и γ (параметар који одређује облик изабране кернел функције), неопходно је извршити њихову оптимизацију. Према [13], [14], између наведених параметара постоји јака веза, тако да је препорука да се они оптимизују истовремено, а не одвојено. Оптимизација параметара је извршена уз помоћ алгоритма роја честица [15] и као критеријум оптималности је коришћена средња квадратна грешка [16]. Као кернел функције су прослеђиване: радијална, гаусова, линеарна и полиномна кернел функција. Након оптимизација кернел функција и свих њених пратећих параметара, добијено је да се најбољи резултати добију за SVM алгоритам који има радијалну кернел функцију и добијени су следећи параметри: $C = 10^2$, $\varepsilon = 10^{-3}$ и $\gamma = 10^{-1}$.

Иако је урађено балансирање података, класе и даље нису у потпуности балансиране. Осим тога, SVM модел је јако осетљив на улазне податке и неретко од скупа улазних података, зависи и тачност класификације SVM. Како би се добили што објективнији

результати, спроведен је експеримент описан у наставку. Узорци који се прослеђују SVM као улазни подаци су прослеђивани на следећи начин:

- од 19 испитаника код којих је рађен ЕРЦП, на случајан начин се бира 80% испитаника (15 испитаника) који се прослеђују SVM алгоритму за обуку и валидацију, док је на преостала 4 узорка вршено тестирање добијеног модела
- од 77 испитаника код којих није рађен ЕРЦП на случајан начин се изабере 40 испитаника. Од тих 40 испитаника, на случајан начин се бира 80% испитаника (32 испитаника) који се прослеђују SVM алгоритму за обуку и валидацију, док је на преосталих 8 узорака вршено тестирање добијеног модела.

Као оцена успешности одабраног класификационог модела, уобичајено се користе стандардизоване мере и оцене којима се квантификује рад пројектованог система за класификацију и предикцију. Под оцењивањем, односно процењивањем рада система за предикцију углавном се мисли на одређивање вредности неких од стандардних мера којима се квантификује његов учинак односно перформансе. Мера квалитета представља потенцијал модела да коректно предвиди класу новог податка. Матрица конфузије (eng. *confusion matrix*) представља детаљан и прегледан приказ бројева исправно и погрешно класификованих узорака на основу којих се могу вршити оцене добијеног модела класификације. Општи облик матрице конфузије за бинарну класификацију је приказан табелом II, где је

- TP (true positive; стварно позитивни) - број узорака који припадају позитивној класи, а додељена им је позитивна класа,
- TN (true negative; стварно негативни) - број узорака који припадају негативној класи, а додељена им је негативна класа,
- FP (false positive; лажно позитивни) - број узорака који припадају негативној класи, а додељена им је позитивна класа,
- FN (false negative; лажно негативни) - број узорака који припадају позитивној класи, а додељена им је негативна класа

Табела II
Општи облик матрице конфузије.

| | | Предвиђена класа | |
|-------------|-----------|------------------|-----------|
| | | Класа = 0 | Класа = 1 |
| Праве класе | Класа = 0 | TN | FP |
| | Класа = 1 | FN | TP |

Табела III представља један пример матрице конфузије SVM класификатора за случај када треба да се спроведе ЕРЦП. Примећујемо да је у овом случају од 40 испитаника којима је довољна само ЛХ процедура, наш класификациони модел то погодио за 39 испитаника, док је за само једног испитаника рекао да му је

неопходан и ЕРЦП. Што се тиче 19 испитаника којима је неопходан и ЕРЦП, наш класификациони модел је одговарајућу класу погодио за 17 испитаника, док је за двојицу испитаника погрешно доделио класу.

Табела III
Матрица конфузије за случајеве када је неопходан ЕРЦП.

| | | Предикција | |
|-------------------|-----------|------------|-----------|
| | | ЛХ | ЛХ и ЕРЦП |
| Стварни резултати | ЛХ | 39 | 1 |
| | ЛХ и ЕРЦП | 2 | 17 |

По завршетку креирања класификационог модела, корисно је тестирати његове перформансе на скупу података који му је непознат, при чему је неопходно да тај скуп података садржи информације о класама. Овакав вид тестирања представља непристрасну оцену генерализације. Најчешће се користи K-слојна унакрсна валидација (eng. *K-fold cross-validation*). K-слојна унакрсна валидација (кросвалидација) је техника евалуације класификационих модела која се изводи тако што се оригинални скуп података дели на k једнаких подскупова. Један подскуп се користи за тестирање, док се сви остали подскупови користе за тренирање. Овај поступак се понавља у k итерација тако да се сваки подскуп користи тачно једном за тестирање. По завршетку свих итерација, издваја се онај модел који је имао најмању грешку класификације, [17].

Већ је напоменуто да је излаз јако осетљив на добијени скуп улазних података, па да би се добили што објективнији резултати поступак описан горе је поновљен у 100 итерација са одабраним SVM моделом. На овај начин су праћене две тачности модела: тачност над обучавајућим скупом података у процесу кросвалидације и тачност над тестним скупом. Тачност, као једна од најчешће коришћених мера за приказивање успешности класификације, представља однос укупног броја коректних предвиђања и укупног броја предвиђања. Математички, тачност записујемо

$$\text{тачност} = \frac{TP + TN}{TP + TN + FP + FN}$$

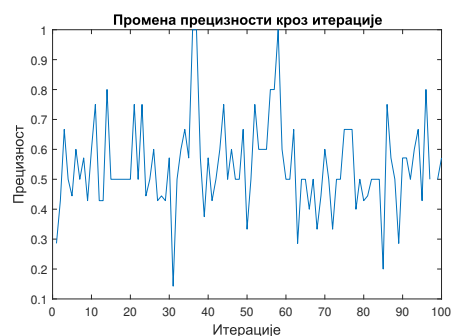
Промена тачности над обучавајућим скупом у процесу кросвалидације за одабрани модел за 100 итерација у зависности од одабраних улазних података дат је на слици 1. Просечна тачност овог модела износи 90.51%.

Промена тачности над тестним скупом за одабрани модел за 100 итерација у зависности од одабраних улазних података дата је на слици 2. Просечна тачност овог модела износи 82.68%.

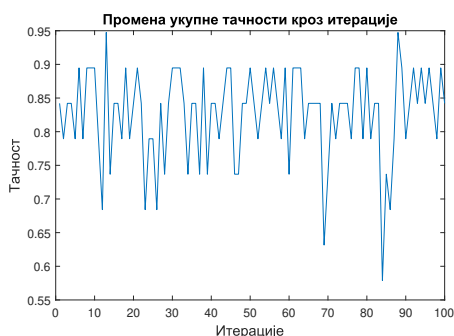
Поред тачности, морамо обратити пажњу на још неколико показатеља успешности класификације. Прецизност је мера слична тачности, али се односи искључиво на једну посматрану класу. Она представља



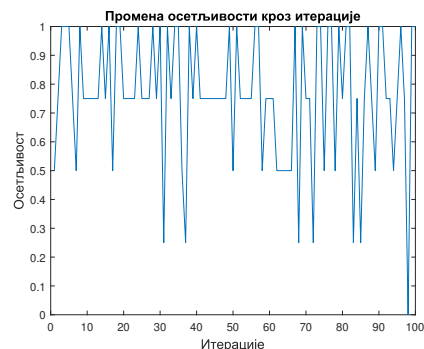
Слика 1. Промена укупне тачности у процесу кросвалидације кроз итерације



Слика 3. Промена прецизности кроз итерације



Слика 2. Промена укупне тачности кроз итерације



Слика 4. Промена осетљивости кроз итерације

однос тачних позитивних предвиђања и укупног броја случајева у којима је класификатор предвидео посматрану класу. Осетљивост приказује однос коректно предвиђених вектора атрибута неке класе и укупног броја правих понављања те класе у скупу података. Специфичност је способност теста да коректно идентификује одсуство неког атрибута, а може да се интерпретира и као процена условне вероватноће да атрибут није идентификован, уз услов да га на посматраној позицији заиста нема. Наведене мере математички можемо записати на следећи начин

$$\begin{aligned} \text{прецизност} &= \frac{TP}{TP + FP}, \\ \text{осетљивост} &= \frac{TP}{TP + FN}, \\ \text{специфичност} &= \frac{TN}{FP + TN}. \end{aligned}$$

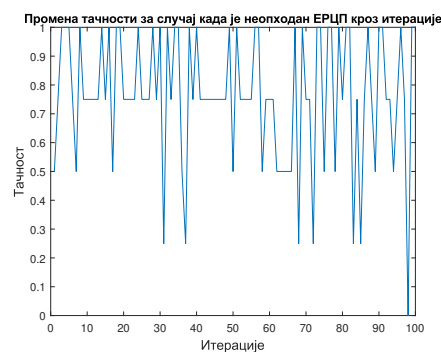
Графици наведених мера прецизности, осетљивости и специфичности су приказани на сликама 3, 4 и 5 респективно. Просечна вредност прецизности износи 73.23%, просечна вредност осетљивости износи 76.25%, док просечна вредност специфичности износи 64.88%.

Као што је већ речено, нама је од посебне важности тачност погађања за случајеве када је неопходна ЕРЦП процедура, односно битно нам је да одговор класификационог алгоритма за случај када треба да се ради ЕРЦП процедура буде тачан. Због тога је по-



Слика 5. Промена специфичности кроз итерације

себно издвојена и тачност модела само за ову класу и приказана је на слици 6. Просечна вредност тачности за овај случај износи 76.25%.



Слика 6. Промена тачности за случај када је неопходан ЕРЦП кроз итерације

Анализирајући вредности за приказане мере, можемо да закључимо да одабрани класификациони СВМ модел даје задовољавајуће резултате. Примећујемо да је укупна тачност модела већа у односу на тачност модела када испитујемо случај када је неопходна и ЕРЦП процедура, али то смо и очекивали пошто та класа представља мањинску класу за овај проблем.

V. ЗАКЉУЧАК

У склопу овог рада пројектован је систем за подршку одлучивању у третману умерене форме билијарног панкреатитиса. Пројектовању система за подршку одлучивању претходила је предобрада података. Приликом самог пријема пацијената прикупљено је преко 80 медицинских параметара, па је било неопходно одабрати најзначајније параметре који би нам указали на присуство калкулозе у жучним водовима. Дискриминантна обележја су одређена коришћењем Пирсоновог коефицијента корелације и као дискриминантна обележја су се издвојили: укупни билирубин, ЦРП, директни билирубин, алкална фосфатаза и гама ГТ. Ваљаност издвојених обележја потврђена је применом стабла одлучивања.

Главни циљ нашег рада је био да минимизујемо лажно негативне закључке у потреби извођења ЕРЦП. За формирање система за подршку одлучивању коришћен је СВМ алгоритам, коме је претходила процедура балансирање података по класама. Оптимизација параметера модела над тестним скупом података је била 82.68%. Пошто нам је од посебне важности била тачна предикција случајева када је неопходан ЕРЦП, издвојена је и просечна тачност модела за овај случај и она износи 76.25%.

Битно је напоменути да лекари у својој клиничкој пракси заиста користе издвојена обележја како би идентификовали присуство калкулозе, а последично следи одлука да ли треба да се ради додатна процедура како би се решио проблем оваквог оболења панкреаса. Оно што се математички не може описати, бар не на малом скупу података, јесте субјективни осећај, знање и искуство искусног лекара из клиничке праксе који приликом ЛХ може да примети да ли у жучним водовима постоји калкулуза и на основу тога да донесе процену да ли је потребан ЕРЦП.

ЗАХВАЛНИЦА

Овај рад је подржан од стране Министарства просвете, науке и технолошког развоја кроз пројекат број 451-03-68/2020-14/200156: „Иновативна научна и уметничка испитивања из домена делатности ФТН-а”

ЛИТЕРАТУРА

- [1] А. Глуховић, “Алгоритам примене лапароскопске холецистектомије и ендоскопске ретроградне холангиопанкреатографије са папилотомијом у третману умерене форме билијарног панкреатитиса.” Докторска дисертација. Нови Сад, 2016.
- [2] Н. Zhou, J. Zhang, Y. Zhou, X. Guo, and Y. Ma, “A feature selection algorithm of decision tree based on feature weight,” *Expert Systems with Applications*, vol. 164, p. 113842, 2021.

- [3] Q. Qiu, Y.-j. Nian, Y. Guo, L. Tang, N. Lu, L.-z. Wen, B. Wang, D.-f. Chen, and K.-j. Liu, “Development and validation of three machine-learning models for predicting multiple organ failure in moderately severe and severe acute pancreatitis,” *BMC gastroenterology*, vol. 19, no. 1, pp. 1–9, 2019.
- [4] M. A. Hall, “Correlation-based feature selection for discrete and numeric class machine learning,” in *Proceedings of the Seventeenth International Conference on Machine Learning, ICML '00*, (San Francisco, CA, USA), p. 359–366, Morgan Kaufmann Publishers Inc., 2000.
- [5] A. Geron, “Hands-on machine learning with scikit-learn, keras tensorflow.” O’Reilly, 2019.
- [6] K. G. (auth.), *Meta-Learning in Decision Tree Induction. Studies in Computational Intelligence 498*, Springer International Publishing, 1 ed., 2014.
- [7] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Monterey, CA: Wadsworth and Brooks, 1984.
- [8] G. E. Batista, R. C. Prati, and M. C. Monard, “A study of the behavior of several methods for balancing machine learning training data,” *ACM SIGKDD explorations newsletter*, vol. 6, no. 1, pp. 20–29, 2004.
- [9] V. Sugumaran and K. Ramachandran, “Effect of number of features on classification of roller bearing faults using svm and psvm,” *Expert Systems with Applications*, vol. 38, no. 4, pp. 4088–4096, 2011.
- [10] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [11] B. Schölkopf, A. J. Smola, F. Bach, et al., *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.
- [12] F. Dedov, *The Python Bible Volume 4: Machine Learning (Neural Networks, Tensorflow, Sklearn, SVM)*. The Python Bible Series, Independently Published, 2019.
- [13] V. Cherkassky and Y. Ma, “Practical selection of svm parameters and noise estimation for svm regression,” *Neural networks*, vol. 17, no. 1, pp. 113–126, 2004.
- [14] Z. Shaowu, W. Lianghong, Y. Xiaofang, and T. Wen, “Parameters selection of svm for function approximation based on differential evolution,” in *International Conference on Intelligent Systems and Knowledge Engineering 2007*, pp. 529–535, Atlantis Press, 2007.
- [15] Z. Kanović, M. R. Rapaić, and Z. D. Jeličić, “Generalized particle swarm optimization algorithm-theoretical and empirical analysis with application in fault detection,” *Applied Mathematics and Computation*, vol. 217, no. 24, pp. 10175–10186, 2011.
- [16] Y. Ren and G. Bai, “Determination of optimal svm parameters by using ga/pso,” *Journal of computers*, vol. 5, no. 8, pp. 1160–1168, 2010.
- [17] C.-W. Hsu, C.-C. Chang, C.-J. Lin, et al., “A practical guide to support vector classification,” 2003.

ABSTRACT

This paper presents one solution for the decision support system in the treatment of moderate form of biliary pancreatitis. Based on parameters and indications from clinical practices, the system for directing doctors in selecting methods and procedures for treating this pancreas ailment has been developed. Due to specific medical procedures, the data set is relatively small, so the choice of relevant features was confirmed through two formalisms: the correlation analysis and the decision tree. Based on the selected features, the final solution is reached by the theory of supporting vectors. The proposed solution has found its application in clinical practice.

Decision support system based on artificial intelligence in the treatment of moderate form of biliary pancreatitis

Anja Buljević, Aleksandar Gluhović, Mirna N. Kapetina, Aleksandar Knežević, Zoran D. Jeličić