



AGH

**AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE**

Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii
Biomedycznej

Katedra Automatyki i Inżynierii Biomedycznej

Rozprawa doktorska

System wspomagania decyzji dla wczesnego
wykrywania dysplazji oskrzelowo-płucnej u
wcześniaków

mgr inż. Marcin Ochab

Promotor: prof. dr hab. inż. Wiesław Wajs

KRAKÓW 2016

Streszczenie

Praca przedstawia system wspomaganie decyzji dla wczesnego wykrywania dysplazji oskrzelowo-płucnej (BPD) u wcześniaków. Standardowa diagnoza choroby dokonywana jest dopiero po czwartym tygodniu od urodzenia. Jak wynika z przeglądu literatury zamieszczonej we wstępie, pomimo wielu prowadzonych prac dotyczących predykcji BPD, aktualnie nie istnieje żaden ogólnie uznany model, możliwy do zastosowania w codziennej praktyce lekarskiej. Na podstawie wykonanych i zamieszczonych w pracy wstępnych badań zaproponowano więc stworzenie odpowiedniego systemu wspomaganie decyzji. Opracowane oprogramowanie pozwala na znalezienie optymalnej metody predykcji przyszłej diagnozy już po pierwszym tygodniu życia dziecka, niezależnie od ilości dostępnych danych. Do konstrukcji bazy wiedzy wykorzystano dane historyczne 109 pacjentów, których masa urodzeniowa była równa lub mniejsza od 1500g. Wykorzystując wszystkie 2^{14} kombinacje czternastu dostępnych parametrów wykonano predykcję standardową metodą regresji logistycznej oraz nigdy wcześniej nie używanym do tego celu algorytmem SVM (ang. support vector machine). Ponad 33 tys. otrzymanych w ten sposób wyników zgromadzono w relacyjnej bazie danych. Na ich podstawie, mając na uwadze wymagania użytkownika oraz dostępne w konkretnym przypadku dane, system jest w stanie wyznaczyć najskuteczniejszą metodę predykcji oraz optymalny dla niej zestaw parametrów. Co ważne, szacowane są także przedziały, w jakich będą się zawierać wartości trafności, czułości i specyficzności proponowanych modeli (poprzez wyznaczenie ich odchyłeń standardowych). Ponadto system potrafi zasugerować, jakie dodatkowe pomiary warto byłoby wykonać w celu maksymalizacji jakości predykcji oraz jak znacznej poprawy można się dzięki temu spodziewać. Stosując wskazania systemu, przy optymalnej dostępności danych, możliwe jest uzyskanie średniej trafności predykcji na poziomie 83,25%. Dzięki opisanym w pracy eksperymentom, możliwe było także dokonanie oceny istotności poszczególnych parametrów dla czułości i specyficzności przewidywania przyszłej diagnozy oraz porównanie własności wyników otrzymanych obiema metodami. W pracy zamieszczono ponadto przykłady ilustrujące problemy stojące na przeszkodzie stworzenia jednego możliwego do zastosowania w praktyce modelu. Tym samym udowodniono zasadność zastosowania tego nowego, oryginalnego podejścia do prognozowania badanego schorzenia. Praca zawiera ponadto opis użytych metod predykcji wraz z przykładami liczbowymi ich zastosowania zawartymi w załączniku.

Abstract

This work presents a decision support system for early medical diagnosis of bronchopulmonary dysplasia (BPD) for extremely premature infants. Since the illness cannot be diagnosed until the 28th day of life it is very important to forecast such a result. As presented in the introduction a significant amount of works is focused in that field. Unfortunately researchers agree that for many reasons there is still a lack of a generally accepted prediction model, which could be used in common medical practice. Based on preliminary results presented in that work author proposed to construct a proper decision support system. Developed software, no matter how much information user can provide to the system, enables him to find the best method of prediction in the given circumstances just after first week of patient's life. For that purpose a knowledge database was created based on the historical data gathered including data on 109 patients with birth weight less than or equal to 1,500 g. As classifiers application considers classic logit regression and support vector machine, which usage in that field is a novelty. The core of the database consists of thirty-three thousands classification results calculated specifically for that system, and obtained by using 2^{14} different combinations of fourteen risk factors. Based on that scores and user demands, the system recommends the best methods and the most suitable parameter subset among those currently available to the user. The program is also able to estimate the accuracy, sensitivity and specificity together with their standard deviations. The user is also given information on which additional parameter is worth adding to his measurement system most and what increase in prediction efficiency it is expected to trigger. The BPD can be predicted by the system with the mean accuracy reaching up to 83.25%. Work presents also a set of examples illustrating the difficulties in obtaining one single model that can be widely used, and thus explaining why a decision support system approach is much more useful in day-to-day clinical practice. In addition, the work discusses the significance of the parameters used and the impact of a chosen method on the sensitivity and specificity. The used classifiers were also widely described in one of the chapters – numerical examples of use were presented in the appendix.

Dziękuję mojej Rodzinie, bez której wsparcia
nie byłoby możliwe powstanie tej pracy.

Spis treści

Wstęp	13
Dysplazja oskrzelowo-płucna	13
Czynniki ryzyka BPD	15
Znane modele predykcyjne BPD	19
Cele pracy	21
Teza pracy	23
1 Dane	25
1.1 Źródła danych	25
1.2 Charakterystyka danych	26
1.2.1 Dane statyczne	26
1.2.2 Dane dynamiczne	31
1.3 Analiza statystyczna parametrów	35
2 Metody predykcji	43
2.1 Regresja logistyczna	44
2.2 Support Vector Machine	46
2.3 Porównanie sposobów działania obu metod	54
3 Wstępne badania	57
3.1 Klasyfikatory regresji logistycznej	59
3.2 Modele oparte o SVM środowiska Matlab	60
3.3 Porównanie wyników Regresji Logistycznej oraz Matlab SVM	62
3.4 Predykcja z wykorzystaniem LIBSVM	62

3.5	Porównanie wyników Regresji Logistycznej oraz LIBSVM	65
3.6	Analiza zdjęć RTG	72
3.7	Wstępne wnioski	77
4	System wspomagania decyzji	81
4.1	Budowa bazy wiedzy	82
4.2	Budowa systemu	87
5	Eksperymenty ilustrujące działanie stworzonego systemu	91
5.1	Sekwencyjna rozbudowa modeli przez system wspomagania decyzji	91
5.2	Dobór cech funkcją <i>sequentialfs</i> pakietu Matlab	92
5.3	Wyniki eksperymentów	93
6	Wnioski	103
6.1	Istotność parametrów	103
6.2	System wspomagania decyzji	106
	Bibliografia	111
	Dodatek A	119
A.1	Przykład liczbowy predykcji z użyciem regresji logistycznej	119
A.2	Przykład liczbowy predykcji z użyciem SVM	122

Spis rysunków

1.1	Histogram masy urodzeniowej	27
1.2	Histogram wskaźnika włósniczkowo-pęcherzykowego	28
1.3	Histogram wieku płodowego	29
1.4	Histogram parametru high94	31
1.5	Histogram parametru low85	32
1.6	Histogram uśrednionego tętna	32
1.7	Histogram trendu tętna	33
1.8	Histogram uśrednionej saturacji hemoglobiny tlenem	33
1.9	Histogram trendu saturacji	34
1.10	Histogram odchylenia standardowego saturacji	34
1.11	Histogram trendu odchylenia standardowego saturacji	35
2.1	Idea maksymalizacji marginesu pomiędzy klasami w metodzie SVM	47
2.2	Wpływ wektorów wspierających na wynik klasyfikacji SVM	49
2.3	Przykład przekształcenia jądrowego umożliwiającego liniową separację danych	51
2.4	Porównanie przykładowych krzywych rozdzielających klasy dla RL i SVM	55
3.1	Przykład optymalizacji ACC na powierzchni $\gamma \times c$	66
3.2	Przykładowe zdjęcia RTG	73
3.3	Histogram ocen zdjęć RTG	75

Spis tablic

1.1	Charakterystyka rozkładów parametrów ciągłych	37
1.2	Porównanie grup z pozytywną i negatywną diagnozą BPD	38
1.3	Wyniki testu współliniowości zmiennych niezależnych	40
1.4	Wartości współczynnika korelacji pomiędzy parametrami	41
3.1	Porównanie najlepszych wyników RL oraz Matlab SVM	63
3.2	Porównanie najlepszych wyników RL oraz LIBSVM	67
3.3	Porównanie odchyłeń standardowych trafności predykcji RL oraz LIBSVM	70
3.4	Lista klasyfikatorów SVM o wartościach TPR i $SPC > 80\%$	71
3.5	Zastosowany system oceny zdjęć RTG	74
4.1	Czterdzieści najlepszych klasyfikatorów zawartych w bazie wiedzy	85
5.1	Wyniki serii eksperymentów z użyciem systemu wspomaganie decyzji	101
A.1	Dane dziesięciu pacjentów użyte w przykładzie liczbowym	120
A.2	Znormalizowane dane użyte w przykładzie liczbowym	122

Spis algorytmów

1	Poszukiwanie najlepszego modelu M.SVM poprzez optymalizację parametru σ	61
2	Poszukiwanie najlepszego modelu LIBSVM poprzez optymalizację parametrów γ i c	65
3	Budowa bazy wiedzy dla metody LIBSVM	83
4	Procedura działania systemu wspomagania decyzji	89
5	Eksperymenty badawcze z użyciem systemu wspomagania decyzji	92

Spis skrótów i oznaczeń

aa – parametr związany z AA

AA – wskaźnik włośniczkowo-pęcherzykowy (ang. alveolar-arterial ratio)

ACC – trafność predykcji (ang. accuracy)

ACC_{dev} – odchylenie standardowe trafności predykcji

AT:RVET(c) – parametr wyznaczany podczas badania dopplerowskiego, określający zależności czasowe pracy serca (ang. corrected acceleration time to right ventricular ejection time ratio)

AUC – pole powierzchni pod krzywą ROC (ang. area under the curve)

BPD – dysplazja oskrzelowo-płucna (ang. bronchopulmonary dysplasia)

bpmmean – parametr odpowiadający uśrednionemu tętnu

bpmmean_tr – parametr odpowiadający trendowi uśrednionego tętna

bweight – parametr odpowiadający masie urodzeniowej (ang. birth weight)

C-SVC – jedna z metod klasyfikacji należących do grupy SVM

DPIP – długość trwania szczytowej wartości ciśnienia wdechowego (ang. duration of peak inspiratory pressure)

ELBW – ekstremalnie niska urodzeniowa masa ciała (ang. Extremely Low Birth Weight)

FN – błędne diagnozy negatywne (ang. False Negatives)

FP – błędne diagnozy pozytywne (ang. False Positives)

gage – parametr odpowiadający wiekowi płodowemu (ang. gestational age)

high94 – parametr związany z saturacją hemoglobiny tlenem we krwi

LBW – niska masa urodzeniowa ciała (ang. Low Birth Weight)

LIBSVM – zewnętrzna biblioteka implementująca metody SVM

low85 – parametr związany z saturacją hemoglobiny tlenem we krwi

MPIP – maksymalne szczytowe ciśnienie wdechowe (ang. maximum peak inspiratory pressure)

M.SVM – biblioteczna implementacja metody SVM pakietu Matlab

ν -SVC (nu-SVC) – jedna z metod klasyfikacji należących do grupy SVM

pCO_2 – ciśnienie parcjalne dwutlenku węgla we krwi tętniczej

pda – parametr związany z występowaniem PDA

pH – ujemny logarytm dziesiętny ze stężenia jonów wodorowych

PIP – szczytowe ciśnienie wdechowe (ang. peak inspiratory pressure)

PDA – przetrwały przewód tętniczy (ang. patent ductus arteriosus)

pO_2 – ciśnienie parcjalne tlenu we krwi tętniczej

RBF – nazwa funkcji – ang. radial basis function

respimv – parametr związany z użyciem respiratora

RL – regresja logistyczna

ROC – krzywa obrazująca zależność czułości od specyficzności (ang. receiver operating characteristic)

SPC – specyficzność predykcji (ang. sensitivity)

SpO_2 – poziom wysycenia tlenem hemoglobiny we krwi

spo2dev – parametr związany z odchyleniem standardowym saturacji hemoglobiny tlenem

spo2dev_tr – parametr związany z trendem odchylenia standardowego saturacji hemoglobiny tlenem

spo2mean – parametr związany z średnią wartością saturacji hemoglobiny tlenem

spo2mean_tr – parametr związany trendem saturacji hemoglobiny tlenem

SMO – nazwa metody optymalizacji – ang. sequential minimal optimization

surfact – parametr związany z użyciem surfaktantu

SVM – maszyna wektorów wspierających (ang. support vector machine)
TN – właściwe diagnozy negatywne (ang. True Negatives)
TP – właściwe diagnozy pozytywne (ang. True Positives)
TPR – czułość predykcji (ang. specificity)
VIF – ang. variance inflation factor
VLBW –bardzo niska masa urodzeniowa ciała (ang. Very Low Birth Weight)
W-K – walidacja krzyżowa



Wstęp

Dysplazja oskrzelowo-płucna

Dysplazja oskrzelowo-płucna (BPD – ang. bronchopulmonary dysplasia, niekiedy nazywana w literaturze zamiennie ang. chronic lung disease) jest przewlekłą chorobą płuc przedwcześnie urodzonych noworodków, którą charakteryzują: stan zapalny płuc, nieprawidłowe ich wzrastanie, niewłaściwy rozwój pęcherzyków płucnych oraz ich unaczynienia [Johnson i Thebaud, 2014]. Jak stwierdzają czasopisma medyczne, przypadłość dotyka ponad jedną trzecią dzieci o masie urodzeniowej mniejszej niż 1000 g [Głowacka i Lis, 2008, Walsh i in., 2006] oraz charakteryzuje się wysoką śmiertelnością (zależnie od wieku płodowego, średnio około 28%) [Oh i in., 2005]. Głównie dotyka wcześniaki o skrajnie małej (ELBW) lub bardzo małej (VLBW) masie ciała tj. 500–1500 g.

Zgodnie z przyjętą na początku bieżącego stulecia przez National Heart, Lung and Blood Institute w ramach programu NICHD (ang. National Institute of Child Health and Human Development) aktualnie uznawaną definicją [Bancalari i Claire, 2016, Jobe i Bancalari, 2001], czas diagnozy BPD zależny jest od wieku płodowego pacjenta. U dzieci urodzonych przed 32 tygodniem ciąży badanie przeprowadza się w 36 tygodniu od momentu poczęcia lub w dniu wypisu do domu, w zależności od tego, co wystąpi jako pierwsze. Noworodki urodzone później diagnozuje się po 28 dobie życia. Głównym kryterium jest tlenozależność (potrzeba sztucznego podawania tlenu) w momencie badania. Biorąc pod uwagę

stężenie tlenu w mieszaninie oddechowej oraz stopień jego suplementacji wyróżnia się trzy stopnie BPD: łagodną, umiarkowaną oraz ciężką. Co prawda, dzięki coraz wyższemu poziomowi opieki medycznej, wprowadzeniu steroidów prenatalnych oraz zastosowaniu surfaktantu po urodzeniu, choroba obecnie najczęściej przybiera łagodniejszą formę choroby przewlekłej (tzw. „nowa” BPD) [De Paepe, 2016, Groothuis i Makari, 2012, Kwinta i Pietrzyk, 2010], to jednak jej skutki są często obserwowane przez wiele lat, aż do wieku dojrzewania czy nawet wczesnego wieku dorosłego [Głowacka i Lis, 2008, Johnson i Thebaud, 2014, Kwinta i in., 2009, Landry i Banbury, 2016]. Następstwami mogą się okazać obniżony wzrost [Sauve i Singhal, 1985, Vohr i in., 1991], zaburzenia neurologiczne jak np. porażenie mózgowe [Skidmore i in., 1990, Vohr i in., 1991] czy też powikłania płucne [Grégoire i in., 1998, Vohr i in., 1991]. Początkowo podstawą definicji BPD były klasyczne cechy radiologiczne zaproponowane przez Northway Jr i in. [1967], tj. najpierw przymglenie pól płucnych, później narastające zmiany niedodmowe, obszary rozdęcia i rozedmy śródmiąższowej, po czym niedodmy i pęcherze rozedmowe, do których ostatecznie dołączają się zmiany włókniejące. W literaturze dostępne są dwa systemy oceny radiogramów: pierwszy zaproponowali Toce i in. [1984], drugi, późniejszy opracowali Weinstein i in. [1994]. Szerszy opis wspomnianych metodyk ewaluacji przedstawiono w Rozdz. 3.6. Mimo, że obecnie zrezygnowano z dawnej definicji dysplazji oskrzelowo-płucnej, dla tzw. „starej” BPD wspomniane cechy radiologiczne dość dobrze korelują ze stanem klinicznym [Głowacka i Lis, 2008]. Niestety dla „nowej” BPD analiz radiologicznych nie przyjmuje się już za tak wiarygodne [Johnson i Thebaud, 2014]. Noack i in. [1993] wspominają np., że dzięki lekko zmodyfikowanej pierwotnej metodzie oceny uzyskali co prawda 93% czułości, ale już tylko 53% specyficzności predykcji korzystając ze zdjęć wykonanych pomiędzy 3, a 23 dniem życia. Co prawda Yuksel i in. [1993] twierdzą, że możliwe jest osiągnięcie nawet 71% specyficzności oraz 88% czułości mając do dyspozycji radiogramy z 7 dnia po narodzinach,

jednakże są oni jednymi z nielicznych wyrażających taką opinię. Wspomniane przededefiniowanie BPD związane jest także z faktem, że z każdym dziesięcioleciem granica przeżywalności noworodków urodzonych przedwcześnie się przesuwa, ponadto im niższy wiek płodowy i mniej rozwinięte płuca, tym ciężiej zaobserwować zmiany radiologiczne. Z pomocą przychodzą tutaj nowe narzędzia obrazujące, jak tomografia komputerowa [Johnson i Thebaud, 2014]. Szczególnie obiecujące wydają się być systemy oceny zdjęć uzyskanych przy pomocy tomografii wysokiej rozdzielczości (np. Shin i in. [2013] uzyskali korelację $r = 0,646, p < 0.001$ z późniejszą diagnozą).

Głównym powodem występowania opisywanych zaburzeń jest przedwczesny poród (wywołany np. zakażeniem wewnątrzmacicznym), a co za tym idzie przerwanie prawidłowego, wewnątrzmacicznego rozwoju płuc oraz ich urazy związane ze stosowaniem wentylacji mechanicznej [Ramos i in., 2016] i tlenoterapii. Z jednej strony procedury te przyczyniają się do powstania BPD, z drugiej jednak są nieodzowne, aby np. przy wystąpieniu zespołu zaburzeń oddychania (RDS – ang. respiratory distress syndrome) [Johnson i Thebaud, 2014] dać pacjentowi szansę przeżycia. Na rozwinięcie się schorzenia ma jednak wpływ także wiele innych, do dziś ostatecznie nie zbadanych czynników [Głowacka i Lis, 2008].

Czynniki ryzyka BPD

Wydaje się być logicznym, że im wcześniej przed planowanym terminem dojdzie do porodu tym mniej rozwinięte i bardziej podatne na uszkodzenia okażą się płuca. Dlatego też jako jeden z głównych predyktorów choroby wymienia się w literaturze wiek płodowy [Bancalari i Claire, 2016, Bhering i in., 2007, Bhutani i Abbasi, 1992, Corcoran i in., 1993, Cunha i in., 2005, Groothuis i Makari, 2012, Kim i in., 2005, Laughon i in., 2011, Oh i in., 2005, Sinkin i in., 1990]. Drugim często wspomnianym elementem [Ambalavanan i in., 2008, Bancalari i Claire, 2016, Corcoran i in., 1993, Kim i in., 2005, Laughon i in., 2011, Oh i in.,

2005, Rojas i in., 1995, Sinkin i in., 1990] jest masa urodzeniowa, która zgodnie z oczekiwaniami jest wyraźnie skorelowana z wiekiem płodowym (Tab. 1.4). Trzecie bardzo popularne w literaturze przekonanie [Bancalari i Claire, 2016, Bhering i in., 2007, Corcoran i in., 1993, Cunha i in., 2005, Farstad i in., 2011, Groothuis i Makari, 2012, Kwinta i in., 2008b, Marshall i in., 1999, Oh i in., 2005, Rojas i in., 1995, Sosenko i Bancalari, 2012] mówi, że przetrwały przewód tętniczy (PDA, szerzej opisany w Rozdz. 1.2.1), powodując zwiększenie przepływu płucnego oraz obrzęk śródmiąższowy przyczynia się do występowania BPD. W kilku pracach [Ambalavanan i in., 2008, Cunha i in., 2005, Farstad i in., 2011, Marshall i in., 1999, Oh i in., 2005] podaje się, że dobrym predyktorem schorzenia jest także fakt podania surfaktantu podczas pobytu w szpitalu. Jest to co prawda wtórny i niemierzalny parametr, ale niezaprzeczalnie zawiera on pewną nieuchwytną wiedzę o stanie pacjenta, posiadaną przez zalecającego go lekarza. Także potrzeba użycia respiratora jest wymieniana jako dobry prognostyk BPD [Bhering i in., 2007, Groothuis i Makari, 2012, Kwinta i in., 2008b, Laughon i in., 2011, Ramos i in., 2016]. Wynika to ze wspomnianego wcześniej ryzyka uszkodzenia płuc podczas wentylacji mechanicznej. Można jednak traktować taką informację również jako pewnego rodzaju estymację ogólnego stanu chorego. Wielu autorów wstępnie brało pod uwagę także ogólną ocenę stanu zaraz po urodzeniu, wyrażoną w skali Apgar, jak się jednak zawsze okazywało, nie były to istotne dane. Interesującym jest jednakże fakt, że ryzyko wystąpienia przypadłości wymienia się jako wyższe wśród pacjentów płci męskiej [Ambalavanan i in., 2008, Corcoran i in., 1993, Farstad i in., 2011, Laughon i in., 2011, Oh i in., 2005]. Co więcej, jak opisują Laughon i in. [2011], na zagrożenie wpływ ma nawet pochodzenie etniczne chorego. Pomimo, że dotychczas potwierdzają to tylko wstępne badania [Lavoie, 2016, Somaschini i in., 2012], badacze są zgodni, iż rozwinięcie się choroby jest uwarunkowane również genetycznie [Groothuis i Makari, 2012, Johnson i Thebaud, 2014, Kwinta i in., 2008a, Sosenko i Bancalari,

2012]. Wszystkie dotychczas wspomniane czynniki i parametry można by określić jako statyczne, ponieważ mogą one zostać jednorazowo wyznaczone w pierwszej dobie życia pacjenta.

Dla odróżnienia, istnieje szereg cech, których ciągłe pomiary mogą także wskazywać na ryzyko wystąpienia BPD – na potrzeby tej pracy nazwano je parametrami dynamicznymi. Główną z takich wielkości jest stężenie tlenu w mieszance oddechowej oznaczane symbolem FiO_2 . W naturze stężenie tlenu w powietrzu wynosi 20,9%. Jeżeli w celu zachowania prawidłowej saturacji (wysycenia tlenem hemoglobiny we krwi) wymagane jest większe stężenie tlenu, mówi się o wspomnianej wcześniej tlenozależności pacjenta, której wystąpienie w odpowiednim dniu życia jest jednoznaczne z pozytywną diagnozą BPD. Dlatego w literaturze [Kim i in., 2005, 2001, Romagnoli i in., 1998, Subhedar i in., 1998] wspomina się, że analiza wcześniejszych wartości nastawy respiratora jaką jest FiO_2 może być dobrym predyktorem przyszłej diagnozy. Dla przykładu Laughon i in. [2011] biorą pod uwagę dzienne wartości średnie, natomiast Ambalavanan i in. [2008] wynik otrzymany przy przyjęciu na oddział. Jako najlepszy wskaźnik często wymienia się szczytową zanotowaną wartość FiO_2 [Corcoran i in., 1993, Cunha i in., 2005]. W literaturze wspomina się także o powiązanim parametrze (więcej w Rozdz. 1.2.1), którym jest wskaźnik włośniczkowo-pęcherzykowy (AA od ang. alveolar-arterial ratio) definiujący jakość zachodzącego procesu oddychania. Subhedar i in. [1998] wspominają, że parametr ten jest zbyt skorelowany z wiekiem płodowym i masą urodzeniową, aby mógł być użyty do predykcji. Stoch [2007] natomiast nie zaobserwował takowej korelacji czy też zjawiska współliniowości parametrów (przeprowadzając analizę wartości współczynników VIF – więcej w Rozdz. 1.3), jednakże pośród 14 badanych przez autora wartości, AA wykazywał najniższą siłę predykcji. Innym ważnym parametrem odczytywanym z respiratora, o którym wspominają Cunha i in. [2005], Romagnoli i in. [1998], Ryan i in. [1996], Subhedar i in. [1998] jest szczytowe ciśnienie wdechowe (PIP,

od ang. peak inspiratory pressure). Corcoran i in. [1993] wprowadzają jeszcze maksymalną wartość PIP (MPIP od ang. maximum peak inspiratory pressure) oraz długość trwania szczytowej wartości (DPIP od ang. duration of peak inspiratory pressure). Inną grupą parametrów, na którą zwrócili uwagę badacze [Cunha i in., 2005, Subhedar i in., 1998] są wyniki gazometrii krwi pacjenta. Autorzy ci wymieniają ciśnienie parcjalne dwutlenku węgla we krwi tętniczej (pCO_2), ciśnienie parcjalne tlenu we krwi tętniczej (pO_2) oraz ujemny logarytm dziesiętny ze stężenia jonów wodorowych (pH) jako ważne czynniki w badaniu ryzyka BPD. Wśród parametrów dynamicznych Stoch [2007] w swojej pracy wymienia także średnią wartość tętna z pierwszego tygodnia życia jako jeden z siedmiu najważniejszych czynników. Autor badał także istotność trendu tętna wyrażonego jako stosunek średniego tętna w pierwszym dniu do pierwszego tygodnia życia, jednak parametr ten okazał się mało istotny przy predykcji BPD. Częściej spotykanym w literaturze [Askie i in., 2003, Geary i in., 2008, Jobe, 2011, Kinsella i in., 2006, Singer i in., 1992] jest zagadnienie ryzyka BPD zależnie od poziomu wysycenia tlenem hemoglobiny we krwi (SpO_2). Autorzy głównie biorą pod uwagę fakt i długość trwania dopuszczenia do spadku lub wzrostu wartości parametru poza określone przedziały, a także średnie (lub mediany) za dane okresy. Zagadnienie jest ściśle związane z parametrem FiO_2 , zwiększając bowiem zawartość tlenu w mieszance oddechowej można próbować podwyższyć SpO_2 dożądanego poziomu. Jak wspomniano wcześniej nie jest to jednak obojętne dla ryzyka rozwoju BPD. Stoch [2007] otrzymuje ciekawe wyniki biorąc pod uwagę także trendy średnich oraz odchylenia standardowe wartości. Kuenzel [2011] natomiast opisuje nawet analizę szeregów czasowych SpO_2 w kontekście BPD. Oh i in. [2005] oraz Bhering i in. [2007] traktują także masę pacjenta jako zmienną w czasie, zauważając że utrata masy poniżej 15% może być ważnym predyktorem BPD. Subhedar i in. [1998] na podstawie badania dopplerowskiego analizują także

zależności czasowe pracy serca w kontekście BPD (AT:RVET(c) od ang. corrected acceleration time to right ventricular ejection time ratio).

Znane modele predykcyjne BPD

Można spotkać wiele publikacji dotyczących predykcji BPD. Nie licząc drobnych wyjątków, wszystkie z nich oparte są na użyciu regresji logistycznej. Jak szerzej opisano w Rozdz. 2.1 metoda ta posiada wiele zalet, jest prosta i niewymagająca co do założeń. Stąd zapewne jej wyjątkowa popularność wśród lekarzy, którzy stanowią gros badaczy zajmujących się predykcją opisywanej choroby. Jednymi z pierwszych, którzy podejmowali próby predykcji BPD byli Sinkin i in. [1990] osiągając jedynie 64% czułości predykcji w 12 godzinie życia oraz 75% w 10 dniu. Natomiast korzystając z wieku płodowego oraz parametru C_L (od ang. pulmonary compliance lub lung compliance, określającego zdolność płuc do rozciągania i poszerzania) 143 pacjentów, Bhutani i Abbasi [1992] byli w stanie osiągnąć 78% dodatniej wartości predykcyjnej (PPV) oraz 100% ujemnej (NPV). Kolejnymi autorami, którzy podejmowali ten temat byli Corcoran i in. [1993]. Korzystając z danych statycznych oraz informacji uzyskanych z respiratora na temat 412 osób byli oni w stanie przeprowadzić predykcję po 72 godzinach od porodu, uzyskując specyficzność równą 88%, ale już tylko 65% czułości. Z kolei Ryan i in. [1996] przeprowadzili predykcję po 4 dniach od urodzenia korzystając z danych 204 osób. Dzięki informacjom z respiratora, parametrom statycznym oraz wartościom AA byli w stanie osiągnąć 90% czułości i 88% specyficzności. W tym samym okresie na podstawie 116 próbek Rozycki i Narla [1996] osiągnęli 73% czułości oraz 83% specyficzności predykcji w 8 godzinie życia. Po 14 dniach osiągnano jeszcze lepsze wyniki – 82% czułości i 89% specyficzności. Autorzy korzystali z wielu parametrów, min. tych otrzymanych z respiratora oraz z badań gazometrycznych. Nieco później Subhedar i in. [1998] opisali predykcję z wykorzystaniem podobnych parametrów, dodając wyniki badań dopplerowskich.

Analizę przeprowadzono na danych 83 osób, 24 godziny po narodzinach, osiągając $AUC=0.89$ (od ang. area under the curve, pole pod krzywą ROC). W tym samym roku Romagnoli i in. [1998] na podstawie danych 149 osób opisali predykcję po trzecim dniu z wynikiem 97% specyficzności i 92,7% czułości ($AUC=0,97$). Po siódmym dniu czułość pozostawała bez zmian, natomiast specyficzność obniżyła się do 95,5%. Autorzy wykorzystali niektóre dane statyczne oraz te pobrane z respiratora. Już po wspomnianej zmianie definicji BPD w 2001 r. Kim i in. [2005] opublikowali, że korzystając z 10 różnych statycznych parametrów oraz danych z respiratora 197 osób, po 7 dniach byli w stanie osiągnąć 88.9% czułości oraz 86.2% specyficzności ($AUC=0,91$). Z kolei Bhering i in. [2007] na podstawie wieku płodowego, wystąpienia PDA, faktu użycia respiratora oraz spadku masy ciała 247 pacjentów, po 7 dniach byli w stanie przewidzieć wystąpienie PDA z trafnością równą 93,7% ($AUC=0,906$). Znacznie później Laughon i in. [2011] przeprowadzili zakrojone na szeroką skalę badania analizujące wiek płodowy, masę urodzeniową, płeć, pochodzenie etniczne, fakt użycia respiratora oraz FiO_2 aż 3636 osób. W 7 dniu życia osiągnięto $AUC=0,771$ i wynik ten rósł z czasem aż do 0,823 w 28 dniu. Wszystkie wymienione dotychczas wyniki zostały osiągnięte przy zastosowaniu regresji logistycznej. Jedynym wyjątkiem okazał się Stoch [2007], który opracował porównanie wyników predykcji BPD metodą klasyczną (RL) z innymi, osiągniętymi dzięki zastosowaniu przez niego sztucznych sieci neuronowych. Dla regresji osiągnięta trafność wynosiła ok. 90% ($AUC=0,95$), natomiast wśród sieci neuronowych najlepsze wyniki zapewniły sieci radialne dając ok. 93% trafności ($AUC=0,95$). Ponadto badacz położył duży nacisk na sposób doboru parametrów. Wszystkie inne prace opierają się na prostym doborze cech poprzez wyznaczenie ich istotności statystycznej oraz ewentualnie w niektórych przypadkach eliminacji wstecznej lub postępującej. Jak udowodnił autor, zastosowanie bardziej wysublimowanych metod selekcji zmiennych niezależnych ma znaczący wpływ na osiągnięte wyniki predykcji.

Cele pracy

Przytoczone wyniki sprawiają wrażenie często tak obiecujących, że wydawać by się mogło, że problem predykcji BPD można uznać za rozwiązany. Jak jednak zauważają Laughon i in. [2011], z pośród wielu proponowanych modeli predykcyjnych żaden nie mógł zostać wprowadzony do użytku w codziennej praktyce lekarskiej. Dla przykładu Ryan i in. [1996] wspominają o braku możliwości pomiaru parametru AA u wszystkich pacjentów, nawet na potrzeby własnych badań. Jak piszą, najczęściej przeszkodą są kuriozalne problemy organizacyjne czy też techniczne. Tymczasem Subhedar i in. [1998] zaznaczają, że zaproponowany przez nich model może być niemożliwy do zastosowania w praktyce ze względu na należące do rzadkości wykonywanie pomiarów parametru AT:RVET(c). Ponadto wyrażają oni obawy co do podejmowania decyzji klinicznych dla konkretnych pacjentów na podstawie prezentowanego modelu. Ich uwagę zwraca fakt, że wszystkie dane pochodziły z badań w jednej placówce medycznej. Nie ma więc pewności co do możliwości uogólnienia zaobserwowanych zależności poza badanym środowiskiem. Co ważniejsze, patrząc na przytaczane wcześniej wyniki widać, że badacze nie są zgodni co do tego, jakie parametry dają jednoznacznie najlepsze wyniki predykcji. Jak zauważają Laughon i in. [2011] istotność parametrów jest inna zależnie od tego, w którym dniu po porodzie przeprowadzamy predykcję. Co więcej, autorzy potwierdzają powyższe obawy twierdząc, że wyniki predykcji mogą być zależne od przyjętych w danej jednostce opieki zdrowotnej procedur medycznych. W dodatku niemałym problemem są różne możliwości techniczne placówek co do wykonywania pomiarów, które są wymagane do zastosowania proponowanych przez badaczy modeli. Nawet w wypadku wielu szpitali klinicznych jest to poważna bariera, nie wspominając już o mniejszych jednostkach, w których często pacjent spędza pierwsze dni lub w najlepszym wypadku godziny życia. Oprócz tego, badając przytoczoną historię predykcji BPD narzucają się pytania dotyczące monogeniczności stosowanych metod. Zastosowanie regresji

logistycznej w tym kontekście jest dość dobrze zbadane, jednak brak jest szerszej perspektywy będącej punktem odniesienia do wspomnianych wyników. Jedyne Stoch [2007] wprowadza pewien kontrast w postaci zastosowania sztucznych sieci neuronowych. Wajs i in. [2016b] opracowali co prawda autorską metodę klasyfikacji opartą na analizie statystyk, natomiast ma ona zastosowanie tylko dla dwóch zmiennych i przy bardzo ograniczonym zbiorze uczącym. Ciekawym byłoby natomiast zastosowanie i porównanie wyników także innych popularnych metod. Z tego też powodu w pracy zdecydowano się na wykorzystanie relatywnie nowego i cenionego algorytmu SVM (ang. Support Vector Machine) [Cortes i Vapnik, 1995], którego wyniki zestawiono z efektami równoległe wykonanych obliczeń RL. Mając na uwadze przytoczone prace badaczy, dostępność danych oraz opinię lekarzy [Kruczek i Kwinta, 2007] zdecydowano się na predykcję po 7 dniu życia pacjenta. Kładąc duży nacisk na optymalny dobór zestawu cech, zbadano wszystkie 2^{14} kombinacje czternastu dostępnych parametrów gromadząc wyniki w bazie danych. Przeprowadzone dogłębne porównanie rezultatów nie dowiodło jednak wyższości którejs z metod lub modeli, wręcz przeciwnie – wykazano komplementarność obu metod. W przypadkach, gdy jeden z algorytmów daje słabsze wyniki np. ze względu na brak części danych, drugi może okazać się godnym zamiennikiem pozwalającym na utrzymanie parametrów predykcji na odpowiednim poziomie. Z drugiej jednak strony, poszerzanie i tak już bogatego wachlarza proponowanych modeli predykcyjnych może prowadzić do zagubienia się w gąszczu dostępnych możliwości. Autor uznał więc za kluczowe opracowanie odpowiedniego systemu wspomagania decyzji systematyzującego wiedzę na temat algorytmów, metod i modeli przewidujących wystąpienie BPD w taki sposób, aby pozwalał on personelowi medycznemu na łatwe uzyskanie informacji adekwatnych do konkretnej sytuacji. Zaprezentowana w pracy aplikacja pozwala użytkownikowi na zdefiniowanie odpowiednich preferencji oraz danych, które są w konkretnym przypadku dostępne. Bazując na pokaźnym zbiorze 33 tys. zbadanych

modeli predykcyjnych, system pozwala na znalezienie najskuteczniejszego, który na podstawie posiadanych przez użytkownika pomiarów jest w stanie przewidzieć diagnozę BPD. Uwzględniany jest przy tym dopuszczony przez użytkownika maksymalny błąd oraz dokonuje się oszacowanie możliwej do osiągnięcia trafności, czułości i specyficzności. Ponadto oprogramowanie informuje, które z parametrów nie są niezbędne dla predykcji BPD i mogłyby być ewentualnie pominięte. Jednocześnie lekarz jest powiadamiany o pomiarach, które warto byłoby dodatkowo wykonać w celu maksymalizacji skuteczności predykcji. Poza tym praca omawia serię eksperymentów, którym został poddany proponowany system wspomagania decyzji. Ich wyniki pozwoliły na dokładniejszą analizę generowanych zaleceń oraz ocenę istotności poszczególnych parametrów dla predykcji BPD.

Teza pracy:

Możliwe jest zbudowanie systemu wspomagnia decyzji dla predykcji dysplazji oskrzelowo-płucnej po pierwszym tygodniu życia wcześniaków w oparciu o metody regresji logistycznej oraz wektorów wspierających (SVM).

Rozdział 1

Dane

1.1 Źródła danych

Badaną grupę stanowili pacjenci Oddziału Intensywnej Terapii Noworodka Polsko-Amerykańskiego Instytutu Pediatrii Collegium Medicum Uniwersytetu Jagiellońskiego. W pracy posłużono się danymi 67 chorych zebranych w latach 2004–2006 przez dr inż. Pawła Stocha wykorzystanymi w jego pracy doktorskiej [Stoch, 2007]. W celu zwiększenia liczebności próby autor niniejszej pracy poszerzył zbiór danych o dodatkowych 42 pacjentów z lat 2006–2010. Nowe dane dynamiczne zostały zebrane dzięki autorskiemu oprogramowaniu stworzonemu specjalnie do tego celu. W 2006 r. w języku C wykonano specjalny driver monitora medycznego Philips M8002A, który umożliwia bezpośredni eksport informacji do centralnej bazy danych szpitala, komunikując się z urządzeniem poprzez sieć ethernet. Natomiast w roku 2008 wykonano odpowiednią aplikację w języku Java o analogicznym działaniu, pozwalającą eksportować w czasie rzeczywistym parametry życiowe chorych ze szpitalnych monitorów General Electric Solar 8000i. Pozostałe dane użyte w pracy pochodzą ze wspomnianej szpitalnej bazy danych, która zawiera historię chorób pacjentów.

1.2 Charakterystyka danych

Pracę oparto na danych pacjentów urodzonych przedwcześnie, przyjętych na oddział nie później niż w drugiej dobie życia. Dodatkowymi warunkami były pełna dostępność danych dynamicznych z pierwszego tygodnia życia, masa urodzeniowa wynosząca 1500 g lub mniej oraz przeżycie do 4 tygodnia. Wśród zebranych danych, 109 pacjentów spełniało powyższe kryteria, z czego u 46 (42%) z nich po czwartym tygodniu życia zdiagnozowano BPD. Do dalszej analizy wybrano wszystkich pacjentów spełniających wspomniane założenia, przyjętych w badanych okresach czasu, co daje podstawy do twierdzenia o reprezentatywności badanej grupy jako podzbioru populacji.

W analizie brano pod uwagę 14 parametrów, które oznaczono symbolami: *bweight*, *aa*, *gage*, *pda*, *respimv*, *surfact*, *high94*, *low85*, *bpmmean*, *bpmmean_tr*, *spO2mean*, *spO2mean_tr*, *spO2dev* oraz *spO2dev_tr*. Wybór powyższych cech podyktowany był ich dostępnością. Autor analizował również parametry gazometryczne [Wajs i in., 2016a], jednak ze względu na bardzo ograniczoną ilość danych nie zdecydowano się na ich użycie w tej pracy. Podjęto także próby zebrania informacji pochodzących z respiratorów (autorowi głównie zależało na parametrze FiO_2). Niestety ze względów techniczno-praktycznych nie było możliwości zapisu takich danych z odpowiednimi znacznikami czasowymi w sposób gwarantujący wiarygodność. Szczegółowe opisy oraz charakterystyki każdego z parametrów zamieszczono poniżej.

1.2.1 Dane statyczne

Tym mianem określamy informacje pobrane ze szpitalnej bazy danych dotyczące wykonanych jednorazowo pomiarów, diagnoz lub wykonanych procedur. Część z nich wykonano zaraz po urodzeniu, inne po przyjęciu dziecka na oddział, jeszcze inne po upłygnięciu pierwszego tygodnia życia.

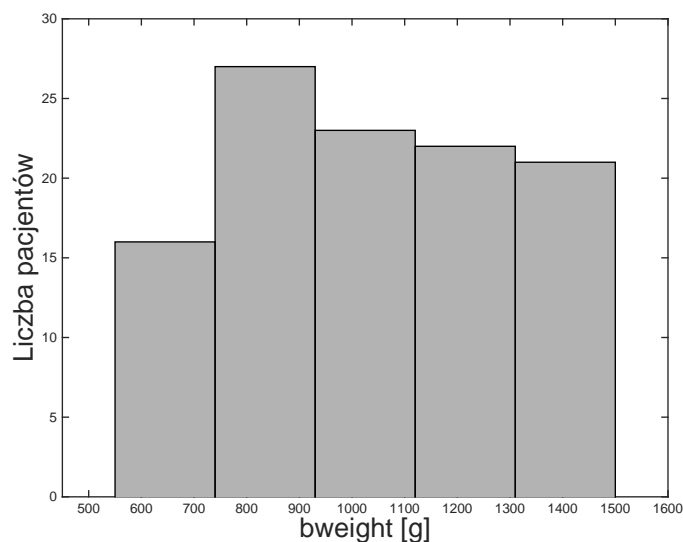
a) Wielkości o wartościach rzeczywistych (ciągłych):

- masa urodzeniowa (ang. birth weight) [ozn. *bweight*].

Jest to jeden z najważniejszych parametrów pozwalających ocenić ogólny stan noworodka. Im wartość jest większa tym lepsze są rokowania dla pacjenta. Zgodnie z klasyfikacją WHO (ang. World Health Organization) [Wardlaw i in., 2005] wśród dzieci urodzonych przedwcześnie z małą masą urodzeniową wyróżnia się następujące grupy:

- mała masa ciała (LBW) 1500–2499 g,
- bardzo mała masa ciała (VLBW) 1000–1499 g,
- skrajnie mała masa ciała (ELBW) 500–999 g.

Dla badanych przypadków wartości masy urodzeniowej zawierały się pomiędzy 550 a 1500g. Jak wspomniano powyżej, w niniejszym opracowaniu nie brano pod uwagę pacjentów o małej masie (LBW), ponieważ u noworodków z powyższej grupy rozwój BPD jest bardzo mało prawdopodobny.



Rysunek 1.1: Histogram masy urodzeniowej

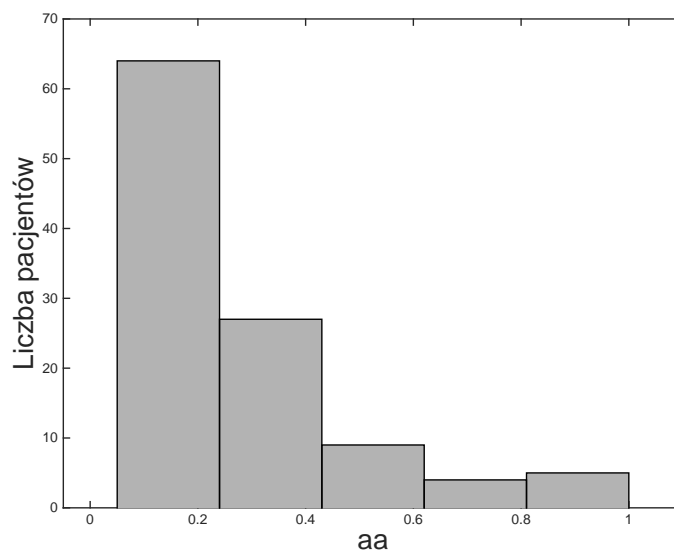
- wskaźnik włośniczkowo-pęcherzykowy (ang. alveolar-arterial ratio) zmierzony jednorazowo podczas przyjęcia na oddział [ozn. aa].

Parametr jest wskaźnikiem jakości procesu oddychania u pacjenta.

Wyznacza się go ze wzoru [Gilbert i Keighley, 1974, Stoch, 2007]:

$$AA = \frac{pO_2}{p_{ATM} \cdot FiO_2 - pCO_2}, \quad (1.1)$$

gdzie: pO_2 to ciśnienie parcjalne tlenu, p_{ATM} to ciśnienie atmosferyczne, pCO_2 to ciśnienie parcjalne dwutlenku węgla, FiO_2 to stężenie tlenu w mieszance oddechowej. Parametr może przyjmować wartości z przedziału $[0,1]$. Wśród badanych wynosił on od 0,05 do 1.



Rysunek 1.2: Histogram wskaźnika włośniczkowo-pęcherzykowego

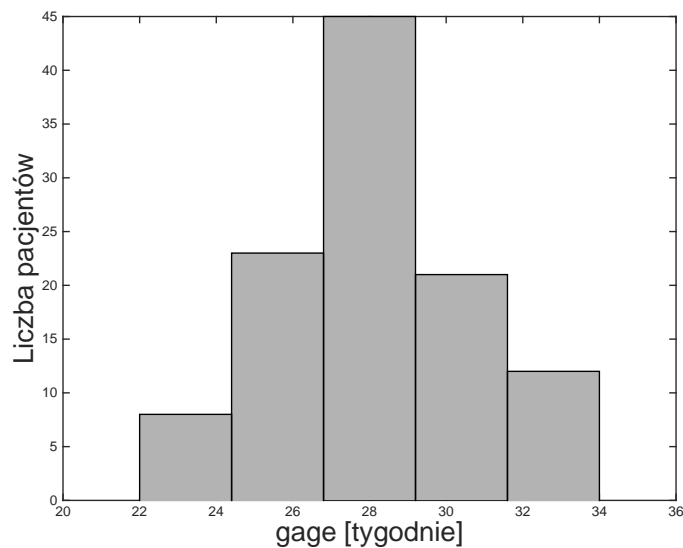
- wiek płodowy (ang. gestational age) [ozn. $gage$].

Jest to kolejny z podstawowych parametrów charakteryzujących ogólny stan noworodka. Za prawidłowy wiek płodowy u człowieka przyjmuje się narodziny między 38 do 42 tygodniem. Poniżej 37 tygodnia dziecko klasyfikowane jest jako wcześniak, przy czym powyżej 34 tygodnia

określane jest często jako "późny wcześniak". Zgodnie z klasyfikacją wyróżnia się cztery stopnie wcześniactwa:

- poród skrajnie przedwczesny <28 tyg.,
- poród bardzo przedwczesny 28–31 tyg.,
- poród umiarkowanie przedwczesny 32–33 tyg.,
- poród miernie przedwczesny 34–36 tyg.

Wśród badanych przypadków wartości parametru zawierały się pomiędzy 22 a 34 tygodniem.



Rysunek 1.3: Histogram wieku płodowego

b) Wielkości dichotomiczne (przyjmujące tylko dwie wartości):

- przetrwały przewód tętniczy (ang. patent ductus arteriosus) [ozn. *pda*].

Jest on naczyniem łączącym tętnicę płucną z łukiem aorty [Yates, 2012], które pozostało z okresu życia płodowego. Normalnie umożliwia ono większości krwi z prawej komory serca omińnięcie płuc, które stawiają większy opór będąc wypełnione płynem. Pozwala to na odciążenie prawej komory serca do czasu porodu. Do 8 godzin po

rozwiązaniu ściana naczyń powinna obkurczyć się samoistnie pod wpływem wzrostu stężenia parcjalnego tlenu we krwi tętniczej i tym samym zamknąć jego światło. wcześniaki są szczególnie narażone na to, że proces ten nie przebiegnie prawidłowo i naczynie pozostanie drożne powodując przeciek pomiędzy lewą i prawą częścią serca. Może to prowadzić do przeciążenia objętościowego lewego przedsionka serca, nadciśnienia płucnego lub nawet do nieodwracalnego stwardnienia płuc (zespół Eisenmengera). Wśród badanej grupy u 47% pacjentów zdiagnozowano PDA.

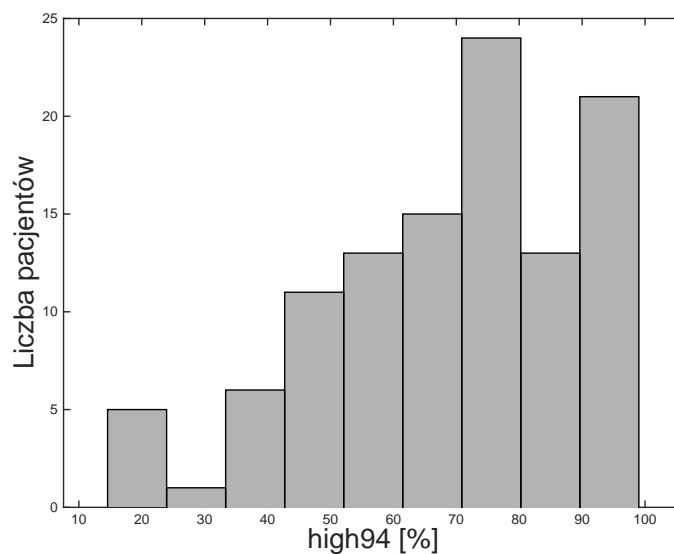
- użycie respiratora w pierwszym tygodniu życia [ozn. *respimv*].
Parametr określa czy wystąpiła potrzeba zastosowania wymuszonej wentylacji mechanicznej przy użyciu respiratora. Zaistnienie takiej sytuacji oznacza ogólnie gorszy stan noworodka. Co więcej zbyt intensywna sztuczna wentylacja zwiększa ryzyko wystąpienia schorzeń płuc, między innymi badanej w tej pracy dysplazji oskrzelowo-płucnej [Ramos i in., 2016]. W 49% z badanych przypadków zalecono wspomnianą procedurę medyczną.
- zastosowanie surfaktantu w pierwszym tygodniu życia [ozn. *surfact*].
Jest to substancja obniżająca napięcie powierzchniowe, tym samym stabilizując pęcherzyki płucne, występująca naturalnie u zdrowych noworodków. Jest ona produkowana przez komórki wyściełające pęcherzyki płucne płodu, począwszy od 26 tygodnia ciąży. U wcześniaków bardzo często występuje jej niedobór, który sprawia zapadanie się i sklejanie pęcherzyków płucnych, powodując niewydolność oddechową i niedobór tlenu we krwi tętniczej. Dlatego też, gdy zachodzi taka potrzeba, substancja jest sztucznie podawana noworodkowi. Wśród badanej grupy w 38% przypadków zaistniała powyższa konieczność.

1.2.2 Dane dynamiczne

Powyższym mianem określamy informacje bazujące na szeregach czasowych uzyskanych podczas całodobowych pomiarów w pierwszym tygodniu życia pacjentów. W niniejszym opracowaniu korzystano z danych dotyczących tętna oraz saturacji tlenem hemoglobiny we krwi. Na ich podstawie wyznaczono odpowiednie parametry brane pod uwagę w algorytmach predykcji:

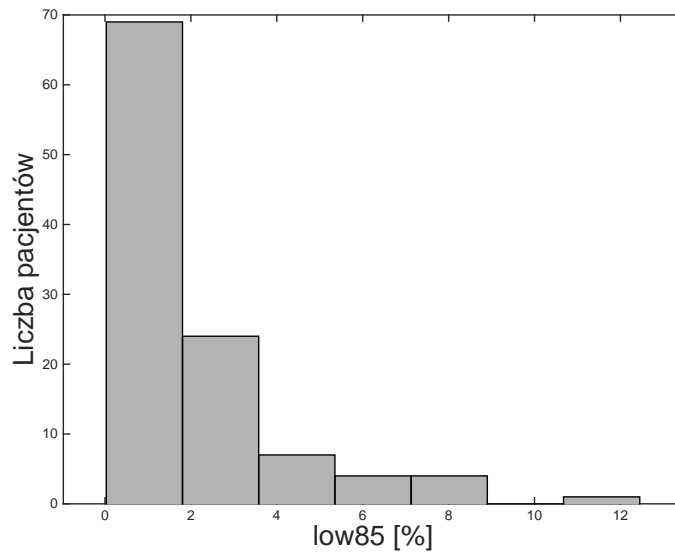
- procent czasu, podczas którego saturacja wynosiła ponad 94% [ozn. *high94*] [Jobe, 2011, Kinsella i in., 2006].

Parametr określa jak często wysycenie hemoglobiny tlenem było w normie. Zaobserwowane wartości zawierały się w przedziale 14,56%–99,02%.



Rysunek 1.4: Histogram parametru high94

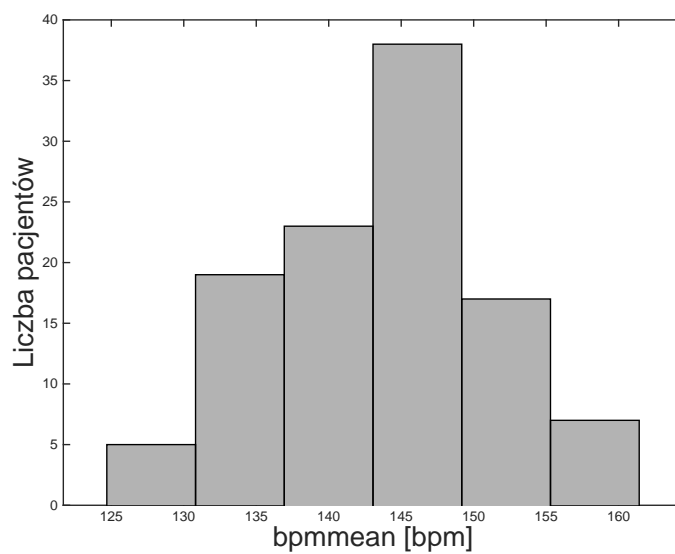
- procent czasu, podczas którego saturacja przyjmowała wartości ekstremalnie niskie, wynosząc mniej niż 85% [ozn. *low85*] [Kinsella i in., 2006]. Zarejestrowano wartości: 0,03%–12,45%.



Rysunek 1.5: Histogram parametru low85

- średnia tygodniowa wartości tętna [ozn. *bpmmean*].

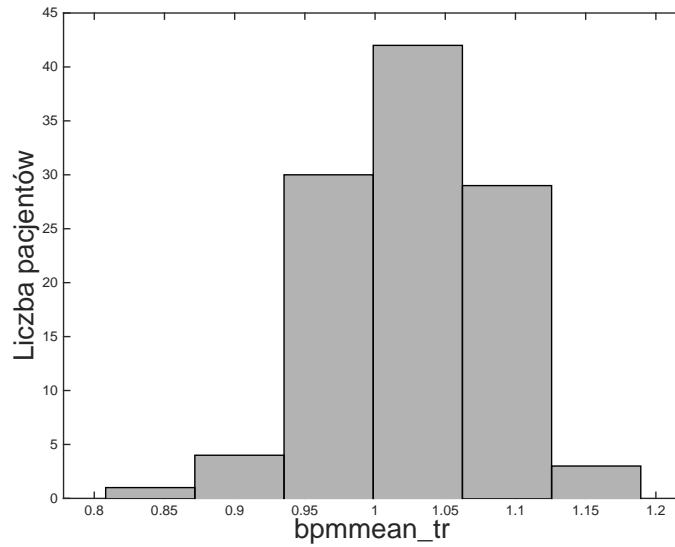
Zakres obserwowanych wartości: 124,69–161,42.



Rysunek 1.6: Histogram uśrednionego tętna

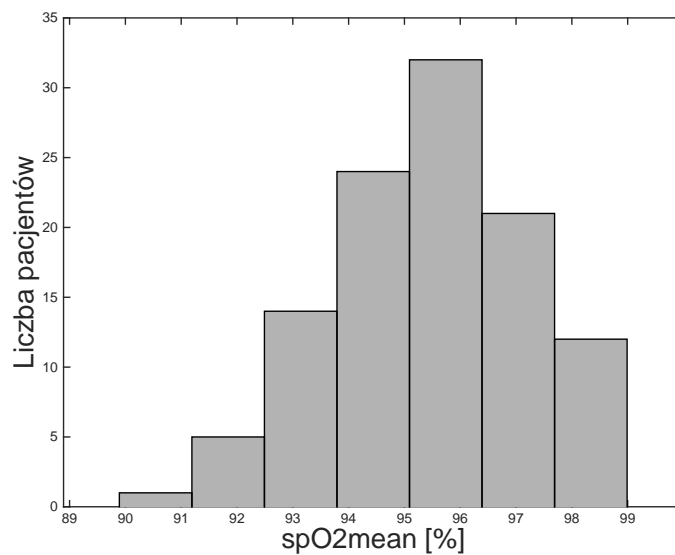
- stosunek wartości średnich tętna z pierwszego tygodnia do pierwszego dnia [ozn. *bpmmean_tr*].

Parametr może być interpretowany jako trend zmian tętna pacjenta, który wynosił od 0,81 do 1,19.



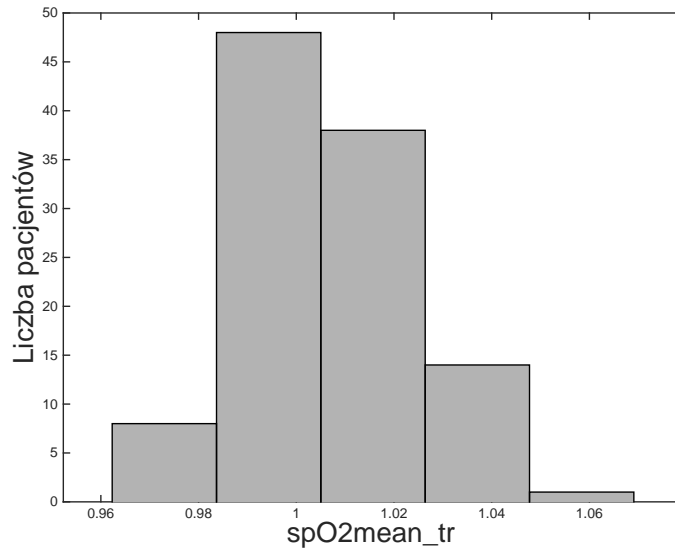
Rysunek 1.7: Histogram trendu tętna

- średnia wartość saturacji hemoglobiny tlenem podczas pierwszego tygodnia [ozn. $spO2mean$]. Zaobserwowano wartości: 89,89%–98,99%.



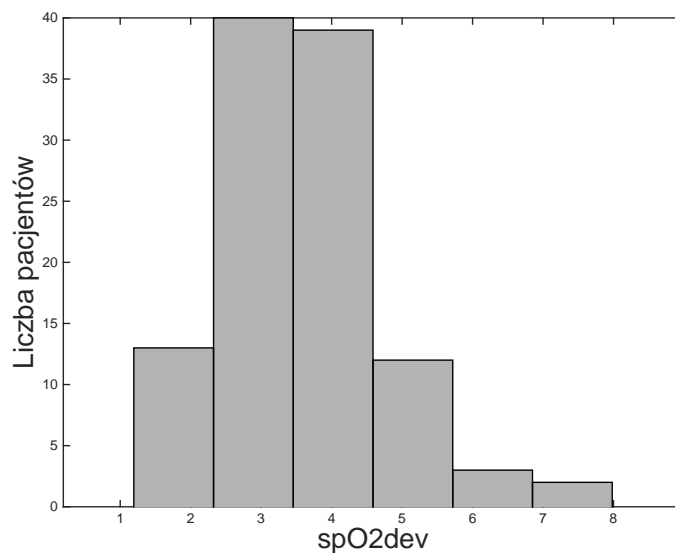
Rysunek 1.8: Histogram uśrednionej saturacji hemoglobiny tlenem

- stosunek wartość średnich saturacji hemoglobiny tlenem z pierwszego tygodnia do pierwszego dnia [ozn. $spO2mean_tr$]. Parametr może być interpretowany jako trend saturacji. Zaobserwowano wartości: 0,96–1,07.



Rysunek 1.9: Histogram trendu saturacji

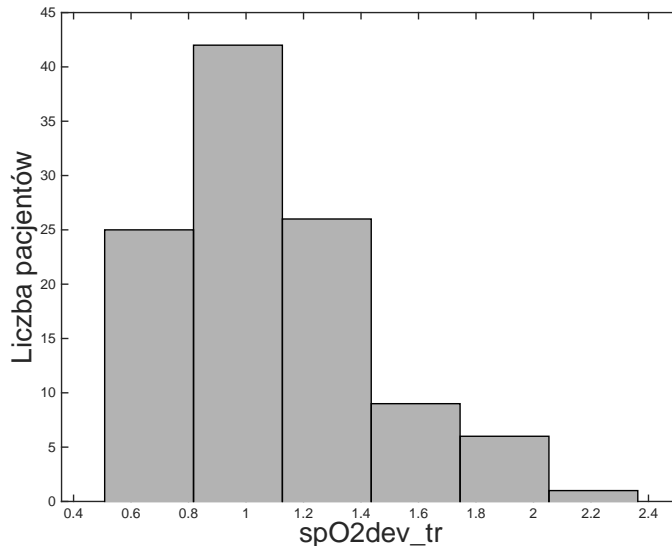
- odchylenie standardowe wysycenia hemoglobiny tlenem podczas pierwszego tygodnia [ozn. $spO2dev$]. Wartości zawierały się w przedziale 1,19–7,98.



Rysunek 1.10: Histogram odchylenia standardowego saturacji

- stosunek odchylenia standardowego wysycenia hemoglobiny tlenem z pierwszego tygodnia do pierwszego dnia [ozn. *spO2dev_tr*]

Wartości zawierały się w przedziale 0.51–2.36.



Rysunek 1.11: Histogram trendu odchylenia standardowego saturacji

1.3 Analiza statystyczna parametrów

W Tab. 1.1 porównano wartości maksymalne, średnie oraz minimalne parametrów przyjmujących wartości rzeczywiste. Ponadto wykorzystując pakiet statystyczny SPSS w wersji 23, przeprowadzono test normalności W Shapiro-Wilka [Shaphiro i Wilk, 1965] dla rozkładów wspomnianych cech. Jako hipotezę zerową przyjęto pochodzenie próbek z populacji o rozkładzie normalnym. Istotność statystyki W na przyjętym poziomie $p \leq 0,05$ oznacza odrzucenie hipotezy zerowej. Jak więc wynika z tabeli, należy odrzucić hipotezę zerową o normalności rozkładu dla parametrów: *bweight*, *aa*, *gage*, *high94*, *low85*, *spO2mean_tr*, *spO2dev*, *spO2dev_tr*.

W Tab. 1.2 przedstawiono natomiast średnie wartości parametrów odpowiednio dla grupy pacjentów ze zdiagnozowaną BPD oraz dla tych bez dysplazji oskrzelowo-płucnej. Ze względu na wykazany wcześniej brak podobieństwa do

rozkładu normalnego zamiast testu T Studenta [Fay i Proschan, 2010] wykonano rangowy test U Manna-Whitney’a [Mann i Whitney, 1947, Tadeusiewicz i in., 1993]. Test ten może być zastosowany tylko do zmiennych o charakterze ciągłym lub porządkowym. Wynik pozwala na określenie, czy istnieją ważne statystyczne różnice parametrów wśród dwóch grup podzielonych ze względu na diagnozę BPD (czy różnica median obu rozkładów jest istotna). Jako hipotezę zerową zakładamy, że dla danego parametru mediany obu grup są równe. Analizując ostatnią kolumnę Tab. 1.2 widać, że hipotezę taką należy odrzucić dla wszystkich parametrów oprócz *bpmmean*, *spO2mean_tr* oraz *spO2dev_tr*. Dla zmiennych o charakterze dychotomicznym wykonano natomiast testy chi-kwadrat (ang. Pearson’s chi-squared test) oraz Fishera (ang. Fisher’s exact test) [Ludbrook, 2008, Moczko i in., 1998]. Jako hipotezę zerową założono identyczny rozkład cech w obu grupach. Oba testy dały zgodne ze sobą wyniki – nie było podstaw do jej odrzucenia jedynie dla parametru *respimv*. Oznacza to, że pozostałe nie wytłuszczone w Tab. 1.2 parametry posiadają statystycznie istotne różnice rozkładów pomiędzy grupą pacjentów ze zdiagnozowaną BPD, a pozostałymi badanymi. Parametry te można więc rozważać jako potencjalnie dobre predyktory schorzenia.

Tablica 1.1: Charakterystyka rozkładów parametrów ciągłych

Parametr	Wartość minimalna	Wartość średnia	Wartość maksymalna	test W S-W p
Masa urodzeniowa [g]	550	1033	1500	0,001
Wskaźnik <i>aa</i>	0,05	0,26	1,00	<0,001
Wiek płodowy [tydz.]	22	28	34	0,007
Parametr <i>high94</i> [%]	14,56	69,51	99,02	0,001
parametr <i>low85</i> [%]	0,03	1,99	12,45	<0,001
Średnie tętno [bpm]	124,69	143,39	161,42	0,449
Trend tętna	0,81	1,03	1,19	0,207
Średnia saturacja [%]	89,89	95,42	98,99	0,453
Trend saturacji	0,96	1,00	1,07	<0,001
Odchylenie std. saturacji	1,19	3,54	7,98	0,001
Trend odchylenia std. saturacji	0,51	1,09	2,36	<0,001

Porównanie wartości maksymalnych, średnich i minimalnych parametrów ciągłych, w ostatniej kolumnie wynik testu normalności W Shapiro-Wilka. Wartości wytłuszczone ($p > 0,05$) oznaczają rozkład normalny parametru.

Tablica 1.2: Porównanie grup z pozytywną i negatywną diagnozą BPD

Parametr	Pacjenci z BPD	Pacjenci bez BPD	p
Masa urodzeniowa – <i>bweight</i> [g]	891	1136	<0,001
Wskaźnik <i>aa</i>	0,19	0,32	<0,001
Wiek płodowy – <i>gage</i> [tydz.]	26,6	29,1	<0,001
Wystąpienie PDA – <i>pda</i> [%]	70	30	0,001
Zastosowanie respiratora – <i>respimv</i> [%]	54	44	0,34
Podanie surfaktantu – <i>surfact</i> [%]	57	24	<0,001
Parametr <i>high94</i> [%]	59,68	76,68	<0,001
Parametr <i>low85</i> [%]	2,95	1,29	<0,001
Średnie tętno – <i>bpmmean</i> [bpm]	145	141	<0,001
Trend tętna – <i>bpmmean_tr</i>	1,02	1,03	0,17
Średnia saturacja – <i>spO2mean</i> [%]	94,51	96,08	<0,001
Trend saturacji – <i>spO2mean_tr</i>	1,00	1,01	0,33
Odchylenie std. saturacji – <i>spO2dev</i>	4,12	3,12	<0,001
Trend odchylenia std. saturacji – <i>spO2dev_tr</i>	1,12	1,08	0,78

Porównano wartości średnie parametrów wśród pacjentów z pozytywną oraz negatywną diagnozą BPD. Dla wielkości dychotomicznych (tj. PDA, podanie surfaktantu, zastosowanie respiratora) podano procent wystąpienia wartości pozytywnej, zamiast wielkości średnich. W ostatniej kolumnie podano wartości *p* osiągnięte dla: testu U Manna-Whitney’a – dla zmiennych o wartościach ciągłych oraz testu Fishera – dla zmiennych dychotomicznych. Przez wytłuszczenie zaznaczono zmienne, dla których nie stwierdzono statystycznie istotnych różnic w rozkładach pomiędzy grupami z i bez BPD.

Na prawidłowe wyniki regresji logistycznej negatywny wpływ ma współliniowość zmiennych niezależnych. Zjawisko takie zachodzi, gdy wzajemna korelacja pomiędzy dwoma parametrami jest zbyt duża. Wyróżnia się także zjawisko wielowspółliniowości, kiedy to jedna z rozważanych cech jest liniową kombinacją kilku innych. W obu przypadkach prowadzi to do niestabilności płaszczyzny wyznaczonej przez cechy, czego efektem mogą być nieprzewidywalne wyniki metod regresyjnych. W celu sprawdzenia czy podobna zależność zachodzi obliczono współczynniki VIF (ang. variance inflation factor) [Larose, 2006] zgodnie ze wzorem:

$$VIF_i = \frac{1}{1 - R_i^2}, \quad (1.2)$$

gdzie R_i^2 nazywamy współczynnikiem wielokrotnej determinacji dla regresji liniowej i -tej zmiennej na pozostałe zmienne niezależne modelu [Draper i Smith, 2014, Stoch, 2007]. Wskazuje on, jaka część zmienności cechy zależnej została wyjaśniona poprzez model regresyjny. Wartość współczynnika wyznacza się ze wzoru:

$$R_i^2 = \frac{\sum_{k=1}^m (\hat{x}_{i,k} - \bar{x}_i)}{\sum_{k=1}^m (x_{i,k} - \bar{x}_i)}, \quad (1.3)$$

przy czym $x_{i,k}$ są wartościami i -tej ($i = 1 \dots n$) cechy pacjenta k , $\hat{x}_{i,k}$ jest odpowiednią wartością cechy wyznaczoną poprzez regresję liniową przy wykorzystaniu pozostałych $n - 1$ cech różnych od i , \bar{x}_i to średnia arytmetyczna cechy i obliczona dla wszystkich m pacjentów, m jest liczbą pacjentów, a n liczbą cech. Prezentowane w Tab. 1.3 wartości VIF można interpretować jako miarę wzrostu wariancji szacowanego współczynnika regresji, spowodowanego współliniowością danej zmiennej objaśniającej z innymi. Mimo braku konkretnego kryterium oceny, przyjmuje się, że $VIF \geq 5$ może być traktowany jako wskaźnik występowania wspomnianego problemu [Larose, 2006]. Wyniki wyraźnie wskazują na to, że parametry *low85*, *spO2mean*, *spO2dev* oraz *high94* są współliniowe. Silny związek pomiędzy *spO2mean* i *high94* oraz *low85* i *spO2dev* można niewątpliwie za-

Tablica 1.3: Wyniki testu współliniowości zmiennych niezależnych

Parametr	wartość VIF	wartość VIF*
<i>bpmmean_tr</i>	1,085	1,069
<i>bpmmean</i>	1,265	1,205
<i>spO2dev_tr</i>	1,276	1,258
<i>spO2mean_tr</i>	1,312	1,293
<i>respimv</i>	1,511	1,231
<i>surfact</i>	1,519	1,491
<i>aa</i>	1,787	1,749
<i>bweight</i>	2,204	2,141
<i>gage</i>	2,360	2,205
<i>spO2dev</i>	6,699	2,078
<i>low85</i>	7,745	–
<i>high94</i>	18,744	1,902
<i>spO2mean</i>	23,833	–

Cechy uszeregowano według wartości parametru VIF. Wartości większe od 5 zostały wytłuszczone, ponieważ wskazują one na wystąpienie współliniowości z innymi parametrami. Przez VIF* oznaczono ten sam współczynnik wyznaczony po wykluczeniu parametrów *low85* oraz *spO2mean*.

obserwować w Tab. 1.4, w której zamieszczono współczynniki wzajemnej korelacji Pearsona wszystkich parametrów. Dla wspomnianych cech odpowiednio wynoszą one aż 0,967 i 0,894. Dla potwierdzenia powtórnie obliczono współczynniki VIF* w Tab. 1.3, tym razem z wykluczeniem po jednej z cech ze wspomnianych powyżej par (w przykładzie wykluczone cechy wybrano losowo). Zgodnie z oczekiwaniami eliminując parametry *low85* oraz *spO2mean*, wartości VIF* są znacznie niższe od założonego progu równego 5. Można więc wnioskować, że problem współliniowości przestaje wtedy występować. Ponadto uwagę może zwrócić fakt, że widoczne w Tab. 1.4 parametry *gage* oraz *bweight* są także wyjątkowo mocno powiązane – współczynnik korelacji wynosi dla nich aż 0,750.

Tablica 1.4: Wartości współczynnika korelacji Pearsona pomiędzy parametrami

	bweight	gage	respimv	aa	pda	surfact	spO2mean	spO2dev	low85	high94	bpmmean	spO2mean_tr	spO2dev_tr	bpmmean_tr
bweight	1,000	0,750	-0,160	0,078	-0,121	-0,121	0,216	0,002	-0,002	0,198	0,023	0,255	-0,189	0,028
gage	0,750	1,000	-0,276	0,216	-0,100	-0,077	0,325	-0,064	0,039	0,324	-0,103	0,288	-0,226	-0,095
respimv	-0,160	-0,276	1,000	-0,624	0,296	0,453	-0,400	0,341	0,282	-0,346	0,047	0,025	0,024	-0,038
aa	0,078	0,216	-0,624	1,000	-0,229	-0,489	0,503	-0,457	-0,374	0,477	-0,084	0,006	-0,010	0,131
pda	-0,121	-0,100	0,296	-0,229	1,000	0,378	-0,126	0,231	0,296	-0,125	-0,000	0,170	-0,057	-0,073
surfact	-0,121	-0,077	0,453	-0,489	0,378	1,000	-0,339	0,430	0,421	-0,277	0,035	0,045	-0,068	0,009
spO2mean	0,216	0,325	-0,400	0,503	-0,126	-0,339	1,000	-0,473	-0,510	0,967	-0,139	0,209	-0,080	0,019
spO2dev	0,002	-0,064	0,341	-0,457	0,231	0,430	-0,473	1,000	0,894	-0,425	0,273	0,018	0,072	0,094
low85	-0,002	0,039	0,282	-0,374	0,296	0,421	-0,510	0,894	1,000	-0,448	0,109	-0,029	0,078	0,046
high94	0,198	0,324	-0,346	0,477	-0,125	-0,277	0,967	-0,425	-0,448	1,000	-0,184	0,192	-0,112	0,072
bpmmean	0,023	-0,103	0,047	-0,084	-0,000	0,035	-0,139	0,273	0,109	-0,184	1,000	0,131	-0,009	0,123
spO2mean_tr	0,255	0,288	0,025	0,006	0,170	0,045	0,209	0,018	-0,029	0,192	0,131	1,000	-0,365	-0,197
spO2dev_tr	-0,189	-0,226	0,024	-0,010	-0,057	-0,068	-0,080	0,072	0,078	-0,112	-0,009	-0,365	1,000	0,113
bpmmean_tr	0,028	-0,095	-0,038	0,131	-0,073	0,009	0,019	0,094	0,046	0,072	0,123	-0,197	0,113	1,000

Wytłuszczonym drukiem zaznaczono parametry silnie skorelowane, o współczynniku korelacji powyżej 0,65.

Rozdział 2

Metody predykcji

W rozdziale przedstawiono opis metod predykcji zastosowanych w prezentowanym systemie wspomagania decyzji. Pierwszą z nich jest regresja logistyczna (*RL*), na której to opiera się gros publikacji podejmujących próby predykcji BPD. Głównym powodem tej popularności jest zapewne prostota i niewymagana głębsza wiedza matematyczna, co jest niezaprzeczalnym atutem dla osób o wykształceniu medycznym. Jedynym zanotowanym przez autora wyjątkiem jest zastosowanie sztucznych sieci neuronowych w [Stoch, 2007]. Nie odnotowano natomiast w literaturze prób zastosowania do tego celu popularnej w innych dziedzinach metody wektorów wspierających (*SVM*). Dlatego też pierwszymi pracami wspominającymi o takim zastosowaniu były [Ochab i Wajs, 2014a,b, 2016]. Z tego też powodu w niniejszej pracy położno duży nacisk na porównanie wyników otrzymanych metodą SVM z tymi otrzymywanymi przy użyciu RL.

Przed zastosowaniem rozważanych metod wszystkie dane zostały znormalizowane do przedziału $[-1,1]$.

2.1 Regresja logistyczna

Statystyczne metody regresji stosuje się w celu badania wzajemnych związków pomiędzy wartościami odpowiednich cech i późniejszego przewidywania wartości jednej z nich na podstawie pozostałych znanych. Zmiennymi niezależnymi (objaśniającymi) nazywamy te z nich, których wartości znamy i potrafimy zmierzyć, natomiast zmienną zależną (objaśnianą) nazywamy tą cechę, której wartości chcemy przewidywać. Problemem obliczeniowym, który należy rozwiązać jest odpowiednia estymacja parametrów modelu regresji dla danego zbioru danych. Jednym z takich modeli (w tym wypadku nieliniowym) jest regresja logistyczna, która może zostać użyta w sytuacji, gdy zmienna objaśniana występuje w skali dychotomicznej (przyjmuje tylko dwie wartości). Właśnie ta właściwość jest powodem popularności tej metody w medycynie, gdzie najczęściej przewidywana jest diagnoza pozytywna lub negatywna. Drugą ważną cechą jest brak jakichkolwiek założeń względem charakteru czy rozkładu zmiennych objaśniających, co pozwala na jej zastosowanie bez głębszej analizy.

W rozważanym modelu prawdopodobieństwo p_k pozytywnej diagnozy BPD dla k -tego pacjenta tj. wartość zmiennej zależnej $\tilde{y}_k^1 = 1$, pod warunkiem, że wektor zmiennych objaśniających tj. wartości odpowiednich pomiarów (cech) wykonanych na pacjencie będzie równy $X_k = (x_{1,k}, x_{2,k}, \dots, x_{n,k})$, definiujemy jako [Hosmer Jr i in., 2013]:

$$p_k = P(\tilde{y}_k = 1 | X_k) = \frac{e^{a_0 + \sum_{i=1}^n a_i x_{i,k}}}{1 + e^{a_0 + \sum_{i=1}^n a_i x_{i,k}}}, \quad (2.1)$$

¹oznaczenia \tilde{y}_k oraz y_k są stosowane zamiennie zależnie od tego, czy we wzorach za diagnozę negatywną odpowiada wartość 0 czy -1, diagnozę pozytywną w obu przypadkach oznaczono przez 1. Zastosowano taką notację, aby nie przekształcać wzorów do mniej czytelnej formy.

gdzie: $x_{i,k}$ są wartościami i -tej cechy pacjenta k , a_i są parametrami modelu regresji, n liczbą rozważanych cech.

W odróżnieniu od regresji liniowej, gdzie zakładamy rozkład normalny zmiennych niezależnych oraz z powodu braku gwarancji co do równości wariancji zmiennych, w celu obliczenia współczynników regresji nie można użyć metody najmniejszych kwadratów. Ze wspomnianych powodów najczęściej używana do tego celu jest metoda największej wiarygodności. Polega ona na maksymalizacji funkcji wiarygodności (L) przy użyciu danych uczących, tj. znalezienie takich współczynników a_i modelu, dla których iloczyn prawdopodobieństw warunkowych p_k (2.1) właściwej diagnozy będzie największy (dla wszystkich danych uczących jednocześnie). Ze względów praktycznych łatwiej rozważyć równoważny problem poszukiwania minimum logarytmu naturalnego wspomnianej funkcji (Bishop [2006]):

$$L = \prod_{k=1}^m p_k^{\tilde{y}_k} (1 - p_k)^{(1 - \tilde{y}_k)}, \quad (2.2)$$

po zlogarytmowaniu

$$\ln(L) = \sum_{k=1}^m [\tilde{y}_k \cdot \ln(p_k) + (1 - \tilde{y}_k) \cdot \ln(1 - p_k)], \quad (2.3)$$

gdzie: k jest numerem pacjenta, $\tilde{y}_k \in \{0, 1\}$ jest właściwą diagnozą dla pacjenta k , m jest liczbą pacjentów w zbiorze uczącym.

Za miarę jakości dopasowania modelu przyjmuje się tzw. dewiancję, którą obliczamy wg. wzoru:

$$DEV = -2 \cdot \ln(L). \quad (2.4)$$

Im wartość jest mniejsza, tym lepsze jest dopasowanie modelu.

Znając optymalne wartości a_i oraz korzystając ze wzoru (2.1) możemy obliczyć wartość szacowanego prawdopodobieństwa pozytywnej diagnozy dla pacjenta X .

Przyjmuje się, że aby model regresyjny mógł być rozważany jako wiarygodny, liczba zmiennych niezależnych powinna być od 10 do 20 razy mniejsza od liczebności zbioru danych [Harrell, 2013]. Przykład liczbowy opisywanej metody zamieszczono w dodatku A.1.

2.2 Support Vector Machine

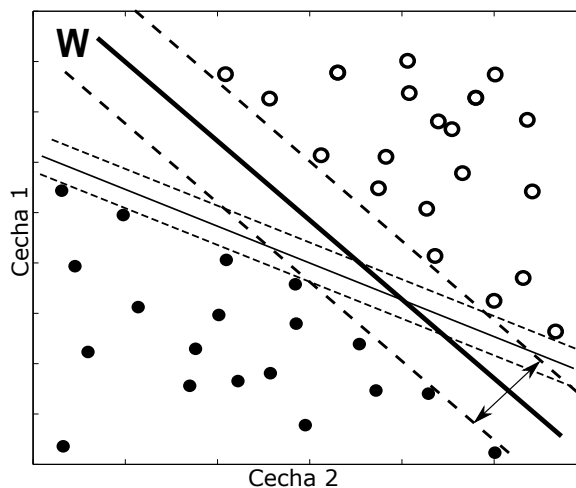
Cortes i Vapnik [1995] zaprezentowali metodę SVM po raz pierwszy w formie zbliżonej do aktualnej. W języku polskim nazywana jest dwojako maszyną wektorów nośnych lub wspierających. Jest klasyfikatorem zakładającym podział danych uczących D na dwie klasy $\{1, -1\}$ [Burges, 1998]:

$$D = \{(X_k, y_k) | X_k \in R^n, y_k \in \{1, -1\}\}_{k=1}^m, \quad (2.5)$$

gdzie m jest liczbą pacjentów w zbiorze uczącym, n jest liczbą rozważanych cech. Poszukiwana jest taka hiperpłaszczyzna W :

$$W \cdot X + b = 0, \quad (2.6)$$

która rozdzieli klasy zapewniając maksymalny możliwy margines między nimi. Przykład zawierający dwie cechy przedstawiono na Rys. 2.1.



Rysunek 2.1: Przykład obrazujący ideę maksymalizacji marginesu pomiędzy klasami rozdzielanymi hiperpłaszczyzną W w metodzie SVM.

Wskazany problem jest równoważny z zagadnieniem minimalizacji funkcji T :

$$T(W, b, \varepsilon) = \frac{\|W\|^2}{2} + c \cdot \sum_{k=1}^m \varepsilon_k, \quad (2.7)$$

przy ograniczeniach:

$$y_k(W \cdot \phi(X_k) + b) \geq 1 - \varepsilon_k, \quad (2.8)$$

gdzie: $\varepsilon_k \geq 0$ jest zmienną relaksującą, $c > 0$ jest współczynnikiem kary za każdy źle sklasyfikowany punkt, ϕ jest funkcją jądrową:

$$K(X_i, X_j) \equiv \phi(X_i) \cdot \phi(X_j). \quad (2.9)$$

W praktyce ze względu na wielowymiarowość wektora W , zazwyczaj rozwiązuje się problem dualny do zdefiniowanego powyżej [Chang i Lin, 2011, Vapnik, 2013]:

$$\min_{\alpha} f(\alpha) \quad (2.10)$$

przy ograniczeniach:

$$Y \cdot \alpha = 0, \quad (2.11)$$

$$0 \leq \alpha_k \leq c, \quad k = 1, \dots, m, \quad (2.12)$$

gdzie:

$$f(\boldsymbol{\alpha}) \equiv \frac{1}{2} \boldsymbol{\alpha}^T \cdot Q \boldsymbol{\alpha} - E^T \cdot \boldsymbol{\alpha} \quad (2.13)$$

oraz $Y = [y_1, \dots, y_m]^T$, $E = [1, \dots, 1]^T$ jest wektorem jedynek, Q jest dodatnio półokreśloną² macierzą o wymiarach $m \times m$, $Q_{ij} = y_i y_j K(X_i, X_j)$.

Powstały dualny problem programowania kwadratowego rozwiązuje się najczęściej metodą Sequential Minimal Optimization (SMO) [Bishop, 2006, Platt i in., 1998]. Po jej zastosowaniu optymalny wektor W można by teoretycznie obliczyć ze wzoru

$$W = \sum_{k=1}^m y_k \alpha_k \phi(X_k), \quad (2.14)$$

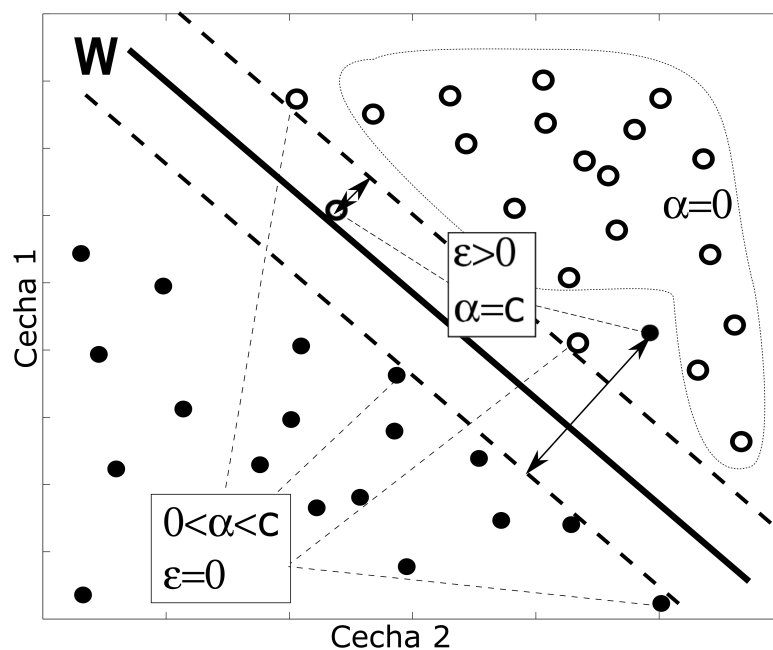
choć w praktyce do klasyfikacji używana jest funkcja decyzyjna

$$f^*(\tilde{X}) = \text{sgn}(W \cdot \phi(\tilde{X}) + b) = \text{sgn}\left(\sum_{k=1}^m y_k \alpha_k K(X_k, \tilde{X}) + b\right), \quad (2.15)$$

pozwalająca na określenie, do której z dwóch klas należy \tilde{X} . Jak widać we wzorze (2.15) wpływ na decyzję klasyfikacyjną mają tylko te X_k , dla których mnożniki Lagrange'a α_i są większe od zera. Właśnie te decydujące wektory nazywamy wektorami wspierającymi (ang. support vectors). Pozostałe X_k , którym odpowiadające $\alpha_k = 0$, są nieistotne dla rozważanego problemu. Jest tak, ponieważ nie leżą one na brzegu klas (Rys. 2.2), a to właśnie punkty brzegowe są kluczowe w procesie klasyfikacji, tj. wyznaczania granicy pomiędzy klasami. Co jednak ciekawe, w konsekwencji dopuszczenia błędu klasyfikacji (poprzez wprowadzenie zmiennych ε_k oraz c , Krzyśko i in. [2008]), wektory wspierające mogą się znajdować nie tylko na granicy marginesu rozdzielającego

²Dotatnie półokreślenie macierzy Q było pierwotnym założeniem twórców metody, natomiast stosuje się przekształcenia jądrowe nie spełniające tego warunku jak, np. używane w tej pracy funkcje sigmoidalne. Jak twierdzą Lin i Lin [2003] zastosowanie takich może być jednak powodem problemów ze zbieżnością algorytmu SMO. Autorzy proponują więc odpowiednie modyfikacje algorytmu, aby uniknąć wspomnianych problemów.

klasy, ale także wewnątrz niego, a nawet poza nim po drugiej stronie płaszczyzny rozdzielającej. Punkty leżące na granicy będą charakteryzować wartości $\varepsilon_k = 0$ oraz $0 < \alpha_k < c$. Pozostałe, leżące wewnątrz marginesu lub po drugiej stronie płaszczyzny będą posiadać $\varepsilon_k > 0$ i wartość tę można rozumieć jako karę proporcjonalną do naruszenia marginesu. Mimo to wektory te mają także wpływ na wynik klasyfikacji, chociaż ze względu na niespełnienie wymogu odpowiedniego marginesu (lub wręcz błędu klasyfikacji) ich wpływ został ograniczony zgodnie ze wzorem (2.12), a więc odpowiadające im wartości $\alpha_k = c$.



Rysunek 2.2: Przykład obrazuje wpływ wektorów wspierających na decyzję klasyfikacyjną. Wektory X_k , którym odpowiadają $\alpha_k = 0$ są „elementami wewnętrznymi” klas, nie są więc wektorami wspierającymi i nie mają wpływu na wynik klasyfikacji. Punkty X_k naruszające margines pomiędzy klasami mają ograniczony wartością c wpływ na wynik klasyfikacji i razem z punktami granicznymi biorą udział w kształtowaniu wyniku klasyfikacji. Wartości ε_k można interpretować jako karę za naruszenie wspomnianego marginesu i jest ona proporcjonalna do odległości od jego granicy.

Za brakującą do obliczenia wartości funkcji decyzyjnej (2.15) stałą b , przyjmujemy średnią arytmetyczną wartości b_k :

$$b_k = -y_k \nabla_k f(\boldsymbol{\alpha}) \quad (2.16)$$

biorąc pod uwagę takie k , które spełniają warunek:

$$0 < \alpha_k < c, \quad k = 1, \dots, m \quad (2.17)$$

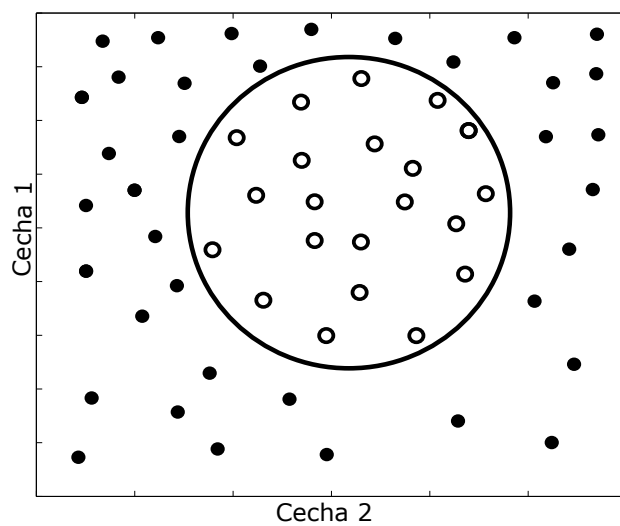
(uwzględniamy tylko punkty leżące na granicach marginesu, ponieważ tworzona płaszczyzna rozdzielająca ma leżeć na jego środku), przy czym $\nabla f(\boldsymbol{\alpha})$ jest gradientem:

$$\nabla f(\boldsymbol{\alpha}) \equiv Q\boldsymbol{\alpha} - E. \quad (2.18)$$

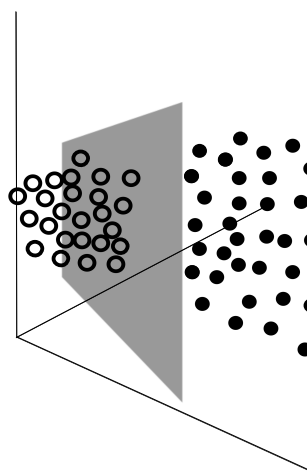
Dzięki użyciu funkcji jądrowej, nawet dla problemów liniowo nieseparowalnych możliwe staje się znalezienie hiperpłaszczyzny W . Funkcja taka zapewnia transformację przestrzeni danych do innej przestrzeni o większej wymiarowości, w której to prawdopodobieństwo separowalności jest znacznie wyższe. Uproszczony przykład takiego przekształcenia przedstawia Rys. 2.3.

Należy zwrócić uwagę na fakt, że w praktyce nigdy nie definiujemy wprost funkcji jądrowej ϕ . W obliczeniach używana jest tylko funkcja K (2.9), która określa jedynie wynik iloczynu skalarnego X_i oraz X_j przetransformowanych do pewnej przestrzeni \mathcal{H} o większej wymiarowości. K może być interpretowana jako miara podobieństwa pomiędzy wektorami wspierającymi (Rys. 2.2), a tymi które chcemy sklasyfikować. Najczęściej używanymi do tego celu funkcjami są:

$$\textit{Sigmoida} : K(X_i, X_j) = \tanh(\gamma X_i^T X_j + r), \quad (2.19)$$



(a) dane w przestrzeni R^2



(b) dane przetransformowane do nowej przestrzeni R^3

Rysunek 2.3: Idea przekształcenia jądrowego umożliwiającego liniową separację danych.

oraz

$$RBF : K(X_i, X_j) = e^{-\gamma \|X_i - X_j\|^2}, \gamma > 0, \quad (2.20)$$

gdzie γ, r są odpowiednio dobranymi współczynnikami przekształcenia jądrowego [Lin i Lin, 2003], a RBF to skrót od ang. radial basis function.

Dzięki właściwości wspomnianej w literaturze jako "kernel trick" [Bishop, 2006, Burges, 1998] hiperpłaszczyzna W nie musi być explicite wyznaczona w celu przeprowadzenia klasyfikacji, dlatego też jak wyżej wspomniano, nie ma potrzeby określania funkcji ϕ , która mapuje dane do nowej (najczęściej nieskończenie wymiarowej) przestrzeni \mathcal{H} [Burges, 1998]. Najwyraźniej widoczne jest to w równaniu (2.15), gdzie ϕ oraz W znika, a w zamian pojawia się K oraz α otrzymana z rozwiązania problemu dualnego, w którym ϕ wcale nie występuje. Ponieważ w pracy korzystano z sigmoidalnej funkcji K (2.19), której odpowiadająca funkcja ϕ istnieje jedynie w przestrzeni nieskończenie wymiarowej, nie było możliwe zamieszczenie wykresów analizowanych danych przed i po transformacji nawet dla tylko dwóch cech³. Z tego też powodu Rys. 2.3 oraz 2.1 nie zawierają prawdziwych danych i należy je traktować jedynie ideowo.

W literaturze, zaprezentowana powyżej klasyczna forma metody klasyfikacji SVM[Vapnik, 2013] bywa nazywana mianem C-SVC, ze względu na występujący we wzorze (2.7) parametr c . Spotyka się także późniejsze jej rozwinięcie, określane symbolem ν -SVC (lub nu-SVC)[Schölkopf i in., 2000], które swoją nazwę zawdzięcza zmiennej ν , którą niejako zastąpiono parametr c przekształcając wzory 2.7 oraz 2.8 do postaci:

$$T(W, b, \varepsilon, \rho) = \frac{\|W\|^2}{2} + \nu \cdot \rho + \frac{1}{m} \sum_{k=1}^m \varepsilon_k, \quad (2.21)$$

³Warto zauważyć, że błędnym jest częste skojarzenie jakoby funkcja K jedynie dodawała kolejne wymiary "uwypuklając" dane.

przy ograniczeniach:

$$y_k(W \cdot \phi(X_k) + b) \geq \rho - \varepsilon_k, \quad (2.22)$$

gdzie $\nu \in (0, 1]$, $\rho \geq 0$. Obie metody działają analogicznie i przy dobraniu odpowiadających sobie parametrów c i ν dają identyczne wyniki. Nowa zmienna posiada jednak dwie specyficzne własności. Jak udowodnili Schölkopf i in. [2000], ν jest jednocześnie:

- dolnym ograniczeniem liczby wektorów wspierających, których liczba wynosi co najmniej νm ,
- górnym ograniczeniem liczby naruszających margines pomiędzy klasami wektorów X_k , których liczba wynosi co najwyżej νm .

W obu przypadkach m jest liczbą elementów w zbiorze uczącym. Korzystając z powyższej metody, można więc intuicyjnie regulować możliwości generalizacji klasyfikatora za pomocą zmiennej ν :

- Wybierając wartości ν bliskie jedności wymagamy istnienia wielu wektorów wspierających i, co za tym idzie godzimy się na wiele naruszeń marginesu oraz większy błąd uczenia. Oznacza to możliwość osiągnięcia szerszego marginesu pomiędzy klasami, ale także dokładniejsze dopasowanie się klasyfikatora do danych i tym samym mniejsze jego zdolności generalizacyjne. Ze względu na dużą liczbę wektorów wspierających model taki będzie znacznie bardziej skomplikowany obliczeniowo.
- Wybierając ν bliskie zeru godzimy się na małą liczbę wektorów wspierających, jednocześnie wymagając, aby margines był możliwie rzadko naruszany. Jest to równoznaczne z większymi możliwościami generalizacji klasyfikatora i węższym marginesem pomiędzy klasami. Także błąd uczenia w tym przypadku będzie mniejszy.

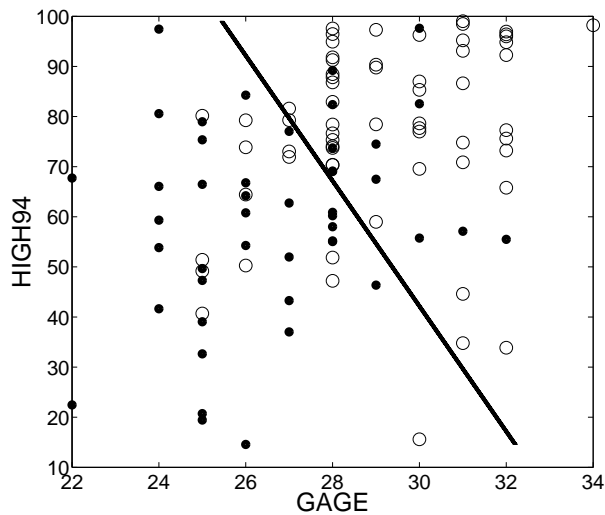
Zmiany parametru c w metodzie C-SVC mają podobny wpływ, nie dają jednak pewności, w jaki dokładnie sposób oddziałują na te właściwości. Z drugiej strony, jak zauważają Chang i Lin [2001], problem zdefiniowany jako ν -SVC często wymaga większego nakładu obliczeń przy rozwiązaniu. Autorzy zaznaczają jednak, że używając C-SVC nakład ten także rośnie dla bardzo dużych wartości c .

W załączniku A.2 zamieszczono przykład liczbowy predykcji BPD z zastosowaniem metody SVM w klasycznej formie (C-SVC).

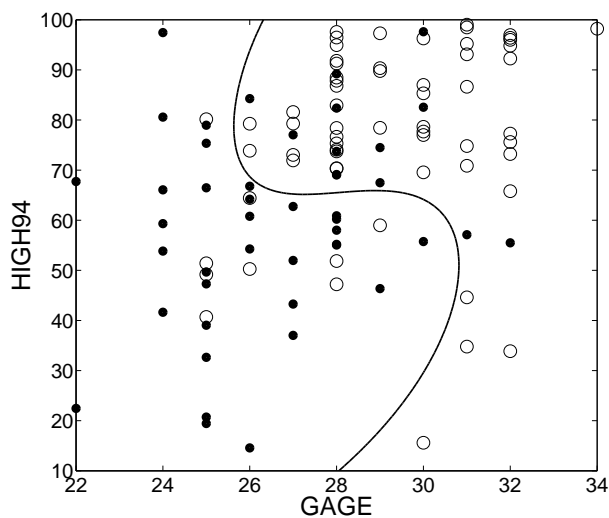
2.3 Porównanie sposobów działania obu metod

Jak omówiono powyżej, metoda regresji logistycznej zwraca prawdopodobieństwo tego, że przyszła diagnoza będzie pozytywna. Aby więc otrzymać funkcję decyzyjną, należy zdefiniować odpowiedni próg prawdopodobieństwa, poniżej którego diagnozy uznamy za negatywne. Zależnie od niego zmieniać będzie się czułość kosztem specyficzności (dokładne definicje tych wartości przedstawiono w następnym rozdziale). W celu oceny takiego modelu często wykonuje się tzw. wykres ROC (ang. receiver operating characteristic, Krzyśko i in. [2008]), który jest zależnością czułości od $(1 - \text{specyficzności})$, gdzie zmieniającym się parametrem jest wartość progu decyzyjnego. Pole pod wykresem ROC, oznaczane symbolem AUC (od ang. area under curve), jest najczęściej używanym parametrem określającym jakość modelu predykcyjnego. Jak jednak przedstawiono powyżej, w wyniku działania klasycznej metody SVM otrzymujemy nie prawdopodobieństwo, a konkretną klasę do której należy badany punkt. W związku z tym, aby mieć możliwość porównania wyników obu metod należy operować pojęciami trafności, czułości oraz specyficzności (przy progu decyzyjnym dobranym tak, aby trafność predykcji była możliwie największa). Dokładną metodykę oceny modeli przedstawiono w kolejnym rozdziale.

Na Rys. 2.4 dla przykładowych dwóch zmiennych przedstawiono poglądowo jak różni się granica podziału klas każdej z metod.



(a) Regresja logistyczna



(b) Support Vector Machine – krzywa jest przybliżeniem hiperpłaszczyzny W przetransformowanej przez funkcję ϕ

Rysunek 2.4: Rezultaty uczenia klasyfikatorów obiema metodami, korzystając z dwóch zmiennych niezależnych ($gage$, $high94$) [Ochab i Wajs, 2014b].

Rozdział 3

Wstępne badania

Proces projektowania systemu wspomagania decyzji został poprzedzony wstępnymi analizami, które miały na celu zbadanie słuszności powstania takiego systemu, jak również wykluczenie zbędnych, a czasowo wymagających obliczeń, których wyniki mogły się okazać mało użyteczne w praktyce. Do wszelkich kalkulacji wykorzystano środowisko Matlab R2013a. Ze względu na małą z perspektywy statystycznej próbkę jaką jest 109 pacjentów (choć w ujęciu medycznym, jak można porównać z pracami wymienionymi we wstępie, jest to całkiem znacząca ilość danych, biorąc pod uwagę trudność zdobycia oraz czas oczekiwania na skompletowanie takiego zbioru) dołożono szczególnych starań, aby proponowane modele predykcyjne posiadały jak największe możliwości generalizacji. Z tego też względu, w celu maksymalnego wykluczenia możliwości, że klasyfikator jedynie „nauczy się na pamięć” znanych przypadków, zdecydowano się na zastosowanie metody zbliżonej do Jackknife [Jones, 1974]: dla każdej badanej kombinacji cech oraz metody, obliczenia powtarzano 30 razy. W każdej takiej iteracji usuwano dane 10 (lub w niektórych przypadkach dla porównania także 30) losowo wybranych pacjentów. Powstały zbiór poddawany był walidacji krzyżowej – każdy pojedynczy pacjent po kolei traktowany był jako testowy, podczas gdy pozostałe dane służyły procesowi uczenia. W ten sposób otrzymano średnie

wartości oraz odchylenia standardowe trafności diagnozy (ACC , ang. accuracy), czułości (TPR , ang. sensitivity) oraz specyficzności (SPC , ang. specificity). Dzięki powyższemu podejściu uzyskano oszacowanie, jak dużego błędu można się spodziewać stosując tak wyuczony model do predykcji u przyszłych pacjentów. Wspomniane parametry obliczono na podstawie następujących wzorów:

$$ACC_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}, \quad (3.1)$$

$$TPR_i = \frac{TP_i}{TP_i + FN_i}, \quad (3.2)$$

$$SPC_i = \frac{TN_i}{TN_i + FP_i}, \quad (3.3)$$

$$ACC = \frac{1}{n} \sum_{i=1}^n ACC_i, \quad (3.4)$$

$$ACC_{dev} = \sqrt{\frac{\sum_{i=1}^n (ACC_i - ACC)^2}{n - 1}}, \quad (3.5)$$

gdzie: TP (ang. True Positives) określa liczbę poprawnie sklasyfikowanych osób z rzeczywistej pozytywnej klasy, FP (ang. False Positives) wskazuje liczbę przypadków błędnie przyporządkowanych, pochodzących z klasy negatywnej, FN (ang. False Negatives) równa się liczbie niewłaściwie ocenionych danych z pozytywnej klasy, TN (ang. True Negatives) oznacza liczbę poprawnie wskazanych przypadków z klasy negatywnej, i jest numerem iteracji Jackknife, $n = 30$ jest liczbą wszystkich iteracji.

W pracy jako główny parametr porównawczy pomiędzy modelami przyjęto ACC . Jednak, podczas gdy trafność diagnozy jest w praktyce procentem właściwych diagnoz, pozostałe dwa parametry są o tyle istotne, że pozwalają na ocenę przydatności modelu w konkretnych specyficznych okolicznościach. W sytuacji, kiedy dużo bardziej niebezpieczna dla zdrowia pacjenta jest pomyłka klasyfikująca go jako zdrowego, mimo że jest on naprawdę chory, niż odwrotnie –

należy zwrócić uwagę na czułość. W przeciwnym wypadku kluczowa będzie specyficzność [Akobeng, 2007].

3.1 Klasyfikatory regresji logistycznej

Ze względu na fakt, że praktycznie cała literatura na temat predykcji BPD (z drobnymi wyjątkami) oparta jest o regresję logistyczną (Rozdz. 2.1), pierwsze obliczenia wykonano właśnie tą metodą korzystając z funkcji *glmfit* oraz *glmval*. Ponieważ wymagany dla metody nakład obliczeniowy jest niewielki oraz nie ma żadnych parametrów, których wartości można by optymalizować dla lepszego wyniku, zdecydowano się na przeliczenie wszystkich 2^{14} kombinacji rozważanych parametrów. Otrzymany zbiór wyników był bardzo dobrą bazą porównawczą, tak do następnie otrzymywanych wyników innymi metodami, jak i rezultatów publikowanych już w literaturze. Na tym etapie najlepszym otrzymanym wynikiem była trafność na poziomie 82,79% dla modelu zawierającego 6 parametrów (średnia wartość $AUC = 0,84$). Co więcej, także wśród modeli 5 oraz 4 elementowych udało się znaleźć takie, które zapewniały średnią trafność powyżej 82%. Natomiast najlepsze modele 2 i 3 parametrowe były w stanie osiągnąć ponad 81% skuteczności. Co warte zauważenia, rozbudowywanie modeli ponad 6 parametrów nie dawało lepszych rezultatów, a wręcz obniżało trafność predykcji. Opisywane w literaturze wyniki predykcji wspominają głównie o współczynniku AUC , a dla dalej badanych modeli nie wyznaczano tego współczynnika (powody przedstawiono w Rozdz. 2.3). Z tego względu zdecydowano się na jego obliczenie dla najlepszego z modeli regresji logistycznej, tworząc w ten sposób swoisty punkt odniesienia. Wnikliwy czytelnik zauważył zapewne, że prezentowane we wstępie wyniki predykcji po 7 dniu życia pacjenta są nieznacznie lepsze – Romagnoli i in. [1998]: $SPC=95,5\%$, $TPR=92,7\%$, $AUC=0,97$, Kim i in. [2005]: $SPC=86,2\%$, $TPR=88,93\%$, $AUC=0,91$, Bhering i in. [2007]: $ACC=93,7\%$, $AUC=0,91$ oraz najbardziej wiarygodni ze względu na znacznie większą ilość danych Laughon

i in. [2011]: AUC=0,77. Należy mieć jednak na uwadze, że we wszystkich wymienionych pracach modele zawierały dane z respiratora. Najważniejszym wśród nich był parametr FiO_2 , którego to wartość w 28 dniu jest właściwie równoznaczna z diagnozą BPD. Można więc domniemywać, że brak tego potencjalnie ważnego parametru w prezentowanej pracy może być wyjaśnieniem nieznacznie słabszych wyników.

3.2 Modele oparte o SVM środowiska Matlab

W kolejnym etapie badań wykorzystano biblioteczną funkcję *svmtrain* pakietu Matlab implementującą metodę SVM (Rozdz. 2.2, dla odróżnienia użycie tej implementacji w dalszej części pracy oznaczane jest symbolem M.SVM). Ze względu na dużą złożoność obliczeniową i co za tym idzie ogromne wymagania czasowe wstępnie testowano kilka arbitralnie wybranych modeli, przekształceń jądrowych oraz ich parametrów. Stwierdzono, że najbardziej obiecujące wyniki można otrzymać używając RBF¹ (2.20) jako funkcji jądrowej z parametrem $\sigma = 20$. W środowisku Matlab powyższy parametr zdefiniowano wg. wzoru:

$$\gamma = \frac{1}{2\sigma^2}, \quad (3.6)$$

gdzie γ jest parametrem równania (2.20). Spośród dostępnych metod rozwiązujących postawiony problem programowania kwadratowego, najbardziej obiecujące wyniki wydawał się osiągać Optimization Toolbox, który to generował lepsze wyniki niż inne dostępne algorytmy tj. Sequential Minimal Optimization [SMO] oraz metoda najmniejszych kwadratów (ang. Least squares). Dopiero dla wspomnianej wyżej wstępnie wybranej metody zdecydowano się na wykonanie obliczeń dla wszystkich 2¹⁴ modeli. Otrzymane wyniki predykcji zawierały się pomiędzy

¹Co ciekawe niewiele gorsze wyniki udawało się uzyskać stosując proste funkcje wielomianowe.

40,37%, a 78,90% trafności (dla porównania wyniki RL to 57,63% do 82,79%).

Algorytm 1 Procedura poszukiwania najlepszego modelu M.SVM poprzez optymalizację parametru σ [Ochab i Waajs, 2014b]

```
for each 120_best_feature_combinations do
  for  $\sigma = 10, 15, 20, 25, 30$  do
    for 1...30 do
      patients = delete_10_random(all_patients);
      for each test_patient of patients do
        learn_patients = patients - test_patient;
        predict(test_patient, learn_patients);
      end for
      calculate ACC, TPR, SPC;
    end for
    for ACC, TPR, SPC do
      calculate mean_value and dev;
    end for
  end for
  best_results[feature_combination]=best(meanACC);
end for
```

Ze względu na ograniczone zasoby mocy obliczeniowej dopiero z 2^{14} otrzymanych wyników wybrano 120 najlepszych modeli (których ACC wynosiło 77,06%–78,90%) poddanych dalszej optymalizacji. Dla każdego z wybranych zestawów zmiennych objaśniających obliczenia powtarzano 5 razy, za każdym razem używając innego parametru $\sigma \in \{10, 15, 20, 25, 30\}$. Dokładny algorytm przedstawiono powyżej (Algorytm 1). W celu zbadania wpływu ilości danych uczących na jakość predykcji procedurę powtórzono raz jeszcze, eliminując za każdym razem aż 30 próbek – zamiast 10 (ze 109) podczas stosowania metody Jackknife.

3.3 Porównanie wyników Regresji Logistycznej oraz Matlab SVM

Wyniki otrzymane w metodą *RL* (Rozdz. 3.1) oraz te obliczone dzięki optymalizacji modeli SVM (Rozdz. 3.2) przedstawiono w Tab. 3.1. Jak widać, dzięki poszerzeniu obszaru poszukiwań o zdefiniowaną gamę parametrów σ , poziom skuteczności predykcji SVM podniósł się z 78,90% do 80,15%. Porównując jednak wyniki obu metod widać wyraźnie, że regresja logistyczna znacznie lepiej radzi sobie z prawidłowym przewidywaniem diagnozy BPD. Algorytm ten z łatwością osiąga 82% średniej skuteczności przy wielu różnych konfiguracjach zmiennych niezależnych modelu. Nawet dla najlepszego znalezionego klasyfikatora SVM, składającego się z 3 parametrów i osiągniętego 80,15% skuteczności, metoda RL była w stanie osiągnąć wyższy współczynnik $ACC = 81,29\%$. Ponadto model RL oparty jedynie na dwóch cechach był skuteczniejszy (80,30%) od najlepszego SVM (80,15%). Co dość zaskakujące, na tym etapie badań im większy model support vector machine (pod względem liczby parametrów), tym gorsze obserwowano rezultaty. Analizując wszystkie wyniki (widoczne jest to także w Tab. 3.1) widać, że w znacznej większości przypadków usuwając tylko 10 próbek podczas testu, zamiast 30, osiągnano lepsze wyniki. Może być to przesłanką do twierdzenia, że modele odznaczają się dobrą własnością generalizacji i nie zachodzi efekt przeuczenia (kiedy to klasyfikator „uczy się na pamięć danych”).

3.4 Predykcja z wykorzystaniem LIBSVM

Użycie bibliotecznych funkcji SVM pakietu Matlab nie dało satysfakcjonujących efektów. Jednym z możliwych powodów niepowodzenia mogły być znacznie utrudniona możliwości dostrajania poszczególnych parametrów w wykorzystywanych wzorach (Rozdz. 2.2) oraz ograniczony wpływ na pracę procedur rozwiązujących powstały problem programowania kwadratowego. Z tego też powodu

3.4. Predykcja z wykorzystaniem LIBSVM

Tablica 3.1: Porównanie najlepszych wyników otrzymanych podczas obliczeń z Rozdz. 3.1 oraz Rozdz. 3.2 [Ochab i Wajs, 2014b]

Liczba parametrów	BWEIGHT	GAGE	RESPMV	AA	PDA	SURFACT	SPO2MEAN	SPO2DEV	LOW85	HIGH94	BPMMEAN	BPMMEAN_TR	SPO2DEV_TR	SPO2MEAN_TR	Liczba usuniętych próbek w Jackknife	ACC	TPR	SPC	Metoda	Najlepsze wyniki pod względem ACC
6	•		•	•	•			•			•	•			10	82,79%	84,20%	81,73%	RL	najlepsze wyniki względem ACC
															30	80,85%	82,72%	79,36%	RL	
															10	70,01%	64,56%	74,01%	SVM $\sigma=30$	
															30	69,00%	61,91%	74,00%	SVM $\sigma=30$	
6	•			•			•				•	•	•		10	82,74%	85,55%	80,82%	RL	
															30	82,52%	84,87%	80,81%	RL	
															10	67,11%	62,06%	70,73%	SVM $\sigma=30$	
															30	66,71%	58,43%	72,41%	SVM $\sigma=30$	
6	•	•		•	•			•			•				10	82,67%	87,50%	79,09%	RL	
															30	81,12%	85,78%	77,62%	RL	
															10	70,48%	65,77%	73,83%	SVM $\sigma=30$	
															30	69,90%	64,58%	73,69%	SVM $\sigma=30$	
5	•			•				•			•	•			10	82,59%	84,65%	81,07%	RL	
															30	81,46%	81,53%	81,30%	RL	
															10	72,10%	64,12%	77,88%	SVM $\sigma=20$	
															30	71,40%	63,94%	76,69%	SVM $\sigma=30$	
4	•			•				•			•				10	82,01%	84,38%	80,34%	RL	
															30	80,64%	80,08%	81,02%	RL	
															10	75,38%	71,32%	78,31%	SVM $\sigma=15$	
															30	74,68%	69,58%	78,16%	SVM $\sigma=25$	
3	•				•			•							10	81,29%	88,70%	75,90%	RL	
															30	81,05%	88,52%	75,34%	RL	
															10	80,15%	78,73%	81,17%	SVM $\sigma=30$	
															30	79,19%	79,84%	78,71%	SVM $\sigma=15$	
2	•				•										30	81,06%	84,39%	78,62%	RL	
															10	80,30%	83,48%	78,01%	RL	
															30	71,38%	52,67%	84,98%	SVM $\sigma=20$	
															10	71,02%	51,68%	84,90%	SVM $\sigma=30$	
4	•			•			•						•		30	79,69%	78,67%	80,47%	RL	
															10	79,63%	78,11%	80,76%	RL	
															10	79,39%	76,24%	81,67%	SVM $\sigma=25$	
															30	77,50%	72,36%	81,15%	SVM $\sigma=30$	

zdecydowano się na zbadanie możliwości szeroko komentowanej w piśmiennictwie biblioteki LIBSVM [Chang i Lin, 2011]. W przedstawionych obliczeniach zastosowano wersję 3.17 użytą jako skompilowane rozszerzenie środowiska Matlab.

Podczas prób na kilku arbitralnie wybranych zestawach cech stwierdzono, że bardziej obiecujące wyniki osiągnęte są przy zastosowaniu sigmoidy (2.19) jako funkcji jądrowej. Pomimo, że Lin i Lin [2003] sugerują ostrożność w jej użyciu i proponują, aby w miarę możliwości używać funkcji RBF(2.20), w tym przypadku sigmoida generuje trafne wyniki oraz pozwala na dużo szybsze znalezienie odpowiedniej hiperpłaszczyzny separującej klasy. Zdecydowano się na użycie metody C-SVC ze względu na takie sugestie w dokumentacji biblioteki, biorąc jednocześnie pod uwagę, że w praktyce obie dostępne metody C-SVC oraz ν -SVC prowadzą do identycznych wyników (Rozdz. 2.2). Stosując wspomniane zalecenia, przeprowadzono optymalizację parametrów γ oraz c (2.19) wykorzystując wartości należące do odpowiednich zbiorów: $\gamma \in \{2^{-15}, 2^{-14}, \dots, 2^3\}$, $c \in \{2^{-5}, 2^{-4}, \dots, 2^{15}\}$. W dalszych badaniach nie zastosowano natomiast skryptu *grid.py*, proponowanego przez autorów biblioteki do przeprowadzenia optymalizacji parametrów oraz walidacji krzyżowej. Zaobserwowano bowiem, że dla opisywanego problemu, metoda ta pozwala znaleźć modele o bardzo obiecującej trafności wyników, sięgającej nawet 83–86%. Niestety charakteryzują się one znaczącą niestabilnością – nawet bardzo drobne zmiany w zbiorze danych, takie jak usunięcie lub dodanie pojedynczej próbki, powodowały ogromny spadek skuteczności. Do praktycznych zastosowań w medycynie kluczowym jest znalezienie takich modeli, które będą mogły mieć zastosowanie z dowolnym zbiorem uczącym (wzbogacanym o nowe dane z każdym kolejnym zdiagnozowanym pacjentem) gwarantując skuteczność predykcji w określonym przedziale błędu. Z tego też powodu, w celu znalezienia optymalnego klasyfikatora, zdecydowano się na zastosowanie analogicznego algorytmu jak w Rozdz. 3.2 opartego na opisywanej już metodzie Jackknife i walidacji krzyżowej. Niestety ze względu na wysoką

złożoność obliczeniową takiego rozwiązania koniecznym okazało się zbadanie jedynie 1348 losowo wybranych modeli składających się z 2 do 14 cech. Pomimo zadanych ograniczeń, obliczenia zajęły ponad tydzień komputerowi osobistemu wyposażonemu w 8-rdzeniowy procesor Intel Core i7. Zastosowany algorytm przedstawiono poniżej:

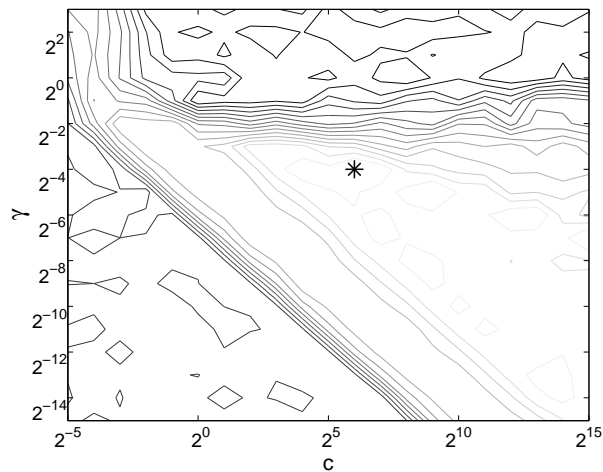
Algorytm 2 Procedura poszukiwania najlepszego modelu LIBSVM poprzez optymalizację parametrów γ i c [Ochab i Wajs, 2016]

```
for each 1348_random_feature_combinations do
  for  $\gamma = 2^{-15}, 2^{-14}, \dots, 2^3 \times c = 2^{-5}, 2^{-4}, \dots, 2^{15}$  do
    for 1...30 do
      patients = delete_10_random(all_patients);
      for each test_patient of patients do
        learn_patients = patients - test_patient;
        predict(test_patient, learn_patients);
      end for
      calculate ACC, TPR, SPC;
    end for
    for ACC, TPR, SPC do
      calculate mean_value and dev;
    end for
  end for
  best_results[feature_combination]=best(meanACC);
end for
```

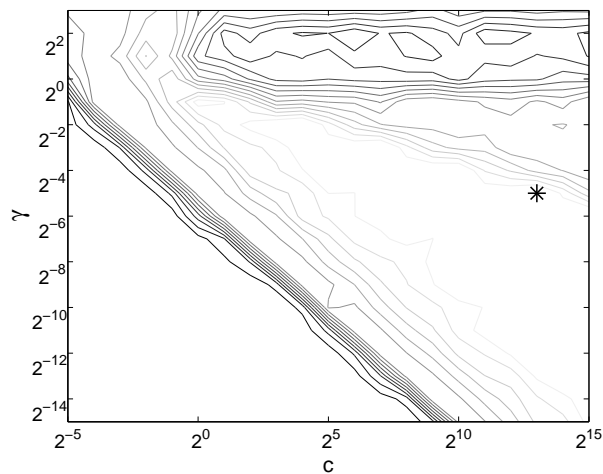
Na Rys. 3.1 przedstawiono wykresy obrazujące proces przeprowadzonej optymalizacji współczynników.

3.5 Porównanie wyników Regresji Logistycznej oraz LIBSVM

Najlepsze rezultaty otrzymane obiema metodami zaprezentowano w Tab. 3.2, te wspomniane w poniższym tekście pogrubiono. Dla każdego z generujących obiecujące wyniki zestawów cech porównano wartości *ACC*, *TPR* oraz *SPC*, zaznaczając odpowiednio wpływ liczby eliminowanych próbek przez metodę Jack-



(a) model 8-parametrowy



(b) model 5-parametrowy

Rysunek 3.1: Przykładowe optymalizowane powierzchnie wg. Algorytm 2, gdzie γ jest współczynnikiem sigmoidy (2.19) oraz c jest współczynnikiem kary (2.7). Linie konturów odpowiadają wartościom ACC . Rysunki zostały wykonane dla odpowiednich wytłuszczonych modeli zawartych w Tab. 3.2 [Ochab i Waajs, 2014a].

knife. Ponadto w miejscach, gdzie wyniki z Rozdz. 3.2 były na odpowiednim poziomie, zamieszczono je oznaczając jako M.SVM. Porównanie odchyleń standardowych trafności diagnoz tych modeli przedstawiono odpowiednio w Tab. 3.3.

3.5. Porównanie wyników Regresji Logistycznej oraz LIBSVM

Tablica 3.2: Porównanie najciekawszych rezultatów klasyfikatorów otrzymanych w wyniku obliczeń prezentowanych w Rozdz. 3.1 oraz 3.4 [Ochab i Wajs, 2014a]

Liczba parametrów	BWEIGHT	GAGE	RESPIMV	AA	PDA	SURFACT	SPO2MEAN	SPO2DEV	LOW85	HIGH94	BPMMEAN	BPMMEAN_TR	SPO2DEV_TR	SPO2MEAN_TR	Liczba usuniętych próbek w Jackknife	ACC	TPR	SPC	Metoda	Najlepsze wyniki pod względem ACC
8	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10	82,60%	80,10%	84,37%	LIBSVM	LIBSVM
															30	79,57%	75,84%	82,17%	LIBSVM	
															10	77,49%	78,62%	76,43%	RL	
															30	76,42%	76,72%	76,09%	RL	
6	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10	82,79%	84,20%	81,73%	RL	RL
															30	80,85%	82,72%	79,36%	RL	
															30	78,49%	72,35%	82,78%	LIBSVM	
															10	78,22%	73,38%	81,77%	LIBSVM	
															10	70,01%	64,56%	74,01%	M.SVM	
30	69,00%	61,91%	74,00%	M.SVM																
6	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10	82,67%	87,50%	79,09%	RL	6-parametry z ACC > 81% dla obu metod
															30	81,12%	85,78%	77,62%	RL	
															10	81,50%	77,98%	84,04%	LIBSVM	
															30	78,22%	72,38%	82,30%	LIBSVM	
															10	70,48%	65,77%	73,83%	M.SVM	
30	69,90%	64,58%	73,69%	M.SVM																
5	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10	82,59%	84,65%	81,07%	RL	5-parametry RL oraz SVM
															10	82,23%	77,13%	85,87%	LIBSVM	
															30	81,46%	81,53%	81,30%	RL	
															30	80,46%	74,31%	84,82%	LIBSVM	
															10	72,10%	64,12%	77,88%	M.SVM	
30	71,40%	63,94%	76,69%	M.SVM																
4	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10	82,01%	84,38%	80,34%	RL	4-parametry RL
															30	80,64%	80,08%	81,02%	RL	
															10	78,78%	72,96%	83,07%	LIBSVM	
															30	77,89%	72,25%	81,97%	LIBSVM	
															10	75,38%	71,32%	78,31%	M.SVM	
30	74,68%	69,58%	78,16%	M.SVM																
4	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10	80,85%	71,95%	87,32%	LIBSVM	4-parametry LIBSVM
															30	80,35%	71,11%	86,96%	LIBSVM	
															10	80,19%	73,27%	85,07%	RL	
															30	79,90%	74,31%	83,94%	RL	
3	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10	81,29%	88,70%	75,90%	RL	3-parametry RL
															30	81,05%	88,52%	75,34%	RL	
															10	80,15%	78,73%	81,17%	M.SVM	
															10	79,83%	78,51%	80,77%	LIBSVM	M.SVM
															30	79,19%	79,84%	78,71%	M.SVM	
															30	79,02%	77,73%	79,68%	LIBSVM	

Tablica 3.2: Porównanie najciekawszych rezultatów klasyfikatorów otrzymanych w wyniku obliczeń prezentowanych w Rozdz. 3.1 oraz 3.4 cd. [Ochab i Wajs, 2014a]

Liczba parametrów	BWEIGHT	GAGE	RESPIMV	AA	PDA	SURFACT	SPO2MEAN	SPO2DEV	LOW85	HIGH94	BPMMEAN	BPMMEAN_TR	SPO2DEV_TR	SPO2MEAN_TR	Liczba usuniętych próbek w Jackknife	ACC	TPR	SPC	Metoda	Najlepsze wyniki pod względem ACC
3		•				•									10	80,57%	78,83%	81,82%	LIBSVM	3-parametry LIBSVM
										•					10	79,66%	78,94%	80,19%	M.SVM	
															30	79,47%	79,70%	79,25%	LIBSVM	
															30	79,43%	79,08%	79,54%	M.SVM	
															10	79,40%	87,20%	73,57%	RL	
															30	79,32%	86,36%	73,98%	RL	
2		•				•									30	81,06%	84,39%	78,62%	RL	2-parametry RL
															10	80,30%	83,48%	78,01%	RL	
															10	73,57%	57,54%	85,22%	LIBSVM	
															30	73,22%	55,87%	85,28%	LIBSVM	
															30	71,38%	52,67%	84,98%	M.SVM	
															10	71,02%	51,68%	84,90%	M.SVM	
2		•					•								30	77,35%	72,50%	80,41%	RL	2-parametry LIBSVM
															10	76,85%	69,73%	81,95%	RL	
															10	76,01%	56,68%	90,09%	LIBSVM	
															30	73,04%	58,38%	83,77%	LIBSVM	
															30	73,04%	58,38%	83,77%	LIBSVM	

Najlepszy pod względem *ACC* znaleziony klasyfikator SVM osiągnął 82,60% skuteczności, dokonując oceny na podstawie ośmiu cech. Niestety na tle innych wyników odchylenie standardowe trafności było bardzo znaczące i wyniosło aż 5,15. Natomiast parametry $TPR = 80,10\%$ oraz $SPC = 84,37\%$ należałoby zakwalifikować jako satysfakcjonujące, zważywszy na fakt, że tylko pięć z badanych 1348 klasyfikatorów tego typu było w stanie osiągnąć wynik wyższy niż 80% dla obu tych parametrów jednocześnie (porównanie zamieszczono w Tab. 3.4). Jako drugi w Tab. 3.2 zamieszczono opisywany już w Rozdz. 3.1 najlepszy model RL osiągający $ACC = 82,79\%$, $TPR = 84,20\%$ oraz $SPC = 81,73\%$. Pomimo, że obliczenia wykonane dzięki bibliotece LIBSVM dla tych samych zmiennych objaśniających, dają o ponad 8% lepsze wyniki niż te z Rozdz. 3.2 (M.SVM), RL w tej konfiguracji jest nadal o 4% skuteczniejsza. Można jednak znaleźć i takie modele (np. kolejne w Tab. 3.2 5- i 6-parametry),

które dla obu algorytmów predykcji są w stanie osiągnąć ponad 81% trafności w diagnozie. Szczególnie ciekawym wydaje się być ten 5-elementowy, który okazał się wspólnym najlepszym modelem tej wielkości dla obu metod, osiągając 82,23% trafności dla SVM oraz 82,59% dla RL. Pomimo, że korzystając z metody wektorów wspierających TPR wynosił jedynie 77,13%, proponowany zbiór cech wydaje się być rozsądnym wyborem biorąc pod uwagę $SPC = 85,87\%$ dla SVM oraz $TPR = 84,65\%$ i $SPC = 81,07\%$ dla RL. Należy jednak wspomnieć, że tak w tym przypadku jak i większości innych odchylenie standardowe trafności predykcji było prawie dwa razy większe dla SVM niż RL (w tym przypadku 3,04 do 1,49 – Tab. 3.3). Dalsza część Tab. 3.2 przedstawia przegląd najlepszych wyników wg. liczności zastosowanych zmiennych niezależnych:

- modele 4-parametrowe:

Regresja logistyczna (RL) jest w stanie osiągnąć 82,01% ACC , podczas gdy alternatywna metoda jedynie 80,85%. Jednak to właśnie zastosowanie tego drugiego zestawu parametrów (wytłuszczone w Tab. 3.3) zwraca uwagę swoją prostotą, niskim odchyleniem standardowym ACC (około 1.4) oraz trafnością ponad 80% dla obu metod.

- modele 3-parametrowe:

Uzyskano 81.29% trafności dla RL oraz 80.57% dla SVM. Wyniki te są szczególnie interesujące ze względu na fakt, że dla obu algorytmów wartość odchylenia standardowego ACC nie przekroczyła poziomu 1,5. Dla tej grupy nie było jednak możliwe znalezienie zestawu cech gwarantujących skuteczność powyżej 80% dla obu metod jednocześnie.

- modele 2-parametrowe:

Dla tak małej liczby cech rezultaty SVM na poziomie 76,01% ACC są całkowicie niesatysfakcjonujące, nie wspominając o bardzo wysokiej dewiacji wynoszącej 5,81. W tym kontekście zaskakujący może być fakt, że

regresja logistyczna pozwala osiągać rezultaty predykcji na poziomie 80,30% korzystając z tak małej ilości danych.

Korzystając z wszystkich 14 cech model SVM osiągnął skuteczność 76,86%, natomiast dla RL wartość ta wyniosła 77,55%. Ponadto analizując wspomniane tabele można zauważyć prawidłowość, że w wynikach z użyciem regresji logistycznej większa jest zazwyczaj wartość TPR niż SPC , podczas gdy dla support vector machine zależność jest odwrotna.

Tablica 3.3: Wielkości odchyłeń standardowych trafności predykcji dla modeli z Tab. 3.2 [Ochab i Wajs, 2014a]

Liczba cech	Metoda	Liczba wykluczonych próbek alg. w Jackknife			
		10		30	
		ACC	ACC_{dev}	ACC	ACC_{dev}
8	LIBSVM	82,60%	5,15	79,57%	4,95
	RL	77,49%	1,51	76,42%	2,75
6	RL	82,79%	1,11	80,85%	2,52
	LIBSVM	78,22%	1,96	78,49%	2,74
6	RL	82,67%	1,19	81,12%	2,56
	LIBSVM	81,50%	4,89	78,22%	5,06
5	RL	82,59%	1,49	81,46%	2,51
	LIBSVM	82,23%	3,04	80,46%	2,44
4	RL	82,01%	1,73	80,64%	2,60
	LIBSVM	78,78%	2,68	77,89%	4,36
4	LIBSVM	80,85%	1,39	80,35%	2,28
	RL	80,19%	1,43	79,90%	2,89
3	RL	81,29%	1,74	81,05%	2,36
	LIBSVM	79,83%	1,73	79,02%	3,34
3	LIBSVM	80,57%	1,48	79,47%	3,06
	RL	79,40%	1,28	79,32%	2,30
2	RL	80,30%	1,62	81,06%	2,38
	LIBSVM	73,57%	2,40	73,22%	3,01
2	RL	76,85%	1,14	77,35%	2,41
	LIBSVM	76,01%	5,81	73,04%	2,70

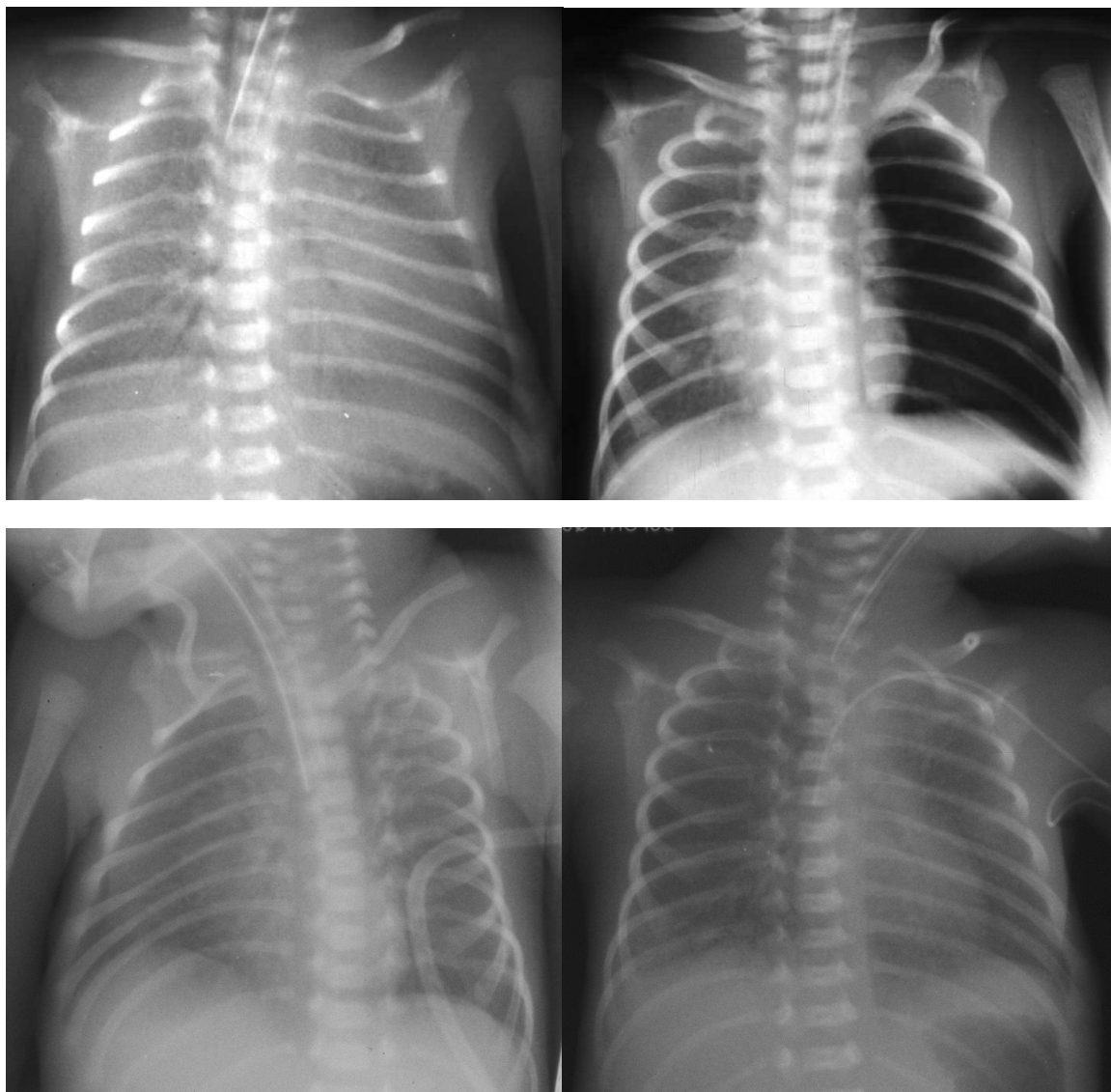
3.5. Porównanie wyników Regresji Logistycznej oraz LIBSVM

Tablica 3.4: Lista znalezionych klasyfikatorów SVM o wartościach TPR i SPC wyższych niż 80% jednocześnie [Ochab i Wajs, 2014a]

Liczba cech	BWEIGHT	GAGE	RESPIMV	AA	PDA	SURFACT	SPO2MEAN	SPO2DEV	LOW85	HIGH94	BPMMEAN	BPMMEAN_TR	SPO2DEV_TR	SPO2MEAN_TR	Liczba usuniętych próbek w Jackknife	ACC	TPR	SPC	Method	ACC_{dev}
8		•	•	•	•	•		•	•	•					10	82,60%	80,10%	84,37%	LIBSVM	5,15
															30	79,57%	75,84%	82,17%	LIBSVM	4,95
															10	77,49%	78,62%	76,43%	RL	1,51
															30	76,42%	76,72%	76,09%	RL	2,75
5		•	•			•					•	•			10	82,18%	83,68%	83,68%	LIBSVM	2,95
															10	81,62%	87,21%	77,47%	RL	1,51
															30	80,73%	85,99%	76,56%	RL	2,27
															30	79,79%	78,02%	81,03%	LIBSVM	3,42
11	•	•		•	•		•	•	•		•	•	•	•	10	81,41%	80,65%	82,01%	LIBSVM	3,33
															30	80,81%	75,82%	84,28%	LIBSVM	3,78
															10	79,83%	83,83%	76,90%	RL	2,22
															30	78,29%	80,05%	76,87%	RL	3,02
11		•	•	•	•	•	•	•	•			•	•		10	80,50%	80,17%	80,68%	LIBSVM	2,73
															30	77,26%	77,54%	76,91%	LIBSVM	3,07
															10	76,99%	80,30%	74,53%	RL	1,21
															30	76,00%	79,87%	72,71%	RL	3,00
9		•	•	•	•	•		•				•	•	•	10	80,38%	80,46%	80,31%	LIBSVM	2,65
															10	78,12%	82,49%	75,02%	RL	1,26
															30	77,58%	77,15%	77,67%	LIBSVM	4,29
															30	76,83%	83,04%	71,87%	RL	2,93

3.6 Analiza zdjęć RTG

Ze względu na sugestie lekarzy [Kruczek i Kwinta, 2007] oraz literaturę [Noack i in., 1993, Yuksel i in., 1993] odnotowującą użycie danych pozyskanych podczas badania rentgenograficznego przy predykcji diagnozy BPD, zdecydowano się na próbę ich dołączenia do głównego zbioru. Jak wynika z rejestrów szpitalnych, 66 pacjentom (spośród badanych 109) w pierwszym tygodniu życia wykonano zdjęcia RTG klatki piersiowej; wszystkie z tych zdjęć okazały się wykonane przed czwartym dniem życia. Należy tu nadmienić, że zdjęcia tak małych pacjentów nie są tak ustandaryzowane jak typowe zdjęcia osób starszych, które świadome badania są w stanie odpowiednio się ustawić (Rys. 3.2). Ze względu na fakt, że automatyczna ocena nie jest łatwa nawet dla zunifikowanych obrazów [Tadeusiewicz i Ogiela, 2004], do wstępnej analizy zdecydowano się z niej zrezygnować, a dzięki konsultacji ze specjalistą radiologiem zdjęcia zostały ocenione ręcznie. W piśmiennictwie wspomniane są dwa systemy takiej oceny w kontekście BPD. Pierwszy z nich nazywany systemem Edwardsa [Toce i in., 1984] został opracowany jeszcze w 1984 r. i zakłada ocenę zdjęcia w skali od 0 do 2 w pięciu kategoriach, tj. zaburzeń układu sercowo-naczyniowego (ang. Cardiovascular abnormalities), hiper-ekspansji płuc (ang. Hyperexpansion), rozedmy (ang. Emphysema), zwłóknień oraz nieprawidłowości śródmiąższowych (ang. Fibrosis/interstitial abnormalities) oraz ogólnej subiektywnej oceny stanu, który sugeruje radiogram. Ostatecznym wynikiem oceny jest suma wyników każdej z pięciu kategorii, co daje wynik z zakresu 0–10. Drugi, nieco nowszy, system oceny [Weinstein i in., 1994] zakłada oceny z zakresu 1–6 przedstawione w Tab. 3.5. Zgodnie z analizą zawartą we wspomnianym artykule można założyć, że oba systemy oceny radiogramów generują zbliżone do siebie wyniki. Biorąc jednak pod uwagę praktyczne zastosowania oraz potencjalne możliwości późniejszej implementacji w celu automatycznej analizy zdjęć, zdecydowano się na użycie drugiego ze wspomnianych systemów. Podczas ręcznej analizy zgromadzonego materiału

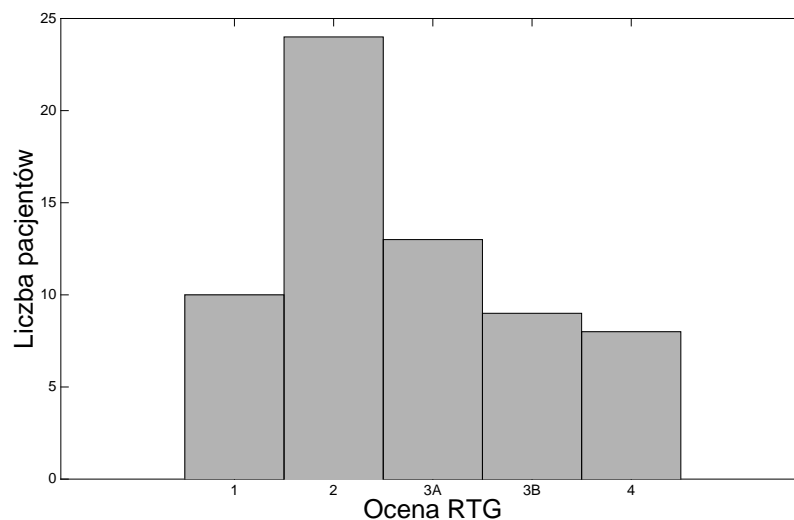


Rysunek 3.2: Przykładowe zdjęcia RTG

Tablica 3.5: Zastosowany system oceny radiogramów dla diagnozy dysplazji oskrzelowo-płucnej.[Weinstein i in., 1994]

Ocena	Opis
1	Słabo widoczne zagęszczenia dające obraz mlecznego szkła (<i>ang. Faint, not well definable opacities giving a hazy appearance to the lung</i>)
2	Wyraźne linijne lub siateczkowate zagęszczenia, zlokalizowane głównie centralnie (<i>ang. Definable linear-reticular opacities which are primarily located centrally</i>)
3	Pogrubiłe linijne lub siateczkowate zagęszczenia, które mogą obejmować (szerzyć się na) części obwodowe płuc; zmiany mogą się ze sobą zlewać w części centralnej (<i>ang. Coarser linear-reticular opacities which may extend to the periphery of the lung; coalescence may be present centrally</i>)
4	Zmiany obserwowane w stopniu trzecim oraz bardzo małe, ale wyraźne zmiany torbielowate (<i>ang. Very small but definable cystic changes in addition to the changes seen in grade 3</i>)
5	Obszary zacięń i torbielowate są w przybliżeniu równej objętości (torbiele są większe niż dla stopnia 4) (<i>ang. Opaque and cystic areas are of approximately equal volume; the cysts are larger than in grade 4</i>)
6	Obszary torbielowate są większe od obszarów zacięń; płuca o obrazie pęcherzykowatym. (<i>ang. Greater cystic areas than opaque areas; the lung has a bubbly appearance.</i>)

RTG, we współpracy z doświadczonym radiologiem [Kurcz, 2013], stwierdzono, że u pacjentów w wieku poniżej 1 tygodnia w praktyce mają zastosowanie tylko noty z zakresu 1–4 z dostępnej skali. Jest to podyktowane faktem, że nie minęła wystarczająca ilość czasu, aby wykształciły się zmiany klasyfikowane jako 5 czy 6. Natomiast wyraźnie zaobserwowano dwie, wyróżniające się pod względem intensywności występujących zmian, grupy pośród tych pacjentów, którym przyporządkowano ocenę równą 3. Z tego względu zdecydowano o rozszerzeniu skali i zastąpieniu wspomnianej noty przez 3A oraz 3B. Co więcej, jakość dwóch ze zdjęć nie pozwoliła na dokonanie prawidłowej oceny, tak więc ostatecznie w dalszej analizie wzięto pod uwagę tylko 64 pacjentów. Histogram otrzymanej w ten sposób nowej zmiennej niezależnej przedstawia Rys. 3.3.



Rysunek 3.3: Histogram ocen zdjęć RTG

Korzystając z powyższych danych powtórzono eksperymenty z Rozdz. 3.1 oraz Rozdz. 3.4. Mogłoby się wydawać, że osiągnięto bardzo interesujące wyniki – $ACC=89,33\%$ dla LIBSVM oraz $87,52\%$ dla RL. Jak się jednak okazało, niestety także wcześniej przebadane modele nie uwzględniające oceny radiologicznej na powstałym 64-elementowym zbiorze uczącym osiągały dużo wyższe noty. Wyda-

jące się więc tak dobrymi wyniki, są jedynie efektem drastycznego zmniejszenia grupy badanych pacjentów i co za tym idzie spadku zdolności generalizacyjnych modeli. Innymi słowy doszło do „nauczenia się wszystkich przypadków na pamięć”. Mimo to, w celu zbadania wpływu badanego dodatkowego parametru na jakość predykcji, pokuszono się na przeprowadzenie dodatkowej analizy:

Wybrano najlepsze znalezione dotychczas zestawy zmiennych niezależnych zamieszczone w Tab. 3.2 oraz Tab. 3.4. Z nowo otrzymanych wyników, wyselekcjonowano modele zawierające te parametry i porównano je z identycznymi poszerzonymi o ocenę radiologiczną. W ten sposób sprawdzono w jaki sposób wpływa na wynik modelu (uznanego wcześniej za dobry) dodanie dodatkowego nowego parametru. W większości wypadków dodanie go nie zwiększało trafności – tylko w 9 z 28 przypadków zaobserwowano poprawę.

Pomimo że powyższa analiza nie dała rozstrzygającego wyniku, można domniemywać, że nowo badana cecha nie niesie ze sobą kluczowej dla predykcji BPD informacji. Ponadto, jak ocenił radiolog, „parametr taki jest owszem swego rodzaju oceną prawdopodobieństwa rozwoju BPD na podstawie nasilenia zmian patologicznych w polach płucnych u noworodków w 1 tygodniu życia. Jednak w tak wczesnym stadium równie prawdopodobne jest, że obrazy przedstawiają zmiany zapalne, RDS, zapalenia zachłystowe itd., a nie wczesne fazy BPD” [Kurcz, 2013]. W podobnym tonie wypowiadają się Toce i in. [1984], twierdząc, że dopiero po 14 dniach istnieje realna możliwość rozróżnienia pomiędzy sobą zmian patologicznych. Jako optymalny czas dla takiej oceny autorzy wymieniają 21 dzień po porodzie. Ze względu na powyższe fakty oraz brak możliwości pozyskania większej ilości danych postanowiono porzucić dalsze badania z użyciem zmiennych związanych z obrazami RTG.

3.7 Wstępne wnioski

Przytoczone wstępne badania jednoznacznie potwierdzają prezentowane w literaturze wyniki [Bhering i in., 2007, Kim i in., 2005, Romagnoli i in., 1998] mówiące, że dzięki regresji logistycznej możliwa jest predykcja BPD w 7 dniu życia z trafnością powyżej 82%. Ponadto udowodniono, że metoda SVM może być także z powodzeniem stosowana do przewidywania wspomnianej przypadłości. Korzystanie z bibliotecznych funkcji pakietu Matlab nie daje co prawda najlepszych rezultatów w tej dziedzinie, jednak bardzo dobrym rozwiązaniem, dającym znacznie szerszy wpływ na przeprowadzane obliczenia, jest zastosowanie biblioteki LIBSVM. Pomimo zbadania tylko losowej części (1348 z 2^{14}) możliwych do skonstruowania modeli SVM, otrzymano niejednokrotnie wyniki lepsze niż stosując RL. Generalnie można stwierdzić, że dla małej liczby cech lepsze wyniki osiąga się używając regresji logistycznej. Im więcej parametrów branych jest pod uwagę, tym więcej zyskuje SVM, osiągając już bardzo wyraźną przewagę nad RL w modelach z więcej niż sześcioma parametrami. Co prawda, w przypadku regresji odchylenie standardowe otrzymywanych wyników jest znacząco niższe, niż dla drugiej metody. Nie oznacza to jednak, że nie można znaleźć konkretnych modeli SVM, dla których ACC_{dev} będzie zbliżone do tej odpowiadającej LR [Tab. 3.3]. Kolejną cechą różniącą obie metody jest stosunek czułości do specyficzności. Dla badanego zagadnienia wyraźnie wyższą specyficznością charakteryzują się klasyfikatory SVM, podczas gdy czułość jest domeną regresji logistycznej [Tab. 3.2]. Optymistyczny jest fakt, że w większości przypadków dla obu metod ACC rosło i jednocześnie ACC_{dev} malało, gdy liczbę próbek uczących zwiększano (wyniki dla Jackknife = 10 są lepsze) [Tab. 3.2, 3.3, 3.4]. Oznacza to, że oba typy klasyfikatorów mają dobrą zdolność do generalizacji i nie następuje efekt tzw. przeuczenia, kiedy to odtwarzane są tylko wcześniej zapamiętane dane uczące. Analizując wspomniane tabele, można wyraźnie zauważyć częściej powtarzające się cechy: *gage*, *low84*, *surfact* i *pda*. Zmienne te owszem są wspomniane w

literaturze jako predyktory BPD [Ambalavanan i in., 2008, Bancalari i Claire, 2016, Bhering i in., 2007, Bhutani i Abbasi, 1992, Corcoran i in., 1993, Cunha i in., 2005, Farstad i in., 2011, Geary i in., 2008, Groothuis i Makari, 2012, Jobe, 2011, Kim i in., 2005, Kinsella i in., 2006, Kwinta i in., 2008b, Laughon i in., 2011, Marshall i in., 1999, Oh i in., 2005, Rojas i in., 1995, Singer i in., 1992, Sinkin i in., 1990, Sosenko i Bancalari, 2012, Stoch, 2007], jednak pomijając parametr *gaze* trudno stwierdzić, które z nich miałyby być kluczowe. Wynik takich poszukiwań byłby zależny od metody, preferencji co do wysokiej specyficzności lub czułości oraz dopuszczalnego błędu predykcji. Brak jednoznacznej odpowiedzi na to pytanie w praktyce może jednak okazać się pozytywnym. Bowiernie podczas przeprowadzonej analizy historii pacjentów, bardzo częstym problemem okazywał się brak części danych. Najczęściej nie były to błędy archiwistyczne, lecz konsekwencje rozlicznych sytuacji takich jak, np. przeniesienie pacjenta z innej placówki w, której nie było odpowiedniego sprzętu do dokonania wszystkich możliwych pomiarów. Na potrzeby tego opracowania należało takiego pacjenta wykluczyć ze zbioru danych, jednak w praktycznym zastosowaniu brak danej informacji nie powinien wykluczać możliwości dokonania oceny rokowań. Mając do dyspozycji wiele modeli opartych o różne parametry, zawsze istnieje wariant znalezienia najlepszego z tych możliwych do zastosowania w danej sytuacji.

Z przedstawionej powyżej analizy wynika jednoznacznie, że tak regresja logistyczna jak i metoda support vector machine, może być z powodzeniem użyta do predykcji BPD. Jak się jednak okazuje, rozwiązanie problemu doboru odpowiednich cech nie jest proste ani jednoznaczne. Zmienne niezależne generujące najlepsze klasyfikatory nie są takie same dla każdej z metod, co więcej, są one różne w obrębie jednej metody zależnie od tego czy kluczowa jest czułość, specyficzność czy może trafność diagnozy. Ponadto mając na uwadze praktyczne zastosowania w codziennej praktyce medycznej, należy zauważyć, że dostępność pomiarów konkretnych cech, w każdym wypadku może być inna z powodu bardzo

trywialnych ograniczeń. W takich wypadkach nawet sprawdzone modele, mogą okazać się bezużyteczne ze względu na brak danych. Zatem lepszym rozwiązaniem jest zastosowanie mniej skutecznej metody, ale mimo wszystko dającej wynik obarczony określonym błędem predykcji, niż brak jakiegokolwiek prognozy. Z tego też powodu wydaje się zasadnym stworzenie systemu wspomagania decyzji, zdolnego do zaproponowania metody predykcji odpowiedniej do danej sytuacji. Mając na uwadze dostępne dane, możliwość wystąpienia bardzo poważnych skutków ubocznych terapii [Ng, 1993, Ryan i in., 1996] (w takim przypadku ważniejsze byłoby SPC niż TPR [Akobeng, 2007]) oraz największy dopuszczalny błąd predykcji, system taki mógłby zaproponować wybór odpowiednich cech oraz metody. Ponadto, skoro losowo wybrane modele SVM dały tak obiecujące wyniki, warto byłoby zbadać wszystkie możliwe kombinacje cech.

Rozdział 4

System wspomaganie decyzji

Ze względu na brak jednego, ogólnie uznanego za sprawdzony i możliwy do zastosowania w praktyce modelu predykcji BPD oraz opierając się na wnioskach zawartych w poprzednim rozdziale stwierdzono, że dla prezentowanego problemu najtrafniejsze mogłoby się okazać zastosowanie systemu wspomaganie decyzji. Wobec nieznaleszenia w literaturze wzmianek o jakichkolwiek pracach, w których wykorzystywano by podobne podejście, zdecydowano się podjąć próbę zbudowania takiego systemu. Głównym założeniem była konstrukcja umożliwiająca praktyczne zastosowanie w codziennej pracy lekarza, dzięki generowaniu odpowiednich wskazówek podczas podejmowania kluczowych decyzji. Lekarz byłby informowany, jaką metodę predykcji najlepiej użyć w danej sytuacji, które z mierzonych parametrów będą najlepszymi predyktorami BPD oraz na ile można ufać otrzymanemu wynikowi. Ponadto program taki mógłby doradzać, które z możliwych pomiarów warto byłoby dodatkowo wykonywać w celu osiągnięcia lepszych wyników predykcji w przyszłości. Stworzony system wspomaganie decyzji został przedstawiony w [Ochab i Waajs, 2016].

4.1 Budowa bazy wiedzy

Rdzeń bazy wiedzy konstruowanego systemu wspomaganie decyzji zbudowano ze średnich wartości wyników predykcji BPD wykonanej w sposób opisany w Rozdz. 3. W celu łatwego przeszukiwania zgromadzonych danych według żądanych kryteriów wszystkie informacje umieszczono w relacyjnej bazie danych MySQL. Ponadto w celu przyspieszenia działania procedur kwerendujących zdefiniowano odpowiednie indeksy dla kluczowych kolumn powstałej tabeli. Każdy wiersz odpowiada konkretnemu zestawowi parametrów oraz zastosowanej metodzie. Oprócz wspomnianych informacji zawiera on także wartości ACC , ACC_{dev} , TPR , SPC oraz optymalne wartości parametrów γ i c dla metody SVM (lub σ dla M.SVM).

Ze względu na to, że regresja logistyczna jest jedyną, szeroko opisywaną w piśmiennictwie (w kontekście predykcji BPD) metodą predykcji, za celowe uznano umieszczenie jej wyników w konstruowanej bazie. Ponieważ na potrzeby rozważań opisanych w Rozdz. 3.1 przeprowadzono już badanie wszystkich 2^{14} kombinacji zmiennych niezależnych stosując RL, nie było potrzeby wykonania dodatkowych obliczeń tym algorytmem. Wszystkie 16384 wyniki mogły zostać wprowadzone do bazy bez jakichkolwiek modyfikacji.

Opierając się na obiecujących wynikach oraz wnioskach (przedstawionych w Rozdz. 3.7) dotyczących predykcji metodą SVM, zdecydowano się na umieszczenie w bazie także wyników otrzymanych wspomnianym sposobem. Kluczowym było zawarcie tych, otrzymanych dzięki bardzo obiecującej bibliotece LIBSVM. Jak wspomniano w Rozdz. 3.4, ze względu na wymagany znaczny nakład obliczeniowy dotychczas zbadano jedynie losowo wybraną część modeli. Mając jednak na uwadze jak największą funkcjonalność systemu oraz duże prawdopodobieństwo znalezienia jeszcze lepszych modeli predykcyjnych niż dotąd, zdecydowano się na użycie trzech komputerów osobistych i wykonanie trwających ponad dwa tygodnie obliczeń na wszystkich 2^{14} kombinacjach cech. Algorytm 3 przedstawia wykonany

proces obliczeniowy – analogiczny do tego z Rozdz. 3.4. Zdecydowano się także załączyć 1150 najlepszych wyników otrzymanych dzięki bibliotecznym funkcjom Matlab, pomimo że były one znacznie gorsze od tych uzyskanych przez LIBSVM. Ostatecznie zgromadzono ponad 33 tys. wyników predykcji BPD pokrywających wszystkie kombinacje wspomnianych w Rozdz. 1.2 zmiennych objaśniających dla obu badanych metod (Rozdz. 2).

Algorytm 3 Budowa bazy wiedzy dotyczącej metody LIBSVM poprzez optymalizację parametrów γ i c dla każdej z możliwych kombinacji cech [Ochab i Wajs, 2016]

```
for each of  $2^{14}$  feature_combinations do
  for  $\gamma = 2^{-15}, 2^{-14}, \dots, 2^3 \times c = 2^{-5}, 2^{-4}, \dots, 2^{15}$  do
    for 1..30 do
      patients = delete_10_random(all_patients);
      for each test_patient of patients do
        learn_patients = patients - test_patient;
        predict(test_patient, learn_patients);
      end for
      calculate ACC, TPR, SPC;
    end for
    for ACC, TPR, SPC do
      calculate mean_value and dev;
    end for
  end for
  best_results[feature_combination]=best(meanACC);
end for
```

Ze względów praktycznych wszelkie wyniki związane z LIBSVM umieszczone w bazie wiedzy zostały wygenerowane całkowicie od nowa. Nie korzystano z obliczeń wykonanych na potrzeby Rozdz. 3.4. Co za tym idzie, wnikliwy czytelnik, może zauważyć, że niektóre dane zawarte w Tab. 3.2 oraz Tab. 3.4 różnią się nieznacznie od tych prezentowanych w Tab. 4.1. Właśnie w związku z pewnym losowym rozrzutem otrzymywanych wyników wartość odchylenia standardowego ACC dla metod SVM jest tak duża. Efekt ten byłby znacznie mniej widoczny, gdyby nie zastosowanie wspomnianej metody Jackknife, która poprzez losowe usuwanie próbek [Algorytm 3] wprowadza pewne fluktuacje. Badanie

odpowiednio dużej grupy pacjentów zmniejszyłoby w znacznym stopniu wpływ wspomnianej losowości i pozwoliłoby na ustabilizowanie się wyniku na pewnym poziomie. Jednak w sytuacji, gdy nie ma dostępu do danych większej liczby pacjentów o konkretnym pasującym do badania profilu, lepszym rozwiązaniem wydaje się zbadanie, jak bardzo wyniki mogą od siebie odstawać, niż pozostawienie przypadkowi, w którą część przedziału możliwych do osiągnięcia wyników trafimy. Dokładnie takim przypadkiem byłoby zastosowanie wspomnianego w Rozdz. 3.4 skryptu *grid.py* dostarczanego wraz z biblioteką LIBSVM. Teoretycznie pozwalał on na predykcję z trafnością dużo wyższą niż 90%, jednak nawet po usunięciu pojedynczego pacjenta ze zbioru uczącego, okazywało się, że dla danego zbioru cech nie było możliwe osiągnięcie trafności nawet o 10% niższej od pierwotnie wskazanej. Przez użycie metody Jackknife świadomie więc pogorszono teoretycznie możliwe do osiągnięcia wyniki predykcji, dając tym samym ogląd jakiej skuteczności metody można się spodziewać w zastosowaniach praktycznych (niekoniecznie korzystających z tych samych danych uczących). Dla RL opisany efekt także występował, jednak jego nasilanie było na tyle małe, że pomimo odchyłeń poszczególnych pojedynczych wyników w iteracjach Jackknife, ostateczne uśrednione wyniki pozostawały takie same przy każdej próbie uruchomienia programu obliczeniowego. Zanotowaną różnicę w zachowaniu się wyników obu metod można zapewne wyłumaczyć faktem, że SVM dysponując przestrzenią nieskończenie wymiarową (Rozdz. 2.2) ma dużo większe możliwości dopasowania się do danych uczących, co jak widać nie zawsze jest korzystne – szczególnie, gdy operuje się na małej ilości danych.

Zestawienie najlepszych znalezionych klasyfikatorów pod względem ACC, których $ACC_{dev} < 4$ przedstawiono w Tab. 4.1. Jak można zauważyć, zgodnie z oczekiwaniami przeszukanie całej przestrzeni kombinacji cech pozwoliło na odkrycie kolejnych modeli SVM, generujących jeszcze bardziej interesujące wyniki niż te prezentowane w Rozdz. 3.5. Wynikiem o najwyższej trafności prognozy

przy zachowaniu arbitralnie wybranego kryterium $ACC_{dev} < 4$ był model oparty na dziewięciu cechach i osiągający $ACC = 83,29\%$ ($TPR = 79,08\%$, $SPC = 86,4\%$). Co prawda $ACC_{dev} = 3,63$ nie należy do najniższych, ale w porównaniu do najlepszego wyniku z poprzedniego rozdziału ($ACC_{dev} = 5,15$ i $ACC = 82,60\%$) obserwuje się i tak znaczną poprawę. Jeszcze ciekawszym wydaje się wynik drugi na liście – $ACC = 83,25\%$, $TPR = 78,82\%$, $SPC = 86,49\%$ – przy zachowaniu $ACC_{dev} = 2,16$, co jest wartością bardzo niską jak na zastosowanie metody SVM. Co prawda takie odchylenie standardowe trafności jest nadal większe niż dla regresji logistycznej, jednakże są to już wartości porównywalne, przy trafności wyższej o $0,46\%$ w stosunku do RL. Pośród wielu nowych wyników pojawiających się we wspomnianej tabeli, szczególną uwagę zwraca także czwarty z klasyfikatorów, którego specyficzność wynosi aż $88,92\%$. Jest to bez wątpienia dobry wynik, lepszy od dotychczas prezentowanych o ponad $4,55\%$. Jak widać, wśród ponad 33 tys. wszystkich modeli, baza wiedzy dysponuje ponad czterdzieściami takimi, które pozwalają osiągnąć trafność na poziomie przekraczającym 82% (przy zachowaniu $ACC_{dev} < 4$). Można je więc stosować wymiennie zależnie od sytuacji.

Tablica 4.1: Czterdzieści najlepszych pod względem ACC klasyfikatorów, dla których $ACC_{dev} < 4$ zawartych w bazie wiedzy [Ochab i Wajns, 2016]

Liczba parametrów	BWEIGHT	GAGE	RESPIMV	AA	PDA	SURFACT	SPO2MEAN	SPO2DEV	LOW85	HIGH94	BPMMEAN	BPMMEAN_TR	SPO2DEV_TR	SPO2MEAN_TR	ACC_{dev}	ACC	TPR	SPC	Metoda
9	•	•	•	•	•	•	•	•	•	•	•	•	•	•	3,63	83,29%	79,08%	86,4%	LIBSVM
8	•	•	•	•	•	•	•	•	•	•	•	•	•	•	2,16	83,25%	78,82%	86,49%	LIBSVM
8	•	•	•	•	•	•	•	•	•	•	•	•	•	•	3,88	83,23%	76,54%	88,07%	LIBSVM
9	•	•	•	•	•	•	•	•	•	•	•	•	•	•	2,98	83,11%	75,07%	88,92%	LIBSVM

Tablica 4.1: Czterdzieści najlepszych pod względem ACC klasyfikatorów, dla których $ACC_{dev} < 4$ zawartych w bazie wiedzy cd.[Ochab i Wajs, 2016]

Liczba parametrów	BWEIGHT	GAGE	RESPIMV	AA	PDA	SURFACT	SPO2MEAN	SPO2DEV	LOW85	HIGH94	BPMMEAN	BPMMEAN_TR	SPO2DEV_TR	SPO2MEAN_TR	ACC_{dev}	ACC	TPR	SPC	Metoda
7	•	•	•	•	•			•		•					2,6	83,04%	82,63%	83,28%	LIBSVM
8	•			•	•		•		•	•	•		•		3,39	82,96%	78,12%	86,5%	LIBSVM
8	•			•	•		•		•		•	•		•	2,39	82,88%	79,07%	85,64%	LIBSVM
6	•			•	•			•			•	•			1,11	82,79%	84,2%	81,73%	LOGIT
9	•			•	•		•	•		•	•	•		•	2,3	82,78%	77,84%	86,45%	LIBSVM
6	•	•			•	•		•			•				1,19	82,67%	87,5%	79,09%	LOGIT
8	•			•	•			•	•	•	•	•		•	2,69	82,63%	78,59%	85,54%	LIBSVM
11	•	•		•	•		•	•	•	•	•	•	•		3,12	82,6%	78,65%	85,55%	LIBSVM
5	•				•			•		•	•				1,49	82,59%	84,65%	81,07%	LOGIT
5	•				•			•			•	•			1,32	82,56%	85,25%	80,54%	LOGIT
6	•				•			•			•	•	•		1,43	82,53%	83,82%	81,54%	LOGIT
7	•			•	•			•	•	•				•	2,7	82,52%	76,91%	86,6%	LIBSVM
6	•				•			•			•	•	•		1,64	82,52%	84,87%	80,81%	LOGIT
7	•			•	•			•	•	•	•	•			2,11	82,42%	78,69%	85,07%	LIBSVM
8	•			•	•		•		•	•	•			•	3,44	82,41%	76,2%	86,96%	LIBSVM
8	•			•	•		•		•		•	•	•		1,52	82,4%	78,65%	85,12%	LIBSVM
6	•				•	•		•	•					•	3,25	82,32%	80,03%	83,97%	LIBSVM
10	•			•	•		•	•	•	•	•	•	•		2,44	82,31%	77,86%	85,56%	LIBSVM
8	•			•	•		•		•	•	•			•	2,41	82,28%	77,78%	85,63%	LIBSVM
7	•				•		•		•		•	•	•		1,41	82,27%	83,35%	81,47%	LOGIT

Tablica 4.1: Czterdzieści najlepszych pod względem ACC klasyfikatorów, dla których $ACC_{dev} < 4$ zawartych w bazie wiedzy cd.[Ochab i Wajs, 2016]

Liczba parametrów	BWEIGHT	GAGE	RESPIMV	AA	PDA	SURFACT	SPO2MEAN	SPO2DEV	LOW85	HIGH94	BPMMEAN	BPMMEAN_TR	SPO2DEV_TR	SPO2MEAN_TR	ACC_{dev}	ACC	TPR	SPC	Metoda
7	•	•	•				•	•			•	•			1,73	82,24%	83,18%	81,51%	LOGIT
8	•	•	•	•			•		•		•	•		•	2,37	82,23%	79,19%	84,49%	LIBSVM
7	•	•	•	•	•			•			•		•		3,75	82,22%	78,11%	85,1%	LIBSVM
8	•			•	•	•		•			•	•	•	•	1,85	82,21%	79,62%	84,07%	LIBSVM
5	•			•	•			•			•				1,27	82,2%	86,06%	79,38%	LOGIT
6	•			•	•			•			•			•	1,76	82,2%	86,31%	79,25%	LOGIT
7	•	•	•	•		•		•			•			•	1,99	82,19%	75,2%	87,17%	LIBSVM
6	•			•			•		•	•	•				1,79	82,18%	84,76%	80,28%	LOGIT
6	•	•	•					•			•	•			1,69	82,17%	83,66%	81,08%	LOGIT
7	•			•			•		•	•	•	•	•		1,08	82,11%	82,69%	81,63%	LOGIT
8	•	•	•	•		•	•		•	•				•	2,24	82,08%	75,32%	87,01%	LIBSVM
9	•	•	•	•		•	•	•			•	•		•	2,25	82,04%	77,62%	85,27%	LIBSVM
8	•	•	•	•	•		•	•						•	2,93	82,04%	80,13%	83,4%	LIBSVM
5	•			•				•			•		•		1,96	82,04%	82,97%	81,35%	LOGIT
9	•			•		•		•	•	•	•	•	•	•	2,55	82,01%	78,33%	84,71%	LIBSVM
4	•			•				•		•					1,73	82,01%	84,38%	80,34%	LOGIT

4.2 Budowa systemu

Mając do dyspozycji opisaną w poprzednim rozdziale bazę wiedzy umieszczoną w bazie danych MySQL wykonano w języku PHP odpowiednie oprogramowanie

do jej analizy. Program opracowano jako aplikację Web, nie tylko ze względu na prostotę implementacji, ale głównie z zamysłem łatwego udostępnienia dowolnej liczbie zainteresowanych użytkowników. Prawie wszystkie wykonywane operacje, których główną częścią jest przeszukiwanie bazy wiedzy, oparte są na formułowaniu odpowiednich zapytań SQL do bazy danych.

W pierwszym etapie poszukiwane są najlepsze możliwe do zastosowania modele, wykorzystujące pomiary uprzednio zdefiniowanej przez użytkownika grupy n cech. Wyniki są sortowane względem ACC i grupowane wg zastosowanej metody predykcji. Jeżeli ACC_{dev} najlepszego wyniku którejś z metod jest wyższe niż arbitralnie wybrany poziom 3,5, wyświetlany jest odpowiedni komunikat, po czym dla danej metody poszukiwany jest model o odpowiednio niższej wartości tego parametru. Jeżeli na etapie wyboru parametrów poszukiwań użytkownik zdefiniował żądany maksymalny poziom ACC_{dev} , poprzedni krok jest pomijany i prezentowane są od razu wyniki spełniające wskazane kryterium.

W następnym etapie system sprawdza, który z niewybranych parametrów byłby w stanie zmaksymalizować trafność predykcji w wypadku jego dodania. Do uprzednio zdefiniowanej grupy cech dodawany jest w pętli jeden dodatkowy, w każdej rundzie inny, parametr. Dla każdego z nowo powstałych zestawów $n + 1$ cech powtarzane jest poszukiwanie analogiczne jak w etapie pierwszym. Jeżeli trafność predykcji, w którychś przypadkach będzie wyższa od otrzymanej w poprzedniej fazie poszukiwań, najlepszy z takowych będzie zaprezentowany użytkownikowi. Podobnie jak w etapie pierwszym, dla nowo powstałego, dającego najlepsze rezultaty predykcji, zestawu $n+1$ cech prezentowane są najlepsze modele oparte na każdej z metod.

Ostatnim etapem działania systemu jest poszukiwanie pośród zdefiniowanych n cech takiej, której pomiary nie zwiększają w żaden sposób jakości predykcji. Poszukiwane i prezentowane są te z n parametrów, które nie wystąpiły w żadnym z sugerowanych modeli w etapie pierwszym i drugim dla którejkolwiek z metod.

Uproszczoną logikę działania systemu opisuje Algorytm 4. Przykładowa odpowiedź systemu zaprezentowana została w przykładzie 1.

Algorytm 4 Uproszczony algorytm działania systemu wspomagania decyzji w pseudo SQL [Ochab i Wajs, 2016]

```
select MODEL where FEATURES in AVAIL_FEATURES order by ACC group by METHOD;
print best result for each method;
foreach METHOD if (DEV >3.5)
  select MODEL where FEATURES in AVAIL_FEATURES and METHOD and DEV <3.5 order by ACC;
  print alternative model for given method with lower DEV;
end for
foreach not AVAIL_FEATURES as ADD_FEATURE
  select EXT_MODEL where FEATURES in AVAIL_FEATURES+ADD_FEATURE order by ACC;
end for
select ADD_FEATURE FROM not AVAIL_FEATURES where max(ACC) in EXT_MODEL;
print propose additional feature which give max improvement;
select FEATURE from AVAIL_FEATURES where FEATURE not in (MODEL or EXT_MODEL);
print features which can be omitted;
```

Przykład 1 Przykładowa odpowiedź systemu wspomaganie decyzji [Ochab i Waajs, 2016]

Wybrałeś dostępne pomiary: *bweight*, *aa*, *spO2mean*, *low85*, *bpmmean*, *bpmmean_tr*.

- Najlepszy wynik osiągniesz metodą LIBSVM:
 - 77.47% trafności predykcji przy $ACC_{dev} = 3.44$, $TPR = 65.42\%$, $SPC = 86.27\%$,
 - dla tej metody sugerowanymi parametrami są:
bweight, *spO2mean*, *bpmmean*.

 - Użycie alternatywnej metody RL pozwoli osiągnąć:
 - 76.8% trafności przy $ACC_{dev} = 1.37$, $TPR = 78.73\%$, $SPC = 75.33\%$,
 - dla tej metody sugerowanymi parametrami są:
spO2mean, *bpmmean*.

 - Warto rozważyć wykonanie dodatkowego pomiaru *gage*,
zwiększyłyby to trafność predykcji o 3.28%:
 - ◊ Przy założeniu $ACC_{dev} < 3.5$ najlepszy wynik dałaby metoda LIBSVM osiągając:
 - 80.75% trafności przy $ACC_{dev} = 1.34$, $TPR = 71.59\%$, $SPC = 87.43\%$,
 - dla tej metody sugerowanymi parametrami są:
bweight, *gage*, *low85*, *bpmmean*, *bpmmean_tr*.

 - ◊ Przy założeniu $ACC_{dev} < 3.5$ użycie alternatywnej metody RL pozwoli osiągnąć:
 - 80.19% trafności przy $ACC_{dev} = 1.43$, $TPR = 73.27\%$, $SPC = 85.07\%$,
 - dla tej metody sugerowanymi parametrami są:
gage, *low85*, *bpmmean*, *bpmmean_tr*.

 - W przedstawionej sytuacji pomiar parametru *aa* jest bezcelowy,
nie zwiększa on trafności predykcji.
-

Rozdział 5

Eksperymenty ilustrujące działanie stworzonego systemu

W celu zaprezentowania działania systemu oraz analizy generowanych przez niego odpowiedzi zdecydowano się na przeprowadzenie szeregu doświadczeń. Wykonanie tych eksperymentów pozwoliło na analizę istotności badanych cech w procesie predykcji BPD i co za tym idzie porównanie z dotychczas publikowanymi wynikami. Ponadto pomogły one wykazać, że użytkownik korzystający z wspomnianego systemu jest w stanie osiągnąć znacznie lepsze wyniki predykcji niż korzystając z innych standardowych metod.

5.1 Sekwencyjna rozbudowa modeli przez system wspomaganie decyzji

Aby w przystępny dla czytelnika sposób przedstawić zakres odpowiedzi jakie generuje powstały system, przeprowadzono serię eksperymentów. Na początku wybrano jedną cechę nazwaną startową, po czym postawiono systemowi zadanie dobrania drugiej, dzięki której, powstałby najlepszy dostępny 2-parametrowy model predykcyjny. W kolejnej próbie wygenerowano zapytanie o rozbudowę

powstałego w poprzedniej próbie klasyfikatora o następną zmienną niezależną – tak, aby powstał możliwie najlepszy klasyfikator 3-parametrowy. Doświadczenie powtarzano analogicznie dotąd, aż system nie był w stanie zaproponować żadnej pojedynczej cechy, która pozwalałaby na polepszenie trafności predykcji – w praktyce nigdy nie było to więcej niż osiem zapytań. Wspomnianą serię testów wykonano 14 razy, za każdym razem wybierając inną cechę jako startową. Poniżej zaprezentowano algorytm wykonanych doświadczeń (Alg. 5):

Algorytm 5 Algorytm wykonanych eksperymentów

```
for each variable of all_explanatory_variables do
  model:=variable;
  repeat
    new_variable:=system→upgrade_model(model);
    model→ add(new_variable);
    Print model, method, ACC, ...;
  until new_variable != NULL;
end for
```

5.2 Dobór cech funkcją *sequentialfs* pakietu Matlab

W celu udowodnienia, że użycie prezentowanego systemu wspomagania decyzji pozwala na uzyskanie lepszych efektów predykcji BPD, niż przy użyciu innych niewyrafinowanych metod dostępnych za pośrednictwem pakietu Matlab przeprowadzono poniższy test. Jako przykład wybrano funkcję *sequentialfs*, która w sposób zautomatyzowany dokonuje selekcji postępującej cech, maksymalizując w każdym kroku jakość predykcji poprzez rozszerzenie modelu o kolejną zmienną niezależną. Funkcja działa sekwencyjnie, aż do momentu kiedy żaden z dostępnych parametrów nie wprowadza poprawy wyniku. Ocena dokonywana jest na podstawie definiowanej przez użytkownika funkcji klasyfikującej, przy użyciu k -krotnej walidacji krzyżowej (W-K). W opisywanym przypadku w każdej

iteracji dane zostały podzielone losowo na dziesięć części, po czym każda z nich była kolejno traktowana jako testowa, podczas gdy pozostałe pełniły rolę uczącą. Używając RL jako klasyfikatora, W-K została zastosowana 20-krotnie za każdym razem losowo dzieląc dane, co pozwoliło osiągnąć zadowalający i stabilny wynik. Jak się jednak okazało, używając metody LIBSVM za każdym razem otrzymywano inny wynik, pomimo zastosowania ponad 50 iteracji W-K. Jak zbadano, dla tej metody, trafniejsze jest zastosowanie eliminacji wstecznej, oferowanej zresztą przez tą samą funkcję *sequential fs*. Niestety jej zastosowanie wiązało się także z niemożnością jednoczesnej optymalizacji parametrów γ i c . Z tego względu, na podstawie zebranych wcześniej doświadczeń, arbitralnie wybrano wartości $c = 0$ oraz $g = -1$, wiedząc że generują one dość dobre wyniki. Wspomnianą optymalizację parametrów wykonano dopiero na ostatecznie otrzymanym najlepszym modelu. Taki zabieg pozwolił na miarodajne porównanie wyniku tej metody doboru cech z wynikami generowanymi przez system wspomagania decyzji.

5.3 Wyniki eksperymentów

Wyniki serii eksperymentów opisanych w Rozdz. 5.1 zamieszczono w Tab. 5.1 – każda kolumna zawiera rezultat n -tego kolejnego w danej serii. Celem każdego z badań było wyznaczenie najbardziej obiecującej cechy, o którą warto byłoby rozszerzyć model, zdefiniowany przez parametry w poprzedzających kolumnach danego wiersza. Poniżej każdej z propozycji umieszczono przewidywane wartości trafności predykcji nowo powstałych klasyfikatorów. Cechy, które na dowolnym dalszym etapie okazały się za zbędne zostały zaznaczone przekreśleniem. Niektóre z prezentowanych ciągów są jedynie dwuelementowe – wynika to z faktu, że jedynie dodanie dwóch lub więcej cech jednocześnie do poprzedzającego modelu spowodowałoby zwiększenie trafności predykcji. W miejscach, gdzie $ACC_{dev} > 3,5$ dla danej metody nie podawano wartości ACC, zastępując ją myślnikiem

ze względu na bezcelowość stosowania takiego klasyfikatora. Przedstawiona wizualizacja pozwala na wysunięcie wielu wniosków dotyczących predykcji BPD oraz istotności konkretnych parametrów dla jej jakości:

- Wszystkie dotychczas opisywane wyniki sugerują wniosek, że modele oparte na małej liczbie parametrów (poniżej 7) w większości przypadków dają znacznie lepsze efekty korzystając z RL. Ponadto pomimo istnienia nielicznych wyjątków, ogólnie można stwierdzić, że modele SVM charakteryzują się zazwyczaj około dwa razy większą wartością ACC_{dev} ,
- Jak można zaobserwować, cztery z cech, tj. *aa*, *bmpmean_tr*, *spO2dev_tr*, *spO2mean_tr* w żadnej z serii samodzielnie nie zwiększają skuteczności predykcji. Niosą one ze sobą pewne informacje, ponieważ wspomniane parametry występują w wykazach najlepszych wyników (Tab. 3.2, 3.4, 4.1), choć głównie wśród modeli wieloparametrycznych. Jednak w kontekście tego eksperymentu można by mniemać, że nie są one na tyle istotne, by było możliwe ich samodzielne wykorzystanie (bez posiłkowania się informacjami pochodzących z większej liczby innych cech).
- Parametry *high94* oraz *spO2dev* podczas trzeciej kolejnej analizy zostały oznaczone przez system jako zbędne. Porównując tę informację z Tab. 3.2, 3.4, 4.1 potwierdza się fakt, że powyższe zmienne niezależne zaczynają mieć wpływ na jakość predykcji dopiero w modelach zbudowanych z ponad 5 cech, a w rozważanych eksperymentach z Rozdz. 5.1 w praktyce analizowano głównie mniejsze modele. Wspomniane dwie cechy można by zatem scharakteryzować analogicznie jak te w poprzednim punkcie. Ponadto mając na uwadze analizę współliniowości wykonaną w Rozdz. 1.3, należałoby używać tych cech z odpowiednią ostrożnością. Co jednak zastanawiające, para ta pomimo wzajemnej współliniowości występuje aż w 9 z 40 najlepszych modeli zamieszczonych w Tab. 4.1 .

- Parametr *spO2mean* został użyty jedynie raz, a *respimv* dwa razy jako cecha zwiększająca zdolności predykcyjne modeli. Szczególnie ciekawym przypadkiem jest ten drugi parametr, który bardzo rzadko można spotkać w Tab. 3.2 oraz 4.1. Można by więc zakładać, że nie jest on warty większej uwagi, czy nawet odnotowywania w bazie danych. Jednak jak wskazał stworzony system, w sytuacji, gdy lekarz nie dysponowałby żadnymi zaawansowanymi pomiarami, a jedynie wagą (*bweight*) i wiekiem płodowym (*gage*), to informacja o wcześniejszym użyciu respiratora (*respimv*) jest znacznie cenniejsza i daje trafniejszą informację o stanie pacjenta w kontekście BPD niż nie tylko *gage*, ale i wszystkie inne rozważane cechy w pojedynkę. Podobnie cenną informacją okazałby się pozornie niezbyt istotny parametr *spO2mean*, podczas gdy posiadano by jedynie wiedzę o średnim tętnie (*bpmmean*). Jak można sobie jednak wyobrazić, jest to sytuacja znacznie mniej prawdopodobna, niż w poprzednim możliwym do zaistnienia w praktyce przykładzie.
- Masa urodzeniowa (*bweight*) została zaproponowana jako dopełniająca cecha przez system aż 12 razy. Co jednak bardzo specyficzne, zawsze następowało to w ostatnim z etapów ulepszania klasyfikatora. Wyraźnie widać więc, że informacja ta może być użyta jedynie do minimalnego polepszenia trafności predykcji już dobrze działającego modelu. Potwierdza to fakt, że w Tab. 3.2, 3.4 i 4.1 cecha ta występuje jedynie w 6 z 54 prezentowanych zestawów cech. Wyjaśnieniem takiej sytuacji może być bardzo silna korelacja z *gage* (0,75, Tab. 1.4), który w aż w 11 na 12 przypadków istniał już w rozszerzonym modelu. Mimo tak mocnej korelacji, nie stwierdzono jednak wzajemnej współliniowości tej pary.
- Fakt występowania przetrwałego przewodu tętniczego (*pda*) wydaje się być natomiast znacznie bardziej wartościową informacją niż inne wcześniej wspomniane parametry. System wygenerował 11 odpowiedzi sugerujących

dodanie *pda* do zestawu cech, przy czym następowało to pomiędzy piątym, a siódmym etapem serii. Potwierdzeniem tego przypuszczenia jest występowanie tej zmiennej objaśniającej we wszystkich czterdziestu modelach przedstawionych w Tab. 4.1 oraz w wielu tych z Tab. 3.2 oraz 3.4.

- Jak wykazały eksperymenty zawarte w rozdziale 5.1, na trzecim miejscu pod względem istotności okazały się być (sugerowane przez system 11 razy) parametry *low85* oraz *bpmmean*. Proponowane one były odpowiednio między 3 a 6 oraz 2 a 6 etapem serii. Nie pozostawiającym złudzeń potwierdzeniem dla słuszności użycia tych cech może być ich wystąpienie odpowiednio 38 i 45 razy na 54 prezentowane zestawy parametrów we wspomnianych poprzednio tabelach.
- Dla wspomnianego cyklu eksperymentów drugą najważniejszą zmienną objaśniającą okazała się *surfact*. Zaproponowano ją 10 razy, za każdym razem na 3 lub 4 etapie rozszerzania modeli. Natomiast w analizowanych tabelach cecha ta pojawiała się z umiarkowaną częstością, głównie zauważalną wśród modeli o małej liczności parametrów.
- Bezspornie najważniejszym parametrem dla predykcji BPD wydaje się parametr *gage*. Dość, że jego propozycje jako posiłkową cechę zaobserwowano 11 razy zawsze na drugim lub trzecim etapie, to znalazł się on bez wyjątku w każdym z modeli zawartych w Tab. 3.2, 3.4 oraz 4.1.

Podsumowując serię badań z Rozdz. 5.1, klasyfikator RL był w stanie osiągnąć trafność na poziomie 82.67% korzystając z sześciu lub siedmiu cech zachowując jednocześnie wartość ACC_{dev} poniżej 2. System zaproponował tylko cztery razy metodę SVM jako lepszą, przy czym najwyższy zaobserwowany wynik trafności predykcji osiągnięto właśnie tą metodą korzystając z 7 parametrów – $ACC = 83,04\%$, $ACC_{dev} = 2,6$, $TPR = 82,63\%$, $SPC = 83,28\%$. Co może być zaskakujące, do osiągnięcia 81,3% ACC wystarczyło użycie tylko trzech

parametrów i RL.

W drugim badaniu (Rozdz. 5.2) do wyboru cech użyto selekcji postępującej. W wypadku regresji logistycznej jako wynik otrzymano model składający się z: *gage*, *surfact*, *pda*, *bpmmean* oraz *high94*. Taki pięcioelementowy klasyfikator był w stanie osiągnąć trafność na poziomie 81,57% przy $ACC_{dev} = 1,49$ ($TPR = 84,03\%$, $SPC = 79,78\%$). Jest to dość dobry wynik, jednak nawet w specyficznym poprzednim eksperymencie (także dodającym cechy po jednej) znaleziono model tego samego rozmiaru osiągający: $ACC = 82,2\%$, $ACC_{dev} = 1,27$, $TPR = 86,06\%$, $SPC = 79,38\%$, co jest lepszym wynikiem, tak pod względem trafności, jak i czułości. Co warte zauważenia, w przeciwieństwie do metody selekcji postępującej, system wspomaganie decyzji był w stanie nadal rozbudowywać model maksymalizując ACC, aż do sześciu cech osiągając ostatecznie $ACC = 82,67\%$, podczas gdy algorytm selekcji postępującej nie był w stanie kontynuować pracy. Dla porównania, najlepszym możliwym pięcioparametrowym rezultatem (przedstawionym w Tab. 4.1) jest $ACC = 82,59\%$, $TPR = 84,65\%$, $SPC = 81,07\%$.

Jak wspomniano wcześniej, dla metody SVM selekcja postępująca okazała się nietrafna. Było to zapewne spowodowane bardzo małymi przyrostami ACC dla niewielkiej liczby parametrów, co zgodnie z wcześniejszymi badaniami jest charakterystyczne dla SVM. Głównym problemem jest jednak zapewne duże odchylenie standardowe wyników, przez które nie są one powtarzalne (szerzej przedyskutowano to zjawisko w Rozdz. 4.1). W zamian zastosowano algorytm eliminacji wstecznej, co doprowadziło do zbudowania predyktora, w którego skład wchodziły: *bweight*, *gage*, *aa*, *surfact*, *spO2mean*, *spO2dev*, *high94*, *bpmmean* and *spO2mean_tr*. Osiągnięto tym samym $ACC = 78,27\%$, $ACC_{dev} = 4,02$, $TPR = 68,56\%$, $SPC = 85,50\%$. W podobnym zadaniu wyeliminowania cech obniżających jakość predykcji system wspomaganie decyzji zaproponowałby

pierwszy z zestawów cech z Tab. 4.1 osiągając $ACC=83,29\%$. Jak pokazano w wypadku metody SVM, różnice pomiędzy wynikami oferowanymi przez eliminację wsteczną, a prezentowany system są znacznie większe niż dla RL. Zapewne optymalizacja parametrów γ oraz c lub użycie odpowiednio zaawansowanego algorytmu doboru parametrów modelu [Cios i in., 2005] dałoby bardziej zbliżone efekty do tych oferowanych przez prezentowany system.

Jasnym wnioskiem serii badań z Rozdz. 5.1, który przedstawiono na początku podrozdziału, jest wyższość regresji logistycznej nad SVM jako klasyfikatora. Należy jednak zauważyć, że ze względu na przebieg eksperymentów (rozbudowę modeli jedynie o jeden parametr na raz) rozważano modele o stosunkowo nielicznej grupie cech. Natomiast jak zauważono podczas wielu opisanych wcześniej badań, metoda SVM daje dobre wyniki dla sześciu i więcej-parametrowych modeli. Aby dopełnić więc obrazu, należy do wspomnianej analizy dołączyć obserwacje wynikające z przeglądu zamieszczonego w Tab. 4.1:

- Jednoznacznie wyższą specyficzność oferują modele oparte na SVM (do ponad 88,9%), podczas gdy dokładnie przeciwną zależność można zaobserwować dla RL i czułości predykcji (do 87,5%). SVM daje specyficzność aż o 7,2% wyższą. Różnica czułości na korzyść RL nie jest już jednak tak wielka mimo, że wynosi blisko 4,9%. W zastosowaniach praktycznych może to być bardzo przydatna właściwość, pozwalająca na łatwy dobór metody ze względu na specyfikę sytuacji pacjenta. Tym bardziej, że jak piszą Ryan i in. [1996] terapia BPD może mieć bardzo wiele niebezpiecznych efektów ubocznych, nie można więc jej stosować „na wszelki wypadek”. Dlatego też wysoka specyficzność metody SVM w praktyce może okazać się sporą zaletą.
- W Tab. 4.1 aż 7 najlepszych oraz 25 z 40 wszystkich prezentowanych to modele SVM. Mając do dyspozycji wszystkie rozważane cechy metoda pozwala osiągnąć $ACC=83,29\%$, $ACC_{dev} = 3,63$, $TPR=79,08\%$ oraz

SPC=86,4% co jest o 0,5% lepszym wynikiem pod względem trafności od RL. Wartość ACC_{dev} jest dość duża, w związku z tym system zaproponowałby alternatywny model SVM składający się z *aa*, *pda*, *low85*, *high94*, *bpmmean*, *bpmmean_tr* and *spO2dev_tr*. Stosując taki otrzymano by odchylenie standardowe trafności na poziomie 2,16 oraz ACC= 83.25%, TPR= 78,82% i SPC= 86,49%. Nadal jest to lepszy wynik niż ten możliwy do osiągnięcia przy zastosowaniu RL: ACC=82,79%, $ACC_{dev} = 1,11$, TPR= 84,2%, SPC= 81,73%. Warty zanotowania jest fakt, że w powyższym przykładzie system zaznaczyłby parametry *bweight*, *respimv* oraz *surfact* jako nieistotne. Jest to o tyle ciekawe, że szczególnie *bweight* i *surfact* w kontekście eksperymentów z Rozdz. 5.1 wymieniono jako nie kluczowe, ale jednak uzupełniające modele, dając poprawę ich skuteczności. Można to wytłumaczyć faktem, że jak już wspomiano dla małej liczby parametrów głównie rozważaną w eksperymentach metodą była RL, która charakteryzuje się całkowicie innym podejściem niż druga z metod. SVM preferowany jest, gdy system ma do dyspozycji wszystkie cechy – stąd powstały dysonans. Wbrew pozorom, że utrudnia to znalezienie najlepszego ogólnie uznanego modelu predykcji BPD, jest to bardzo korzystna cecha. Mając do dyspozycji system wspomaganie decyzji i metody o tak różnym „spojrzeniu”, niezależnie od dostępności danych lekarz otrzymuje zawsze możliwość wybrania metody adekwatnej do realnej sytuacji.

- Wielokrotnie wspomniano, że stosowanie metod SVM w predykcji BPD wiąże się z wyższą wartością ACC_{dev} . Zazwyczaj jest ona dwukrotnie większa niż dla RL (mieści się w przedziale 2,0–3,5), jednak istnieją i takie modele, czego dowodem jest występowanie dwóch w Tab. 4.1, dla których wartość jest mniejsza nawet od 2,0. Jest wysoce prawdopodobnym, że odchylenie standardowe ACC dla tej grupy metod mogłoby być znacznie

niższe dla większej próbki (Rozdz. 4.1). Niestety dotychczas nie było możliwości przeprowadzenia takich badań.

Tablica 5.1: Wyniki serii eksperymentów opisanych w Rozdz. 5.1 [Ochab i Wajs, 2016].

Parametr startowy	Cechy maksymalizujące ACC zaproponowane przez system w n -tym eksperymencie						
	1	2	3	4	5	6	7
<i>bweight</i>	\Rightarrow <i>respimv</i> 74.05% 76.1%						
<i>gage</i>	\Rightarrow <i>surfact</i>	\Rightarrow <i>low85</i>	\Rightarrow <i>bpmmean</i>	\Rightarrow <i>pda</i>	\Rightarrow <i>bweight</i>		
<i>respimv</i>	\Rightarrow <i>bweight</i> 74.05% 76.1%						
<i>aa</i>	\Rightarrow <i>gage</i>	\Rightarrow <i>surfact</i>	\Rightarrow <i>low85</i>	\Rightarrow <i>bpmmean</i>	\Rightarrow <i>pda</i>	\Rightarrow <i>respimv</i>	
<i>pda</i>	\Rightarrow <i>gage</i>	\Rightarrow <i>surfact</i>	\Rightarrow <i>low85</i>	\Rightarrow <i>bpmmean</i>	\Rightarrow <i>bweight</i>		
<i>surfact</i>	\Rightarrow <i>gage</i>	\Rightarrow <i>low85</i>	\Rightarrow <i>bpmmean</i>	\Rightarrow <i>pda</i>	\Rightarrow <i>bweight</i>		
<i>spO2mean</i>	\Rightarrow <i>bpmmean</i>	\Rightarrow <i>gage</i>	\Rightarrow <i>surfact</i>	\Rightarrow <i>pda</i>	\Rightarrow <i>low85</i>	\Rightarrow <i>bweight</i>	
<i>spO2dev</i>	\Rightarrow <i>gage</i>	\Rightarrow <i>surfact</i>	\Rightarrow <i>low85</i>	\Rightarrow <i>bpmmean</i>	\Rightarrow <i>pda</i>	\Rightarrow <i>bweight</i>	
<i>low85</i>	\Rightarrow <i>gage</i>	\Rightarrow <i>surfact</i>	\Rightarrow <i>bpmmean</i>	\Rightarrow <i>pda</i>	\Rightarrow <i>bweight</i>		
<i>high94</i>	\Rightarrow <i>gage</i>	\Rightarrow <i>surfact</i>	\Rightarrow <i>low85</i>	\Rightarrow <i>bpmmean</i>	\Rightarrow <i>pda</i>	\Rightarrow <i>bweight</i>	
<i>bpmmean</i>	\Rightarrow <i>spO2mean</i>	\Rightarrow <i>gage</i>	\Rightarrow <i>surfact</i>	\Rightarrow <i>pda</i>	\Rightarrow <i>low85</i>	\Rightarrow <i>bweight</i>	
<i>bpmmean_tr</i>	\Rightarrow <i>gage</i>	\Rightarrow <i>surfact</i>	\Rightarrow <i>low85</i>	\Rightarrow <i>bpmmean</i>	\Rightarrow <i>pda</i>	\Rightarrow <i>bweight</i>	
<i>spO2dev_tr</i>	\Rightarrow <i>high94</i>	\Rightarrow <i>gage</i>	\Rightarrow <i>surfact</i>	\Rightarrow <i>low85</i>	\Rightarrow <i>bpmmean</i>	\Rightarrow <i>pda</i>	\Rightarrow <i>bweight</i>
<i>spO2mean_tr</i>	\Rightarrow <i>spO2dev</i>	\Rightarrow <i>gage</i>	\Rightarrow <i>surfact</i>	\Rightarrow <i>low85</i>	\Rightarrow <i>bpmmean</i>	\Rightarrow <i>pda</i>	\Rightarrow <i>bweight</i>
	SVM ACC	RL ACC	SVM ACC	RL ACC	SVM ACC	RL ACC	SVM ACC

Tabela przedstawia sugerowane przez system wspomaganie decyzji dodatkowe cechy, o które warto rozbudować dany model. Każda kolumna zawiera rezultat n -tego eksperymentu w danej serii. W każdym wierszu kolumny poprzedzające wynik definiują model, który został wprowadzony na wejściu systemu. Na każdy wynik składa się proponowany nowy dodatkowy parametr oraz znajdujące się poniżej przewidywane wartości trafności predykcji nowo powstałego klasyfikatora – po lewej SVM, po prawej RL (wartość większa została wytłuszczona). Cechy, które na dowolnym dalszym etapie zostały uznane przez system za zbędne zostały zaznaczone przekreśleniem (przypadkowo we wszystkich wypadkach nastąpiło to na etapie trzecim).

Rozdział 6

Wnioski

W pracy pokazano, że możliwe jest stworzenie systemu wspomaganie decyzji wspomagającego predykcję dysplazji oskrzelowo-płucnej po pierwszym tygodniu życia pacjenta. Jak udowodniono, w systemie takim, obok sprawdzonej już metody regresji logistycznej, także algorytm SVM może zostać zastosowany z powodzeniem. Zaproponowana nowa, do tej pory nie wykorzystywana w predykcji BPD, metoda okazała się rozwiązaniem nie tyle konkurencyjnym, co raczej komplementarnym do tej dotychczas stosowanej. Autor ma tutaj na myśli szczególnie zastosowania praktyczne, w których możliwości wykonania wszelkich adekwatnych pomiarów są niejednokrotnie bardzo ograniczone. Jak jednak wykazano, optymalny zbiór zmiennych niezależnych stosowanych do przewidywania przyszłej diagnozy najczęściej jest różny dla każdej z tych metod. Dzięki wprowadzeniu dodatkowej metody, znacznie zwiększono więc szansę znalezienia skutecznego sposobu predykcji w przypadkach ograniczonej liczby pomiarów.

6.1 Istotność parametrów

Bez najmniejszych wątpliwości można stwierdzić, że najważniejszą cechą w predykcji BPD jest wiek płodowy. *Gage* był podstawą praktycznie wszystkich

wymienianych jako warte uwagi modelach, niezależnie od stosowanego algorytmu i jest to obserwacja zgodna z wnioskami wcześniej cytowanych badaczy [Bancalari i Claire, 2016, Bhering i in., 2007, Bhutani i Abbasi, 1992, Corcoran i in., 1993, Cunha i in., 2005, Groothuis i Makari, 2012, Kim i in., 2005, Laughon i in., 2011, Oh i in., 2005, Sinkin i in., 1990]. Dalsze rozważania nie dają już jednak tak jednoznacznych konkluzji. Obok parametru FiO_2 , którego wpływu nie było możliwości zbadać w tej pracy, jako kolejny najważniejszy z parametrów wymienia się w literaturze *pda* [Bancalari i Claire, 2016, Bhering i in., 2007, Corcoran i in., 1993, Cunha i in., 2005, Farstad i in., 2011, Groothuis i Makari, 2012, Kwinta i in., 2008b, Marshall i in., 1999, Oh i in., 2005, Rojas i in., 1995, Sosenko i Bancalari, 2012]. Patrząc na Tab. 4.1, która przedstawia 40 najlepszych pod względem trafności klasyfikatorów, z których każdy zawiera *pda*, można potwierdzić tę tezę. Należy wziąć jednak pod uwagę, że są to modele bardzo rozbudowane pod względem liczby parametrów. Kiedy natomiast popatrzeć na tabelę 3.2, obrazującą najlepsze modele pogrupowane według ich złożoności, okazuje się, że *pda* jest w czołówce, ale jednak na ex aequo 3 miejscu razem z *surfact* i *bpmmean*, pozostając za *low85* oraz *gage*. Ponadto *pda* w modelach 2- i 3-parametrowych, które są domeną RL, w ogóle nie występował. Także w tabeli 5.1 obrazującej eksperyment z użyciem systemu wspomaganie decyzji uwzględniający głównie „małe” modele RL, parametr ten występuje w 12 na 14 przypadków, jednak zawsze na dalszych pozycjach. Dokładnie przeciwna sytuacja ma miejsce w przypadku parametru *low85*, który osiąga lepsze noty dla mniej skomplikowanych modeli (zazwyczaj RL), niż w ogólnym rankingu zawierającym głównie rozbudowane predyktory (w większej części SVM). Natomiast bardzo podobnie do *pda* można scharakteryzować wspomniany już *bpmmean*, który jednak wypada nieco lepiej w sekwencyjnym teście systemu wspomaganie decyzji (Rozdz. 5.1). Pomimo tak obiecujących wyników *bpmmean* oraz *low85* nie są zbyt często wspominane w literaturze [Jobe, 2011, Stoch, 2007]. Znacznie częściej wymienia się zmienną

surfact [Ambalavanan i in., 2008, Cunha i in., 2005, Farstad i in., 2011, Marshall i in., 1999, Oh i in., 2005], która nie tylko jak wcześniej wspomniano znalazła się w czołówce modeli ułożonych wg. liczby użytych parametrów (Tab. 4.1), ale także w teście systemu wspomagania decyzji (Tab. 5.1) zajmowała kluczowe miejsce w tworzeniu kolejnych dość prostych modeli RL. Pomimo tego, w rankingu najlepszych predyktorów (Tab. 4.1) ta zmienna związana z użyciem surfaktantu znalazła się aż na przedostatnim miejscu, nie będąc potrzebną w skomplikowanych modelach bez względu na stosowany algorytm. Zmienna *aa* natomiast, jest w literaturze tylko dwukrotnie komentowana [Stoch, 2007, Subhedar i in., 1998] i po głębszej analizie wskazuje się na jej bezużyteczność. Także w prezentowanych w pracy badaniach cecha ta praktycznie nie występuje, gdy rozważać modele RL (niezależnie od liczby użytych parametrów). Co jednak niezwykle zaskakujące, w Tab. 4.1 zawierającej najtrafniejsze modele występuje ona aż w 21 z 25 predyktorów SVM. Kolejnym zaskakującym faktem jest wykazana mała użyteczność pomiaru masy urodzeniowej pacjenta, która (zgodnie z Tab. 5.1) w najlepszym wypadku potrafi poprawić trafność modelu o niecałe 0,5%. Tymczasem, jak wcześniej wspomniano w literaturze [Ambalavanan i in., 2008, Bancalari i Claire, 2016, Corcoran i in., 1993, Kim i in., 2005, Laughon i in., 2011, Oh i in., 2005, Rojas i in., 1995, Sinkin i in., 1990], szeroko komentuje się odpowiadający jej parametr *bweight* jako ważny czynnik ryzyka BPD. Inną niezbyt istotną cechą okazała się *spO2dev*, która znalazła się w grupie sporadycznie używanych zmiennych niezależnych wśród najtrafniejszych modeli, pomimo wystąpienia w kilku dość obiecujących predyktorach w tabeli 3.2. Nie jest to zaskakujące, gdyż w rozdziale 1.3 wykazano, że parametr ten jest współliniowy z wielokrotnie używaną cechą *low85*. Zależność tę można zaobserwować także w Tab. 5.1, gdzie w 7 wierszu, po dodaniu do modelu parametru *low85* system wskazał, że pomiar *spO2dev* nic już nie wnosi i może zostać usunięty. Do tej samej grupy mało istotnych parametrów należy zaliczyć *spO2mean*, który zdarza się nieczęsto w

modelach wyszczególnionych w tabelach. Jeszcze mniej istotne są parametry *respimv*, *bpmmean_tr*, *spO2mean_tr* oraz *spO2dev_tr*. Jak wykazano w rozdziale 1.3, dla tych zmiennych nie stwierdzono statystycznie istotnych różnic w rozkładach pomiędzy grupami z i bez BPD. Można by się więc pokusić o stwierdzenie, że analiza saturacji hemoglobiny we krwi oraz tętna w postaci trendów nie jest dobrym kierunkiem poszukiwań dla predykcji BPD. Zgodnie z [Askie i in., 2003, Geary i in., 2008, Jobe, 2011, Kinsella i in., 2006, Singer i in., 1992] oraz prezentowanymi wynikami analiza przebiegów czasowych tych wartości ma sens, ale głównie w postaci średniej tętna (*bpmmean*) oraz procenta czasu, podczas którego saturacja jest niższa od 85% (*low85*). Udowodniono, że pojedynczo zmienne *bpmmean_tr*, *spO2mean_tr*, *spO2dev_tr*, *spO2dev*, *spO2mean* charakteryzują się małą istotnością. Patrząc jednak na Tab. 4.1, zawierającą najskuteczniejsze predyktory, nie można się jednak oprzeć wrażeniu, że występują one w bardzo wielu najtrafniejszych modelach. Błędem byłoby więc wykluczenie powyższych zmiennych z procesu przygotowania bazy wiedzy tworzonego systemu wspomaganie decyzji tylko na podstawie wykonanych na wstępie analiz statystycznych. W ten sposób stracono by szansę na odkrycie chociażby 9 najlepszych modeli predykcyjnych SVM. Są jednakże i takie parametry, jak *bweight* oraz *respimv*, które bez większej szkody dla jakości predykcji mogłyby zostać pominięte w tym procesie, pomimo licznego ich poparcia w literaturze [Bhering i in., 2007, Groothuis i Makari, 2012, Kwinta i in., 2008b, Laughon i in., 2011, Ramos i in., 2016].

6.2 System wspomaganie decyzji

Powyższe podsumowanie istotności poszczególnych parametrów wyraźnie pokazuje, że stawiając sobie za cel wyjście poza rozważania teoretyczne, niemożliwe jest zdecydowanie się na stosowanie jednego najlepszego modelu, który byłby odpowiedni do zastosowania w praktyce. Z kolei mnogość obiecujących moż-

liwości zależnie od warunków jest tak wielka, że nie sposób dokonać wyboru modelu ręcznie. Powinien on być podyktowany głównie dostępnością do danych, uzależnioną od placówki medycznej w której leczy się pacjenta, ale także od ilości aktualnie dostępnego sprzętu [Subhedar i in., 1998]. Niejednokrotnie noworodki po wstępnej diagnozie w lokalnych placówkach są dopiero w późniejszym czasie przewożone do szpitali klinicznych. Nawet, gdyby istniała tam możliwość wykonania pomiarów wszystkich parametrów, nierealnym wydaje się przeniesienie danych ciągłych pomiędzy szpitalami, które zazwyczaj posiadają odmienne niekompatybilne systemy komputerowe. W wypadku danych statycznych sytuacja jest prostsza, chociaż należy mieć na uwadze, że każda ręczna ingerencja w dane (np. podczas ich przepisywania) może wprowadzać pewien błąd. Właśnie z tego rodzaju problemami spotkał się autor podejmując nieudaną próbę zebrania danych nastaw respiratorów opisaną w Rozdz. 1.2. Zastosowanie proponowanego systemu pozwala jednak na wiele więcej, niż odpowiednie dostosowanie modelu do typu i ilości posiadanych danych. W przeciwieństwie do pozostałych cytowanych prac, autor nie miał na celu uzyskania maksymalnej możliwej trafności dla jednego konkretnego przypadku. Intencją było umożliwienie oszacowania, z jak dużym błędem predykcji powinien się liczyć użytkownik, niezależnie od tego czy dysponuje identycznymi danymi uczącymi. Jak opisano dokładniej w rozdziale 4.1, wiąże się to z zaniżeniem otrzymywanych wyników, ale jednocześnie prowadzi do rezultatów bliższych tym otrzymywanym w praktyce. Ponadto, mając do dyspozycji kilka modeli o różnym stosunku czułości do specyficzności, lekarz uzyskuje komfort wyboru tego bardziej wg niego odpowiedniego. Jest to tym ważniejsze, że często stosowana przy podejrzeniu BPD terapia kortykosteroidami może mieć wiele niebezpiecznych efektów ubocznych [Ng, 1993, Ryan i in., 1996]. W takiej sytuacji byłoby więc bardzo porządane zastosowanie metody o wysokiej specyficzności [Akobeng, 2007]. Szczególnie przydatne okażą się wtedy nie stosowane nigdy wcześniej modele oparte o metodę SVM, które

charakteryzują się przewagą specyficzności na czułością. Stosując dotychczas proponowane modele regresji logistycznej, otrzymywano dokładnie przeciwną, w tym wypadku niekorzystną, zależność¹.

Autor zdaje sobie sprawę z faktu, że stworzony system jest dość prosty i istnieje szerokie pole do jego rozbudowy. Mając na uwadze, że nowa definicja [Bancalari i Claire, 2016, Jobe i Bancalari, 2001] wprowadza trzystopniową diagnozę BPD, bardzo ważne byłoby wprowadzenie poza analizą wystąpienia schorzenia, także badanie poziomu jego intensywności (dotychczas nie było to możliwe ze względu na brak dostępu do takich informacji). Celowym także wydawałoby się uzupełnienie danych o kolejne parametry związane z nastawą respiratorów oraz gazometrią krwi. Można mniemać, że największy skok jakościowy predykcji zapewniłoby użycie parametru FiO_2 . Wartym rozważenia wydaje się także użycie informacji o płci oraz pochodzeniu etnicznym pacjenta. Ponadto warto byłoby zaproponować taką zmianę struktury systemu, aby dodawanie kolejnych danych uczących nie wymagało czasochłonnego uaktualniania całej bazy danych. Bardzo korzystną byłaby także możliwość umieszczania w bazie informacji niekompletnych, które mimo to system byłby w stanie w pewien sposób wykorzystać. Niewątpliwie ciekawym okazałoby się zastosowanie i wzajemne porównanie większej liczby metod predykcji, np. sieci neuronowych czy algorytmu drzew decyzyjnych. Szczególnie przydatnym byłoby zintegrowanie w systemie narzędzia, pozwalającego poza wyborem optymalnego modelu, także na wykonanie odpowiadających mu obliczeń, wykonując tym samym predykcję przyszłej diagnozy dla konkretnego przypadku.

Jak pokazują wykonane przez autora badania, pomimo wieloletnich prac badaczy, problem zastosowania predykcji BPD w praktyce pozostawia nadal szerokie pole do dalszych działań badawczych. Prezentowana praca proponuje jednak pewien nowy sposób podejścia do opisywanego problemu, który jak się

¹W wypadku regresji logistycznej stosunek ten można co prawda próbować zniwelować dobozem, nieoptymalnej względem trafności, wartości odcięcia (Rozdz. 2.3).

wydaże zbliża do powstania w przyszłości narzędzi, mogących mieć zastosowanie w codziennej pracy lekarza.

Bibliografia

- Akobeng, A. K. (2007). Understanding diagnostic tests 1: sensitivity, specificity and predictive values. *Acta Paediatrica*, 96(3):338–341.
- Ambalavanan, N., Van Meurs, K. P., Perritt, R., Carlo, W. A., Ehrenkranz, R. A., Stevenson, D. K., Lemons, J. A., Poole, W. K., and Higgins, R. D. (2008). Predictors of death or bronchopulmonary dysplasia in preterm infants with respiratory failure. *Journal of Perinatology*, 28(6):420–426.
- Askie, L. M., Henderson-Smart, D. J., Irwig, L., and Simpson, J. M. (2003). Oxygen-saturation targets and outcomes in extremely preterm infants. *New England Journal of Medicine*, 349(10):959–967.
- Bancalari, E. and Claure, N. (2016). *Bronchopulmonary Dysplasia*, chapter Bronchopulmonary Dysplasia: Definitions and Epidemiology, str. 167–182. In Bhandari [2016].
- Bhandari, V., editor (2016). *Bronchopulmonary Dysplasia*. Springer.
- Bhering, C. A., Mochdece, C. C., Moreira, M. E., Rocco, J. R., and Sant’Anna, G. M. (2007). Bronchopulmonary dysplasia prediction model for 7-day-old infants. *Jornal de pediatria*, 83(2):163–170.
- Bhutani, V. K. and Abbasi, S. (1992). Relative likelihood of bronchopulmonary dysplasia based on pulmonary mechanics measured in preterm neonates during the first week of life. *The Journal of pediatrics*, 120(4):605–613.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167.
- Chang, C.-C. and Lin, C.-b. (2001). Training v-support vector classifiers: theory and algorithms. *Neural computation*, 13(9):2119–2147.
- Chang, C.-C. and Lin, C.-J. (2011). Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Cios, K. J., Mamitsuka, H., Nagashima, T., and Tadeusiewicz, R. (2005). Computational intelligence in solving bioinformatics problems. *Artificial Intelligence in Medicine*, 35(1):1–8.
- Corcoran, J., Patterson, C., Thomas, P., and Halliday, H. (1993). Reduction in the risk of bronchopulmonary dysplasia from 1980–1990: results of a multivariate logistic

- regression analysis. *European journal of pediatrics*, 152(8):677–681.
- Cortes, C. and Vapnik, V. (1995). Support vector machine. *Machine learning*, 20(3):273–297.
- Cunha, G. S., Mezzacappa-Filho, F., and Ribeiro, J. D. (2005). Risk factors for bronchopulmonary dysplasia in very low birth weight newborns treated with mechanical ventilation in the first week of life. *Journal of tropical pediatrics*, 51(6):334–340.
- De Paepe, M. E. (2016). *Bronchopulmonary Dysplasia*, chapter Pathology of Bronchopulmonary Dysplasia, str. 149–164. In Bhandari [2016].
- Draper, N. R. and Smith, H. (2014). *Applied regression analysis*. John Wiley & Sons.
- Farstad, T., Bratlid, D., Medbø, S., and Markestad, T. (2011). Bronchopulmonary dysplasia—prevalence, severity and predictive factors in a national cohort of extremely premature infants. *Acta paediatrica*, 100(1):53–58.
- Fay, M. P. and Proschan, M. A. (2010). Wilcoxon-mann-whitney or t-test? on assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics surveys*, 4:1.
- Geary, C., Caskey, M., Fonseca, R., and Malloy, M. (2008). Decreased incidence of bronchopulmonary dysplasia after early management changes, including surfactant and nasal continuous positive airway pressure treatment at delivery, lowered oxygen saturation goals, and early amino acid administration: a historical cohort study. *Pediatrics*, 121(1):89–96.
- Gilbert, R. and Keighley, J. (1974). The arterial-alveolar oxygen tension ratio. an index of gas exchange applicable to varying inspired oxygen concentrations. *The American review of respiratory disease*, 109(1):142.
- Głowacka, E. and Lis, G. (2008). Dysplazja oskrzelowo-płucna—wczesne i odległe następstwa w zakresie układu oddechowego. *Pneumonologia i Alergologia Polska*, 76(6):437–445.
- Grégoire, M.-C., Lefebvre, F., and Glorieux, J. (1998). Health and developmental outcomes at 18 months in very preterm infants with bronchopulmonary dysplasia. *Pediatrics*, 101(5):856–860.
- Groothuis, J. R. and Makari, D. (2012). Definition and outpatient management of the very low-birth-weight infant with bronchopulmonary dysplasia. *Advances in therapy*, 29(4):297–311.
- Harrell, F. E. (2013). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer Science & Business Media.
- Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied logistic regression*, volume 398. John Wiley & Sons.
- Jobe, A. H. (2011). The new bronchopulmonary dysplasia. *Current opinion in pediatrics*, 23(2):167.

-
- Jobe, A. H. and Bancalari, E. (2001). Bronchopulmonary dysplasia. *American journal of respiratory and critical care medicine*, 163(7):1723–1729.
- Johnson, B.-H. and Thebaud, B. (2014). *Workbook in practical neonatology*, chapter Bronchopulmonary Dysplasia. Elsevier Health Sciences.
- Jones, H. L. (1974). Jackknife estimation of functions of stratum means. *Biometrika*, 61(2):343–348.
- Kim, Y. D., Kim, E. A.-R., Kim, K.-S., Pi, S.-Y., and Kang, W. (2005). Scoring method for early prediction of neonatal chronic lung disease using modified respiratory parameters. *Journal of Korean medical science*, 20(3):397–401.
- Kim, Y. D., Kim, K. S., Kim, E. A. R., Lee, J. J., Park, S. J., and Pi, S. Y. (2001). Perinatal risk factors for the development of bronchopulmonary dysplasia in premature infants less than 32 weeks’ gestation. *Journal of the Korean Society of Neonatology*, 8(1):78–93.
- Kinsella, J. P., Greenough, A., and Abman, S. H. (2006). Bronchopulmonary dysplasia. *The Lancet*, 367(9520):1421–1431.
- Kruczek, P. and Kwinta, P. (2007). Klinika Chorób Dzieci Uniwersytetu Jagiellońskiego w Krakowie (komunikacja osobista).
- Krzyśko, M., Wołyński, W., Górecki, T., and Skorzybut, M. (2008). *Systemy uczące. Rozpoznawanie wzorców, analiza skupień i redukcja wymiarowości*. WNT, Warszawa.
- Kuenzel, L. (2011). Predicting and undestanding bronchopulmonary dysplasia in permature infants. *Stanford Undergraduate Research Journal*, 10:36–43.
- Kurcz, J. (2013). Zakład Radiologii Ogólnej, Zabiegowej i Neuroradiologii Katedry Radiologii Uniwersytetu Medycznego we Wrocławiu (komunikacja osobista, 23.08.2013).
- Kwinta, P., Bik-Multanowski, M., Mitkowska, Z., Tomasik, T., Legutko, M., and Pietrzyk, J. J. (2008a). Genetic risk factors of bronchopulmonary dysplasia. *Pediatric research*, 64(6):682–688.
- Kwinta, P., Grudzień, A., Pawlik, D., Olechowski, W., Lauterbach, R., and Pietrzyk, J. (2008b). [prevalence and risk factors of bronchopulmonary dysplasia among extremely low birth weight newborns of regional birth cohort of south-east poland]. *Przegląd lekarski*, 66(1-2):14–20.
- Kwinta, P. and Pietrzyk, J. J. (2010). Preterm birth and respiratory disease in later life. *Expert review of respiratory medicine*, 4(5):593–604.
- Kwinta, P., Tomasik, T., Klimek, M., Cichocka-Jarosz, E., Lis, G., and Pietrzyk, J. (2009). Health status at the age of 5-7 years of preterm infants with and without bronchopulmonary dysplasia. *Przegląd lekarski*, 66(1-2):21.
- Landry, J. S. and Banbury, S. P. (2016). *Bronchopulmonary Dysplasia*, chapter Pulmonary Function in Survivors of Bronchopulmonary Dysplasia, str. 281–295. In Bhandari [2016].
- Larose, D. T. (2006). *Data mining methods & models*. John Wiley & Sons.

- Laughon, M. M., Langer, J. C., Bose, C. L., Smith, P. B., Ambalavanan, N., Kennedy, K. A., Stoll, B. J., Buchter, S., Laptook, A. R., Ehrenkranz, R. A., et al. (2011). Prediction of bronchopulmonary dysplasia by postnatal age in extremely premature infants. *American journal of respiratory and critical care medicine*, 183(12):1715–1722.
- Lavoie, P. M. (2016). *Bronchopulmonary Dysplasia*, chapter Genetics of Bronchopulmonary Dysplasia, str. 109–127. In Bhandari [2016].
- Lin, H.-T. and Lin, C.-J. (2003). A study on sigmoid kernels for svm and the training of non-psd kernels by smo-type methods. *submitted to Neural Computation*, str. 1–32.
- Ludbrook, J. (2008). Analysis of 2×2 tables of frequencies: matching test to experimental design. *International journal of epidemiology*, 37(6):1430–1435.
- Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, str. 50–60.
- Marshall, D. D., Kotelchuck, M., Young, T. E., Bose, C. L., Kruyer, L., and O’Shea, T. M. (1999). Risk factors for chronic lung disease in the surfactant era: a north carolina population-based study of very low birth weight infants. *Pediatrics*, 104(6):1345–1350.
- Moczko, J. A., Bręborowicz, G. H., and Tadeusiewicz, R. (1998). Statystyka w badaniach medycznych.
- Ng, P. (1993). The effectiveness and side effects of dexamethasone in preterm infants with bronchopulmonary dysplasia. *Archives of disease in childhood*, 68(3 Spec No):330.
- Noack, G., Mortensson, W., Robertson, B., and Nilsson, R. (1993). Correlations between radiological and cytological findings in early development of bronchopulmonary dysplasia. *European journal of pediatrics*, 152(12):1024–1029.
- Northway Jr, W. H., Rosan, R. C., and Porter, D. Y. (1967). Pulmonary disease following respirator therapy of hyaline-membrane disease: bronchopulmonary dysplasia. *New England Journal of Medicine*, 276(7):357–368.
- Obozinski, G. (2006). Skmsmo – support kernel machine solved by sequential minimal optimization. <http://imagine.enpc.fr/~obozinsg/code.html>.
- Ochab, M. and Wajs, W. (2014a). Bronchopulmonary dysplasia prediction using support vector machine and libsvm. In *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, volume 2 of *Annals of Computer Science and Information Systems*, str. 201–208. IEEE.
- Ochab, M. and Wajs, W. (2014b). Bronchopulmonary dysplasia prediction using support vector machine and logit regression. *Information Technologies in Biomedicine*, 4:365–374.
- Ochab, M. and Wajs, W. (2016). Expert system supporting an early prediction of the bronchopulmonary dysplasia. *Computers in biology and medicine*, 69:236–244.

-
- Oh, W., Poindexter, B., Perritt, R., Lemons, J., Bauer, C., Ehrenkranz, R., Stoll, B., Poole, K., and Wright, L. (2005). Neonatal research network. association between fluid intake and weight loss during the first ten days of life and risk of bronchopulmonary dysplasia in extremely low birth weight infants. *J Pediatr*, 147(6):786–790.
- Platt, J. et al. (1998). Sequential minimal optimization: A fast algorithm for training support vector machines.
- Ramos, L. M., Najrana, T., and Sanchez-Esteban, J. (2016). *Bronchopulmonary Dysplasia*, chapter Invasive Mechanical Ventilation in the Pathogenesis of Bronchopulmonary Dysplasia, str. 27–54. In Bhandari [2016].
- Rojas, M. A., Gonzalez, A., Bancalari, E., Claire, N., Poole, C., and Silva-Neto, G. (1995). Changing trends in the epidemiology and pathogenesis of neonatal chronic lung disease. *The Journal of pediatrics*, 126(4):605–610.
- Romagnoli, C., Zecca, E., Tortorolo, L., Vento, G., and Tortorolo, G. (1998). A scoring system to predict the evolution of respiratory distress syndrome into chronic lung disease in preterm infants. *Intensive care medicine*, 24(5):476–480.
- Rozycki, H. J. and Narla, L. (1996). Early versus late identification of infants at high risk of developing moderate to severe bronchopulmonary dysplasia. *Pediatric pulmonology*, 21(6):345–352.
- Ryan, S., Nycyk, J., and Shaw, B. (1996). Prediction of chronic neonatal lung disease on day 4 of life. *European journal of pediatrics*, 155(8):668–671.
- Sauve, R. S. and Singhal, N. (1985). Long-term morbidity of infants with bronchopulmonary dysplasia. *Pediatrics*, 76(5):725–733.
- Schölkopf, B., Smola, A. J., Williamson, R. C., and Bartlett, P. L. (2000). New support vector algorithms. *Neural computation*, 12(5):1207–1245.
- Shapiro, S. and Wilk, M. (1965). An analysis of variance test for normality. *Biometrika*, 52(3):591–611.
- Shin, S.-M., Kim, W. S., Cheon, J.-E., Kim, H. S., Lee, W., Jung, A. Y., Kim, I.-O., and Choi, J. H. (2013). Bronchopulmonary dysplasia: new high resolution computed tomography scoring system and correlation between the high resolution computed tomography score and clinical severity. *Korean journal of radiology*, 14(2):350–360.
- Singer, L., Martin, R. J., Hawkins, S. W., Yamashita, T. S., Benson-Szekely, L. J., and Carlo, W. A. (1992). Oxygen desaturation complicates feeding in infants with bronchopulmonary dysplasia after discharge. *Pediatrics*, 90(3):380–384.
- Sinkin, R. A., Cox, C., and Phelps, D. L. (1990). Predicting risk for bronchopulmonary dysplasia: selection criteria for clinical trials. *Pediatrics*, 86(5):728–736.
- Skidmore, M. D., Rivers, A., and Hack, M. (1990). Increased risk of cerebral palsy among very low-birthweight infants with chronic lung disease. *Developmental Medicine & Child Neurology*, 32(4):325–332.
- Somaschini, M., Castiglioni, E., Volonteri, C., Cursi, M., Ferrari, M., and Carrera, P. (2012). Genetic predisposing factors to bronchopulmonary dysplasia: preliminary

- data from a multicentre study. *The Journal of Maternal-Fetal & Neonatal Medicine*, 25(sup4):119–122.
- Sosenko, I. and Bancalari, E. (2012). New developments in the pathogenesis and prevention of bronchopulmonary dysplasia. *The Newborn Lung: Neonatology Questions and Controversies: Expert Consult-Online and Print*, str. 217–233.
- Stoch, P. (2007). Zastosowanie narzędzi statystycznych i matematycznych metod sztucznej inteligencji do predykcji wystąpienia dysplazji oskrzelowo-płucnej u noworodków. *Praca doktorska*, Akademia Górniczo-Hutnicza, Kraków, str. 60–72.
- Subhedar, N., Hamdan, A., Ryan, S., and Shaw, N. (1998). Pulmonary artery pressure: early predictor of chronic lung disease in preterm infants. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, 78(1):F20–F24.
- Tadeusiewicz, R., Izworski, A., and Majewski, J. (1993). *Biometria*. Wydawnictwa AGH.
- Tadeusiewicz, R. and Ogiela, M. R. (2004). *Medical image understanding technology*. Springer.
- Toce, S. S., Farrell, P. M., Leavitt, L. A., Samuels, D. P., and Edwards, D. K. (1984). Clinical and roentgenographic scoring systems for assessing bronchopulmonary dysplasia. *American journal of diseases of children*, 138(6):581–585.
- Vapnik, V. (2013). *The nature of statistical learning theory*. Springer Science & Business Media.
- Vohr, B. R., Coll, C. G., Lobato, D., Yunis, K. A., O’Dea, C., and Oh, W. (1991). Neurodevelopmental and medical status of low-birthweight survivors of bronchopulmonary dysplasia at 10 to 12 years of age. *Developmental Medicine & Child Neurology*, 33(8):690–697.
- Wajs, W., Wais, P., Ochab, M., and Wojtowicz, H. (2016a). Arterial blood gases forecast optimization by artificial neural network method. In *Information Technologies in Medicine*, str. 433–444. Springer.
- Wajs, W., Wojtowicz, H., Wais, P., and Ochab, M. (2016b). Statistical method for the problem of bronchopulmonary dysplasia classification in pre-mature infants. In *Intelligent Decision Technologies 2016*, volume 56 of *Smart Innovation, Systems and Technologies*, str. 165–179. Springer.
- Walsh, M. C., Szefer, S., Davis, J., Allen, M., Van Marter, L., Abman, S., Blackmon, L., and Jobe, A. (2006). Summary proceedings from the bronchopulmonary dysplasia group. *Pediatrics*, 117(Supplement 1):S52–S56.
- Wardlaw, T., Blanc, A., Zupan, J., and Ahman, E. (2005). Low birthweight: Country, regional and global estimates. *United Nations Children’s Fund and World Health Organization*.
- Weinstein, M. R., Peters, M. E., Sadek, M., and Palta, M. (1994). A new radiographic scoring system for bronchopulmonary dysplasia. *Pediatric pulmonology*, 18(5):284–289.

-
- Yates, R. (2012). *Rennie & Robertson's textbook of neonatology*, chapter Cardiovascular disease. Elsevier Health Sciences.
- Yuksel, B., Greenough, A., and Karani, J. (1993). Prediction of chronic lung disease from the chest radiograph appearance at seven days of age. *Acta Paediatrica*, 82(12):944–947.

Dodatek A

A.1 Przykład liczbowy predykcji BPD metodą regresji logistycznej

Obliczenia wykonano dla dwuparametrowego modelu zamieszczonego w Tab. 3.2, który opiera się na zmiennych niezależnych *gage* oraz *surfact*. Dla czytelności przykładu wzięto pod uwagę rzeczywiste dane tylko 10 pacjentów, zawarte w Tab. A.1.

Zgodnie ze wzorami (2.1) oraz (2.3) wyznaczamy logarytm naturalny funkcji L :

$$p_k = P(\tilde{y}_k = 1 | X_k) = \frac{e^{a_0 + \sum_{i=1}^n a_i x_{i,k}}}{1 + e^{a_0 + \sum_{i=1}^n a_i x_{i,k}}}, \quad (\text{A.1})$$

$$\ln(L) = \sum_{k=1}^m [\tilde{y}_k \cdot \ln(p_k) + (1 - \tilde{y}_k) \cdot \ln(1 - p_k)], \quad (\text{A.2})$$

podstawiamy do wzoru (A.2) \tilde{y}_k tj. wynik diagnozy BPD k -tego pacjenta

$$\begin{aligned} \ln(L) = & 0 \cdot \ln(p_1) + 1 \cdot \ln(1 - p_1) + 0 \cdot \ln(p_2) + 1 \cdot \ln(1 - p_2) + \\ & 0 \cdot \ln(p_3) + 1 \cdot \ln(1 - p_3) + 0 \cdot \ln(p_4) + 1 \cdot \ln(1 - p_4) + \\ & 0 \cdot \ln(p_5) + 1 \cdot \ln(1 - p_5) + 1 \cdot \ln(p_6) + 0 \cdot \ln(1 - p_6) + \\ & 1 \cdot \ln(p_7) + 0 \cdot \ln(1 - p_7) + 1 \cdot \ln(p_8) + 0 \cdot \ln(1 - p_8) + \\ & 1 \cdot \ln(p_9) + 0 \cdot \ln(1 - p_9) + 1 \cdot \ln(p_{10}) + 0 \cdot \ln(1 - p_{10}), \end{aligned} \quad (\text{A.3})$$

po wykonaniu mnożeń otrzymujemy

$$\begin{aligned} \ln(L) = & \ln(1 - p_1) + \ln(1 - p_2) + \ln(1 - p_3) + \\ & \ln(1 - p_4) + \ln(1 - p_5) + \ln(p_6) + \\ & \ln(p_7) + \ln(p_8) + \ln(p_9) + \ln(p_{10}), \end{aligned} \quad (\text{A.4})$$

Tablica A.1: Dane dziesięciu pacjentów użyte w przykładzie liczbowym

nr pacjenta – k	$gage - x_{1,k}$	$surfact - x_{2,k}$	$BPD - \tilde{y}_k$
1	26	1	0
2	28	0	0
3	31	0	0
4	30	1	0
5	34	0	0
6	25	0	1
7	28	0	1
8	29	1	1
9	24	1	1
10	30	1	1

podstawiamy p_k ze wzoru (A.1), przy czym $n = 2$, bo rozważamy dwa parametry, $x_{1,k}$ to wartość parametru $gage$ k -tego pacjenta, $x_{2,k}$ to wartość parametru $surfact$ k -tego pacjenta (w przykładzie nie przeprowadzono normalizacji parametrów – dla regresji logistycznej nie jest ona zresztą konieczna)

$$\begin{aligned}
\ln(L) = & \ln\left(1 - \frac{e^{a_0+a_1x_{1,1}+a_2x_{2,1}}}{1 + e^{a_0+a_1x_{1,1}+a_2x_{2,1}}}\right) + \ln\left(1 - \frac{e^{a_0+a_1x_{1,2}+a_2x_{2,2}}}{1 + e^{a_0+a_1x_{1,2}+a_2x_{2,2}}}\right) + \ln\left(1 - \frac{e^{a_0+a_1x_{1,3}+a_2x_{2,3}}}{1 + e^{a_0+a_1x_{1,3}+a_2x_{2,3}}}\right) + \\
& \ln\left(1 - \frac{e^{a_0+a_1x_{1,4}+a_2x_{2,4}}}{1 + e^{a_0+a_1x_{1,4}+a_2x_{2,4}}}\right) + \ln\left(1 - \frac{e^{a_0+a_1x_{1,5}+a_2x_{2,5}}}{1 + e^{a_0+a_1x_{1,5}+a_2x_{2,5}}}\right) + \ln\left(\frac{e^{a_0+a_1x_{1,6}+a_2x_{2,6}}}{1 + e^{a_0+a_1x_{1,6}+a_2x_{2,6}}}\right) + \\
& \ln\left(\frac{e^{a_0+a_1x_{1,7}+a_2x_{2,7}}}{1 + e^{a_0+a_1x_{1,7}+a_2x_{2,7}}}\right) + \ln\left(\frac{e^{a_0+a_1x_{1,8}+a_2x_{2,8}}}{1 + e^{a_0+a_1x_{1,8}+a_2x_{2,8}}}\right) + \ln\left(\frac{e^{a_0+a_1x_{1,9}+a_2x_{2,9}}}{1 + e^{a_0+a_1x_{1,9}+a_2x_{2,9}}}\right) + \\
& \ln\left(\frac{e^{a_0+a_1x_{1,10}+a_2x_{2,10}}}{1 + e^{a_0+a_1x_{1,10}+a_2x_{2,10}}}\right),
\end{aligned} \tag{A.5}$$

po podstawieniu za x otrzymujemy

$$\begin{aligned}
\ln(L) = & \ln\left(1 - \frac{e^{a_0+26\cdot a_1+a_2}}{1 + e^{a_0+26\cdot a_1+a_2}}\right) + \ln\left(1 - \frac{e^{a_0+28\cdot a_1}}{1 + e^{a_0+28\cdot a_1}}\right) + \ln\left(1 - \frac{e^{a_0+31\cdot a_1}}{1 + e^{a_0+31\cdot a_1}}\right) + \\
& \ln\left(1 - \frac{e^{a_0+30\cdot a_1+a_2}}{1 + e^{a_0+30\cdot a_1+a_2}}\right) + \ln\left(1 - \frac{e^{a_0+34a_1}}{1 + e^{a_0+34a_1}}\right) + \ln\left(\frac{e^{a_0+25\cdot a_1}}{1 + e^{a_0+25\cdot a_1}}\right) + \\
& \ln\left(\frac{e^{a_0+28\cdot a_1}}{1 + e^{a_0+28\cdot a_1}}\right) + \ln\left(\frac{e^{a_0+29\cdot a_1+a_2}}{1 + e^{a_0+29\cdot a_1+a_2}}\right) + \ln\left(\frac{e^{a_0+24\cdot a_1+a_2}}{1 + e^{a_0+24\cdot a_1+a_2}}\right) + \\
& \ln\left(\frac{e^{a_0+30\cdot a_1+a_2}}{1 + e^{a_0+30\cdot a_1+a_2}}\right).
\end{aligned} \tag{A.6}$$

W wyniku poszukiwania minimum funkcji $\ln(L)$ otrzymujemy:

$$a_0 = 10,87, \quad (\text{A.7})$$

$$a_1 = -0,39, \quad (\text{A.8})$$

$$a_2 = 0,53, \quad (\text{A.9})$$

$$(\text{A.10})$$

Po podstawieniu współczynników a_i do wzoru (A.1) otrzymujemy gotowy wzór na prawdopodobieństwo pozytywnej diagnozy BPD dla pacjenta o $x_1 = \textit{gage}$ oraz $x_2 = \textit{surfact}$:

$$p = P(\tilde{y} = 1|X) = \frac{e^{10,87-0,39 \cdot x_1+0,53 \cdot x_2}}{1 + e^{10,87-0,39 \cdot x_1+0,53 \cdot x_2}}. \quad (\text{A.11})$$

Dla przykładu prawdopodobieństwo wystąpienia BPD u pacjenta urodzonego w 32 tygodniu ciąży, któremu nie podawano surfaktantu będzie wynosić:

$$p = P(\tilde{y} = 1|x_1 = 32, x_2 = 0) = \frac{e^{10,87-0,39 \cdot 32+0,53 \cdot 0}}{1 + e^{10,87-0,39 \cdot 32+0,53 \cdot 0}} = \frac{e^{-1.61}}{1 + e^{-1.61}} = 0,1666 = 16,6\%. \quad (\text{A.12})$$

A.2 Przykład liczbowy predykcji BPD metodą SVM

Znormalizowane dane z Tab. A.1 pochodzące z poprzedniego przykładu zamieszczono w Tab. A.2.

Tablica A.2: Dane 10 pacjentów z przykładu A.1 znormalizowane do przedziału $[-1,1]$

nr pacjenta – k	$gage - x_{1,k}$	$surfact - x_{2,k}$	$BPD - y_k$
1	-0,6	1	-1
2	-0,2	-1	-1
3	0,4	-1	-1
4	0,2	1	-1
5	1	-1	-1
6	-0,8	-1	1
7	-0,2	-1	1
8	0	1	1
9	-1	1	1
10	0,2	1	1

Jako funkcję jądrową wybrano sigmoide (2.19). Dla prostoty obliczeń zakładamy, że wartości parametrów przyjmują wartości $\gamma = 2$, $r = -4$ oraz $c = 2$. Zgodnie ze wzorami (2.13–2.12) definiujemy problem dualny:

$$\min_{\alpha} = \frac{1}{2} \alpha^T \cdot Q \alpha - E^T \cdot \alpha \quad (\text{A.13})$$

przy ograniczeniach:

$$Y \cdot \alpha = 0, \quad (\text{A.14})$$

$$0 \leq \alpha_k \leq c, \quad k = 1, \dots, m. \quad (\text{A.15})$$

gdzie $Y = [y_1, \dots, y_m]^T$, $E = [1, \dots, 1]^T$ jest wektorem jedynek, Q jest macierzą o wymiarach $m \times m$, $Q_{ij} = y_i y_j K(X_i, X_j)$. Funkcja jądrowa ma postać:

$$K(X_i, X_j) = \tanh(\gamma X_i^T X_j + r). \quad (\text{A.16})$$

Po podstawieniu macierz Q będzie miała postać:

$$Q = \tanh \left(\begin{bmatrix} -1,28 & -5,76 & -6,48 & -2,24 & -7,20 & 5,04 & 5,76 & 2,00 & 0,80 & 2,24 \\ -5,76 & -1,92 & -2,16 & -6,08 & -2,40 & 1,68 & 1,92 & 6,00 & 5,60 & 6,08 \\ -6,48 & -2,16 & -1,68 & -5,84 & -1,20 & 2,64 & 2,16 & 6,00 & 6,80 & 5,84 \\ -2,24 & -6,08 & -5,84 & -1,92 & -5,60 & 6,32 & 6,08 & 2,00 & 2,40 & 1,92 \\ -7,20 & -2,40 & -1,20 & -5,60 & 0 & 3,60 & 2,40 & 6,00 & 8,00 & 5,60 \\ 5,04 & 1,68 & 2,64 & 6,32 & 3,60 & -0,72 & -1,68 & -6,00 & -4,40 & -6,32 \\ 5,76 & 1,92 & 2,16 & 6,08 & 2,40 & -1,68 & -1,92 & -6,00 & -5,60 & -6,08 \\ 2,00 & 6,00 & 6,00 & 2,00 & 6,00 & -6,00 & -6,00 & -2,00 & -2,00 & -2,00 \\ 0,80 & 5,60 & 6,80 & 2,40 & 8,00 & -4,40 & -5,60 & -2,00 & 0 & -2,40 \\ 2,24 & 6,08 & 5,84 & 1,92 & 5,60 & -6,32 & -6,08 & -2,00 & -2,40 & -1,92 \end{bmatrix} \right). \quad (\text{A.17})$$

Ostatecznie problem przyjmie postać:

123

$$\min_{\alpha} = \frac{1}{2} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \\ \alpha_9 \\ \alpha_{10} \end{bmatrix}^T \begin{bmatrix} -0,86 & -1,00 & -1,00 & -0,98 & -1,00 & 1,00 & 1,00 & 0,96 & 0,66 & 0,98 \\ -1,00 & -0,96 & -0,97 & -1,00 & -0,98 & 0,93 & 0,96 & 1,00 & 1,00 & 1,00 \\ -1,00 & -0,97 & -0,93 & -1,00 & -0,83 & 0,99 & 0,97 & 1,00 & 1,00 & 1,00 \\ -0,98 & -1,00 & -1,00 & -0,96 & -1,00 & 1,00 & 1,00 & 0,96 & 0,98 & 0,96 \\ -1,00 & -0,98 & -0,83 & -1,00 & 0 & 1,00 & 0,98 & 1,00 & 1,00 & 1,00 \\ 1,00 & 0,93 & 0,99 & 1,00 & 1,00 & -0,62 & -0,93 & -1,00 & -1,00 & -1,00 \\ 1,00 & 0,96 & 0,97 & 1,00 & 0,98 & -0,93 & -0,96 & -1,00 & -1,00 & -1,00 \\ 0,96 & 1,00 & 1,00 & 0,96 & 1,00 & -1,00 & -1,00 & -0,96 & -0,96 & -0,96 \\ 0,66 & 1,00 & 1,00 & 0,98 & 1,00 & -1,00 & -1,00 & -0,96 & 0 & -0,98 \\ 0,98 & 1,00 & 1,00 & 0,96 & 1,00 & -1,00 & -1,00 & -0,96 & -0,98 & -0,96 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \\ \alpha_9 \\ \alpha_{10} \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}^T \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \\ \alpha_9 \\ \alpha_{10} \end{bmatrix} \quad (\text{A.18})$$

przy ograniczeniach:

$$\alpha_6 - \alpha_2 - \alpha_3 - \alpha_4 - \alpha_5 - \alpha_1 + \alpha_7 + \alpha_8 + \alpha_9 + \alpha_{10} = 0, \quad (\text{A.19})$$

$$0 \leq \alpha_k \leq 1, \quad k = 1, \dots, m. \quad (\text{A.20})$$

Po zastosowaniu algorytmu SMO (do obliczeń wykorzystano bibliotekę SKMsmo [Obozinski, 2006]) otrzymujemy współczynniki α_k będące rozwiązaniem zdefiniowanego problemu:

$$\boldsymbol{\alpha} = \begin{bmatrix} 2,000 \\ 2,000 \\ 2,000 \\ 2,000 \\ 1,016 \\ 1,763 \\ 2,000 \\ 2,000 \\ 1,253 \\ 2,000 \end{bmatrix} \quad b = 0,346 \quad (\text{A.21})$$

Biblioteka zwraca także wartość b , można ją jednak łatwo wyznaczyć ze wzorów (2.16–2.18):

$$\nabla f(\boldsymbol{\alpha}) \equiv Q\boldsymbol{\alpha} - E = \begin{bmatrix} -1,21 \\ -1,05 \\ -0,71 \\ -1,05 \\ 0,35 \\ -0,35 \\ -0,95 \\ -0,95 \\ -0,35 \\ -0,95 \end{bmatrix}, \quad (\text{A.22})$$

wyznaczamy

$$b_k = -y_k \nabla_k f(\boldsymbol{\alpha}) \quad (\text{A.23})$$

dla k takich, że:

$$0 < \alpha_k < 2, \quad k = 1, \dots, m. \quad (\text{A.24})$$

Do obliczenia wartości b bierzemy więc pod uwagę średnią b_k gdzie $k = \{5, 6, 9\}$, czyli

$$b = \frac{b_5 + b_6 + b_9}{3} = \frac{0,35 + 0,35 + 0,35}{3} = 0,35. \quad (\text{A.25})$$

Biorąc pod uwagę zaokrąglenia, wynik otrzymany ręcznie zgadza się z tym wygenerowanym przez bibliotekę.

Aby przeprowadzić predykcję przy pomocy otrzymanego modelu, wystarczy obliczyć wartość funkcji decyzyjnej (2.15):

$$f^*(\tilde{X}) = \text{sgn}\left(\sum_{k=1}^m y_k \alpha_k K(X_k, \tilde{X}) + b\right). \quad (\text{A.26})$$

Podobnie jak w poprzednim przykładzie, prognozowana diagnoza wystąpienia BPD u pacjenta urodzonego w 32 tygodniu ciąży, któremu nie podawano surfaktantu, będzie wyrażona wzorem:

$$\tilde{X} = \begin{bmatrix} 0,6 \\ 0 \end{bmatrix}, \quad (\text{A.27})$$

$$\begin{aligned} f^*(\tilde{X}) = & \operatorname{sgn}(1,21 \cdot \tanh(2 \cdot -0,6 \cdot 0,6 - 4) + \\ & 1,05 \cdot \tanh(2 \cdot -0,2 \cdot 0,6 - 4) + \\ & 0,71 \cdot \tanh(2 \cdot 0,4 \cdot 0,6 - 4) + \\ & 1,05 \cdot \tanh(2 \cdot 0,2 \cdot 0,6 - 4) + \\ & -0,35 \cdot \tanh(2 \cdot 1 \cdot 0,6 - 4) + \\ & -0,35 \cdot \tanh(2 \cdot -0,8 \cdot 0,6 - 4) + \\ & -0,95 \cdot \tanh(2 \cdot -0,2 \cdot 0,6 - 4) + \\ & -0,95 \cdot \tanh(2 \cdot 0 \cdot 0,6 - 4) + \\ & -0,35 \cdot \tanh(2 \cdot -1 \cdot 0,6 - 4) + \\ & -0,95 \cdot \tanh(2 \cdot 0,2 \cdot 0,6 - 4) + \\ & 0,35), \end{aligned} \quad (\text{A.28})$$

po uproszczeniach:

$$\begin{aligned} f^*(\tilde{X}) = & \operatorname{sgn}(1,21 \cdot \tanh(-4,72) + 1,05 \cdot \tanh(-4,24) + 0,71 \cdot \tanh(-3,52) + \\ & 1,05 \cdot \tanh(-3,76) - 0,35 \cdot \tanh(-2,8) - 0,35 \cdot \tanh(-4,96) - \\ & 0,95 \cdot \tanh(-4,24) - 0,95 \cdot \tanh(-4) - 0,35 \cdot \tanh(-5,2) - \\ & 0,95 \cdot \tanh(-3,76) + 0,35), \end{aligned} \quad (\text{A.29})$$

dalej:

$$f^*(\tilde{X}) = \operatorname{sgn}(-1,21 - 1,05 - 0,71 - 1,05 + 0,35 + 0,35 + 0,95 + 0,95 + 0,35 + 0,95 + 0,35), \quad (\text{A.30})$$

ostatecznie otrzymujemy:

$$f^*(\tilde{X}) = \operatorname{sgn}(0,23) = 1. \quad (\text{A.31})$$

Stworzony model przewidział więc pozytywną diagnozę BPD dla rozważanego przykładu testowego. Otrzymany wynik jest przeciwny do tego z poprzedniego przykładu. Nie należy się tym jednak sugerować. Parametry funkcji jądrowej dobrano bowiem tak, aby łatwo było wykonać prezentowane obliczenia, pokazując jednocześnie schemat działania metody – nie jest to więc model optymalny. Ponadto wykorzystano niewielką liczbę przykładów uczących.