**Skład grupy:**

Grzegorz Denert

Michał Dorosz

**Tytuł projektu:**

Prognozowanie przebiegu ostrego zapalenia trzustki na podstawie rzeczywistych danych klinicznych pacjentów

**Opis projektu:**

Ostre zapalenie trzustki (OZT) jest to stan zapalny gruczołu spowodowany przedwczesną aktywacją proenzymów w komórkach pęcherzykowych trzustki, czego efektem jest uszkodzenie miąższu narządu oraz wytworzenie się stanu zapalnego, nierzadko o ogólnoustrojowym charakterze. Rozpoznanie OZT stawia się, gdy występują dwa z trzech poniższych kryteriów: (1) typowy obraz kliniczny - ostry, silny ból zlokalizowany w nadbrzuszu, często promieniujący do pleców, (2) ponad trzykrotny wzrost aktywności enzymów trzustkowych w surowicy powyżej górnej granicy normy, (3) wyniki badań obrazowych charakterystyczne dla OZT. OZT jest najczęstszą gastroenterologiczną przyczyną hospitalizacji ze wskazań pilnych. Ze względu na znaczą zachorowalność oraz śmiertelność bardzo ważna jest szybka diagnoza i wdrożenie leczenia w przeciągu 24 godzin od wystąpienia pierwszych objawów. Zgodnie z kliniczną klasyfikacją z Atlanty wyróżniamy trzy stopnie ciężkości choroby: łagodną, umiarkowaną i ciężką. U większości pacjentów występuje forma łagodna choroby (80%), w której obserwuje się powrót do zdrowia w przeciągu kilku dni do tygodnia. Niestety u około 5-20% chorych dochodzi do ciężkiego OZT, kiedy to śmiertelność wynosi od 10% do 50%. Ocena przebiegu i stopnia zaawansowania choroby już na początkowym etapie jest zatem niezwykle istotna w zaplanowaniu sposobu leczenia pacjenta. Chorzy z przewidywaną ciężką postacią OZT wymagają intensyfikacji leczenia oraz hospitalizacji w oddziałach intensywnego nadzoru.

**Cele projektu:**

Stworzenie z pomocą programu GeNIe oraz API smile systemu, który będzie wspierał w ocenie przebiegu i stopniu zaawansowania choroby.

**Dane:**

W projekcie wykorzystane zostały dane 120 pacjentów hospitalizowanych w Klinice Gastroenterologii i Chorób Wewnętrznych Uniwersyteckiego Szpitala Klinicznego w Białymstoku w latach 2018-2022. Zgromadzone dane obejmują następujące czynniki:

1. Demograficzne – wiek, płeć
2. Etiologiczne - kamica żółciowa, nadmierne spożywanie alkoholu, hipertriglicerydemia, hiperkalcemia, rak trzustki, przyjmowane leki, czynniki jatrogenne
3. Dane kliniczne - wzrost, masa ciała, choroby współistniejące
4. Wyniki badań laboratoryjnych - CRP, WBC, HCT, PLT, BUN, kreatynina, prokalcytonina, albuminy, parametry równowagi kwasowo-zasadowej, gazometria krwi, bilirubina, AST, ALT, GTP, ALP
5. Wyniki badań obrazowych - TK, MR, USG

Ze 120 przypadków 79 obejmuje pacjentów z łagodnym przebiegiem, 19 z umiarkowanym oraz 22 z ciężkim.

**Przygotowanie danych do analizy:**

Tak jak w przypadku większości danych rzeczywistych w zbiorze występowały wartości brakujące. Do uzupełnienia ich wykorzystana została metoda interpolate z biblioteki Pandas.

Część kolumn była opisywana wartościami „TAK”, „NIE” zostały one zamienione na odpowiednio „1” oraz „0”.

Kolumna Skali Balthazara zawierała wartości zapisane literowo zamieniliśmy je na wartości numeryczne zgodnie z ich hierarchią A-1, B-2, C-3, D-4, E-5.

Kolumny wypełnione w niej niż 50 % zostały odrzucone. Ostateczny zbiór danych zawierał 120 wierszy oraz 113 kolumn.

Po załadowaniu danych do programu GeNIe wykonaliśmy dyskretyzację zmiennych ciągłych.

**Uczenie struktury sieci**

Zastosowano dwa algorytmy uczenia Greedy Thick Thinning oraz Bayesian Search.

Obraz zawierający tekst, diagram, krąg, zrzut ekranu

Opis wygenerowany automatycznieObraz zawierający tekst, diagram, zrzut ekranu, krąg

Opis wygenerowany automatycznie

Rys. Struktura sieci wyuczona z użyciem algorytmu Bayesian Search z uwzględnieniem jedynie istotnych węzłów z punktu widzenia analizy wrażliwości.

Rys. Struktura sieci wyuczona z użyciem algorytmu Greedy Thick Thinning z uwzględnieniem jedynie istotnych węzłów z punktu widzenia analizy wrażliwości.

Ostatecznie w projekcie została wykorzystana sieć wyuczona z użyciem algorytmu Bayesian Search.

**Uczenie parametrów sieci**

Uczenie parametrów sieci przeprowadziliśmy najpierw dla obu wyuczonych modeli, aby móc porównać wyniki

Dokładność dla każdej metody prezentowała się następująco:

OZT = 0.716667 (86/120)

State1 = 0.974684 (77/79)

State2 = 0 (0/19)

State3 = 0.409091 (9/22)

Obraz zawierający tekst, zrzut ekranu, linia, Czcionka

Opis wygenerowany automatycznieMacierz pomyłek:

Współczynniki AUC:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Greedy Thick Thinning Uniformize | Greedy Thick Thinning  Keep original  (confidence  =1) | Greedy Thick Thinning  Keep original  (confidence  =100) | Bayesian Search  Uniformize | Bayesian Search  Keep original  (confidence  =1) | Bayesian Search  Keep original  (confidence  =100) |
| State0 | 0.636308 | 0.636308 | 0.636308 | 0.636308 | 0.636308 | 0.636308 |
| State1 | 0.701407 | 0.713392 | 0.701407 | 0.701407 | 0.713392 | 0.713392 |
| State2 | 0.823284 | 0.824212 | 0.823284 | 0.823284 | 0.824212 | 0.824212 |

**Krzywe ROC**Obraz zawierający tekst, diagram, linia, Wykres

Opis wygenerowany automatycznieObraz zawierający tekst, linia, Wykres, diagram

Opis wygenerowany automatycznieObraz zawierający tekst, diagram, linia, Wykres

Opis wygenerowany automatycznie

**Interfejs użytkownika**

Obraz zawierający tekst, zrzut ekranu, Czcionka, numer

Opis wygenerowany automatycznieZ wykorzystaniem API smile oraz biblioteki streamlit stworzyliśmy interfejs użytkownika, który na podstawie wpisanych parametrów pacjenta oblicza prawdopodobieństwo ciężkiego przebiegu OZT oraz informuje o tym.