

Analiza zbioru danych "Patient-level dataset to study the effect of COVID-19 in people with Multiple Sclerosis" z wykorzystaniem nadzorowanego algorytmu uczenia maszynowego One-Class SVM(Support Vector Machines)

Anna Mrozek, Bartosz Panek i Przemysław Jura

ARTICLE INFO

Keywords:
One-Class SVM
COVID-19

STRESZCZENIE

Artykuł analizuje zbiór danych pacjentów ze stwardnieniem rozsianym (SM), którzy przeszli COVID-19, przy użyciu nadzorowanego algorytmu One-Class SVM. Celem jest identyfikacja przypadków o zwiększonym ryzyku ciężkiego przebiegu infekcji. Umożliwi to zidentyfikowanie charakterystycznych cech pacjentów bardziej narażonych na powikłania.

1. Wprowadzenie

Pandemia COVID-19 miała znaczący wpływ na osoby z chorobami przewlekłymi, w tym na pacjentów ze stwardnieniem rozsianym (SM). SM jest przewlekłą chorobą autoimmunologiczną, która wpływa na układ nerwowy i może prowadzić do trwałego uszkodzenia neuronów oraz ograniczenia funkcji motorycznych oraz poznawczych.[1][2] Ze względu na charakter choroby i stosowane terapie immunosupresyjne, pacjenci z SM mogą być bardziej narażeni na cięższy przebieg infekcji wirusowych, w tym COVID-19.[3][4] Analiza danych od pacjentów z SM, którzy przeszli COVID-19, może dostarczyć ważnych informacji na temat ryzyka powikłań i zidentyfikować czynniki przyczyniające się do poważniejszych objawów, co ostatecznie może pomóc w lepszym monitorowaniu i opiece nad tą grupą pacjentów.

2. Cel

Celem badania jest wykorzystanie nadzorowanego algorytmu One-Class SVM do analizy zbioru danych pacjentów ze stwardnieniem rozsianym, którzy przeszli COVID-19, w celu zidentyfikowania cech pacjentów o zwiększonym ryzyku ciężkiej infekcji. Oczekuje się, że wyniki analizy dostarczą wskazówek do opracowania bardziej ukierunkowanych strategii opieki i środków zapobiegawczych dla pacjentów ze stwardnieniem rozsianym w kontekście potencjalnych przyszłych pandemii lub innych zagrożeń wirusowych.

3. Przegląd literatury

W literaturze medycznej i naukowej, od początku pandemii COVID-19, wiele badań skupiało się na analizie wpływu wirusa SARS-CoV-2 na osoby cierpiące na choroby przewlekłe i autoimmunologiczne, takie jak stwardnienie rozsiane (SM). Wyniki tych badań sugerują, że pacjenci z SM, zwłaszcza ci stosujący terapie immunosupresyjne,

mogą być bardziej narażeni na cięższy przebieg COVID-19 oraz związane z nim powikłania[5][6]. U pacjentów z SM często obserwuje się obniżoną odporność oraz większą podatność na infekcje, co wynika zarówno z samej choroby, jak i z efektów leków tłumiących układ odpornościowy[7]. Dodatkowo w badaniach zwraca się uwagę na takie czynniki ryzyka jak wiek, płeć, rodzaj stosowanego leczenia oraz inne choroby towarzyszące, które mogą zwiększać ryzyko ciężkiego przebiegu infekcji u tej grupy pacjentów.

W ostatnich latach coraz większą popularność zyskuje zastosowanie algorytmów uczenia maszynowego, takich jak Support Vector Machines (SVM), do analizy danych medycznych. Algorytmy te pozwalają na wykrywanie wzorców i anomalii w dużych, zróżnicowanych zbiorach danych[8]. W szczególności algorytm One-Class SVM, używany do wykrywania anomalii, jest przydatny do identyfikacji przypadków wysokiego ryzyka w danych medycznych, gdzie przypadki nietypowe (np. pacjenci bardziej narażeni na ciężki przebieg COVID-19) występują rzadko. Badania pokazują, że One-Class SVM dobrze sprawdza się w analizie populacji pacjentów, gdy dostęp do dużych zbiorów danych o przypadkach zdrowych jest ograniczony, co często stanowi wyzwanie w analizie medycznej[9].

W kontekście badań nad COVID-19 i SM, algorytmy nadzorowanego uczenia maszynowego, takie jak SVM, umożliwiają dokładniejszą analizę czynników ryzyka oraz lepsze prognozowanie ciężkiego przebiegu choroby. Wyniki takich analiz mogą być przydatne nie tylko dla lekarzy, ale także dla decydentów zajmujących się zdrowiem publicznym, ponieważ pozwalają na opracowanie bardziej ukierunkowanych strategii opieki dla osób z SM, które są bardziej narażone na zagrożenia związane z wirusami, takimi jak COVID-19.

4. Metodologia

W badaniu zastosowano algorytm nadzorowanego uczenia maszynowego One-Class SVM w celu identyfikacji pacjentów z podwyższonym ryzykiem ciężkiego przebiegu COVID-19 w grupie osób ze stwardnieniem rozsianym.

Analiza obejmowa \acute ła cechy kliniczne pacjent \acute ow, takie jak wiek, p \acute leć, typ leczenia oraz choroby w \acute s \acute sp \acute l \acute stniej \acute ące, aby zidentyfikować wzorce zwi \acute azane z wi \acute eksz \acute ą podatnośc \acute ią na powik \acute łania.

4.1. Dataset

Stwardnienie rozsiane (MS) to przewlekła choroba autoimmunologiczna, która wywo \acute łuje stan zapalny w obr \acute bie o \acute srodkowego uk \acute ładu nerwowego. Choroba prowadzi do r \acute żnych stopni utraty funkcji przez uszkodzenia mieliny oraz wł \acute okien nerwowych [10]. Osoby z MS s \acute ą bardziej podatne na infekcje z powodu z \acute łożonego dzia \acute łania samej choroby, jej leczenia i naturalnego przebiegu [11]. Z inicjatywy COVID-19 and MS Global Data Sharing Initiative (GDSI) zbadano, jak leki immunosupresyjne lub immunomoduluj \acute ące wp \acute ływaj \acute na COVID-19 i jego przebieg u os \acute ob z MS. GDSI mia \acute ła na celu zwi \acute kszenie skali zbierania danych dotycz \acute ących COVID-19 i dostarczenie spo \acute łeczno \acute ci zwi \acute azanej z MS informacji opartych na danych podczas pandemii [12]. W ramach GDSI wybrano kluczowe zmienne obejmuj \acute ące informacje o COVID-19, stopniu jego ci \acute ężko \acute ci, leczeniu, dane demograficzne, historię i nasilenie MS, stosowanie lek \acute ow modyfikuj \acute ących przebieg choroby (DMT), choroby w \acute s \acute sp \acute l \acute stniej \acute ące i wybrane zachowania zwi \acute azane ze stylem życia, takie jak palenie tytoniu. Globalna spo \acute łeczno \acute ć MS w \acute s \acute pr \acute acowa \acute ła, przekazuj \acute ąc dokumentację statusu COVID-19 u os \acute ob z MS za po \acute średnictwem centralnej platformy udost \acute ępnionej przez QMENTA [13].

Ten zbi \acute or danych zosta \acute ł zebrany za pomoc \acute ą narz \acute ędzia do szybkiego wprowadzania danych, które umo \acute żliwia \acute ło klinicystom, osobom ze stwardnieniem rozsianym (PwMS) lub ich przedstawicielom bezpo \acute średnie wprowadzanie informacji do centralnej platformy GDSI. Narz \acute ędzie to zawiera \acute ło kwestionariusz oparty na wcze \acute śniej ustalonych zmiennych i nie gromadzi \acute ło bezpo \acute średnich danych osobowych, aby chronić prywatno \acute ć u \acute żytkownik \acute ow. Narz \acute ędzie zosta \acute ło wy \acute łączone 3 lutego 2022 roku.

Zbi \acute or danych obejmuj \acute e informacje o 1141 osobach ze stwardnieniem rozsianym (PwMS). Aby zapewnić zgodno \acute ć danych z wytycznymi HIPAA, przeprowadzono proces deidentyfikacji. Po zebraniu danych dokonano oceny ryzyka zwi \acute azanego z małymi kom \acute orkami (SCRA), klasyfikuj \acute ąc zmienne na trzy kategorie: bezpo \acute średnie identyfikatory, zmienne wrażliwe i identyfikatory po \acute średnie. Bezpo \acute średnie identyfikatory to zmienne, które mog \acute ą jednoznacznie zidentyfikować osob \acute , zmienne wrażliwe to te, które respondent mo \acute że chciać zachować w tajemnicy, natomiast identyfikatory po \acute średnie mog \acute ą zidentyfikować osob \acute , je \acute śli s \acute ą po \acute łączone z danymi z innych zbior \acute ow.

Poniewa \acute ż w danych nie zbierano imion pacjent \acute ow, deidentyfikacja skupi \acute ła się na datach i wieku pacjent \acute ow. Daty w kolumnie „stop-or-end-date-combined” zosta \acute ły przesuni \acute ęte o losow \acute ą liczb \acute ę dni (mi \acute ędzy -15 a 15), aby uniemo \acute żliwić identyfikację na podstawie dat. Wiek pacjent \acute ow sklasyfikowano w cztery grupy: 0 dla os \acute ob w wieku 0–17 lat, 1 dla os \acute ob mi \acute ędzy 18 a 50 lat, 2 dla os \acute ob mi \acute ędzy 51 a 70 lat, oraz 3 dla os \acute ob powy \acute żej 71 lat. Dzi \acute ęki temu żadne dok \acute ładne

warto \acute ści wieku powy \acute żej 90 lat nie zosta \acute ły ujawnione. Po klasyfikacji zmiennych i wdro \acute żeniu odpowiednich \acute rodk \acute ow ostro \acute żno \acute ci dane zosta \acute ły zdeidentyfikowane i spe \acute łniaj \acute ą standardy HIPAA, zachowuj \acute ąc jednocze \acute śnie warto \acute ć badawcz \acute ą. Ponadto, aby zapewnić ochron \acute ę prywatno \acute ci, zastosowano techniki takie jak K-anonimizacja oraz r \acute óżnorodno \acute ć.

Zbi \acute or danych obejmuj \acute e zestaw z g \acute ory okre \acute ślonych zmiennych (n=47), takich jak p \acute leć, kategoria wiekowa, typ MS, wynik EDSS, status palenia oraz kategoria BMI. Te zmienne dostarczaj \acute ą informacji o demografii pacjent \acute ow, ich stanie klinicznym oraz symptomach zwi \acute azanych z COVID-19. Szczeg \acute ółowy opis typ \acute ow zmiennych i ich statystyki znajduje się w sekcji „Opis Danych”.

4.2. Opis metody

4.3. Opis przeprowadzonych oblicze \acute n

4.4. Wykorzystane metryki oceny

5. Wyniki

6. Opis wynik \acute ow

7. Podsumowanie

8. Bibliografia

- [1] <https://diag.pl/pacjent/artykuly/jakie-sa-pierwsze-objawy-i-sposoby-leczenia-stwardnienia-rozsianego/>
- [2] <https://www.medicover.pl/o-zdrowiu/stwardnienie-rozsiane-objawy-przyczyny-i-leczenie,6619,n,192>
- [3] https://pl.wikipedia.org/wiki/Stwardnienie_rozsiane
- [4] <https://ptsr.org.pl/strona/133,covid-19-a-sm>
- [5] Sormani MP, De Rossi N, Schiavetti I, Carmisciano L, Cordioli C, Radaelli M, et al. Disease modifying therapies and COVID-19 severity in multiple sclerosis. *Ann Neurol*. 2021 Apr
- [6] Louapre C, Collongues N, Stankoff B, Giannesini C, Papeix C, Bensa C, et al. Clinical characteristics and outcomes in patients with coronavirus disease 2019 and multiple sclerosis. *JAMA Neurol*. 2020 Sep
- [7] Simpson-Yap S, De Brouwer E, Kalincik T, Rijke N, Hillert JA, Walton C, et al. Associations of Disease-Modifying Therapies With COVID-19 Severity in Multiple Sclerosis. *Neurology*. 2021 Nov
- [8] Schiff MA, Rae-Grant A, Gilden D, Franklin GM. Practice guideline: Disease-modifying therapies for adults with multiple sclerosis: Report of the guideline development, dissemination, and implementation subcommittee of the American Academy of Neurology. *Neurology*. 2019 Jan
- [9] Erfani P, Mitchell AJ, Hameed S, Heydarpour P, Ghaffaripour R, Sahraian MA. Systematic review of health-related quality of life in multiple sclerosis patients: The impact of pharmacological treatments and lifestyle. *J Neurol Sci*. 2016 Dec
- [10] Calabresi PA. Diagnosis and management of multiple sclerosis. *Am Fam Physician*. 2004 Nov
- [11] Montgomery S, Hillert J, Bahmanyar S. Hospital admission due to infections in multiple sclerosis patients. *Eur J Neurol*. 2013 Aug

[12] Peeters LM, Parciak T, Walton C, Geys L, Moreau Y, De Brouwer E, et al. COVID-19 in people with multiple sclerosis: A global data sharing initiative. *Mult Scler Houndmills Basingstoke Engl.* 2020 Sep

[13] Simpson-Yap S, De Brouwer E, Kalincik T, Rijke N, Hillert JA, Walton C, et al. Associations of Disease-Modifying Therapies With COVID-19 Severity in Multiple Sclerosis. *Neurology.* 2021 Nov