

dr hab. inż. Izabela Sówka, prof. uczelni

**Politechnika Wroclawska
Wydział Inżynierii Środowiska
Katedra Inżynierii Ochrony Środowiska**



Plac Grunwaldzki 13
50-377 WROCLAW
tel. (71) 320 25 60
e-mail: izabela.sowka@pwr.edu.pl

R E C E N Z J A

rozprawy doktorskiej mgr inż. Mariusza Filaka

pt.

Predykcja stężeń zanieczyszczeń powietrza w perceptronowych modelach regresyjnych

opracowana w oparciu o uchwałę nr 71/2021/2022 Rady Dyscypliny Naukowej Inżynieria środowiska górnictwo i energetyka na Wydziale Infrastruktury i Środowiska Politechniki Częstochowskiej w roku akademickim 2021/2022 z dnia 4. lipca 2022 r. (pismo nr R.WIIS.BOD.41.2022.3 kierownik dyscypliny naukowej Inżynieria środowiska, górnictwo i energetyka dr hab. inż. Iwony Zawiei, prof. PCz.).

Praca doktorska mgr inż. Mariusza Filaka dotyczy badań nad zastosowaniem neuronowych modeli regresyjnych w prognozowaniu stężeń zanieczyszczeń powietrza atmosferycznego ze szczególnym z uwzględnieniem podniesienia poziomu dokładności predykcji poprzez uwzględnienie wielu modeli predykcyjnych dla różnych podokresów jakości powietrza.

Realizacja pracy w określonym zakresie jest bez wątpienia bardzo istotna w świetle szeroko dyskutowanej tematyce związanej ze złą jakością powietrza na terenach miejskich oraz wpływie zanieczyszczeń powietrza na zdrowie człowieka. Opracowanie oraz przetestowanie narzędzi, które mogłyby wspierać ocenę poziomu zanieczyszczenia na danym obszarze oraz umożliwić podjęcie właściwych działań zapobiegawczych minimalizujących stopień narażenia populacji ludzkiej na szkodliwe substancje można uznać za priorytetowe i niezbędne w kontekście ochrony zdrowia publicznego.

Rozprawa doktorska Pana mgr inż. Mariusza Filaka została przygotowana w formie i układzie obejmującym: streszczenie (w języku polskim i angielskim); część wstępną, a w niej syntetyczne omówienie wpływu zanieczyszczeń powietrza na zdrowie człowieka oraz kwestii związanej z monitoringiem zanieczyszczeń powietrza w Polsce

i problematyką kompletności danych; część teoretyczną obejmującą charakterystykę wybranych elementów systemu prawnego w kontekście przepisów regulujących funkcjonalnie monitoringu powietrza w ramach Państwowego Monitoringu Środowiska, opis wpływu zanieczyszczeń powietrza na zdrowie człowieka i charakterystykę sztucznych sieci neuronowych, a także syntetyczny opis w kontekście metod modelowania jakości powietrza; cel i tezy badawcze; część badawczą wraz z charakterystyką punktów pomiarowych/stacji monitoringu jakości powietrza oraz danych wykorzystanych w przeprowadzanych przez doktoranta analizach, opis wyboru typu sieci neuronowej, architektury sieci MLP oraz metodyki badań (przygotowanie danych, metodyka modelowania, podział zbiorów pełnozakresowych na podzbiory, ocena dokładności modeli, epizody); wyniki badań i ich dyskusję w kontekście modeli RVS oraz PVS dla wybranych zanieczyszczeń powietrza (O_3 , NO, NO_2 , SO_2 , PM_{10} , CO); podsumowanie; wnioski; bibliografię; spis tabel oraz rysunków.

W ramach prowadzonych prac Pan mgr inż. Mariusz Filak sformułował następujące tezy pracy doktorskiej (cyt.):⁴

- archiwalne dane pomiarowe mogą być wykorzystane do modelowania aktualnych stężeń zanieczyszczeń;

- do modelowania predykcyjnego można wykorzystać sztuczne sieci neuronowe, które stanowią wygodne i praktyczne narzędzie do modelowania, w porównaniu do tradycyjnych metod;

- istnieje możliwość zmniejszenia błędów modelowania regresyjnego, poprzez zastąpienie jednej sieci neuronowej przez kilka sieci, obejmujących różne sektory zakresu stężeń modelowanego zanieczyszczenia;

oraz jej główny cel polegający na (cyt.): ‘poprawieniu dokładności predykcji stężeń zanieczyszczeń powietrza w neuronowych modelach regresyjnych poprzez zastosowanie wielu modeli predykcyjnych, utworzonych dla różnych podzakresów stężeń zanieczyszczeń powietrza.’

W przygotowanej w sposób syntetyczny części wprowadzającej pracy (rozdział 1) nazwanej przez mgr inż. Mariusza Filaka wstępem doktorant zebrał w syntetyczny sposób informacje dotyczące wpływu zanieczyszczeń powietrza na zdrowie człowieka (co zostało rozwinięte następnie w rozdziale 2.2) oraz monitoringu zanieczyszczeń powietrza w Polsce i problematyką kompletności danych

W rozdziale 2, tak jak wspomniano wcześniej, autor syntetycznie omówił elementy systemu prawnego w kontekście przepisów regulujących funkcjonalnie monitoringu powietrza w ramach Państwowego Monitoringu Środowiska (rozdział 2.1), opisał wpływu

zanieczyszczeń powietrza na zdrowie człowieka (rozdział 2.2), rozwinął charakterystykę i historię rozwoju i zastosowań sztucznych sieci neuronowych (rozdział 2.3), a także w krótki sposób opisał metody modelowania jakości powietrza wraz z zaznaczeniem znaczenia modyfikacji i opracowania autorskiej metodyki zmierzającej do poprawy jakości modelowania. Rozdział nr 3 stanowi cel i tezy pracy.

W rozdziale 4 (część badawcza), obok charakterystyki stacji pomiarowych, danych wykorzystanych w analizach, opisu zmienności stężeń wybranych zanieczyszczeń powietrza (rozdział 4.1) bardzo cennymi i wskazującymi na dojrzałość doktoranta są zwłaszcza rozdziały i dyskusja nad wyborem typu sieci neuronowej (rozdział 4.2), architektury sieci MLP (rozdział 4.3). Dodatkowo doktorant w sposób wyczerpujący i opisał metodykę badań (rozdział 4.4) obejmującą przygotowanie danych, metodykę modelowania, podział zbiorów pełnozakresowych na podzbiory, wraz z oceną dokładności modeli. Wszystkie następujące po sobie rozdziały są ułożone w sposób logiczny i stanowią spójną całość.

Szczególnie interesującym i wskazującym na dojrzałość naukową jest rozdział 5 oraz 6, w których odnaleźć można wyniki badań, ich dyskusję oraz podsumowanie. Autor bardzo dokładnie i w skrupulatny sposób przedstawił wyniki modelowania przy zastosowaniu modeli RVS (rozdział 5.1) oraz PVS (rozdział 5.3) jasno zaznaczając wnioski z przeprowadzonych analiz. Opisy i grafiki w kontekście wyników modelowania, informacje nt. oszacowanych wartości błędów tego typu pracy mają cenny, choć podstawowy charakter i chociaż autor dużą część w tym fragmentach pracy poświęcił na opisy wyników to wg mojej opinii brakuje w nich uwzględnienia głębszej dyskusji nad zjawiskami, procesami, które mogły wpłynąć na różnice, które uzyskano pomiędzy wartościami rzeczywistymi (stężenia mierzone) oraz uzyskanymi w wyniku zastosowania modeli, zarówno RVS, jak i PVS. Częściowo brak tak pogłębionej dyskusji rekompensuje rozdział 6, w którym autor podsumowuje wyniki swoich badań oraz analiz i przedstawia wyniki wskazujące na zwiększenie dokładności predykcji, zwłaszcza w przypadku modeli RVS (tabela 30). Jednocześnie na skutek uzyskanych wyników autor potwierdza zasadność zastosowania kilku sieci neuronowych do węższych zakresów stężeń jednocześnie wskazując iż modele podzakresowe wykazują zmienną poprawę dokładności predykcji, w porównaniu do dokładności modeli pełnozakresowych. Uzyskana przez autora dokładność modelowania okazała się różna w zależności od liczby submodeli (modeli podzakresowych) i poziomu stężeń charakteryzujących dany podzakres. Ostatecznie zrealizowane przez autora dysertacji analizy pozwoliły na potwierdzenie założonych tez badawczych oraz sformułować wnioski końcowe (rozdział 7), z których jako najważniejsze wymienić należy m.in. (cyt.):

- modele podzakresowe wykazują zróżnicowaną poprawę dokładności predykcji, w porównaniu do dokładności modeli pełnozakresowych ;

- dokładność modelowania zależy od liczby modeli podzakresowych i poziomu stężeń charakteryzujących dany podzakres. Na ogół największe wartości błędów występują przy modelach dedykowanych do podzakresów z najwyższymi poziomami stężeń, a najniższe – przy modelach dedykowanych dla najniższych podzakresów reprezentujących najniższe poziomy stężeń;
- dla modeli podzakresowych utworzonych po wstępnym sortowaniu przypadków według rosnących wartości stężeń rzeczywistych modelowanego zanieczyszczenia (modele RVS) uzyskano znacznie lepsze wyniki predykcji w porównaniu do modeli podzakresowych utworzonych po wstępnym sortowaniu przypadków według rosnących wartości stężeń predykcyjnych modelowanego zanieczyszczenia (modele PVS);
- liczba podzakresów stężeń dla modeli PVS nie może być zbyt duża. Podział pełnozakresowego zbioru przypadków na 8 podzakresów nie zawsze skutkuje poprawą dokładności aproksymacji w tak utworzonych modelach podzakresowych. '

W tak zwięźle w opisany powyżej przez mnie sposób mgr inż. Mariusz Filak zaprezentował wyniki rozważań i analiz – zgodnie z założonym celem głównym dysertacji doktorskiej – zmierzających do wskazania przydatności zastosowania neuronowych modeli regresyjnych uwzględniających wiele modeli predykcyjnych utworzonych dla różnych podzakresów stężeń zanieczyszczeń powietrza do poprawy dokładności predykcji stężeń zanieczyszczeń powietrza, co mogłoby wg mojej opinii, po przeprowadzeniu analiz jeszcze na szerszą skalę, stanowić przyczynek do wzbogacenia informacji w warunkach polskich w kontekście ocen stanu jakości powietrza atmosferycznego.

Rozprawa doktorska mgr inż. Mariusza Filaka jest przygotowana w przemyślany i syntetyczny sposób – poniżej zwrócę tylko uwagę na kilka zagadnień natury polemicznej:

1. We wstępie rozprawy (str. 2, akapit 2) autor pracy pisze iż (cyt): '*(...) Stacje monitoringu powietrza dokonują pomiarów w sposób ciągły, a wyniki rejestrowane są w postaci 1-godzinnych wartości średnich. (...)*' Wg mojej opinii to błąd stylistyczny. To nie stacje dokonują pomiarów tylko pomiary są wykonywane przy zastosowaniu konkretnych analizatorów/ urządzeń pomiarowych.
2. Pomimo, że autor użył spisu skrótów i oznaczeń zabrakło w nim skrótu PMŚ (Państwowy Monitoring Środowiska), którego używa na stronie nr 3 dysertacji doktorskiej.
3. W części teoretycznej pracy (rozdział 2.1., str. 4) wskazanym byłoby aby w akapicie nr 2 łącznie z informacjami dotyczącymi wartości minimalnych % danych

dotyczących pomiarów ciągłych zamieścić również informacje nt. poziomów dopuszczalnych zanieczyszczeń np. w postaci tabeli.

4. W rozdziale nr 2.2. (akapit 1, str. 5) autor używa sformułowania (cyt): *' (...) Do efektywnego rozwiązania problemu zanieczyszczenia powietrza konieczne są stanowcze działania podjęte we wszystkich obszarach odpowiedzialnych za emisję zanieczyszczeń. Do obszarów tych można zaliczyć przemysł, energetykę, transport, gospodarstwa domowe, ale również ogół społeczeństwa, którego postawa i świadomość ma fundamentalne znaczenie dla powodzenia działań mających na celu ograniczenie emisji zanieczyszczeń powietrza (...)*' Wg mojej opinii zdania słowo 'obszar' nie jest użyte we właściwym znaczeniu zwłaszcza w kontekście 'ogółu społeczeństwa'.
5. W rozdziale 2.2. (akapit 2, str. 5) autor wspomina o 'rozmaitych defektach zdrowotnych'. Uważam to stwierdzenie za niefortunne i proszę o przytoczenie przykładów wyników badań w tym zakresie.
6. W rozdziale 2.2. (akapit 1 str. 7) autor używa sformułowania (cyt): *' Pozostałe zanieczyszczenia gazowe, których źródło nie zawsze jest pochodzenia antropogenicznego (...)* Proszę zatem o przedstawienie i klasyfikację źródeł zanieczyszczeń gazowych.
7. W rozdziale 2.3. (akapit 1 str. 8) autor używa sformułowania (cyt): *' Do przeprowadzenia analizy wykorzystano dane pochodzące z lat 2011-2016, zarejestrowane na dwóch stacjach monitoringu powietrza. Obie stacje znajdowały się na terenie województwa śląskiego. Wybór konkretnych stacji monitoringu powietrza podyktowany był stosunkowo wysoką kompletnością danych zarejestrowanych na tych stacjach. (...)* Proszę o doprecyzowanie informacji dot. wyboru stacji i kompletności danych oraz porównanie % kompletności danych z innymi przykładowymi stacjami pomiarowymi nieuwzględnionymi w analizie.
8. W rozdziałach dotyczących modeli RVS (rozdział 5.1) oraz PVS (rozdział 5.3) autor dużą część poświęcił na opisy wyników. Brakuje w nich jednak uwzględnienia głębszej dyskusji nad zjawiskami, procesami, które mogły wpłynąć na różnice, które uzyskano pomiędzy wartościami rzeczywistymi (stężenia mierzone) oraz uzyskanymi w wyniku zastosowania modeli, zarówno RVS, jak i PVS. W wielu miejscach dotyczących opisów wyników analiz odznacza się bardziej sprawozdawczy charakter opisów, w miejsce których powinna wg mojej opinii zostać poprowadzona pogłębiona dyskusja. Częściowo te braki rekompensują jednak zapisy rozdziału nr 6.

W pracy doktorskiej, można odnaleźć kilka błędów natury redakcyjnej np. autor nie używa indeksu dolnego przy symbolach oznaczających frakcję pyłu PM10 oraz PM2.5, co powszechnie stosuje się w publikacjach naukowych. Także w kilku miejscach pracy, zwłaszcza we wstępie i części teoretycznej brakuje wg mojej opinii odwołań do źródeł literaturowych (np. str. 1, 2, 4, 8).

Oceniając pracę 'całościowo' pragnę jednak podkreślić iż wskazane przeze mnie pewnego rodzaju niedociągnięcia, czy niejasności, a także uchybienia redakcyjne nie pomniejszają w żaden sposób osiągnięć naukowych w niej przedstawionych.

W n i o s k i

Mgr inż. Mariusz Filak w pracy doktorskiej, przygotowanej pod opieką promotora: dr hab. Szymona Hoffmana oryginalnie rozwiązał problem naukowy. Doktorant syntetycznie podsumował stan wiedzy w tematyce swojej pracy doktorskiej, co potwierdziło jego określony poziom ogólnej wiedzy teoretycznej w poruszanej tematyce badań i analiz. Doktorant określił jasno cel i zakres dysertacji doktorskiej. Zrealizowane przez doktoranta prace oraz analizy pozwoliły wykazać zasadność stosowania neuronowych modeli regresyjnych uwzględniających wiele modeli predykcyjnych utworzonych dla różnych podokresów stężeń zanieczyszczeń powietrza do poprawy dokładności predykcji stężeń zanieczyszczeń powietrza. Poprzez zastosowanie wyżej wymienionych metod modelowania i przeprowadzone badania doktorant umożliwił m.in. praktyczne posługiwanie się nimi w określonych warunkach, co przy analizach na szerszą skalę w warunkach polskich ma potencjał aplikacyjny umożliwiający efektywniejsze zarządzanie jakością powietrza atmosferycznego.

Reasumując, stwierdzam iż praca doktorska mgr inż. Mariusza Filaka spełnia wymagania stawiane dysertacjom doktorskim przez ustawę z roku 2003 o stopniach naukowych i tytule naukowym oraz stopniach i tytule w zakresie sztuki (Dz. U. Nr 65, poz. 595 z późn. zm.). W związku z powyższym wnoszę do Wysokiej Rady Dyscypliny Naukowej Inżynieria środowiska górnictwo i energetyka na Wydziale Infrastruktury i Środowiska Politechniki Częstochowskiej o dopuszczenie mgr inż. Mariusza Filaka do publicznej obrony przedstawionej pracy doktorskiej.



Wrocław, 12. września 2022 r.