



dr hab. inż. Katarzyna Bizon, prof. PK  
Katedra Inżynierii Chemicznej i Procesowej  
[katarzyna.bizon@pk.edu.pl](mailto:katarzyna.bizon@pk.edu.pl), tel.: 12 628 2735

Kraków, 23 sierpnia 2024 r.

**Recenzja rozprawy doktorskiej**  
**Pani mgr inż. Pauliny Muchel**  
**pt. „Modelowanie zmienności składu gazu ziemnego metodą sztucznych sieci neuronowych”**  
**wykonanej pod kierunkiem Pani dr hab. inż. Jolanty Szoplik, prof. ZUT**

**Podstawa opracowania recenzji**

Podstawę formalną sporządzenia recenzji rozprawy doktorskiej autorstwa Pani mgr inż. Pauliny Muchel stanowi pismo ZUT/RDICH/32/2024 Pani Przewodniczącej Rady Dyscypliny Inżynieria Chemiczna Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie prof. dr hab. inż. Zofii Lendzion-Bieluń z dnia 25 czerwca 2024 r.

Przedstawiona do recenzji rozprawa doktorska została zrealizowana na Wydziale Technologii i Inżynierii Chemicznej Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie pod kierunkiem Pani dr hab. inż. Jolanty Szoplik, prof. ZUT. Nadanie stopnia doktora przewiduje się w dziedzinie nauk inżynieryjno-technicznych w dyscyplinie *inżynieria chemiczna*.

**Cel i zakres oraz ogólna ocena merytoryczna rozprawy**

Uczenie maszynowe i sztuczne sieci neuronowe znajdują coraz szersze zastosowanie do rozwiązywania problemów inżynierskich, w tym zagadnień z zakresu inżynierii chemicznej i procesowej. Rosnąca popularność takiego podejścia modelowego do opisu zazwyczaj bardzo złożonych problemów fizycznych i chemicznych wynika z kilku aspektów. Pierwszym z nich jest fakt, że nieodłączną cechą sieci neuronowych jest bazowanie na danych – zazwyczaj eksperymentalnych – charakteryzujących zachowanie badanego obiektu/systemu. To zaś eliminuje konieczność formułowania, a następnie numerycznego rozwiązywania klasycznych równań modelowych, które zazwyczaj przyjmują postać równań różniczkowych. Do pozostałych czynników należy zaliczyć przede wszystkim postępujący w ostatnich dziesięcioleciach gwałtowny wzrost mocy obliczeniowych komputerów oraz możliwości magazynowania i przetwarzania dużych zbiorów danych.



Prognozowanie składu gazu ziemnego przepływającego w złożonej sieci rurociągów przy użyciu modeli symulacyjnych uwzględniających procesy zachodzące wewnątrz gazociągu, warunki pogodowe czy też zmienne w czasie zapotrzebowanie odbiorców na gaz jest doskonałym przykładem zagadnienia praktycznie niemożliwego do rozwiązania w czasie rzeczywistym w przypadku zastosowania klasycznego podejścia do modelowania procesów. Jednocześnie, ze względu na ogromne znaczenie gospodarcze gazu ziemnego, postępującą dywersyfikację dostawców wynikającą z uwarunkowań geopolitycznych i potencjalny wzrost liczby punktów wejścia do systemu gazowego (m.in. poprzez lokalne wprowadzanie biometanu do sieci gazowniczej) konieczne jest opracowywanie efektywnych modeli predykcyjnych umożliwiających bieżącą prognozę składu gazu i jego właściwości (np. ciepła spalania).

W związku z powyższym, Pani mgr inż. Paulina Muchel jako wiodący cel pracy postawiła opracowanie i weryfikację modelu bazującego na sztucznej sieci neuronowej do prognozowania udziałów wybranych składników gazu ziemnego w punktach pomiarowych rozlokowanych na sieci gazowniczej. W celu realizacji głównego celu pracy oraz weryfikacji postawionych hipotez badawczych, Doktorantka w pierwszym kroku przeprowadziła badania wstępne obejmujące: (i) analizę zmienności składników gazu ziemnego w różnych punktach pomiarowych systemu gazowniczego w okresie jednego roku, (ii) analizę statystyczną zmienności składników gazu na przestrzeni czterech lat oraz (iii) analizę wpływu temperatury otoczenia na skład mieszaniny gazowej transportowanej siecią przesyłową. Następnie w ramach badań zasadniczych, ukierunkowanych na opracowanie jak najlepszego modelu typu MLP (ang. *multilayer perceptron* – perceptron wielowarstwowy), przeanalizowała wpływ szeregu czynników na dokładność prognozy składu mieszaniny gazowej. Badania te obejmowały cztery etapy, w których Doktorantka skupiła się kolejno na ocenie: (i) wpływu liczności zbioru danych użytych w procesie uczenia na dokładność prognozy składu gazu ziemnego; (ii) wpływu liczby i typu funkcji aktywacji oraz metody uczenia na dokładność prognozy; (iii) wpływu architektury sieci na dokładność prognozy oraz (iv) możliwości zastosowania sztucznej sieci neuronowej do prognozowania składu gazu ziemnego z domieszką wodoru. Uzupełnienie badań mających na celu opracowanie efektywnego modelu MLP do prognozowania składu gazu ziemnego stanowi wykorzystanie przez Doktorantkę uzyskanego wyniku prognozy do wyznaczenia ciepła spalania transportowanej w rurociągu mieszaniny gazowej.

Przedstawiona do recenzji rozprawa doktorska liczy 188 stron i ma standardowy układ. Zasadnicza część opracowania składa się z: syntetycznego wprowadzenia (rozdz. 1) nakreślającego problematykę będącą przedmiotem rozprawy; studium literaturowego (rozdz. 2); definicji problemu badawczego, hipotez badawczych oraz celu i zakresu pracy (rozdz. 3 i 4); części badawczej (rozdz. 5), którą zamyka podsumowanie i wnioski końcowe oraz zestawienia osiągniętych celów pracy (rozdz. 6 i 7). Uzupełnienie rozprawy stanowi zestawienie dorobku naukowego Doktorantki; wykaz stosowanych oznaczeń; polsko- i angielskojęzyczne streszczenie; bibliografia, na którą składa się aż 200 pozycji literaturowych oraz wykaz rysunków i tabel.



W obszernym rozdziale poświęconym przeglądowi piśmiennictwa i omówieniu aktualnego stanu wiedzy, Doktorantka skoncentrowała się na aspektach związanych z transportem i charakterystyką gazu ziemnego oraz modelach matematycznych bazujących na sztucznych sieciach neuronowych, w tym na metodach ich uczenia. W tej części rozprawy przedstawione zostały również przykłady zastosowania sztucznych sieci neuronowych zarówno do rozwiązywania problemów z obszaru inżynierii chemicznej i procesowej, jak i wielu innych dziedzin. Dominującą część podrozdziału dotyczącego zastosowania modeli sztucznych sieci neuronowych stanowi przegląd prac naukowych dotyczących analizy systemów transportu gazu ziemnego, który stanowił podstawę do sformułowania problemu badawczego, hipotez badawczych i celu pracy. Elementy te zostały bardzo precyzyjnie zdefiniowane w rozdziałach 3 i 4. Główna, bardzo ogólna teza pracy sformułowana przez Doktorantkę zakłada możliwość wytrenowania sztucznej sieci neuronowej do prognozowania składu gazu ziemnego w zależności od wybranych czynników kalendarzowych oraz pogodowych. Tezie tej towarzyszy aż osiem założeń/hipotez szczegółowych.

Zasadnicza badawcza część rozprawy składa się z omówienia wyników badań wstępnych oraz wyników badań zasadniczych. Wszystkie dane pomiarowe użyte zarówno na etapie badań wstępnych jak i zasadniczych zostały pozyskane (na podstawie umowy o współpracy i wymianie informacji) od polskiego Operatora Gazociągów Przesyłowych GAZ-SYSTEM S.A. Badania przeprowadzone przez Doktorantkę miały natomiast wyłącznie charakter obliczeniowy. W ramach badań wstępnych przeprowadzona została m.in. analiza zmienności w czasie składu gazu ziemnego w sześciu punktach sieci gazowej zlokalizowanych w rejonie północno-zachodniej Polski oraz analiza statystyczna danych pochodzących ze stacji gazowej zlokalizowanej w powiecie czarnkowsko-trzcieńskim. Następnie w ramach badań zasadniczych opracowane zostały przy użyciu programu STATISTICA różne modele typu wejście-wyjście bazujące na płytkiej sieci neuronowej typu MLP i przeanalizowany został wpływ szeregu czynników (m.in. licznosc zbioru danych użytych na etapie uczenia, typ funkcji aktywacji neuronu, architektura sieci) na właściwości predykcyjne poszczególnych modeli. Ponadto opracowana sieć MLP została wykorzystana do predykcji wartości ciepła spalania transportowanej w rurociągu mieszaniny gazowej.

Zakres przeprowadzonych przez Doktorantkę prac o charakterze obliczeniowym jest niezwykle obszerny, a dobór zastosowanych metod badawczych jest w mojej opinii dobrze przemyślany. Wartościowym elementem jest wykorzystanie – poza różnymi algorytmami do uczenia sieci neuronowych – zarówno do analizy surowych danych pomiarowych jak i wyników prognoz szerokiej gamy metod statystycznych. Pewne wątpliwości budzi jednak możliwość praktycznego zastosowania opracowanej sieci i jej zdolność do generalizacji, przede wszystkim ze względu na bardzo ograniczony (tj. użyto wyłącznie czynniki pogodowe i kalendarzowe) zestaw zmiennych w warstwie wejściowej. Pomimo tej ogólnej uwagi o charakterze dyskusyjnym, stwierdzam, że sformułowany cel rozprawy został w pełni osiągnięty przez Doktorantkę.



Głównym osiągnięciem Pani mgr inż. Pauliny Muchel jak również naukowym elementem nowości rozprawy jest opracowanie i zweryfikowanie modelu sztucznej sieci neuronowej służącego do prognozowania składu mieszaniny gazu ziemnego w danym punkcie sieci gazowej, która jest zasilana przez wielu dostawców. Zaproponowane narzędzie symulacyjne w postaci sieci MLP może mieć znaczenie użytkowe, gdyż potencjalnie może wspomóc/zastąpić badania chromatograficzne prowadzone w punktach pomiarowych usytuowanych na rurociągach gazowych. Biorąc pod uwagę rosnącą popularność sieci neuronowych oraz istotne znaczenie gospodarcze gazu ziemnego można zatem stwierdzić, że rozprawa stanowi rozwiązanie bardzo aktualnego problemu wpisującego się w ramy współczesnej inżynierii chemicznej. W mojej opinii przedstawiony w rozprawie materiał świadczy o wysokich kompetencjach badawczych Doktorantki. Ich dodatkowym i wymiernym potwierdzeniem jest dorobek naukowy Pani mgr inż. Pauliny Muchel, na który składa się 6 artykułów naukowych (w tym 3 opublikowane w czasopiśmie znajdujących się w bazie JCR) oraz 6 wystąpień konferencyjnych. Warty podkreślenia jest również udział Doktorantki w rocznym programie stażowym „Energia dla przyszłości” zorganizowanym przez Ministerstwo Energii wspólnie z Polską Grupą Energetyczną S.A., Polskim Koncernem Naftowym ORLEN S.A. oraz Polskim Górnictwem Naftowym i Gazownictwem S.A.

#### **Uwagi merytoryczne i dyskusyjne**

Mimo ogólnej pozytywnej oceny rozprawy, przyjęte na etapie konstruowania modelu sztucznej sieci założenia jak i pewne niedociągnięcia wymagają wyjaśnienia i szerszego przedstawienia. W związku z powyższym proszę o odniesienie się do następujących uwag:

1. Wartościowym uzupełnieniem podrozdziału (5.2.1) byłyby: (i) mapa obrazująca położenie geograficzne analizowanych sześciu punktów pomiaru składu gazu, punktów transgranicznych (w analizowanym regionie) i terminalu LNG, jak również (ii) zbiorcze graficzne porównanie zmienności zawartości metanu w poszczególnych punktach pomiarowych.
2. Zgodnie ze schematem (rys. 5-14, str. 121) i komentarzem zamieszczonym na wstępie podrozdziału 5.3 zmiennymi wyjściowymi sieci były: typ dnia tygodnia, miesiąc, dzień miesiąca oraz temperatura. Zastosowanie takich zmiennych wejściowych w znacznym stopniu ogranicza potencjalnie zastosowanie modelu do celów predykcyjnych. Opracowanie efektywnego narzędzia symulacyjnego do praktycznych zastosowań powinno uwzględniać również strukturę samej sieci, w tym np. takie elementy jak: (i) odległość analizowanego punktu pomiarowego od punktów transgranicznych i terminalu LNG wraz ze składem gazu w tym punkcie; (ii) położenie/odległość od punktów/odbiorców charakteryzujących się dużym całorocznym (przemysł) jak i okresowym (duże ośrodki miejskie w sezonie zimowym) zapotrzebowaniem na gaz. W mojej opinii zastosowane w modelu zmienne wejściowe są w gruncie rzeczy zmiennymi zależnymi, na które wpływ ma szereg właściwych zmiennych niezależnych.



3. Czy wszystkie wyniki uzyskane w ramach części zasadniczej badań dotyczą tylko jednego punktu pomiarowego zlokalizowanego w powiecie czarnkowsko-trzcieńskim? Jeżeli tak to proszę o odpowiedź czym był umotywowany ten wybór oraz czy były podjęte próby opracowania sieci dla innych punktów pomiarowych.
4. Analiza i interpretacja wyników zamieszczonych w podrozdziale 5.3.1. budzi pewne wątpliwości, tj.:
  - a) Po pierwsze, o ile analiza wpływu liczności zbioru danych zastosowanych do trenowania na wyniki prognoz składu gazu ziemnego jest w pełni zasadna, nasuwa się pytanie czym uzasadnione jest przyjęcie tak zróżnicowanych okresów czasu, z których pochodziły dane? W mojej opinii zbiór  $N_4$  charakteryzuje się nie tylko większą liczbą punktów pomiarowych, ale przede wszystkim większą zmiennością w czasie poszczególnych zmiennych wyjściowych niż np. zbiór  $N_1$ . To w oczywisty sposób przekłada się na błędy dla podzbioru walidacyjnego i testowego (rys. 5-15 i 5-16).
  - b) Dalej na str. 131 Doktorantka pisze, że „Na podstawie wyników zamieszczonych na wykresach na rys. 5-17-5.20 stwierdzono, że zależności pomiędzy prognozowanymi składnikami gazu ziemnego najlepiej zostały odwzorowane za pomocą modelu MLP 17-40-5 trenowanego na zbiorze  $N_1 = 8760$  przypadków.” Brak jest jednak graficznego porównania z wynikami pomiarów, a zatem skąd taki wniosek?
  - c) Podzbiór  $N_1$  obejmuje dane pochodzące z roku 2018, podzbiór  $N_4$  to dane zebrane w latach 2017-2020, natomiast weryfikacja modeli była wykonana dla podzbioru  $N_5$ , czyli danych z roku 2021. Biorąc pod uwagę, że na przestrzeni lat może ulegać zmianie lokalne zapotrzebowanie na gaz (np. na skutek rozwoju przemysłu/ośrodków miejskich) bardziej zasadne byłoby przypisanie do podzbioru  $N_1$  danych pochodzących z roku 2020; to samo dotyczy pozostałych podzbiorów np. zbiór  $N_2$  powinien być skonstruowany z danych pochodzących z lat 2019-2020. Czy zatem przytoczone w jednej z postawionych hipotez badawczych (str. 83) zmniejszenie błędu w skutek zwiększenia liczności zbioru nie jest jednak efektem zmniejszenia „luki czasowej” pomiędzy okresem próbkowania i prognozowania składu gazu?
5. Na str. 135 zamieszczono stwierdzenie brzmiące: „im mniejszy udział danego składnika w gazie, tym większy dla niego błąd prognozy  $MAPE_j$ .” Powszechną techniką stosowaną w uczeniu maszynowym jest normalizacja zmiennych wejściowych i wyjściowych, która umożliwia zredukowanie błędu uczenia. Czy taka operacja była wykonywana lub były poczynione inne próby redukcji bardzo wysokich błędów prognoz ( $MAPE_j$  dla azotu i zbioru  $N_4$  wynosi prawie 200%, Tabela 5-9)?
6. W rozprawie dokładność prognozy składu gazu ziemnego uzyskana za pomocą modelu MLP oceniana jest na wiele sposobów, m.in. przy użyciu współczynników korelacji  $R$ , sumy kwadratów błędów  $SSE$  czy też błędów średnich  $MAPE_{\bar{s}}$  i  $nRMSE_{\bar{s}}$ . Wszystkie wspomniane miary błędów mają jednak charakter globalny, tj. odnoszą się do błędu w skali roku lub kilku lat. Brak jest natomiast w pracy graficznej analizy porównawczej wartości chwilowych składu gazu. Jedynie porównanie graficzne danych rzeczywistych z danymi prognozowanymi zostało zamieszczone na rys. 5-32 i 5-33



i dotyczy ono ciepła spalania. Proszę o wyjaśnienie przyczyny braku prezentacji wyżej wspomnianych porównań graficznych.

### **Pozostałe uwagi**

Pod względem językowym i edytorskim opracowanie zostało przygotowane bardzo starannie i nie zawiera istotnych błędów. Poniżej pozwolę sobie przytoczyć nieliczne nieścisłości i uchybienia natury redakcyjnej:

1. We wprowadzeniu na str. 13 pojawia się stwierdzenie „W niniejszej pracy doktorskiej zamiast czasochłonnych i kosztownych badań laboratoryjnych lub symulacji (...) zaproponowano opracowanie modelu (...) metodą sztucznych sieci neuronowych.” Sformułowanie to jest nieprecyzyjne, ponieważ sugeruje brak konieczności przeprowadzania badań eksperymentalnych lub symulacyjnych, a przecież każda sztuczna sieć neuronowa to model oparty na danych. Doktorantka wprawdzie nie wykonała samodzielnie pomiarów, jednak w mojej opinii należało w tym miejscu podkreślić, że model został skonstruowany w oparciu o dane udostępnione przez Operatora Gazociągów Przesyłowych GAZ-SYSTEM S.A.
2. Na str. 21 pojawia się błędne stwierdzenie, iż „obecnie dominującą technologią produkcji wodoru jest elektroliza wody”. Zgodnie z powszechnie dostępnymi raportami udział elektrolizy w produkcji wodoru wynosi obecnie kilka procent, a dominującymi technologiami są te wykorzystujące jako surowiec gaz ziemny (47% w 2021 r.) i węgiel (27% w 2021 r.).
3. W Tabeli 2-1 zamieszczonej na str. 26 brak jest źródeł literaturowych dla podanych w niej składów poszczególnych gazów.
4. Na str. 60 w odniesieniu do wejścia i wyjścia z sieci poprawnym terminem jest „zmienna”, a nie „parametr”.
5. Stwierdzenie na str. 72 „Sztuczne sieci neuronowe ze względu na zdolność modelowania złożonych problemów nieliniowych bez konieczności programowania (...)” jest nieprecyzyjne. W przypadku używania gotowych pakietów obliczeniowych programowanie oczywiście nie jest konieczne, jednak to samo dotyczy rozwiązywania modeli mechanistycznych sformułowanych w oparciu o zasady zachowania. Bardziej właściwe byłoby w tym miejscu stwierdzenie, że sieci neuronowe nie wymagają formułowania równań modelowych wywodzących się z fizyki problemu.
6. Poczynając od rys. 5-3 (str. 97) w legendach rysunków i na osiach wzory związków chemicznych w udziałach składników powinny być zapisywane przy użyciu indeksu dolnego ( $y_{\text{CH}_4}$  itd.). Analogicznie uwaga dotyczy zapisu symbolu użytego do oznaczenia przedziału ufności (np. rys. 5-9 na str. 105).





Powyższe drobne uchybienia i nieścisłości nie wpływają na wartość merytoryczną rozprawy doktorskiej i zostały wskazane z obowiązku recenzenta. Proszę, aby Doktorantka nie ustosunkowywała się do nich w trakcie publicznej obrony rozprawy.

#### **Wniosek końcowy**

Reasumując stwierdzam, iż przedłożona do recenzji rozprawa doktorska Pani mgr inż. Pauliny Muchel pt. „*Modelowanie zmienności składu gazu ziemnego metodą sztucznych sieci neuronowych*” stanowi oryginalne rozwiązanie problemu naukowego oraz potwierdza, że Doktorantka posiada umiejętności badawcze oraz wiedzę teoretyczną w zakresie dyscypliny *inżynieria chemiczna*.

W związku z powyższym stwierdzam, że rozprawa Pani mgr inż. Pauliny Muchel spełnia wymogi formalne stawiane rozprawom doktorskim określone w art. 187 ustawy z dnia 20 lipca 2018 r. *Prawo o szkolnictwie wyższym i nauce* (Dz. U. 2018 poz. 1668 z późn. zm.) i wnoszę do Rady Dyscypliny Naukowej Inżynieria Chemiczna Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie o dopuszczenie Doktorantki do dalszych etapów przewodu doktorskiego.



Katarzyna Gsa