

**Prof. dr hab. inż. Jędrzej Trajer**  
**Instytut Inżynierii Mechanicznej**  
**SGGW w Warszawie**  
**ul. Nowoursynowska 166**  
**02-787 Warszawa**

Warszawa, 28.12.2022

**Recenzja rozprawy doktorskiej mgr inż. Dominiki Sierackiej**  
**pt. „Identyfikacja parametrów charakterystycznych w plonowaniu nasion wybranych**  
**odmian konopi przemysłowych (*Cannabis sativa* L.) z użyciem metod sztucznej inteligencji”**

Przedmiotem recenzji jest rozprawa doktorska mgr inż. Dominiki Sierackiej pt. „Identyfikacja parametrów charakterystycznych w plonowaniu nasion wybranych odmian konopi przemysłowych (*Cannabis sativa* L.) z użyciem metod sztucznej inteligencji” wykonanej na stopień doktora w dziedzinie nauk inżynieryjno – technicznych, w dyscyplinie Inżynieria Mechaniczna. Rozprawa doktorska została wykonana w Katedrze Inżynierii Biosystemów Uniwersytetu Przyrodniczego w Poznaniu pod kierunkiem dra hab. inż. Macieja Zaborowicza, prof. UPP. Recenzję wykonano na podstawie uchwały Rady Naukowej Dyscypliny Inżynieria Mechaniczna Uniwersytetu Przyrodniczego w Poznaniu z dnia 10 listopada 2022 roku.

**1. Ogólna charakterystyka pracy**

Praca liczy 82 strony maszynopisu formatu A4, łącznie ze spisem treści oraz spisem tabel i rysunków. Praca została napisana według wymogów stawianych naukowym pracom eksperymentalnym i składa się z 8 rozdziałów i wykazu literatury zawierającej 95 pozycje, w większości obcojęzyczne (65 pozycji angielskojęzycznych i 30 polskojęzycznych) oraz źródła internetowe w liczbie 9. Duża część cytowanej literatury (45 pozycji) pochodzi z ostatnich 5 lat. Świadczy to o aktualności badań i znajomości wykorzystania najnowszej wiedzy w rozwiązaniu poruszonego problemu.

**2. Merytoryczna ocena pracy**

Rozprawa doktorska mgr inż. Dominiki Sierackiej dotyczy identyfikacji parametrów uprawowych nasion konopi przemysłowych i ich praktycznego zastosowania w budowie neuronowego modelu plonowania, z wykorzystaniem sztucznych sieci neuronowych. Kluczowym w budowie efektywnego modelu prognostycznego jest bowiem poprawny dobór predyktorów. Nie ma tu

skutecznych metod wyboru zmiennych wejściowych. Stosuje się różne procedury wspomagające wybór i redukcję liczebności zmiennych wejściowych, takie jak np. analizę wrażliwości, analizę składowych głównych czy wykresy macierzowe, nie są to jednak sposoby w pełni niezawodne. Nie uwzględniają choćby wzajemnych powiązań i oddziaływań na zmienną zależną, stąd zawsze testowanie różnych modeli daje ostateczną ich akceptację. Dużą pomoc w wyborze predyktorów i weryfikację modelu prognostycznego może przynieść pomoc eksperta, który zna dane zjawisko, w tym przypadku musi wykazać się dobrą znajomością uprawy konopi przemysłowych. Powyższe przesłanki świadczą, że jest to problem badawczy, który wymaga wykorzystania niekonwencjonalnych analiz. Tematyka tych badań dotyczy więc sposobu pozyskania danych, ich analizy dla potrzeb oceny plonu nasion konopi oraz stworzenia modelu prognostycznego. Doktorantka zaproponowała nowoczesne narzędzia analizy jakimi są sztuczne sieci neuronowe, które z powodzeniem wykorzystane mogą być w budowie modeli prognostycznych oraz analizę wrażliwości pomocną w identyfikacji parametrów charakterystycznych w plonowaniu nasion wybranych odmian konopi przemysłowych.

Uważam, że wybór tematyki rozprawy można uznać za właściwy i odpowiadający wymaganiom ważności tematyki badawczej stawianym pracom doktorskim i mieści się ona w zakresie wnioskowanej dziedziny nauk inżynieryjno – technicznych, w dyscyplinie Inżynieria Mechaniczna, dotyczy bowiem takich zagadnień jak: produkcja roślinna i wykorzystanie narzędzi informatycznych do analizy tej produkcji. Dyscyplina Inżynieria Mechaniczna w obecnym kształcie podziału nauki powstała z połączenia kilku wcześniej dyscyplin naukowych, a wśród nich m.in. Inżynierii produkcji i Inżynierii rolniczej. Inżynieria mechaniczna jest dyscypliną naukową, której domeną jest budowanie i eksploataowanie systemów mechanicznych łącznie z systemami produkcji i przetwarzania surowców pochodzenia biologicznego (wg. PAUza, Nr 504/2020). Opracowane neuronowe modele typu RBF mogą być z powodzeniem wykorzystane do oceny plonowania nasion konopi przemysłowych dla odmian Białobrzaska i Henola uprawianych w Polsce, co nie jest bez znaczenia przy polityce zwiększania produkcji wysokiej jakości nasion. Od 2013 do 2018 roku nastąpił bowiem aż 70% wzrost areału pod przemysłową uprawę konopi w Europie i dalej rośnie. Doktorantka zajęła się ważnym i aktualnym problemem oceny plonowania nasion konopi przemysłowych, które mają zarówno wartość przemysłową, jak i leczniczą. Jak wynika z przedstawionego przeglądu literatury dotychczas brak jest efektywnych rozwiązań, które pomogłyby w prognozowaniu plonów konopi przemysłowych.

Autorka przedstawiła problem naukowy w formie pytania, w którym brakuje określenia trudności i niewiedzy w rozwiązaniu danego zagadnienia. Dopiero dalsze postawione pytania doprecyzowują to określając cele i uwarunkowania rozwiązania w szczegółach. Temat pracy i sformułowana hipoteza badawcza wskazują, że celem głównym pracy jest identyfikacja parametrów charakterystycznych w plonowaniu nasion wybranych odmian konopi przemysłowych, natomiast wniosek pierwszy odnosi się do możliwości wytworzenia predykcyjnego modelu do oceny plonowania nasion konopi. Brak tu jest więc konsekwencji. Powstaje więc pytanie czy większe znaczenie ma sama identyfikacja predyktorów, czy stworzenie modelu predykcyjnego?

Na podstawie analizy celów rozprawy postawionych w zakresie pracy można stwierdzić, że praca ma charakter naukowy. Doktorantka podjęła się działań, mających na celu wytworzenie metodyki, umożliwiającej pozyskanie danych, ich dobór i odpowiednią reprezentację w zbiorach uczących. Autorka przetestowała różnego rodzaju typy sieci neuronowych, takie jak: rekurencyjna sieć neuronowa PNN, sieć uogólnionej regresji GRNN, sieć o radialnych funkcjach bazowych RBF oraz perceptrony wielowarstwowe MLP trój i czterowarstwowe. Brak uzasadnienia dlaczego tylko te i dlaczego te sieci wybrano do testowania. Na przykład, wielkość zbioru danych nie pozwala tu na stworzenie modelu prognostycznego z wykorzystaniem sieci typu MLP (mówi o tym wymiar Wapnika dla własności uogólniających sieci neuronowych typu MLP). Natomiast zastosowanie sieci RBF zamiast MLP spowoduje, że sieć neuronowa znajdzie aproksymację lepiej dopasowaną do lokalnych właściwości zbioru danych, ale gorzej ekstrapoluje. Sieć RBF realizuje aproksymację typu lokalnego ale umożliwia lepsze odwzorowanie cech charakterystycznych modelowanego procesu. Wyniki symulacji potwierdziły, że najlepsze okazały się sieci typu RBF. Oznacza to, że predykcja z ich wykorzystaniem nie może być zbyt uogólniona poza obszar danych.

Zaproponowana przez doktorantkę metodyka badań składała się z kilku etapów, do których zaliczyć należy: wybór odpowiednich odmian konopi uprawianych w Polsce, zebranie danych dotyczących plantacji konopi, wybór istotnych parametrów agrotechnicznych i wegetacyjnych na plon konopi i stworzenie zbioru uczącego do budowy neuronowego modelu prognostycznego. Ostatnimi etapami badań było testowanie sieci neuronowych dla różnych zmiennych wejściowych. Redukcję predyktorów dokonano na podstawie analizy wrażliwości. Ostatecznie uzyskano zbiór danych z 336 przypadkami i 24 predyktorami, który posłużył do stworzenia trzech typów modeli: z dwoma zmiennymi wyjściowymi (plon z hektara i siła kiełkowania nasion), z jedną zmienną wyjściową plon z hektara oraz siła kiełkowania nasion. Modele te zbudowano z wykorzystaniem, zaimplementowanego w pakiecie *Statistica* ver.7.1, algorytmu *Automatycznego projektanta sieci*.

Następnie Doktorantka próbowała poprawić te modele, w opcji oprogramowania *Projekt sieci użytkownika*, dobierając przez siebie parametry sieci i uczenia. Weryfikację sieci neuronowych dokonano na podstawie wyników uzyskanych z sieci w porównaniu z rzeczywistymi dla zbiorów: uczącego, walidacyjnego i testowego. Wyznaczono wartości jakości (utożsamianych ze współczynnikiem korelacji) i błędów dla tych zbiorów. Zbyt wysokie wartości jednego z typów lub wszystkich błędów lub duże rozbieżności pomiędzy jakościami dla poszczególnych zbiorów wskazują na potrzebę modyfikacji struktury modelu, zestawu zmiennych wejściowych bądź wskazują na błędnie przeprowadzoną procedurę uczenia. Wyniki z wykorzystaniem *Automatycznego projektanta sieci* pokazują, że model ze zmienną wyjściową plon wymaga poprawy (wartości jakości sieci dla zbiorów są odstające, a jakość dla zbioru walidacyjnego niepoprawna). Sieci typu RBF bywają nadmiernie wrażliwe na nawet nieliczne błędy w danych uczących. Czy to sprawdzono w tym przypadku? Doktorantka postanowiła poprawić wszystkie trzy modele, co udało się Jej osiągnąć, jeśli idzie o wskaźniki oceny. Natomiast stworzone sieci są bardziej rozbudowane (liczba neuronów w warstwie ukrytej 9), co sugerować może ograniczone możliwości rozszerzenia wyników poza zasięg zbioru danych. Na jakiej podstawie Doktorantka przyjęła taką liczbę neuronów w warstwie ukrytej, czy istnieje tutaj dowolność?

Wyniki badań oraz ich analiza zostały przedstawione przejrzysto i adekwatnie do sformułowanego celu. W dyskusji wyników Doktorantka słusznie zauważa, że w procesie modelowania neuronowego ważnym jest przeprowadzenie analizy wrażliwości, która umożliwia uproszczenie struktury sieci (zmienne wejściowe mało znaczące mogą być usuwane z modelu). Zły bowiem dobór złożoności architektury sieci i tym samym jej nieadekwatność w stosunku do złożoności problemu może doprowadzić do zupełnie nieprzydatnych rezultatów, a mianowicie braku zadawalającej generalizacji. Z jednej strony zostało to wykorzystane w uproszczeniu sieci poprzez redukcję zmiennych wejściowych, natomiast przy tworzeniu przy użyciu funkcji *Projektanta sieci* strukturę tę rozbudowano poprzez zwiększenie liczby neuronów do 9 w warstwie ukrytej. Dyskusyjne jest więc stwierdzenie, że te sieci mają lepsze własności przewidywania niż stworzone z wykorzystaniem *Automatycznego projektanta sieci*. Jak podaje sama Doktorantka, przy wyborze sieci neuronowych, jest to problem regresyjny. Regresja odnosi się do przewidywania z zakresu danych modelowanego systemu, natomiast predykcja w znaczeniu uogólniania poza zakres danych. Rozdział 6 pod tytułem „Dyskusja” sugeruje dyskusję uzyskanych wyników, natomiast dotyczy analizy stanu wiedzy na temat wykorzystania sztucznej inteligencji w

produkcji rolniczej. Powinien więc być podsumowaniem przeglądu literatury i stanowić podstawę sformułowania problemu badawczego.

### **3. Uwagi szczegółowe**

Moje uwagi krytyczne odnoszące się do monografii dotyczą następujących kwestii:

- Doktorantka analizując zagadnienie identyfikacji parametrów charakterystycznych w plonowaniu nasion wybranych odmian konopi i tworzeniu neuronowych modeli używa pojęcie predykcja i nie precyzuje tego pojęcia. Czym innym jest przewidywanie wyników w problemie regresyjnym (aproksymacja), a czym innym w problemie prognostycznym rozumianym jako przewidywanie przyszłych stanów zjawisk poza zakres danych (uogólnianie).
- Proszę o wyjaśnienie co Doktorantka rozumie pod pojęciem „efektywnej predykcji” w określeniu problemu badawczego oraz określeniu „modelu neuronowego wspierającego predykcję” w sformułowanej hipotezie. Czy problem i hipoteza nie powinny być trafniej sformułowane ?
- Doktorantka w wielu miejscach używa terminu zmienne, zamiast zmienne wejściowe lub predyktory, a na stronie 50 w opisie podziału zbioru danych pisze o nim jako o zbiorze uczącym. Opis pod rysunkiem 18 jest nieadekwatny do przedstawianego modelu neuronowego.
- Wniosek 1 nie odnosi się do sformułowanej hipotezy, a wniosek 3 sugeruje, że tylko sieci neuronowe typu RBF opisują zjawiska nieliniowe – co nie jest prawdą.

### **4. Podsumowanie**

Powyższe uwagi nie obniżają pozytywnej oceny pracy, stanowią jedynie wskazówki dla Doktorantki dotyczące publikacji wybranych jej części. W ramach ocenianej pracy zastosowała prawidłowe, nowoczesne narzędzia badawcze i wykazała się bardzo dobrą znajomością zagadnień dotyczących zastosowań sztucznej inteligencji, a także dociekliwością badacza i umiejętnością logicznego wnioskowania.

Opracowane neuronowe modele umożliwiają nie tylko wyznaczenie procenta wykiełkowanych nasion i plonu z hektara, a także identyfikację i znaczenie parametrów charakterystycznych w plonowaniu nasion konopi przemysłowych, co może być wykorzystane w praktyce ich uprawy.

## 5. Ocena końcowa

Podsumowując uważam, że przedstawiona do oceny rozprawa doktorska mgr inż. Dominiki Sierackiej pt. „Identyfikacja parametrów charakterystycznych w plonowaniu nasion wybranych odmian konopi przemysłowych (*Cannabis sativa L.*) z użyciem metod sztucznej inteligencji”, spełnia ustawowe wymagania stawiane rozprawom doktorskim w dziedzinie nauk inżynieryjno – technicznych, w dyscyplinie Inżynieria Mechaniczna i może być dopuszczona do publicznej obrony.



Jędrzej Trajda