

Wrocław, 4.01.2024

Dr hab. inż. Wojciech Golimowski, prof. UEW

RECENZJA

rozprawa doktorska pt.

„Prognozowanie plonu i zawartości białka w nasionach grochu siewnego (*Pisum sativum* L.) z zastosowaniem sztucznych sieci neuronowych oraz regresji wielorakiej w oparciu o wyniki doświadczeń odmianowych ”

Pana mgr inż. Patryka Hara

1. Wstęp

Prognozowanie plonów jest istotnym i ważnym działaniem w produkcji rolniczej. Predykcja produkcji roślinnej ma znaczenie logistyczne i przekłada się bezpośrednio na efekt ekonomiczny gospodarstw rolnych. Duża ilość czynników, istotnie wpływających na plon oraz potrzeba uzyskiwania wiarygodnych danych jest powodem do opracowania coraz nowszych technik modelowania. Sztuczne sieci neuronowe są narzędziem pozwalającym na wielowymiarową analizę danych, opisujących nieograniczoną ilość zmiennych niezależnych w celu uzyskania konkretnej informacji.

Wraz z rozwojem nowych i bardziej precyzyjnych technik uprawy roślin oraz wysokiego poziomu zmechanizowania procesów produkcji roślinnej istnieje potrzeba analizy wielu danych. W konsekwencji zmieniają się metody zarządzania gospodarstwami rolnymi, w tym redukcja osób zaangażowanych w prace polowe. Z tego powodu ciągłe monitorowanie upraw jest trudne a predykcja plonów obciążona dużym błędem. Prognozowanie plonów w małych gospodarstwach nie jest tak istotne jak w przypadku dużych przedsiębiorstw rolnych, do których zarządzania zatrudnia się wykształcone osoby, nie posiadające doświadczenia przekazywanego z pokolenia na pokolenie. W efekcie dostęp do precyzyjnych narzędzi analitycznych jest wysoce pożądanym. Informacja o wielkości plonów jest niezbędna do zaplanowania przestrzeni magazynowych na poziomie producenta rolnego, szacowania wielkości produkcji w skali kraju bezpośrednio wpływa na cenę produktów pochodzenia roślinnego.

Zmiana klimatu, odmiany roślin, stosowane nawozy, środki ochrony roślin oraz żyzność gleb są czynnikami zmieniającymi się w każdym okresie produkcyjnym. Zastosowanie uczenia maszynowego w celu uwzględniania tych zmian, a zwłaszcza sztucznych sieci neuronowych

do analizy tak dużej liczby zmiennych niezależnych jest w pełni uzasadnione. Dlatego stwierdzam, że przedmiot badań podjętych przez Pana mgr inż. Patryka Hara wpisuje się w aktualną problematykę badań naukowych w dziedzinie nauk rolniczych, w dyscyplinie rolnictwo i ogrodnictwo.

2. Ogólny opis recenzowanej rozprawy doktorskiej

Przedłożona do recenzji rozprawa doktorska pt. „Prognozowanie plonu i zawartości białka w nasionach grochu siewnego (*Pisum sativum* L.) z zastosowaniem sztucznych sieci neuronowych oraz regresji wielorakiej w oparciu o wyniki doświadczeń odmianowych” Pana mgr inż. Patryka Hara jest wynikiem badań empirycznych podzielonych na trzy etapy. Wyniki badań każdego z etapów zostały opublikowane w wysoko punktowanych czasopismach naukowych o zasięgu międzynarodowym. Praca właściwa została zawarta na 48 stronach maszynopisu w języku polskim, do której załączono trzy recenzowane publikacje naukowe. Maszynopis składa się z jedenastu rozdziałów, chronologicznie występujących po sobie. Na wstępie Doktorant zamieścił wykaz symboli i akronimów oraz streszczenie w języku polskim i angielskim, co znacząco ułatwiło zrozumienie treści zawartej w pracy. Właściwa część pracy zaczyna się od rozdziału „Wprowadzenie”, w którym opisano, w sposób syntetyczny, genezę pracy. Następny rozdział „Przegląd aktualnego stanu wiedzy” został poświęcony analizie stanu wiedzy nt. upraw grochu i uczenia maszynowego, w tym sztucznych sieci neuronowych. W rozdziale trzecim zawarto dane bibliograficzne publikacji, stanowiących podstawę rozprawy doktorskiej. W czwartym rozdziale uzasadniono potrzebę podjęcia pracy naukowej, sformułowano cel pracy, ukierunkowanej na rozwiązanie problemu badawczego, sformułowanego w postaci trzech pytań. Hipotetycznie przyjęto trzy tezy, które poddano confirmacji na podstawie wykonanych badań empirycznych. Kolejne trzy rozdziały są poświęcone omówieniu wyników badań. W rozdziale pierwszym z trzech opisano wybór zmiennych niezależnych, niezbędnych do prognozowania plonów, w drugim prognozowano zawartość białka w grochu za pomocą modelu ANN oraz MLR, w trzeciej wykonano analizę porównawczą skuteczności sieci neuronowych i wielorakiej regresji liniowej w przewidywaniu plonu grochu siewnego. W rozdziale ósmym omówiono wyniki badań w odniesieniu do aktualnego stanu wiedzy. W rozdziale dziewiątym podsumowano pracę i wyciągnięto wnioski. Pozostałe dwa rozdziały stanowią dane bibliograficzne i trzy załączniki.

3. Ocena pracy pod względem formalnym

Praca została napisana poprawnie w języku polskim. Doktorant wykazał się dużą starannością, gdyż nieliczne błędy gramatyczne i stylistyczne nie wpływają na trudność w zrozumieniu tekstu, dlatego nie wymagają żadnego komentarza. Autoreferat powstał na podstawie opublikowanych wyników własnych badań i analizy 89 pozycji literaturowych w przeważającym stopniu anglojęzycznych z renomowanych czasopism naukowych. Analizując daty publikacji stwierdzam, że w większości są to publikacje aktualne z ostatnich 5 lat. Wyniki badań Doktoranta zostały opublikowane w prestiżowych czasopismach naukowych o zasięgu międzynarodowym. Bez wątplenia wyniki opublikowanych badań, podzielonych na trzy etapy, są podstawą do sformułowania odpowiedzi na postawione pytania w autoreferacie, które stanowiły problem badawczy do rozwiązania.

Układ pracy jest logiczny. Streszczenie zawiera skrótowe informacje o celu pracy, metodycznym opisie doświadczenia, wynikach oraz wnioskach uzyskanych po zrealizowaniu całego planu badawczego. W genezie pracy, opisanej we wstępie, Doktorant podkreślił znaczenie i istotę stosowania uczenia maszynowego, w tym sieci neuronowych. Przedstawił szeroki zakres ich stosowania skupiając się przede wszystkim na obszarze rolniczym. Zaprezentował walory użytkowe oraz kierunek bezpośredniego wykorzystania modeli do generowania informacji niezbędnych do planowania produkcji roślinnej.

Analiza stanu wiedzy została ograniczona do dwóch obszarów, omówiono wyczerpująco znaczenie upraw grochu siewnego i uczenia maszynowego, w tym sieci neuronowych do generowania modeli, używanych do prognozowania plonu grochu siewnego. Doktorant odniósł się do informacji literaturowych w celu opisanie obszaru badawczego, zdefiniował istotne elementy badanego zagadnienia, co ważne, krytycznie odniósł się do aktualnego stanu badań naukowych, udowadniając konieczność podjęcia badań empirycznych. Jedyną, niezrozumiałą rzeczą od strony edytorskiej jest zapis „opracowanie własne, fot. Bartłomiej Sawicz” pod rysunkiem nr 1. W jakim zakresie było wykonane opracowanie zdjęcia przez Doktoranta? Zastrzeżenie mam również do ostatniego akapitu w rozdziale 2.2., w którym Doktorant użył sformułowania „ nadal istnieje wyraźna luka w zakresie prognozowania plonów „, w dalszej części pracy jest uzasadnienie, jednak uważam, że w tym miejscu należało tę lukę zdefiniować, a nie ograniczyć się tylko do zasygnalizowania istnienia pewnych braków wiedzy.

Przed sformułowaniem kluczowych tez, pytań i czynności Doktorant w rozdziale czwartym przedstawił znaczenie wyników badań własnych. Podkreślił praktyczną potrzebę predykcji plonów roślin przed ich zbiorem oraz rzetelność uzyskiwanych informacji. Wskazał na ogromną liczbę czynników mogących w sposób nieprzewidywalny istotnie zmienić szacowany plon. Dlatego celem pracy Doktoranta była budowa modeli prognozujących zawartość białka i plon grochu siewnego. Sprecyzował również cel pomocniczy, którym było wykonanie analizy wrażliwości sieci neuronowych zmiennych niezależnych, które w największym stopniu wpływają na ilość i jakość plonu nasion grochu. Zgodnie z przyjętymi standardami cel pomocniczy jest składową celu głównego. Sądzę, z formalnego punktu widzenia, że należało użyć konstrukcji cel główny i cel dodatkowy. Proszę o komentarz do mojej uwagi. Doktorant za zmienne zależne przyjął plon ziarna i udział w nim białka. Założył, 20 dniowy okres wyprzedzający datę zbioru grochu. Do badań własnych użył bazy danych COBORU zawierającą wyniki badań odmianowych z okresu ostatnich 5 lat. Ponadto za cel pomocniczy przyjął wskazanie zmiennych niezależnych, które w istotny sposób powodowały zmianę plonowania. Osiągnięcie celu badań skutkowało pozyskaniem niezbędnych danych do rozwiązania problemu badawczego sformułowanego w postaci trzech pytań:

1. Czy możliwe jest zbudowanie modeli prognozujących plon i zawartości białka w nasionach grochu na 20 dni przed zbiorem za pomocą sztucznych sieci neuronowych?
2. Czy za pomocą modeli neuronowych zostanie uzyskana większa dokładność predykcji plonu i zawartości białka w nasionach grochu odmian ogólnoużytkowych w stosunku do modelu wielorakiej regresji liniowej?
3. Czy analiza wrażliwości sieci neuronowej umożliwia wykonanie wartościowania czynników wpływających na plon grochu i zawartość białka w nasionach?

Tak sformułowane pytania, w mojej ocenie, nie wymagają podejmowania badań empirycznych, ponieważ odpowiedzi na nie można znaleźć w literaturze. Natomiast uzupełnienie tego rozdziału o hipotetyczne odpowiedzi na te pytania istotnie precyzują ich znaczenie. Dlatego z formalnego punktu widzenia logicznie został sformułowany problem badawczy. W tym miejscu nasuwa się pytanie, czy wyniki badań mają charakter uniwersalny czy dotyczą wyłącznie grochu siewnego? Zarówno w pytaniach, jak również w hipotezach, ta kwestia nie została doprecyzowana.

Doktorant swój pierwszy etap badań ukierunkował na wybór zmiennych niezależnych, dokonując analizy danych literaturowych, których skrótowy opis umieścił w piątym rozdziale autoreferatu, a szczegółowy w publikacji pt. : „Selection of independent variables for crop yield

prediction using artificial neural network models with remote sensing data”. W wyniku krytycznej analizy stanu wiedzy Doktorant w sposób obiektywny dokonał oceny jakościowej zmiennych niezależnych. Doszedł do wniosku, że nie tylko liczebność zmiennych ma znaczenie ale również ich jakość. Wskazał konieczność wykonywania selekcji czynników, która wynika z niekompletnych baz danych i skrajnie odstępujących od trendu wartości. Ponadto wskazał możliwości wykorzystania wskaźników związanych z produktywnością roślin. Dokonał podziału tych wskaźników na cztery kategorie: pierwsza dotyczy wzrostu roślin, druga relacji pomiędzy fotosyntezą a przyrostem biomasy, trzecia wegetacji roślin a czwarta zawartości chlorofilu. Wskaźniki te uznał jako zmienne niezależne. Podkreślił również ogromne znaczenie selekcyjnego doboru danych do budowy modeli z uwagi na liczne błędy w bazach danych, wynikające z wykonywania pomiaru i rejestrowania tych danych. W w/w publikacji przeanalizował również trendy w tworzeniu modeli predykcyjnych. W podsumowaniu wskazał na problemy praktycznego zastosowania modeli. Co jest istotne, zdefiniował źródła problemów dowodząc dobrej znajomości narzędzi, którymi posługiwał się w dalszych etapach swoich badań.

W rozdziale szóstym Doktorant opisał wyniki badań opublikowanych w publikacji pt. „Prediction of Protein Content in Pea (*Pisum sativum* L.) Seeds Using Artificial Neural Networks”. Celem podjętych badań było opracowanie modeli predykcji białka w grochu siewnym przy wykorzystaniu sztucznych sieci neuronowych i porównaniu ich skuteczności do modeli regresyjnych. Doktorant budował modele na dużych bazach danych, obejmujących pięć okresów wegetacyjnych 11 odmian grochu uprawianego na 1155 poletkach w 7-miu lokalizacjach na terenie Polski. Bazy danych pochodziły z ksiąg polowych COBORU, a dane meteorologiczne ze stacji w zakładach doświadczalnych COBORU. Do budowy modeli wykorzystał dane z 1040 poletek a ich poziom precyzji analizowania zweryfikował na 115 poletkach. Doktorant wykazał przewagę sieci neuronowych nad podstawową analizą statystyczną. Precyzja predykcji modeli regresyjnych była znacznie gorsza od modeli neuronowych. Podstawą do wyciągnięcia wniosków było porównanie wielkości błędów MAPE (błąd średni bezwzględny procentowy) i MAE (błąd średni bezwzględny). Dla modeli RS błędy te były prawie trzy i pół krotnie większe względem modeli neuronowych N1. Doktorant wskazał wysoce istotny wpływ ilości magnezu w glebie względem pozostałych 19-nastu analizowanych czynników. Wyniki zaprezentowane w publikacji w całości wpisują się w plan badawczy rozprawy doktorskiej.

W rozdziale siódmym Doktorant omówił trzeci etap badań, wpisany w plan rozprawy doktorskiej. Skrótowo omówił wyniki badań, które szczegółowo opublikował w artykule pt. „Prediction of Pea (*Pisum sativum* L.) Seeds Yield Using Artificial Neural Networks”. Za cel badań przyjął analizę jakościową predykcji dwóch modeli: model neuronowy N2 i regresyjny R2. Do badań użył te same dane, na których zbudował modele do szacowania zawartości białka w grochu siewnym. Wyniki jednoznacznie wskazywały na przeważającą skuteczność modeli neuronowych względem modeli regresyjnych. Współczynnik determinacji wynosił odpowiednio 0,83 i 0,58 a błąd średniokwadratowy 0,44 t/ha i 6,4 t/ha. Wielkość błędu predykcji plonów dla modelu regresyjnego skutkuje wykluczeniem go w dalszych badaniach. Doktorant, w wyniku wykonanych badań empirycznych, doszedł do wniosku, że na wielkość plonu najistotniejszy wpływ miała data początku dojrzewania. Wyniki przedstawionych w publikacji badań wpisują się w całości rozprawę doktorską Pana mgra inż. Patryka Hara.

Reasumując, pod względem formalnym opisane w autoreferacie wyniki badań empirycznych w całości są zgodne z tytułem rozprawy doktorskiej i są podstawą do sformułowania wniosków, dając tym samym jednoznaczne odpowiedzi na postawione pytania badawcze. Opisane wyniki badań, kształt pracy, plan badawczy, postawione tezy i ich confirmacja zostały wykonane w sposób prawidłowy. Doktorant wykazał się dużymi umiejętnościami w posługiwaniu się nowoczesnymi narzędziami badawczymi, ponad przeciętną starannością przygotowania maszynopisu oraz udowodnił, publikując prace naukowe w czasopismach wysoko punktowanych, że jest dobrze przygotowany do realizacji przyszłych badań naukowych. Przedstawiona do recenzji praca pod względem formalnym nie budzi zastrzeżeń, co jest podstawą do pozytywnej oceny.

4. Ocena pracy pod względem merytorycznym

Rozprawa doktorska Pana mgra inż. Patryka Hara dotyczyła predykcji plonów grochu siewnego na 20 dni przed zbiorem. Obiektem badań było zastosowanie sztucznych sieci neuronowych do analizy zmiennej zależnej jakościowej (zawartość białka) i ilościowej (plonu ogólnego) grochu siewnego. Niewątpliwie przedmiot badań jest innowacyjny i niezwykle ważny w rolnictwie zrównoważonym, gdzie rośliny strączkowe odgrywają coraz większe znaczenie przy zmianowaniu upraw. Białko jest podstawowym składnikiem w diecie organizmów żywych, dlatego niezwykle ważne jest poznanie czynników mających istotny wpływ na plonowanie roślin wysokobiałkowych.

Analizując tytuł pracy, w którym Doktorant do prognozowania plonów i zawartości białka stosuje dwie metody, klasyczne narzędzia statystyczne w postaci regresji liniowej i nowoczesne w postaci sztucznych sieci neuronowych bazując na wynikach doświadczeń odmianowych. Tak sformułowany tytuł sugeruje wykonanie równoległych badań przy założeniu, że obie metody mogą charakteryzować się wysoką skutecznością. Analizując dalej sformułowany problem badawczy w postaci trzech pytań precyzyjnie wskazuje na zastosowanie sztucznych sieci neuronowych w odniesieniu do modeli wielorakiej regresji liniowej. W tym miejscu należy wyjaśnić tę niespójność między tytułem pracy a sformulowanym problemem badawczym.

Problem badawczy, którego rozwiązania się podjął Doktorant jest niezwykle trudny z uwagi na liczebność czynników powodujących zmiany wielkości plonowania. W prezentowanych badaniach analizowano aż 19 zmiennych niezależnych, ponadto dokonał ich klasyfikacji pod względem istotności wpływu na plonowanie jakościowo-ilościowe. Przy tak skonstruowanym planie badawczym nie tylko udało się opracować wysoce wartościowe narzędzia do prognozowania, ale również wskazać poziom istotności zmiennych niezależnych. Dane te pozwalają na kontynuowanie badań w celu zwiększania dokładności plonowania wykluczając część zmiennych wpływających na poziom błędu statystycznego.

Badania wykonano na podstawie pięciu okresów wegetacji roślin uprawianych w siedmiu lokalizacjach na 1155 poletkach jedenastu odmian grochu siewnego. Tak duża ilość danych bez wątplenia była podstawą do wykonania rzetelnych badań i uzyskania wysoce wartościowych wyników. Model zbudowano na zbiorze 1140 poletek a weryfikacje jego skuteczności wykonano na 115 poletkach, które nie stanowiły danych do budowy modeli. Taki plan badań gwarantuje bez wątplenia rzetelną ocenę skuteczności opracowanych modeli do predykcji plonów. Część metodyczna badań wymaga wyjaśnienia, na jakiej podstawie dokonano podziału liczebności poletek w grupie A i B?

Za cel pracy Doktorant przyjął zbudowanie modeli do prognozowania plonu grochu siewnego oraz poziomu zawartego w nim białka. Założył 20-dniowy czas prognozowania przed zbiorem. W tym miejscu należy wyjaśnić, na jakiej podstawie został wskazany tak precyzyjny okres. Cel został osiągnięty. Doktorant zbudował dwa modele N1 przy użyciu sieci neuronowych i RS przy użyciu wielorakiej regresji liniowej do prognozowania białka. Kolejne dwa modele N2 i RS2 analogicznie do prognozowania ilościowego grochu. Model N1 (przewidujący procentową zawartość białka) cechował się błędem MAPE na poziomie 2,72%. W przypadku predykcji plonu grochu błąd MAPE dla modelu N2 wyniósł 7,98%, w odniesieniu do danych literaturowych był wynikiem wysoce satysfakcjonującym. Tym samym empirycznie

dowiódł prawdziwości hipotetycznie przyjętych tez, które są odpowiedzią na postawiony problem badawczy. Ponadto dowiódł wysoce istotną przewagę nad modelami RS i RS2 budowanych za pomocą wielorakiej regresji liniowej, gdzie błąd MAPE wyniósł odpowiednio 8,85 i 148,59%. Współczynnik determinacji R^2 analogicznie był wyższy dla modeli N1 i N2 i wyniósł odpowiednio 0,80 i 0,83. Na tak dużą ilość zmiennych, czynników mających wpływ na wartość prognozowanych plonów ilościowo jakościowych dopasowanie modelu do danych rzeczywistych jest bardzo wysokie. Modele RS i RS2 charakteryzowały się istotnie niższym poziomem dopasowania, gdzie R^2 wyniósł odpowiednio 0,34 i 0,58. Błąd aproksymacji modelu RMSE dla N1 i N2 wyniósł odpowiednio 0,595 i 0,650.

Drugim, równoległym efektem badań było sklasyfikowanie poziomu istotności czynników powodujących zmianę zawartości białka w grochu oraz wielkość plonu. Z pośród 19 zmiennych to skład mineralny gleby, a dokładniej ilość magnezu, fosforu i wapnia najistotniej wpływały na parametr jakościowy. Najmniej istotny był okres kwitnienia i odmiana. Ciekawy również był wysoce mało istotny wpływ opadów, plasując ten czynnik dopiero na 12 miejscu. Na plon ilościowy najistotniejszy wpływ miał okres dojrzewania roślin, data zbioru oraz wielkość opadów. W mojej ocenie są to ciekawe dane do wykonania badań optymalizacyjnych, ukierunkowanych na prognozowania ilości całkowitego białka.

Praca pod względem merytorycznym jest wysoce wartościowa. Pan mgr inż. Patryk Hara dowiódł, że potrafi posługiwać się nowoczesnymi metodami badawczymi oraz potrafi dobrze zaplanować badania. Wyniki uzyskanych badań stanowią duży wkład w rozwój nauki w dziedzinie nauk rolniczych w dyscyplinie rolnictwo i ogrodnictwo. Podjął się trudnego zagadnienia, wykazał się dużą precyzją i starannością w dążeniu do uzyskania rzetelnych wyników badań. Jest to podstawą do wystawienia oceny pozytywnej pracy pod względem merytorycznym.

Wniosek końcowy

Na podstawie przedłożonej do oceny rozprawy doktorskiej pt. „Prognozowanie plonu i zawartości białka w nasionach grochu siewnego (*Pisum sativum* L.) z zastosowaniem sztucznych sieci neuronowych oraz regresji wielorakiej w oparciu o wyniki doświadczeń odmianowych” stwierdzam, że Pan mgr inż. Patryk Hara spełnia wszystkie wymagania stawiane rozprawom doktorskim, określone w art. 187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (tekst jednolity Dz. U. z 2023 r., poz. 742). Przedmiotem rozprawy doktorskiej jest, bez wątpienia, oryginalne rozwiązanie problemu naukowego,

wpisujące się w dyscyplinę rolnictwo i ogrodnictwo, w dziedzinie nauk rolniczych. Na tej podstawie wnoszę do Rady Naukowej Instytutu Technologiczno-Przyrodniczego – Państwowego Instytutu Badawczego o dopuszczenie Pana mgr inż. Patryka Hara do dalszego postępowania w sprawie nadania stopnia doktora. Ponadto, ponad przeciętna wartość merytoryczna i wysoka ocena formalna rozprawy doktorskiej są argumentami przemawiającymi za wnioskowaniem do Rady Naukowej Instytutu Technologiczno-Przyrodniczego – Państwowego Instytutu Badawczego o wyróżnienie rozprawy doktorskiej Pana mgr inż. Patryka Hara.

Wojciech Golimowski

4.01.2024

Dr hab. inż. Wojciech Golimowski, prof. UEW