Wstęp – cel zakres

Wprowadzenie teoretyczne – sieci neuronowe, klasyfikacja, regresja, kan i mlp, miary oceny klasyfikatorów

Przegląd literatury/istniejące rozwiązania – ocenianie skuteczności (jak oni oceniali)

Opis mojego rozwiązania

Eksperymenty - Dyskusja/wyniki

Podsumowanie

Spis treści

[1. Wstęp 2](#_Toc206657253)

[1.1 Cel pracy 2](#_Toc206657254)

[1.2 Zakres pracy 2](#_Toc206657255)

[2. Wprowadzenie teoretyczne 2](#_Toc206657256)

[2.1 Uczenie nienadzorowane i nadzorowane 2](#_Toc206657257)

[2.2 Sieci neuronowe 3](#_Toc206657258)

[2.3 Wczesne zatrzymywanie 3](#_Toc206657259)

[2.4 Regresja i klasyfikacja 4](#_Toc206657260)

[2.5 Metody używane przy porównywaniu modeli 4](#_Toc206657261)

[2.6 KAN i MLP 5](#_Toc206657262)

[3. Istniejące rozwiązania 7](#_Toc206657263)

[3.1 Drzewo decyzyjne 7](#_Toc206657264)

[3.2 K najbliższych sąsiadów 8](#_Toc206657265)

[4. Zbiory danych używane w pracy 8](#_Toc206657266)

[4.1 Breast Cancer Wisconsin (Diagnostic) (Wolberg 1995) 9](#_Toc206657267)

[4.1.1 Opis danych 9](#_Toc206657268)

[4.1.2 Opis wartości 9](#_Toc206657269)

[4.2 Iris (Anderson 1935) 10](#_Toc206657270)

[4.2.1 Opis danych 11](#_Toc206657271)

[4.2.2 Opis wartości 11](#_Toc206657272)

[4.3 Wine (Stefan Aeberhard 1994) 11](#_Toc206657273)

[4.3.1 Opis danych 11](#_Toc206657274)

[4.3.2 Opis wartości 11](#_Toc206657275)

[5. Opis eksperymentów 12](#_Toc206657276)

[6. Wyniki 12](#_Toc206657277)

[6.1 Hipoteza 1: Przy takiej samej ilości parametrów KANy osiągną wyższą dokładność niż MLP. 13](#_Toc206657278)

[6.2 Hipoteza 2: Bardziej elastyczne KANy nie osiągają akceptowalnej dokładności ze względu na zbyt małą ilość epok 18](#_Toc206657279)

[6.3 Hipoteza 3: Bardziej złożone KANy nie są w stanie w pełni nauczyć się przy zadanych parametrach. 24](#_Toc206657280)

[7. Wnioski 30](#_Toc206657281)

[8. Bibliografia 30](#_Toc206657282)

# Wstęp

Jednym z najpopularniejszych zagadnień tej dekady jest sztuczna inteligencja. Oprócz zastosowań generatywnych wykazuje ona niesamowity potencjał w zadaniach klasyfikacji i regresji. W czerwcu 2024 zaproponowany został alternatywny do najczęściej używanego MLP sposób tworzenia modeli do klasyfikacji i regresji: KAN (Ziming i inni 2024).

## Cel pracy

W artykule wprowadzającym KAN nacisk kładziony jest na obszary, gdzie KAN jest lepszy od MLP. Celem tej pracy jest porównanie modeli KAN i MLP dla zadań klasyfikacji na łatwo dostępnych i często używanych przez innych naukowców danych testowych w celu niezależnej weryfikacji potencjalnej przewagi KAN nad MLP.

## Zakres pracy

[TODO]

# Wprowadzenie teoretyczne

Dla prostszego zrozumienia omawianych zagadnień zdefiniowane i wyjaśnione zostaną używane pojęcia i koncepty.

## Uczenie nienadzorowane i nadzorowane

Uczenie nienadzorowane to proces uczenia maszynowego, którego celem jest podzielenie zbioru danych na grupy podobnych danych bez potrzeby nadawania im klas przed uczeniem (Wu 2022).

Uczenie nadzorowane, w kontekście uczenia maszynowego, jest to proces, który pozwala na dopasowanie wyników dawanych przez modele regresji lub klasyfikacji do oczekiwanych wyników. W trakcie uczenia model ma dostęp do danych wejściowych (np. wektor liczb) oraz oczekiwanego wyniku dla tych danych (np. konkretna wartość liczbowa) (Mirjalili i Sawka 2019).

W tej pracy stosowane było uczenie nadzorowane.

## Sieci neuronowe

W informatyce sieci neuronowe to sposób przeprowadzania obliczeń inspirowany biologicznymi sieciami neuronowymi m.in. tymi, które można znaleźć w ludzkim mózgu (Hardesty 2017). Sieć składa się ze sztucznych neuronów: funkcji matematycznych również skonstruowanych na wzór neuronów w ludzkim mózgu, które w zależności od siły otrzymanego bodźca przekazują go dalej lub nie a w trakcie życia organizmu (m.in. w trakcie uczenia się) mogą zmieniać swoją organizację umożliwiając adaptację (Costandi 2016). W zależności od implementacji sztuczne neurony uczą się poprzez zmianę zmiennych koniecznych do obliczenia sygnału wyjściowego. Zmienne te mają różne zastosowania w zależności od implementacji. W uczeniu nadzorowanym najpopularniejszą metodą dopasowywania wyników sieci neuronowych do poprawnych wyników jest propagacja wsteczna. Pozwala ona wyznaczyć kierunek zmian jakie powinny zajść w wartościach zapisanych w neuronach by kolejny wynik podany przez sieć dla tych samych danych był bardziej zbliżony do oczekiwanego wyniku (Bergmann i Stryker 2024).

## Wczesne zatrzymywanie

Jest to sposób na zapobieganie zbytniemu dopasowaniu się modelu do danych testowych w uczeniu korzystającym z propagacji wstecznej. Jeśli model jest bardzo plastyczny (ma duże możliwości dokładnego dopasowania się do danych) może zbyt silnie dopasować się do danych treningowych co prezentuje Rysunek 2.1.

Obraz zawierający diagram, zrzut ekranu, linia, Wykres

Zawartość wygenerowana przez AI może być niepoprawna.

Rysunek . Przykład modelu, który jest zbyt mało plastyczny by dopasować się do danych (lewo) modelu dobrze dopasowanego do danych (środek) oraz modelu nadmiernie dopasowanego (prawo). Źródło: Kate Crawford https://www.youtube.com/watch?v=fMym\_BKWQzk&t=640s

We wczesnym zatrzymywaniu model uczony jest na danych treningowych, ale poprawność jego dopasowania sprawdzana jest na odrębnym zbiorze danych – danych walidacyjnych i to najlepszy wynik na tych danych decyduje o tym, który z modeli jest najlepszy. Dzięki temu modele zbyt dobrze dopasowane do danych treningowych dostają gorszą ocenę niż modele mogące dobrze uogólnić i dać odpowiedzi dla wcześniej niewidzianych danych.

## Regresja i klasyfikacja

Regresja opisuje proces, w którym model przewiduje wartość zmiennej na podstawie dostarczonych danych np. określenie wzrostu osoby na podstawie jej wieku i genotypu (Cornell Bowers Computer Science 2022).

Klasyfikacja polega na przypisaniu danej próbki z ogółu danych (najczęściej jest to kolekcja cech pojedynczego obiektu lub wydarzenia) do jednej z poprzednio zdefiniowanych klas. W uczeniu maszynowym próbki reprezentowane są jako wektory liczb i ich klasyfikacja zachodzi na podstawie podobieństwa badanej próbki do wzorca klasy („idealnego” reprezentanta danej klasy) lub stwierdzenia, że badana próbka posiada cechy uznawane za charakterystyczne dla danej klasy. (Pelikant 2022)

## Metody używane przy porównywaniu modeli

Podczas eksperymentów jako wartość pozwalającą stwierdzić, który z modeli jest lepszy wybrano dokładność (accuracy). Dokładność klasyfikacji jest jedną z najbardziej intuicyjnych, prostych i najczęściej używanych metryk określania jakości klasyfikatora. Oblicza się ją jako procent ilości poprawnie zaklasyfikowanych próbek dzielonej przez ilość wszystkich próbek, najwyższą możliwą do uzyskania dokładnością jest 100%. Należy jednak brać pod uwagę, że ta metoda porównywania klasyfikatorów ma również wady: w przypadku zbiorów danych, gdzie jedna klasa jest znacznie liczniejsza niż inne, model może osiągnąć wysoką dokładność poprzez ciągłe przewidywanie tej dominującej klasy. Dokładność nie dostarcza informacji o tym, jak model radzi sobie z poszczególnymi klasami, co jest szczególnie ważne w przypadku problemów wieloklasowych (o więcej niż dwóch klasach) lub gdy niektóre błędy są bardziej kosztowne niż inne (jak na przykład przy klasyfikacji chorób w medycynie). Dodatkowo należy pamiętać, że dokładność klasyfikacji jest niewrażliwa na rodzaje błędów i traktuje je wszystkie tak samo, nie rozróżniając pomiędzy błędami typu I (fałszywie pozytywnymi) i błędami typu II (fałszywie negatywnymi), co może nie pozwalać na wybranie odpowiedniego modelu w wielu zastosowaniach, gdzie różne typy błędów mają różne konsekwencje (jak między innymi wyżej wymieniona medycyna). Takie zerojedynkowe podejście do poprawności nie pozwala też uwzględnić pewności modelu co do podanej odpowiedzi.

Z tego względu różne dziedziny stosują różne podejścia. Współczynnik Giniego i test Kołmogorowa-Smirnowa są szeroko stosowane w branży oceny kredytowej. Czułość i swoistość są szeroko stosowane w epidemiologii i medycynie. Precyzja i czułość (recall) są szeroko stosowane w wyszukiwaniu informacji. (Hand 2012)

Mając te ograniczenia na uwadze, często zaleca się użycie dodatkowych metryk oceny modelu, takich jak wyżej wymienione precyzja, czułość czy specyficzność, czy też wartość F1, krzywe ROC, AUC lub inne, które mogą dostarczyć bardziej zróżnicowany obraz jakości modelu, szczególnie w bardziej złożonych przypadkach lub dla nierównomiernie rozłożonych zbiorów danych (Fawcett 2006) (Powers 2011) (Swets 1988).

Niemniej jednak, autorzy prac naukowych bardzo często poprzestają na podaniu tylko dokładności wypracowanych modeli, więc aby umożliwić porównania z innymi dostępnymi na rynku rozwiązaniami właśnie tą metrykę wybrano do porównania.

## KAN i MLP

Autorzy oryginalnego artykułu badawczego porównują komunikację z KANami do komunikacji z LLM, przy czym funkcje matematyczne, w ich porównaniu, zastępują zdania w języku naturalnym. Jako, że KAN dąży do reprezentacji danych treningowych przez równanie matematyczne, jeśli zostanie ono odpowiednio uproszczone ma ono szansę przyjąć formę zrozumiałą dla człowieka. (Liu 2024)

Podstawą teoretyczną dla sposobu działania KANów jest twierdzenie Kolmogorov-Arnolda, które mówi, że każdą wielowymiarową (o wielu zmiennych) ciągłą funkcję można przedstawić za pomocą skończonej liczby funkcji jednowymiarowych i dodawania. Mówiąc dokładniej każda funkcja operująca na n różnych parametrów może być przedstawiana jako 2 \* n + 1 ciągłych funkcji, z których każda operuje na sumie ciągłych funkcji operujących na pojedynczym parametrze.

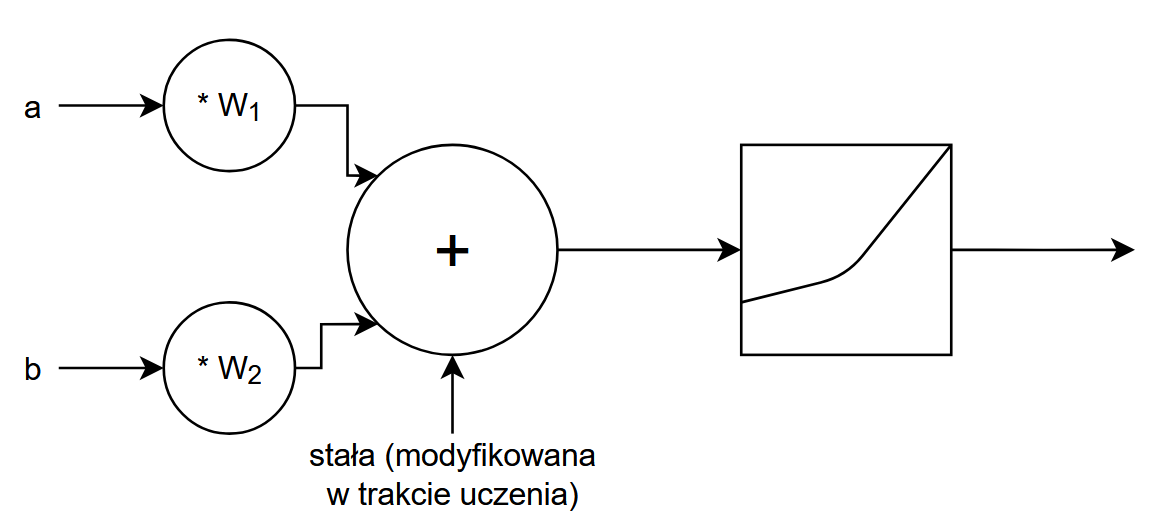
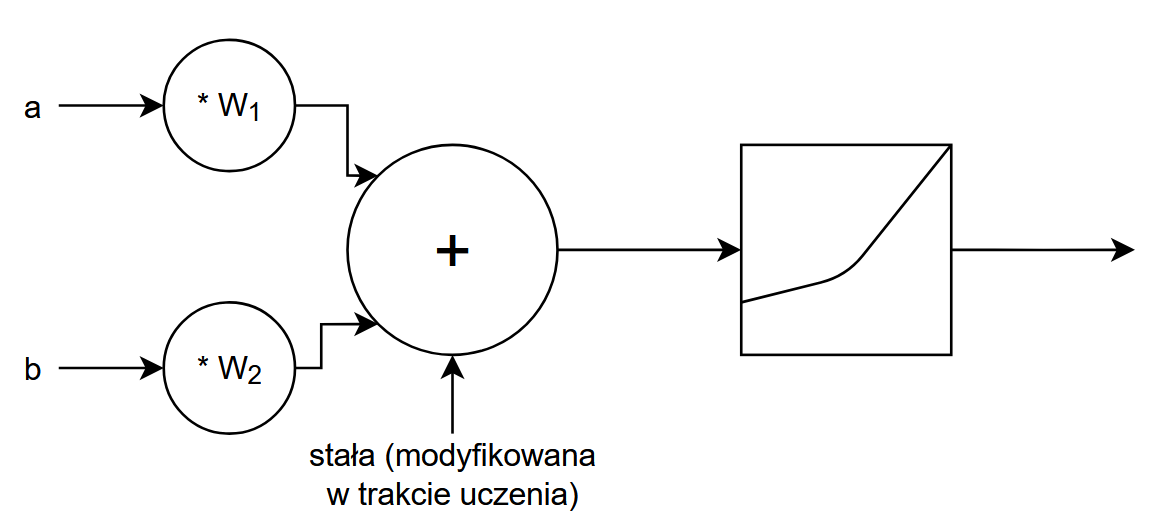
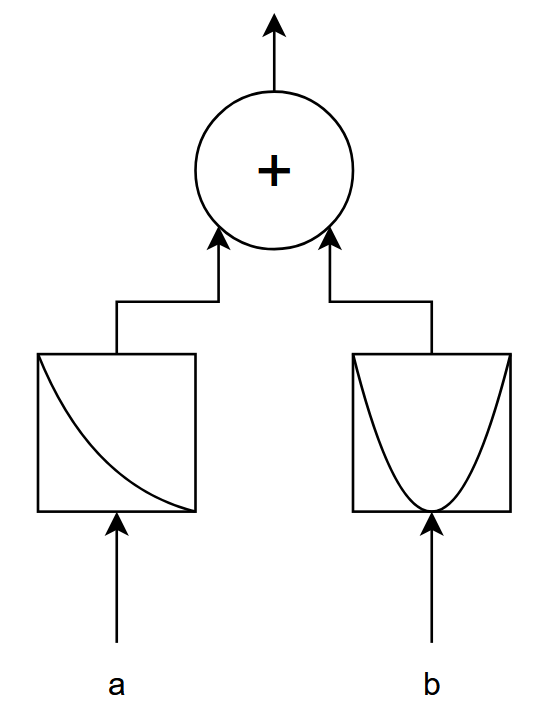
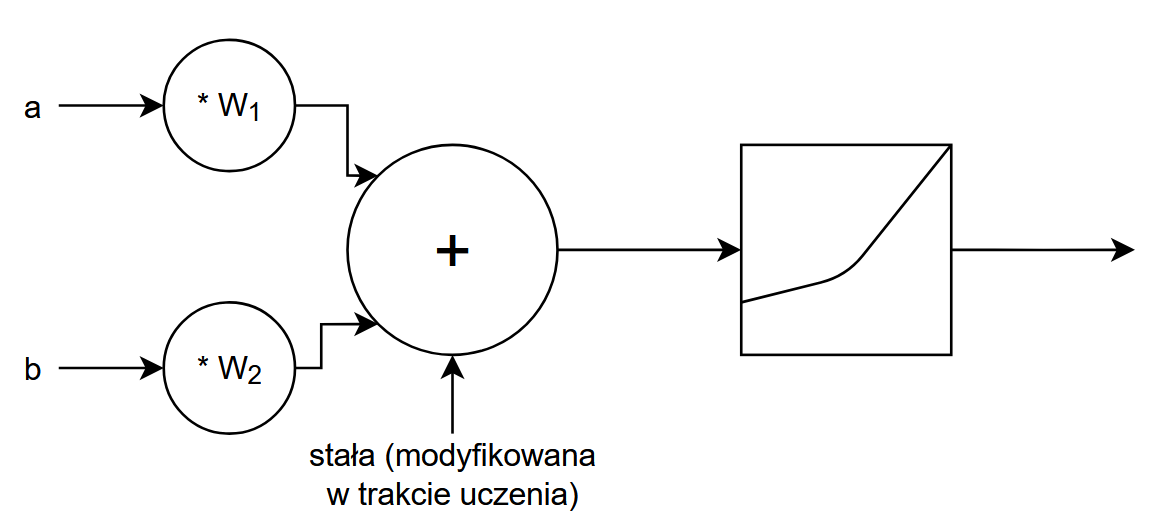
Dla przykładu funkcja operująca na zmiennych a i b może być przedstawiona jako:

Gdzie to funkcje operujące na sumie funkcji operujących na poszczególnych parametrach (np. ).

W uogólnionej wersji twierdzenie to przedstawiane jest jako (Kołmogorow 1957):  
Obraz zawierający tekst, Czcionka, biały, pismo odręczne

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Równanie . Twierdzenie Kołmogorowa-Arnolda



Rysunek . Porównanie działania KAN (lewo) i MLP (prawo) gdzie W to wagi

Pojedynczy neuron MLP (multilayer perceptron) działa mnożąc wszystkie pojedyncze dane wejściowe przez ich wagi. Każdy neuron połączony jest z każdym neuronem z warstwy sąsiedniej. Po przemnożeniu wszystkie wyniki są sumowane razem ze stałą wartością progową (bias) i ta wartość jest następnie używana jako wejście do funkcji progowej. Funkcja ta, nazywana również funkcją aktywacji, jest liniowa. Z tego względu, mimo iż MLP składa się z warstwy wejścia, dowolnej ilości warstw ukrytych oraz warstwy wyjściowej, dowiedziono, że model o dowolnej ilości warstw ukrytych można uprościć do jednej warstwy ukrytej, jeśli warstwa ta będzie odpowiednio szeroka (Kurt, Maxwell i Halbert 1989).

Przy trenowaniu MLP zmieniane są wagi dla poszczególnych wejść podczas gdy funkcja aktywacji pozostaje stała (jest ustawiana przez programistę przed rozpoczęciem uczenia). Z kolei w KANach każde z wejść trafia bezpośrednio do funkcji (można powiedzieć, że wszystkie wagi są równe 1). Wszystkie wyniki są do siebie dodawane. W trakcie treningu w KANach zmieniane są funkcje aktywacji (Serano 2024).

W 1989 to twierdzenie zostało uznane za nieistotne w kontekście sieci uczących się przez Federico Girosi i Tomaso Poggio w 1989 (Girosi 1989). Dowiedli oni, że twierdzenie nie jest zawsze prawdziwe, jeśli funkcja wewnętrzna (Równanie 2.1 reprezentuje ją jako hpq) musi być gładka co pozwala na generalizację oraz jest jednym z koniecznych warunków do wykorzystania metody najszybszego spadku (funkcja musi być różniczkowalna) (Tomczyk 2024). Mimo tych uwag, Ziming Liu i współbadacze postanowili posłużyć się często używaną w uczeniu głębokim zależnością, że więcej warstw pozwala na bardziej elastyczne dopasowania, aby uzyskać niedokładne odwzorowania zależności w świecie rzeczywistym (akceptują nie stuprocentowe dopasowanie). (Liu 2024)

# Istniejące rozwiązania

[wstęp]

## Drzewo decyzyjne

Jednym z najprostszych do interpretacji modeli jest drzewo decyzyjne. Jak nazwa wskazuje pomagają podjąć decyzje używając serii prostych pytań, która pozwala brać pod uwagę kontekst decyzji. Główną metodą ich kreacji jest przygotowanie określonych scenariuszy: rozważane są różne możliwe zależności między osobą decydującą, środowiskiem i podejmowanymi działaniami tak by znać konsekwencje każdej możliwej decyzji. W rzeczywistości przewidzenie wszystkiego co może się wydarzyć często nie jest możliwe co pozostawia drzewa decyzyjne jako narzędzie wspomagające rozważania i symulacje wydarzeń zamiast narzędzi do podejmowania decyzji. (Von Winterfeldt i Edwards 1986)  
W aspekcie uczenia maszynowego drzewa decyzyjne mogą być używane do zadań klasyfikacji i regresji. Algorytmy budujące drzewa przeważnie bazują na algorytmach pozwalających na podzielenie przestrzeni opisywanej przez posiadane cechy (Hastie, Tibshirani i Friedman 2001).

Obraz zawierający diagram, szkic, tekst, rysowanie

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek . Podział dwuwymiarowej przestrzeni cech wykonany przez dzielenie binarne algorytmem CART Źródło: (Hastie, Tibshirani i Friedman 2001)

W zależności od parametrów obranych przez algorytm przy tworzeniu drzewa przestrzeń danych może być dzielona na dowolnie małe obszary. W nomenklaturze drzew decyzyjnych każda linia rozgraniczająca obszary jest węzłem i odpowiada weryfikacji jaką należy przeprowadzić na podanych danych a niepodzielony obszar, zwany liściem, reprezentuje wynik drzewa.

Mimo niezaprzeczalnych zalet drzew decyzyjnych należy pamiętać, że drzewa decyzyjne mogą też być obarczone znacznymi problemami. Jednym z ważniejszych jest stronniczość algorytmów wybierających atrybuty dające największy przyrost informacji tym atrybutom kategorialnym, które mają więcej kategorii (Deng, Runger i Tuv 2011).

## K najbliższych sąsiadów

Algorytm K najbliższych sąsiadów (KNN od K nearest neighbours) to algorytm pozwalający na klasyfikację bez konieczności uczenia. Na bazie zbioru treningowego umieszcza każdy z dostępnych punktów danych w przestrzeni, gdzie każdy wymiar odpowiada jednej cesze a każdy nowy przypadek klasyfikuje określając najczęściej występującą klasę w K (K to liczba naturalna; jednymi z najczęściej używanych są 3 i 5), najbliższych do nowego, punktów (Fix i Hodges 1951) (Cover i Hart 1967).

Ten sposób klasyfikacji jest łatwy w zrozumieniu i wytłumaczeniu, ale nie jest bez swoich wad. Wraz z rosnącą ilością wymiarów dostępne przykłady rozłożone będą coraz bardziej sporadycznie przez co wybrane K może znacznie wpłynąć na ostateczną klasyfikację danych co ilustruje Rysunek 3.2.

Obraz zawierający zrzut ekranu, Wielobarwność

Zawartość wygenerowana przez AI może być niepoprawna.

Rysunek . Nowy punkt danych reprezentowany zieloną kropką będzie sklasyfikowany jako czerwony trójkąt przy K = 3 i niebieski kwadrat przy K = 5. Źródło: Antti Ajanki https://en.wikipedia.org/wiki/File:KnnClassification.svg

Znaczne problemy może też prezentować nierówność rozłożenia klas: jeśli przykładów jednej z klas będzie znacznie więcej to więcej tych punktów może być liczone przy większym analizowanym obszarze (większych K). Klasyczny algorytm KNN patrzy po równo na zmianę w każdym wymiarze, jeśli w badanych danych jedna z cech jest lepszym predykatorem klasy może to być uwzględnione tylko zmieniając algorytm na np. ważony KNN, gdzie z kolei należy wytrenować lub ręcznie ustawić wagi.

Przy przygotowywaniu danych pod KNN należy również zadbać o ich normalizację tak by zmiany w dużych liczbowo wartościach jak np. pensja miały podobny efekt co zmiany w małych liczbowo wartościach jak np. wiek.

# Zbiory danych używane w pracy

[wstęp]

## Breast Cancer Wisconsin (Diagnostic) (Wolberg 1995)

Dane dostępne na stronie repozytorium Uniwersytetu Kalifornijskiego, Irvine (link: <https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic>) oraz poprzez sklearn.datasets.load\_breast\_cancer.

### Opis danych

Dane zostały pozyskane poprzez analizę zdigitalizowanych obrazów mas piersi uzyskanych za pomocą aspiracyjnej cienkoigłowej (fine needle aspirate (FNA)). Przykład masy nowotworowej pobrany tą metodą jest zobrazowany na Rys. 4.1.

Wartości numeryczne w zestawie danych opisują jądra komórkowe w trzech wymiarach stosując algorytm opisany w "Robust Linear Programming Discrimination of Two Linearly Inseparable Sets" autorstwa K. P. Bennett i O. L. Mangasarian z Optimization Methods and Software 1 z 1992 na stronach 23-34.

Obraz zawierający plaster miodu, czarne i białe

Zawartość wygenerowana przez AI może być niepoprawna.

Rys. . Komórki nowotworu niezłośliwego pobrane za pomocą aspiracji cienkoigłowej; żródło: M. W. Teague, W. H. Wolberg, W. N. Street, O. L. Mangasarian, S. Labremont, and D. L. Page.; Cancer Cytopathology 81: 129-135, 1997.

Dane zostały zebrane przez doktora Wolberg’a w Szpitalu Uniwersyteckim w Madison w Wisconsin i udostępnione przez Olvi Mangsarian’a 15 lipca 1992. Dane zbierane były od stycznia 1989 do listopada 1991. (Rosly 2018)

### Opis wartości

Wartości w kolumnie Diagnosis: B (Benign: niezłośliwy), M (Malignant: złośliwy).

Tabela . Nazwy i typ danych kolumn dla danych Breast Cancer Wisconsin (Diagnostic) razem z informacją czy zawierają brakujące dane

|  |  |  |
| --- | --- | --- |
| Nazwa kolumny | Typ danych w kolumnie | Czy zawiera brakujące wartości? |
| ID | Kategorialna | Nie |
| Diagnosis | Kategorialna | Nie |
| radius1 | Liczbowa | Nie |
| texture1 | Liczbowa | Nie |
| perimeter1 | Liczbowa | Nie |
| area1 | Liczbowa | Nie |
| smoothness1 | Liczbowa | Nie |
| compactness1 | Liczbowa | Nie |
| concavity1 | Liczbowa | Nie |
| concave\_points1 | Liczbowa | Nie |
| symmetry1 | Liczbowa | Nie |
| fractal\_dimension1 | Liczbowa | Nie |
| radius2 | Liczbowa | Nie |
| texture2 | Liczbowa | Nie |
| perimeter2 | Liczbowa | Nie |
| area2 | Liczbowa | Nie |
| smoothness2 | Liczbowa | Nie |
| compactness2 | Liczbowa | Nie |
| concavity2 | Liczbowa | Nie |
| concave\_points2 | Liczbowa | Nie |
| symmetry2 | Liczbowa | Nie |
| fractal\_dimension2 | Liczbowa | Nie |
| radius3 | Liczbowa | Nie |
| texture3 | Liczbowa | Nie |
| perimeter3 | Liczbowa | Nie |
| area3 | Liczbowa | Nie |
| smoothness3 | Liczbowa | Nie |
| compactness3 | Liczbowa | Nie |
| concavity3 | Liczbowa | Nie |
| concave\_points3 | Liczbowa | Nie |
| symmetry3 | Liczbowa | Nie |
| fractal\_dimension3 | Liczbowa | Nie |

## Iris (Anderson 1935)

Dane dostępne na stronie repozytorium Uniwersytetu Kalifornijskiego, Irvine (link: <https://archive.ics.uci.edu/dataset/53/iris>) oraz poprzez sklearn.datasets.load\_iris.

### Opis danych

Zestaw danych powstał, gdy Edgar Anderson zebrał wymiary działki kielicha oraz płatków trzech gatunków irysów (Iris setosa, Iris virginica i Iris versicolor) w celu zmierzenia ich różnic morfologicznych. Został spopularyzowany w zadaniu klasyfikacji przez statystyka Ronalda Fisher’a (Fisher 1936).

### Opis wartości

Wartości w kolumnie class to 0, 1 i 2 co odpowiada Iris setosa, Iris versicolor i Iris virginica.

Tabela . Nazwy i typ danych kolumn dla danych Iris razem z informacją czy zawierają brakujące dane

|  |  |  |
| --- | --- | --- |
| Nazwa kolumny | Typ danych w kolumnie | Czy zawiera brakujące wartości? |
| sepal length | Liczbowa | Nie |
| sepal width | Liczbowa | Nie |
| petal length | Liczbowa | Nie |
| petal width | Liczbowa | Nie |
| class | Kategorialna | Nie |

## Wine (Stefan Aeberhard 1994)

Dane dostępne na stronie repozytorium Uniwersytetu Kalifornijskiego, Irvine (link: https://archive.ics.uci.edu/dataset/109/wine) oraz poprzez sklearn.datasets.load\_wine.

### Opis danych

Zestaw danych jest wynikiem analizy chemicznej win wyprodukowanych z winogron z trzech różnych odmian.

### Opis wartości

Wartości w kolumnie class to 1, 2 i 3 co odpowiada trzem odmianom winogron.

Tabela . Nazwy i typ danych kolumn dla danych Wine razem z informacją czy zawierają brakujące dane

|  |  |  |
| --- | --- | --- |
| **Nazwa kolumny** | **Typ danych w kolumnie** | **Czy zawiera brakujące wartości?** |
| **class** | Kategorialna | Nie |
| **alcohol** | Liczbowa | Nie |
| **malicacid** | Liczbowa | Nie |
| **ash** | Liczbowa | Nie |
| **alcalinity\_of\_ash** | Liczbowa | Nie |
| **magnesium** | Liczbowa | Nie |
| **total\_phenols** | Liczbowa | Nie |
| **flavanoids** | Liczbowa | Nie |
| **nonflavanoid\_phenols** | Liczbowa | Nie |
| **proanthocyanins** | Liczbowa | Nie |

# Opis eksperymentów

Aby umożliwić porównanie metod tworzenia klasyfikatorów opracowany został program zbierający dane o dokładności modeli utworzonych poszczególnymi metodami. Do obliczenia ilości parametrów został użyty wzór przedstawiony w „KAN or MLP: A Fairer Comparison” (Runpeng Yu 2024)

Obraz zawierający tekst, Czcionka, zrzut ekranu, algebra

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek . Fragment oryginalnego tekstu omawianego artykułu zawierający wzory pozwalające na obliczenie ilości parametrów KANów i MLP

Program pozwalał na ustawienie ilości epok przez jakie model będzie trenowany oraz cierpliwości, która reprezentowała ilość kolejnych epok przez które jeśli dokładność na zbiorze walidacyjnym nie wzrosła o ustawiane kolejnym parametrem minimum uczenie było przerywane (wczesne zatrzymywanie). Podczas uczenia pięciokrotnie tworzono każdy z modeli (by móc obliczyć średnią pozwalającą na ograniczenie wpływu oryginalnych wartości parametrów). Każdy z modeli osiągał różną ilość parametrów poprzez zmianę właściwości KANa: Ilość podziałów domeny danych (reprezentowany jako G we wzorze na ilość parametrów (Rysunek 5.1) kontrolowana argumentem grid) oraz stopień splinów (Rysunek 5.1 oznacza go jako K a w implementacji oznaczony jest jako k) jak i szerokość i ilość warstw między wejściową i wyjściową (przekazywane jako lista argumentem width).

Zbierana była dokładność na zbiorze testowym dla KANa o określonych parametrach, MLP o 2 warstwach ukrytych, których szerokość nie dopuszcza przekroczenia łącznej ilości parametrów KANa oraz MLP posiadającego warstwy ukryte szersze od MLP opisanego powyżej o 1. Oryginalny zestaw danych dzielony był na zbiór testowy (25% danych) walidacyjny (15% danych) oraz treningowy (pozostałe 60% danych).

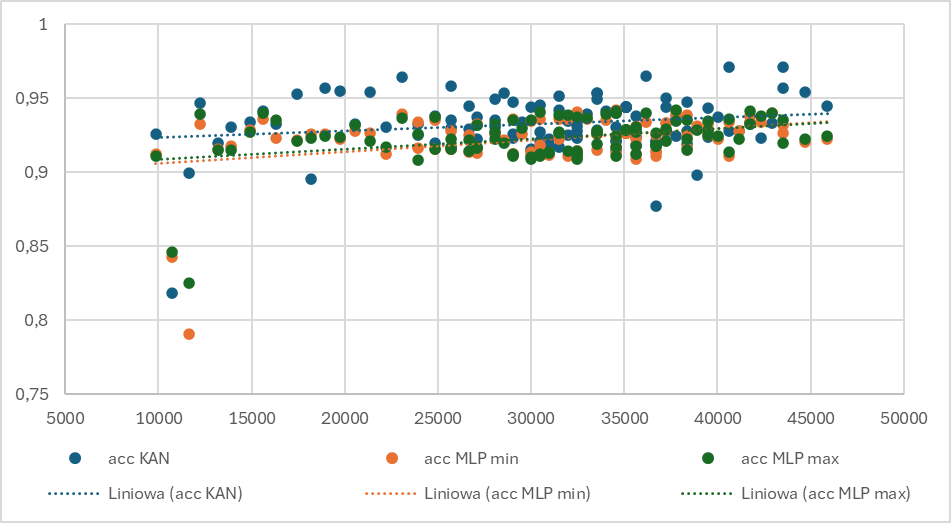
# Wyniki

[wstęp]

## Hipoteza 1: Przy takiej samej ilości parametrów KANy osiągną wyższą dokładność niż MLP

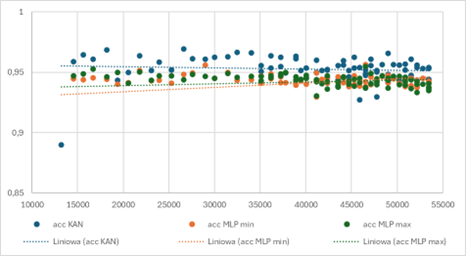
Zebrano dane dla zbiorów danych opisanych w poprzednim rozdziale dla 150 epok, 5 epok cierpliwości i 0,00001 (0,001%) minimalnej zmianie dokładności.

Wynikiem powyższych eksperymentów są poniższe grafy. Na osi Y można odczytać dokładność (maksymalna dokładność to 1) a na osi X ilość parametrów KANa. Przedstawione w podpisach grid i k odnoszą się do argumentów użytych przy tworzeniu modeli KAN. Powtarzające się podpisy danych odpowiadają:  
acc KAN – dokładność KAN na zbiorze testowym  
acc MLP min – dokładność MLP o szerokości warstw niepozwalającej przekroczyć ilości parametrów KAN na zbiorze testowym  
acc MLP max - dokładność MLP o szerokości warstw pozwalającej na minimalnie przekroczenie ilości parametrów KAN na zbiorze testowym

Obraz zawierający tekst, zrzut ekranu, linia, Wykres

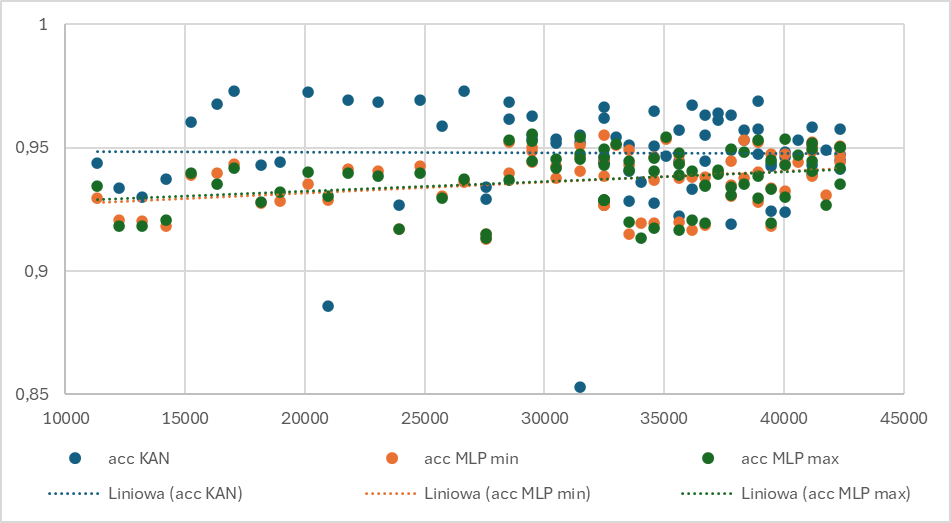
Zawartość wygenerowana przez AI może być niepoprawna.

Rysunek 6.1 Breast Cancer Wisconsin grid = 3, k = 2 (lewo) Breast Cancer Wisconsin grid = 3, k = 3 (prawo)

Obraz zawierający tekst, zrzut ekranu, linia, Wykres

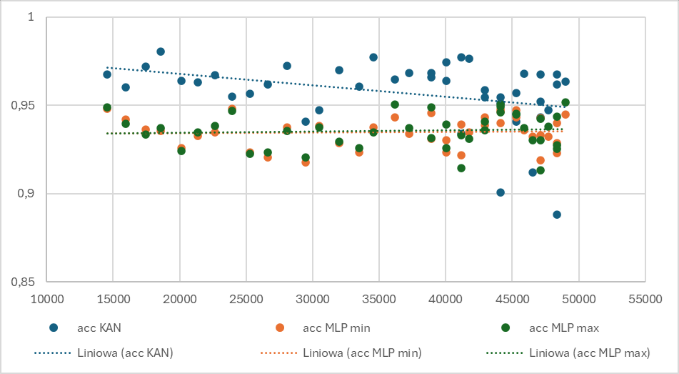
Zawartość wygenerowana przez AI może być niepoprawna.

Rysunek 6.2 Breast Cancer Wisconsin grid = 3, k = 4 (lewo) Breast Cancer Wisconsin grid = 3, k = 5 (prawo)

Obraz zawierający zrzut ekranu, linia, Wykres, tekst

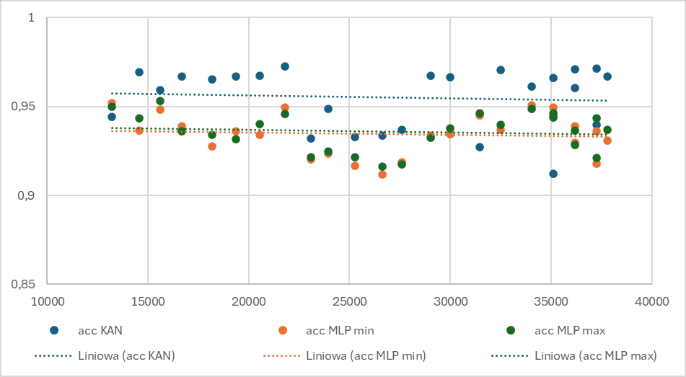
Zawartość wygenerowana przez AI może być niepoprawna.

Rysunek 6.3 Breast Cancer Wisconsin grid = 5, k = 2 (lewo) Breast Cancer Wisconsin grid = 5, k = 3 (prawo)

Obraz zawierający zrzut ekranu, tekst, linia, Wykres

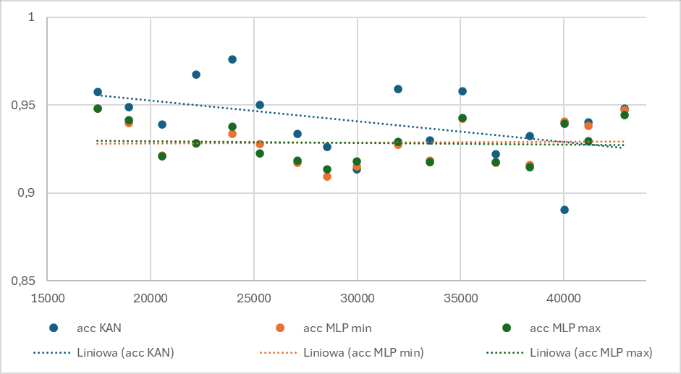
Zawartość wygenerowana przez AI może być niepoprawna.

Rysunek 6.4 Breast Cancer Wisconsin grid = 5, k = 4 (lewo) Breast Cancer Wisconsin grid = 5, k = 5 (prawo)

Obraz zawierający zrzut ekranu, tekst, linia, diagram

Zawartość wygenerowana przez AI może być niepoprawna.

Rysunek 6.5 Breast Cancer Wisconsin grid = 9, k = 2 (lewo) Breast Cancer Wisconsin grid = 9, k = 3 (prawo)

Obraz zawierający tekst, zrzut ekranu, diagram, linia

Zawartość wygenerowana przez AI może być niepoprawna.

Rysunek 6.6 Breast Cancer Wisconsin grid = 9, k = 4 (lewo) Breast Cancer Wisconsin grid = 9, k = 5 (prawo)

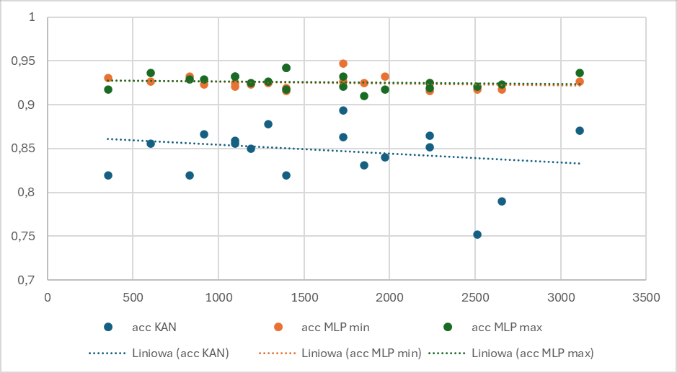
Średnio, dla zbioru danych Breast Cancer Wisconsin KANy osiągały lepsze wyniki niż MLP dla mniejszej ilości parametrów, ale nie zawsze dla większej ilości parametrów.

Obraz zawierający tekst, zrzut ekranu, linia, Wykres

Zawartość wygenerowana przez AI może być niepoprawna.Obraz zawierający tekst, zrzut ekranu, numer, linia

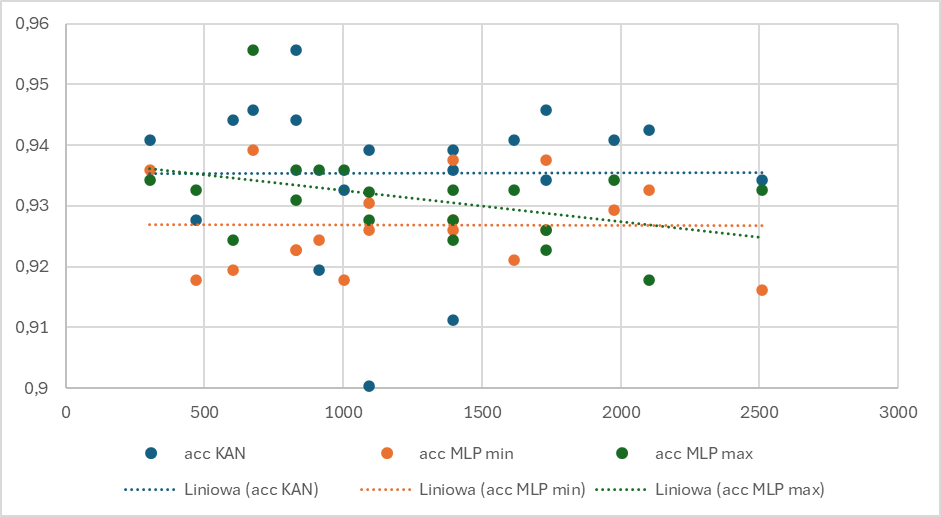
Zawartość wygenerowana przez AI może być niepoprawna.

Rysunek 6.7 Iris grid = 3, k = 2 (lewo) Iris grid = 3, k = 3 (prawo)

Obraz zawierający tekst, zrzut ekranu, linia, Wykres

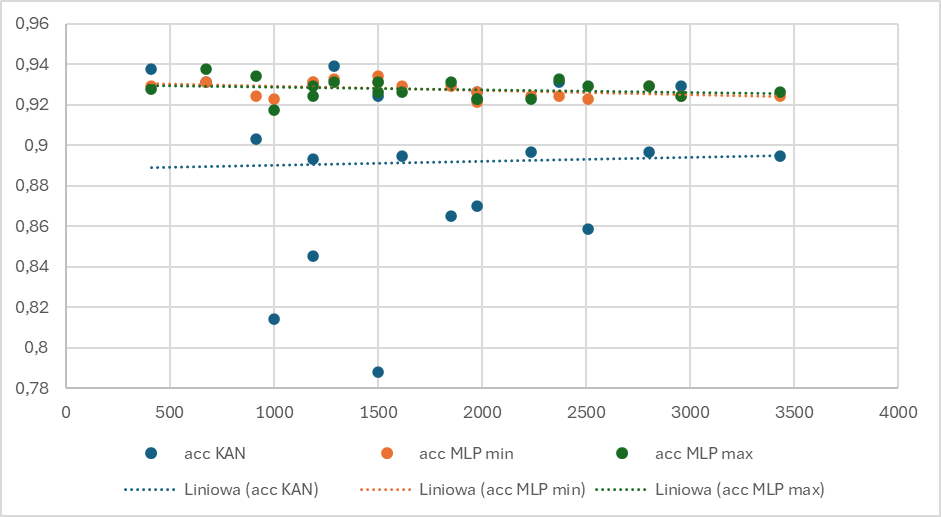
Zawartość wygenerowana przez AI może być niepoprawna.

Rysunek 6.8 Iris grid = 3, k = 4 (lewo) Iris grid = 3, k = 5 (prawo)

Obraz zawierający tekst, zrzut ekranu, diagram, linia

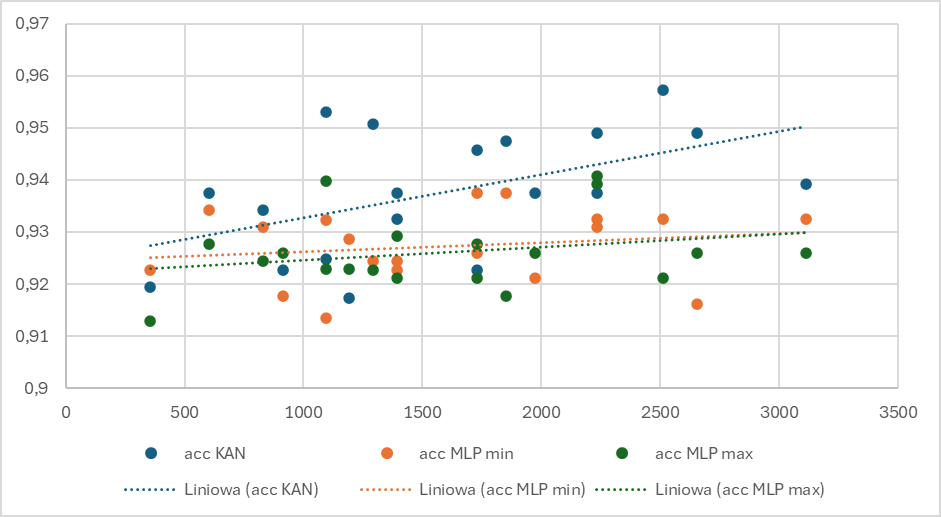
Zawartość wygenerowana przez AI może być niepoprawna.

Rysunek 6.9 Iris grid = 5, k = 2 (lewo) Iris grid = 5, k = 3 (prawo)

Obraz zawierający tekst, zrzut ekranu, linia, Wykres

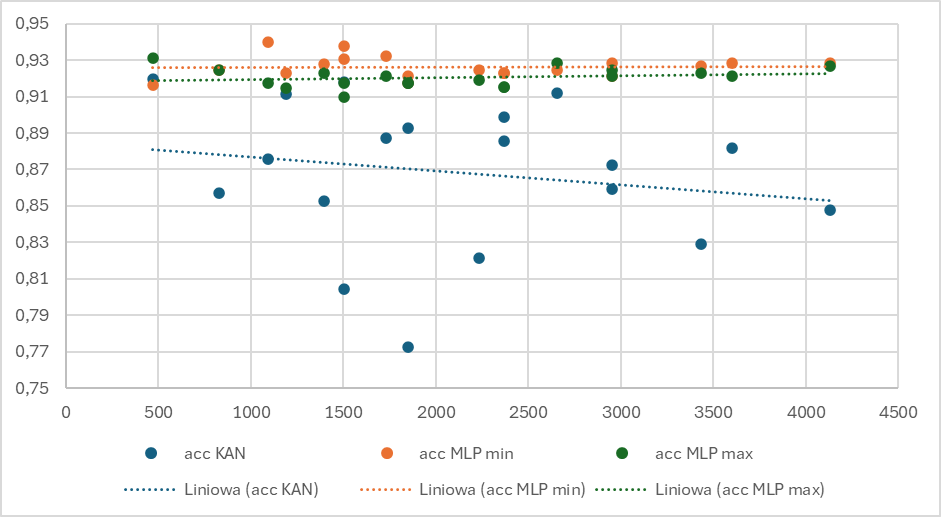
Zawartość wygenerowana przez AI może być niepoprawna.

Rysunek 6.10 Iris grid = 5, k = 4 (lewo) Iris grid = 5, k = 5 (prawo)

Obraz zawierający tekst, zrzut ekranu, linia, Wykres

Zawartość wygenerowana przez AI może być niepoprawna.

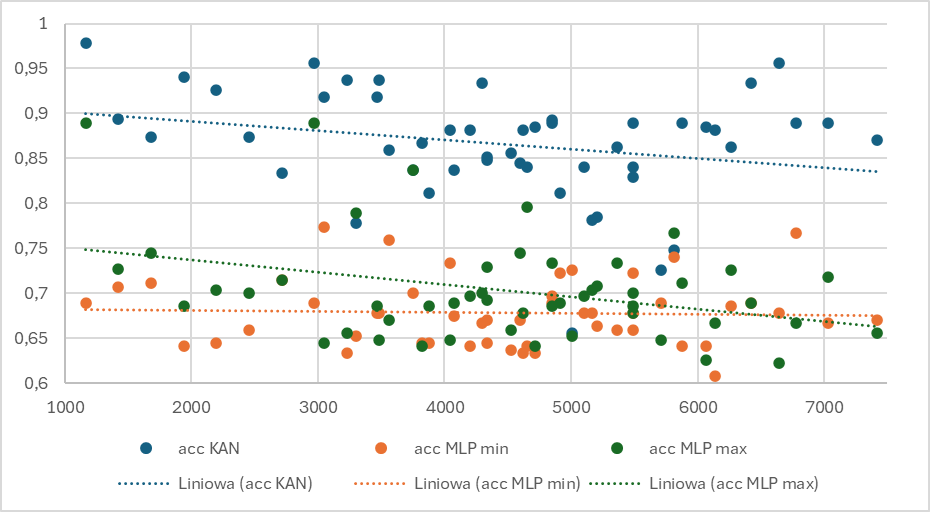
Rysunek 6.11 Iris grid = 9, k = 2 (lewo) Iris grid = 9, k = 3 (prawo)

Obraz zawierający tekst, zrzut ekranu, linia, Wykres

Zawartość wygenerowana przez AI może być niepoprawna.

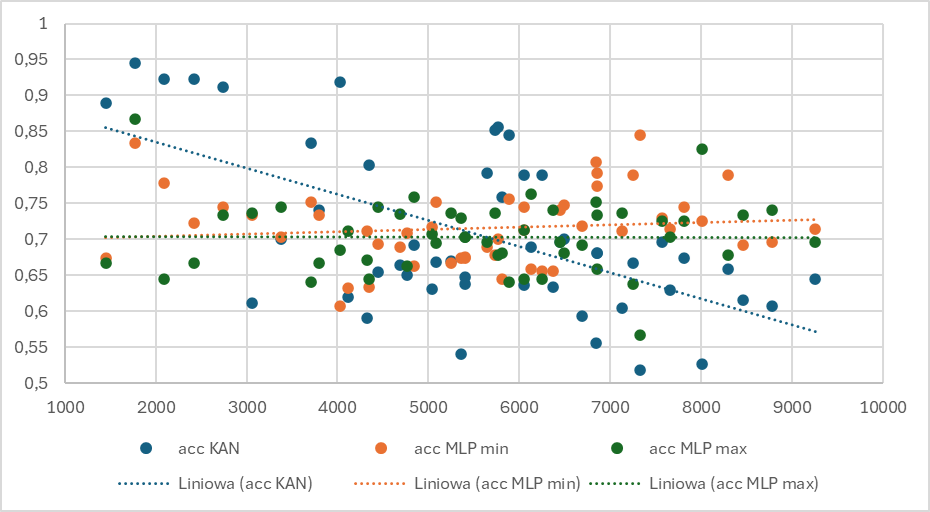
Rysunek 6.12 Iris grid = 9, k = 4 (lewo) Iris grid = 9, k = 5 (prawo)

Dla zbioru danych Iris dla k 4 i 5 KANy mają wyniki niższe niż MLP. Dotatkowo tendencja do gorszego wyniku KANów o większej ilości parametrów zauważalna na zbiorze Breast Cancer Wisconsin widoczna jest również na zbiorze Iris.

Obraz zawierający zrzut ekranu, Wykres, linia, diagram

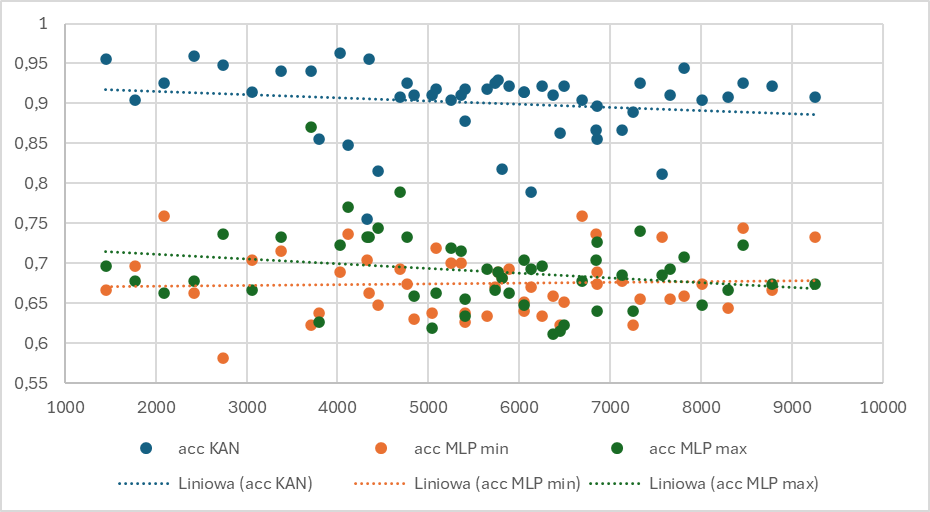
Zawartość wygenerowana przez AI może być niepoprawna.

Rysunek . Wine grid = 3, k = 2 (lewo) Wine grid = 3, k = 3 (prawo)

Obraz zawierający zrzut ekranu, tekst, diagram, Wykres

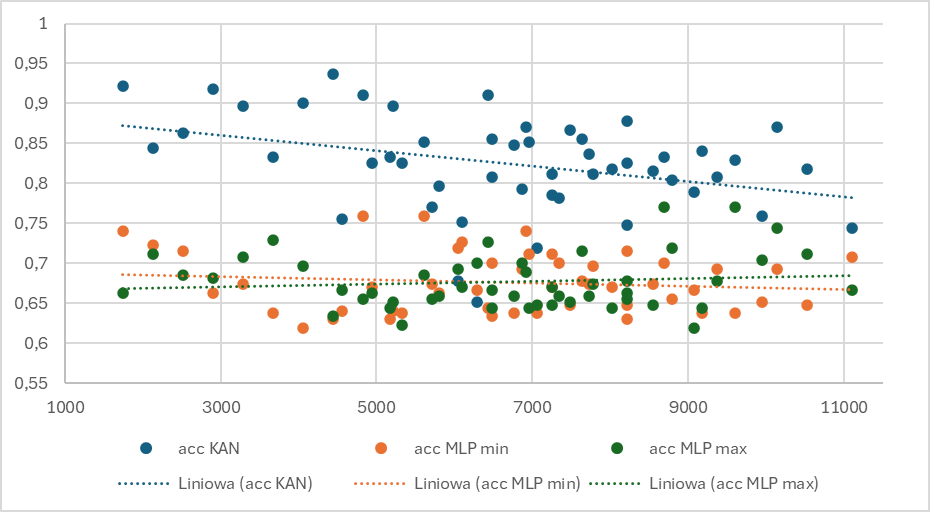
Zawartość wygenerowana przez AI może być niepoprawna.

Rysunek . Wine grid = 3, k = 4 (lewo) Wine grid = 3, k = 5 (prawo)

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

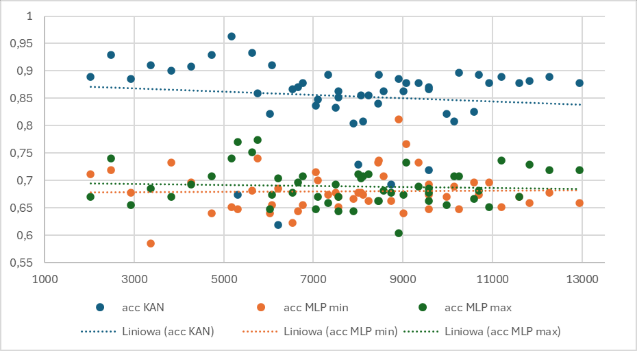
Zawartość wygenerowana przez AI może być niepoprawna.

Rysunek . Wine grid = 5, k = 2 (lewo) Wine grid = 5, k = 3 (prawo)

Obraz zawierający zrzut ekranu, tekst, Wykres, linia

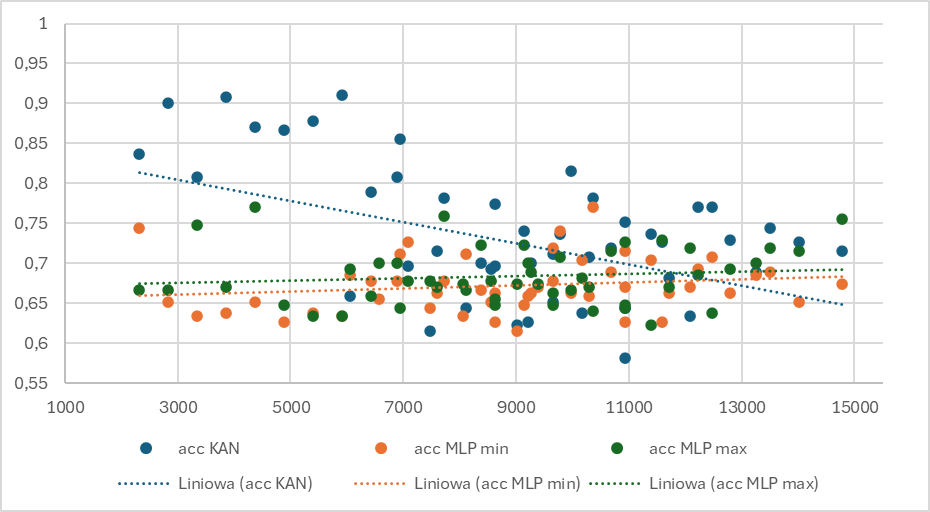
Zawartość wygenerowana przez AI może być niepoprawna.

Rysunek . Wine grid = 5, k = 4 (lewo) Wine grid = 5, k = 5 (prawo)

Obraz zawierający zrzut ekranu, tekst, Wykres, diagram

Zawartość wygenerowana przez AI może być niepoprawna.

Rysunek . Wine grid = 9, k = 2 (lewo) Wine grid = 9, k = 3 (prawo)

Obraz zawierający zrzut ekranu, tekst, Wykres, linia

Zawartość wygenerowana przez AI może być niepoprawna.

Rysunek . Wine grid = 9, k = 4 (lewo) Wine grid = 9, k = 5 (prawo)

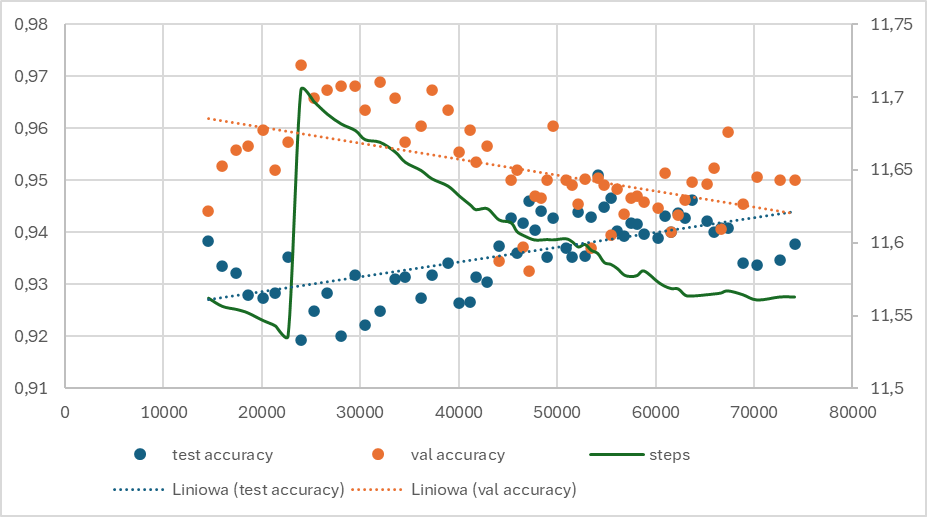
Podobnie jak dla Breast Cancer Wisconsin KANy z większą ilością parametrów radzą sobie gorzej niż te z mniejszą ilością parametrów. Szczególnie widoczne dla k równego 5 szersze i głębsze modele radzą sobie gorzej niż MLP.

Dla sieci z większą ilością parametrów (szersze warstwy i głębsze modele) KANy radzą sobie gorzej niż MLP.

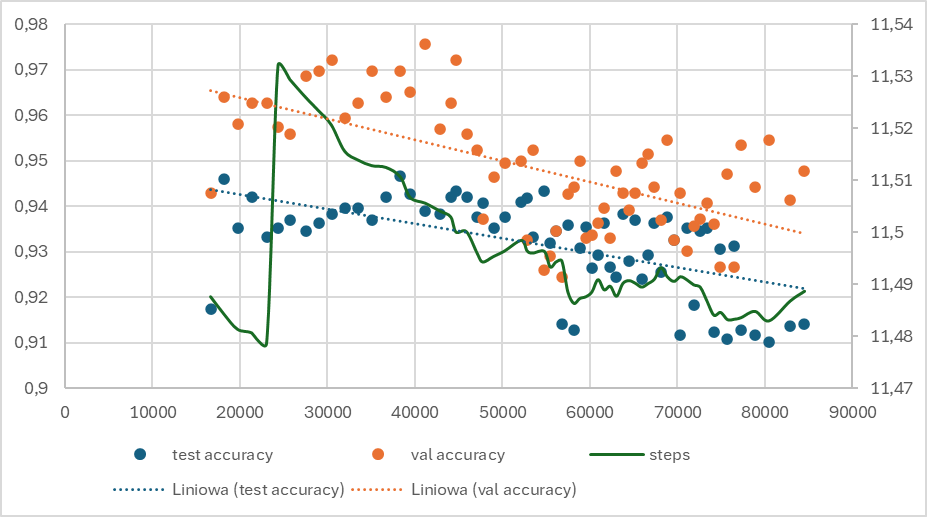
Aby zweryfikować możliwe przyczyny tego zjawiska utworzono kolejną hipotezę:

## Hipoteza 2: Bardziej elastyczne KANy nie osiągają akceptowalnej dokładności ze względu na zbyt małą ilość epok

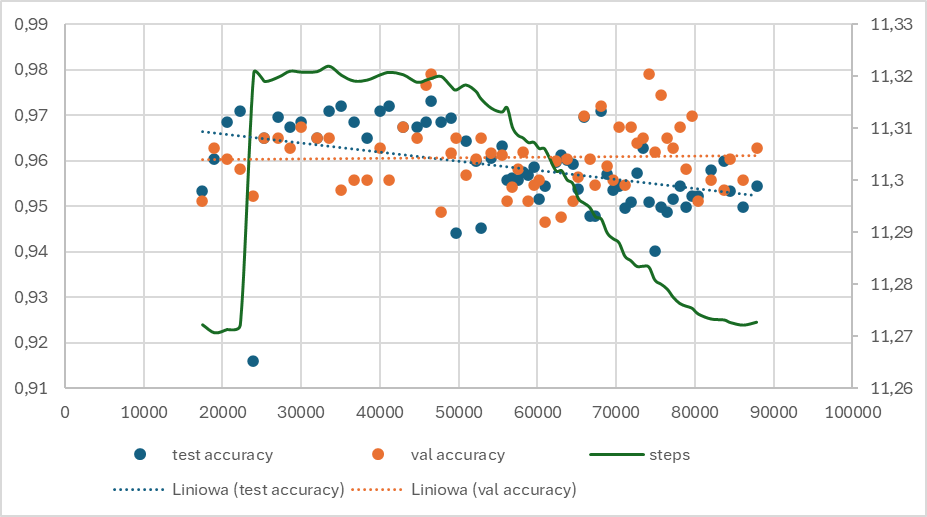
By ją zweryfikować porównano dokładność KANów na zbiorze walidacyjnym w momencie wykonania wczesnego zatrzymania z otrzymaną dokładnością na zbiorze testowym i uzyskano poniższe wyniki. Na wykresach zieloną linią (steps) zilustrowano średnią ilość epok jaką miał do dyspozycji każdy z trenowanych modeli przed przerwaniem uczenia (skala znajduje się po prawej) podczas gdy punkty oznaczają dokładność na zbiorze testowym (niebieskie – test accuracy) i zbiorze walidacyjnym (żółte – val accuracy). Wszystkie dane dotyczą KANów.



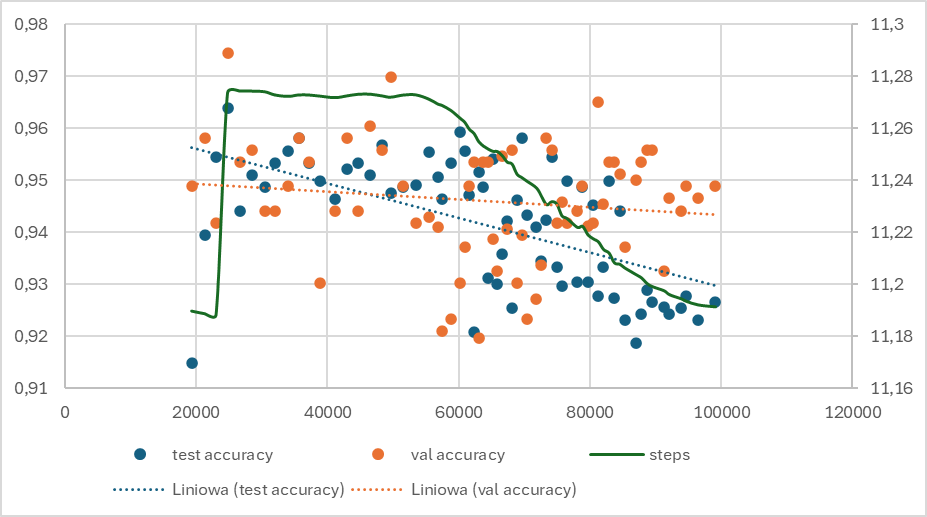
Rysunek . Cancer grid = 5, k = 4



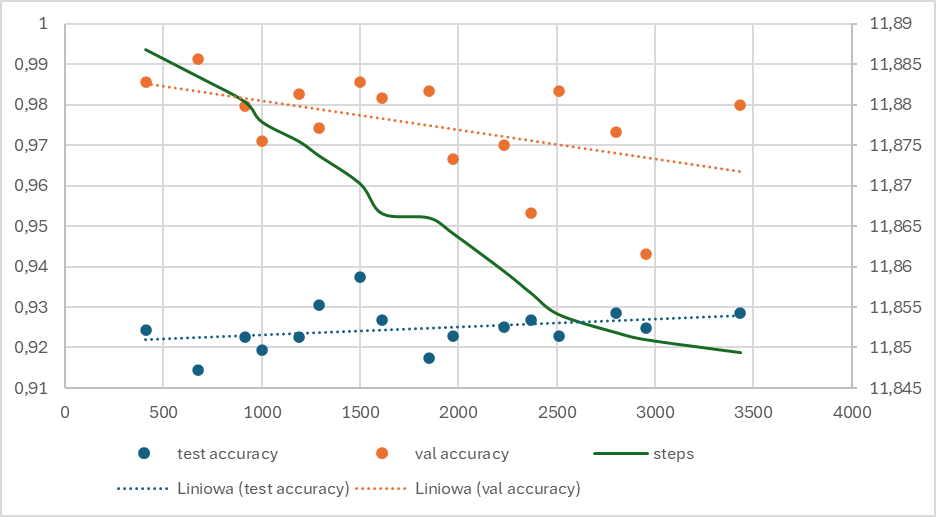
Rysunek . Cancer grid = 5, k = 5



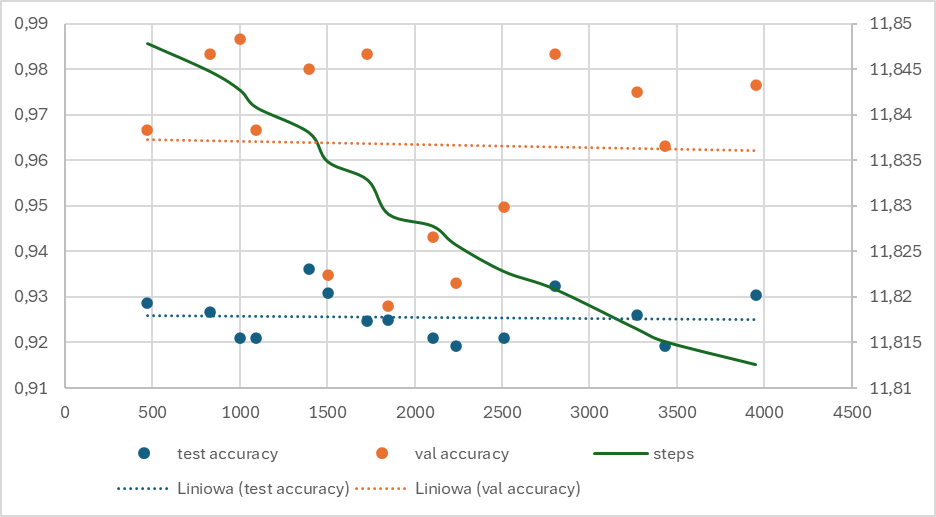
Rysunek . Cancer grid = 9, k = 4



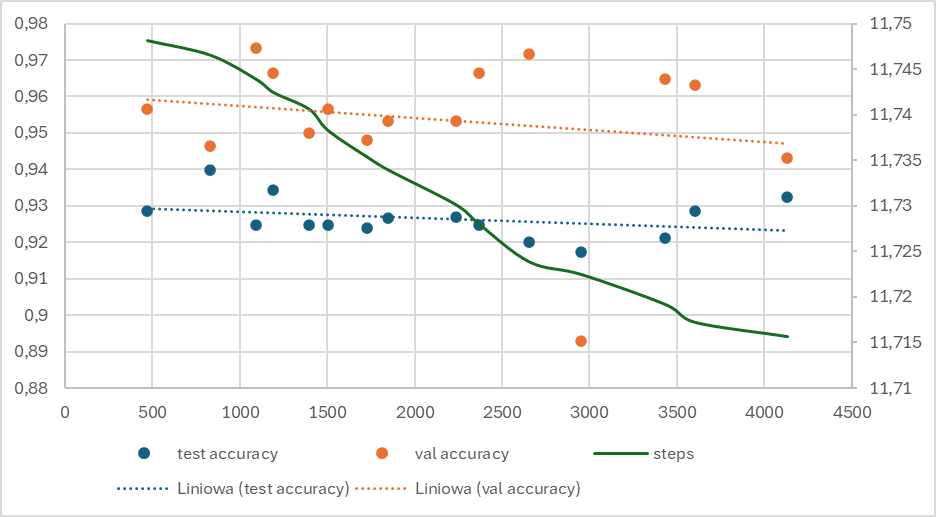
Rysunek . Cancer grid = 9, k = 5



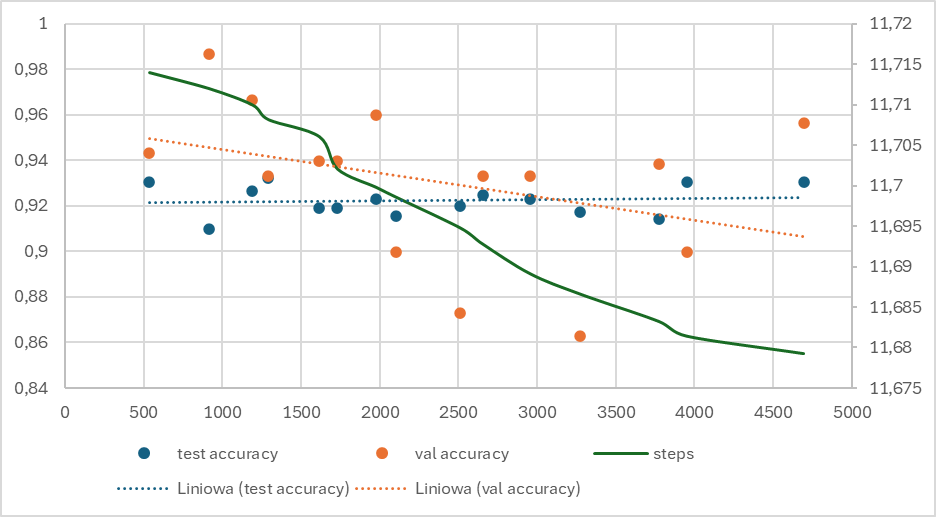
Rysunek . Iris grid = 5, k = 4



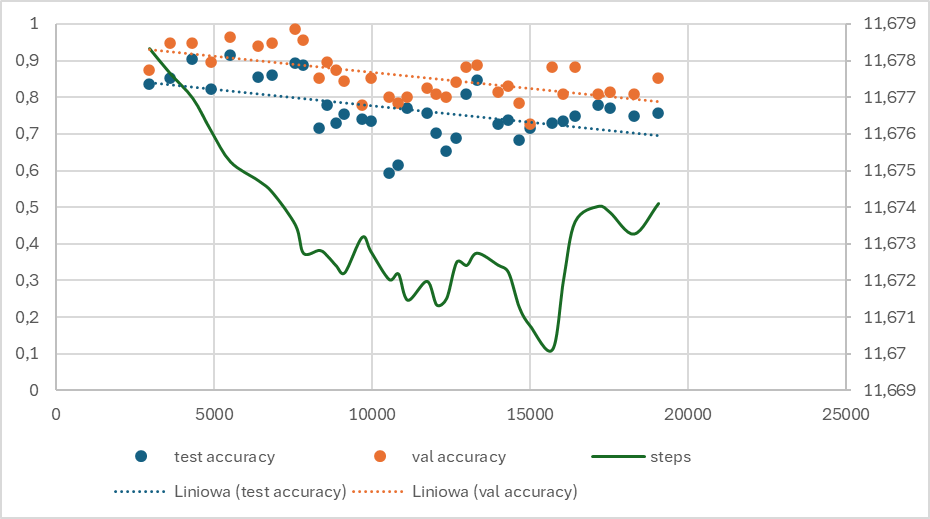
Rysunek . Iris grid = 5, k = 5



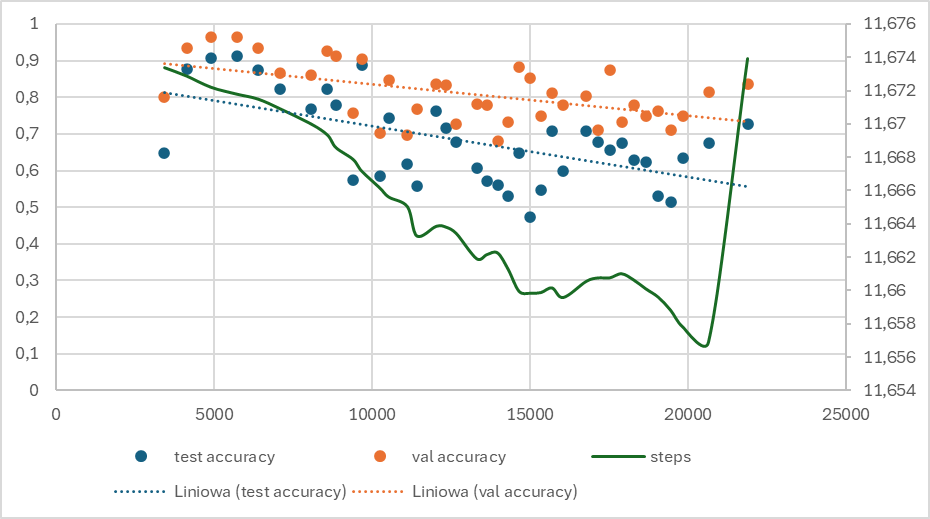
Rysunek . Iris grid = 9, k = 4



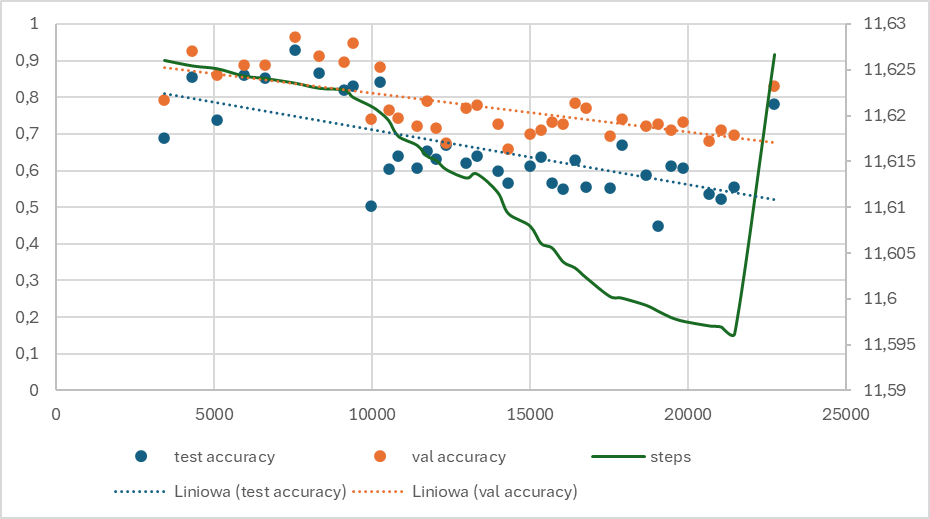
Rysunek . Iris grid = 9, k = 5



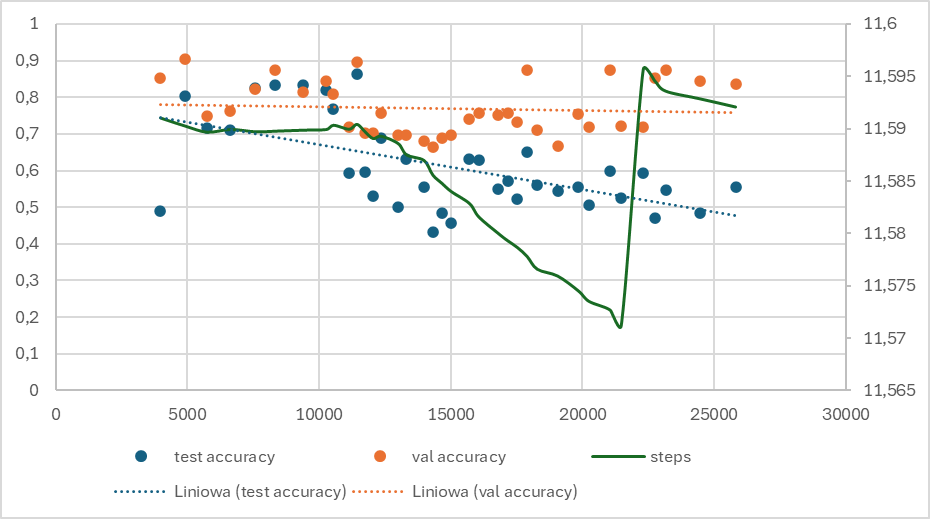
Rysunek . Wine grid = 5, k = 4



Rysunek . Wine grid = 5, k = 5



Rysunek . Wine grid = 9, k = 4



Rysunek . Wine grid = 9, k = 5

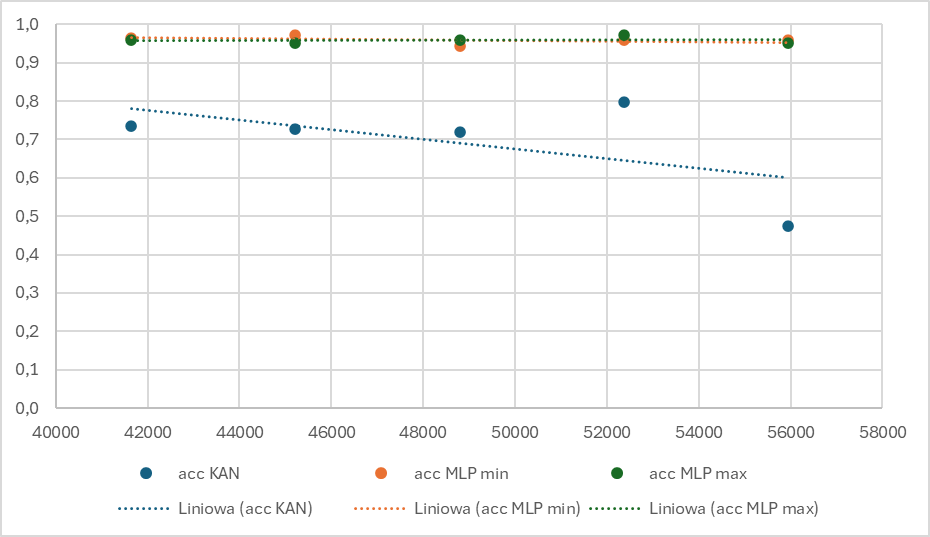
Dokładności na zbiorze testowym są zbliżone do dokładności na zbiorze walidacyjnym jednak dla wszystkich zbiorów danych oprócz najprostszego – Iris dokładność na obu zbiorach spada wraz ze wzrostem ilości parametrów. Wszystkie modele są uczone przez mniej niż 12 kroków co może sugerować, że wartość cierpliwości jest zbyt restrykcyjna co uniemożliwia lepsze nauczenie bardziej skomplikowanych modeli, jako że w teorii każdy skomplikowany model jest w stanie przewidywać przynajmniej równie dobrze co prostszy model.

## Hipoteza 3: Bardziej złożone KANy nie są w stanie w pełni nauczyć się przy zadanych parametrach.

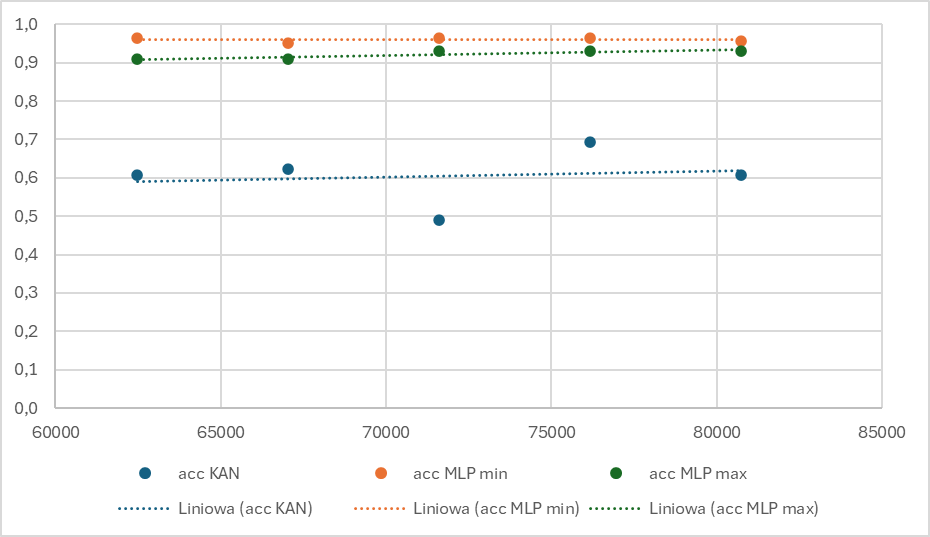
Aby zweryfikować tę hipotezę zmienione zostały parametry używane przy uczeniu.

Nowe wartości to cierpliwość równa 4000, tolerancja równa 0, ilość epok równa 100000. Dodatkowo testy dokładności KANów i MLP przeprowadzane były co 50 epok.

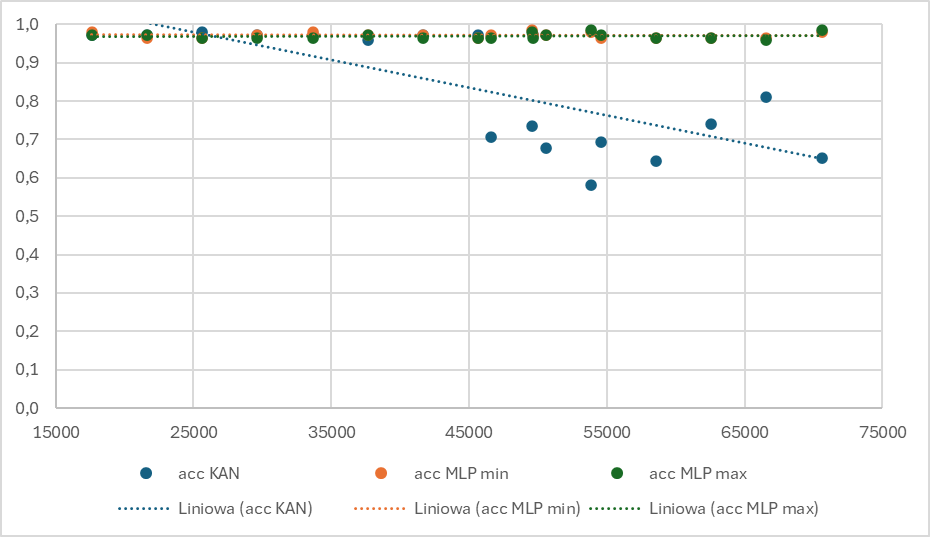
Wyniki tego eksperymentu zostały przedstawione na poniższych wykresach gdzie acc KAN to maksymalna dokładność na zbiorze testowym osiągnięta przez model o danej liczbie parametrów a acc MLP min i max to maksymalna dokładność na zbiorze testowym osiągnięta przez model MLP o liczbie parametrów maksymalnie zbliżonej do liczby parametrów KANa dla MLP min i o szerokości warstw o 1 większej w przypadku MLP max.



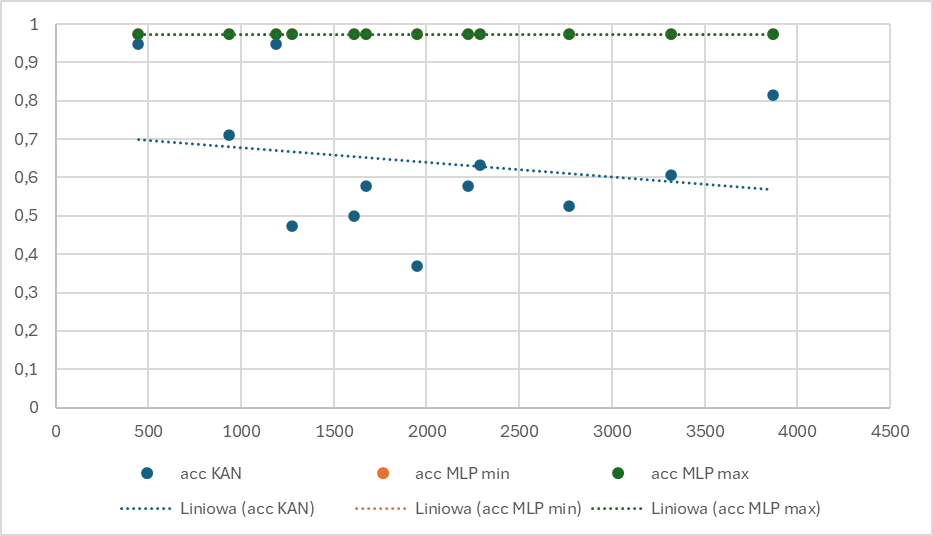
Rysunek . Cancer grid = 5, k = 4



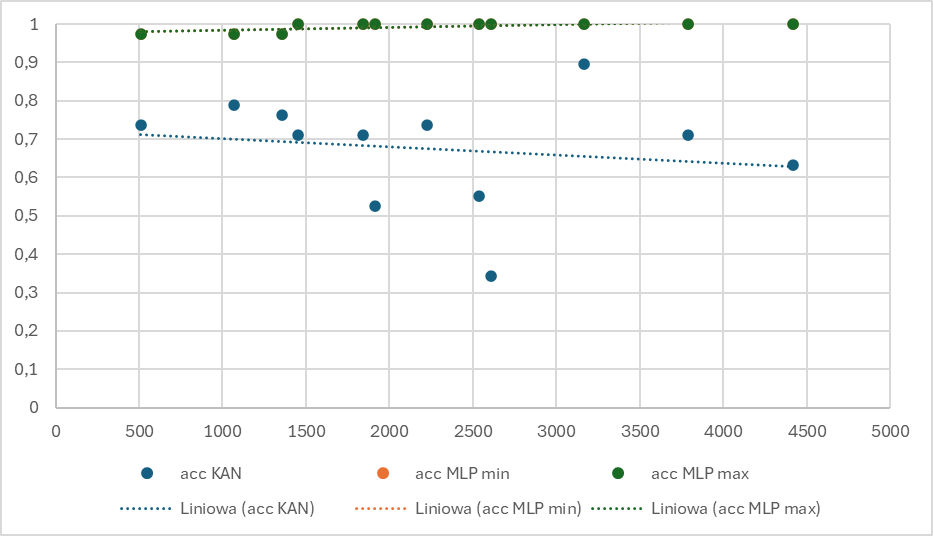
Rysunek . Cancer grid = 5, k = 5



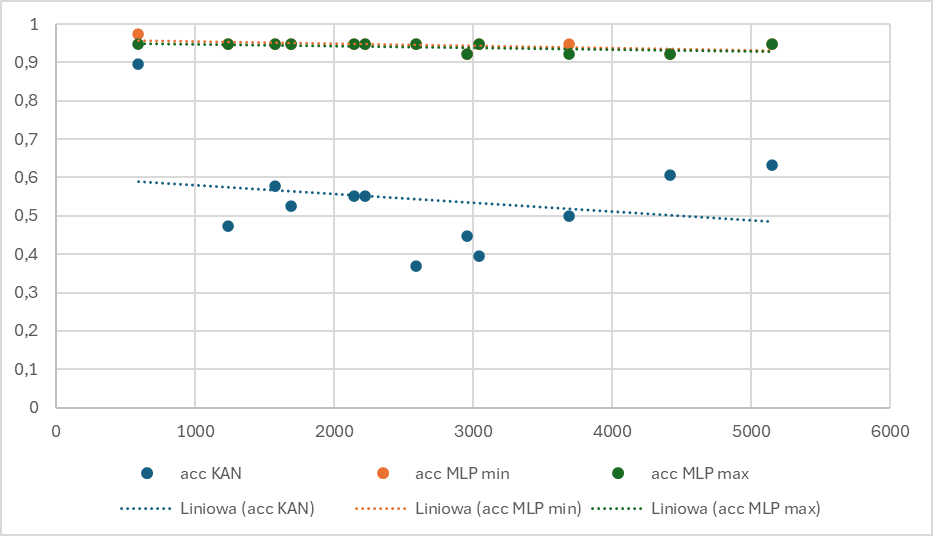
Rysunek . Cancer grid = 9, k = 4



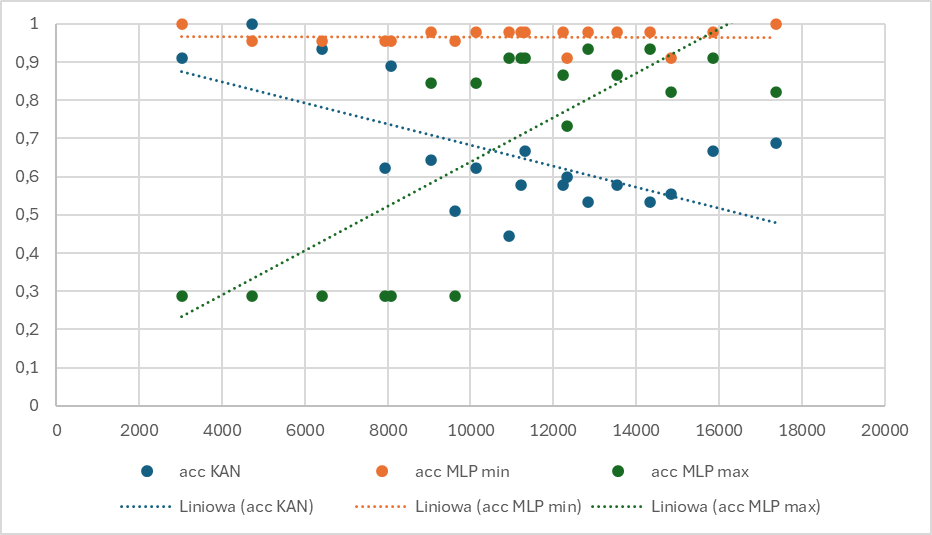
Rysunek . Iris grid = 5, k = 4



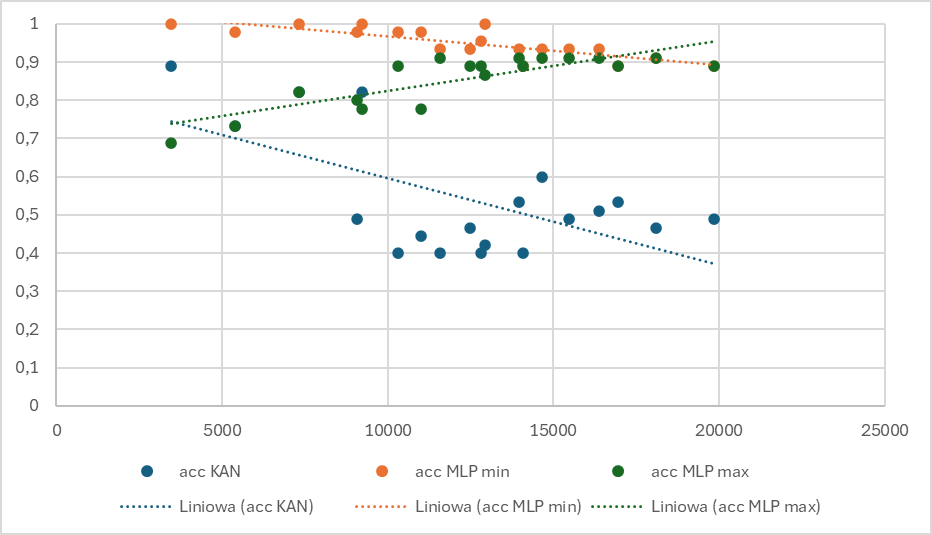
Rysunek . Iris grid = 5, k = 5



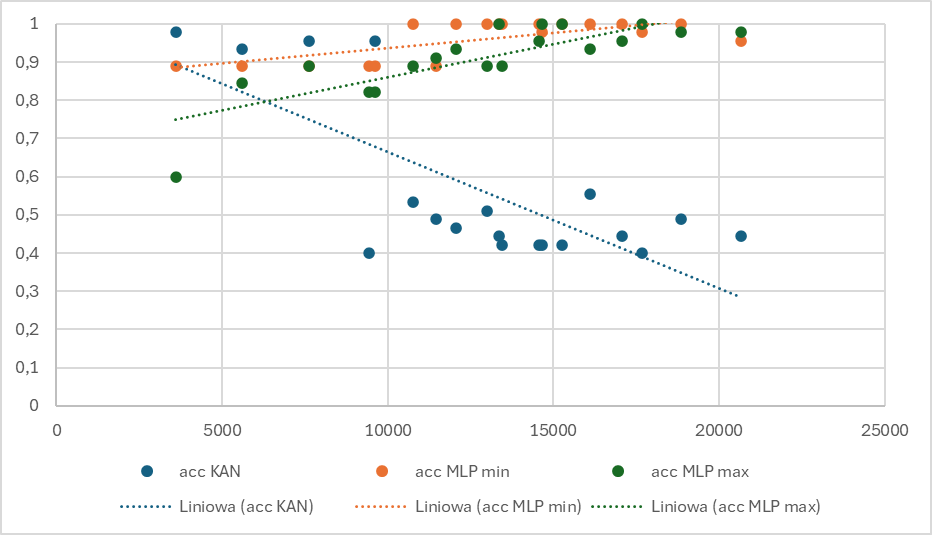
Rysunek . Iris grid = 9, k = 5



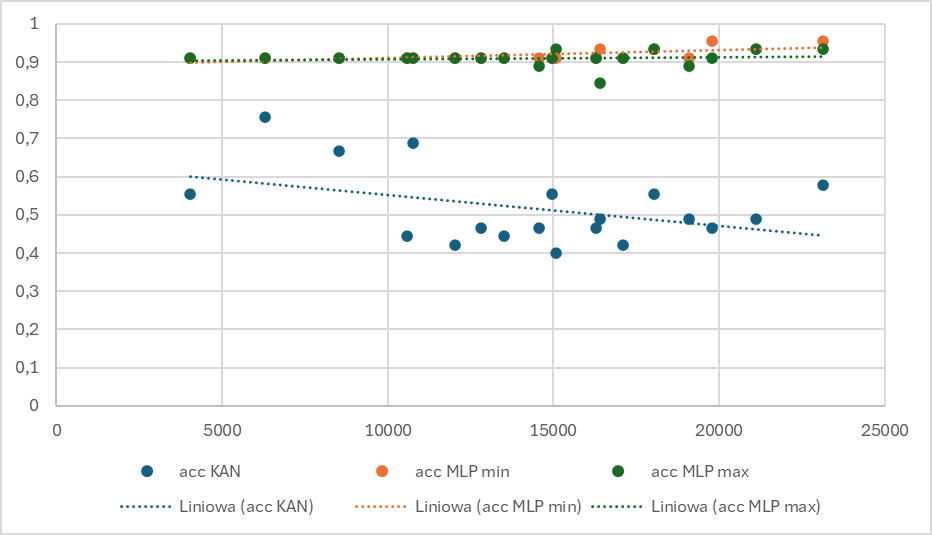
Rysunek . Wine grid = 5, k = 4



Rysunek . Wine grid = 5, k = 5



Rysunek . Wine grid = 9, k = 4



Rysunek . Wine grid = 9, k = 5

Mimo iż dla każdego zestawu danych można utworzyć KAN o dokładności na zbiorze testowym równej 100% ich wyniki nie są tak pewne jak MLP, które z dużą dozą przewidywalności osiągają wysoki poziom dokładności, kiedy to nie wszystkie konfiguracje KANa pozwalają na osiągnięcie tak wysokiego wyniku.

Niektóre z modeli uczonych w ten sposób osiągnęły wyniki niższe niż te uczone w pierwszym eksperymencie. Najprawdopodobniej modele te są przeuczone a optymalny moment ich nauczenia został pominięty w trakcie sprawdzania co uniemożliwiło utworzenie najlepszego możliwego modelu.

Tabela . Tabela przedstawiająca które modele KAN osiągnęły dokładność 100%

|  |  |  |  |
| --- | --- | --- | --- |
| Zestaw danych | Wartość grid | Wartość k | Ilość modeli o dokładności 100% na zbiorze testowym jakie udało się utworzyć z tą parametryzacją |
| Iris | 5 | 2 | 1 |
| Iris | 5 | 3 | 1 |
| Iris | 9 | 2 | 2 |
| Cancer | 5 | 2 | 2 |
| Cancer | 5 | 3 | 3 |
| Wine | 3 | 2 | 11 |
| Wine | 5 | 2 | 2 |
| Wine | 5 | 3 | 7 |
| Wine | 5 | 4 | 1 |
| Wine | 9 | 2 | 3 |

# Wnioski

[todo]

Porównując wyniki pierwszego i ostatniego eksperymentu można zauważyć, że w przeciwieństwie do MLP KANy można łatwo przeuczyć jednocześnie znacznie obniżając ich zdolności do poprawnej klasyfikacji. Z tego powodu KANy należałoby testować co epokę co znacznie opóźnia proces uczenia przy dużych zestawach danych.

# Bibliografia

Adam Pelikant. „Analityczne przetwarzanie danych.” Łódź: Politechnika Łódzka, 2022.

Andrey Kołmogorow. „On the representation of continuous functions of several variables by superpositions of continuous.” *Proceedings of the USSR Academy of Sciences*, 1957: tom 114. strony 953-956.

Arkadiusz Tomczyk. „Uczenie głębokie.” Politechnika Łódzka. *Wykład.* Łódź, 2024.

Cornell Bowers Computer Science. „Lecture 2 Notes.” *cs.cornell.edu.* 2022. https://www.cs.cornell.edu/courses/cs4780/2022sp/notes/LectureNotes02.html (data uzyskania dostępu: Lipiec 24, 2025).

Danny Coomans, Olivier de Vel Stefan Aeberhard. „Comparative analysis of statistical pattern recognition methods in high dimensional settings.” *Pattern Recognition*, 1994: 1065-1077.

Dave Bergmann i Cole Stryker. „What is backpropagation?” *ibm.com.* 02 Lipec 2024. https://www.ibm.com/think/topics/backpropagation (data uzyskania dostępu: Lipiec 30, 2025).

David Hand. „Assessing the Performance of Classification Methods.” *International Statistical Review 80.3*, 2012: 400-414.

David M. W. Powers. „Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation.” *Journal of Machine Learning Technologies*, 2011: 37-63.

Detlof Von Winterfeldt i Ward Edwards. „Decision Trees.” W *Decision Analysis and Behavioral Research*, autor: Detlof Von Winterfeldt i Ward Edwards, 63-65. Cambridge, London, New York, New Rochelle: Cambridge University Press, 1986.

Edgar Anderson. „The irises of the Gaspé Peninsula.” *Bulletin of the American Iris Society*, 1935: 59: 2-5.

Evelyn Fix i Joseph L. Hodges. *Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties.* Randolph Field, Texas: USAF School of Aviation Medicine, 1951.

Federico, and Tomaso Poggio Girosi. „Representation properties of networks: Kolmogorov's theorem is irrelevant.” *Neural Computation 1.4*, 1989: 465-469.

Hornik Kurt, Stinchcombe Maxwell i White, Halbert. „Multilayer feedforward networks are universal approximators.” *Neural Networks Volume 2, Issue 5*, 1989: 359-366.

Houtao Deng, George Runger i Eugene Tuv. „Bias of Importance Measures for Multi-valued Attributes and Solutions.” *Lecture Notes in Computer Science*, 2011.

John A. Swets. „Measuring the Accuracy of Diagnostic Systems.” *Science 240*, 1988: 1285-1293.

Larry Hardesty. „Explained: Neural networks.” *MIT News.* 14 Kwiecień 2017. https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414 (data uzyskania dostępu: Lipiec 20, 2025).

Liu Ziming i inni. „KAN: Kolmogorov-Arnold Networks.” *arXiv.* 30 Kwiecień 2024. https://arxiv.org/abs/2404.19756 (data uzyskania dostępu: Lipiec 29, 2025).

Louis Serano. „Kolmogorov-Arnold Networks (KANs) - What are they and how do they work?” *YouTube.* 03 grudzień 2024. https://www.youtube.com/watch?v=myFtp5zMv8U (data uzyskania dostępu: marzec 16, 2025).

Moheb Costandi. *Neuroplasticity.* The MIT Press, 2016.

R. Leardi, C. Armanino, S. Lanteri M. Forina. „PARVUS: An extendable package of programs for data exploration, classification and correlation.” *Journal of Chemometrics*, 1988: 191-193.

Raul, and Raúl Rojas Rojas. „The backpropagation algorithm.” *Neural networks: a systematic introduction*, 1996: 149-182.

Ronald Fisher. „The use of multiple measurements in taxonomic problems.” *Annals of Human Genetics*, 30 Czerwiec 1936: 179–188.

Rosaida, Mokhairi Makhtar, Mohd Khalid Awang, Mohd Isa Awang and Mohd Nordin Abdul Rahman Rosly. „Analyzing performance of classifiers for medical datasets.” *International journal of engineering and technology*, 15 Luty 2018: 136-138.

Thomas M. Cover i Peter E. Hart. „Nearest neighbor pattern classification.” *IEEE Transactions on Information Theory. 13*, 1967: 21-27.

Tom Fawcett. „An introduction to ROC analysis.” *Pattern Recognition Letters 27(8)*, 2006: 861-874.

Trevor Hastie, R. Tibshirani i J. Friedman. „Tree-Based Methods.” W *Elements of Statistical Learning*, autor: Trevor Hastie, R. Tibshirani i J. Friedman, 266-279. New York: Springer, 2001.

Vahid Mirjalili i Krzysztof Sawka. *Python: uczenie maszynowe.* Gliwice: Helion, 2019.

Wei Wu. „Unsupervised Learning.” *https://na.uni-tuebingen.de/.* 11 Maj 2022. https://na.uni-tuebingen.de/ex/ml\_seminar\_ss2022/Unsupervised\_Learning%20Final.pdf (data uzyskania dostępu: Sierpień 01, 2025).

Weihao Yu, and Xinchao Wang Runpeng Yu. „KAN or MLP: A Fairer Comparison.” *arXiv.* 17 08 2024. https://arxiv.org/abs/2407.16674 (data uzyskania dostępu: 12 07, 2024).

William, Mangasarian,Olvi, Street,Nick, Street,W. Wolberg. „Breast Cancer Wisconsin (Diagnostic).” *UCI Machine Learning Repository.* 1995. https://doi.org/10.24432/C5DW2B.

Ziming Liu. „A Google Algorithms Seminar TechTalk.” *YouTube.* 04 czerwiec 2024. https://www.youtube.com/watch?v=JuPwfQlPUt0 (data uzyskania dostępu: marzec 16, 2025).