

## Streszczenie

Niniejsza rozprawa doktorska stanowi jednotematyczny cykl publikacji naukowych dotyczących zagadnień wykorzystania metod sztucznej inteligencji w celu usprawnienia procesu diagnostycznego w medycynie. Zagadnienia obejmowały problematykę zastosowania metod sztucznej inteligencji do konstrukcji wybranych systemów wspomagania decyzji medycznych. Uwzględniając powyższe, w pracy sformułowano hipotezę, która zakłada, że

*Możliwe jest wykorzystanie różnych metod sztucznej inteligencji do analizy danych medycznych i automatyzacji wybranych procesów diagnostycznych, pozwalające na uzyskanie akceptowalnych wyników z dokładnością i efektywnością nie gorszą niż innych istniejących metod znanych z literatury.*

Hipoteza została uprawdopodobniona przez realizację następujących zadań:

**Zadanie 1. Zastosowanie metod sztucznej inteligencji do klasyfikacji cukrzycy typu 1 na podstawie danych uzyskanych za pomocą nieinwazyjnych pomiarów aktywności fizycznej**

Zadanie zostało zrealizowane przez zastosowanie dziesięciu najpopularniejszych metod sztucznej inteligencji do klasyfikacji cukrzycy typu 1. Wyniki uzyskane przez każdy z wybranych algorytmów zostały zwalidowane za pomocą metryk wydajnościowych, a następnie porównane w celu wyboru optymalnego algorytmu do klasyfikacji.

**Zadanie 2. Opracowanie metody pozwalającej na automatyczne, jednoczesne rozpoznawanie i zliczanie czerwonych i białych krwinek oraz płytek krwi na podstawie zdjęć mikroskopowych z wykorzystaniem głębokich sieci neuronowych**

Zadanie zrealizowano przez wytrenowanie sieci RetinaNet do rozpoznawania i klasyfikacji trzech rodzajów komórek krwi, a następnie manualną ocenę wyników klasyfikacji czerwonych i białych krwinek oraz płytek krwi i obliczenie metryk wydajnościowych dla uzyskanych wyników. Ustalono optymalną wartość progową do zliczania wszystkich typów komórek jednocześnie, a otrzymane wyniki porównano z wynikami innych autorów zajmujących się tematyką liczenia komórek krwi.

**Zadanie 3. Opracowanie aplikacji umożliwiającej automatyzację procesu oceny, składania i identyfikacji sekwencji genomowych uzyskanych za pomocą nowych metod sekwencjonowania przy użyciu narzędzi korzystających z metod uczenia maszynowego**

Zadanie zostało zrealizowane dzięki wykonaniu implementacji aplikacji-serwera NanoForms, umożliwiającego ocenę jakości danych z sekwencjonowania nanoporowego oraz Illumina, a także składanie genomów bakteryjnych metodami de novo i hybrydową. Serwer został skonfigurowany oraz przekazany do niekomercyjnego użytku publicznego, a kod źródłowy aplikacji-serwera został udostępniony w postaci otwartego oprogramowania.

**Zadanie 4. Implementacja w języku Python klasyfikatora opartego na logice rozmytej i programowaniu ekspresji genów, służącego do generowania wysoce interpretowalnych reguł rozmytych**

Zadanie zrealizowano przez wykonanie implementacji klasyfikatora GPR w języku Python, który osiąga bardzo wysokie wyniki pod względem dokładności i pola pod krzywą ROC. W implementacji algorytmu zaproponowano wiele usprawnień pierwotnej wersji, a całość oprogramowania udostępniono na licencji gwarantującej dostęp do kodu źródłowego.

**Zadanie 5. Opracowanie narzędzia pozwalającego na eksperymentalne porównanie wybranych rozmytych algorytmów opartych na regułach do klasyfikacji danych medycznych**

Zadanie zostało zrealizowane przez przygotowanie porównania wyników metryk wydajnościowych osiąganych przez wybrane algorytmy oparte na regułach, przeprowadzonego na zbiorach danych medycznych. Przeprowadzono wiele analiz statystycznych i porównawczych, pozwalających na wybór najlepszych algorytmów regułowych w zastosowaniach medycznych.

**Słowa kluczowe:** sztuczna inteligencja, algorytmy interpretowalne, diagnostyka medyczna, metody sekwencjonowania następnej generacji, głębokie sieci neuronowe

## Abstract

This manuscript-based doctoral thesis addresses using artificial intelligence (AI) methods to improve the medical diagnostic process. The aim of the dissertation was to present the possibility of using artificial intelligence methods to enhance the quality and efficiency of the medical diagnostic process. Based on this premise, the paper formulates a hypothesis, which assumes that

*It is possible to use a variety of artificial intelligence methods to analyze medical data and automate selected diagnostic processes, which will produce interpretable results with accuracy and efficiency not worse than other existing methods known from the literature.*

The hypothesis was substantiated by achieving the following goals:

### **Goal 1. Application of artificial intelligence methods to classify type 1 diabetes based on data obtained by noninvasive physical activity measurements**

This goal was met by applying ten of the most popular artificial intelligence methods to classify type 1 diabetes. The results obtained by each of the selected algorithms were validated using performance metrics and then compared to select the optimal algorithm to solve this problem.

### **Goal 2. Development of a method to automatically and simultaneously identify and count red and white blood cells and platelets from microscopic images using deep machine learning methods**

The goal was achieved by training RetinaNet to recognize and classify three types of blood cells, then manually evaluating the results of classifying red and white blood cells and platelets, and calculating performance metrics for the results obtained. The optimal threshold value was determined to count all cell types simultaneously, and the results obtained were compared with those of other authors working on blood counting.

### **Goal 3. Development of an application to automate the process of evaluating, assembling, and identifying genomic sequences obtained by new sequencing methods using machine learning-based tools**

This goal was achieved by developing the NanoForms application-server, which allows the quality assessment of the ONT and Illumina sequencing data, as well as the

assembly of bacterial genomes using de novo and hybrid methods. The server was configured and made available for noncommercial public use, and the source code of the application-server is open source.

**Goal 4. A Python implementation of a classifier based on fuzzy logic and gene expression programming to generate highly interpretable fuzzy rules**

This goal was met by proposing a novel Python-based implementation of the GPR classifier, which achieves very high results in terms of accuracy and area under the ROC curve. The implementation of the algorithm includes several improvements to the original version, and all the software was made available under a license that guarantees access to the source code.

**Goal 5. Developing a tool to experimentally compare selected fuzzy rule-based algorithms for medical data classification**

This goal was achieved by designing an experimental comparison of the performance metrics achieved by selected rule-based algorithms, carried out on medical data sets. A number of statistical and comparative analyzes were performed, allowing the selection of the best rule-based algorithms for medical applications.

**Keywords:** artificial intelligence, interpretable algorithms, medical diagnostics, next-generation sequencing methods, deep neural networks